La naturaleza de la probabilidad. Una revisión histórico-epistemológica

César Sáenz Castro

1. Historia y filosofía de la probabilidad

EXISTEN características poco usuales en el desarrollo histórico de la probabilidad en comparación a otras ramas matemáticas tales como la geometría o aritmética. Un enfoque matemático de la probabilidad empezó a surgir hace poco más de tres siglos, mucho después que el hombre tuviera las primeras experiencias con el azar. Un gran número de paradojas acompañó el desarrollo conceptual indicando la disparidad entre intuiciones y enfoques formales. Un hito importante fue abandonar la tarea de formalizar una interpretación específica y concentrarse en estudiar la estructura de la probabilidad. Una fundamentación matemática sólida se estableció por Kolmogorov en 1933 pero no clarificó la naturaleza de la probabilidad. Todavía hoy existen distintos enfoques filosóficos que despiertan controversia.

1.1. Tardía y dual emergencia de la probabilidad

¿Por qué no hubo teoría de probabilidad en Occidente antes de Pascal, en el siglo XVII, a pesar de que en todas las civilizaciones se utilizaban aparatos y juegos de azar? Hacking (1975) describe como «ausente familia de ideas» a este hecho y al analizar las razones de esta ausencia considera insuficientes o irrelevantes cada una de las explicaciones que se han dado, consideradas individualmente:

1) Se ha argumentado que una visión determinista del mundo excluye el pensamiento probabilístico; sin embargo, una conjetura alternativa y mejor es que el pensamiento determinista, causal, es esencial para la formación de los conceptos de azar y probabilidad y por eso el modelo de causación mecánica y el modelo probabilístico emergen en el mismo período histórico, el siglo XVII.

2) Las loterías y los dados constituyen una buena forma de consultar a los dioses directamente, sin sacerdotes intermediarios, pero entonces resulta impío intentar computar lo que los dioses dicen, es decir, el papel de los dados en la adivinación podría excluir investigaciones críticas de las leyes de la aleatoriedad; sin embargo, mucha gente impía y culta era aficionada a los juegos de azar (Hacking pone como ejemplo a Marco Aurelio) y no por eso reflexionaron sobre la aritmética del azar.

3) Para concebir las leyes de la probabilidad necesitamos tecnología del azar, aparatos aleatorios

que permitan generar ejemplos empíricos fácilmente comprensibles; las primeras experiencias aleatorias siempre emplean lo que Neyman (1950, citado en Hacking, 1975) llamó un Conjunto de Probabilidad Fundamental (CPF) de alternativas igualmente probables; sólo después de que el individuo comprenda esta idea puede progresar a conjuntos cuyas alternativas no son equiprobables. Se sugiere que en la edad antigua no existían CPF que nos diesen idea de equiprobabilidad: por ejemplo, los más antiguos de los dados cúbicos conocidos, hallados en tumbas egipcias datadas como anteriores al 2000 a. de C., no proporcionan un conjunto de 6 probabilidades iguales porque no son de tamaño uniforme, ni en el material ni en la forma de numerar sus caras (si bien en muchos de ellos los números de 1 a 6 están dispuestos de forma que las caras opuestas sumen 7, igual que en los dados modernos). Sin embargo, argumenta Hacking, aunque no mucho, existía material aleatorio adecuado, por ejemplo, se conservan dados de marfil muy antiguos en el Museo de Antigüedades del Cairo que están muy bien equilibrados.

4) Hay dos motivos por los que una ciencia se desarrolla: en respuesta a problemas que ella misma crea y en respuesta a problemas que le son propuestos desde fuera, problemas derivados, sobre todo, de necesidades económicas. Pues bien, sólo muy recientemente la teoría de probabilidad ha sido capaz de crear sus propios problemas y generar sus propios programas de investigación; históricamente, el estímulo vino de otras disciplinas: en el S. XVII el establecimiento de los seguros y anualidades impulsaron a la estadística, en el S. XVIII la teoría de la medida se desarrollaba con fuerza sobre todo al servicio de la astronomía, en el S. XIX se creaba la biométrica para análisis de datos biológicos, en el S. XX las necesidades de la agricultura y medicina motivan el desarrollo de la teoría probabilística; con todo, esta explicación economicista no tiene en cuenta que algunas destrezas de cálculo de anualidades ya eran utilizadas en la época romana.

5) La matemática en Occidente no era suficientemente rica en ideas y capacidad de cálculo para generar una matemática del azar; faltaba, sobre todo, un álgebra combinatoria porque hay que esperar a 1666 a que Leibniz publique su *Ars Combinatoria*. Por contra, los indios y árabes que tenían un buen sistema de numeración también desarrollaron antes terminología y cálculos probabilísticos (el término *hazard* es tan árabe como el término álgebra). Desde una perspectiva educativa, conviene subrayar el paralelismo psico-histórico que se desprende del dato de que si las técnicas combinatorias fueron necesarias para la aparición histórica de la probabilidad, Piaget e Inhelder (1951) establecen que es necesario que el niño posea el esquema combinatorio, que forma parte del pensamiento intelectual más avanzado, para que pueda comprender el concepto de probabilidad.

En definitiva y como corolario, Hacking establece que la conjunción de diversos factores tales como «la impiedad», la existencia de aritmética, un diferente concepto de causalidad y el desarrollo comercial, debería conducir a la formación de la matemática de la probabilidad. Como dato confirmatorio encuentra que hace 2000 años la India tenía un avanzado sistema de mercado, tenía un buen sistema de numeración, y tanto su piedad como sus teorías de las causalidad no seguían moldes europeos; pues bien, en esta sociedad se encuentran rastros de una teoría de la probabilidad desconocida en Occidente. Con todo y aunque los dados son uno de los más viejos pasatiempos humanos, el hecho histórico es que no se conocen matemáticas de la aleatoriedad hasta el Renacimiento y que ninguna de las explicaciones de este hecho es concluyente.

De acuerdo a la leyenda, la probabilidad comenzó en 1654 cuando el jansenista Pascal resolvió los dos célebres problemas que le propuso el mundano Caballero de Méré y envió su solución a Fermat (en realidad, los dos problemas llevaban ya algún tiempo en circulación entre los estudiosos de la época). Lo que sí es verdad es que la segunda mitad del S. XVII es el tiempo del nacimiento de la probabilidad: en 1657 Huygens escribió el primer libro de texto sobre la probabilidad que se ha publicado. Por esas fechas Pascal hizo la primera aplicación de razonamiento probabilístico a problemas distintos de los juegos de azar e «inventó» la teoría de la decisión: el pensador francés no duda en apostar por la existencia de Dios, ya que, por pequeña que sea la probabilidad de que ello ocurra, la ganancia es infinita si es que, en efecto, Él existe; por lo que tal juego tiene la propiedad de tener una esperanza positiva. Aunque Pascal estableció estas consideraciones con la intención de predicar la religión, muestran como subproducto algo interesantísimo desde la perspectiva de la matemática: establecen el modo en que la aritmética aleatoria puede ser parte de un arte de razonamiento general y hacen posible comprender que la estructura de pensamiento sobre juegos de azar se puede transferir a una teoría de la inferencia que no está basada en un escenario de azar. En el libro *Logic* de *Port Royal* se mencionan medidas numéricas de algo que hoy día se llama probabilidad. Simultánea pero independientemente, Leibniz pensaba en aplicar una métrica de las probabilidades a problemas legales y en desarrollar la combinatoria. John Graunt publicó en 1662 el primer conjunto extenso de inferencias estadísticas extraídas de los registros de mortalidad.

La probabilidad que se desarrolla en tiempos de Pascal es esencialmente dual: tiene que ver, a la vez, con frecuencias estables a largo plazo y con grados de creencia; es simultáneamente, estadística y epistemológica. La dualidad de la probabilidad está bien ilustrada por los fundadores de la teoría: el problema de Pascal de dividir el dinero de una apuesta cuando hay que interrumpir el juego, es de naturaleza aleatoria; su argumento de decisión sobre la existencia de Dios es de grado de creencia. Huygens escribió sobre todo de problemas aleatorios. El *Logic* finaliza con una discusión sobre el concepto de creencia razonable. ¿A qué necesidad histórica se debió que estas dos familias de ideas fácilmente distinguibles confluyeran en una sola? ¿Cómo se hizo posible este concepto dual de probabilidad?

Hacking (1975) afirma que los filósofos han analizado esta dualidad de la probabilidad desde hace tiempo: Carnap distinguía entre probabilidad inductiva y probabilística; Poisson aprovechaba las palabras *chance y probabilité* para hacer la misma distinción; Condorcet sugirió facilidad para el concepto aleatorio y motivo de creencia para el concepto epistemológico; Russell usó credibilidad para el último.

