ORIGEN DE LA TEORÍA DE LA PROBABILIDAD TEOREMA DE BAYES

Miguel A. Gómez Villegas
Universidad Complutense de Madrid

1.- LAS DISTINTAS ETAPAS DEL CÁLCULO DE PROBABILIDADES Y LA ESTADÍSTICA

Aunque nos vamos a referir al origen de la Teoría de la Probabilidad con un especial hincapié en el teorema de Bayes, es interesante empezar distinguiendo las siguientes etapas en la historia del Cálculo de Probabilidades. De acuerdo con Maistrov (1974), podemos hablar de "La Prehistoria de la Teoría de la Probabilidad", que abarcaría un período de tiempo que se pierde en las nieblas de la antigüedad y que llegaría hasta el siglo XVI, con los trabajos de Cardano, Paccioli, Tartaglia. etc.

En este período se utilizan conceptos rudimentarios de probabilidad, azar y aleatoreidad, que básicamente están relacionados con aspectos adivinatorios y religiosos.

Rabinovitch (1973) recoge ejemplos de usos de mecanismos aleatorios mencionados en la literatura talmúdica y rabínica: para obtener lotes que eran usados en ceremonias religiosas, para repartir las ganancias entre los sacerdotes de los templos y para determinar los turnos de asistencia a los mismos.

Una aproximación muy elemental y temprana a la primera estadística la constituyen los primitivos censos de población, por ejemplo el realizado por Moisés al pueblo Israelita antes de huir de Egipto.

Es importante señalar la sorpresa que produce, vistos los logros alcanzados por los griegos en matemáticas y en la ciencia en general, que no percibieran la simetría presente en los juegos de azar que les permitiera desarrollar una teoría axiomática de la probabilidad, análoga a la que introdujeron en la Geometría. David (1955, 1962) ha señalado como posible justificación, la imperfección existente en los dados de la época. Es conocido que los primeros dados eran los astrágalos de los animales, pequeños huesos de las articulación de las patas, y con los que, hasta la llegada de la era del plástico, todavía era frecuente ver jugar a los muchachos. Se ha citado por Samburski (1956) que un popular juego con astrágalos presenta leyes de regularidad en cuanto a las frecuencias que no fueron señaladas por los griegos.

Una idea mantenida por David (1955) es que el estar ligado estos fenómenos a los aspectos religiosos no facilitó el tratamiento científico de los mismos. Kendall (1956), quizás con mayor razón, señala como motivo, la ausencia de la noción de sucesos aleatorios en las teorías que explicaban el mundo real, y que por lo tanto no necesitaron de una noción de probabilidad que permitiera su tratamiento.

Sambursky (1956) abunda en esta idea y expone que tanto Platón como Aristóteles, limitaron el estudio de los sabios griegos a la regularidad que existía en las matemáticas y en los cielos. A título de ejemplo, Aristóteles clasificaba los sucesos en los que ocurren necesariamente, los que ocurren la mayor cantidad de veces y los sucesos impredecibles o no conocidos. Al utilizar como gradación únicamente el todo o la nada, no necesitaban introducir la probabilidad como medida de la incertidumbre.

Esta clasificación de los griegos fue aceptada, como tantas otras cosas, por los romanos y hubo que esperar a los escolásticos, que tuvieron que reconciliar las doctrinas aristotélicas con las Sagradas Escrituras, para que aparecieran los sucesos aleatorios; así Santo Tomás de Aquino distingue entre la ciencia del conocimiento cierto, de la opinión o del conocimiento probable, y la ciencia de lo accidental o del azar. Una interpretación de la medida de lo contingente, mediante un concepto rudimentario de frecuencia, ha permitido relacionar sus ideas con lo que hoy llamaríamos una aproximación lógica al concepto de probabilidad.

La segunda etapa sería la de "Los Orígenes de la Teoría de la Probabilidad como Ciencia" que abarcaría los siglos XVII y XVIII y que recogería las figuras de Cardano, Galíleo, Pascal, Fermat y Huygens.

Durante este período se introducen los primeros conceptos y teoremas de la probabilidad: el de la adición y la multiplicación, y se obtienen las primeras aproximaciones a la equiprobabilidad y a la esperanza matemática. En este tiempo la Teoría de la Probabilidad encuentra sus aplicaciones primeras a la demografía. Incluiría este periodo lo que tradicionalmente se ha conocido como el comienzo de la teoría; la correspondencia entre Pascal y Fermat sobre el problema de los puntos o como han de repartirse las ganancias de un juego cuando éste ha de ser interrumpido. Continuaría con el tratado de Huygens "Reckoning at Games of Chance" y podemos decir que los fundamentos están consolidados, cuando Pascal escribe su tratado "Arithmetical Triangle" sobre el triángulo aritmético y sus aplicaciones en 1655.

El tercer período "El del Desarrollo de la Teoría de la Probabilidad y la Estadística" iría desde la publicación por Jacob Bernoulli del "Ars Conjectandi" en 1708, hasta la obtención en 1866 de la ley de los grandes números por Chebychef, el originador de la Escuela de Probabilidad de San Petersburgo. Tendría entre otras las figuras de los Bernoulli, De Moivre, Bayes, Laplace, Gauss, Condorcet, Poisson y Quetelet.

