

LA PROBABILIDAD Y SUS INTERPRETACIONES

Segundo Gutiérrez Cabria

INSTITUTO NACIONAL DE ESTADISTICA

1.— INTRODUCCION

La teoría de la probabilidad está en la base del método estadístico. La forma clásica de la ley científica, la relación funcional determinista que liga las magnitudes que en ella figuran, es un caso límite, una esquematización permitida cuando puede despreciarse la interacción universal que introduciría rigurosamente un hábito de incertidumbre y, en consecuencia, una relación sólo probable.

Ante lo probable, el hombre formado en la lógica clásica, se encuentra, no sólo ante una técnica nueva, sino ante un modo nuevo de pensar que implica un modo nuevo de juzgar, de prever y de actuar. Si logra dominar esta nueva lógica de lo probable, se encontrará ricamente recompensado, pues el pensamiento probabilista se adapta infinitamente mejor a la diversidad de criaturas y circunstancias, que constituyen el universo, que los esquemas demasiado perfectos del análisis clásico.

La probabilidad pertenece al mundo físico y no puede reducirse a una simple relación entre casos favorables y casos posibles, ni siquiera a la teoría de la medida, como podría deducirse de exposiciones que figuran en ciertos manuales.

“El cálculo de Probabilidades es seguramente una rama de las matemáticas, afirma Ullmo, pero en el sentido en que lo es también la Mecánica Racional; estas dos disciplinas no aportan su nombre, sino porque responden a ciertas propiedades del universo concreto, donde se encuentran aquí masas, fuerzas, allí condiciones de experiencias que justifican la introducción de la probabilidad”. (1).

(1) J. Ullmo : Prólogo al libro de J. Mothes, “Previsions et Décisions Statistiques dans l'Entrepise”.

¿Dónde termina el Cálculo de Probabilidades, y dónde empieza la Estadística? . Oigamos a un profesor de la Sorbona, Daniel Dugué, y observemos la terminología por él empleada: “Si se quiere fijar una frontera entre Estadística aplicada y Estadística teórica o Cálculo de Probabilidades (y tales líneas son siempre imprecisas), digamos que la Estadística aplicada es el estudio en la realidad de resultados adquiridos en lo abstracto por el modelo que constituye el Cálculo de probabilidades. Ciencia crítica de resultados experimentales, la Estadística aplicada no puede conservar la existencia de una infinidad de tales resultados, mientras que, por el contrario, el objeto del pensamiento (lógico), que es la Estadística teórica, puede perfectamente admitirlos en sus razonamientos.” (1)

El problema que aquí suscitamos será replanteado más tarde, pues está ligado a distintas formas de convergencia: el estadístico aplicado empleará primordialmente la convergencia en probabilidad, como señala el propio Dugué, mientras que el probabilista, se ceñirá de modo especial a la noción de estabilidad ergódica que es la “convergencia casi cierta”.

El Cálculo de probabilidad, por cuanto impregna todo el razonamiento inductivo estadístico, se nos presenta como un útil de difícil y delicado manejo. “A medida que la Estadística prueba su utilidad y se impone en los más diversos dominios —advirtió el Papa a los Estadísticos—, se ponen de manifiesto las dificultades de su correcta utilización y los escollos a los que se exponen los que manejan el azar, y, ante todo, lo delicado que es fijar exactamente el hecho preciso que servirá de base a las investigaciones. Aislar los diferentes factores en los que se desea examinar su papel causal, es así como se prueban las cualidades profesionales del Estadístico y la necesidad de métodos precisos” (2)

Convendrá, pues, detenemos un momento para contemplar sucintamente la lógica de la probabilidad.

2.— LOGICA PROBABILISTICA Y LOGICA ESTADISTICA.

Lo característico de la Lógica clásica, o Lógica de dos valores, consistió en el principio del “tertium non datur”, según el cual entre el ser y el no ser, el sí y el nó, la afirmación y la negación, “no hay término medio”.