En el enfoque epistemológico de la probabilidad hay dos escuelas de pensamiento dominantes: 1) en las primeras décadas de este siglo, se prestó mucho interés a la teoría avanzada por Jeffreys (1933), según la cual la probabilidad conferida a una hipótesis por algún tipo de evidencia es una relación lógica entre dos proposiciones: la probabilidad de h a la luz de e es el grado en el que e implica lógicamente a h; 2) por otro lado está la teoría que De Finetti (1937) llamó probabilidad personal o subjetiva; en esta teoría la probabilidad que tú asignas a una proposición particular depende de tu propio juicio personal, pero el conjunto de todas tus asignaciones de probabilidad debe estar sometido a reglas rigurosas de coherencia interna. Independientemente de aceptar la teoría lógica o personal, ambas son plenamente epistemológicas, interesadas en la credibilidad de proposiciones a la luz de un juicio o evidencia.

En el enfoque aleatorio de la probabilidad hay una familia de teorías estadísticas que se centran en el estudio de la tendencia que muestran algunos fenómenos experimentales o naturales, a producir frecuencias estables a largo plazo en ensayos repetidos. La probabilidad de salir «caras» es una propiedad de la moneda como lo es su masa, y la estabilidad de las frecuencias en ensayos repetidos es un hecho objetivo de naturaleza independiente del conocimiento de cualquier persona sobre ello.

Es interesante analizar el caso de Jacques Bernoulli (1654-1705) que es visto como subjetivista por unos, como logicista por otros y frecuencialista por otros. Se le ha llamado subjetivista porque introdujo la palabra 'subjetivo' al reflexionar sobre la probabilidad; otros dicen que anticipa la teoría de probabilidades logicista de Carnap y por fin, hay algunos que le consideran el precursor de la versión frecuencialista en virtud de su ley de los grandes números. Aunque se considera que estas tres concepciones de la probabilidad son virtualmente incompatibles, muchos investigadores afirman que se pueden encontrar los orígenes de todas ellas en el trabajo de Bernoulli. La verdad de la cuestión puede ser que se sintió atraído por todas y cada una de esas ideas aparentemente incompatibles pero que suponen, cada una de ellas, una interpretación específica de la probabilidad. En todo caso, conviene señalar el dato significativo de que las teorías de hoy ya se pueden distinguir en el nacimiento del concepto de probabilidad.

En el Renacimiento lo que se llamó entonces probabilidad era un atributo de opinión y se contraponía a conocimiento que se podía obtener sólo mediante demostración. Así, surgió una dualidad entre ciencia (conocimiento) y opinión (creencia); había «altas ciencias», como matemáticas, mecánica, astronomía y filosofía, que buscaban verdades absolutas y «bajas ciencias», como medicina, astrología y alquimia, que producían opiniones basadas en evidencia empírica (Hacking, 1975). Galileo consideró la probabilidad como «ciencia baja», basada en la opinión. Una opinión, en principio, tenía tanto peso como cualquiera otra y sólo era más probable si estaba soportada por alguna autoridad; por ejemplo, en el caso de enfoques contrapuestos a un problema, la solución tenía que basarse en opiniones de las escrituras y en las enseñanzas de la Iglesia. Para que emergiese nuestra moderna versión matemática de la probabilidad, tenía que cambiar el concepto de lo que se consideraba evidencia aceptable. Antes del S. XVII se consideró la probabilidad como una materia de aprobación más que un cálculo matemático. Finalmente, la idea de evidencia experimental empezó a ganar respetabilidad en el S. XVII gracias a los trabajos de Pascal y Huygens. Mientras las ciencias clásicas intentaban deducir efectos a partir de Primeras Causas, la nueva ciencia intentaba inducir causas a partir de efectos observados. Aquí descansan las semillas de nuestra estadística.

Queremos señalar otra característica de la disciplina estadística que puede haber sido un obstáculo para el desarrollo temprano de conceptos formales de probabilidad. Ante una situación de incertidumbre, gobernada por los dados que interpretan el juicio divino, tiene para el hombre el mayor interés el siguiente resultado, el número que saldrá en el siguiente lanzamiento. El hombre no puede conseguir ninguna predicción definitiva y sólo puede especular a partir de patrones de resultados previos o confiar en la divina voluntad. Sin embargo para progresar en la formalización del concepto de probabilidad hay que considerar el siguiente resultado sólo como representativo de resultados futuros o hipotéticos. Sólo esta transformación del problema lo hace abordable pero no da una respuesta a la cuestión original. La probabilidad de 1/6 no dice nada acerca de que número se obtendrá realmente y si saldrá un «cinco» en la siguiente tirada de un dado. Sin embargo y sorprendentemente, la probabilidad de 1/6 constituye algún conocimiento indirecto para una tirada específica. Este aspecto de la probabilidad, todavía motivo de debate filosófico, es un obstáculo importante para la comprensión de los alumnos.

Con todo, no conviene sobreestimar la tardía conceptualización de la probabilidad porque esto también ocurre en otras disciplinas. El desarrollo científico basado en un enfoque físico causal enfrentado a un enfoque deístico estuvo marcado por grandes controversias incluso sobre las ideas que hoy nos parecen más naturales (no hay más que recordar los problemas de Galileo con la Iglesia Católica). La geometría euclidiana no supuso una temprana clausura conceptual ni de la geometría ni del pensamiento axiomático, tal como a veces se dice, en cuanto que sólo proporcionó reglas de construcción pero no conceptos formales; el problema del axioma de las paralelas no se clarificó hasta los trabajos de Gauss y Lobachevski en el S. XIX. En aritmética no se consiguió una axiomatixación de los números hasta Peano, apenas hace 100 años. También es verdad que un nivel de desarrollo conceptual similar en probabilidad se alcanzó todavía más tarde puesto que la axiomatización de la probabilidad data de 1933.

1.2. Hitos en la historia de la probabilidad

Después de analizar el hecho crucial de la emergencia tardía y dual de la probabilidad y de sus posibles causas, conviene volver sobre nuestros pasos. Vestigios de situaciones probabilísticas se pueden encontrar en las antiguas culturas de la India, Babilonia y Egipto. Entre los objetos más antiguos que se conocen usados en juegos de azar se encuentra el astrágalo, un hueso pequeño del pie. Los soldados romanos jugaban con esos huesos. Es posible que los primitivos dados se hicieran alisando las superficies rugosas del astrágalo hasta que estuviesen regulares. Con todo, mientras se usaron huesos auténticos no se podía garantizar la regularidad de la caída puesto que influía el tipo de hueso de animal que se usaba y su desgaste.

Por contra, los cubos de cerámica que se usaron en Babilonia 3000 años A.C. eran dados casi perfectos. Es natural pensar que se obtuvo considerable experiencia estadística del lanzamiento de dados o de la extracción de judías o granos contenidos en urnas que se realizaban en ceremonias religiosas; sin embargo, el progreso conceptual basado en la regularidad de la caída del dado fue muy lento. Es posible que se enseñase a los sacerdotes a manipular la caída del dado para lograr el resultado deseado como la interpretación del juicio divino y que se considerase impío y por tanto susceptible de castigo cualquier especulación en las leyes del azar en cuanto suponían intromisión en los misterios de la deidad.

Se atribuye a Cardano la primera referencia al proceso de abstracción que va desde la experiencia aleatoria al concepto teórico de probabilidad; por primera vez se encuentra una idealización explícita de equiprobabilidad basada en la abstracción de un dado normal. Cardano (1501-1576) analiza el lanzamiento de un dado en Liber d*e ludo aleae*: «La mitad del número total de caras representa la igualdad; así, son iguales las probabilidades de que un punto dado salga en tres tiradas, cuando el circuito total es de 6, y de que salga uno de tres puntos dados en una tirada. Por ejemplo, puedo conseguir tan fácilmente 1, 3 ó 5 como 2, 4 ó 6. Las apuestas hay que hacerlas de acuerdo con esta igualdad si el dado está equilibrado» (citado en Hacking, 1975, p. 54).

Por «son iguales las probabilidades de que un punto dado salga en tres tiradas...», Cardano parece referirse a lo que hoy en día se expresa mediante el concepto de esperanza matemática ($np=3.1/6=1/2$); en este sentido sí que son iguales las dos cantidades. Sin embargo, la probabilidad de conseguir al menos una vez un punto dado en 3 tiradas es $1-(5/6)^3=91/216$, que es menor que 1/2 (la probabilidad de que salga uno de tres puntos dados en una tirada). El argumento de Cardano es por tanto, un híbrido de equiprobabilidad y esperanza. Con todo, es dudoso que realmente hubiera hecho una abstracción desde las frecuencias empíricas al concepto teórico de probabilidad puesto que no intentó definir el concepto explícitamente, más bien parece que Cardano simplemente evalúa una probabilidad específica.

Un siglo después Pascal (1623-1662) y Fermat (1601-1665) consiguieron un auténtico y crucial progreso en la conceptualización de la probabilidad como denota su famosa correspondencia de 1654 (que no fue publicada hasta 1679) donde aparecen resueltos dos problemas específicos planteados por de Méré. La historia de la relación entre Pascal (el austero jansenista) y el Caballero de Méré (el hombre de mundo) resulta tan familiar que es, quizá, el único acontecimiento en la historia de la probabilidad que se puede considerar de conocimiento general. Mucho menos conocida es la relación entre Pascal y Roannez (otro aristócrata mundano y con talento matemático) a quién Hacking llama «la partera del cálculo de probabilidades»; el círculo de Roannez que engloba a Pascal, Fermat, de Méré y el propio Roannez, es el caldo de cultivo primigenio donde surge la teoría de probabilidades matemática (un buen nacimiento, el placer del cuerpo y el rigor del espíritu en fructífero contubernio).

