

Historia
de la
Teoría de la
Probabilidad

Pablo Salinero Ruiz
Historia de las Matemáticas

ÍNDICE

LA PROBABILIDAD Y LOS JUEGOS DE AZAR	1
Precursores	1
Girolamo Cardano y Niccolo Tartaglia	1
Galileo Galilei	2
Blaise Pascal y Pierre de Fermat	2
Christiaan Huygens	3
EL DESARROLLO DE LA TEORÍA DE LA PROBABILIDAD.....	4
Primeras investigaciones en demografía	4
Primera definiciones de la probabilidad	4
Teoremas básicos de la Teoría de la Probabilidad.....	5
Los teoremas del límite	6
El problema de la ‘Ruina del Jugador’	8
La paradoja de San Petersburgo	8
LA PROBABILIDAD MODERNA	9
La Teoría de la medida de errores	9
Formación del concepto de ‘variable aleatoria’	9
La Ley de los Grandes Números	10
El Teorema Central del Límite	11
La interpretación subjetiva de la probabilidad.....	12
La axiomatización de la probabilidad.....	12
APLICACIONES DE LA TEORÍA DE LA PROBABILIDAD	13
La probabilidad geométrica.....	13
Control de calidad	14
Procesos Estocásticos	14
Martingalas.....	15
Método de Montecarlo	15

1.- LA PROBABILIDAD Y LOS JUEGOS DE AZAR

PRECURSORES

La probabilidad matemática tiene sus orígenes en los juegos de azar, principalmente los juegos con dados y cartas, muy populares desde tiempos antiguos. Los primeros estudios “científicos” sobre fenómenos aleatorios se centraban en dos problemas:

1. Contabilizar el número de posibles resultados de lanzar un dado varias veces.
2. Distribuir las ganancias entre jugadores cuando el juego se interrumpía antes de finalizar, conocido como el ‘problema del reparto de apuestas’.

Una respuesta al primer problema se puede encontrar en el poema *De Vetula*, de Richard de Fournival (1200–1250), donde afirma correctamente que si se lanzan tres dados hay 216 combinaciones posibles y calcula correctamente los diferentes valores para la suma de los tres dados. Aunque ahora puede parecer una cuestión trivial, en aquella época no lo era, y otros autores erraron al intentar resolverla, generalmente porque no tenían en cuenta las posibles permutaciones de una misma combinación.

El segundo problema fue abordado por Luca Pacioli (1445–1517), quien en 1487 propuso estos dos similares problemas particulares: un juego en el que el premio es de 22 ducados que consiste en alcanzar 60 puntos se interrumpe cuando un equipo lleva 50 puntos y el otro 30; y tres arqueros que compiten por un premio de 6 ducados lanzan flechas hasta que uno de ellos haga 6 dianas, siendo interrumpidos cuando el primero de ellos lleva 4 dianas, el segundo 3 y el tercero 2. ¿Cómo deben repartirse los premios entre los contendientes? Pacioli propuso que el premio debería ser repartido en función de las victorias obtenidas anteriormente: así, el premio del primer problema se dividía en $60 \times 5/8$ ducados para el primer equipo y en $60 \times 3/8$ para el segundo; para el problema de los arqueros, el premio se dividía en la proporción $4/9$, $3/9$ y $2/9$. Como más tarde se pondría de manifiesto, esta solución es incorrecta.

GIROLAMO CARDANO Y NICCOLO TARTAGLIA

La primera obra importante relacionada con el cálculo de probabilidades en juegos de azar fue el *Libro de los Juegos de Azar*, de Girolamo Cardano (1501–1576), escrito en 1565, aunque no publicado hasta 1663. Cardano era un jugador empedernido y su obra es más bien un manual para jugadores; contiene descripciones de juegos y las precauciones a tomar para que los rivales no hagan trampas, y sólo una pequeña parte está dedicada al estudio del azar: problemas tales como calcular todos los resultados posibles al lanzar dos o tres dados y las frecuencias con que aparecían, hallar la probabilidad de que al lanzar un dado una serie de veces salga un determinado número al menos una vez, o calcular las frecuencias de los valores de la suma de las caras de una tirada de dos dados. En la resolución de estos problemas, Cardano trabajó con los conceptos de la definición clásica de la probabilidad, aunque no los definió. En concreto, Cardano introdujo la idea de asignar una probabilidad p entre 0 y 1 a un suceso cuyo resultado se desconoce, considerando el número total de resultados y el número de resultados favorables, y esbozó de una forma rudimentaria lo que ahora se conoce como la “ley de los grandes números”, al afirmar que si la probabilidad de un suceso es p , después de un número n grande de repeticiones lo más razonable es apostar a que ocurrirá alrededor de np veces. Sin embargo, Cardano no alcanzó a reconocer la importancia teórica de estos conceptos, ya que consideraba estas relaciones como meramente aritméticas, más que como una medida de la posibilidad de ocurrencia de un suceso aleatorio.

Cardano se había ocupado previamente del problema del reparto de apuestas. En 1539 escribió que la solución de Pacioli era incorrecta porque al considerar tan sólo el número de juegos ganados por cada equipo, no contaba cuántos juegos debían ganar para hacerse con el premio; como solución propuso que si n es el número de juegos totales y a y b

los juegos ganados por cada equipo, el premio debía repartirse de la siguiente manera $[1+2+\dots+(n-b)]: [1+2+\dots+(n-a)]$. Esta solución es, en general, incorrecta y sólo da resultados válidos en casos particulares. El problema del reparto de apuestas también fue abordado por Niccolo Tartaglia (1499–1557), quien en 1556 publicó un libro sobre aritmética en el que criticaba la solución dada por Pacioli («Si un bando ha ganado 10 puntos y el otro ninguno, entonces todo el premio sería para el primer jugador, pero esto no tiene ningún sentido») y dio su propio solución: si un equipo ha ganado a puntos y el otro b , se juega a n puntos y el premio total es P , las ganancias deberían repartirse de la forma $(P/2) \pm P[(a-b)/n]$, siendo la cantidad mayor para el equipo que tenga más victorias. Sin embargo, Tartaglia fue consciente de que su solución no era la correcta y le dio un carácter más jurisdiccional que matemático.

GALILEO GALILEI

Galileo Galilei (1564–1642) también se dedicó a resolver problemas sobre dados. Su obra *Sobre la Puntuación en Tiradas de Dados* calculaba el número de resultados posibles tirando tres dados. A pesar de que ya se sabía desde mucho tiempo antes que hay 216 posibilidades diferentes, Galileo fue el primero que llegó a esta conclusión a través del simple cálculo $216=6^3$. Luego atacaba el problema de calcular de cuántas maneras diferentes se puede lograr cada una de las puntuaciones entre 3 y 18. Para hacer esto, Galileo numeró los dados —primero, segundo, tercero— y fue considerando cada una de las combinaciones de los tres dados que sumaban una determinada cantidad, pero sólo entre 3 y 10. Galileo encontró que sólo hay una manera de obtener tres puntuaciones iguales, tres maneras de obtener dos puntuaciones iguales y otra diferente, y seis maneras de obtener tres puntuaciones diferentes. Su conclusión fue que es preferible apostar por el 10 antes que por el 9 porque el 10 se puede obtener de 27 maneras por 25 del 9. El resto de posibilidades —de 11 a 18— se obtenían sin cálculos, directamente por simetría: 18 igual que 3, 17 igual que 4, etc. A pesar de la simplicidad del problema, Galileo reconoció que quedó exhausto.

Sin embargo, la principal contribución de Galileo a la teoría de la probabilidad fue la creación de la teoría de la medida de errores. Según Galileo, los errores de medida son inevitables y los clasificó en dos tipos: los errores ‘sistemáticos’, debidos a los métodos y las herramientas de medida; y los errores ‘aleatorios’, que varían impredeciblemente de una medida a otra. Esta clasificación sigue en vigor actualmente. Galileo fue muy cuidadoso al analizar las propiedades de los errores aleatorios y estableció que son más frecuentes los errores pequeños que los grandes; que los errores por defecto son tan frecuentes como los errores por exceso; y que la mayoría de las mediciones se agrupan alrededor del verdadero valor. Con estas ideas, Galileo no sólo contribuyó al desarrollo de la teoría de la probabilidad, sino que puso las bases para el nacimiento de la estadística.