Una primera negación de este principio se observa en el ejemplo siguiente. Consideremos dos sucesos S y S' . Podemos apostar sobre S condicionando nuestra apuesta al resultado de S' . Si ocurre S' se gana o se pierde, pero si no ocurre, se anula la apuesta. El suceso S condicionado a S' es, pues, una entidad lógica que puede tener tres valores: *verdad*, si S y S' son verdaderos; *falsedad*, si S' es cierto y S falso;

(1) D. Dugué: *Traité de Statistique Theorique et Appliqués*—Massons et Comp. 1958

(2) Sesión 22 del I. I. Estadística.

nulidad, si S' es falso. Basados en los sucesos condicionados, se puede, pues, elaborar una lógica trivalente.

Ya algunos filósofos habían puesto en tela de juicio el principio del "tertium non datur". De nuestro Luis Vives escribió Menéndez y Pelayo lo siguiente: "Su sentido de la probabilidad era el de los antiguos Académicos, combinado con la teoría de la "epagoge" o inducción socrática, que él amplió hasta convertirla en la inducción moderna. De esos procedimientos modestos y desdafiados, de las verosimilitudes, conjeturas e hipótesis, de aquellos razonamientos que sin aspirar a la certeza ni a la evidencia, se contentan con ser más verosímiles que los argumentos contrarios, esperaba el filósofo valenciano el futuro progreso de las ciencias, mucho más que de la esgrima escolástica o del ejercicio de la disputa. (1)

Kant, continuador, en este aspecto, de Vives, introduce y articula ya de un modo expreso y razonado, ese "término medio" en su sistema lógico y Hegel intercala, de modo claro, entre el ser y no ser el devenir o llegar a ser. El paso de la "lógica bivalente" a la "plurivalente" abrió el camino de la "lógica probabilística".

Hoy se postula por distintos autores, principalmente de la escuela subjetivista, la necesidad de confrontar la diversidad de posibilidades del mundo real con una "lógica continua".

"Desde el punto de vista lógico —escribe Bruno de Finetti—, la teoría de la probabilidad sería solamente una lógica polivalente en una escala de modalidades continua, superpuesta sobre una lógica de dos valores. Esto equivale a decir que, para cada suceso, se admiten sólo dos resultados (tres para sucesos condicionales), pero que sólo tiene un significado formal; la infinidad de modalidades intermedias, no se enfrenta, a este respecto, a la insuficiencia de la lógica de dos valores, sino que sirve únicamente para medir nuestra duda cuando aún no sabemos cuál de las dos modalidades adjetivas es correcta" (2). En sendos trabajos (publicados en "Comptes Rendus Soc.Sci.Lett. Varsovie", CIII, Vol. XXIII y XXV, 1930), J. Lukasiewicz y S. Mazurkiewicz, mantienen la tesis de una lógica de escala de modalidades continua, pero sin estar concebida como superpuesta sobre una lógica de dos valores, como quiere De Finetti. En su trabajo "La théorie des probabilités est-elle une logique généralisée?" (3), J. Hosiasson hace una crítica profunda de este aspecto debatido por los partidarios de la lógica continua.

Así como la lógica clásica es el instrumento necesario de todo razonamiento en que sólo hay dos alternativas para cada suceso, ocurrir o no ocurrir, así también la lógica de lo probable, la lógica de una escala continua de valores, es el instrumento

(1) M. Menéndez y Pelayo: Discurso de recepción en la real Academia de Ciencias Morales y Políticas, pág. 145.

(2) B. de Finetti: La Previsión, ses lois, ses sources subjectives— An. de l'Institut Poincaré, Vol. 7, 1937.

(3) Actes du Congrès Int. de Phil. Scientifique, Vol. IV, 1935.

necesario de todo razonamiento en que entra, visible u ocultamente, un grado de duda, un juicio de certeza o de posibilidad práctica, o, finalmente, una estimación de la verosimilitud de un acontecimiento cualquiera. Todo lo que no sea un hecho históricamente cierto, lo que constituya un consejo para el futuro, lo que no caiga bajo las leyes implacables de la macrofísica, constituye un juicio de probabilidad, basado en los principios del Cálculo de probabilidades. En el Cálculo de Probabilidades se funda, pues lo más importante de nuestro pensamiento, “sin él —diremos con Poincaré— la Ciencia sería imposible”.