El primer problema que de Méré planteó a Pascal fue: ¿Al lanzar dos dados, cuántos lanzamientos son necesarios para tener una probabilidad de 0.5 de conseguir al menos un «doble seis»? Méré tenía dos respuestas posibles, 24 ó 25; la primera respuesta la basaba en un cálculo aritmético mientras la segunda la fundamentaba en la experiencia; pensaba que su problema mostraba una contradicción de la propia aritmética. El razonamiento del Caballero de Méré era: Consideremos una situación en la que tenemos 1 posibilidad entre N de ganar en un juego. Sea n el número de ensayos requeridos para tener la probabilidad 0.5 de ganar. Aparentemente la regla que aplica de Méré es que n/N es constante: por ejemplo, en el caso de un dado con el que intentamos conseguir un «seis», N es 6 y n es 4, es decir, n/N es 2/3. En el lanzamiento de dos dados, N es 36 y por tanto n debe ser 24. Pascal mediante una enumeración exhaustiva de posibilidades mostró que la probabilidad de conseguir un «doble seis» en 24 tiradas de dos dados es 0.491 mientras que en 25 tiradas es 0.505.

Este razonamiento del Caballero es de rabiosa actualidad porque nuestros alumnos también se dejan llevar por el sesgo de «la regla de tres», si se les propone la siguiente cuestión: «Hay que decidir entre dos juegos. En el juego 1, el jugador gana si hay al menos un «seis» en 4 tiradas de un dado; en el juego 2, el jugador gana si hay al menos un «doble seis» en 24 tiradas de dos dados. ¿Que juego prefieres? «La solución normativa es:

$$P(\text{ganar en el juego 1}) = 1-(5/6)^4 = 671/1296 = 0.508 > 0.5$$
$$P(\text{ganar en el juego 2}) = 1-(35/36)^{24} = 0.491 < 0.5$$

Los alumnos, como haría de Méré, rechazan esta solución y defienden la equivalencia de los dos juegos mediante el argumento de que 24 es a 36 (el número de casos para dos dados) como 4 es a 6 (el número de casos para un dado). Forma parte de la leyenda probabilística que de Méré ganó gracias al juego 1 pero perdió toda su fortuna en el juego 2. Es difícil de creer que los jugadores observaran una diferencia en las probabilidades de los dos juegos a pesar de su gran experiencia práctica. El razonamiento de de Méré se basa quizá en un conflicto teórico entre una enumeración directa del conjunto de probabilidad fundamental (CPF) y la regla de casos favorables a posibles. Esta regla produce soluciones correctas si es aplicada a sucesos favorables que son elementos simples del espacio muestral. Pero en los juegos anteriores se aplica la regla a series de 4 ó 24 ensayos como casos favorables los cuales no son, evidentemente, elementos del mismo espacio muestral. Además hay una confusión con el valor esperado del juego: el número esperado de «seises» en una serie de 4 ensayos es 4.(1/6); el número esperado de «dobles seises» en una serie de 24 ensayos es 24.(1/36), por tanto el número esperado de «éxitos» es igual en ambos juegos.

El segundo problema que de Méré planteó a Pascal (el problema de la división de premios) se refiere al reparto equilibrado de premios si un juego tiene que pararse o finalizarse antes de lo previsto. Al comienzo del juego dos jugadores A y B apuestan la misma cantidad; por ejemplo, se trata del lanzamiento sucesivo de una moneda normal y A apuesta a «caras» y B a «cruces». El jugador que gane primero un cierto número de puntos, fijado de antemano, gana la cantidad total apostada. Sin embargo, el juego tiene que ser interrumpido antes que cualquiera de los jugadores haya alcanzado el número requerido de puntos y el premio tiene que dividirse. Si, por ejemplo, se requieren 5 puntos para ganar y la puntuación en el momento de parar el juego es 4 a 3 favorable al jugador A ¿cuál es la división razonable de premios? Este es un viejo y famoso problema desde el S. XIII del que se han dado numerosas soluciones, casi siempre no estadísticas.

Lo mismo ocurre siempre que hemos propuesto este problema a nuestros alumnos. La mayoría de sus soluciones supone realizar un reparto proporcional a 4 y a 3. Les recordamos que el jugador que hubiese ganado se habría llevado todo el dinero, independientemente de los juegos que hubiese ganado el perdedor y que, por lo tanto, no es razonable hacer tal reparto proporcional. También les insistimos en que el reparto no depende de lo que «ha pasado» sino de lo que «puede pasar». Pero todas estas reflexiones no les sirven de mucha ayuda, aunque siempre hay algún alumno que llega a la solución normativa: tres partes para A y una para B.

Pascal y Fermat basaron su enfoque del problema estableciendo el escenario de lo que sucedería si el juego continuaba y si las probabilidades de los jugadores eran iguales en cada ensayo o repetición del juego. Los premios deberían dividirse proporcionalmente a la probabilidad de ganar si el juego continuaba hasta el final, es decir, A debería llevarse los 3/4 del premio y B el resto. En efecto:

P(gane jugador B) = P(B gane dos siguientes repeticiones del juego) = (1/2).(1/2) =1/4

P(gane jugador A) = 3/4

Si nos hemos detenido en el análisis de estos dos históricos problemas es por un doble motivo: en primer lugar, porque muestran cómo problemas que exigen un sencillo cálculo de probabilidades pero resultan muy difíciles para nuestros alumnos también lo eran para los ilustres pioneros de la teoría de probabilidades; en segundo lugar, porque muestran cómo la relación aritmética-probabilidades no siempre es tranquila: en estos dos problemas, la fuerte tendencia de los alumnos (y también del Caballero de Méré) a utilizar la regla de tres o el reparto proporcional dificulta la comprensión probabilística de ambos problemas. Es una señal del cuidado educativo que hay que tener porque la intuición estadística no se adiestra naturalmente, implícitamente, en la escuela sino, quizá, todo lo contrario. En todo caso, el tipo de razonamiento que hay que utilizar para resolver los dos problemas, es justamente el que nos gustaría que usase una persona que haya tenido varios años de entrenamiento en matemáticas y teoría de probabilidades.

Aunque Cardano y Galileo realizaron cálculos probabilísticos y Pascal y Fermat exploraron interesantes problemas de probabilidad, el científico que sintetizó ideas de modo sistemático y que realizó generalizaciones desde las soluciones de problemas fue Christian Huygens (1629-1695); él fue el primero que estableció formalmente la idea de esperanza matemática, por ejemplo. Su libro De *ratiociniis in aleae ludo* fue publicado en 1657 y no fue reemplazado durante más de 50 años hasta que Jacques Bernoulli lo incorporó en parte a su obra maestra *Ars conjectandi* (1713).

Puede parecer que la noción de esperanza matemática debía emerger más fácilmente que la noción de probabilidad. Desde una perspectiva aleatoria, la esperanza es la ganancia media en una larga serie de juegos similares. Podemos «ver» realmente las ganancias o pérdidas de una opción persistente. Traducimos la ganancia total en ganancia media y observamos la esperanza con mucha más facilidad que la probabilidad. Sin embargo, el mismo concepto de media es nuevo en 1650; antes de esa fecha, un jugador podía notar que una estrategia era más ventajosa que otra pero hay un salto entre este hecho y el conocimiento cuantitativo de la esperanza matemática.

El libro de Huygens sobre los juegos de azar tiene el mismo objetivo de rigor que un tratado moderno y llega a deducir mediante una forma muy elaborada de razonamiento que el valor de un juego donde hay p posibilidades de obtener a y q posibilidades de obtener b, equivale a $(pa+qb)/(p+q)$. Aunque el autor holandés no habla de esperanza (una denominación que surge de la traducción latina de su libro donde aparece el término *expectatio)*, tiene el mérito de haber usado este concepto. Utilizó la probabilidad como un concepto elemental no definido y lo justificó en referencia a los juegos reales de azar. Con Huygens se desarrolló una rama de aplicaciones estadísticas de la probabilidad en cuanto que estableció tablas de mortalidad, definió conceptos teóricos como tiempo medio de vida y trató las frecuencias de la misma manera que las probabilidades.

Los problemas que aparecen al final del libro fueron objeto de estudio durante varias generaciones de probabilistas. Algunos problemas son ambiguos lo que refleja de nuevo la dificultad de interpretación de los enunciados probabilísticos y lo mal establecido que estaba el propio lenguaje de la teoría de probabilidades. Por ejemplo, un problema decía: «Tres jugadores A, B y C meten 12 fichas en una bolsa de las que 4 son blancas y 8 negras. El ganador es quién primero extraiga una ficha blanca. El orden en que extraen los jugadores es primero A, luego B, después C, después A y así hasta terminar ¿Qué relación hay entre las probabilidades de ganar que tiene cada jugador?». Jacques Bernoulli ya señaló que hay al menos tres diferentes interpretaciones: primera, cada vez que se extrae una ficha negra se devuelve a la bolsa; segunda, las extracciones son sin reemplazamiento; tercera, podemos suponer que cada uno de los tres jugadores comienza con su propia bolsa de doce fichas y las va extrayendo sin reemplazamiento. En la correspondencia que se cruzó Huygens con otros estudiosos de la época aparece la ambigüedad de la interpretación de unos y otros aunque parece que Huygens se inclina por la primera interpretación.