BLAISE PASCAL Y PIERRE DE FERMAT

El desarrollo de la teoría de la probabilidad experimentó un gran avance en Francia a mediados del siglo XVII con la correspondencia que mantuvieron Blaise Pascal (1623–1662) y Pierre de Fermat (1601-1665) durante 1654. Antoine Gombaud, caballero de Méré, filósofo y literato que jugaba compulsivamente, pidió a Pascal que le resolviese el problema del reparto de apuestas. Pascal y Fermat lo resolvieron correctamente por medios diferentes pero equivalentes, aunque el desconocimiento de la teoría general les hizo pensar que no lo eran. El acierto de ambos consistió en darse cuenta de que el reparto de las apuestas debe hacerse en función de la probabilidad de ganar que tuviese cada jugador en el momento de interrumpirse el juego. Para hallar la solución correcta se valieron de una rigurosa metodología desconocida hasta entonces; sin embargo, Pascal falló en su intento de extender el procedimiento al caso en que hubiera tres o más jugadores.

Once años más tarde, en 1665, Pascal publicaba su *Tratado sobre el Triángulo Aritmético*, la más importante contribución realizada hasta entonces en el campo de la combinatoria. El libro comienza con la construcción de lo que se dio en llamar el triángulo de Pascal, aunque era conocido desde hacía más de 500 años en diversas partes del mundo. El triángulo es de la siguiente forma:

Fila/Columna	0	1	2	3	4	5	6	7	8
0	1								
1	1	1							
2	1	2	1						
3	1	3	3	1					
4	1	4	6	4	1				
5	1	5	10	10	5	1			
6	1	6	15	20	15	6	1		
7	1	7	21	35	35	21	7	1	
8	1	8	28	56	70	56	28	8	1

donde el valor de la k -ésima entrada de la n -ésima fila es el número combinatorio $\binom{n}{k}$.

Pascal demostró algunas propiedades importantes del triángulo: cada elemento es la suma de todos los elementos de la columna anterior hasta la fila anterior (es decir, $\binom{n}{k} = \sum_{j=k-1}^{n-1} \binom{j}{k-1}$); la suma de todos los elementos de la fila n -ésima es 2^n ; y $\binom{n}{k} : \binom{n}{k+1} = (k+1) : (n-k)$. Para demostrar estos argumentos usaba algo parecido al principio de inducción, pues demostraba un caso y, a continuación, que eso implicaba el caso inmediatamente siguiente. La última gran propiedad del triángulo que demostró Pascal fue que $\binom{n}{k} = \frac{n \cdot (n-1) \cdot \dots \cdot (n-k+1)}{k!}$, demostrándolo por inducción e identificando ese número como el número de combinaciones de k elementos en un conjunto de n elementos. Finalmente, Pascal aplicó todos estos resultados para producir una solución sistemática del problema del reparto de apuestas: si al jugador A le faltan r juegos para ganar y al jugador B le faltan s (con $r+s \geq 1$), las apuestas deberían dividirse de manera que al jugador A le correspondiera una parte proporcional al cociente entre $\sum_{k=0}^{s-1} \binom{n}{k}$ y 2^n , donde $n=r+s-1$.

Pascal aplicó los razonamientos probabilísticos sobre la toma de decisiones a la teología y trató de demostrar la existencia de Dios. Su argumento era el siguiente: Dios existe o no existe; si no existe, da igual creer en él que no creer; si existe, creer que no existe provoca la condenación eterna, mientras que creer trae la salvación. Como la salvación es preferible a la condenación (en términos probabilísticos, la ganancia es mayor), una persona ‘razonable’ actuará como si Dios existiera, aunque crea que la probabilidad de que exista es pequeña.

CHRISTIAAN HUYGENS

Los trabajos de Pascal y Fermat fueron continuados por el científico holandés Christiaan Huygens (1629–1695). Su interés por la probabilidad nació en 1655 durante el transcurso de un viaje a París, donde coincidió con otros científicos y discutió con ellos el problema del reparto de apuestas. Fue así como Huygens entró en conocimiento de las obras de Pascal y Fermat y sus métodos. En 1656 salía publicado su tratado *Sobre los Cálculos en los Juegos de Azar*, el cual constaba de un breve prefacio y 14 proposiciones. En las tres primeras, Huygens introducía el concepto de esperanza matemática para variables aleatorias que toman dos o tres valores, definida como la ganancia media si se repitiera el juego muchas veces; la palabra ‘esperanza’ apareció por primera vez en la historia en la traducción al latín del original en holandés. En las seis siguientes proposiciones Huygens proponía su solución al problema del reparto de apuestas, muy similar a la de Pascal, pero lo llevó más allá, pues Huygens fue capaz de extender el problema al caso de tres jugadores; sobre esto último, no dio una solución general, sino que indicó cómo aplicar al caso general los casos particulares previamente resueltos.

Las otras cuatro proposiciones trataban sobre problemas varios. En particular, en la proposición 11 del libro aparece un problema planteado por De Méré a Pascal y Fermat— ¿cuántas veces hay que lanzar dos dados para que sea

más probable ganar que perder apostando a que saldrá al menos un doble 6?— que tanto Cardano como De Méré habían resuelto incorrectamente (18 y 24 tiradas, respectivamente) y que Huygens fue el primero en resolver correctamente: 25 tiradas. Al final del libro, se incluían cinco problemas propuestos sin solución para el lector. En 1687 se publicó un libro anónimo en el que se resolvía el primero se esbozaba la solución del resto. Más tarde, se comprobó que el autor de ese libro era el filósofo holandés de origen judeo-portugués Benito Spinoza.

2.- EL DESARROLLO DE LA TEORÍA DE LA PROBABILIDAD

Hacia el último cuarto del siglo XVII, existía ya un gran volumen de conocimientos sobre sucesos aleatorios, acompañados por un buen número de problemas correctamente planteados y resueltos. Pero todos estos conocimientos se enfocaban a problemas concretos limitados a los juegos de azar, y ningún estudioso había sido capaz de dar una definición general de la probabilidad.

PRIMERAS INVESTIGACIONES EN DEMOGRAFÍA

El primer estudioso de la probabilidad que no se centró en juegos de azar fue el comerciante inglés John Graunt (1620–1675), quien en 1662 abordó problemas sobre demografía —o “política aritmética”, como se la llamaba entonces. Graunt se propuso encontrar un método preciso para estimar la edad media de los habitantes de Londres mediante la edad de defunción; haciendo esto introdujo el concepto de ‘frecuencia de un suceso’, remarcando el cierto grado de aleatoriedad presente en las proporciones obtenidas. También demostró que en Londres la proporción de nacimientos de niños y niñas no era igual sino 14:13 y elaboró la primera tabla de mortalidad. Graunt comprendió que cuántas más observaciones hacía más precisos eran los resultados, anticipando el principio estadístico de estabilidad de las medias.

Las ideas de Graunt fueron recogidas por William Petty (1623–1687), quien elaboró un análisis comparativo entre la mortalidad de los Londres y la de París, basándose en los datos de los hospitales de caridad, y por el astrónomo Edmund Halley (1656–1742), quien presentó en 1693 una tabla de mortalidad de la ciudad de Breslau (Alemania) — Halley prefería ciudades pequeñas con pocos movimientos migratorios antes que ciudades grandes como Londres, París o Dublín— e introdujo el concepto de longitud de vida tal que la frecuencia con que se superaba o no se alcanzaba era la misma; en otras palabras, el concepto de ‘mediana’. En los trabajos de Halley también se pueden encontrar las bases de los teoremas de la suma y la multiplicación de probabilidades y de la ley de los Grandes Números, aunque no los enunció explícitamente. Las obras de Halley tuvieron gran influencia en demografía y los seguros.

PRIMERAS DEFINICIONES DE LA PROBABILIDAD

Sólo cuando los fenómenos aleatorios dejaron de enfocarse como casos particulares y se intentó ver los conceptos generales que había detrás de ellos, las nociones de probabilidad mejoraron en su definición. El primero en dar la definición clásica de probabilidad fue Jakob Bernoulli (1654–1705) en su obra *El Arte de Predecir* —publicada póstumamente en 1713—, muy influenciado por los trabajos de Graunt y Petty, que habían demostrado las ventajas de incluir en sus tablas no sólo los números absolutos, sino también las proporciones respecto del total. Más adelante, el matemático francés exiliado en Inglaterra Abraham De Moivre (1667–1754) aceptó la definición dada por Bernoulli y la reformuló en términos modernos: «una fracción en la que el numerador es igual al número de apariciones del suceso y el denominador es igual al número total de casos en los que es suceso pueda o no pueda ocurrir. Tal fracción expresa la probabilidad de que ocurra el suceso». La definición clásica de la probabilidad, en su forma actual, está basada en el concepto de equiprobabilidad de los resultados, basado a su vez en la simetría. Se supone que un experimento se puede descomponer en n sucesos equiprobables y mutuamente excluyentes E_1, \dots, E_n , llamados sucesos ‘elementales’. Así, la probabilidad de suceso aleatorio A es el número del intervalo $[0,1]$ que expresa el cociente entre los m sucesos elementales que componen A y el número total n de posibles sucesos elementales. El principal escollo que encuentra esta in-

interpretación de la probabilidad es la dificultad de descomponer un suceso en sucesos elementales equiprobables; siendo fácil para problemas sencillos, como los de cartas, dados o urnas, es casi imposible para problemas más complejos.

