

## EL CONCEPTO DE PROBABILIDAD

ROBERTO TORRETTI\*

El concepto de probabilidad ocupa en la vida diaria un lugar privilegiado. Cada noche los servicios nacionales de meteorología anuncian la *probabilidad* de que llueva al día siguiente. Las empresas farmacéuticas estiman, con tal o cual admitida *probabilidad* de error, la *probabilidad* de que sus medicinas tengan el efecto curativo que prometen, así como la *probabilidad* de que provoquen ciertos indeseables efectos “secundarios”. Nuestros médicos sabrían decirnos —si nos diésemos el trabajo de preguntarles— cuál es la *probabilidad* de que sea falso el resultado positivo de un examen de laboratorio a que nos sometieron, y también la *probabilidad* de que, si es verdadero, logremos, con el tratamiento que nos prescriben, superar esa condición. Las empresas industriales establecen con métodos de control de calidad la *probabilidad* de que uno de sus productos les sea devuelto por defectuoso dentro del perío-

---

\* Dedico este ensayo a Carla Cordua, con todo mi cariño. Me da muchísima alegría tener la oportunidad de hacer un aporte a la *Festschrift* en su honor, a la vez que lamento no saber escribir sobre los temas a que ella ha contribuido magistralmente. En fin, la probabilidad nos concierne a todos y cabe esperar que este panorama del debate filosófico al respecto no carezca de interés para los lectores de este volumen. El trabajo fue escrito en dos partes. La primera, referente a las concepciones frecuentista y personalista de la probabilidad, es el texto, ligeramente corregido, de la conferencia que dicté en la Universidad de Puerto Rico en Río Piedras el 8 de octubre de 2001, y fue redactada —junto con las notas que lo acompañan y los dos apéndices— antes de esa fecha. La otra parte, relativa a la probabilidad entendida como propensión física, fue escrita en abril de 2002. En el tiempo intermedio leí el lúcido libro de Donald Gillies, *Philosophical Theories of Probability* (Londres, 2000), que recomiendo vivamente a todos los interesados en este asunto. Su clara y penetrante exposición crítica de las ideas propensionistas me ha hecho posible explicarlas aquí mucho mejor —o al menos así espero— que en mi libro *Creative Understanding* (Chicago, 1990). Doy aquí efusivamente las gracias al profesor Gillies por el muy oportuno regalo de un ejemplar de su espléndida obra.

do de garantía, y las agencias de publicidad determinan mediante encuestas demoscópicas la *probabilidad* de que tal o cual campaña de anuncios aumente, por ejemplo, en un diez por ciento la fracción del mercado que controlan sus respectivos clientes. De menor consecuencia, aunque no menos llamativos, son las estimaciones que a veces se publican de la *probabilidad* de que un mono, tecleando al azar en una computadora, escriba el primer párrafo del *Quijote*, o la *probabilidad* de que una taza de café con leche se divida espontáneamente en un estrato negro de café y otro blanco de leche.

Dada esta eminencia de la probabilidad en la cultura moderna no es raro que sea tema de intenso debate filosófico. Como es habitual, este no ha conducido a un acuerdo. Peor aún, las desaveniencias entre los filósofos que han reflexionado seriamente sobre el tema son más profundas que lo normal en la tradición metafísica. En efecto, mientras los metafísicos discuten si existe la libertad, o si es posible conocer con la razón lo que no se percibe con los sentidos, en general concuerdan en cuanto al significado de estos términos. En las disputas sobre la probabilidad, en cambio, lo que se pone radicalmente en cuestión es el sentido mismo de la palabra. Esta circunstancia, por cierto, refuerza el interés filosófico del concepto, pues la filosofía, ante todo, se ceba en la perplejidad. Pero no facilita la solución y el presente trabajo no pretende ofrecerla. Aquí se trata solo de elucidar críticamente las tres opciones más populares que se presentan en el debate contemporáneo sobre el concepto de probabilidad y algunas de las razones aducidas en favor y en contra de cada una.

Pero antes de ir al grano, quiero señalar otro factor que motiva una reflexión filosófica sobre el concepto de probabilidad, además de su carácter aporético y su importancia vital. El predicado 'probable' como se usa hoy admite y demanda valores numéricos, tomados normalmente del intervalo real entre cero y uno (o entre cero y cien, si, según la práctica vulgar, las probabilidades se expresan como porcentajes). Cuantificada así, la probabilidad está presente sin duda en el pensamiento europeo del siglo XVII, y hay uno que otro indicio suyo en documentos medievales; mas no, ciertamente, en la antigüedad grecorromana ni en otras civilizaciones distintas de la europea.<sup>1</sup> El adjetivo latino *probabilis*, que

<sup>1</sup> Se ha querido ver una excepción en el siguiente episodio del *Mahabharata*, que data a más tardar del siglo IV de nuestra era. El rey Rtuparna le dice a su auriga Vahuka, que ha detenido en el aire el carro en que van los dos: "¿Ves ese árbol?

originalmente significaba 'digno de aprobación', 'loable', fue adoptado por Cicerón como equivalente del griego *πιθανός*, 'persuasivo', 'creíble', 'verosímil', o sea, 'probable'.<sup>2</sup> Un predicado con estas connotaciones naturalmente invita a comparar: tal aseveración puede parecerse probable o verosímil en el mismo grado o en un grado mayor o menor que tal otra. Sexto Empírico, el médico del siglo II de nuestra era cuyos libros de epistemología son el principal documento del escepticismo antiguo, usa el adjetivo *πιθανός* o el correspondiente sustantivo abstracto *πιθανότης* ('probabilidad') en 143 ocasiones; en seis de ellas asocia estos términos a la idea de igualdad, y se vale en otras cinco del comparativo *πιθανώτερος* para decir que tal alternativa es "más probable" que tal otra.<sup>3</sup> Pero en ningún caso compara Sexto más de dos opciones, ni le asigna

---

Tiene cincuenta millones de hojas y dos mil noventa y cinco frutos; y las hojas y frutos en el suelo superan en ciento uno el número que cuelga del árbol". Bajan a tierra y Vahuka —quien de hecho es el ex rey Nala, que perdió su reino jugando a los dados— se pasa la noche contando. Asombrado de comprobar que Rtuparna había acertado, Nala-Vahuka le pregunta: "¿Qué arte te ha enseñado esto?" Rtuparna se saca del cinturón una bolsa de cuero que contiene tres dados y le dice: "Eso no es todo: con los dados puedo hacer que salga el número que tú pidas" ¿Era Rtuparna un experto en probabilidades que estimaba el follaje de los árboles por inferencia estadística? Quizás. Como ocurre con otras supuestas anticipaciones del pensamiento moderno en la literatura de la India, el texto deja bastante libertad a la imaginación del lector. Una cosa me parece clara, sin embargo: si Rtuparna hizo su estimación mientras sobrevolaban el árbol, no pudo basarlo en una muestra de una rama cogida al azar, como sugiere Hacking (1975, p. 7).

<sup>2</sup> El siguiente pasaje de Cicerón ilustra bien la transición semántica mencionada: "Si quicquid acciderit specie probabile, si nihil se offeret quot sit probabilitati illi contrarium, utetur eo sapiens, ac sic omnis ratio vitae gubernabitur. Etenim is quoque qui a vobis sapiens inducitur multa sequitur probabilia, non comprehensa neque percepta neque adsensa sed similia veri; quae nisi probet, omnis vita tollatur. Quid enim? Conscedens navem sapiens num comprehensum animo habet atque perceptum se ex sententia navigaturum? Qui potest? Sed si iam ex hoc loco proficiscatur Puteolos stadia triginta probo navigio, bono gubernatore, hac tranquillitate, probabile <ei> videatur se illuc venturum esse salvum." (*Lucullus*, 31). En griego, *πιθανός* deviene un término técnico de la epistemología justamente con Carneades (s. II a.C.), a quien Cicerón remite en el capítulo citado; conviene si recordar la definición de la retórica por Aristóteles: "ἔστω δὴ ἡ ῥητορικὴ δύναμις περὶ ἕκαστον τοῦ θεωρῆσαι τὸ ἐνδεχόμενον πιθανόν (*Rhet.* 1355b).

<sup>3</sup> Particularmente instructivo me parece el texto de *Adversus mathematicos*, 8.333a; véase asimismo 7.445.1 y *Pyrrhoniae hypotyposes*, 2.79.9–11, 1.183.5, 1.190.4. El comparativo *πιθανώτερος* ('más probable', 'más creíble') figura en *Pyrrh. hyp.* 1.222.12, 2.79.8, y en *Adv. mathem.* 7.173.5, 9.86.3, 9.311.2. También, ciertamente, en Aristóteles; por ejemplo, *Top.*, 105<sup>a</sup>16; *Eth. Nich.*, 1096<sup>b</sup>5; *Rhet.*, 1355<sup>a</sup>38.

valores numéricos a la respectiva *πιθανότης*, ni mucho menos contempla la posibilidad, tan familiar para nosotros, de sumar o multiplicar probabilidades. En este respecto hay una diferencia notable entre la *probabilitas* o *πιθανότης* de los antiguos y nuestro concepto de probabilidad. Es razonable pensar que éste proviene de aquél, pero no sin ruptura; tal como nuestro concepto de *ciencia* proviene sin duda, aunque con profundas alteraciones, del concepto aristotélico de *ἐπιστήμη*. Ahora bien, como nadie en su sano juicio querría atribuirle a nuestros antepasados homínidos conceptos como *algoritmo*, *campo de fuerzas*, *superávit estructural* o *privacidad*, estamos hoy convencidos de la historicidad de los conceptos. Sin embargo —y tal vez por lo mismo— es muy útil que tengamos a mano, en el concepto de probabilidad, una pieza clave de la razón teórica y la razón práctica actuales, cuyo origen histórico es incuestionable y casi puede fecharse. Porque es más elusivo, ambiguo y quizás más complejo que otros conceptos básicos como el primero de mis ejemplos, es más nítido que otros, como el último, y cumple una función más decisiva en nuestro pensamiento que conceptos como los otros dos que nombré, la noción moderna de probabilidad constituye un asunto especialmente atractivo para cualquiera que se interese en la historia y la historicidad de la razón.

El contraste entre la *πιθανότης* de los antiguos y el moderno concepto cuantitativo de probabilidad es visible ya en un texto de Galileo, “Sopra le scoperte dei dadi”, que anticipa la idea de probabilidad como propensión, introducido por Popper en el debate filosófico actual. Galileo lo redactó presumiblemente entre 1613 y 1623, antes de que nacieran los creadores del cálculo de probabilidades, Blaise Pascal, Christiaan Huygens y Jacques Bernoulli; pero solo vino a publicarse póstumamente en 1718, cuando ya los tres habían fallecido. Compara cuantitativamente la ventaja que algunos números tienen sobre otros en juegos de dados. Analiza la *zara*, que se jugaba con tres dados, pero lo que dice es aplicable también a nuestro *crap*, que se juega con dos. Todos sabemos que el 7 sale mucho más a menudo que el 11. Para Galileo, la ventaja del 7 se explica porque este número es *más fácil* de formar. En efecto, mientras el 11 sólo se forma de dos maneras, esto es, si sale el 5 en un dado y el 6 en el otro o viceversa, el 7 se obtiene con seis resultados distintos: 6 y 1, 5 y 2, 4 y 3, 3 y 4, 2 y 5, 1 y 6. Me parece que la facilidad (*facilità*) de que habla Galileo está concebida como un carácter objetivo del proceso físico, en cuyo conocimiento se basará la expectativa subje-

tiva de los jugadores. Como dirá Leibniz: “Lo que es fácil para la cosa, es probable (creíble) para la mente”.<sup>4</sup> Para evaluar la facilidad de cada resultado obtenible con dos o tres dados, Galileo atiende a la facilidad de los resultados obtenibles con uno. Si cada dado es perfectamente simétrico, esta facilidad se puede dar por sentada, pues entonces tiene que ser igualmente fácil que caiga hacia arriba cualquiera de los seis números representados en sus caras. Si hay seis resultados igualmente fáciles con un dado, hay treinta y seis igualmente fáciles con 2; de esos treinta y seis, dos forman el 11 y seis forman el 7; por lo tanto, el 7 es exactamente tres veces más fácil de obtener que el 11. Supongamos que tú apuestas al 11 y yo al 7. Para que tu expectativa de ganar sea igual a la mía, el premio que recibes, si ganas, tiene que ser tres veces mayor que el que recibiría yo. Aunque estas ideas nos parecen hoy muy claras y fáciles, no hay ninguna constancia de que alguien las haya tenido en la Antigüedad. Sin embargo, no es difícil ver una conexión entre la cuantificación de la expectativas que podemos basar en el pensamiento de Galileo y la igualdad de las probabilidades de que hablaba Sexto. Pasar de dos alternativas en equilibrio a 6, 36, 216,... fue el gran salto. El paso siguiente, consistente en asignar a cada alternativa equiprobable la misma fracción positiva menor que 1 —1/2 si son dos, 1/6 si son seis, etc.— es mucho más fácil y quizás tenga relación con la práctica de dividir el pozo acumulado cuando hay que interrumpir un juego de azar. Si tú apostaste a que sale primero el 11 y yo a que sale el 7 y decidimos parar el juego antes de que haya salido ninguno de los dos, es justo que te lleves tres cuartos del pozo y yo un cuarto.<sup>5</sup>

<sup>4</sup> *Quod facile est in re, id probabile est in mente* (Leibniz, *SS*, VI.II, p. 492). El paso de la facilidad física a la plausibilidad epistémica ya está presente, me parece, en el pasaje de Cicerón citado en la nota 2: “Sed si iam ex hoc loco proficiscatur Puteolos stadia triginta probo navigio, bono gubernatore, hac tranquillitate, probabile <ei> videatur se illuc venturum esse salvum.”