En este tiempo se resuelve el problema de la duración de un juego, el problema de la ruina, se llega al concepto de probabilidad condicionada, se introducen la distribución binomial y la binomial negativa y se obtienen la ley de los grandes números de Bernoulli y la distribución de la suma de variables uniformes.

Es en esta época cuando tiene lugar el nacimiento de la Teoría de la Estadística, en la aproximación con que ahora la conocemos, a partir, principalmente, del estudio y desarrollo de los datos que suministran los astrónomos y los demógrafos.

El primer paso en esta línea lo da John Graunt en 1662 con "Natural and Political observations made upon the Bills of Mortality" en el que estudia las probabilidades de morir a distintas edades y saca conclusiones, a partir de los boletines que publicaban las distintas parroquias londinenses, de lo que ocurría con toda la población; dando así una primera y rudimentaria aproximación al concepto de inducción, que es lo que va a diferenciar a la Teoría de la Probabilidad de la Inferencia Estadística.

Los primeros probabilistas

Antes de descender a un mayor detalle sobre la contribución de los distintos autores, debemos hablar de dos principales, aunque no únicos, conceptos de probabilidad: El concepto de probabilidad "objetivo" y el concepto de probabilidad "subjetivo".

El primero es utilizado para describir propiedades de mecanismos aleatorios, tales como juegos de azar, lanzamientos de una moneda etc. En él, las probabilidades están derivadas de consideraciones de simetría o estimadas a partir de frecuencias relativas. Se puede resumir diciendo que están basadas en un juego idealizado con un número finito de resultados igualmente posibles, de tal manera que la probabilidad de un suceso es el cociente entre el número de resultados favorables al mismo y el número total de casos posibles.

El segundo es usado para medir la opinión sobre la ocurrencia de un suceso, parte de suponer unas probabilidades iniciales o a priori que se combinan con cierta información obtenida a partir de datos muestrales y que nos llevan a determinar una nueva probabilidad final o a posteriori. La manera de pasar de una a otra probabilidad, es mediante el concurso del teorema de Bayes, que en su versión continua dice:

$$\pi(\theta / x) = \frac{\pi(\theta) f(x / \theta)}{\int \pi(\theta) f(x / \theta) d\theta}$$

Dos textos en los que pueden verse las distintas aproximaciones al concepto de probabilidad son los de Fine (1973) y Barnett (1982).

Podríamos decir, que de acuerdo con su ámbito de aplicación, la Teoría de la Probabilidad, en el siglo XVII se aplica a los juegos en sentido amplio y a la Jurisprudencia. En el siglo XVIII, en la forma del análisis de datos y de la inferencia inductiva se aplica a la teoría del seguro. En el siglo XIX a la Sociología, la Física, la Biología y la Psicología, para terminar en el XX con aplicaciones a la Agronomía, las encuestas, los contrastes médicos, el deporte y prácticamente todos los campos de las ciencias experimentales.

La primera figura a la que queremos referirnos es a Cardano. Girolamo Cardano (1501-1576) era hijo ilegítimo de un abogado milanés que también explicó matemáticas. Estudió medicina en Padua, siendo rechazado para formar parte del colegio de médicos de Milan, formalmente por su origen ilegítimo, pero posiblemente por su carácter y lengua afilada, en 1536 publicó "sobre las malas prácticas de la medicina en uso común" en la que arremetía contra los abusos que cometían sus colegas.

Practicó como médico rural, en 1532 viajó a Milán como profesor de matemáticas publicando dos libros sobre esta materia. Posteriormente escribió sobre medicina, astronomía, física, juegos, y la inmortalidad del alma. En 1562 se trasladó a Bolonia como profesor de medicina. En 1570, por sus escritos es acusado de hereje y arrestado, prohibiéndosele hablar en público y pasando sus libros al índice. A pesar de todo continuó escribiendo "De Vida Propia Liber" (El libro de mi vida). Muere en Roma a la edad de 75 años.

Como es sabido a él se deben diversos descubrimientos como la suspensión Cardan, la junta Cardan. Para nosotros su contribución más importante es "Liber de Ludo Aleae" (El libro de los Juegos de Azar), publicado póstumamente en 1663, está traducido en el libro de De Mora 1989. La obra comienza con una

Figura 1.
Girolamo Cardano (1501 - 1576)

autobiografía en la que se mezclan consejos moralizantes con anécdotas. En el aparece la expresión.

$$\frac{r^n}{t^{n-r}}$$

como medida de la posibilidad de ganar un juego, con **t** casos igualmente posibles de los cuales r son favorables a un suceso **A** y **t-r** al suceso complementario de **A**, cuando el juego se repite **n** veces.

A título de ejemplo y para poner de manifiesto el carácter de Cardano recogemos del Capítulo Tercero de su libro un extracto de "Con quienes y cuándo conviene más jugar".