Hay otra consideración sobre el concepto de esperanza matemática que tiene que ver con el concepto de esperanza de vida y que es muy instructiva sobre la sutileza de los conceptos probabilísticos. En 1662, John Graunt usó los datos de natalidad y mortalidad de Londres para realizar inferencias acerca de diversas variables (llegó a establecer, por ejemplo, que la proporción de nacimientos de niños/niñas era 1.05) . Aunque Huygens recibió una copia del libro no le concedió mucha importancia hasta que su hermano Ludwig, que había leído el libro de Graunt, le preguntó cual sería la esperanza de vida de un niño recién nacido según las tablas de Graunt. Como no conoce el término esperanza de vida, Ludwig escribe: *la question est jusqu'à quel âge doit vivre naturellement un enfant aussitot qu'il est conçue.* El propio Ludwig realiza un cálculo del tipo $(ap+bq)/(p+q)$ y llega a la conclusión de que 18.2 años es la esperanza de vida de un niño recién concebido.

Christian le matiza este resultado explicando que aunque la esperanza de vida sea 18.2 años, esto no significa que se espera que la mayoría de los recién nacidos vivan 18.2 años sino que la mayoría de ellos morirán bastante antes: «Imagina que las personas fueran todavía más débiles en su infancia que lo son ahora y que 90 de cada 100 muriesen antes de los 6 años pero los que pasasen esta edad fuesen matusalenes y viviesen en media 150 años». En este caso la esperanza de vida de un recién nacido sería aproximadamente de 18 años pero cualquiera que apueste a que un recién nacido concreto no pasará de los 6 años tiene una enorme ventaja sobre otro que apueste lo contrario. La dificultad del problema surge del propio enunciado oscuro de Ludwig, *jusqu'à quel âge on doit vivre naturellement.* Hoy día, la gran disminución de la mortalidad infantil hace que la edad esperada de vida y la edad mediana de vida estén muy próximas pero en el siglo XVII la edad esperada era 18.2 años mientras que la edad mediana era 11 años, según los datos de Graunt. De la correspondencia de los dos hermanos parece deducirse que aunque Ludwig dio la respuesta de 18 años el número que realmente buscaba era 11.

Leibniz (1646-1716) no hizo una contribución formal fundamental a la teoría de probabilidades pero tuvo un profundo interés en el tema. Hacking (1975) afirma que fue el primer filósofo de la probabilidad. Fue el primero en decir que la teoría de la probabilidad podía ser una rama de la lógica comparable a la teoría de la deducción y la intentó axiomatizar como una ciencia inferencial pura. Antes de dominar el trabajo de Pascal, de Huygens y de otros matemáticos, había intentado desarrollar una aritmética de la probabilidad que no estaba basada en juegos de azar y por tanto tenía más aplicaciones potenciales. Escribió la primera monografía de la teoría combinatoria (Ars *Combinatoria*) y observó su relación con la teoría de probabilidades. Predijo que una teoría de juegos generalizada debería ser el fundamento para la toma de decisiones en situaciones de incertidumbre.

De nuevo, y con el pensamiento puesto en nuestros sufridos alumnos, debemos insistir en la dificultad de los problemas probabilísticos, aún en los aparentemente fáciles. Es arriesgado cuestionar el rigor de razonamiento de Leibniz y sin embargo, uno de nuestros mejores matemáticos «se despistó» al calcular probabilidades en el lanzamiento de dos dados. Leibniz creía que las sumas 11 y 12 tenían iguales probabilidades, porque cada una de ellas puede conseguirse solamente con una combinación de dos dados; su fallo fue no darse cuenta de que 12 sólo puede hacerse de una manera (6+6), mientras que 11 puede hacerse de dos maneras (5+6 ó 6+5), siendo de esta forma doblemente fácil lograr la suma 11 que la 12.

El libro de J. Bernoulli *Ars conjectandi* presenta las innovaciones conceptuales más decisivas en la historia temprana de la probabilidad. El autor estuvo trabajando en el libro durante 20 años y aunque probó el teorema clave (la ley de los grandes números) en 1692 no quedó satisfecho y no lo publicó. Por fin el libro fue publicado a título póstumo en 1713 por su sobrino Nicolás. Tiene cuatro partes. La primera es una versión mejorada del libro de Huygens sobre juegos de azar; Bernoulli tiene un gran talento para dar explicaciones intuitivas de conceptos técnicos, así, explica la esperanza matemática como la esperanza de conseguir lo mejor menos el temor de conseguir lo peor y presenta de modo gráfico e impactante la ley de la adición de probabilidades para sucesos disjuntos. En la segunda parte Bernoulli presenta la teoría de combinaciones y en la tercera aplica los resultados encontrados a la resolución de nuevos problemas sobre juegos de azar. Es en la última parte del libro, titulada «Aplicaciones de lo anterior a problemas económicos, morales y civiles», donde Bernoulli revoluciona la teoría de probabilidades; la revolución es doble: por primera vez se declara explícitamente una concepción subjetiva de la probabilidad y se prueba el primer teorema límite.

La primera ley de los grandes números, que Bernoulli llamó *teorema aureum*, supuso un decisivo progreso conceptual en cuanto que estableció el fundamento sólido para enlazar las frecuencias relativas y las probabilidades. El problema planteado es el siguiente: supongamos que lanzamos, sucesivamente, una moneda con probabilidad p de salir cara ¿Qué podemos decir sobre la frecuencia relativa de caras en una sucesión larga de lanzamientos de la moneda? Este problema es el origen de los teoremas límite en probabilidades: las leyes de los grandes números y el teorema central del límite. J. Bernoulli demuestra que la frecuencia relativa de caras en n lanzamientos de una moneda regular (p=1/2) «converge» a 1/2. La convergencia de Bernoulli (convergencia en probabilidad) tiene el siguiente sentido:

$$P[\,|(\text{n}^\circ\ \text{caras}/n) - 1/2\,| > \varepsilon\,] \to 0 \text{ cuando } n \to \infty\ \forall\varepsilon > 0$$

En definitiva, este teorema prueba que las frecuencias relativas convergen en probabilidad a la probabilidad subyacente. Bernoulli afirma que este teorema nos permite averiguar a posteriori lo que no podemos determinar a priori, esto es, averiguarlo a partir de los resultados observados en numerosos ejemplos similares. Laplace (1749-1827) generalizó el resultado de Bernoulli a cualquier valor de p. Borel en 1909 lo generalizó a un tipo de convergencia más fuerte (convergencia casi segura).

Bernoulli da un giro desde el concepto de esperanza de Huygens al de probabilidad que se convierte en el concepto central y se enriquece con un tratamiento sistemático de la combinatoria. Sus ideas filosóficas, sin embargo, todavía se pueden definir como un determinismo metafísico. Todos los fenómenos (tiempo atmosférico, el lanzamiento de dados, los eclipses, etc.) están gobernados por leyes determinísticas. El azar sólo se explica por nuestro limitado conocimiento de esas leyes y las aplicaciones de la probabilidad se deberían restringir a los juegos porque la mayoría de los fenómenos son tan complejos que es inútil estudiar los casos posibles que conforman el espacio muestral.

Hay algunas indicaciones de que el concepto de distribución de probabilidad aparece en el S. XVIII. De Moivre (1667-1754) fue el primero en encontrar la función que hoy se llama la función de densidad normal: estudiando cómo en n lanzamientos de una moneda regular, se desvía la frecuencia relativa de caras de la probabilidad teórica de 1/2, obtuvo la primera versión del teorema central del límite y la Ley Normal. En términos modernos, buscaba una distribución de las desviaciones de las frecuencias relativas H/n de la probabilidad subyacente p para un tamaño de muestra fijado n. Lo resolvió para $p=1/2$ deduciendo la distribución límite para $n \to \infty$. Para de Moivre la densidad normal servía sólo como una herramienta para la aproximación numérica y no tenía significado probabilístico propio.

Es curioso el hecho de que hay vestigios de inferencia estadística antes de cualquier intento fundamentado de definición del concepto de probabilidad y que estos vestigios aparecen en relación a la más importante y repetida de las experiencias aleatorias cual es el nacimiento de un niño. Arbuthnot (1667-1735) analizó las estadísticas de nacimientos de Londres durante 80 años sucesivos y encontró que nacían más niños que niñas cada año. Si la probabilidad de nacimiento de un varón fuese $p=1/2$, la probabilidad de un suceso tal sería muy pequeña. Por ello rechazó la hipótesis $p=1/2$, sustituyéndola por la hipótesis $p>1/2$. Esta fue, quizá, la primera prueba de significación. La justificación de $p>1/2$ para Arbuthnot, estaba en la voluntad divina de compensar el mayor número de fallecimientos de los hombres como consecuencia de accidentes laborales y guerras, con el fin de mantener el equilibrio entre los sexos y asegurar así la monogamia. Si Dios no existe, argumentó, no hay motivo especial para que no sean iguales las probabilidades de niño o niña, por tanto los datos empíricos confirman «la Voluntad de Dios en acción».