<sup>5</sup> Según se lo que tradicionalmente se cuenta, el cálculo de probabilidades nació con la solución de un problema de este género por Pascal, en carta a Fermat del miércoles 29 de julio de 1654. Se trata de dividir equitativamente un pozo de 64 monedas entre dos jugadores, que llamaré Juan y Diego, cuando el juego se interrumpe en las circunstancias siguientes: (i) según las reglas convenidas el primero en enterar tres partidas ganadas se lleva el pozo completo; (ii) ambos jugadores tienen la misma probabilidad de ganar cada partida; (iii) cada partida dura hasta que un jugador la gana (no hay empates); (iv) el juego va a interrumpirse después de tres partidas, de las que Juan ha ganado dos y Diego una. Pascal razona así: Si Juan gana la cuarta partida, se lleva las 64 monedas; si la gana Diego, ambos jugadores tienen la misma expectativa de ganar el total después la partida siguiente; por lo tanto, si el juego tu-

Durante sus primeros dos siglos la teoría de las probabilidades descansa en la noción de equiprobabilidad. A ella remite la clásica definición de Laplace: *La probabilidad de un evento es el cociente entre el número de casos "igualmente posibles" favorables a ese evento y el total de todos los casos igualmente posibles.*<sup>6</sup> (Por ejemplo, la probabilidad de que salga el 7 en una jugada de *crap* es el cociente entre 6 y 36). Para que esta definición no sea circular, hay que adoptar 'igualmente posible' o 'equiprobable' como término primitivo. La equiprobabilidad es manifiesta y se explica por sí misma en las situaciones aleatorias donde hay simetría entre los varios desenlaces posibles.<sup>7</sup> Pero en la mayoría de las situaciones reales no hay simetría. Además, como parece indicar la paradoja de Bertrand, habría casos en que la presencia de distintas simetrías daría lugar a asertos de equiprobabilidad incompatibles.<sup>8</sup> Por eso, desde

viera que interrumpirse después de cuatro partidas lo justo sería que cada jugador se lleve 32 monedas. Pascal concluye entonces que, después de la tercera partida, Juan tiene derecho en todo caso a las 32 monedas que le pertenecerían quienquiera que gane la cuarta; como tiene la misma probabilidad que Diego de ganar la cuarta partida, lo justo es que las otras 32 monedas, cuyo destino se jugaría en esa partida, se repartan por partes iguales, entre los dos. Por lo tanto, al interrumpirse el juego en las condiciones descritas bajo el número (iv), le tocan 48 monedas a Juan y 16 a Diego.

<sup>6</sup> "La teoría de las probabilidades consiste en reducir todos los eventos de un mismo género a cierto número de casos igualmente posibles —esto es, tales que estemos igualmente indecisos sobre su existencia— y a determinar el número de casos favorable al evento cuya probabilidad se busca. La proporción entre este número y el de los casos posibles es la medida de esta probabilidad, la cual, pues, no es otra cosa que una fracción cuyo numerador es el número de los casos favorables, y cuyo denominador es el número de todos los casos posibles" (Laplace 1795, pp. viii-ix). Obsérvese que para el determinista Laplace, los asertos de equiprobabilidad no reflejan la simetría real de una situación física, sino sólo una simetría aparente producto de nuestra ignorancia.

<sup>7</sup> Lorraine Daston (1988) sostiene que la preeminencia de la noción de equiprobabilidad en la etapa clásica de la investigación matemática de las probabilidades se debe a que ésta estuvo motivada por el deseo de conocer las condiciones en que los contratos aleatorios (seguros, apuestas) resultan equitativos, para lo cual tiene que haber igualdad entre las expectativas de las partes. Seguramente tiene razón; pero ello no impide que los asertos de equiprobabilidad se hayan basado en la percepción de simetrías físicas, donde estas existían.

<sup>8</sup> Explico esta paradoja en el apéndice I. Conviene destacar que no todos se han dejado afectar por ella. Al final del apéndice I me refiero a Jaynes (1973). Más recientemente, en un estudio profundo y original sobre los fundamentos de la probabilidad, el filósofo y matemático chileno Rolando Chuaqui (1991) devuelve su primacía clásica a la equiprobabilidad y las distribuciones uniformes. *Vide infra*, nota 33.

mediados del siglo XIX la definición de Laplace pierde prestigio y se busca entender la probabilidad de otro modo. En 1843 Cournot formula la concepción frecuentista, elaborada con más precisión por Ellis (1849, 1856) y perfeccionada en el siglo XX por von Mises (1928, 1931, 1964) y Reichenbach (1935, 1949). En 1921, Keynes formula la concepción logísta, que luego culmina y naufraga en la obra de Carnap (1950, 1952, 1962, 1971/80). En 1928, de Finetti demuestra el teorema de representación que cimienta la concepción personalista, adoptada en la segunda mitad del siglo XX por Savage (1954), Lindley (1965) y muchos otros estadísticos. En los años 50, Popper (1957, 1959) revive la idea galileana de *facilità*, concibiendo la probabilidad como propensión.

• • •

Dicho toscamente, el frecuentismo identifica la probabilidad de un evento con la frecuencia relativa con que se presentan, *a la larga*, eventos como ese entre los eventos de su clase. Por ejemplo, la probabilidad de que un varón de 40 años de edad que fuma 60 cigarrillos diarios desde los 20 muera de cáncer pulmonar antes de los 50 no es sino la frecuencia relativa de tales defunciones entre todos los varones de 40 años de edad que han fumado con el mismo tesón. Esta nueva manera de entender la probabilidad prospera a lo largo de todo el siglo XIX con la devoción de las elites europeas a lo "colectivo" y empieza a flaquear solo después del fracaso de los regímenes colectivistas del XX. Como motivación intelectual hay que citar la toma de conciencia de ese género de fenómenos que Denis Poisson (1837) subsume bajo su "ley de los grandes números": en muy diversas áreas de la experiencia se repiten sucesos que, aunque inconexos y al parecer aleatorios, exhiben proporciones estables a largo plazo.<sup>9</sup> Ya en el siglo XVII llamó la atención la constancia de la propor-

<sup>9</sup> No tengo a mano un ejemplar del tratado de Poisson (1837). El pasaje siguiente, tomado de la introducción al mismo, está traducido de otro libro, que lo cita en inglés: "En muchos campos diferentes, los fenómenos empíricos parecen obedecer cierta ley general, que cabe llamar la Ley de los Grandes Números. Esta ley dice que las proporciones numéricas derivadas de la observación de un número muy grande de sucesos similares permanece prácticamente constante, con tal que esos sucesos estén gobernados en parte por factores constantes, en parte por factores variables cuyas variaciones son irregulares y no causan un cambio sistemático en una dirección definida. Ciertos valores de estas proporciones son característicos de cada clase dada de sucesos. Al aumentar la longitud de la serie de observaciones las proporciones derivadas de tales observaciones se aproximan más y más a esas constantes características. Cabría esperar que las reproduzcan exactamente si fuera posible hacer series de observaciones de longitud infinita."

ción en que nacen niñas y niños. En un filme que muestra un enjambre de espermios dando caza a un óvulo impresiona ante todo el desorden de esta carrera, cuyo ganador parece ser completamente imprevisible. Solo si este es portador del cromosoma Y podrá el óvulo fecundado convertirse en un machito. La meiosis de una célula masculina produce dos espermios, uno solo de los cuales porta ese cromosoma. Sin embargo, en las poblaciones humanas nacen más varones que mujeres, en una proporción sorprendentemente estable, sobre el 51%. Por ejemplo, en 1935, nacieron en Suecia 45,682 niños y 42,591 niñas. Basándose en estas cifras, Harald Cramér escribe: "La probabilidad estimada de un nacimiento masculino es  $p^* \dots = 0.5175082$ " (Cramér 1946, p. 447). Un frecuentista como Cramér no pretende con esta afirmación atribuir a cada espermio portador del cromosoma Y una facilidad para alcanzar y penetrar el óvulo ligeramente mayor que la de sus rivales cargados con el cromosoma X. Del espermio ganador no sabemos gran cosa. Tampoco se barrunta una conexión causal entre la respectiva composición química de los cromosomas X e Y y una diversa aptitud de sus portadores para alcanzar un óvulo. Un frecuentista solo habla de *probabilidad* a propósito de colecciones numerosas de casos, de secuencias muy numerosas de sucesos; y emplea el término para referirse a la proporción estable que a la larga ocupan los casos o sucesos de un cierto tipo en el total de esas colecciones o secuencias.

La ley de los grandes números empírica a que me he referido evoca por su nombre y su contenido varios teoremas de la teoría matemática de las probabilidades que también suelen llamarse "leyes de los grandes números".<sup>10</sup> El primero de ellos fue demostrado por Jacques Bernoulli en la etapa fundacional de la teoría de las probabilidades.<sup>11</sup> El teorema de

<sup>10</sup> Véase P. Révész (1968). Las más conocidas son la llamada ley *débil* de los grandes números, o teorema de Bernoulli, explicada en el texto a continuación, y la ley *fuerte* demostrada a principios del siglo XX por Cantelli y llamada así porque implica pero no equivale a la anterior. Con la nomenclatura y el simbolismo que emplearé para explicar la ley débil, la ley fuerte puede enunciarse así: La probabilidad  $p$  de que la frecuencia relativa  $f_n$  sea igual a  $p$  converge al límite 1 cuando  $n$  crece indefinidamente:

$$\lim_{n \rightarrow \infty} (|f_n - p| = 0) = 1$$

<sup>11</sup> La demostración está contenida en la Parte IV de la obra póstuma de Bernoulli (1713). Por este resultado, Jacques Bernoulli ocupa en la historia de la teoría de las probabilidades, iniciada por Galileo, Pascal y Huygens, un lugar compara-

Bernoulli concierne a una serie indefinida de experimentos de un cierto tipo que llamaré E. Los experimentos E son completamente independientes entre sí, y en cada uno de ellos puede ocurrir, con la misma probabilidad  $p$ , cierto evento que llamaré *éxito* y simbolizaré con 1. Si en un determinado experimento no se produce el éxito digo que ha ocurrido un *fracaso*, que simbolizo con 0. Entonces, según las reglas del cálculo de probabilidades, basadas en la concepción clásica pero admitidas por todas las concepciones, la probabilidad de fracaso es  $1 - p$ .<sup>12</sup> El prototipo clásico del experimento E consiste en sacar a ciegas una bola de cierto color de una urna en que hay un cierto número de bolas, bien mezcladas, del mismo material y tamaño, pero de distintos colores. El experimento se repite utilizando otra urna igual a la primera o retornando a ésta la bola extraída y sacudiéndola bien. Consideremos ahora el experimento E', consistente en realizar un determinado número  $n$  de veces —simultánea o sucesivamente— el experimento E. Para describir el resultado de cada experimento E' asignamos un orden a los  $n$  experimentos E que lo forman; pero, como se verá, este orden es del todo indiferente a la conclusión obtenida por Bernoulli.<sup>13</sup> En cada experimento E' el éxito ocurre un cierto número  $r$  de veces, y  $n - r$  veces el fracaso. El resultado puede describirse entonces como una serie de unos y ceros. La proporción  $r/n$  entre el número  $r$  de unos y el número  $n$  de experimentos E es la *frecuencia relativa* del éxito en el experimento E'. La designo con  $f_n$ . Espontáneamente, esperamos que, si  $n$  es grande,  $f_n$  ande cerca de  $p$ ; en otras palabras, que la probabilidad del éxito en cada experimento no difiera mucho de la frecuencia relativa en una serie larga de repeticiones del mismo experimento. El teorema de Bernoulli se refiere precisamente a la probabilidad de que esta diferencia exceda una cierta cantidad  $\epsilon$

ble al de Isaac Newton en la historia de la dinámica, iniciada por Galileo, Descartes y Huygens.

<sup>12</sup> Dentro de la concepción clásica, el éxito es o bien (i) uno de  $n$  resultados equiprobables que puede tener el experimento, y entonces  $p = 1/n$  y la probabilidad de fracaso es  $1 - 1/n = 1 - p$ , o bien (ii) la unión de  $r$  de esos  $n$  resultados, y entonces  $p = r/n$  y la probabilidad de fracaso es  $(n - r)/n = 1 - r/n = 1 - p$ .

<sup>13</sup> Típicamente, se trata del orden temporal, si E' consiste en repeticiones sucesivas de E. Si la repetición es simultánea podemos ordenar los E por su posición en el espacio, o alfabéticamente, por el nombre de los experimentadores o, si un mismo experimentador ejecuta simultáneamente varios experimentos E, por el órgano —mano izquierda, pie derecho, etc.— utilizado principalmente en cada ejecución. Es claro, entonces, que el teorema no podría depender del orden asignado a los experimentos.

fijada arbitrariamente, tan pequeña como se desee. El teorema dice que la probabilidad de que  $|f_n - p|$  sea mayor que  $\varepsilon$  es tanto menor cuanto mayor sea  $n$ . Más precisamente, dice que, cuando  $n$  crece indefinidamente, la probabilidad  $\mathbf{P}(|f_n - p| > \varepsilon)$  converge al límite 0. En otras palabras, el teorema de Bernoulli dice que, dados  $p$  y  $\varepsilon$ , entonces, para cualquier cantidad  $\delta$ , por pequeña que sea, existe un entero  $M$  tal que, para todo  $n > M$ ,  $\mathbf{P}(|f_n - p| > \varepsilon) < \delta$ . El teorema permite calcular cuál es el mínimo  $M$  de veces que hay que repetir el experimento  $E$  para asegurar con probabilidad  $1 - \delta$  que la frecuencia relativa de éxitos  $f_M$  caerá dentro del intervalo  $(p - \varepsilon, p + \varepsilon)$ . Supongamos, por ejemplo, que  $E$  consiste en sacar a ciegas una bola roja de una urna bien mezclada en que hay 55 bolas rojas y 45 de otros colores. Entonces, el teorema de Bernoulli nos permite calcular cuántas repeticiones del experimento  $E$  hacen falta para confiar en un 95% que el porcentaje de bolas rojas extraídas estará entre 52 y 58. Como en otros grandes descubrimientos matemáticos, el golpe de genio de Bernoulli estuvo en dar con un enunciado a la vez interesante y demostrable. La demostración misma es laboriosa pero sencilla. Se trata de determinar cuáles son las distintas secuencias posibles de éxitos o fracasos en  $n$  experimentos, agrupar las que exhiben una misma frecuencia relativa de éxitos, calcular la probabilidad de cada grupo y pasar al límite  $n \rightarrow \infty$ . Un mayor desafío ofrece el problema inverso planteado por Bernoulli: dada la frecuencia relativa de éxitos  $f_n$  en un experimento repetitivo  $E^n$ , determinar, con un pequeño margen de error  $\varepsilon$  y un alto grado de confiabilidad  $1 - \delta$ , la probabilidad de éxito  $p$ . No es exagerado decir que el desacuerdo entre frequentistas y personalistas tiene sobre todo que ver con la respectiva actitud hacia este problema, y por lo tanto no existiría si el mismo tuviera una solución que los convenciese a todos.