> Las condiciones para jugar deben ser: Jugar en raras ocasiones, durante breves espacios de tiempo, en lugar apropiado, con pequeñas apuestas y en ocasiones apropiadas, como un banquete festivo. La persona con la que se juegue debe ser el rey, o un sacerdote de noble virtud, o un pariente por la sangre o por el matrimonio. Jugar con jugadores profesionales es lo más vergonzoso y (como he dicho) es peligroso, el lugar más decente es la casa propia, o la de algún amigo, donde no pueda entrar el público. Los juristas juegan en desventaja, lo mismo que los médicos, pues por una parte, parecería que disponen de demasiados ocios y por otra si ganan, son vistos como jugadores y si son vencidos se considera que sabrán de su arte tanto como del juego. El mismo juicio se les aplica si quieren practicar la música.

La siguiente contribución en la que queremos fijarnos es debida a Galileo (1564-1642). El célebre pisano en "Sopra le Scoperte dei Dadi" (Un Descubrimiento Concerniente a los Dados) da respuesta a la observación de que si uno juega con tres dados, la suma 9 y la suma 10 se pueden obtener de "seis" formas básicas diferentes:

(6,2,1) (5,3,1) (4,3,2) (3,3,3) (4,1,4) (5,2,2)
(6,3,1) (6,2,2) (5,3,2) (5,4,1) (4,4,2) (4,3,3)

y sin embargo si uno lanza un número alto de veces los tres dados, es más ventajoso apostar a 10 que a 9. La explicación de Galileo es del todo correcta, el 9 puede obtenerse mediante 25 descomposiciones igualmente simétricas, mientras que el 10 se obtiene mediante 27.

Como ya hemos dicho es admitido con frecuencia que, el comienzo de la

Teoría de la Probabilidad lo constituyen las cartas intercambiadas de julio a octubre de 1654 entre Pascal y Fermat, se conservan tres de Pascal a Fermat, y cuatro de Fermat a Pascal, aunque hubo más.

De esta manera, podríamos decir que el primer paso en la Historia de la Probabilidad se debe a los matemáticos italianos (básicamente el concepto de equiprobabilidad y variantes), mientras que el segundo lo dan los matemáticos franceses.

El origen de la correspondencia, es el problema de la interrupción de un juego o problema de los puntos; dos jugadores están de acuerdo en jugar hasta que uno gane s veces, si el juego se interrumpe cuando uno de los jugadores ha ganado $a<s$ partidas y el otro $b<s$, la cuestión es cómo deberían repartirse las ganancias.

Por cierto que existe una divertida leyenda que recogemos del libro ya citado de Maistrov (1974) pag. 70: Un caballero francés, el caballero De Méré, era un ardiente jugador que quería hacerse rico con el juego, para lo cual diseñaba complicadas reglas que le permitieran ganar dinero. Propuso arrojar un dado 4 veces consecutivas y apostar por la aparición de al menos un 6. De Méré suponía que mediante este procedimiento ganaría con mas frecuencia que perdería y propuso a Pascal calcular la probabilidad de ganancia.

Observamos que la probabilidad pedida es:

$$1 - P\{\text{no sacar 6 en 4 tiradas}\} = 1 - \left(\frac{5}{6}\right)^4 = 0'518$$

siendo notable el que experimentalmente detectaran sucesos con probabilidad que difiere muy poco de 0'5.

Por lo visto, los contrincantes de De Méré acabaron descubriendo que su procedimiento era ganador, por lo que este cambió de enunciado y propuso arrojar dos dados 24 veces y apostar a que dos cincos saldrían por lo menos una vez. En lenguaje actual se trata de calcular, para una variable **X** con distribución Binomial de parámetros n = 24 y $\theta = \frac{1}{36}$, la probabilidad del suceso $X > 1$ que viene dada por

$$P\{X \geq 1\} = 1 - P\{X = 0\} = 1 - \left(\frac{35}{36}\right)^{24} = 0'491$$

Esta vez el caballero De Méré se equivocó y la apuesta era desfavorable para él. Anécdotas aparte, De Méré cuyo verdadero nombre era Antoine Gambaud, tuvo muy buena reputación entre los principales matemáticos de su tiempo, mantuvo correspondencia con Huygens, Pascal y Fermat llegando a corregir algún error cometido por Pascal en relación con el problema ya citado de la interrupción de un juego.

La siguiente figura, a la que vamos a dedicar nuestra atención es Pierre Fermat (1601-1665). Nació en Beaumont-de-Lomagne en la Gascuña francesa de una familia de ricos comerciantes, estudiando leyes en las universidades de Toulouse y Orleans y matemáticas en la de Burdeos. Vivió como abogado y jurista teniendo las matemáticas como entretenimiento.

Figura 2.
Pierre Fermat (1601 - 1665)

Fue parlamentario y un gran erudito. Sabía griego, latín, inglés y español. Como es conocido, su fama descansa mas bien en sus estudios sobre teoría de números, con su famoso teorema tan de actualidad en este momento.