3.— *DESARROLLO DE LA PROBABILIDAD.*

Desde muy pronto se dió cuenta la humanidad de que se encontraba frente a dos clases de fenómenos; los previsibles, como la salida y puesta del sol, y los imprevisibles, como el resultado de una batalla. La estabilidad de los primeros estimuló su estudio; los otros fueron confiados al Horóscopo.

En otra parte de este número de la Revista se estudia profusamente el desenvolvimiento histórico de la Teoría de la Probabilidad. Anotemos dos particularidades en este desarrollo: primero, su divorcio, hasta tiempos muy próximos a los actuales, del paralelo desenvolvimiento de la Estadística de los Aritméticos Políticos; segundo, su orientación, después de Laplace, hacia las aplicaciones, y esto en distintos sentidos: teoría de errores y método de los mínimos cuadrados que condujeron al descubrimiento de la curva normal; matemática actuarial como aplicación de la probabilidad a las estadísticas de mortalidad; biometría y sociometría; mecánica estadística y genética.

En cambio, los conceptos quedaron estancados. Todo el cálculo probabilístico se basó en la definición de Laplace cuya endeblez, a pesar de las mejoras introducidas por Bertrand y Poincaré, la hace insuficiente para elaborar sobre ella una teoría científica de la probabilidad.

Entonces las investigaciones se dirigieron a considerar la estabilidad de las frecuencias relativas como modelo empírico de la probabilidad matemática. Como consecuencia del Formalismo, implantado a principios de siglo por Hilbert, otra tendencia ha sido la axiomatización: la Probabilidad será una función matemática de una cierta clase, un término indefinido dentro de un sistema formal. Pero cuando queremos aplicar el modelo formal al mundo real, cuando hablamos acerca de la probabilidad de ciertos específicos sucesos o de cierta clase de sucesos, o cuando tomamos la probabilidad como norma en las decisiones de la vida, nos vemos obligados a pensar con más precisión sobre la noción misma de probabilidad. Debemos procurar encontrar alguna conexión entre esta entidad abstracta, que satisface ciertas estipulaciones matemáticas, y el contenido pragmático, el significado real de la palabra probabilidad y sus sinónimos en el lenguaje ordinario de las ciencias, de los negocios

y de los medios de comunicación escritos y hablados. Han sido dados varios tipos de conexión: la frecuencalista o empírica, la lógica y la subjetiva, son las más importantes. De cada una de estas doctrinas nos ocuparemos a continuación, así como del método axiomático.

Good (1), propone cuatro tipos de medidas para clasificar las teorías sobre la probabilidad:

- 1.— La teoría puede depender o no de un sistema axiomático.
- 2.— La probabilidad puede ser o no “objetiva”, en el sentido de que puede suponerse o no que exista, independientemente de las apreciaciones particulares de los individuos.
- 3.— La teoría puede ser frecuencalista o apoyarse en los grados de creencia.
- 4.— La probabilidad puede ser asociada o no a un número.

4.— INTERPRETACION LOGICISTA.

La interpretación logicista de la probabilidad consiste en considerarla como una relación lógica entre una proposición y un cuerpo de doctrina, entre una aserción y otra aserción o un conjunto de aserciones que presentan la evidencia. De acuerdo con esta interpretación, decir que la probabilidad de A, basados en la evidencia de B, es $2/3$, es muy parecido a decir que la proposición S está implicada en sus dos terceras partes en la S': Leonard Savage llama a esta interpretación *necesaria*, “porque, según ella, la probabilidad de A, basada en la evidencia de B, es una necesidad lógica, deducida de la estructura lógica de las proposiciones. A y B”.