Me parece que hay una afinidad muy clara entre el teorema de Bernoulli y la formulación rigurosa de la concepción frequentista por Richard von Mises. Según esta, sólo puede atribuirse probabilidad a un evento en cuanto el mismo pertenece a lo que von Mises llama un *colectivo*. Con este término, von Mises se refiere a una larga serie de observaciones en que el evento se repite con una frecuencia relativa más o menos estable. Una serie de este tipo constituye un colectivo si hay buenas razones para creer que la frecuencia relativa del evento tendería a un límite fijo si las observaciones continuasen indefinidamente. Ese límite es *la probabilidad del evento considerado dentro del colectivo dado*. Si tal

límite no existe, no tiene sentido atribuir una probabilidad al evento. Hasta aquí, todo parece muy claro. Pero von Mises tuvo que imponer un requisito más a sus colectivos, debido a la simple consideración siguiente: si una ruleta produce alternativamente números rojos y negros la frecuencia relativa del rojo converge obviamente al límite  $\frac{1}{2}$ ; sin embargo, no puede decirse que  $\frac{1}{2}$  sea la probabilidad de que salga el rojo en una cierta jugada, pues, conforme a las condiciones prescritas, si acaba de salir el negro es seguro que sale el rojo, y también es seguro que vuelve a salir el negro en la jugada subsiguiente. Por eso, para que una larga serie de observaciones constituya un colectivo en el sentido de von Mises no basta que la frecuencia relativa de cada tipo de eventos observados converja a un límite fijo, y von Mises exige además que dicho límite no varíe cuando, en vez de la secuencia completa, se considera una subsecuencia formada por lo que llamaré selección tópica; esto es, una subsecuencia seleccionada atendiendo exclusivamente al lugar ( $\tau\acute{o}\pi\omicron\varsigma$ ) que ocupa cada uno de sus miembros dentro de la secuencia total. (Así, formamos por selección tópica la subsecuencia que comprende todos los eventos cuyo lugar corresponde a un número par —segundo, cuarto, sexto,...— o a un número primo —segundo, tercero, quinto, séptimo, undécimo—, etc., sin atender a su respectiva índole, cero o uno, éxito o fracaso). Como es obvio, para decidir si una secuencia de observaciones  $S$  es o no un colectivo en el sentido de von Mises, no basta examinar las subsecuencias finitas que pueden extraerse por selección tópica de un segmento inicial de  $S$ , por largo que sea.

Según von Mises la relación entre el teorema de Bernoulli y su concepción frequentista no consiste en que ésta se funde en aquél; antes bien, para él, el teorema es aplicable a la realidad sólo en los términos de su concepción frequentista. Como el teorema y sus generalizaciones y fortalecimientos posteriores son la médula misma de la teoría matemática de las probabilidades, von Mises habría diseñado su concepción frequentista para asegurar la aplicabilidad de esta teoría. Según él, esta provee el núcleo matemático del estudio científico de los fenómenos que exhiben proporciones estables a largo plazo e ilustran, por lo tanto, la ley empírica de los grandes números; al modo como, digamos, la geometría de Riemann, especificada por las ecuaciones de campo de Einstein, es el núcleo matemático del tratamiento científico de los fenómenos gravitacionales en la astronomía y la cosmología de hoy.

El frequentismo ha sido severamente criticado, incluso por sus seguidores. Así, Carl Hempel, el más agudo de los empiristas lógicos, en la te-

sis doctoral que le supervisó el frecuentista Reichenbach, hace ver que, si la probabilidad es el límite de la frecuencia relativa en una secuencia infinita, los asertos de probabilidad no pueden verificarse empíricamente y, por ende, para un empirista lógico, no quieren decir nada.<sup>14</sup> La razón es muy simple. Si  $S$  es una secuencia empírica, digamos, de un billón de observaciones, en que la frecuencia relativa de éxitos  $f_{10^{12}} \approx p$ ,  $S$  puede ser el segmento inicial de una secuencia infinita  $\Sigma$  en que la frecuencia relativa de éxitos converge a cualquier límite  $q \neq p$ . Wolfgang Stegmüller (1973, p. 37). objeta que la definición frecuentista de probabilidad banaliza la ley matemática de los grandes números: un teorema difícil de probar y aparentemente profundo pasa a ser una tautología obvia. En verdad, no es fácil enunciar la ley fuerte de un modo razonable, utilizando el vocabulario frecuentista. La ley fuerte implica que, si  $\sigma$  es una secuencia infinita de experimentos del mismo tipo y la probabilidad del resultado  $\lambda$  en cada experimento es  $p$ , entonces la probabilidad  $P$  de que la frecuencia relativa de los  $\lambda$  converja al límite  $p$  es igual a 1. La probabilidad  $P$  de que una cierta secuencia de experimentos exhiba una determinada frecuencia límite se infiere pues de la probabilidad dada  $p$  de un cierto tipo de resultado en cada experimento singular de la secuencia. Para el frecuentista, empero, la probabilidad en un experimento singular es solo una manera de hablar sobre la frecuencia límite en el colectivo a que el experimento pertenece. Para él, la probabilidad dada  $p$  es idéntica, por definición, a la frecuencia límite de los  $\lambda$  en  $\sigma$ . La probabilidad inferida  $P$  también debe identificarse con una frecuencia límite, a saber, el límite —en una secuencia  $\Sigma$  de colectivos como  $\sigma$  en que la probabilidad de  $\lambda$  es  $p$ — de la frecuencia relativa de los colectivos en que la frecuencia límite de los  $\lambda$  es  $p$ . Como todo colectivo en que la probabilidad de  $\lambda$  es  $p$  tiene, *por definición*, una frecuencia límite  $p$  de casos  $\lambda$ , no es sorprendente que  $P = 1$ . Pero, a diferencia de la ley fuerte de los grandes números, esta ecuación no es un resultado matemático sorprendente, logrado con mucho trabajo, sino una consecuencia trivial de la definición frecuentista de probabilidad. Recordamos situaciones parecidas en geometría, donde las propiedades no triviales de una figura valen trivialmente para sus casos degenerados. Parecería, entonces, que la teoría de los colectivos no elu-

<sup>14</sup> Hempel (1934, 1935). El segundo trabajo está ahora disponible en traducción inglesa, en Hempel (2000), pp. 89–123.

cida el concepto científico de probabilidad en toda su generalidad, sino solo un caso especial, degenerado, del mismo.

Creo, sin embargo, que la peor objeción contra el frecuentismo no es matemática, sino moral. Si la probabilidad, por definición, no es sino el límite a que converge la frecuencia relativa en una secuencia infinita o, a lo menos, la fracción a que se acerca en una secuencia larguísima; si *no tiene sentido hablar de la probabilidad de un suceso singular*, ¿qué queda de la probabilidad *como guía de la vida*? ¿De qué le vale a un enfermo averiguar que, en una secuencia larguísima de intervenciones quirúrgicas como aquella que el médico le recomienda para librarse de los terribles dolores que lo aquejan, es prácticamente seguro que un 80% de los pacientes se sana, un 5% se muere y el 15% restante queda parálítico? Estos números pueden ser de interés para el empleado de la compañía de seguros que tiene que resolver si autoriza o no la operación (en sus cálculos, deberá tener en cuenta además cuánto cuesta la operación y cuánto el cuidado del paciente en su condición actual y si queda parálítico). Pero al enfermo individual, solo pueden procurarle una angustiada incertidumbre. Ciertamente no es eso lo que los pacientes entendemos cuando los médicos nos hablan de probabilidades, ni, me atrevo a decir, lo que estos, en su mayoría, creen estarnos diciendo. Resta, por último, la dificultad epistemológica insinuada arriba, cuando mencioné el problema inverso de Bernoulli. Sólo tenemos acceso a secuencias finitas de observaciones. ¿Cómo inferir válidamente, desde ellas, los límites a que convergirán las frecuencias relativas si esas secuencias se prolongan al infinito?

• • •

Estos problemas se resuelven de una sentada con la concepción personalista o subjetivista de la probabilidad. Ella fue adoptada independientemente, poco antes de 1930, por Bruno de Finetti y Frank Ramsey, quienes descubrieron, cada uno por su cuenta, el hecho fundamental que la hace viable.<sup>15</sup> De Finetti pudo además demostrar un hermoso teorema matemático —el llamado Teorema de Representación<sup>16</sup>— que ha sido luego fortalecido y sin el cual me parece que la concepción personalista

<sup>15</sup> Véase de Finetti (1931), Ramsey (1931).

<sup>16</sup> De Finetti (1930). De Finetti comunicó el teorema al Congreso Internacional de Matemáticos celebrado en Bolonia en 1928, como consta en sus actas, publicadas en 1932. Véase el apéndice II.

no habría alcanzado nunca su popularidad actual. De Finetti es un maestro de la presentación desembozada de ideas radicales:

### LA PROBABILIDAD NO EXISTE

El abandono de creencias supersticiosas sobre la existencia del Flogisto, el Éter Cósmico, el Espacio y el Tiempo Absolutos, ..., o las Hadas y las Brujas, constituyó un paso esencial en el camino hacia el pensamiento científico. También la Probabilidad, si se la considera como algo dotado de algún género de existencia objetiva, es una concepción no menos descaminada, un intento ilusorio de exteriorizar y materializar nuestras verdaderas creencias probabilísticas.

(De Finetti 1974, vol. I, p. x)

Según de Finetti, una aseveración objetiva es *verdadera* o *falsa*, mas no *probable*. La probabilidad es un atributo de nuestras opiniones subjetivas sobre aquellos asuntos acerca de los cuales no podemos o no queremos hacer una aseveración objetiva.<sup>17</sup> El valor numérico de las probabilidades mide el grado de confianza que cada opinión inspira, ahora y aquí, a quien la profesa. A la probabilidad así concebida, de Finetti dio en llamarla *previsión*. Como es obvio, la previsión puede referirse a sucesos singulares y no implica ni supone un quimérico paso al infinito. Aunque las previsiones de los sujetos son por cierto subjetivas, en un importante género de casos, si los sujetos son coherentes, sus previsiones tenderán al acuerdo intersubjetivo (conforme al susodicho Teorema de Representación). Pero antes de referirme a esto, tengo que explicar en qué consiste el hecho fundamental descubierto por de Finetti y Ramsey.

Se trata de lo siguiente: si nuestra confianza se cuantifica, esto es, si se traduce en previsiones del modo muy razonable que explicaré enseguida, las que forme cada persona sobre la realización de eventos lógicamente relacionados entre sí tienen que regirse, so pena de incoherencia, por las reglas familiares del cálculo de probabilidades. "Entonces —dice de Finetti— el cálculo de probabilidades no es más que *la teoría matemática que enseña a ser coherentes*" (de Finetti 1974, vol. I, p. 305). De

<sup>17</sup> Esto puede deberse a que el asunto en cuestión está indeterminado, ya sea intrínsecamente —por ejemplo, ¿cuántos átomos de uranio se desintegrarán en este pequeño trozo de roca durante las próximas seis horas?— o a nuestros ojos —por ejemplo, ¿qué equipo de fútbol triunfará en el próximo campeonato?— ; pero también cabe tener opiniones probables sobre asuntos perfectamente determinados —por ejemplo, ¿cuál es el trillonésimo dígito de la expansión decimal de la base de los logaritmos naturales?

Finetti ha dado dos definiciones de previsión, distintas pero equivalentes. La más sencilla es esta: La previsión del sujeto  $X$  sobre la ocurrencia del evento  $E$  es la cantidad  $p$  de pesos que  $X$  está dispuesto a pagar a cambio de la seguridad de recibir \$1 si ocurre  $E$ .<sup>18</sup> Las previsiones así definidas pueden ser todo lo arbitrarias que se quiera, sin otra restricción que esta: so pena de incoherencia,  $X$  no puede adoptar previsiones tales que, si está obligado a apostar de acuerdo con ellas, pierde seguramente, pase lo que pase. Por lo tanto,  $p$  no puede nunca ser mayor que 1, pues si lo fuera,  $X$  perdería la cantidad  $p$  si no ocurre  $E$  y la cantidad  $p - 1$  si ocurre. Además, si  $E$  y  $F$  son dos eventos que mutuamente se excluyen, y las previsiones de  $X$  sobre la ocurrencia de  $E$  y de  $F$  son respectivamente  $p$  y  $q$ , su previsión sobre la ocurrencia del evento  $(E \vee F)$  tiene que ser igual  $p + q$ , si no quiere verse puesto en la situación de perder pase lo que pase.<sup>19</sup> Ahora bien, si designamos con  $\mathbf{p}(E)$  la probabilidad de un evento  $E$ , la teoría matemática de las probabilidades para un espacio de eventos finito queda caracterizada precisamente por la doble condición:

- I.  $0 \leq \mathbf{p}(E) \leq 1$  para todo evento  $E$ .
- II.  $\mathbf{p}(E \vee F) = \mathbf{p}(E) + \mathbf{p}(F)$  para dos cualesquiera eventos  $X$  y  $Y$  cuya conjunción  $X \wedge Y$  es imposible.

<sup>18</sup> La otra definición es esta: El sujeto  $X$  es contratado para evaluar la probabilidad de un suceso incierto. Sea  $E$  el indicador de ese suceso, esto es,  $E = 1$  si el suceso ocurre y  $E = 0$  si no ocurre.  $X$  expresará su evaluación mediante un número real  $\mathbf{p}(E)$ . Su honorario se fija en una cantidad de dinero igual a  $H - k(E - \mathbf{p}(E))^2$  pesos, donde  $H$  y  $k$  son números reales fijados de modo que a  $X$  le valga la pena hacer este trabajo con el máximo cuidado de que es capaz. Obviamente, el honorario recibido es máximo si  $\mathbf{p}(E) = E$ ; pero, como  $E$  es incierto, si  $X$  se gana la vida haciendo este tipo de evaluaciones, preferirá asignar a  $\mathbf{p}(E)$  un valor entre 0 y 1, tanto mayor cuanto más posible le parezca la realización del suceso. Por definición,  $\mathbf{p}(E)$  es la previsión de  $X$  sobre la ocurrencia del suceso en cuestión.

<sup>19</sup> Doy un ejemplo. En una carrera de ocho caballos participan solo dos yeguas, Nicéfora y Mesalina. Supongamos que la previsión de  $X$  sobre el triunfo de Nicéfora es  $p = 0.6$ , y sobre el de Mesalina es  $q = 0.2$ , pero que su previsión de que ganará una yegua es  $0.75 \neq p + q = 0.8$ . Si  $X$  actúa de acuerdo con estas previsiones tendrá que tomar apuestas por el triunfo de una yegua, en la razón de 3 a 1; contra Nicéfora, en la razón de 2 a 3, y contra Mesalina, en la de 1 a 4. Si le apuestan simultáneamente \$1 a que gana una yegua, \$3 a que no gana Nicéfora y \$4 a que no gana Mesalina,  $X$  recibirá un total de \$8, pero tendrá que desembolsar, si gana Mesalina, los \$4 apostados a las alternativas ganadoras más \$5 de premio, o sea, \$9; también \$9, si gana Nicéfora, y \$10 si gana un potro. Los desembolsos netos de  $X$  oscilan, pues, entre \$1 y \$2, pero significan una pérdida en todo caso.

Es claro, entonces, que las previsiones de una persona que sea *coherente* en el sentido de de Finetti son efectivamente *probabilidades* en el sentido de la teoría matemática.