El otro protagonista del célebre epistolario fue Blas Pascal (1623-1662). Nació en Clermont-Ferrant, su madre murió cuando él tenía 3 años, encargándose su padre de su educación. Era un niño prodigio, a los 11 años compuso un tratado sobre sonidos, a los 17 escribió su obra capital "Essai pour les Coniques" en el que sienta las bases de la geometría proyectiva. En 1646 se convierte al jansenismo[1] y alrededor de 1654 participa en las reuniones del padre Mersenne que serían el germen de la Academia Francesa. En esta época informa a sus amigos acerca de su manuscrito sobre el triangulo aritmético; este trabajo fue publicado postumamente en 1655 bajo el titulo "Traité du Triangle Arithmétique". En 1658 se retira a la abadía de Port-Royal desde donde escribe sus "Cartas Provinciales" y su "Apología de la Religión Cristiana" no volviendose ya a dedicar a las matemáticas.

Figura 3.
Blas Pascal (1623 - 1662)

Desde el punto de vista del Cálculo de Probabilidades, que aquí nos ocupa, su contribución mas importante es al epistolario. Está dedicado a resolver el problema de los puntos y en él cada uno de los autores lo hace mediante un procedimiento diferente; así Fermat utiliza la técnica de lo que hoy llamaríamos recorridos aleatorios, mientras que Pascal lo resuelve por recurrencia, poniéndolo en

[1] Los jansenistas defendían la predestinación a salvarse o condenarse del hombre, estuvieron enfrentados con los jesuitas que defendían las tesis contrarias.

relación con su triángulo aritmético definido mediante la siguiente ecuación recurrente:

$$t_{m,n} = t_{m-1,n} + t_{m,n-1} \qquad m = n = 0,1...,$$

con las condiciones iniciales

$$t_{m,-1} = t_{-1,n} = 0 \qquad t_{0,0} = 1 \qquad m = n = 1, 2, ...,$$

A través de la correspondencia puede verse que es Fermat quien ocupa una posición preponderante. A continuación, y como ejemplo, tomamos este pequeño resumen de la carta enviada por Pascal a Fermat el miercoles 20 de diciembre de 1654.

"....El método de Vd. es muy seguro y es el primero que me vino a la mente en esta investigación; pero como el trabajo de las combinaciones es excesivo, he encontrado un resumen de las mismas y propiamente otro método mucho más corto y claro, que desearía poderle exponer aquí en pocas palabras: pues en lo sucesivo querría abrirle mi corazón si es posible, tanto placer siento de ver nuestro acuerdo, veo bien que la verdad es la misma en Toulouse que en París"

Pascal se dio cuenta de la relación de su famoso triángulo con la expansión binomial y con el número de combinaciones, él es quien introdujo el término "combinación".

Figura 4.
Jacob Bernoulli (1654 - 1705)

Digamos también que Pascal fue el primero en obtener las sumas finitas de las potencias enteras de las progresiones aritméticas en forma recursiva, su expresión explícita fue descubierta por Jacob Bernoulli en 1713.

Por último, con referencia a Pascal, queremos decir también que en 1670 escribió "Infini-Rien" cuatro páginas manuscritas, no ordenadas con anotaciones al margen y que han provocado grandes discusiones entre los expertos, siendo conocidas también como "Le Pari de Pascal" (La Apuesta) y que constituyen un antecedente de la moderna Teoría de la Decisión. En ellas trata de demostrar, aplicando la Teoría de la Probabilidad, la conveniencia de "llevar una vida piadosa" tanto si uno cree en la existencia de

Dios como si no es así.

La probabilidad inversa

El que puso los primeros cimientos fué el ya citado Jacob Bernoulli (1654-1705). La familia Bernoulli se trasladó a Basilea en 1570 huyendo de la persecución de los herejes protagonizada por las campañas del Duque de Alba.

Los Bernoulli fueron, sin duda alguna, la familia de más renombre en la historia de las matemáticas; hasta doce Bernoulli han contribuido a la historia de las matemáticas y de la física y de ellos cinco al menos han escrito sobre probabilidad. Esto le ha hecho decir a Stigler (1986), que simplemente por una cuestión de azar, era inevitable que alguno de ellos tenía que ser el padre de la cuantificación de la incertidumbre.

Actualmente hay un Daniel Bernoulli profesor de Geología en la Universidad politécnica de Zurich. Para mayor abundamiento a alguno de los Bernoulli se le conoce con varios nombres.

Como decíamos, Jacob nació en Basilea en 1654, estudio filosofía y teología graduándose en esta última en 1676. Viajó a través de Suiza, Francia, Holanda e Inglaterra contactando con los más prominentes matemáticos de la época, como Leibniz, Euler, De Moivre, Montmort, etc. A su vuelta a Basilea en 1682, dio lecciones sobre Física; en 1687 fue nombrado catedrático de Matemáticas, explicando el naciente análisis infinitesimal. Dió clases a su hermano John y a su sobrino Nicolás. Se dedicó también a ecuaciones diferenciales estudiando la cicloide, la catenaria, la curva de la vela, la lemniscata, la elástica etc. se dedicó también al estudio de la Mecánica, las series infinitas, y es en el estudio de estas donde llegó a descubrir los números de Bernoulli.