Basándose en los trabajos de Graunt y Petty, Bernoulli resolvió incluso la cuestión de cómo hallar la probabilidad de ocurrencia de un suceso aun siendo imposible contar los casos favorables: «Aquí hay otro camino disponible para alcanzar el resultado deseado. Lo que no se puede hallar *a priori* se puede obtener *a posteriori*, es decir, mediante la observación múltiple de los resultados de pruebas similares...» De esta manera, Bernoulli introdujo el concepto de probabilidad ‘frecuentista’ o ‘estadística’: asignar como probabilidad de un suceso el resultado que se obtendría si el proceso se repitiera en condiciones similares un número grande de veces. Sin embargo, estas condiciones son demasiado vagas para servir como base para una definición científica rigurosa. En primer lugar, se menciona un ‘número grande’ de veces, pero no se da ninguna indicación sobre cuál es ese número lo suficientemente grande; no se describe con precisión qué se entiende por condiciones similares —si las condiciones fuesen siempre exactamente las mismas se obtendría siempre el mismo resultado—; tampoco se especifica cuál es la máxima desviación admitida respecto del resultado teórico; además, sigue habiendo sucesos que no pueden plantearse suponiendo la posibilidad de repetirlos muchas veces.

Precisamente, fueron la necesidad de precisar qué se entiende por un ‘número grande’ de repeticiones del experimento y la tolerancia del resultado obtenido respecto del teórico, lo que llevaron a Jakob Bernoulli a idear, en su forma más intuitiva y básica, la Ley de los Grandes Números.

TEOREMAS BÁSICOS DE LA TEORÍA DE LA PROBABILIDAD

Los tres teoremas básicos que hicieron posible el desarrollo de la probabilidad tal y como la conocemos hoy fueron los teoremas de la suma, de la multiplicación y de la probabilidad total. Aunque ni Fermat ni Pascal ni Huygens los idearon, en sus escritos aparecen ya de una forma implícita y utilizados correctamente.

Teorema de la Suma.- Pascal dio a entender implícitamente que sabía cómo calcular los casos favorables de un suceso A si conocía los casos favorables de unos A_j disjuntos cuya unión es A (es decir, si los A_j son una partición de A). Jakob Bernoulli también fue consciente de ello, y fue más allá al darse cuenta de que la probabilidad de la unión no es la suma de las probabilidades si los sucesos no son disjuntos, aunque no supo dar la razón. Previamente, Cardano había expresado un resultado similar en términos de casos en vez de probabilidades. No fue ninguno de ellos quien formuló finalmente el teorema de la suma de las probabilidades, sino el reverendo inglés Thomas Bayes (1702–1761), cuyo trabajo fue leído póstumamente, en 1763. En esta obra, Bayes da la primera definición explícita de sucesos disjuntos —él los llamó ‘inconsistentes’— y enunció la fórmula ahora conocida:

$$P\{A \cup B\} = P\{A\} + P\{B\} - P\{A \cap B\}$$

Teorema de la Multiplicación.- Del mismo modo, el teorema de la multiplicación de probabilidades era conocido por casi todos los estudiosos a través de cálculos particulares. Sin embargo, fue Abraham De Moivre (1667–1754) el primero que los enunció con autoridad. En la introducción a su obra *Doctrina de las Posibilidades* de 1711, De Moivre presentó el importante concepto de independencia de sucesos aleatorios; así, escribió: «Diremos que dos sucesos son independientes, si uno de ellos no tiene ninguna relación con el otro» y procedió a definir los sucesos dependientes: «Dos sucesos son dependientes si están ligados el uno al otro y la probabilidad de ocurrencia de uno de ellos influye en la probabilidad de ocurrencia del otro». Una vez hecho esto, De Moivre lo aplicó al cálculo de probabilidades: «la probabilidad de ocurrencia de dos sucesos dependientes es igual a la probabilidad de ocurrencia de uno de ellos dividida por la probabilidad de que el otro ocurra si el primero ya ha ocurrido. Esta regla puede generalizarse para varios sucesos». El caso de varios sucesos lo describía así: «Se necesita elegir uno de ellos como el primero, otro como el segundo, y así. Luego, la ocurrencia del primero debe considerarse independiente de todas las demás; el segundo debe considerarse con la condición de que el primero ha ocurrido: el tercero con la condición de que tanto el primero como el segundo han ocurrido, y así. De aquí, la probabilidad de las ocurrencias de todos los sucesos es igual al producto de todas las probabilidades». O en notación moderna:

$$P\{A_1 \cap A_2 \dots A_n\} = P\{A_1\} \cdot P\{A_2 | A_1\} \cdot P\{A_3 | A_1 \cap A_2\} \dots \cdot P\{A_n | A_1 \cap \dots A_{n-1}\}$$

De Moivre acompañó sus afirmaciones con ejemplos resueltos. También fue consciente de que lo verdaderamente difícil de este teorema es descubrir cuándo dos sucesos son o no independientes.

Teorema de Bayes. - El trabajo de De Moivre obtuvo una gran difusión, así que el mérito de Bayes no fue tanto la originalidad sino expresar la probabilidad condicional en función de la probabilidad de la intersección:

$$P\{A | B\} = \frac{P\{A \cap B\}}{P\{B\}}$$

Además, el honor del teorema que lleva su nombre no es completamente suyo, ya que él no estaba en condiciones de formular con probabilidades totales. Fue Pierre-Simon Laplace (1749-1827) quien desarrolló la mayor parte del teorema de Bayes en su *Experiencia en la Filosofía de la Teoría de la Probabilidad* (1812). Sea A un suceso que ocurre en conjunción con uno y sólo uno de los n sucesos disjuntos B_1, \dots, B_n . Si se sabe que el suceso A ha ocurrido, ¿cuál es la probabilidad de que el suceso B_j también? Laplace respondió de la siguiente manera: «La probabilidad de existencia de una de esas causas es igual a una fracción con un numerador igual a la probabilidad del suceso que se sigue de esta causa y un denominador que es la suma de las probabilidades similares relativas a todas las posibles causas. Si estas diferentes causas *a priori* no son equiprobables, entonces en lugar de tomar la probabilidad del evento que sigue a cada causa, se toma el producto de esta probabilidad veces la probabilidad de la causa». Esta enrevesada receta puede escribir en notación moderna:

$$P\{B_i | A\} = \frac{P\{B_i\} \cdot P\{A | B_i\}}{\sum_{i=1}^{\infty} P\{A | B_i\} \cdot P\{B_i\}}$$

Laplace aplicó el teorema a problemas de la mecánica celeste, la estadística médica e, incluso, a la jurisprudencia. Al igual que los otros dos teoremas, el Teorema de Bayes ya se usaba anteriormente, aunque nunca había sido formulado.

LOS TEOREMAS DEL LÍMITE

La Ley de los Grandes Números. - Jakob Bernoulli era consciente de que las frecuencias observadas se acercaban a un cálculo previo de su probabilidad al aumentar el número de repeticiones del experimento. Pero él quería encontrar una prueba científica que no sólo demostrara que al aumentar el número de observaciones se podía estimar la probabilidad auténtica con cualquier grado de precisión deseado, sino que permitiera calcular explícitamente cuántas observaciones eran necesarias para garantizar esa precisión de que el resultado queda dentro de un intervalo predeterminado alrededor de la verdadera solución. El experimento que consiste repetir una prueba con la misma probabilidad de éxito un número grande de veces recibió el nombre de ‘experimento de Bernoulli’ y, más adelante, con la creación del concepto de variable aleatoria, la variable que contabiliza el número de éxitos en N pruebas se llamó ‘Bernoulli’ o ‘binomial’.

Consciente de que en las situaciones de la vida real, la certeza absoluta (probabilidad 1) es imposible de alcanzar, Bernoulli introdujo la idea de la ‘certeza moral’: para que un resultado fuese moralmente cierto, debía tener una probabilidad no menor que 0.999, mientras que un resultado con probabilidad no mayor que 0.001 se consideraría ‘moralmente imposible’. Fue para determinar la certeza moral de un suceso para lo que Bernoulli formuló su teorema, la Ley de los Grandes Números.