El Teorema de Representación concierne a lo que de Finetti llama "eventos intercambiables". Pensemos en una serie de  $n$  sucesos individuales cualesquiera, clasificados como "éxitos" y "fracasos" (donde  $n$  es un entero positivo). Ella constituye, para un determinado sujeto  $X$ , una *sucesión de eventos intercambiables* si y solo si, para cada entero no negativo  $k \leq n$ , la previsión de  $X$  de que ocurren exactamente  $k$  éxitos depende únicamente de  $k$  y de  $n$  (y no del orden en que se suponga los éxitos de alternan con los fracasos).<sup>20</sup> Lo que nos interesa aquí no es el teorema mismo, cuyo enunciado preciso es difícil,<sup>21</sup> sino una consecuencia o corolario del mismo que se puede explicar y justificar informalmente sin mucho trabajo. Siguiendo a de Finetti, llamo  $\omega_k^{(n)}$  a la previsión de que ocurran  $k$  éxitos en una sucesión de  $n$  eventos intercambia-

bles. Hay exactamente  $\frac{n!}{(n-k)!k!} = \binom{n}{k}$  modos diferentes como  $n$  even-

tos sucesivos pueden repartirse entre  $k$  éxitos y  $(n-k)$  fracasos. Sea  $A$  uno cualquiera de ellos. La previsión  $\mathbf{p}(A)$  de que justamente ha de ocurrir  $A$  es entonces igual a  $\omega_k^{(n)} / \binom{n}{k}$ . Sea  $A'$  la sucesión de eventos intercambia- bles que se produce si la secuencia  $A$  es seguida por el éxito  $E$ . La previsión  $\mathbf{p}(A')$  de que ocurrirá  $A$  es igual a  $\omega_{k+1}^{(n+1)} / \binom{n+1}{k+1}$ . Con estos datos po-

<sup>20</sup> En otras palabras, la sucesión  $E_1, E_2, \dots, E_n$  es una sucesión de eventos intercambia- bles (para  $X$ ) si y solo si, para cada entero no negativo  $k \leq n$ , la probabilidad de que haya exactamente  $k$  eventos positivos y que estos sean precisamente los primeros,  $E_1, E_2, \dots, E_k$ , es —a juicio de  $X$ — idéntica a la probabilidad de que los únicos eventos positivos sean  $E_{f(1)}, E_{f(2)}, \dots, E_{f(k)}$ , donde  $f$  es cualquier función inyectiva creciente de  $\{1, 2, \dots, k\}$  en  $\{1, 2, \dots, n\}$ . A modo de ejemplo, consideremos una bolsa que contiene muchas bolas blancas y negras, en una proporción desconocida. Un niño con los ojos vendados extrae bolas una por una. Cada bola extraída se devuelve inmediatamente a la bolsa después de anotar su color. Cualquiera que fuese nuestra previsión de que, digamos, en las próximas 100 extracciones aparezcan 43 bolas blancas, ella sería normalmente indiferente al orden de aparición. Y otro tanto vale para la probabilidad de que las bolas blancas sean 44 ó 42 ó cualquier otro número entre 0 y 100. Análogamente, no me parece que la previsión de ninguna persona sobre el porcentaje de varones entre los próximos 100.000 nacidos vivos dependa en nada del orden en que se sucedan los nacimientos de uno y otro sexo.

<sup>21</sup> En el apéndice II reproduzco el enunciado del teorema y la demostración del mismo debida a William Feller (1971).

demos computar la previsión de que la particular sucesión  $A$  de  $n$  eventos que incluyen  $k$  éxitos sea sucedida inmediatamente por  $E$ . Esta cantidad, comúnmente simbolizada  $\mathbf{p}(E|A)$ , es la probabilidad (o previsión) de  $E$  bajo la condición  $A$ , o probabilidad condicional de  $E$  dado  $A$ . De Finetti la define como la probabilidad que uno le atribuye a  $E$  si uno piensa que, además de la información de que ahora dispone, uno se enterará de que  $A$  se ha cumplido (y de nada más).<sup>22</sup> Si uno es coherente,<sup>23</sup>

$$\mathbf{p}(E|A) = \frac{\mathbf{p}(A')}{\mathbf{p}(A)} = \frac{\omega_{k+1}^{(n+1)}}{\omega_k^{(n)}} \frac{n!}{(n-k)!k!} \frac{(n+1-(k+1))(k+1)!}{(n+1)!} = \frac{k+1}{n+1} \frac{\omega_{k+1}^{(n+1)}}{\omega_k^{(n)}}$$

Si  $n$  es un número muy grande, la previsión de que hay  $k+1$  éxitos en una sucesión de  $n+1$  eventos intercambia- bles no puede ser muy distinta de la previsión de que hay  $k$  éxitos entre los primeros  $n$ . Asimismo, el cociente  $(k+1)/(n+1)$  diferirá apenas de  $k/n$ . Por lo tanto, si  $n \gg 1$ ,  $\mathbf{p}(E|A) \approx k/n$ , la frecuencia relativa registrada en la sucesión. En otras palabras, si un sujeto es coherente, cualesquiera que sean sus previsiones  $\omega_k^{(n)}$  sobre la frecuencia relativa de éxitos en una secuencia indefinidamente creciente de eventos intercambia- bles, su previsión de que el próximo evento será un éxito tiende a coincidir con la frecuencia relativa observada a medida que aumenta el número  $n$  de eventos registrados. Como lo mismo vale para cualquier sujeto coherente, las previsiones de todos ellos tienden a concordar si concuerdan sus datos.

Según de Finetti, este resultado explica por qué los estadísticos que comparten un mismo banco de datos tienen previsiones similares, y en general, cómo es que aprendemos de la experiencia. No porque corrija- mos nuestras previsiones a la luz de la información que nos va llegando.

<sup>22</sup> De Finetti 1974, vol. 1, p. 134. Conviene advertir, sí, que la probabilidad que una persona le atribuye a  $E$  si *piensa que llegará a saber* que  $A$  se cumple *no es necesariamente igual* a la probabilidad que la misma persona le atribuiría a  $E$  si *efectivamente supiera* que  $A$  se cumple. El propio de Finetti hace esta observación (1974, vol. 2, p. 203), pero no parece hacerse cargo del peligro, a mi modo de ver manifiesto, de que ella prive a su famoso teorema de casi toda importancia en la práctica científica real.

<sup>23</sup> Recordemos que  $\mathbf{p}(A')$  es la suma en pesos que  $X$  está dispuesto a pagar a cambio de ganar \$1 si ocurre  $A$  seguido de  $E$ . Por otra parte,  $X$  está dispuesto a pagar  $\mathbf{p}(A)$  para ganar \$1 si ocurre  $A$  y  $\mathbf{p}(E|A)$  para ganar \$1 una vez que sepa que ocurrió  $A$ . Por lo tanto, antes de estar informado sobre  $A$ ,  $X$  está dispuesto a pagar  $\mathbf{p}(A) \times \mathbf{p}(E|A)$  pesos para ganar \$1 si ocurre  $A$  y enseguida ocurre  $E$ . Si  $X$  es coherente, esta cantidad tiene que ser igual a  $\mathbf{p}(A')$ .

“Nada puede obligarlo a uno a reemplazar su previsión inicial, ni puede haber una justificación para tal reemplazo”.<sup>24</sup> Pero a medida que se conocen más sucesos, nuestras previsiones iniciales *bajo la condición* de estos sucesos deben reemplazar a nuestras previsiones iniciales incondicionadas. Por lo tanto, cuando quiera tenemos que habérmolas con eventos que juzgamos intercambiables, “una experiencia suficientemente rica nos lleva siempre a considerar probables distribuciones y frecuencias futuras cercanas a las que se han observado”.<sup>25</sup> De este modo, “la teoría subjetivista resuelve completamente el problema de la inducción si hay intercambiabilidad, lo cual corresponde al caso más considerado habitualmente, y conduce a las mismas conclusiones que generalmente se admiten o se demuestran mediante razonamientos vagos e imprecisos”.<sup>26</sup> Adviértase, sin embargo, que todo este “aprender de la experiencia” presupone que el aprendiz conciba los sucesos futuros como pertenecientes a la misma secuencia intercambiable que los eventos observados. La previsión puede aprovechar la información sólo si es enlazada con ella por conceptos que proveen un esquema para describir el campo de aplicación de aquella en los mismos términos empleados para transmitir esta. Al proyectar así creativamente el pasado hacia el futuro, nuestro entendimiento extiende el dominio de la objetividad. Aprender de la experiencia a la manera de de Finetti no le enseña a nadie qué sucesos individuales son eventos intercambiables ni qué esquemas descriptivos demandan una revisión.<sup>27</sup> Me da la impresión que de Finetti, no obstante su genio y su agudeza crítica, ha permanecido cautivo de un positivismo ingenuo, para el cual el mundo está hecho de eventos bien deslindados en sí mismos de una vez por todas, cuya adecuada clasificación viene dada con ellos. Por otra parte, si el análisis del acontecer en even-

<sup>24</sup> De Finetti 1974, vol. 2, p. 211. Denotemos con  $\mathbf{p}_t$  las previsiones del sujeto a la hora  $t$ . Sea  $E$  un evento desconocido a las horas  $t = 1$  y  $t = 2$ . Sea  $\mathcal{A}$  toda la información pertinente a  $E$  que el sujeto averigua entre  $t = 1$  y  $t = 2$ . De Finetti da por supuesto, sin demostrarlo, que  $\mathbf{p}_2(E) = \mathbf{p}_1(E|\mathcal{A})$ . Un argumento de David Lewis explicado por Paul Teller (1973, pp. 222 ss.) demuestra que, si el sujeto es coherente, tiene que asumir esta ecuación. Este argumento, claro está, es inaplicable si, como ocurre generalmente en la vida real, la información adquirida entre  $t = 1$  y  $t = 2$  incluye al menos un suceso  $Q$  tal que  $\mathbf{p}_1(Q)$  no está definida, o lleva a revisar el mismísimo esquema conceptual en términos del cual se concibió el dominio de  $\mathbf{p}_1$ .

<sup>25</sup> De Finetti 1937, en Kyburg y Smokler 1980, p. 102.

<sup>26</sup> De Finetti 1937, en Kyburg y Smokler 1980, p. 105.

<sup>27</sup> Hasta este punto, el presente párrafo es una traducción libre de pasajes tomados de Torretti (1990), pp. 217–18.

tos individuales y la clasificación de estos bajo conceptos generales son el fruto de iniciativas humanas, ya no cabe sostener que el origen subjetivo de las previsiones las separa tajantemente de otras aseveraciones científicas. Antes bien, todo invita a reconocer en la probabilidad un atributo más que nuestro entendimiento proyecta sobre las situaciones objetivas constituidas por él, consistente en la disposición cuantificable de éstas a dar lugar o no a tales o cuales eventos, a evolucionar así o asá. Esta es precisamente la idea propuesta en 1957 por Karl Popper, que paso a examinar.

• • •

En su *Lógica de la investigación científica* (1935) Popper habla extensamente de la probabilidad, defendiendo una filosofía frecuentista. Veinte años después se ha convencido de que esta posición es insostenible porque niega la realidad de las probabilidades de sucesos individuales: la próxima jugada de la ruleta, el ingreso de Fulana en una universidad determinada, el efecto curativo de tal dosis de radiación sobre el cáncer de Zutano. Según Popper no se la puede negar, no por la razón moral que expuse arriba, sino porque la mecánica cuántica y sus afines, que se cuentan entre las teorías físicas más fructíferas y mejor corroboradas, presuponen la asignación de probabilidades a sucesos singulares y a propiedades y relaciones de objetos individuales.

Popper (1957, p. 67) propuso originalmente su nuevo concepto de probabilidad con estas palabras:

Todo arreglo experimental es *capaz de producir*, si repetimos el experimento muy a menudo, una secuencia con frecuencias que dependen de este particular arreglo experimental. Estas frecuencias virtuales pueden llamarse probabilidades. Pero como resulta que las probabilidades dependen del arreglo experimental, se las puede considerar como *propiedades de este arreglo*. *Caracterizan la disposición, o la propensión*, del arreglo experimental para generar ciertas frecuencias características *cuando el experimento es repetido a menudo*.<sup>28</sup>

<sup>28</sup> Aduciendo las dificultades con que tropieza la atribución de probabilidades a sucesos singulares —de las que hablaré enseguida—, Donald Gillies (2000) propone tomar al pie de la letra estas palabras de Popper y entender que la probabilidad de un suceso no es otra cosa que *la propensión de un arreglo experimental a producir ese resultado con cierta frecuencia característica cuando el experimento pertinente se repite a menudo*. Este modo de ver contraría patentemente el propósito declarado de la concepción propensionista popperiana. Además la cantidad física igualada así a la ‘probabilidad’ está lejos de ser una cantidad definida. En efecto, la

Un experimento ejecutado con un arreglo de esta clase se dice *aleatorio*.

La idea de Popper no es del todo nueva. Está implícita, como vimos, en un texto de Galileo. Además, una nota redactada por Peirce en 1910 y publicada póstumamente en 1933, contiene este pasaje notable:

Tengo, pues, que definir el significado del siguiente aserto: la *probabilidad* de que el dado, tirado desde un cubilete, exhiba un número divisible por tres es igual a un tercio. Este aserto significa que el dado tiene una cierta "potencialidad" (*a certain "would-be"*); y decir que el dado tiene una "potencialidad" es decir que tiene una propiedad, bastante parecida a cualquier *hábito* que pudiese tener un hombre. Solo que la "potencialidad" del dado es presumiblemente tanto más simple y más definida que el hábito humano, cuanto la composición homogénea y la forma cúbica del dado es más simple que la naturaleza del sistema nervioso y el alma del hombre; y así como, para definir un hábito de una persona, sería necesario describir cómo la llevaría a conducirse y en qué clase de ocasiones —aunque esta descripción de ningún modo implicaría que el hábito *consiste* en esa acción— así, para definir la "potencialidad" del dado es necesario decir cómo llevaría al dado a conducirse en una ocasión que saque a luz todas las consecuencias de la "potencialidad"; y esta aseveración no implicará por sí sola que la "potencialidad" del dado *consiste* en este comportamiento.

(Peirce CP, 2.664; cf. 8.225s.)

En cierto modo, la *facilità* de Galileo, el *would-be* de Peirce y la propensión de Popper constituyen una versión cuantitativa moderna del ser en potencia aristotélico. Cabe preguntarse pues:

- (i) ¿Cuál es el sujeto de la potencialidad cuantificada? En otras palabras, ¿qué cosa, situación o proceso real es el portador de las probabilidades?
- (ii) ¿Cuál es el acto en que la potencialidad se realiza? En otras palabras, ¿qué es lo probable, en virtud de la probabilidad?

---

frecuencia relativa  $f_n(A)$  de cierto resultado  $A$  en una serie de  $n$  realizaciones de un experimento aleatorio difiere necesariamente de la frecuencia relativa  $f_{n+1}(A)$  obtenida al repetir una vez más el experimento (a menos que  $f_n(A) = 0$ ). A mi modo de ver, sin embargo, la imprecisión inevitable en la aplicación de los conceptos científicos a la experiencia no recomienda ni justifica la adopción de conceptos científicos intrínsecamente imprecisos.

- (iii) ¿Cómo se mide la probabilidad? En particular, ¿qué hechos observables podrían confirmar que su valor es tal o cual, con tal o cual margen de error?