Posteriormente, en 1713 aparece la obra capital, en relación al tema que nos ocupa, "Ars Conjectandi" (El Arte de la Conjetura). La relación entre los Bernoulli no fue del todo fluida, como en toda familia, esto explica el que siendo su hermano John el sucesor en la cátedra de Basilea de Jacob, hubiera que esperar a que un sobrino de ambos, Nicolás (1687-1759) se encargara de la publicación de la obra.

El libro está dividido en cuatro partes: La primera es el tratado "De Ratiociniis in Ludo Aleae" de Huygens, con anotaciones de Jacob, que ocupa las primeras 71 páginas; la segunda contiene la doctrina de permutaciones y combinaciones, hasta la página 137; la tercera estudia el tratamiento probabilístico de varios juegos de azar, hasta la página 209, para terminar en la cuarta parte, páginas 250 a 239, con la aplicación de las ideas anteriores a temas civiles, morales y económicos.

En las tres primeras partes, establece la regla de la multiplicación de la probabilidad, la distribución binomial, la probabilidad de ganar un juego como


Figura 5. Pierre Simon de Laplace

suma de una serie infinita, y resuelve los mismos mediante la combinación de estas reglas y el criterio de la esperanza matemática.
En la cuarta parte, introduce el "grado de certeza moral" aproximándose a Fisher y a la probabilidad subjetiva, cuando dice:

"Hay certeza moral cuando la probabilidad es casi igual a la certeza total.... así, si una cosa es considerada moralmente cierta cuando tiene 0'999 de certeza, otra será moralmente imposible cuanto tenga 0'001 de certeza"

Como se ve, anticipa valores para los niveles de significación, que posteriormente Fisher fijaría en 0'1, 0'05 y 0'01.

Precisamente, en relación con este problema, introduce la ley de los grandes números que lleva su nombre. Bernoulli era consciente del hecho, de que la incertidumbre disminuye, a medida que aumenta el número de observaciones y quiso demostrar el principio, de que la "certidumbre moral" acerca de una proporción podría aproximarse a la "certeza total", aumentando el número de observaciones.

En notación actual si X_i $i = 1,....,n$ tiene distribución de Bernoulli de parámetro θ, es decir

$$P\{X_i = 1/\theta\} = \theta \qquad\qquad P\{X_i = 0/\theta\} = 1 - \theta$$

para cualquier $\in$ positivo arbitrario, se tiene que

$$\lim_{n \to \infty} P\left\{\left|\frac{\sum x_i}{n} - \theta\right| < \in / \theta\right\} = 1$$

Es decir si no conocemos el valor de θ piénsese en la probabilidad con que una moneda sale cara, podemos tirarla un número grande de veces y estar "prácticamente seguros" de que $\sum \frac{x_i}{n}$ se parecerá a θ, entendiendo por esto, que para cualquier **c** todo lo grande que queramos, **c** =100, 1000....se cumple

$$P\left\{\left|\frac{\sum x_i}{n} - \theta\right| \leq \in\right\} > c.P\left\{\left|\frac{\sum x_i}{n} - \theta\right| > \in\right\}$$

El libro de Bernoulli, termina bruscamente con una çota inferior para N=25.500 con c=1000. Se ha especulado, sobre si esto le pareció un valor muy grande, y por eso no publicó su tratado en vida.

La siguiente figura que queremos tratar en relación con la probabilidad inversa es la de Bayes. Thomas Bayes nació presumiblemente en 1701 en Londres, la fecha exacta no se conoce, era hijo de Joshua Bayes, uno de los seis primeros ministros no conformistas ordenados públicamente en Inglaterra. Fue educado por prefectores y Barnard (1956) (incluido en Pearson y Kendall (1770)) especula con que quizás le diera clase de matemáticas De Moivre, la más alta cota en la Teoría de la Probabilidad de su tiempo. Fue ministro protestante en Tunbridge Wells, cerca de Londres. De él solo se conserva un discurso sobre tema religioso y un escrito que publicó bajo el seudónimo de John Noon titulado "An introduction to the Doctrine of Fluxions, and a Defense of the Mathematicians against the objections of the Analyst" que es una defensa del método de las diferenciales frente a los ataques que le hace Berkeley, el autor del "Analyst".

Presumiblemente, este artículo le valió el ingreso en la Royal Society, la Academia de Ciencias Inglesa. Se sabe que se retiró de su ministerio en 1752 y que murió en Tunbridge Wells el 17 de Abril de 1761, estando enterrado en Bunhill Fields, un cementerio londinense donde también descansan Richard Price, la persona por mediación de la cual conocemos la contribución a la probabilidad de Bayes, Daniel Defoe y otros hombres ilustres.

Como decíamos la fama de Bayes se debe a que Richard Price, amigo suyo y matemático de prestigio, publicó "An Essay toward solving a problem in the doctrine of chances", el ensayo fue leído ante la Royal Society el 23 de diciembre de 1763 y publicado en las "Philosophical Transactions" de la misma en 1764, dos años y medio después del fallecimiento de Bayes.