Para ilustrar este concepto, Bernoulli propuso el siguiente ejemplo: una urna con 30.000 bolas blancas y 20.000 negras, aunque el observador no lo sabe, pues lo que quiere es determinar la proporción entre bolas blancas y negras, sacando una de cada vez, anotando el resultado —éxito si es blanca y fracaso si es negra— y reintroduciéndola en la urna. Sea N el número de observaciones, X el número de éxitos y $p = r/(r+s)$ la probabilidad de éxito en cada prueba, siendo r el número de bolas blancas y s el de bolas negras. El teorema de Bernoulli afirma, en terminología moderna, que dada cualquier pequeña fracción ε (que Bernoulli siempre escribía en la forma $1/(r+s)$) y dado cualquier número entero positivo grande c , se puede hallar un número $N = N(c)$ tal que la probabilidad de que X/N difiera de p no más de

ε es mayor que c veces la probabilidad de que X/N difiera de p más de ε . Con símbolos:

$$P\left\{\left|\frac{X}{N} - p\right| \leq \varepsilon\right\} > c \cdot P\left\{\left|\frac{X}{N} - p\right| > \varepsilon\right\}$$

O como escriben los libros modernos:

$$\forall \varepsilon > 0 \forall c \in \mathbb{Z}^+ \exists N \text{ tal que } P\left\{\left|\frac{X}{N} - p\right| > \varepsilon\right\} < \frac{1}{c+1}$$

En su ejemplo, para $c=1.000$, Bernoulli obtuvo como resultado que eran necesarias 25.550 observaciones. La intuición le decía que no hacían falta tantas y, por ello, lo intentó con otros valores de c . Desilusionado por sentir que había fracasado en su intento de cuantificar la certeza moral, Bernoulli no incluyó en su libro las aplicaciones prometidas. El que sí lo hizo fue su sobrino Niklaus (1687–1759), que aplicó el resultado de su tío a registros de 14.000 nacimientos y llegó a la inesperada conclusión de que la frecuencia de nacimientos de niños es mayor que la de niñas, en la proporción de 18:17. Este resultado fue confirmado años después por Laplace

El Teorema Central del Límite.- La ley de los Grandes Números planteó el problema de estimar la suma de un subconjunto de términos de una expresión binomial. La principal dificultad era calcular la probabilidad de que el número de éxitos de un suceso dado de probabilidad p en n pruebas estuviera entre A y B . Jakob Bernoulli había demostrado que esta probabilidad era $\sum_{A < k < B} \binom{m}{k} \cdot p^k \cdot (1-p)^{m-k}$, siendo éste una parte de la expansión $1 = (p+(1-p))^m$. Lo más difícil era calcular $\binom{m}{k} = \frac{m!}{k!(m-k)!}$, pues $m!$ se hace muy grande cuando m es grande. Bernoulli recurrió a estimaciones muy poco precisas, pero suficientes para probar su teorema. De Moivre quiso ser algo más preciso y recurrió a una expresión asintótica y demostró que $m! \approx B \cdot e^{-m} \cdot m^{m+\frac{1}{2}}$, con B constante. Para determinar el valor de esta constante, construyó la siguiente expansión: $\ln B = 1 - \frac{1}{12} + \frac{1}{360} - \frac{1}{1260} + \frac{1}{1680} - \dots$, y halló que B es aproximadamente 2.5074, pero no quedó satisfecho porque no logró vincular este número a ninguna constante matemática conocida. De Moivre pidió ayuda a su amigo James Stirling (1692–1770), quien demostró que $B = \sqrt{2\pi}$. Con este dato, De Moivre calculó una tabla para la función $m!$ con m entre 10 y 900, y enunció un resultado que, en notación moderna, dice:

$$P\left\{X = \frac{n}{2} + t\right\} \approx P\left\{X = \frac{n}{2}\right\} \cdot e^{-(2t^2/n)} = \frac{2}{\sqrt{2\pi n}} \cdot e^{-(2t^2/n)}$$

De Moivre dibujó la gráfica de esta curva, introduciendo el importantísimo concepto de ‘distribución normal’ y demostró que esta curva es simétrica en torno al máximo y que los puntos de inflexión están a distancia $\frac{1}{2}\sqrt{n}$ de este máximo. El hecho de que el promedio de una muestra aleatoria independiente convenientemente normalizada es aproximadamente normal es lo que se conoce como el Teorema Central del Límite. Ahora ya estaba en condiciones De Moivre de resolver el problema planteado por la ley de los grandes números:

$$\sum P\left\{X = \frac{n}{2} + t\right\} = \frac{2}{\sqrt{2\pi n}} \int_0^k e^{-(2t^2/n)} dt$$

De Moivre trabajó con el valor de $k = \frac{1}{2}\sqrt{n}$ y obtuvo que la probabilidad de que el número de ocurrencias de un experimento binomial cayera dentro de un intervalo de radio $\frac{1}{2}\sqrt{n}$ era 0.6827; luego repitió el cálculo para otros múltiplos de \sqrt{n} . Finalmente, De Moivre encontró que \sqrt{n} era unidad cuya distancia del centro debe ser medida. Así, la precisión de una estimación de probabilidad aumenta igual que la raíz cuadrada del número de experimentos; en otras palabras, De Moivre acababa de descubrir la importancia de la ‘varianza’.

De Moivre repitió el experimento de Bernoulli y obtuvo que bastaban 6.498 observaciones. Aunque mejoró el

método de Bernoulli, sin embargo, no llegó a reconocer la importancia de la curva que había encontrado y no pudo aplicar su resultado a otros problemas.

EL PROBLEMA DE LA ‘RUINA DEL JUGADOR’

El problema de la ‘ruina del jugador’ tuvo un papel importantísimo en el desarrollo de la teoría de la probabilidad, pues era extraordinariamente complejo para la época y exigió la creación de nuevos métodos para su resolución; además, sirvió como trampolín para el desarrollo de los procesos estocásticos. El problema de la ‘ruina del jugador’ fue propuesto por primera vez por Huygens en su libro y consiste en lo siguiente: los jugadores A y B tienen a y b francos, respectivamente. Después de cada juego, el ganador le quita un franco al perdedor. La probabilidad de ganar de A es p y la de B es $q = 1-p$. ¿Cuál es la probabilidad p_A de que el jugador A arruine totalmente al jugador B? El problema intrigó a muchos de los científicos más importantes del momento, como Jakob y Niklaus Bernoulli, De Moivre y Laplace. Jakob Bernoulli criticó la formulación y solución numérica del problema que dio Huygens, argumentando que restringía la posibilidad de encontrar una regla general. Los resultados obtenidos por estos matemáticos demostraron que eran capaces de afrontar problemas de sucesos muy complicados y que sabían manejar los teoremas de la suma, la multiplicación y la probabilidad total, incluso antes de ser formulados.

La solución fue hallada por De Moivre y Niklaus Bernoulli de forma independiente en 1710–1711. Según De Moivre,

$$p_A = \frac{\left(\frac{q}{p}\right)^a - 1}{\left(\frac{q}{p}\right)^{a+b} - 1}, \text{ y } p_B = \frac{\left(\frac{p}{q}\right)^b - 1}{\left(\frac{p}{q}\right)^{a+b} - 1}$$

Y el número esperado N de juegos que son necesarios para que un jugador arruine al otro es,

$$E\{N\} = \frac{bp_A - ap_B}{p - q}$$

LA PARADOJA DE SAN PETERSBURGO

Una vez definido matemáticamente el concepto de esperanza, la idea general era que la manera más razonable de tomar una decisión que involucrara probabilidades sería optar por aquella que tuviera una mayor esperanza. Pero esta manera de pensar quedó desacreditada cuando en 1713 Niklaus Bernoulli propuso el siguiente juego: el jugador A lanza una moneda; si sale cara en la primera tirada, paga 2 ducados al jugador B; si la primera cara sale en la segunda tirada, le paga 4 ducados; y así, si la primera cara sale en la tirada n -ésima, le paga 2^n ducados. ¿Cuánto pagaría el jugador B al jugador A para jugar a este juego? Si se calcula la esperanza matemática, tenemos $\sum_{n=1}^{\infty} 2^n \cdot \frac{1}{2^n} = \sum_{n=1}^{\infty} 1 = \infty$, y, por tanto, B

podría pagar a A cualquier cantidad, porque su ganancia seguiría siendo infinita, pero, sin embargo, la probabilidad de ganar sólo 2 ducados es de 1 a 1, y la de ganar más de 10 es de 1 contra 7. Este ejemplo —conocido como la paradoja de San Petersburgo por ser la ciudad donde discutieron el problema Niklaus Bernoulli y su hermano Daniel (1700–1782)— puso de manifiesto que no siempre la opción más razonable es la más correcta matemáticamente. La solución a esta paradoja consistió en la introducción del concepto de ‘esperanza moral’ (en contraposición a la ‘esperanza matemática’) o ‘utilidad’, que consiste en dar prioridad al sentido común y las circunstancias personales o particulares sobre el resultado matemático. Así, una persona rica, con mucho dinero que perder, podría permitirse el lujo de hacer de jugador B, mientras que un pobre podría arriesgarse a hacer de jugador A, porque la posibilidad de perder dinero es muy pequeña; sin embargo, los papeles no serían intercambiables. Esta paradoja tuvo mucha importancia en el desarrollo de la teoría de la probabilidad aplicada a los problemas relacionados con los seguros: pagar una gran cantidad al asegurado pero con una probabilidad muy pequeña de que eso ocurra.