A primera vista, parecería que para Peirce el portador de las probabilidades en su ejemplo es el dado mismo. Como bien señala Gillies (2000, p. 118), esta respuesta a la pregunta (i) es inaceptable: la probabilidad de que salga un número divisible por tres al tirar el dado, no depende solo de la construcción de este y el modo como están pintadas sus caras, sino también del cubilete y como se lo sacude con el dado dentro, y del modo de lanzar el dado sobre la mesa. Por eso Popper, en el pasaje citado, dice que las probabilidades son una propiedad de *un arreglo experimental*, esto es, una configuración de diversas cosas interrelacionadas que genera cierto tipo de eventos cuando se la administra en cierta forma.<sup>29</sup>

Si la decisión de concebir probabilidad como propensión se toma, con Popper, *porque* así podremos asignar probabilidades a los sucesos individuales literalmente y no solo por metonimia, la pregunta (ii) admite obviamente una sola respuesta: *probables* son *los sucesos individuales* que pueden resultar de *un experimento* ejecutado mediante el *arreglo experimental* portador de la probabilidad; cada uno de esos resultados actualiza, cuando ocurre, una potencialidad del arreglo experimental y la distribución de probabilidades de que este es portador mide justamente esas potencialidades.

La pregunta (iii) es más difícil e importante que las otras dos y demanda por ello una atención más prolongada. Esbozaré una respuesta a la luz de lo que sabemos sobre los conceptos cuantitativos de la física. ¿Cómo se mide, por ejemplo, la masa de un planeta, o la resistencia de un circuito eléctrico, o el calor específico de un compuesto químico? En todos estos casos, se adopta un modelo del objeto físico en cuestión —esto es, una representación simplificada e idealizada del mismo— que sea a su vez una realización —esto es, un "modelo" en la acepción que se da a

---

<sup>29</sup> No creo, sin embargo, que Peirce esté cometiendo un error cuando le atribuye el *would-be* al dado, sino que se expresa con cierta laxitud, como conviene, por lo demás en una nota que el autor no preparó para la imprenta. Es inverosímil que Peirce no se diera cuenta, por ejemplo, de que un dado que cae por un tubo vertical tan estrecho que le deja solo un grado de libertad no tiene la más mínima probabilidad de "hacer" un número distinto del que aparece en la cara que da hacia la parte superior del tubo.

esta palabra en la llamada "teoría de modelos"— de una determinada teoría física.

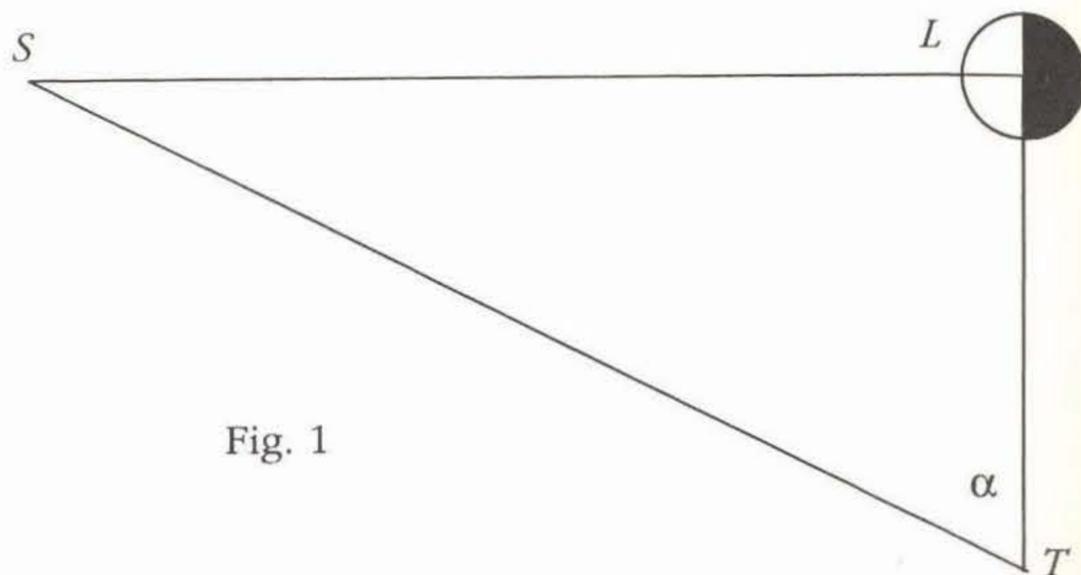


Fig. 1

Un sencillo ejemplo histórico ayudará a entender lo que está en juego aquí. Para determinar cuánto mayor es en realidad el disco del Sol que el de la Luna, Aristarco de Samos supuso implícitamente que la Tierra, la Luna y el Sol son cuerpos inmersos en un espacio euclidiano. Supuso además que, en este espacio, la luz se difunde por líneas rectas desde el Sol a la Tierra y la Luna y desde esta a la Tierra. Como la Luna cubre ceñidamente al Sol en los eclipses totales, los supuestos antedichos implican que los discos de ambos astros están en la misma proporción que sus respectivas distancias a la Tierra. Cuando la luna  $L$ , vista desde la tierra  $T$ , aparece dividida en dos mitades, una oscura y la otra iluminada por el sol  $S$ , los segmentos rectos  $LT$  y  $LS$  forman los catetos de un triángulo rectángulo con hipotenusa  $TS$  (Fig. 1). Por lo tanto, el cociente  $LT:TS$  entre la distancia de la Luna a la Tierra y la distancia de la Tierra al Sol es igual al coseno del ángulo  $\sphericalangle LTS$ , que llamaré  $\alpha$ . Un observador colocado sobre la Tierra puede fácilmente medir el ángulo  $\alpha$  entre la recta  $TL$  en dirección de la Luna y la recta  $TS$  en dirección del Sol; si bien la precisión del resultado dependerá, claro, de la calidad de sus instrumentos. Aristarco obtuvo  $\alpha = 87^\circ$ . Si este valor fuera correcto, la distancia  $TS$  entre la Tierra y el Sol sería 19,11 veces mayor que la distancia  $LT$  entre la Tierra y la Lu-

na.<sup>30</sup> Por lo tanto, concluye Aristarco, el disco del Sol es una 19 veces mayor que el de la Luna. En verdad, es alrededor de 400 veces mayor. Pero el error de Aristarco no viene de su razonamiento ni del modelo geométrico adoptado por él, sino que resulta simplemente del valor  $\alpha = 87^\circ$  que obtuvo de la observación. Los textos de astronomía actuales ponen  $\alpha = 89^\circ 50'$ . En rigor, el coseno de este ángulo está más cerca de  $1/344$  que de  $1/400$ . Para que  $LT:TS = (1/\cos \alpha) \approx 400$ , hay que poner  $\alpha = 89^\circ 51' 24''$ . Sin embargo, es inútil expresar estas cantidades con tanta precisión, puesto que las distancias de la Tierra al Sol y a la Luna son variables y el ángulo  $\alpha$  no tiene el mismo valor cada vez que la Luna se presenta dividida en una mitad oscura y otra clara.<sup>31</sup> Se habrá advertido que, en el procedimiento de Aristarco, no se mide directamente la cantidad que se quiere conocer, sino otra ligada con aquella por inferencias autorizadas por el modelo matemático. Lo mismo ocurre normalmente, y en formas mucho más complicadas, con la medición de otras cantidades físicas, como las mencionadas arriba.

Análogamente, para medir la propensión de un arreglo experimental  $\mathcal{S}$  a generar uno u otro de diversos resultados posibles cuando se experimenta con él, concebimos un modelo de la teoría matemática de las probabilidades que represente el espectro de sus resultados alternativos. Casi todos los físicos y matemáticos aceptan que esta teoría quedó satisfactoriamente caracterizada por el sistema axiomático propuesto por Kolmogorov (1933) y aquí me plegaré a esta abrumadora mayoría.<sup>32</sup> Si  $\mathcal{S}$

<sup>30</sup> Aristarco, que no tenía una calculadora de mano ni una tabla de cosenos a su disposición, ofrece la siguiente desigualdad:  $18 < TS:LT < 20$  (*Sobre los tamaños y las distancias del sol y la luna*, Proposición 7; en Heath 1913, p. 376).

<sup>31</sup> Al subestimar en menos de un 3,2% el ángulo  $\alpha$ , que era su dato empírico, Aristarco acabó sobrestimando en un 2.000% el cociente  $LT:TS$ , que era la cantidad buscada. Ello ilustra un importante distinguo relativo a las cualidades de un modelo matemático. Un modelo puede ser perfectamente *adecuado*, en cuanto permite calcular resultados exactos a partir de datos exactos, y sin embargo no ser *idóneo*, si un error en los datos, que es probable o aun inevitable con las técnicas de medición de que se dispone, puede generar un error intolerable en los resultados. Evidentemente, para la empresa del conocimiento, lo que interesa es que los modelos sea idóneos, aunque no sean adecuados.

<sup>32</sup> La discrepancia minoritaria más notable es el rechazo por parte de Bruno de Finetti del axioma **K3** enunciado en la nota 32. Según de Finetti, la probabilidad no tiene que ser una función  $\sigma$ -aditiva. De hecho, no podría serlo, si las probabilidades son grados de creencia.

puede producir solo un número finito de resultados diferentes, se puede usar un modelo de la siguiente versión muy sencilla de la teoría de Kolmogorov:

Un espacio aleatorio es un par  $\langle \Omega, \mathbf{p} \rangle$ , donde  $\Omega$  es un conjunto finito de objetos y  $\mathbf{p}$  es una función definida en el conjunto  $\wp\Omega$  de todos los subconjuntos de  $\Omega$  y con valores en el intervalo  $[0,1] \subseteq \mathbb{R}$ , la cual satisface las dos condiciones siguientes:

**K1.**  $\mathbf{p}(\Omega) = 1$ .

**K2.** Si  $A$  y  $B$  son dos miembros de  $\wp\Omega$  tales que  $A \cap B = \emptyset$ , entonces  $\mathbf{p}(A \cup B) = \mathbf{p}(A) + \mathbf{p}(B)$ .

---

Fetzer (1981, pp. 59–67) propuso un sistema alternativo de axiomas, elaborado conjuntamente con Nute, ajustado a su propia concepción filosófica de la probabilidad como propensión; pero, veinte años después de su publicación, no hay el menor indicio de que vaya a usarse nunca para trabajar con probabilidades en la vida real.

Gillies (2000), a quien debo esta observación, sugiere por su parte un cambio menos radical en los axiomas de Kolmogorov: en vez de definir —con Kolmogorov— la probabilidad condicional  $\mathbf{p}(A|B)$  del evento  $A$  dado el evento  $B$  como el cociente  $\mathbf{p}(A \cap B)/\mathbf{p}(B)$ , Gillies propone tomar la probabilidad condicional como un concepto primitivo, caracterizado por el axioma siguiente:

$$\mathbf{p}(A|B)\mathbf{p}(B) = \mathbf{p}(A \cap B)$$

No logro ver la ventaja de ello. La probabilidad condicional caracterizada por el axioma de Gillies tiene las mismas propiedades que la cantidad homónima definida por Kolmogorov, con una sola excepción: la probabilidad condicional  $\mathbf{p}(A|B)$  de Kolmogorov *no existe* en el caso en que la probabilidad  $\mathbf{p}(B)$  de la condición sea igual a 0; en cambio, en ese caso, la probabilidad condicional de Gillies existe, sí, pero está completamente indeterminada. En efecto, si  $\mathbf{p}(B) = 0$ , entonces  $\mathbf{p}(A \cap B) = 0$ , y por lo tanto, conforme al axioma de Gillies,  $\mathbf{p}(A|B)$  puede tomar cualquier valor. Me parece comprensible que el matemático Kolmogorov haya pensado que una cantidad que, en un caso dado, no tiene un valor definido, debía representarse mediante una función que no está definida en ese caso. Por otra parte, Gillies me ha señalado —en comunicación privada— que “aunque no se pueda calcular  $\mathbf{p}(A|B)$  matemáticamente, podría ser posible obtener una estimación empírica de su valor a partir de datos sobre frecuencias; de modo que no es vacuo postular que  $\mathbf{p}(A|B)$  tiene un valor definido.”

En los últimos años, se han propuesto estructuras más generales que la contemplada por Kolmogorov, aplicables a casos en que conviene suponer que la probabilidad evoluciona en el tiempo, incluso aleatoriamente. Tales extensiones de la teoría de Kolmogorov no discrepan de ella, sino que confirman el rumbo que él imprimió al estudio matemático de las probabilidades hace ya 70 años. Véase, por ejemplo, Gudder 1988, cap. 5, “Generalized Probability Spaces”, y Cap. 6, “Probability Manifolds”.

La función  $\mathbf{p}$  se llama *probabilidad*. Los elementos de  $\wp\Omega$  se llaman *eventos*; los elementos de  $\Omega$  son los *eventos elementales*.<sup>33</sup>

Los axiomas implican que, si  $A'$  es el complemento de  $A$ , entonces  $\mathbf{p}(A') = 1 - \mathbf{p}(A)$ . De aquí se deduce inmediatamente que  $\mathbf{p}(\emptyset) = 0$ .

La representación del dado de Peirce mediante un modelo de la teoría enunciada es muy simple.  $\Omega$  consta de seis elementos, correspondientes a los seis números que es posible hacer con el dado. Si el dado es perfectamente simétrico, la única diferencia entre sus caras es el número de puntitos que aparece en cada una. El procedimiento estándar para sacudir el dado en el cubilete y tirarlo sobre una mesa está pensado de tal modo que esta diferencia no afecte para nada el curso del experimento. Por eso, ningún jugador de dados razonable rechazaría un acuerdo en virtud del cual, por ejemplo, si el resultado obtenido es  $x$ , se lo lea como  $x + 2 \pmod{6}$ , o aplicándole cualquier otra permutación predeterminada del conjunto  $\{1,2,3,4,5,6\}$ . Se podría recurrir asimismo a un segundo experimento aleatorio —por ejemplo, una jugada de ruleta— para determinar cuál de las 720 permutaciones posibles se aplicará en cada ocasión.

---

<sup>33</sup> En la aplicación de los axiomas de Kolmogorov a un arreglo experimental  $\mathcal{S}$ , el conjunto  $\Omega$  representa los distintos resultados experimentales, mutuamente excluyentes, generables por  $\mathcal{S}$ . Si estos forman un continuo y, como es habitual, se acepta una teoría de conjuntos en que vale el axioma de elección de Zermelo, el dominio de definición de la probabilidad  $\mathbf{p}$  no puede ser el conjunto potencia  $\wp\Omega$ , porque  $\mathbf{p}$  se concibe entonces como una medida (en el sentido de Lebesgue) y, bajo el axioma de elección, no todos los subconjuntos de un continuo son medibles. Por eso, en la formulación estándar de la teoría de Kolmogorov, un espacio aleatorio es caracterizado como un triple  $\langle \Omega, \mathcal{B}, \mathbf{p} \rangle$ , donde  $\mathcal{B}$  es un cuerpo de Borel sobre  $\Omega$  (*vide infra*).  $\mathcal{B}$ , que sustituye a  $\wp\Omega$  en el axioma **K2**, es ahora el conjunto de los *eventos* del espacio aleatorio, y la *probabilidad*  $\mathbf{p}$  cumple el axioma siguiente además de los otros dos:

**K3.** Si  $A_1, A_2, \dots$  es una secuencia de elementos de  $\mathcal{B}$ , tales que  $A_b \cap A_k = \emptyset$  excepto si  $b = k$ , entonces  $\mathbf{p}\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} \mathbf{p}(A_k)$ .