Buffon (1707-1788) utilizó un argumento similar para probar que los planetas se originaron de una causa común que supuso fue la colisión del sol con un cometa. En definitiva, este tipo de argumento consiste en evaluar una hipótesis H mediante un suceso observado realmente E, vía la probabilidad condicional $p(E/H)$. Si esta probabilidad es pequeña entonces se rechaza la hipótesis H. Hoy en día, este tipo de argumento no se usa para evaluar una hipótesis simple sino sólo para comparar la plausibilidad de hipótesis competitivas.

El trabajo de Laplace (1749-1827) marcó una culminación en el desarrollo conceptual antiguo de la probabilidad; con él comenzó la edad moderna de la probabilidad. Filosóficamente, sin embargo, su enfoque se basaba todavía en un determinismo mecanicista: «Una inteligencia que comprendiese todo...nada le resultaría incierto y tanto el futuro como el pasado, estarían presentes ante sus ojos... La probabilidad es relativa en parte a (nuestra) ignorancia y en parte a nuestro conocimiento» (Laplace, 1985, p. 27).

Laplace dio la primera definición explícita de probabilidad, la llamada probabilidad clásica: la probabilidad p(A) de un suceso A es igual a la proporción del número de resultados que son favorables al suceso A en relación al número de todos los resultados posibles de la prueba. Esta definición asume implícitamente que los resultados individuales son equiprobables. Laplace formuló el «principio de razón insuficiente» para operativizar la regla; según este principio, debemos asumir que los resultados son equiprobables si no tenemos razón para creer que alguno de los resultados es más probable que otro. Esta primera definición formal no clarifica la naturaleza de la probabilidad en cuanto que para su operatividad se refiere a un principio oscuro filosóficamente y tiene un dominio de aplicación que no engloba los problemas reales. Los intentos posteriores que se hicieron para corregir este principio (basado en consideraciones de indiferencia o invariancia) no tuvieron éxito. Había algunos problemas e ideas que dificultaban el progreso conceptual.

Bayes (1702-1761) había justificado el uso de una distribución uniforme en el intervalo (0,1) para un parámetro binomial p ante la completa ignorancia de los resultados de un experimento de Bernoulli. Su argumentación la utilizó Laplace para formular el principio de razón insuficiente como una guía básica para aplicar su definición de probabilidad. El problema es que en el caso extremo de ignorancia completa, el principio de Laplace se podría utilizar para establecer una equiprobabilidad de todos los casos posibles lo que es claramente una clase de información. Si se toma por absolutamente válido, este principio produce una regla muy ambigua que transforma ignorancia en conocimiento.

El intento de representar los casos equiprobables de un modo auténticamente objetivo causa dificultades. Una manera de justificar el principio de Laplace es buscar simetrías físicas del fenómeno aleatorio en cuestión, por ejemplo, la simetría física del dado conduciría directamente a la equiprobabilidad de sus caras. Sin embargo, hay muchas simetrías físicas posibles, por tanto, una teoría verdaderamente objetiva requiere un procedimiento para elegir una simetría particular y justificar esa elección. Fine (1973) ilustra las dificultades con el ejemplo de la experiencia aleatoria de lanzar dos dados, donde se pueden plantear al menos tres modelos:

Modelo de Maxwell-Boltzmann: los 36 pares (1,1), (1,2),..., (6,5),(6,6), son igualmente probables, de modo que los pares como (2,3) y (3,2) son resultados diferentes.

Modelo de Bose-Einstein: los 21 pares (1,1), (1,2) ,..., (1,6), (2,2), (2,3),....(5,5),(5,6),(6,6), son igualmente probables, de modo que los pares como (2,3) y (3,2) se tratan como resultados idénticos.

Modelo de Fermi-Dirac: los 15 pares (1,2),(1,3), ..(1,6),(2,3),(2,4),..., (5,6), son equiprobables de modo que se eliminan las parejas con los dos componentes iguales.

Para los dados ordinarios, la estadística de Maxwell-Boltzmann es el modelo natural; los dos dados son discernibles (dado azul y dado rojo o dado primero y dado segundo, etc.) y de ese modo la hipótesis de independencia es altamente plausible. Este modelo natural, sin embargo, no es verdad para muchas aplicaciones en física. Según Feller (1973) se han hecho numerosos intentos para probar que las partículas físicas se comportan de acuerdo con la estadística de Maxwell-Boltzmann pero la teoría moderna ha mostrado que esta estadística no se aplica a algunas partículas conocidas. En efecto, la estadística de Bose-Einstein parece apropiada para fotones, núcleos y átomos y la estadística de Fermi-Dirac es apropiada para electrones, protones y neutrones.

Así, el modelo natural para los dados y su simetría subyacente es, con frecuencia, inapropiado para representar fenómenos con partículas físicas lo que reduce considerablemente su campo de aplicación. El ejemplo también muestra que hay diferentes posibles simetrías en la misma situación física; es interesante notar que sobre esta cuestión también surgen problemas pedagógicos. La probabilidad no puede ser una característica inherente de los objetos reales sino sólo un resultado de nuestra tarea de modelar la realidad. El empeño en deducir probabilidades únicas ha llevado a paradojas sin solución. Fine (1973) resume su crítica a la definición de Laplace afirmando que no podemos extraer información (una distribución de probabilidad) de la ignorancia, que el principio de razón insuficiente es ambiguo y su aplicación lleva frecuentemente a inconsistencias y que, en definitiva, el enfoque clásico de la probabilidad ni es una teoría objetiva ni una teoría empírica.

Como cuenta Martín Gardner (1984), los griegos y romanos preferían jugar con tres dados y Platón en sus Leyes (libro 12) menciona que 3 y 18 son, en este caso, las sumas más difíciles por ser las únicas que sólo pueden conseguirse de una forma: (1, 1, 1) y (6, 6, 6); los griegos llamaban «el perro» a la primera forma y «Afrodita» a la segunda. También, en los siglos XVII y XVIII, era más común jugar con tres dados que con dos y por tanto el problema planteado históricamente fue: ¿Cuántas alternativas iguales surgen al lanzar tres dados? Este problema combinatorio dio lugar a interpretaciones muy interesantes: ya hemos comentado que Leibniz cometió el error de considerar que se forma el espacio muestral de resultados igualmente probables con las combinaciones y no con las variaciones; esto es, aceptó la estadística de Bose-Einstein para los dados. Galileo se inclinó por la estadística de Maxwell-Boltzmann con la siguiente argumentación: con tres dados, hay el mismo número de combinaciones que sumen 9 y 12 y que sumen 10 y 11; exactamente hay 6 combinaciones. Sin embargo, se sabe por observaciones sistemáticas de gran número de lanzamientos de tres dados que las sumas 10 y 11 son más ventajosas que las sumas 9 y 12; la explicación es simple, a saber, que las 6 combinaciones que producen la suma de 9 ó 12 se pueden descomponer en 25 variaciones mientras que las 6 combinaciones de 10 y 11 se descomponen en 27 variaciones. Si las variaciones son igualmente probables, entonces 11 es más ventajoso que 12 en la proporción de 27:25.

El argumento de Galileo (1564-1642) parece el primer caso de refutación de una hipótesis estadística por observación a largo plazo. Se tiene la hipótesis de que las combinaciones son equiprobables *versus* la hipótesis de que lo son las variaciones; la primera es inconsistente con los hechos, mientras que la segunda se ajusta a los hechos perfectamente. Desde luego, hubiera sido más simple contrastar las hipótesis observando las frecuencias relativas de las sumas 4 y 3 que se consiguen con una única combinación (1+1+2 ó 1+1+1, respectivamente) pero, a su vez, la suma 4 se consigue con tres variaciones (1+1+2, 1+2+1, 2+1+1) y la suma 3 con 1 variación. Sin embargo, en juegos estándar de 3 dados, tanto la suma 4 como 3 son raras, difíciles de conseguir, por tanto no es fácil registrar la experiencia a largo plazo, y sí lo es con las sumas 9, 10, 11 y 12 que ocurren con mayor frecuencia.

Volviendo a Laplace, su Teorema Central del Límite supone un avance estadístico crucial. Este teorema enuncia, en esencia, que la distribución binomial se aproxima a la distribución normal cuando el número de ensayos se incrementa al infinito. Laplace creía que la ley normal podría jugar un papel análogo a la ley de gravitación universal que explica la mayoría de los fenómenos celestes. Cualquier variable general podría ser explicada por la ley normal descomponiéndola en una suma de cantidades aditivas, las distribuciones de las cuales podrían incluso ser desconocidas.