3.- LA PROBABILIDAD MODERNA

LA TEORÍA DE LA MEDIDA DE ERRORES

La teoría de la medida de errores fue iniciada por Galileo y continuada por otros muchos científicos, en su mayoría astrónomos, como, por ejemplo, Ticho Brahe (1546–1601), que encontró que cada medida tiene un posible error y que la precisión de la medida puede aumentar si se hacen varias medidas y se calcula la media aritmética. Los primeros intentos de construir matemáticamente la teoría de la medida de errores fue hechos por R. Cotes (1682–1716), T. Simpson (1710–1761) y Daniel Bernoulli. Cada uno de ellos tenía una idea diferente sobre la medida de los errores. Cotes opinaba que los errores se distribuyen uniformemente a lo largo el intervalo $(-a,a)$. Simpson creía que los errores pequeños ocurren más frecuentemente que los grandes, pero que están restringidos por un número a , de manera que el error es 0 en los intervalos $(-\infty,-a]$ y $[a,+\infty)$; así, la función de densidad es $x-2a^2y=-a$ en el intervalo $(-a,0)$, y en $(0,a)$ es $x+2a^2y=a$. Daniel Bernoulli fue el primero en poner en duda que la media aritmética fuera la mejor estimación del error y propuso como función de densidad $y = \sqrt{R^2 - (x - \bar{x})^2}$, donde R es conocido y \bar{x} se determina mediante repetidas observaciones. Bernoulli no se dio cuenta de que la integral de esta función no es 1, sino $\left(\frac{\pi}{2}\right)R^2$, por lo que sólo represente una verdadera función de densidad en casos particulares. El trabajo de Bernoulli, no obstante, es importante porque fue el primero en proponer estimar un parámetro desconocido mediante el método de ‘máxima verosimilitud’.

Otro estudioso de la cuestión fue Laplace, que consideraba la teoría de probabilidad más como una disciplina de la ciencia natural que de las matemáticas. Muy dedicado a la astronomía, aplicó a sus investigaciones en teoría de medida de errores. Laplace afirmó los errores de medida observados eran la suma de una gran cantidad de pequeños errores; si estos errores tenían una distribución normal, su suma también debería tenerla. Como estimación del valor desconocido del error a , Laplace sugirió tomar el valor que minimiza la cantidad $\sum_{1 \leq k \leq n} |x_k - a|$, que es igual a la media de las n observaciones realizadas.

Sin embargo, el trabajo de Laplace no alcanzó mucha difusión porque quedó eclipsado por las nuevas ideas presentadas por K. Gauss (1777–1855) y A. Legendre (1752–1833), que propusieron y desarrollaron el método de mínimos cuadrados. Gauss demostró que, bajo ciertas condiciones generales, la función de densidad de los errores de medida tiene la forma $\varphi(\Delta) = \frac{h}{\sqrt{\pi}} \cdot e^{-h^2 \Delta^2}$.

Otra gran contribución fue la realizada por Simeon Poisson (1781–1840). En particular, planteó la siguiente pregunta: ¿es cierto que la media aritmética es siempre mejor que una única observación? La respuesta es no. Como contraejemplo, presentó la siguiente distribución $p(x) = \frac{1}{\pi(1+x^2)}$, $-\infty < x < \infty$. Poisson demostró que la suma de dos variables aleatorias con esta distribución tiene la misma distribución pero con otra escala y luego probó que la media aritmética de variables aleatorias independientes de este tipo tiene exactamente la misma distribución que cualquiera de ellas. 20 años después A. Cauchy (1789–1857) repitió este mismo resultado y la distribución descubierta por Poisson recibió el nombre de Cauchy.

Más tarde, la teoría de errores atrajo la atención de muchos probabilistas rusos, como P. Chebyshev (1821–1894) y A. Markov (1856–1922), que desarrollaron el método de mínimos cuadrados.

FORMACIÓN DEL CONCEPTO DE VARIABLE ALEATORIA

Suele ocurrir que la formación de los conceptos científicos ocurre antes de que sean comprendidos totalmente. Eso también fue lo que pasó con el concepto de ‘variable aleatoria’, uno de los pilares básicos de la teoría de probabilidad moderna.

El concepto de ‘variable aleatoria’ estuvo presente de forma encubierta casi desde el principio de la teoría de

probabilidad. En uno de los problemas de su libro, Huygens introdujo una variable aleatoria que sigue una distribución hipergeométrica. Cuando Galileo habló de los errores ‘aleatorios’ que no se pueden predecir y que varían de medida en medida, en realidad se refería a que esos errores son una variable aleatoria de distribución desconocida. Cuando Bernoulli enunció su ley de los grandes números, al contabilizar el número de bolas blancas extraídas de la urna, ese número de éxitos es una variable aleatoria que toma valores entre 1 y n el número total de pruebas, siguiendo una distribución binomial. Al analizar el problema de la ruina del jugador, también aparece una variable aleatoria, el número de juegos hasta que uno de los jugadores queda arruinado. Y De Moivre fue incluso más lejos al introducir la distribución normal. Sin embargo, ninguno de ellos se dio cuenta que hacía falta introducir un nuevo concepto. Tampoco Gauss, Daniel Bernoulli o Laplace mencionaron en ningún momento esta idea, ni siquiera cuando en el comienzo del siglo XIX aparecieron nuevos problemas en los que siempre aparecía la distribución normal: el tamaño de los órganos de animales de la misma edad, la desviación de los proyectiles en la artillería o en la teoría matemática de la física molecular aplicada a los gases. Además, la preeminencia de la distribución normal sobre todas las demás tuvo el efecto contraproducente de desalentar la investigación sobre las propiedades de la media y la varianza, porque para el caso de la normal es una cuestión bien conocida.

Los primeros pasos en la dirección de introducir la idea de ‘variable aleatoria’ fueron dados por Poisson en 1832 en su libro *Sobre la Probabilidad de los Resultados Promedios de Observaciones*. No utilizó el término ‘variable aleatoria’, pero sí habló de ‘alguna cosa’ que uno puede entender como un conjunto a_1, a_2, \dots, a_n con sus correspondientes probabilidades p_1, p_2, \dots, p_n ; es decir, habló de las variables aleatorias discretas. También consideró variables aleatorias continuas y sus densidades. La palabra ‘variable’ fue utilizada por primera vez por Chebyshev, que asumió implícitamente que todas las variables aleatorias eran independientes y fue A. Liapunov (1857–1918) el primero que usó sistemáticamente el término ‘variable aleatoria’ y especificó que serían independientes cuando fuese necesario. En el comienzo de su obra *Sobre una Afirmación en la Teoría de Probabilidad*, Liapunov dio la definición de función de distribución tal y como la conocemos hoy: $P\{a < \xi < b\} = F(b) - F(a)$.

LA LEY DE LOS GRANDES NÚMEROS

El teorema de Bernoulli fue generalizado por vez primera por Poisson en 1837, quien también introdujo el término ‘ley de los grandes números’. Poisson consideró una sucesión de n pruebas independientes, en cada una de las cuales A puede ocurrir con probabilidad p_k $k=1, \dots, n$. Si μ_n es el número de ocurrencias de A en las n observaciones, entonces: $\forall \varepsilon > 0 \ P\left\{\left|\frac{\mu_n}{n} - \frac{p_1 + p_2 + \dots + p_n}{n}\right| < \varepsilon\right\} \xrightarrow{n \rightarrow \infty} 1$. En 1843 Chebyshev criticó el teorema de Poisson, alegando que no daba una cota explícita para el error y enunció su propia versión, dando una estimación de n para que se cumpliera la desigualdad: $\forall \varepsilon > 0 \ \forall \eta \ P\left\{\left|\frac{\mu_n}{n} - \frac{p_1 + p_2 + \dots + p_n}{n}\right| < \varepsilon\right\} > 1 - \eta$. Sin embargo, estas dos versiones no significaron ningún avance, pues no superaban la idea original de Bernoulli. Sólo cuando en 1867 Chebyshev empezó a considerar variables aleatorias en lugar de sucesos, se produjo ese avance cualitativo.