En él, Price dice que encontró el Ensayo entre los papeles de su amigo muerto Mr. Bayes. El problema que se quiere resolver está allí claramente expuesto:

> Dado el número de veces [r] que un suceso ha ocurrido en un número de repeticiones [n], calcular la probabilidad de que la probabilidad de que ocurra en una repetición esté entre dos valores conocidos [b y f] [1]

En notación actual, si llamamos X al número de veces que un suceso ocurre en **n** pruebas y es θ la probabilidad de que se presente en cada prueba se trata de calcular $P\{b < \theta < f / X = r\}$

[1] Los corchetes son añadidos.

La expresión que Bayes obtuvo (Sección II, Proposición 9 del ensayo) fue

$$\mathrm{P}\{\mathrm{b} < \theta < f / X = r\} = \frac{\int_b^f \binom{n}{r} \theta^r (1-\theta)^{n-r} d\theta}{\int_0^1 \binom{n}{r} \theta^r (1-\theta)^{n-r} d\theta} \qquad [4.1]$$

Hoy diríamos que esta probabilidad a posteriori, viene dada por la integral de la función de densidad "a posteriori "entre b y f $\pi(\theta/x = r)$ mediante

$$\pi(\theta / X = r) = \frac{\pi(\theta) P\{X = r / \theta\}}{\int_0^1 \pi(\theta) P\{X = r / \theta\} d\theta} \qquad [4.2]$$

donde $\pi(\theta) = I_{(0,1)}(\theta)$ es la densidad uniforme en el intervalo [0.1] y $\mathrm{P}\{X = r / \theta\} = \binom{n}{r} \theta^r (1-\theta)^{n-r}$ es la probabilidad de obtener **r** éxitos en **n** repeticiones de experimentos de Bernoulli. Es claro entonces que $P\{b < \theta < f / X = r\} = \int_b^f \pi(\theta / X = r) d\theta$ lo que nos lleva a la expresión [4.1].

Hacemos notar que Bayes trata de aprender de θ, la probabilidad de éxito en una repetición, y para ello se basa en que sabe que en **n** repeticiones se han producido **r** éxitos; ha "invertido" la manera de razonar de Bernoulli. Este podía calcular la probabilidad de que se presenten **r** éxitos en **n** repeticiones, si uno conoce cual es el valor de θ mediante la expresión

$$P\{X = r / \theta\} = \binom{n}{r} \theta^6 (1-\theta)^{n-6}$$

La novedad de Bayes, consiste en llegar a [4.1] y en utilizar como densidad a priori

$$\pi(\theta) = I_{(0,1)}(\theta)$$

El poner en este caso, para la probabilidad de éxito, la densidad uniforme, parece bastante plausible; esta densidad da la misma probabilidad a dos intervalos de la misma amplitud en el intervalo [0,1], donde forzosamente ha de moverse θ.

La polémica que ha venido después, y la crítica a este procedimiento de comportarse, el paradigma bayesiano, radica en aplicar [4.2] a contextos generales, ¿Cuál sería la determinación adecuada de la función π(θ) en otras situaciones?

El ensayo de Bayes, ha sido editado varias veces en lo que va de siglo; una transcripción bastante ajustada puede verse en el libro de Press (1989).

Como se ve Bayes no escribió, que sepamos, el teorema que lleva su nombre y que constituye la versión discreta de la expresión [4.2]. Para ello tuvimos que esperar a la siguiente figura que vamos a tratar.

Pierre Simon de Laplace (1749-1827), nació en Beaumont-en-Auge en la región francesa de Calvados. En 1765 ingresó en la facultad de Artes de Caen. Inicialmente trabajó sobre máximos y mínimos y calculo integral. En 1772 a la edad de 23 años ingresó en la Academia Francesa.

La primera obra que nos interesa de él es "Memoire sur la Probabilité des Causes par les Évènements" publicada en 1774 y a la que nos referiremos posteriormente. En los años siguientes se dedica al estudio de temas de Astronomía, siendo miembro de la Comisión de Pesos y Medidas.

Esta época coincide con la revolución francesa; así la toma de la Bastilla se produce en 1789, la matanza del Campo de Marte es en 1791 y en 1793 es guillotinado Luis XVI. Durante estos años Laplace se traslada a Melun dedicándose a finalizar sus trabajos sobre la "Mecánica Celeste". En 1784, Laplace vuelve a Paris, siendo profesor de la Escuela Normal; en 1795 es abolida la Academia de Ciencias Francesa siendo sustituida por el Instituto Nacional. Laplace era su presidente y como tal en 1796 presenta su informe a Napoleón sobre el progreso de la ciencia. En 1799 Napoleón disuelve el Consejo de los 500 y el Directorio, iniciando el cambio que le llevará a la coronación como emperador en 1804. Laplace en 1802 publica su tercer volumen sobre la Mecánica Celeste y a partir de este momento se dedica con mayor intensidad a sus estudios relacionados con la probabilidad; así en 1820 presenta su "Memoire sur les Approximations des Formules que sont Fonctions de très Grand Nombres" donde obtiene la aproximación de la distribución Binomial a la Normal y en 1812, a la edad de 63 años, escribe su "Teoria Analítica de Probabilidades" de la que se hicieron tres ediciones en vida de Laplace (1812, 1814 y 1820). En ella está contenida lo que hoy en día se conoce como "definición clásica" de la probabilidad: la probabilidad de un suceso **A** es igual al cociente del número de resultados del experimento que son favorables al suceso **A** entre el número de todos los resultados posibles del experimento.