El argumento intuitivo de Laplace sobre la universalidad de la distribución normal fue muy pronto recogido por otros autores. Gauss (1777-1855) usó la distribución normal no sólo como una herramienta para la aproximación sino como una distribución en sí misma. Su enfoque estaba anclado en la teoría del error; al establecer la media como el más apropiado de los valores que reemplazan varias medidas repetidas de una cantidad desconocida, reconocía que primero había que conocer la distribución de los errores de medida. Quetelet (1796-1874) desarrolló la idea del hombre promedio en analogía a la teoría del error. Galton, pariente de Darwin y biólogo como él, escribió «Herencia Natural», en 1889, donde estableció la ley de regresión universal y simuló una demostración práctica de pruebas binomiales y del teorema central del límite. Había entre los investigadores de la época un entusiasmo romántico que Galton supo expresar con gran lirismo: «No conozco casi nada tan apto para impresionar la imaginación como la forma maravillosa de orden cósmico expresada por la 'Ley de Frecuencia del Error'. La ley hubiera sido personificada y deificada por los griegos, si la hubieran conocido. Reina con serenidad y con completa discreción entre la más amplia confusión. Cuanto más abultado el gentío y mayor la aparente anarquía más perfecto es su dominio. Es la suprema ley de la Sinrazón» (citado en Borovcnik, Bentz y Kapadia, 1991, p. 35).

Teniendo presente este entusiasmo es más fácil comprender el título de ley normal. El papel protagonista de esta ley no cambió ni siquiera cuando otras distribuciones como la de Maxwell tuvieron interés en física o cuando Pearson (1857-1936) investigó de modo sistemático otros tipos de distribuciones continuas; este matemático y abogado, influido por Galton, aplicó las probabilidades a la teoría de la evolución darwiniana y realizó estudios sobre la regresión y la correlación. La escuela rusa (Tchebychev, Markov,...) propuso varias generalizaciones del teorema central del límite aportando ideas de la teoría de la medida. Hoy día, hay un cierto decrecimiento de este protagonismo debido al auge de la estadística robusta, la estadística no paramétrica y el análisis exploratorio de datos.

La teoría de la probabilidad desarrolló un importante papel conceptual en física durante las últimas décadas del S. XIX porque algunas nuevas leyes físicas sólo podían describirse en términos probabilísticos (por ejemplo, el segundo principio de la termodinámica). Las aplicaciones estadísticas, especialmente la regresión y la correlación, culminaron en desarrollos biométricos. Sin embargo, este desarrollo está contrapesado por el hecho de que no había adecuada fundamentación salvo el intento de Laplace que, como hemos visto, tenía sus problemas. En el Congreso Matemático de París de 1900, Hilbert formuló un programa para la investigación matemática entre cuyas tareas principales estableció la de axiomatizar satisfactoriamente la probabilidad y la mecánica estadística.

Von Mises, en 1919, fue uno de los pioneros en el trabajo de axiomatización y se basó en la interpretación de la probabilidad como convergencia de frecuencias relativas, siguiendo el teorema de Bernoulli. Su enfoque no tuvo éxito; era demasiado complicado y los problemas filosóficos fueron abrumadores. Por ejemplo, el teorema de Bernoulli sobre la convergencia de las frecuencias a la probabilidad subyacente no implica la convergencia usual sino la convergencia en probabilidad. Entonces ocurre que, o bien esas dos probabilidades que entran en el teorema son del mismo tipo y por tanto no debe usarse una para la definición de la otra, o bien son de un tipo diferente que necesita ser clarificado y definido. Además, la definición de Von Mises se basa en una propiedad de aleatoriedad de secuencias que es un concepto oscuro desde un punto de vista filosófico y difícil de verificar en muchas aplicaciones. El único modo de clarificar la propiedad de aleatoriedad es por medio de la misma probabilidad lo que de nuevo marca una circularidad en el enfoque de Von Mises.

Fue Kolmogorov en 1933 quien finalmente formuló un sistema de axiomas de la probabilidad y dedujo los teoremas usuales que fueron reconocidos inmediatamente. Su enfoque significa la aplicación de principios extraídos de la teoría de la medida que habían ganado importancia al probar varias generalizaciones del teorema central del límite. El enfoque de Kolmogorov no clarificó lo que es la probabilidad, sólo elaboró las propiedades estructurales de la probabilidad y dejó la interpretación del concepto así definido de probabilidad como una cuestión abierta. Conviene señalar que este enfoque fue pensado principalmente como una justificación de la interpretación frecuencialista de la probabilidad (Kolmogorov, 1976). A pesar del éxito de la axiomatización de Kolmogorov, la controversia en fundamentos entre subjetivistas y objetivistas revivió poco más tarde con Jeffreys (1933) y De Finetti (1937).

2. Implicaciones didácticas de la historia y filosofía de la teoría de probabilidades. Paradojas y falacias probabilísticas de interés educativo

Aunque la matemática pretende tratar con verdades universales, su progreso no ha sido siempre diáfano y está salpicado de crisis fundamentales. La historia de la matemática ha revelado muchas paradojas interesantes, algunas de las cuales han servido de acicate de cambios importantes. Las paradojas y falacias, que abundan en la teoría de probabilidades, son instructivas y por eso nos interesan desde la perspectiva didáctica. Es un apasionante y novedoso enfoque que conecta muy bien con los aspectos psicológicos del aprendizaje de las probabilidades.

Hay muchas paradojas y falacias (que se entrecruzan y solapan) y por tanto no es posible dar un listado completo. Siguiendo a Borovcnik, Bentz y Kapadia (1991), hemos preferido analizar en detalle unas cuantas paradojas representativas e importantes para el proceso de enseñanza-aprendizaje de la teoría de probabilidades elemental. Se presentan en cuatro grupos: asignación de probabilidades, esperanza matemática, independencia y dependencia y pensamiento lógico *vs* pensamiento probabilístico.

2.1. Asignación de probabilidades

La asignación de probabilidades mediante la regla de Laplace (casos favorables/casos posibles) tiene dificultades intrínsecas como hemos visto al estudiar el lanzamiento de dos o tres dados. Esta experiencia aleatoria revela que las relaciones entre los conceptos de simetría, equiprobabilidad e independencia, en las que se basa la utilización de la regla de Laplace, no siempre están claras. Ahora analizamos dos ejemplos de probabilidades geométricas: la experiencia aleatoria del lanzamiento de un dardo a una diana y el problema clásico de la cuerda de Bertrand ejemplifican cómo diferentes hipótesis conducen a distintos modelos de asignación de probabilidades en un mismo problema.

Lanzamiento de un dardo a una diana

Es necesario construir un modelo que nos describa ese fenómeno aleatorio. Supongamos que siempre se da en la diana:

$\Omega = \{ (x,y) / x^2 + y^2 \leq 1 \}$ (círculo unidad)

Las partes de Ω es demasiado grande para poder asignar probabilidades a todos los sucesos, por tanto establecemos: A = {subconjuntos de Ω con área}

Se pueden establecer varias hipótesis que originan diferentes modelos de asignación de probabilidades:

Modelo 1: supongamos que el tirador no apunta

$\forall$ A $\in A$, P(A) = área de A / área de Ω = área de A / π (es un equivalente continuo de la regla de Laplace: P(A)= casos favorables / casos posibles).

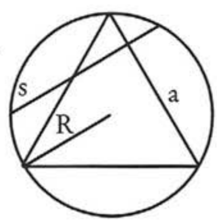
P (acertar en la mitad de arriba) = 1/2
P (acertar en el diámetro vertical) = 0, etc.

Modelo 2: supongamos que el tirador apunta y es buen tirador y por eso queremos hacer más probables los sucesos que están cerca del centro.

Se puede definir para cada punto $(x,y) \in \Omega$ una función f(x,y), densidad del punto (x,y), tal que f(x,y)>0 y disminuye al alejarnos del centro.

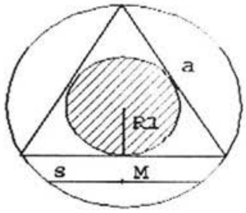
$P(A) = \int_A f(x,y)dxdy$. Obviamente debe cumplirse que $\int_\Omega f(x,y)dxdy = 1$

Problema de la cuerda de Bertrand: Se tiene un triángulo equilátero inscrito en una circunferencia de radio R y se traza una cuerda al azar ¿Cuál es la probabilidad que la longitud de la cuerda s sea mayor que el lado a del triángulo?