En 1909 Émile Borel (1871–1956) demostró que el experimento de Bernoulli llevaba a una afirmación más fuerte que la ley de los grandes números para $p=0.5$. En 1917, el matemático italiano Francesco Cantelli (1875–1966) extendió este hecho para p arbitrario: $P\left\{\lim_{n \rightarrow \infty} \frac{\mu_n}{n} = p\right\} = 1$, afirmación que se conoce como la Ley Fuerte de los Grandes Números. Una generalización de este resultado fue dada por Andrei Kolmogorov (1903–1987) en 1930.

En 1935 A. Khinchine (1894–1959) introdujo el concepto de estabilidad relativa de suma de variables aleatorias: la suma $S_n = \xi_1 + \xi_2 + \dots + \xi_n$, se dice ‘relativamente estable’ si existen constantes $A_n > 0$ tales que para todo $\varepsilon > 0$ y $n \rightarrow \infty$, se cumple $P\left\{\lim_{n \rightarrow \infty} \frac{\mu_n}{n} = p\right\} = 1$. Para variables aleatorias idénticamente distribuidas, Khinchine encontró condiciones necesarias y suficientes; su discípulo A. Bobrov (1912–) extendió este resultado a variables con distribuciones diferen-

tes.

Uno de los resultados relacionados con la ley de los grandes números es el teorema de Glivenko–Cantelli, descubierto por A. Glivenko en 1929–1933, que trata sobre la convergencia de una distribución empírica a una verdadera función de distribución.

EL TEOREMA CENTRAL DEL LÍMITE

Los esfuerzos por lograr una generalización del teorema de De Moivre vinieron desde dos campos: la teoría de medida de errores y la estadística física. Entre los primeros se cuentan Laplace, que en 1809 formuló un teorema de límites para la suma de variables aleatorias distribuidas uniformemente en un intervalo $(-h, h)$ y, considerando un número creciente de variables aleatorias discretas dio una aproximación de una distribución continua a partir de otra discreta; Poisson, que en 1837 dio un teorema local del límite y falló al generalizarlo a la suma de variables aleatorias independientes arbitrarias con varianza finita porque le faltó añadir que debían ser idénticamente distribuidas; y F. Bessel, que coincidió con Laplace en que la suma de un número grande de cantidades aleatorias pequeñas en comparación con la suma tiene una distribución normal, aunque en 1838 dio ejemplos de medidas en las que los errores adoptan otras distribuciones, como, por ejemplo, la medida de ángulos.

En el lado de la estadística, aparecen principalmente los matemáticos rusos de finales del siglo XIX y principios del XX. El primero de ellos fue Chebyshev, quien en 1887 demostró un teorema sobre la acotación de la suma de variables aleatorias con esperanza 0. Para probarlo, Chebyshev ideó el Método de los Momentos, de gran importancia en la estadística. Sin embargo, la demostración tenía algunos errores que fueron corregidos por su discípulo A. Markov, que demostró en 1898 la siguiente versión del Teorema de Chebyshev:

Sea S_n la sucesión de sumas $u_1 + u_2 + \dots + u_n$ y $\Phi_n(x)$ la distribución de u_n .

Si $\forall k$ se cumple $\int_{-\infty}^{\infty} x^k d\Phi_n(x) \rightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^k e^{-x^2/2} dx$, entonces se cumple

$$\forall a, b \ P \left\{ a < \frac{S_n}{\sqrt{\text{Var } S_n}} < b \right\} \rightarrow \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx$$

Este parecía el teorema del límite definitivo, hasta que en 1900 A. Liapunov demostró que si las variables aleatorias tienen momento de orden 3 finito $c_k = E \left\{ \xi_k - a_k \right\}^3$ y $C_n = \sum_{1 \leq k \leq n} c_k$, $B_n^2 = \sum_{1 \leq k \leq n} \text{Var}(\xi_k)$, y $\frac{C_n}{B_n^2} \xrightarrow{n \rightarrow \infty} 0$, entonces las funciones de distribución de S_n convergen a la distribución normal. Al año siguiente, Liapunov probó bastaba con pedir que el momento de orden $2+\delta$ con $\delta > 0$, y probó el mismo teorema pero cambiando 3 por $2+\delta$. Liapunov llegó incluso a calcular el orden de esta convergencia a la normal: $n^{-1/2} \log(n)$.

La demostración de este teorema no exigía la aplicación del método de momentos de Chebyshev y Markov, queriendo recuperar la idea de su maestro, propuso el concepto de variables aleatorias truncadas, para las cuales su versión del Teorema Central del Límite de Liapunov sí necesita el método de momentos.

En 1922 el matemático finlandés Lindeberg fue más allá que Liapunov, al no pedir la existencia de ningún momento salvo el de orden 2. Así, si $\forall \tau > 0 \lim_{n \rightarrow \infty} \frac{1}{B_n} \sum_{1 \leq k \leq n} \int_{|x-a_k| > \tau B_n} (x-a_k)^2 dF(x) = 0$, la distribución de la suma de variables aleatorias, centradas en sus esperanzas y normadas por la raíz cuadrada de la suma de sus varianzas converge a una distribución normal estándar. En 1942 William Feller demostró que la condición suficiente de Lindeberg es también necesaria.

Como corolario al teorema de Lindeberg se obtuvo un resultado esperado desde hacía mucho tiempo: si se tienen variables aleatorias independientes idénticamente distribuidas con varianza finita no nula, se puede aplicar el Teorema Central del Límite a su suma.

En 1927 S. Bernshtein consideró un problema más general. Sea $\xi_1, \xi_2, \dots, \xi_n, \dots$ una sucesión de variables aleato-

rias independientes sin suponer nada acerca de sus esperanzas o varianzas. Bernshtein encontró condiciones necesarias y suficientes para hallar constantes $B_n > 0$ y A_n , tales que la función de distribución $(S_n - A_n)/B_n$ converja a la normal. En 1935, Feller demostró que esas condiciones eran también necesarias, suponiendo que los términos de la suma son uniformemente pequeños. Ese mismo año, A. Khinchine y Paul Lévy (1886–1971) encontraron, de forma independiente, condiciones necesarias y suficientes para la convergencia de la distribución suma de variables aleatorias independientes idénticamente distribuidas a la distribución normal.

Actualmente, las investigaciones se centran en cuestiones tales como cuánto de rápida es la convergencia a la normal, cómo converge un número aleatorio de variables aleatorias, o las relaciones entre el Teorema Central del Límite y la Ley de los Grandes Números.

LA INTERPRETACIÓN SUBJETIVA DE LA PROBABILIDAD

Las interpretaciones clásica y frecuentista de la probabilidad no satisfacían a todos. Así, en el segundo cuarto del siglo XX surgió una nueva interpretación, llamada ‘subjetiva’, según la cual la probabilidad mide el grado de creencia de un individuo en la verdad de una proposición, variando entre 0 (el individuo cree que es falso) a 1 (cree que es cierto). Esta interpretación fue propuesta por primera vez por el filósofo Frank P. Ramsey en su libro *Los Fundamentos de las Matemáticas* de 1931, y el primer matemático que la adoptó fue el estadístico italiano Bruno de Finetti en 1937. La interpretación subjetiva de la probabilidad es más amplia que la frecuentista, pues mientras que ésta sólo funciona con experimentos que se puedan repetir un número grande de veces, aquélla se puede aplicar a cualquier tipo de proposiciones. Además, mientras que los frecuentistas consideran que la probabilidad de un suceso es siempre una constante (que se aproxima repitiendo el proceso), para los subjetivistas la probabilidad de un suceso puede —y debe— variar en función de la nueva información recibida respecto del suceso, manera de proceder que se ajusta más al método científico.

Una crítica que se ha hecho de la interpretación subjetiva es la supuesta arbitrariedad con que se asignan probabilidades a sucesos. Los subjetivistas se defienden afirmando que las normas de la moral y la coherencia obligan a quien asigna la probabilidad de un suceso a actuar del modo más objetivo posible y no de forma caprichosa. Además, se han hecho esfuerzos por convertir esta noción intuitiva en demostraciones formales.

La interpretación subjetiva es muy utilizada en los diseños de modelo probabilísticos de la física cuántica, y sus técnicas se han aplicado con éxito recientemente para filtrar el *spam* del correo electrónico legítimo.

LA AXIOMATIZACIÓN DE LA PROBABILIDAD

La definición de variable aleatoria dada por Poisson era demasiado intuitiva y demasiado apegada a la experiencia práctica como para servir de base a un desarrollo teórico de la teoría de la probabilidad. Era necesaria una formalización más grande de los conceptos probabilísticos y hubo que esperar al desarrollo de las teorías de conjuntos y de la medida que tuvo lugar a finales del siglo XIX y comienzos del XX, debidos principalmente a Georg Cantor (1845–1918), Émile Borel y Henri Lebesgue (1875–1941). Ya en 1909, Borel consideró la importancia de la teoría general de la medida para la construcción de la fundamentación de la teoría de la probabilidad, pero no fue hasta 1933 cuando N. Kolmogorov se propuso construir una teoría de la probabilidad de una manera rigurosa, basándose en axiomas fundamentales, similar el tratamiento que Euclides dio a la geometría.