(Como dicen los matemáticos: la probabilidad es una función  $\sigma$ -aditiva).

Recordemos que, si  $M$  es un conjunto cualquiera, un conjunto  $\mathcal{B} \subseteq \wp M$  es un *cuerpo de Borel* sobre  $M$  si cumple las condiciones siguientes:

**B1**  $M \in \mathcal{B}$ .

**B2** Si  $A \in \mathcal{B}$ , su complemento  $A' = {}_M - A \in \mathcal{B}$ .

**B3** Si  $(A_k)_{k \in \omega}$  es una familia numerable de elementos de  $\mathcal{B}$ , su unión  $\bigcup_{k \in \omega} A_k \in \mathcal{B}$ .

Es claro, entonces, que en el caso ideal en que el dado es simétrico y el procedimiento estándar para jugar con él se aplica rectamente, la función  $\mathbf{p}$  tiene que tomar el mismo valor sobre cada elemento de  $\Omega$ ; como esos elementos son 6, la probabilidad de cada evento elemental tiene que ser  $1/6$ .

No existe un poliedro simétrico con más de veinte caras. Pero un experimento aleatorio que admita cualquier número finito  $N$  de resultados alternativos igualmente fáciles se representa mediante el familiar modelo de la urna que contiene  $N$  bolitas bien mezcladas y que solo difieren entre sí por propiedades que no ejercen la más mínima influencia en el procedimiento de extracción. Cada resultado posible se representa mediante la extracción de una bolita determinada. En las condiciones descritas, la probabilidad de cada extracción no puede ser otra que  $1/N$ .

Sin embargo, en la vida real, los resultados alternativos de un experimento aleatorio no tienen que ser igualmente fáciles.<sup>34</sup> ¿Cómo se establece la probabilidad de cada resultado, cuando el arreglo que los genera no ostenta simetrías evidentes? La probabilidad  $\mathbf{p}(H)$  de un evento elemental  $H$  no es directamente observable; pero, tal como en otras ramas de la ciencia natural, podemos *suponer* que  $\mathbf{p}(H)$  tiene un valor determinado, calcular las consecuencias de esta hipótesis y confrontar las consecuencias así calculadas con la experiencia. La dificultad está en que, dada una probabilidad, sus consecuencias calculables consisten siempre en otras probabilidades, nunca en un hecho mondo y lirondo. Con todo, en virtud de las leyes matemáticas de grandes números, partiendo de cualesquiera probabilidades atribuidas a  $H$ , es posible predecir, con probabilidades muy altas, el valor aproximado de la frecuencia relativa de  $H$  en una secuencia larga de resultados producidos por el arreglo experimental pertinente. La hipótesis probabilística puede considerarse corroborada si tales predicciones se cumplen y descartarse si fracasan.

<sup>34</sup> Me refiero a las alternativas discernibles *prima facie*. Evidentemente, siempre es posible representar un experimento que admite dos resultados claramente discernibles  $A$  y  $B$ , tales que  $A$  es el doble de fácil que  $B$ , mediante un modelo que contemple tres eventos elementales equiprobables, dos de los cuales corresponderían al resultado  $A$  y uno al resultado  $B$ . Este recurso está disponible también para un espectro continuo de resultados, si reemplazamos la teoría clásica de los números reales por el análisis no estándar de Abraham Robinson (1966). En efecto, en virtud de un teorema de Rolando Chuaqui (1991, p. 351), es posible representar cualquier proceso estocástico indexado por un conjunto hiperfinito  $T$  mediante otro infinitesimalmente equivalente, definido sobre un espacio de probabilidad cuyos eventos elementales son equiprobables.

Ya me he referido al teorema de Bernoulli. Sea, como entonces,  $f_n(H)$  la frecuencia relativa del evento  $H$  en una secuencia de  $n$  repeticiones del experimento que lo produce. Sea  $\mathbf{p}(H)$  la probabilidad asignada hipotéticamente a  $H$ . Entonces, es una consecuencia lógica de los axiomas **K1** y **K2** que, a medida que  $n$  aumenta indefinidamente, converge a 0 la probabilidad de que la diferencia absoluta entre  $f_n(H)$  y  $\mathbf{p}(H)$  supere una cantidad arbitrariamente escogida  $\varepsilon > 0$ , por muy pequeña que esta sea. Los propósitos del investigador deciden en cada caso que valor hay que darle a  $n$  y  $\varepsilon$  para que la probabilidad inferida equivalga *prácticamente* a una certeza. Después de Bernoulli se han demostrado otros teoremas que permiten alcanzar la certeza práctica para un determinado propósito con valores más pequeños de  $n$  y  $\varepsilon$ . Pero la idea esencial es la misma: Dada una determinada distribución de probabilidades entre los resultados experimentales posibles, se puede computar la probabilidad de que dicha distribución coincida aproximadamente con la proporción en que se presentan los resultados cuando el experimento se repite  $n$  veces. Al crecer  $n$  sin límite, la probabilidad computada se acerca sin cesar a la certeza, aunque se mejore todo lo que se quiera la aproximación exigida.

Sea  $\mathcal{S}$  un arreglo experimental que produce distintas clases de resultados:  $A_1, \dots, A_n$ . Si ocurre que en toda secuencia larga de resultados de  $\mathcal{S}$  la frecuencia relativa de los resultados de cada clase tiene a estabilizarse en torno a un valor característico de esa clase —esto es, si  $\mathcal{S}$  y sus resultados obedecen a la ley empírica de grandes números— los experimentos efectuados con  $\mathcal{S}$  se consideran aleatorios y se trata de representarlos mediante un modelo de la teoría de probabilidades. Sea  $\langle \Omega, \mathbf{p} \rangle$  el modelo propuesto, con  $\Omega = \{A_1, \dots, A_n\}$ . El siguiente razonamiento ilustra esquemáticamente como este modelo puede ponerse a prueba. Sea  $N$  un número realmente enorme. Entonces es prácticamente seguro que, entre la probabilidad del resultado  $A_k$  y su frecuencia relativa  $f_N(A_k)$  en una secuencia de  $N$  resultados de  $\mathcal{S}$ , la diferencia es insignificante. Si  $N$  es muy grande, generalmente no será posible hacer  $N$  experimentos y observar sus resultados; pero podremos ver una secuencia más breve de, digamos,  $r$  experimentos como una muestra aleatoria tomada de una posible secuencia mayor (como si los resultados correspondiesen, por así decir, a  $r$  bolitas extraídas a ciegas de una urna que contiene  $N$  bolitas). Entonces, el valor observado de  $f_r(A_k)$  permite calcular el valor aproximado de  $f_N(A_k)$  con un margen de imprecisión y una probabilidad de error que

disminuyen según aumenta el cociente  $r/N$  y pueden determinarse exactamente para cada valor de  $r/N$ .

A la luz de lo dicho, parece claramente posible establecer que un determinado modelo probabilístico representa apropiadamente las propensiones de un arreglo experimental específico. Tal conclusión no es infalible, pero su certeza no tiene que ser menor que la de la mayoría de las hipótesis científicas. Hay, con todo, una dificultad que, a juicio de algunos, impediría atribuirle una probabilidad determinada a cada resultado singular generable por un experimento de esa especie. La explicaré mediante un ejemplo. Sea  $A$  un átomo de  $^{226}\text{Ra}$  (el isótopo del elemento radio cuyo núcleo atómico contiene 226 bariones). La física vigente asigna una probabilidad precisa a la desintegración radioactiva espontánea de  $A$  dentro del plazo de un año. Pero el valor de esa probabilidad varía, según que  $A$  sea descrito como (a) un átomo de  $^{226}\text{Ra}$ , (b) un átomo de radio, (c) un átomo de algún elemento radioactivo, (d) un átomo cualquiera. ¿Cuál de esos valores es la verdadera probabilidad de que  $a$  se desintegre espontáneamente en los próximos doce meses? Un físico diría hoy sin vacilar que es la correspondiente a la descripción (a); y durante el breve período en que ya se sabía que la desintegración radioactiva es aleatoria pero aún no se habían descubierto los isótopos habría referido el átomo  $A$  a la clase (b). No creo que ningún físico se haya molestado nunca en calcular las probabilidades correspondientes a las descripciones (c) y (d). En suma, el físico asigna al suceso individual la probabilidad que le corresponde al describirlo como perteneciente a la clase de referencia más estrecha en que está contenido. Ocurre, sin embargo, que en la mayoría de las aplicaciones diarias del concepto de probabilidad no se sabe cuál es la clase de referencia mínima de cada suceso individual de interés, o que esos sucesos son tan idiosincráticos que la clase más estrecha de cada uno no lo contiene más que a él.<sup>35</sup> Ayer (1963) objetó por eso a la atribu-

<sup>35</sup> Un claro ejemplo de ello es la muerte dentro del próximo año de una persona determinada, viva hoy. Las compañías de seguros calculan su probabilidad, clasificando a esa persona por su edad, su sexo, su profesión, su nacionalidad, para ajustar lo mejor posible la prima que cobran por asegurar su vida, al riesgo efectivo de tener que pagar la suma contratada. Por otra parte, no pueden llevar el ajuste demasiado lejos so pena de que el negocio de contratar un seguro se vuelva poco atractivo para los asegurados. Para el asegurador suele ser conveniente otorgar una póliza colectiva a un grupo bien estudiado y de bajo riesgo, sin entrar en los detalles de cada individuo perteneciente a él, asegurando, por ejemplo, sin previo examen médico a todos los empleados de un banco o una universidad, con una prima uniforme, o dependiente solo de la edad.

ción de probabilidades a sucesos individuales; según él, solo tendría sentido atribuirles a las condiciones específicas que se consideren ejemplificadas por ellos. Me parece, sin embargo, que la objeción de Ayer apunta solo a una dificultad en el empleo del concepto de probabilidad; incluso, si se quiere, a una necesidad de darle en el caso de tales sucesos idiosincráticos solamente un sentido pragmático, relativo al contexto de su aplicación. Tal vez, como sugiere Gillies, la concepción personalista de la probabilidad sea la única apropiada a tales casos. (De ahí la gran aceptación que ella tiene entre los practicantes de las ciencias sociales, mientras que los físicos tienden a rechazarla como carente de sentido para ellos). Pero dicha objeción no implica una imposibilidad lógica o física de adjudicar probabilidades a los sucesos individuales. En efecto, si la física actual está en lo cierto, cada átomo radioactivo individual tiene una probabilidad precisa de desintegrarse en el próximo año, la misma que cualquier otro átomo de ese isótopo. Y esta es la probabilidad de un suceso singular e irrepetible, pues cada átomo puede desintegrarse una sola vez.

Popper parece haber entendido su noción propensionista de la probabilidad como una generalización del concepto clásico de disposición causal. Con ello esta última pasa a ser, al revés, una restricción de aquella: "La causalidad es solo un caso especial de propensión: el caso de una propensión igual a 1" (Popper 1990, p. 20). Este modo de ver suscitó una objeción de Paul Humphreys, a mi modo de ver muy ingenua.<sup>36</sup> Humphreys señala con toda razón que la causalidad es una relación asimétrica y que su asimetría se refleja en el orden temporal de causa y efecto. De ello concluye que, si la propensión es una forma generalizada de causalidad, las probabilidades condicionales no pueden ser propensiones. El siguiente caso es típico de toda una serie de ejemplos con que distintos autores han ilustrado el argumento de Humphreys. Si  $\mathbf{p}(B|A)$  es la probabilidad de que un artículo elaborado en cierta fábrica con la máquina  $A$  tenga el defecto  $B$ , y la interpretamos como una propensión de la máquina  $A$  a entregar productos con ese defecto, tendremos que entender la probabilidad condicional inversa  $\mathbf{p}(A|B)$  como una propensión de los productos de esa fábrica que exhiben el defecto  $B$  a ser fabricados

<sup>36</sup> La objeción se publicó primero en la reseña de D.H. Mellor, *A Matter of Chance* por Wesley Salmon (1979), quien cita allí con aprobación una "comunicación privada" de Humphreys. El artículo de Humphreys, "Why propensities cannot be probabilities?" (1985) apareció seis años después.

con la máquina  $\mathcal{A}$ . Esta noción viola obviamente la asimetría temporal característica de la relación de causalidad. Aunque esta llamada “paradoja de Humphreys” suele mencionarse en la literatura como un gran hallazgo, no resiste el menor examen. Desde luego, que la propensión sea una generalización de la causalidad no implica que herede su asimetría. Por ejemplo, la relación general de orden no es asimétrica —pues incluye el orden cíclico o circular— aunque es una generalización del orden lineal; el concepto topológico de función continua prescinde de toda noción métrica, aunque es una generalización del concepto de función continua real o compleja que supone tales nociones. Por otra parte, el que las probabilidades sean propensiones no implica que también tengan que serlo las probabilidades condicionales. En efecto, una probabilidad condicional entre eventos de un espacio aleatorio finito  $\langle \Omega, \mathbf{p} \rangle$  es un cociente de dos valores de la función  $\mathbf{p}$ :

$$\mathbf{p}(A | B) = \frac{\mathbf{p}(A \cap B)}{\mathbf{p}(B)} \quad (\text{siempre que } \mathbf{p}(B) > 0)$$

Y sería muy raro que el cociente de dos cantidades físicas de cierta clase sea una cantidad física de esa misma clase. La confusión de Humphreys y los demás autores que se dejaron ganar por su “paradoja” es comprensible, pues las probabilidades condicionales entre los eventos  $A_1, A_2, \dots$  de un espacio aleatorio  $\langle \Omega, \mathbf{p} \rangle$  y un evento fijo  $B \in \wp \Omega$  (tal que  $\mathbf{p}(B) > 0$ ) son precisamente iguales a las probabilidades de los eventos correspondientes  $A'_1, A'_2, \dots$  de otro espacio aleatorio  $\langle \Omega', \mathbf{p}' \rangle$  definido por las ecuaciones:

$$\Omega' = \{A'_k : A'_k = A_k \cap B \text{ para cada } k \text{ tal que } A_k \in \Omega\}$$

$$\mathbf{p}'(A'_k) = \mathbf{p}(A_k | B)$$

Si el evento  $B$  consiste, como arriba, en ser un producto defectuoso de una determinada fábrica, y  $A_k$  designa el hecho de provenir de una determinada máquina,  $\mathbf{p}'(A'_k)$  mediría la tendencia de los productos defectuosos a venir de la máquina  $A_k$ . No veo inconveniente en reconocer que hay propensiones de este tipo retrospectivo, que los pedantes, si quieren, pueden llamar “retropensiones”. Obviamente, no son disposiciones causales; pero tampoco tienen que serlo las tendencias prospectivas o propensiones propiamente dichas. Según todo lo que se sabe sobre la radioactividad, la desintegración radioactiva de un átomo, cuando se

produce, no es causada por nada y, por lo tanto, la propensión bien conocida de cada átomo radioactivo a desintegrarse no puede entenderse como una disposición causal.