Laplace ha sido criticado por su excesiva adaptabilidad a los cambios de su época, la edición de 1812 de su "Teoría Analítica de las Probabilidades" está dedicada a "Napoleón el Grande" mientras que en la edición de 1814 suprimió esta dedicatoria y escribió "la caida de los imperios que aspiran al dominio universal puede predecirse con probabilidad muy alta por cualquiera versado en el cálculo del azar". Debemos admitir que en esta época era costumbre dedicar los libros a personajes notables y que los años en que vivió Laplace no estaban exentos de riesgo para cualquier personaje público. El hecho es, que en 1814 cuando

Luis XVIII sube al trono, Laplace obtiene el título de marqués. En 1827 a la edad de 78 años muere en París el que alcanzó una de las mas altas cotas en la historia del Cálculo de Probabilidades.

Con referencia a los trabajos de Laplace podríamos clasificarlos en tres etapas: una primera entre 1770 y 1760, en la que estudia ecuaciones en diferencias y series, probabilidad, obtiene el teorema de Bayes en el caso continuo, y trata problemas de diferencias de proporciones entre los nacimientos de Londres y París e introduce una función de pérdida para estimación. Una segunda, entre 1780 y 1805, en la que se dedica a aplicar las matemáticas a la física del sistema solar y que culmina en su "Mecánica Celeste" y una tercera, entre 1805 y 1827, en la que obtiene el teorema central del límite y la justificación bayesiana del método de los mínimos cuadrados.

Con referencia a la contribución de Laplace a la solución del problema de la probabilidad inversa, vamos a recoger únicamente su obtención de la fórmula de Bayes en el caso discreto. El teorema aparece por primera vez en su "Mémoire sur la Probabilité des Causes par les Évènements" (1774) sin embargo nosotros vamos a tomarlo de su "Théorie Analytique" (1812). En su tercer Principio dice:

> "Si un suceso observado puede resultar de **n** causas diferentes; sus probabilidades son respectivamente, como las probabilidades del suceso, deducido de su existencia: y la probabilidad de cada una de ellas es una fracción cuyo numerador es la probabilidad del suceso, en la hipótesis de la existencia de la causa y cuyo denominador es la suma de las probabilidades posibles, relativas a todas las causas".

Es decir si llamamos al suceso E y C_i i=1,...,n a las distintas causas

$$P(C_i / E) = \frac{P(E / C_i)}{\sum_1^n P(E / C_i)} \qquad [4.3]$$

expresión que coincide con la formula de Bayes cuando todas las causas son equiprobables $P(C_i)=\frac{1}{n}$ i =1,...,n y que en general se expresa como:

$$P(C_i / E) = \frac{P(C_i)P(E / C_i)}{\sum_1^n P(C_i)P(E / C_i)} \qquad [4.4]$$

La demostración obtenida por Laplace no difiere en nada de la que aparece en los libros de texto actuales; se basa en la definición de probabilidad condicionada y en el teorema de la probabilidad total.

5.- Comentarios

Como puede verse, la expresión [4.2], Teorema de Bayes en el caso continuo, coincide o, mas exactamente, tiene la misma expresión formal que la [4.4], Teorema de Bayes en el caso discreto, con las modificaciones obvias: sustitución del suceso X=r por el suceso **E** y sustitución de los distintos valores de θ por la causa de ocurrencia del suceso **E**.

Un segundo comentario es, que el Teorema de Bayes sale automáticamente de los axiomas del Cálculo de Probabilidad y que como teorema es por tanto indiscutible. Si por concretar ideas nos centramos solo en la versión discreta, es matemáticamente correcto, decir que si las probabilidades de los sucesos C_i son $P(C_i)$ i=1,...,n entonces si nosotros sabemos que ha ocurrido el suceso **E** las probabilidades $P(C_i/E)$, de que ocurra C_i condicionadas a que ha ocurrido **E**, vienen dadas por la expresión [4.4]. Esta expresión permite ir actualizando las probabilidades pasando de las probabilidades "a priori" o "iniciales" $P(C_i)$ a las probabilidades"a posteriori" o "finales" $P(C_i/E)$.

Analicemos ahora el siguiente ejemplo, éste o similar puede verse en los libros de texto actuales. Supongamos dos urnas, la primera contiene 2 bolas blancas y 3 bolas negras y la segunda 3 bolas blancas y 2 negras. Se extrae una bola al azar sin saber de que urna se trata y resulta ser negra, si llamamos **E** a este suceso, ¿Cuál es la probabilidad de que haya sido sacada de la primera urna?.