La construcción axiomática de la teoría de la probabilidad procede de las propiedades fundamentales de la probabilidad observadas en los ejemplos que ilustran las definiciones clásica y frecuentista. Así, la definición axiomática las incluye como casos particulares y supera las carencias de ambas. De esta manera, la probabilidad pudo desarrollarse como un teoría completamente lógica al mismo tiempo que siguió permitiendo resolver los problemas aplicados de la ciencias modernas y la tecnología.

La definición axiomática da una equivalencia entre los conceptos de la teoría de la medida y los de la teoría de la probabilidad. Se toma un conjunto de medida 1, Ω , cuyos elementos son sucesos elementales (puntos de un espacio de

medida) y se considera una ‘ σ -álgebra’ M , un subconjunto de Ω que satisface que incluye a Ω y a \emptyset , que es cerrado por complementación y por uniones numerables. Luego se define una función P que asigna a un suceso (o conjunto medible en la teoría de la medida) un número entre 0 y 1 (su medida). Así, la terna (Ω, M, P) recibe el nombre de ‘espacio de probabilidad’. No obstante, la teoría de la probabilidad no queda inmersa dentro de la de la medida, porque ésta no posee una noción fundamental de la probabilidad, la independencia de variables aleatorias (equivalente probabilístico de las funciones medibles).

Los axiomas de Kolmogorov que definen la probabilidad son los siguientes:

1. Para cada suceso aleatorio A hay asociado un número no-negativo $P(A)$ que se llama su probabilidad.
2. $P(\Omega)=1$
3. Si los sucesos A_1, A_2, \dots, A_n son mutuamente excluyentes dos a dos, entonces, $P(A_1 + A_2 + \dots + A_n) = P(A_1) + P(A_2) + \dots + P(A_n)$.

Del hecho de que $\Omega = \emptyset \cup \Omega$ y el axioma 3 se deduce que $P(\Omega) = P(\emptyset) \cup P(\Omega)$, y de esto se obtiene una serie de resultados importantes:

$$P(\emptyset) = 0; \forall A P(A^c) = 1 - P(A); \forall A 0 \leq P(A) \leq 1; A \subseteq B \Rightarrow P(A) \leq P(B);$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B); P(A_1 \cup \dots \cup A_n) \leq \sum_{i=1}^n P(A_i).$$

El sistema de axiomas de Kolmogorov es consistente, pues existen objetos reales que los satisfacen. Por ejemplo, $\Omega = \{a_1, a_2, \dots, a_n\}$ y M el conjunto de todos los subconjuntos de Ω , donde $P(a_i) = p_i$ con $i=1, \dots, n$, satisface los axiomas de Kolmogorov.

Sin embargo, el sistema de axiomas es incompleto, pues para un conjunto dado Ω se pueden elegir las probabilidades en la σ -álgebra M de maneras diferentes. Esto no significa que el sistema de axiomas no sea lo bastante razonable, sino que ocurre en ocasiones que es necesario estudiar el mismo conjunto de sucesos aleatorios con diferentes probabilidades; por ejemplo, los posibles resultados de lanzar dos dados equilibrados o de lanzar un dado equilibrado y otro truco son los mismos y, sin embargo, las probabilidades asociadas a estos sucesos son diferentes.

También ocurre en probabilidad que constantemente hay que considerar sucesos que se pueden descomponer en una cantidad infinita de sub-sucesos. Ello exige un nuevo axioma, a elegir de entre estos dos, que son equivalentes: el ‘axioma extendido de la suma de probabilidades’, que dice que si el suceso A es equivalente a la ocurrencia de al menos uno de los sucesos mutuamente excluyentes dos a dos $A_1, A_2, \dots, A_n, \dots$ entonces $P(A) = P(A_1) + P(A_2) + \dots + P(A_n) + \dots$; o el ‘axioma de continuidad’: si la sucesión de sucesos $B_1, B_2, \dots, B_n, \dots$ es tal que cada suceso implica el anterior y su intersección es el vacío, entonces

$$P(B_n) \xrightarrow{n \rightarrow \infty} 0.$$

4.- APLICACIONES DE LA TEORÍA DE LA PROBABILIDAD

LA PROBABILIDAD GEOMÉTRICA

El primero que planteó problemas probabilísticos asociados a la geometría fue J. Arbuthnot (1667–1735), quien, en su traducción al inglés, en 1692, del libro de Huygens, añadió un problema de naturaleza totalmente distinta a los otros: se lanza un cubo, con aristas a, b, c que se cortan en ángulo recto; ¿cuál es la probabilidad de que el cubo caiga sobre el plano que contiene a dos de esas aristas y es perpendicular a la otra? En la resolución de este problema, hubo que identificar casos favorables y posibles con áreas.

El naturalista George Buffon (1707–1788) planteó en 1733 otros dos problemas de probabilidad geométrica, con el objetivo de demostrar que no sólo la aritmética, sino también la geometría, sirve para resolver problemas de probabilidad. El primer problema de Buffon era el siguiente: se divide el suelo en figuras idénticas; se lanza una moneda de radio $2r$, menor que el tamaño de cualquiera de las figuras; ¿cuál es la probabilidad de que la moneda corte al menos el borde de una figura? Para resolverlo, se considera un rectángulo en el plano de lados a y b , tales que $\min(a,b) < 2r$; la probabilidad de la moneda corte al menos uno de los lados de ese rectángulo es $\frac{2r(a+b-2r)}{ab}$. El segundo problema reza de la siguiente manera: un plano es dividido por líneas paralelas a igual distancia, a ; se lanza una aguja de longitud $2r$; ¿cuál es la probabilidad de que la aguja toque alguna de las líneas? Buffon dio una respuesta equivocada, que años más tarde fue corregida por Laplace: $\frac{4r(2a-r)}{\pi a^2}$.

Los principales estudiosos de la probabilidad geométrica durante el siglo XIX fueron G. Lamé (1795–1870), J. Barbier (1797–1856), D. Sylvester (1814–1857) y M. Crofton (1814–1874). Sylvester propuso otro problema clásico de la probabilidad geométrica: se colocan en el plano aleatoriamente cuatro puntos; ¿cuál es la probabilidad de que esos cuatro puntos sean los vértices de un cuadrángulo convexo? Sylvester dio como solución $1 - \frac{4M}{A}$, donde A denota el área de la figura convexa y M la del triángulo que forman los tres primeros puntos colocados.

En el siglo XX, el interés por la probabilidad geométrica aumentó a causa de sus múltiples aplicaciones a la física, la biología, la medicina y la ingeniería.

CONTROL DE CALIDAD

Con el desarrollo de la producción en masa, en los últimos 60–70 años ha aumentado el interés en los problemas de control de calidad. Los orígenes de este tipo de problemas se remontan a 1740, cuando T. Simpson propuso lo siguiente: dado un número de objetos de calidad diferente: n_1 del tipo 1, n_2 del tipo 2, y así; Se eligen aleatoriamente m objetos; hallar la probabilidad de haber sacado m_i del tipo i . El problema fue resuelto en 1846 por M. Ostrogadsky (1801–1862) que trató de aplicarlo a los problemas de intendencia militar con el objetivo de ahorrar tiempo en las comprobaciones. En la resolución, Ostrogadsky aplicó la regla de Bayes, pero hizo la suposición de que la probabilidad de sacar que un objeto del tipo n era siempre $\frac{1}{n+1}$, que es demasiado poco realista. Mucho más lo es el planteamiento actual del problema del control de calidad: cada objeto puede ser defectuoso con probabilidad p ; esta probabilidad, en un sistema de producción bien organizado debería ser pequeña y estable en el tiempo. En este caso, la probabilidad de encontrar m objetos defectuosos de entre n se expresa mediante la fórmula de Bernoulli: $P_n(m) = \binom{n}{m} p^m \cdot (1-p)^{n-m}$.

Los métodos estadísticos de control de calidad alcanzaron gran importancia durante la 2ª guerra mundial, ya que algunos productos no podían ser comprobados sin destruirlo —detonadores, bombas— y, en otros casos, porque una comprobación exhaustiva de la calidad exigía un gran esfuerzo no permitido por las circunstancias bélicas.

PROCESOS ESTOCÁSTICOS

La probabilidad clásica sólo se ocupa de problemas estacionarios en el tiempo, mientras que físicos, biólogos e ingenieros están interesados en procesos que evolucionan con el tiempo. Algunos ejemplos de estos problemas son el número y la duración de las llamadas telefónicas, sistemas dinámicos de población, movimiento de partículas que chocan entre ellas, difusión entre líquidos, decaimiento radiactivo.