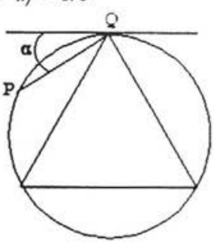


Entre varios modelos de solución, vamos a considerar dos modelos que conducen a dos soluciones distintas:

a) Se elige un punto del círculo M y se traza la cuerda perpendicular al radio que pasa por el punto. Entonces la cuerda está determinada únicamente por su punto medio M. Si M está contenido en el círculo con radio R_1, donde $R_1 = R/2$, entonces s > a, de lo contrario $s \leq a$. Por tanto P(s > a) = Área del círculo de radio R_1 / Área del círculo de radio R = 1/4.



b) Se fija un punto Q de la circunferencia y se elige al azar el otro extremo P. Sea α el ángulo que forma la cuerda así trazada con la tangente en el punto Q. La medida de este ángulo se encuentra en el rango (0,180). Si $60< \alpha <120$ se verifica que $s > a$, por tanto $P(s > a) = 1/3$



Si el azar está determinado por la equiprobabilidad de la definición de Laplace, debería haber un único conjunto de casos posibles y una probabilidad única, sin embargo, cada uno de los dos modelos representa el fenómeno aleatorio y sus probabilidades asociadas. Esto refleja un conflicto intuitivo y supone una contradicción con el principio básico de la definición de Laplace; el concepto de aleatoriedad ni está completamente cubierto por este enfoque ni es significativo sin referencia a un generador real de los sucesos.

2.2. Valor esperado

Ya hemos estudiado que, históricamente, el concepto de esperanza matemática ha sido un enfoque alternativo y a veces más importante que el concepto de probabilidad para resolver problemas de la matemática del azar. Como dicen Borovcnik, Bentz y Kapadia (1991), la Paradoja de San Petersburgo y el ejemplo de las monedas independientes ilustran las dificultades del concepto de esperanza y su relación con la probabilidad. Además, el valor esperado es un concepto clave en la conexión de la probabilidad y la estadística.

Paradoja de San Petersburgo: Dos jugadores A y B lanzan una moneda hasta que aparece una cara por primera vez. Si esto ocurre en el ensayo n entonces el jugador B paga 2^n pta. al jugador A. ¿Qué cantidad debería pagar A a B al comienzo del juego para que éste fuese equitativo?

Si X denota la cantidad que paga B, entonces su espacio muestral es un subconjunto de los números naturales. El valor esperado E(X) no existe porque la serie originada diverge:

$$E(X) = 2.(1/2) + 4.(1/4) + \ldots + 2^n.(1/2^n) + \ldots$$

Así el jugador A tendría que pagar una cantidad infinita de dinero al jugador B antes de comenzar el juego. Huygens introdujo el valor esperado como el precio equitativo de un juego estocástico. En este ejemplo, aunque la posibilidad de una larga secuencia es muy pequeña y tiende a 0, sin embargo el pago esperado es infinito. Por tanto nadie querría jugar un juego como éste en que el premio ganado es realmente una cantidad limitada de dinero.

Para resolver la paradoja, Daniel Bernoulli (1700-1782) propuso promediar las utilidades de los pagos y no los pagos específicos. Definió la utilidad como una función logarítmica de los pagos y estableció una esperanza moral finita como apuesta equitativa; por ello se le considera un precursor de la teoría de la decisión conductual. El concepto de esperanza moral se afirmó, entre otros, con Condorcet pero las circunstancias no estaban maduras para considerar la elección especial de la función de utilidad meramente como uno de varios posibles modelos (una selección de sus textos está en Condorcet, 1974).

Un problema distinto es la longitud esperada del juego que es dos ensayos, $1/p = 1/(1/2)=2$, una corta longitud que supone un pago de $2^2=4$ pts. Esos dos valores de 4 e ∞ son contraintuitivos y muestran que la relación entre diferentes esperanzas es compleja.

Monedas dependientes: Un bolsa contiene 7 monedas: 1 de 100 pta., 3 de 50 pta. y 3 de 10 pta. Se extraen 3 monedas al azar ¿Cuál es el valor esperado de su suma? ¿Es relevante si las monedas extraídas son reemplazadas?

El valor esperado de la extracción de la primera moneda es:

$$E(X_1) = (100 + 3.50 + 3.10) / 7 = 40 \text{ pta.}$$

Si las monedas son reemplazadas este es el mismo valor para la segunda y tercera extracción, por tanto $E(X_2) = E(X_3) = 40$

Si las monedas extraídas no se reemplazan entonces hay que usar los valores esperados condicionados. El valor esperado para la segunda extracción es la media ponderada de los tres valores esperados condicionados:

$$\begin{aligned} E(X_2) &= (1/7).E(X_2|100)+(3/7).E(X_2|50)+(3/7).E(X_2|10) = \\ &= (1/7)\,(180/6)+(3/7)(230/6)+(3/7)(270/6) = 40 \end{aligned}$$

Los cálculos son largos para la tercera extracción pero dan el mismo resultado. Así el valor esperado de la suma de las extracciones de las tres monedas es 120 pta. independientemente de que las extracciones hayan sido con reemplazamiento o sin reemplazamiento. Este ejemplo ilustra una propiedad fundamental de la esperanza matemática, su linealidad. Siendo X_1, X_2, X_3 variables aleatorias cualquiera (dependientes o independientes) y suponiendo que las esperanzas de cada variable existen y son finitas: $E(X_1+X_2+X_3) = E(X_1)+E(X_2)+E(X_3)$.

Este problema caracteriza la diferencia entre probabilidad y esperanza. Desde la perspectiva de la probabilidad, la relación de linealidad resulta intuitivamente inaceptable si las variables aleatorias son dependientes en función de que parece contraria a los cambios en los cálculos probabilísticos que se producen si la extracción de monedas es sin reemplazamiento. Como conclusión didáctica, diremos que hay que equilibrar la importancia que se concede en la enseñanza a los conceptos de valor esperado y de probabilidad ya que la enseñanza tiende a centrarse en el concepto de probabilidad.

2.3. Independencia y probabilidad condicionada

La probabilidad de un suceso puede cambiar si se dispone de nueva información y este hecho se modela por la noción de probabilidad condicionada. Por ejemplo, si en el lanzamiento de un dado sabemos que ha salido un número par, la probabilidad del suceso «que sea un dos» es 1/3 mientras que antes de saber la nueva información la probabilidad asignada sería 1/6.

Si A y B son dos sucesos y $P(B) >0$, se define la probabilidad condicionada de A dado B como: P[ocurra A sabiendo que ha ocurrido B]= $P(A|B)= P(A\cap B)/P(B)$.

Se dice que A es **independiente** de B si $P(A | B)= P(A)$. El conocimiento de que B ocurre no altera la probabilidad de A. Obviamente si A es independiente de B:

$P(A|B)=P(A) \Leftrightarrow P(A \cap B)= P(A)P(B) \Leftrightarrow P(B \mid A)= P(B)$, B es independiente de A.

Se llega así a una nueva definición:

A y B son independientes si $P(A \cap B)= P(A)P(B)$

Esta definición tiene la ventaja de que no requiere como la anterior que $P(A)>0$ ó $P(B)>0$.

La independencia es un concepto clave dentro del modelo de Kolmogorov (aunque no forma parte de los axiomas) porque permite la modelización de experimentos aleatorios de pruebas repetidas lo que a su vez lleva al teorema de Bernoulli y al teorema central del límite. Con todo, abundan las intuiciones inadecuadas: la paradoja de d'Alembert no se puede resolver sin el concepto de independencia; el problema del padre y su hijo muestra que el procedimiento de obtener la información influye realmente en la probabilidad.

Problema de d'Alembert: Se lanzan al aire dos monedas ¿Cuál es la probabilidad de obtener diferentes resultados en las dos monedas?

a) La solución normativa considera que el espacio producto completo es {XX, XC, CX, CC}, con una distribución uniforme; por tanto, la probabilidad pedida es 1/2.

b) D'Alembert en 1754 se opuso a la equiprobabilidad de los cuatro resultados {CC, CX, XC, XX} al lanzar dos monedas y defendió la probabilidad de 1/3 para cada uno de los resultados 0, 1 ó 2 caras. No tuvo en cuenta el concepto de independencia.

Problema del padre y su hijo: Se sabe que un señor tiene dos hijos. Se encuentra con un amigo y le presenta al chico que va con él como su hijo. ¿Cuál es la probabilidad de que el otro hijo sea también varón (V)?

a) Como hay el mismo número aproximadamente de chicos y chicas y los nacimientos son independientes, la información que se da acerca de que uno de los hijos es chico, es irrelevante y por tanto la probabilidad pedida es 1/2.

b) Las cuatro posibles combinaciones son VV, VM, MV, MM. La información dada elimina la combinación MM, por tanto la probabilidad pedida es 1/3.

A esta solución también se puede llegar utilizando probabilidades condicionadas. En efecto, sea A el suceso de que en una familia con dos hijos los dos sean varones, $P(A)=1/4$; sea B el suceso de que una familia con dos hijos tenga al menos uno varón, $P(B)=3/4$; sea $A|B$ el suceso de que tenga dos varones una familia que tiene uno:

$$P(A \mid B) = P(A \cap B)/P(B) = (1/4)/(3/4) = 1/3.$$

Sin embargo, si preguntamos la probabilidad de que una familia tenga dos varones (suceso A) suponiendo que el primer hijo es varón (suceso C), entonces:

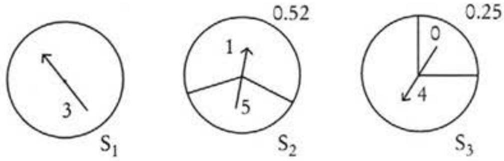
$$P(A \mid C) = P(A \cap C)/P(C) = (1/4)/(1/2) = 1/2$$

Estos dos problemas muestran la sutileza de los conceptos de independencia y de probabilidad condicionada y la necesidad de un tratamiento didáctico muy cuidadoso para que los alumnos los comprendan en profundidad.