Al liberar la idea de propensión de una pretendida conexión conceptual con la causalidad, evitamos también la fatal tentación en que Popper cayó en sus últimos años. Dejando de lado los arreglos experimentales y sus efectos aleatorios repetibles, Popper (1990, p. 17) escribe: “...las propensiones en la física son propiedades de la situación física entera y a veces del modo particular como ésta cambia”. Su discípulo David Miller ha hecho suya con entusiasmo esta tesis, que “asciende” a la probabilidad entendida como propensión del nivel humilde de los conceptos científicos sometidos a control empírico a la soberanía propia de un concepto metafísico. Dice Miller: “En la interpretación propensionista, la probabilidad de un resultado no es una medida de una frecuencia, sino...una medida de la inclinación del presente estado de cosas a realizar ese resultado” (1994, p. 182). Y agrega:

Es lamentable, por eso, que...nos topemos con observaciones [anteriores de Popper] que...pintan las propensiones como “tendencias a producir frecuencias relativas al repetirse condiciones o circunstancias similares”... Las propensiones no están localizadas en cosas físicas, ni tampoco en situaciones locales. Estrictamente, toda propensión (absoluta o condicional) tiene que estar referida a la situación completa del universo (o el cono de luz) en ese tiempo.<sup>37</sup> Las propensiones dependen de la situación hoy, no de otras situaciones, por muy similares que sean. Solo así alcanzamos la especificidad requerida para resolver el problema del caso singular.

(Miller 1994, pp. 185–186)

Vimos ya, en el ejemplo de la radioactividad, que el aserto final es simplemente falso: cualquier átomo de un determinado isótopo tiene “la es-

<sup>37</sup> La expresión “cono de luz” proviene de la Teoría de la Relatividad. El *cono de luz* de un evento o punto espacio-temporal  $E$  es la hipersuperficie espacio-temporal que contiene todos los puntos espacio-temporales conectables con  $E$  por una curva “nula”, y por ende, idealmente, por una señal luminosa propagada en el vacío. Me cuesta entender que papel puede corresponder a la mitad futura del cono de luz de  $E$  en la determinación de las propensiones que pueden manifestarse en ese evento. Por otra parte, el texto de Miller alude a “the complete situation of the universe (or the light-cone) at the time” (cursiva mía). Como el estado de cosas sobre un determinado cono de luz —en la acepción normal del término— no cambia con el tiempo, me inclino a pensar que Miller usa aquí la expresión ‘the light-cone’ en una acepción esotérica, inventada por él.

pecificidad requerida para resolver el problema del caso singular"; en lo que concierne a su posible desintegración radioactiva la situación de cualquiera de ellos es en cada momento no solamente similar, sino idéntica; y la probabilidad de desintegración y la desintegración misma, si ocurre, no dependen para nada de la situación del resto del universo ni dentro, ni fuera, ni sobre el cono de luz (pasado, futuro, o pasado-y-futuro) de ese evento. Por otra parte, es ocioso aducir tales contraejemplos: la tesis de Miller es de esas cuyo solo enunciado la reduce al absurdo. Me interesaba, sin embargo, mencionarla por la analogía que presenta con el tratamiento de la causalidad por John Stuart Mill (1843). Tratando de rescatar para el pensamiento científico este abigarrado concepto de la vida cotidiana, también Mill llegó a la conclusión de que la *causa* de cualquier suceso solo podía consistir en la situación completa del universo que inmediatamente lo precede. Con ello, el insigne filósofo victoriano nos privó definitivamente de toda oportunidad de conocer la causa de nada. Afortunadamente, la noción misma de un estado del universo entero en el momento actual se vino abajo cuando Einstein (1905) relativizó la relación de simultaneidad. Confiemos en que Miller no pretende resucitarla en una versión revisada, como "estado del cono de luz en un tiempo  $t$ ".

Pienso que las consideraciones precedentes responden a las principales objeciones que han hecho valer contra la concepción propensionista de la probabilidad. Naturalmente, la fuerte resistencia que muchos filósofos le oponen no surge de tales detalles técnicos, sino de la negativa positivista a reconocer la posibilidad física como una noción primitiva, tan comprensible para todos como la noción de existencia actual. Si no hay cabida en el mundo para lo posible como tal, mucho menos la habrá para lo posible cuantificado. Y la explicación extensional de lo posible como existencia en algún mundo alternativo, que estuvo de moda en el último tercio del siglo XX, nunca acertó a encontrar criterios viables para medir o siquiera graduar la posibilidad. Pero este no es un asunto que corresponda examinar aquí. Me limito, pues, a manifestarme sorprendido de que tantas personas que la sociedad emplea como pensadores profesionales se nieguen a entender un concepto que un niño de tres años maneja con soltura.

## APÉNDICE I

La siguiente explicación de la paradoja de Bertrand está tomada del *Diccionario de Lógica y Filosofía de la Ciencia* por Jesús Mosterín y Roberto Torretti (Madrid: Alianza, 2002).

Bertrand (1907, pp. 4–5) propuso el siguiente problema: Se traza al azar una cuerda en el círculo  $K$  de centro  $O$  y radio  $r$ ; ¿cuál es la probabilidad de que ella sea menor que el lado de un triángulo equilátero inscrito en  $K$ ? Bertrand ofrece tres soluciones incompatibles (Fig. 2).

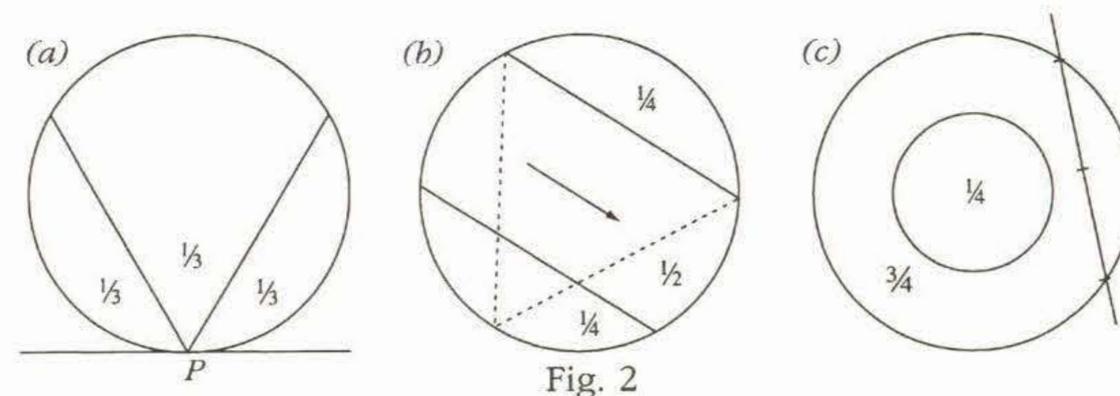


Fig. 2

(a) Si fijamos uno de los extremos de la cuerda, esta determinación no altera la probabilidad buscada, pues, en virtud de la simetría de  $K$ , ella no tiene ninguna influencia favorable o desfavorable sobre el evento en cuestión. Ahora bien, si uno de los extremos de la cuerda se fija, digamos en el punto  $P$ , el azar gobernará su dirección. Constrúyase un triángulo equilátero inscrito con un vértice en  $P$ . Los dos lados del triángulo que concurren en  $P$  forman entre sí y con la tangente que toca a  $K$  en  $P$  tres ángulos iguales. La cuerda será mayor que esos lados solo si cae en el ángulo entre ellos. La probabilidad de que la cuerda trazada al azar caiga en ese ángulo y no en uno de los que cada lado forma con la tangente es igual a  $1/3$ .

(b) Si fijamos la dirección de la cuerda, esta determinación no altera la probabilidad buscada. Ahora bien, la distancia entre el centro de gravedad de un triángulo equilátero y cada uno de sus vértices es igual al doble de la distancia desde ese punto a lado opuesto respectivo. Además, el centro de gravedad de un triángulo equilátero inscrito en  $K$  coincide con el centro  $O$ . Por lo tanto, si la dirección de la cuerda está dada, esta será mayor que el lado de un triángulo equilátero inscrito solo si corta el diámetro perpendicular a esa dirección a una distancia de  $O$  menor que  $1/2r$ .

La probabilidad de que la cuerda trazada al azar cumpla esta condición es igual a  $\frac{1}{2}$ .

(c) Si fijamos el punto medio de la cuerda, esta determinación no altera la probabilidad buscada. Como se vio en el caso (b), para que la cuerda trazada al azar sea mayor que el lado de un triángulo equilátero inscrito es necesario y suficiente que su punto medio diste de  $O$  menos que  $\frac{1}{2}r$ , y por lo tanto que dicho punto medio caiga en el interior de un círculo de superficie igual a la cuarta parte de la superficie de  $K$ ; la probabilidad de este evento es igual a  $\frac{1}{4}$ .

Bertrand concluye: Ninguna de las tres soluciones es falsa, ninguna es correcta; el problema está mal planteado. Con todo, Jaynes (1973) ha defendido la validez exclusiva de la primera solución, argumentando que es la única de las tres que es invariante bajo todas las simetrías del plano euclídeo: rotaciones, traslaciones y transformaciones de escala.

## APÉNDICE II

Daré aquí el enunciado y la demostración del Teorema de Representación de de Finetti, siguiendo a William Feller (1971, pp. 228–29). Pero antes tengo que definir algunos términos.

- (i) En el vocabulario estándar de la estadística y la teoría de las probabilidades se llama *variable aleatoria* a una función con valores reales definida sobre un espacio probabilístico, esto es, sobre un conjunto de eventos a los cuales se han asignado probabilidades; es claro que, si  $\mathbf{X}$  denota una variable aleatoria definida sobre  $\Omega$ , la probabilidad  $\mathbf{p}(\mathbf{X} = k)$  de que  $\mathbf{X}$  tome el valor  $k \in \mathbb{R}$  es igual a la suma de las probabilidades de todos los eventos mutuamente excluyentes  $u \in \Omega$  tales que  $\mathbf{X}(u) = k$ . Si el alcance de  $\mathbf{X}$  contiene solo un número finito o denumerable de valores diferentes,  $\mathbf{X}$  es una variable aleatoria *discreta*. (El concepto de variable aleatoria continua se define en el (iii)).
- (ii) Sea  $S$  un subconjunto del conjunto de eventos  $\Omega$ ; el *indicador* de  $S$  es la variable aleatoria que toma el valor 1 en todos los eventos comprendidos en  $S$  y el valor 0 en los eventos comprendidos en su complemento  $\Omega \setminus S$ .

- (iii) Una *distribución* (normalizada) es una función no decreciente  $F: \mathbb{R} \rightarrow [0,1]$  tal que  $\lim_{x \rightarrow -\infty} F(x) = 0$  y  $\lim_{x \rightarrow \infty} F(x) = 1$ . En particular, la distribución de la variable aleatoria  $\mathbf{X}$ , denotada por  $F_{\mathbf{X}}$ , se define por la ecuación  $F_{\mathbf{X}}(u) = \mathbf{p}(\mathbf{X} \leq u)$  ( $-\infty < u < \infty$ ). Si  $F_{\mathbf{X}}$  es continua y tiene una derivada, continua por segmentos, definida en todo el dominio de  $F_{\mathbf{X}}$  excepto a lo sumo en un número finito de puntos, se dice que  $\mathbf{X}$  es una variable aleatoria *continua*.
- (iv) Sea  $\mathbf{X}$  una variable aleatoria continua. Se dice que la distribución  $F_{\mathbf{X}}$  está *concentrada* en el intervalo  $[a,b]$  si  $\frac{dF_{\mathbf{X}}(x)}{dx} = 0$  para todo  $x \notin [a,b]$ .
- (v) La *expectativa*  $\mathbf{E}(\mathbf{X})$  de la variable aleatoria  $\mathbf{X}$  se define como sigue: (a) si  $\mathbf{X}$  es una variable aleatoria discreta que asume los valores  $x_1, x_2, \dots$  con las probabilidades  $\mathbf{p}_{\mathbf{X}}(x_1), \mathbf{p}_{\mathbf{X}}(x_2), \dots$ , respectivamente,  $\mathbf{E}(\mathbf{X}) = \sum_{k=1}^{\infty} x_k \mathbf{p}_{\mathbf{X}}(x_k)$ , siempre que esta serie converja absolutamente (si la serie  $\sum_{k=1}^{\infty} |x_k| \mathbf{p}_{\mathbf{X}}(x_k)$  es divergente decimos que  $\mathbf{X}$  no tiene una expectativa finita); (b) si  $\mathbf{X}$  es una variable aleatoria continua, con distribución  $f_{\mathbf{X}}$ ,  $\mathbf{E}(\mathbf{X}) = \int_{-\infty}^{\infty} \frac{x dF_{\mathbf{X}}(x)}{dx} dx$ , que podemos escribir  $\int_{-\infty}^{\infty} x dF_{\mathbf{X}}(x)$ .
- (vi) El *r-ésimo momento* de una variable aleatoria  $\mathbf{X}$  es la expectativa de la variable aleatoria  $\mathbf{X}^r$ . Por lo tanto, si  $\mathbf{X}$  es discreta y  $\mathbf{X}^r$  tiene expectativa finita, el *r-ésimo momento* de  $\mathbf{X}$  es  $\mathbf{E}(\mathbf{X}^r) = \sum_{k=1}^{\infty} x_k^r \mathbf{p}_{\mathbf{X}}(x_k)$ . Y si  $\mathbf{X}$  es continua, con distribución  $F_{\mathbf{X}}$ , el *r-ésimo momento* de  $\mathbf{X}$  es  $\mathbf{E}(\mathbf{X}^r) = \int_{-\infty}^{\infty} x^r dF_{\mathbf{X}}(x)$ . Esta última integral suele llamarse también *el r-ésimo momento de la distribución*  $F_{\mathbf{X}}$ . (Con esta nomenclatura, obviamente, el 0-ésimo momento de  $F_{\mathbf{X}}$  es  $\int_{-\infty}^{\infty} dF_{\mathbf{X}}(x) = F_{\mathbf{X}}$ ).
- (vii) Si  $\mathfrak{X} = \langle \mathbf{X}_1, \dots, \mathbf{X}_n \rangle$  es un *n-tuplo* de variables aleatorias, su *distribución n-dimensional* es la función  $F_{\mathfrak{X}}: \mathbb{R}^n \rightarrow [0,1]$  definida por  $F_{\mathfrak{X}}(u_1, \dots, u_n) = \mathbf{p}(\mathbf{X}_1 \leq u_1, \dots, \mathbf{X}_n \leq u_n)$  ( $-\infty < u_i < \infty$ ).