El concepto de proceso estocásticos, basado en la definición axiomática de la probabilidad, es el siguiente: sea Ω el conjunto de sucesos elementales y t un parámetro continuo. Un proceso estocástico es la función de dos argumentos $\xi(t) = \varphi(\omega, t)$ $\omega \in \Omega$; para cada valor de t , la función $\varphi(\omega, t)$ es sólo una función de ω y, por consiguiente, una variable aleatoria; para cada valor fijo de ω , la función $\varphi(\omega, t)$ depende sólo de t y es una función de una variable real, función

que recibe el nombre de ‘realización del proceso estocástico’. Un proceso aleatorio puede verse como la colección de variables aleatorias $\xi(t)$ que dependen del parámetro t o como la colección de las realizaciones del proceso $\xi(t)$. Es natural asignar una medida de probabilidad en el espacio de funciones de las realizaciones para definir el proceso.

Los primeros en estudiar estos procesos fueron físicos como Max Planck y Fokker, y matemáticos como Markov o Slutsky, pero no fue hasta la década de 1930 cuando Kolmogorov y Khinchine construyeron la teoría general y rigurosa de los procesos estocásticos.

MARTINGALAS

En teoría de probabilidad, una martingala es un proceso estocástico discreto, es decir, una sucesión de variables aleatorias X_1, X_2, \dots, X_n que satisfacen $E\{X_{n+1} | X_1, \dots, X_n\} = X_n$, es decir, la esperanza condicional de la próxima observación, dadas todas las anteriores, es igual a la última observación. El término fue adoptado del lenguaje de los jugadores. Una definición algo más general es esta otra: una sucesión Y_1, Y_2, \dots, Y_n es una martingala con respecto a otra sucesión X_1, X_2, \dots, X_n si $E\{Y_{n+1} | X_1, \dots, X_n\} = Y_n \forall n$.

Originalmente, la martingala se refería a una estrategia de juego popular en la Francia del siglo XVIII. Se lanza una moneda; el jugador apuesta 1 a que sale cara; si pierde, apuesta 2 a que sale cara en esa segunda tirada; si vuelve a perder, apuesta 4 en la tercera tirada; así, si pierde la n primeras tiradas, apuesta 2^n en la tirada $(n+1)$ -ésima. Cuando gane, el jugador recupera todas las apuestas anteriores y, además, un beneficio igual a la primera apuesta. Un jugador con infinita riqueza e infinito tiempo a su disposición tiene garantizado ganar alguna vez, así que este se convirtió en un juego muy popular entre la gente que quería enriquecerse rápidamente. El problema para el jugador es que nadie posee una infinita riqueza y el crecimiento exponencial de las apuestas arruina a cualquiera, incluso con una racha de mala suerte relativamente pequeña.

El concepto de martingala fue introducido por Paul Lévy en los años 40, y gran parte del desarrollo original de la teoría fue hecho por Doob. Parte de la motivación de este trabajo fue demostrar la imposibilidad de encontrar estrategias de apuestas que siempre tengan éxito. A pesar de que la definición matemática de martingala es del siglo XX, De Moivre ya hizo uso de una martingala matemática para resolver el problema de la ‘ruina del jugador’.

MÉTODO DE MONTECARLO

El Método de Montecarlo es un método usado en estadística para generar números aleatorios, así llamado por el casino de Mónaco, ya que toma una ruleta como un generador simple de números aleatorios.

La simulación de Montecarlo fue creada para resolver integrales que no se pueden resolver por métodos analíticos, usando números aleatorios. La idea básica del método consiste en escribir la integral requerida como el valor esperado de alguna función con respecto a alguna de distribución de probabilidad, lo cual sugiere una solución estadística al problema de integración. El método proporciona soluciones aproximadas a una gran cantidad de problemas matemáticos, posibilitando la realización de experimentos con muestreos estadísticos en un ordenador. También tiene aplicación a la probabilidad geométrica.

Una de las aplicaciones más importantes del método de Montecarlo tuvo lugar con el trabajo para fabricar la bomba atómica durante la 2ª guerra mundial, ya que había que simular problemas probabilísticos de hidrodinámica concernientes a la difusión de neutrones aleatorios en el material de fusión.

BIBLIOGRAFÍA

LIBROS

- COOKE, ROGER, The History of Mathematics. A Brief Course, John Wiley & Sons Inc, 1997
- DE GROOT, MORRIS H., Probabilidad y Estadística, Addison–Wesley Iberoamericana, 1988
- GNEDENKO, BORIS, Theory of Probability, Gordon & Breach Science Publications, 1997
- GRIMMETT, G.R./STIRZAKER D.R., Probability and Random Processes, Oxford University Press, 1992
- KATZ, VICTOR, A History of Mathematics. An Introduction, Addison–Wesley, 1998
- TODHUNTER, ISAAC, History of the Theory of Probability, Chelsea Publishing Company, 1965

INTERNET

- “The MacTutor History of Mathematics archive”:
www-groups.dcs.st-and.ac.uk/~history
- “Figures from the History of Probability and Statistics”:
www.economics.soton.ac.uk/staff/aldrich/Figures.htm
- “Earliest Known Uses of Some of the Words of Mathematics”:
members.aol.com/jeff570/mathworld.html
- “A short History of Probability and Statistics”
www.leidenuniv.nl/fsw/verduin/stathist/stathist.htm
- “History of Science: Origins of Modern Probability Theory”
www.mala.bc.ca/~johnstoi/darwin/sect4.htm

PROBABILIDAD Y JUEGOS DE AZAR

- Resultados de dados y problema del 'reparto de apuestas'
- Fermat y Pascal (1654) → combinatoria; triángulo aritmético
- Pascal → existencia de Dios; teoría de la decisión
- Huygens (1655) → esperanza matemática
- Paradoja de San Petersburgo (1713) → esperanza 'moral'

PROBABILIDAD Y DEMOGRAFÍA

- Graunt (1662) → estabilidad de las medias
- Halley (1693) → mediana

TEOREMAS BÁSICOS

- Teorema de la suma → Bayes
- Teorema de la multiplicación → De Moivre: independencia
- Teorema de Bayes → Bayes y Laplace

TEOREMAS DEL LÍMITE

- Ley de los Grandes Números → Jakob Bernoulli (~1700)
 - Observaciones necesarias para garantizar convergencia de frecuencias observadas a frecuencia teórica → binomial
- Teorema Central del Límite → De Moivre (1711)
 - De Moivre → $m! \approx B \cdot e^{-m} \cdot m^{m+\frac{1}{2}}$; Stirling → $B = \sqrt{2\pi}$
 - $P\left\{X = \frac{n}{2} + t\right\} \approx P\left\{X = \frac{n}{2}\right\} \cdot e^{-(2t^2/n)} = \frac{2}{\sqrt{2\pi n}} \cdot e^{-(2t^2/n)}$
 - Gráfica → curva 'normal'; varianza

TEORÍA DE MEDIDA DE ERRORES

- Problemas de astronomía → Brahe (media aritmética); Daniel Bernoulli (máxima verosimilitud); Laplace; Gauss y Legendre (método de mínimos cuadrados)
- Distribución de Cauchy → Poisson (1824): la media aritmética no siempre es mejor que una única observación

$$p(x) = \frac{1}{\pi(1+x^2)}, \quad -\infty < x < \infty$$

LA PROBABILIDAD MODERNA

- Variable aleatoria: Poisson (1832): discretas; Chebychev: independientes; Liapunov: definición general, función de distribución
- Ley de los Grandes Números → Generalizaciones: Poisson (1837), Chebyshev (1867), Borel (1909), Cantelli (Ley Fuerte 1917), Kolmogorov (1930), Khinchine (estabilidad de sumas de vv.aa. 1935), Glivenko (convergencia de funciones de distribución 1933)
- Teorema Central del Límite → Generalizaciones: Laplace (1809); Poisson (1837), Bessel (1838), Chebychev (1887), Markov (1898), Liapunov (1900), Lindeberg (condición necesaria para la convergencia a la normal de suma de vv.aa. 1922) Feller (condición suficiente 1942)

INTERPRETACIONES DE LA PROBABILIDAD

- Interpretación clásica → Bernoulli: equiprobabilidad
- Interpretación frecuentista → Bernoulli: observaciones múltiples
- Interpretación subjetiva → De Finetti (1937): grado de creencia
- Interpretación axiomática → Kolmogorov (~1930): teoría de la medida + independencia

APLICACIONES

- Probabilidad geométrica
- Control de calidad
- Procesos estocásticos → problemas que varían en el tiempo
- Martingalas → estrategias de juego
- Método de Montecarlo → generación de números aleatorios