2.4. Pensamiento lógico vs pensamiento probabilístico

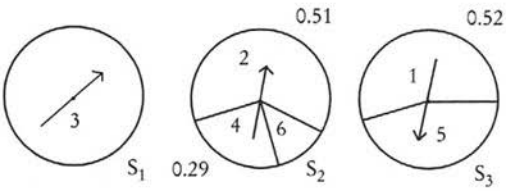
La estructura formal de la probabilidad se establece mediante el enfoque axiomático pero los axiomas no regulan la estructura de las conclusiones. Esas conclusiones tienen una estructura que difiere del razonamiento lógico y causa varias paradojas, por ejemplo, las conclusiones probabilísticas no cumplen la relación de transitividad lo que resulta muy contraintuitivo.

Ruletas intransitivas: Se dispone de las tres ruletas de la figura. ¿Cuál debe elegir un jugador para jugar si ha de competir con otro jugador?



Como $P(S_1 > S_2)=0.52$, $P(S_2 > S_3)=0.61$ y $P(S_1 > S_3)=0.25$, la ruleta S_1 es preferible a S_2, S_2 es preferible a S_3, pero S_1 no es preferible a S_3. No hay transitividad en la elección, cualquier ruleta mejora y es mejorada por otra ruleta, por tanto el jugador que elige segundo tiene ventaja. Esto es una paradoja desde la perspectiva de la lógica ordinaria en donde la transtitividad de las conclusiones es normal. La estocástica no es una forma débil de la lógica sino una forma diferente de razonamiento.

La paradoja de Blythe: Dos jugadores tienen las tres ruletas de la figura para escoger. ¿Cuál es la ruleta mejor? ¿Cambia la elección si un tercer jugador entra en el juego?



Un cálculo sencillo nos da las siguientes probabilidades:

$$P(S_1 > S_2)=0.51,\ P(S_2 > S_3)=0.62 \text{ y } P(S_1 > S_3)=0.52$$

Así S_1 es la mejor elección seguida de S_2. Sin embargo, si se introduce un tercer jugador, un cálculo un poco más largo nos conduce a que S_3, es la mejor elección:

$$P[(S_1 > S_2) \text{ y } (S_1 > S_3)]=0.51 \text{x} 0.52=0.27$$
$$P[(S_2 > S_1) \text{ y } (S_2 > S_3)]=0.29 \text{x} 0.52+0.20=0.35$$
$$P[(S_3 > S_1) \text{ y } (S_3 > S_2)]=0.48 \text{x} 0.8=0.38$$

Este resultado es sorprendente e intuitivamente inaceptable. La peor elección en el juego de dos personas se convierte en la mejor si participan tres jugadores. Las ruletas como aparatos físicos, son completamente independientes pero los resultados estocásticos dependen unos de otros si se trata de comparar probabilidades. Esto no es obvio pero se refleja en los cálculos relevantes; por ejemplo: $P[(S_3 > S_1) \text{ y } (S_3 > S_2)] \neq P(S_3 > S_1).P(S_3 > S_2)$. Es decir, la regla de la multiplicación no se cumple lo que significa que los juegos no son estocásticamente independientes.

Paradoja de Simpson: En 1973 en la Universidad de Berkeley en California, la tasa de admisión de mujeres (un 35% de las solicitantes) era más baja que la de hombres (un 44% de los solicitantes). Investigando la razón de esta discriminación sexual, se encontró que en algunos departamentos las mujeres tenían tasas de admisión más alta que los hombres y en la mayoría de los departamentos tenían tasas de admisión similares. ¿Es posible que las tasas de admisión para todos y cada uno de los departamentos sean más altas para las mujeres y en cambio, considerada la universidad globalmente, como un todo, la tasa de admisión de mujeres sea menor?

Vamos a simplificar el problema asumiendo que hay sólo dos departamentos

	Mujeres		Hombres	
	Admitidos	Rechazados	Admitidos	Rechazados
Departamento 1	2	3	1	2
Departamento 2	3	1	5	2
Universidad	5	4	6	4

En ambos departamentos la proporción de admitidos es más alta para mujeres que para hombres ya que 2/5 > 1/3 y 3/4 > 5/7. Pero para la universidad como un todo, se cumple lo contrario la proporción de admitidos de 5/9 para mujeres es menor que la proporción de 6/10 de hombres admitidos. Hay diferentes tasas de solicitudes para hombres y mujeres, las mujeres solicitan departamentos con baja tasa de admisión y los hombres al contrario. La paradoja surge porque los resultados parecen contrarios a la lógica ordinaria, donde tratar con casos separados es un método apropiado de prueba. Si el caso i) y el ii) cubren todas las posibilidades y son mutuamente excluyentes entonces se prueba una relación si se muestra que se sostiene tanto en el caso i) como en el ii). Esta característica estructural, sin embargo, no se manifiesta en el razonamiento probabilístico.

2.5. Conclusión

Las concepciones erróneas en los ejemplos que acabamos de revisar muestran que hay situaciones donde la intuición no guía la solución formal, incluso el resultado se percibe como paradójico. Esos ejemplos ilustran el salto entre intuición y teoría matemática, entre otras cosas porque el razonamiento estocástico no tiene control empírico fácil para revisar estrategias inadecuadas. Como dicen Borovcnik, Bentz y Kapadia (1991), las paradojas y las falacias destacan las dificultades de comprensión probabilística porque son señales de un conflicto cognitivo entre un nivel intuitivo y un nivel formalizado de razonamiento. En una paradoja, el aspecto objetivo es adecuado aunque intuitivamente inaccesible mientras en una falacia la componente objetiva es inadecuada aunque intuitivamente atractiva.

En resumen, las paradojas y las falacias pueden ser de interés en el aula en cuanto que su estudio y discusión pueden ayudar a: 1) analizar apropiadamente situaciones probabilísticas obscuras o complejas; 2) comprender mejor conceptos básicos en este campo; 3) interpretar formulaciones y resultados más efectivamente; 4) educar la intuición y razonamiento probabilístico; 5) ilustrar las dificultades del quehacer científico ante la presencia de situaciones científicas inesperadas y/o anómalas; 6) combatir el «sedentarismo» intelectual al que son proclives nuestros alumnos como lo son la mayoría de personas (el nomadismo cognitivo, la exploración, ayudan al avance científico).

REFERENCIAS

BERNOULLI, J. (1713). *Ars conjectandi*. Paris: Basle.

BOROVCNIK, M., BENTZ, H.J. y KAPADIA, R. (1991). A Probabilistic perspective. En R. Kapadia y M. Borovcnik (Eds.), *Chance encounters: Probability in education* (pp. 27-71). Amsterdam: Kluwer.

CONDORCET, J.A.C. (1974). *Mathématique et societé.* París: Hermann.

FELLER, W. (1973). *Introducción a la teoría de probabilidades y sus aplicaciones*. México: Limusa-Wiley.

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

FINETTI, B. de (1937). La Prévision: Ses Lois Logiques, Ses Sources Subjetives. *Ann. de l'Inst. H. Poincaré 7*. Versión inglesa: en H.E. Kyburg and H.E. Smokler (Eds.), 1964, *Studies in Subjetive Probability* (pp. 93-158). New York: Wiley.

HACKING, I. (1975). *The emergence of probability.* New York: Cambridge University Press.

JEFFREYS, M. (1933). *Theory of probability*. London: Oxford University Press.

KOLMOGOROV, A. N. (1976). La teoría de probabilidades. En A. D. Aleksandrov, A.N. Kolmogorov y M.A. Laurentiev (Eds.), *La matemática: su contenido, métodos y significado.* (pp. 269-309). Madrid: Alianza Editorial.

LAPLACE, P.S. de (1985). *Ensayo filosófico sobre las probabilidades*. Madrid: Alianza Editorial. (Original de 1814).

MARTIN GARDNER (1984). *Festival mágico-matemático.* Madrid: Alianza Editorial.

PIAGET, J. e INHELDER, B. (1951). *La genése de l'idée de hasard chez l'enfant*. Paris: P.U.F.

Resumen

En este trabajo se estudia el proceso de evolución histórica de las ideas sobre los fenómenos aleatorios, las connotaciones filosóficas que tienen esas ideas, su formalización matemática y la forma en la que situaciones paradójicas surgen en ese desarrollo y tienen profundas implicaciones didácticas.

Palabras clave: probabilidad, historia, epistemología, didáctica.

Abstract

This paper deals with the historical evolution of ideas on probabilistic phenomena, their philosophical implications, their mathematical formulae and how paradoxical situations arise within their development. These situations have important methodological implications.

Key words: Probability, History, Epistemology, Didactic.

César Sáenz Castro
Instituto de Ciencias de la Educación
Universidad Autónoma de Madrid
Ciudad Universitaria de Cantoblanco
28049 MADRID