(viii) Las  $n$  variables aleatorias  $\mathbf{X}_1, \dots, \mathbf{X}_n$  se dicen *intercambiables* si y solo si el  $n$ -tuplo de variables aleatorias  $\langle \mathbf{X}_1, \dots, \mathbf{X}_n \rangle$  tiene la misma distribución  $n$ -dimensional que el  $n$ -tuplo  $\langle \mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(n)} \rangle$ , para cada permutación  $\sigma$  de  $\{1, \dots, n\}$ . Las variables aleatorias de una secuencia infinita  $\mathbf{X}_1, \mathbf{X}_2, \dots$  se dicen *intercambiables*, si  $\mathbf{X}_1, \dots, \mathbf{X}_n$  son intercambiables para cada  $n$ .

(ix) Sea  $a_0, a_1, \dots$  una secuencia de números; el *operador diferencia*  $\Delta$  se define por la relación  $\Delta a_i = a_{i+1} - a_i$ . Aplicando nuevamente el operador  $\Delta$  a la secuencia  $\Delta a_0, \Delta a_1, \dots$  obtenemos otra secuencia con elementos  $\Delta^2 a_i = a_{i+2} - 2a_{i+1} + a_i$ . Si denotamos la identidad con  $\Delta^0$  (esto es, si  $\Delta a_i = a_i$ ) y ponemos  $\Delta^r = \Delta \Delta^{r-1}$ , es fácil comprobar que, para todo entero  $n \geq 0$ ,

$$\Delta^n a_i = \sum_{k=0}^n \binom{n}{k} (-1)^{n-k} a_{i+k}$$

(x) Una secuencia de números  $a_0, a_1, \dots$  se dice *completamente monótona* si, para cualesquiera enteros  $k$  y  $n$  tales que  $0 \leq k \leq n$ ,

$$(-1)^n \Delta^n a_k \geq 0$$

El *Teorema de la Representación* de Bruno de Finetti puede ahora enunciarse así: *Si  $\mathbf{E}_1, \mathbf{E}_2, \dots$  es una secuencia infinita cualquiera de indicadores intercambiables (esto es, de variables aleatorias intercambiables que solo admiten los valores 1 y 0), entonces hay, para cada entero positivo  $n$  una distribución  $n$ -dimensional  $F$  concentrada en  $[0,1]$ , tal que*

$$(1) \quad \mathbf{P}(\mathbf{E}_1 = 1, \dots, \mathbf{E}_k = 1, \mathbf{E}_{k+1} = 0, \dots, \mathbf{E}_n = 0) = \int_0^1 \theta^k (1 - \theta)^{n-k} dF(\theta)$$

y

$$(2) \quad \mathbf{P}\left(\sum_{i=1}^n \mathbf{E}_i = k\right) = \binom{n}{k} \int_0^1 \theta^k (1 - \theta)^{n-k} dF(\theta)$$

Este teorema se deduce casi inmediatamente de un teorema de Hausdorff que Feller demuestra concisa y elegantemente (1971, pp. 225-27) y que aquí presento como lema:

Si  $c_0, c_1, \dots$  son los momentos 0-ésimo, primero, etc., de una distribución  $F$ , entonces  $c_0, c_1, \dots$  forman una secuencia completamente monótona con  $c_0 = 1$ . Recíprocamente, para cada secuencia arbitraria completamente monótona  $c_0, c_1, \dots$  con  $c_0 = 1$  existe una única distribución  $F$  cuyos momentos 0-ésimo, primero, etc., son respectivamente  $c_0, c_1, \dots$

Siguiendo a Feller, usaré la expresión  $p_{k,n}$  para abreviar el lado izquierdo de la ecuación (1),  $\mathbf{P}(\mathbf{E}_1 = 1, \dots, \mathbf{E}_k = 1, \mathbf{E}_{k+1} = 0, \dots, \mathbf{E}_n = 0)$ . Asimismo, pongo  $c_0 = 1$  y, para cada entero positivo  $n$ ,  $c_n = p_{n,n} = \mathbf{P}(\mathbf{E}_1 = 1, \dots, \mathbf{E}_n = 1)$ . No es difícil ver entonces que

$$(3) \quad p_{n-1,n} = p_{n-1,n-1} - p_{n,n} = -\Delta c_{n-1}$$

y, por lo tanto,

$$(4) \quad p_{n-2,n} = p_{n-2,n-1} - p_{n-1,n} = -\Delta^2 c_{n-2}$$

Prosiguiendo de este modo se obtiene, para todo  $k \leq n$ ,

$$(5) \quad p_{k,n} = p_{k,n-1} - p_{k+2,n} = (-1)^{n-k} \Delta^{n-k} c_k$$

Todas estas cantidades son no negativas, y por ende la secuencia  $c_0, c_1, \dots$  es completamente monótona. Por consiguiente, de acuerdo con el lema,  $c_r$  es el  $r$ -ésimo momento de una única distribución  $F$  y la ecuación (1) no hace sino expresar el contenido de la ecuación (5). La ecuación (2) se sigue inmediatamente, pues hay precisamente  $\binom{n}{k}$  modos como pueden ocurrir  $k$  éxitos en  $n$  experimentos.

Universidad de Chile

Profesor Emérito de la Universidad de Puerto Rico

## OBRAS CITADAS

- Aristoteles (1831) *Opera*. Ex recognitione I. Bekkeris edidit Academia Regia Borussica. Berlin: G. Reimer. 2 vols.
- Ayer, A. J. (1963), "Two notes on probability", en Ayer, *The Concept of a Person and other Essays*, London: Macmillan, pp. 188-208.
- Bernoulli, J. (1713), *Ars conjectandi*. Basilea: Impensis Thurnisiorum, fratrum.
- Bertrand, J. (1907), *Calcul des probabilités*, 2<sup>e</sup> édition. Paris: Gauthier-Villars. (Primera edición: 1888)
- Carnap, Rudolf (1950), *The Logical Foundations of Probability*. Chicago: University of Chicago Press.
- Carnap, Rudolf (1952), *The Continuum of Inductive Methods*. Chicago: University of Chicago Press.
- Carnap, Rudolf (1962), *The Logical Foundations of Probability*. Second edition. Chicago: University of Chicago Press.
- Carnap, Rudolf (1971/1980), "A Basic System of Inductive Logic", en R. Carnap y R. C. Jeffrey (1971), pp. 1-166, y R. C. Jeffrey (1980), pp. 7-155.
- Carnap, Rudolf y Richard C. Jeffrey, eds. (1971), *Studies in Inductive Logic and Probability*. Vol. I. Berkeley: University of California Press.
- Chuaqui, Rolando (1991), *Truth, Possibility, and Probability: New Logical Foundations of Probability and Statistical Inference*. Amsterdam: North-Holland.
- Cicerón, *Lucullus*, en Cicero, *De natura deorum, Academica*, with an English translation by H. Rackham. Cambridge MA: Harvard University Press, 1979 (Loeb Classical Library, 268), pp. 464-658.
- Cramér, H. (1946), *Mathematical Methods of Statistics*. Princeton: Princeton University Press.
- Daston, Lorraine (1988), *Classical Probability in the Enlightenment*. Princeton, NJ: Princeton University Press.
- de Finetti, Bruno (1930), "Funzione caratteristica di un fenomeno aleatorio", *Memorie della Reale Accademia dei Lincei*. IV 5: 86-133.
- de Finetti, Bruno (1931), "Sul significato soggettivo della probabilità", *Fundamenta Mathematica*, 17: 298-329.
- de Finetti, Bruno (1937), "La prevision: ses lois logiques, ses sources subjectives", *Annales de l'Institut Henri Poincaré*. 7: 1-68. (Traducido como "Foresight: Its logical laws, its subjective sources" en Kyburg and Smokler 1980, pp. 53-118; mis citas remiten a esta edición).
- de Finetti, Bruno (1974), *Theory of Probability: A critical introductory treatment*. Traducido por A. Machi and A. Smith. New York: Wiley. 2 vols.
- Einstein, Albert (1905), "Zur Elektrodynamik bewegter Körper", *Annalen der Physik*, 17: 891-921.

- Ellis, R. L. (1849), "On the foundations of the theory of probabilities", *Transactions of the Cambridge Philosophical Society*. 8: 1-6.
- Ellis, R. L. (1856), "Remarks on the fundamental principle of the theory of probabilities", *Transactions of the Cambridge Philosophical Society*. 9: 605-607.
- Feller, William (1971), *An Introduction to Probability Theory and its Applications*. Volume II. Second Edition. New York: Wiley, 1971.
- Fetzer, J. H. (1981), *Scientific Knowledge: Causation, Explanation, and Corroboration*. Dordrecht: Reidel.
- Galileo Galilei (1718), "Sopra le scoperte dei dadi", En G. Galilei. *Opere*. Edizione Nazionale. Firenze: Barbera, 1898. Vol. 8, pp. 591-594. (Publicado póstumamente en 1718; redactado tal vez hacia 1620),
- Gillies, Donald (2000), *Philosophical Theories of Probability*. London: Routledge.
- Gudder, Stanley P. (1988), *Quantum Probability*. San Diego, CA: Academic Press
- Hacking, Ian (1975), *The Emergence of Probability: A philosophical study of early ideas about probability, induction and statistical inference*. Cambridge: Cambridge University Press.
- Heath, T. L. (1913), *Aristarchos of Samos: The Ancient Copernicus*. A History of Greek Astronomy to Aristarchus together with Aristarchus's *Treatise on the Sizes and Distances of the Sun and Moon*, a new Greek text with translation and notes. Oxford: Clarendon Press.
- Hempel, Carl G. (1934), *Beiträge zur logischen Analyse des Wahrscheinlichkeitsbegriffs*. Jena: Neuenhahn.
- Hempel, Carl G. (1935), "Über den Gehalt von Wahrscheinlichkeitsaussagen", *Erkenntnis*, 5: 228-260.
- Hempel, Carl G. (2000), *Selected Philosophical Essays*. Chicago: University of Chicago Press.
- Humphreys, Paul (1985), "Why propensities cannot be probabilities?", *Philosophical Review*, 94: 557-570.
- Jaynes, E. T. (1973), "The well-posed problem", *Foundations of Physics*, 4: 477-492.
- Jeffrey, Richard C., ed. (1980), *Studies in Inductive Logic and Probability*. Vol. II. Berkeley: University of California Press.
- Keynes, John Maynard (1921), *A Treatise on Probability*. London: Macmillan, 1921.
- Kolmogorov, A. N. (1933), *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Berlin: Springer.
- Kyburg Jr., H. E. y H. E. Smokler, eds. (1980), *Studies in Subjective Probability*. Huntington, NY: Krieger.
- Laplace, Pierre Simon (1795), *Essai philosophique sur les probabilités*. En *Œuvres complètes de Laplace*, Paris: Académie des Sciences, 1878-1912, vol. 7, pp. i-cliii.
- Leibniz, Gottfried Wilhelm (SS), *Sämtliche Schriften und Briefe*, hrsg. von der Preussischen, bzw. Deutschen Akademie der Wissenschaften, 1923- .

- Lindley, D. V. (1965), *Introduction to Probability and Statistics from a Bayesian Viewpoint*. Cambridge: Cambridge University Press, 1965. 2 vols.
- Mellor, D. H. (1971), *A Matter of Chance*. Cambridge: Cambridge University Press.
- Mill, John Stuart (1843), *A System of Logic Ratiocinative and Inductive*. London: John W. Parker.
- Miller, David (1994), *Critical Rationalism: A Restatement and Defence*. Chicago: Open Court.
- Mosterín, Jesús y Roberto Torretti (2002), *Diccionario de lógica y filosofía de la ciencia*. Madrid: Alianza (en prensa).
- Peirce, C. S. (CP). *Collected Papers*. Edited by C. Hartshorne, P. Weiss and A.W. Burks. Cambridge MA: The Belknap Press of Harvard University Press, 1931–1958. 8 vols.
- Poisson, Denis Siméon (1837), *Recherches sur la probabilité des jugements en matière criminelle et en matière civile*. Paris: Bachelier.
- Popper, Karl R. (1957), "The propensity interpretation of the calculus of probability and the quantum theory", en S. Körner, ed. *Observation and Interpretation*, London: Butterworth, pp. 65–70.
- Popper, Karl R. (1959), "The propensity interpretation of probability", *British Journal for the Philosophy of Science*. **10**: 25–42.
- Popper, Karl R. (1990), *A World of Propensities*. London: Thoemmes.
- Ramsey, Frank P. (1931), "Truth and probability", en Ramsey, *The Foundations of Mathematics and other Logical Essays*, London: Routledge & Kegan Paul, pp. 156–198.
- Reichenbach, Hans (1935), *Wahrscheinlichkeitslehre: Eine Untersuchung über die logischen und mathematischen Grundlagen der Wahrscheinlichkeitsrechnung*. Leiden: A.W. Sijthoff.
- Reichenbach, Hans (1949), *The Theory of Probability: An Inquiry into the Logical and Mathematical Foundations of the Calculus of Probability*. Traducido por E. H. Hutten y M. Reichenbach. Berkeley: University of California Press.
- Révész, Pál (1968), *The Laws of Large Numbers*. New York: Academic Press.
- Robinson, Abraham (1966), *Non-Standard Analysis*. Amsterdam: North-Holland.
- Salmon, Wesley (1979), "Propensities: a Discussion Review of D. H. Mellor *The Matter of Chance*", *Erkenntnis*, **14**: 183–216.
- Savage, Leonard J. (1954), *The Foundations of Statistics*. New York: Wiley.
- Sextus Empiricus, *Pyrrhoniae hypotyposes*, edidit H. Mutschmann. Leipzig: Teubner, 1912.
- Sextus Empiricus, *Adversus mathematicos*, edidit H. Mutschmann et J. Mau. Leipzig: Teubner, 1914/1961.
- Stegmüller, Wolfgang (1973), *Persönliche und statistische Wahrscheinlichkeit*, Zweiter Halbband (Berlin: Springer, 1973)

- Torretti, Roberto (1990), *Creative Understanding*. Chicago: University of Chicago Press.
- Teller, Paul (1973). "Conditionalization and observation", *Synthese*. **26**: 218–258.
- von Mises, Richard (1928), *Wahrscheinlichkeit, Statistik und Wahrheit*. Wien: Springer.
- von Mises, Richard (1931), *Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und theoretischen Physik*. Leipzig: Deuticke.
- von Mises, Richard (1964), *Mathematical Theory of Probability and Statistics*. New York: Academic Press.