Es claro que las dos "causas" que han originado el suceso **E** son C_i ={la bola negra ha sido extraida de la urna i} con i=1,2. Si ahora suponemos que $P(C_i)$ =0`5 la utilización de [4.4] nos lleva a

$$P\{C_i/E\} = 0'6 \qquad\qquad P\{C_2/E\} = 0'4$$

Resultado incontrovertible, es consecuencia del Teorema de Bayes, e intuitivo, puesto que en la primera urna había mas bolas negras que en la segunda y por lo tanto cabe esperar una mayor probabilidad de que la bola haya sido extraida de la primera urna que de la segunda.

El tercer comentario es que así como en este sencillo ejemplo la asignación $P(C_i = 0'5)$ i =1,2 parece plausible, puede ser dudoso que en todas las situaciones en las que hay incertidumbre, uno admita que debe asignar a los sucesos sobre los que esté incierto **A** una probabilidad P(A/H), que dependería del estado de información en que se encuentre cada uno y que cuando se presente un suceso en forma de experimentación x deba pasar a P(A/x,H) mediante lo análogo a [4.2] o [4.4] dependiendo de si uno está en un contexto continuo o discreto. Esto es, de forma muy simplificada, lo que supone la aproximación bayesiana a la inferencia.

6.- Epílogo

En los últimos años, parece haber resurgido el interés por la Historia de la Ciencia en general, lo que tambien ha ocurrido en el campo de la Historia del Calculo de Probabilidades y de la Estadística, como lo demuestra el creciente número de libros y monografías publicadas recientemente.

Entre los libros clásicos dedicados a este tema, debemos citar el de Todhunter (1865), que junto con los de David (1962) y Maistrov (1974) constituyen una bibliografía básica sobre el mismo. Tambien debemos citar la recopilación de artículos de distintas revistas, publicados en forma de libro, en dos volumenes: el de Pearson y Kendall (1970); y el de Kendall y Plackett (1977) y la recopilación de conferencias impartidas por Karl Pearson en el University College entre 1923 y 1933 y editadas por su hijo Egon Pearson en 1978.

Entre los más recientes citaremos el de Stigler (1986), escrito por un estadístico con un gran sentido del humor, en castellano el de De Mora (1989), que reune traducciones comentadas de varios autores clásicos y por último otra vez en inglés, el de Hald (1990).

Naturalmente, la elección de temas y libros elegidos por el autor es subjetiva. Dos aproximaciones recientes y distintas, en las que se recogen la Historia del Cálculo de Probabilidades y la Estadística, son las debidas a Fienberg (1992) y Girón (1994), que contienen autores y bibliografías alternativas.

7.- Referencias

Barnett, V (1982). Comparative Statistical Inference. Wiley: Cluchester.

David, F. N. (1955). Dicing and Gaming (a note on the history of probability). Biometrika, 42, 1-15.

David. F. N. (1962). Games, Gods and Gambling. Charles Griffin: London.

De Mora, M. (1989). Los Inicios de la Teoria de la Probabilidad. Ed. Universidad del Pais Vasco: Vizcaya.

Fienberg, S. E. (1992). The history of statistics: a review essay. Statistical Science. 7, 208-225.

Fine, T. L. (1973). Theories of Probability. Academic Press: New York.

Girón, F. J. (1994). Historia del cálculo de probabilidades: de Pascal a Laplace. En Historia de la Ciencia Estadistica. Real Acad. Cien. Exac Fis. y Nat. Madrid.

Hald, A. (1990). A History of Probability and Statistics and Their Applications Before 1750- Wiley: New York.

Kendall, M. Q. (1956). The beginnings of a probability calculus. Biometrika, 43, 1-14

Kendall, M. and Plackett, R.L. (1977). Studies in the History of Statistics and Probability. Vol II. Charles Griffin: London.

Maistrov, L.E. (1774). Probability Theory: A Historial Sketch. Academic Press: New York.

Pearson, K (1978). The History of Statistics in the 17 th and 18 th centuries, Against the Changing Background of Intellectual, Scientific and Religious Thought. E.S. Pearson ed. Based en lectures delivered 1921-33. McMillan: New York.

Pearson, E. S. and Kendall, M.G. (1970). Studies in the History of Statistics and Probability. Charles Griffin: London.

Press, J. (1989). Bayesian Statistics: Principles, Models and Applications. Wiley: New York.

Stigler, S. M. (1986). The History of Statistics: The Measurement of Uncertainty before 1900. Harvard University Press.

Todhunter, I. (1865). History of the Mathematical Theory of Probability. Cambridge University Press: London. Reeditado por Chelsea. New York, 1949-1962

Rabinovitch, N. L. (1973). Probability and Statistical Inference on Ancient and Medieval Jewish Literature. University of Toronto Press: Toronto.

Sambursky, S. (1956). On the posible and probable in Ancient Greece. Osiris, 12, 35-48.