El primero en formular explícitamente la interpretación lógica de la probabilidad fué John Maynard Keynes, en 1921. Es siempre posible, dice, deducir una proposición desconocida de otra conocida con un grado de certeza pasando de la ignorancia completa a la certeza absoluta. El grado de creencia racional representa la confianza que es razonable depositar en la deducción que nos lleva del conjunto de proposiciones conocidas a las desconocidas. Un grado de creencia racional debe ser interpretado siempre por comparación con un cuerpo de evidencia. En el capítulo III de su libro, Keynes, aborda el tema de si todas las probabilidades son medibles. La conclusión a que llega es, que no es posible cuantificar todas las probabilidades a causa de las múltiples dimensiones que es preciso considerar, el evaluar algunas de ellas, que no pueden ser ponderadas según un patrón común. No es posible, pues, determinar si una conclusión es más probable que otra y, por consiguiente, la probabilidad de tales conclusiones, no pueden situarse sobre una escala numérica.

(1) I.J. Good. *Probability and the Weighing of Evidence*. Hafnarn New York.

Una afirmación en términos de probabilidad es *lógicamente* cierta, si es cierta en su totalidad; de lo contrario, es *lógicamente* falsa. Las afirmaciones en términos de probabilidad son puramente formales, lo mismo que las de la aritmética. Dada una afirmación A y un cuerpo de evidencia B, hay uno y sólo un número real p, comprendido entre cero y uno, tal que puede afirmarse correctamente que la probabilidad de A relativa a B es p.

Rudolf Carnap defendió en parte esta doctrina (1). Se sitúa entre el punto de vista de la probabilidad como relación lógica objetiva entre proposiciones, mantenido por Keynes y Jeffreys, y la posición frecuentalista. Carnap dice que los desacuerdos entre estas dos escuelas proviene de falsa interpretación, pero que no se trata de conceptos distintos. Carnap llama al primer concepto probabilidad₁ y al segundo probabilidad₂. El puente entre los dos puede echarse, de algún modo, con enunciados como éste, debido al propio Carnap: "El valor de la probabilidad₁ de varias hipótesis, puede ser interpretado como la estimación de la frecuencia relativa de la certeza entre ellas" (2).

Otro representante de esta escuela es Harold Jeffreys. Su libro "Theory of Probability" (3) es una de las mayores contribuciones a la teoría de la probabilidad, interpretada como grado de creencia probable. En la pág. 62 dice, en resumen: Es sólo *casi siempre* cierto —es decir, cierto con probabilidad uno— que las caras de una moneda ideal, carente de sesgo, tienden a una frecuencia relativa de 1/2. Por lo tanto, si la frecuencia relativa de 1/2 es utilizada para definir el concepto de probabilidad, en tal definición se recurre a la probabilidad para definir la probabilidad. Por otra parte, no caemos en este círculo vicioso si afirmamos que lo más racional que se puede extraer, del hecho de lanzar una moneda perfecta, es asignar a la obtención de un resultado determinado, un grado de creencia de 1/2.

Luego veremos algunas de las críticas que se han hecho a la doctrina logicista.

5.— INTERPRETACION FRECUENCIALISTA.

La interpretación frecuentalista o empírica de la probabilidad se presenta de varias maneras. El primero en formularla fué Venn, luego von Mises y, más tarde, Reichenbach. La idea básica es identificar la probabilidad con el límite de un frecuencia relativa.

John Venn (4) presenta el caso de un hombre situado sobre la arena de la playa en un día de niebla densísima; las olas vienen amenazadoras sobre él y está

(1) R. Carnap: Logical Foundations of Probability. Univ. of Chicago Press, 1950.

(2) Ybid, pág. 72.

(3) H. Jeffreys: Theory of Probability. Clarendon Press, 3^a ed., 1961.

(4) J. Venn: Logic of Chance. Mac Millan, Londres, 1888.

persuadido de que si cae en ellas perecerá. Oye el toque de una campana de iglesia pero no sabe si proviene del mismo lado del agua o del opuesto, de suerte que al dejarse guiar por su sonido no sabe si se precipita en la corriente o, por el contrario, gana tierra firme. El suceso no admite evidentemente repetición y los casos son igualmente posibles para él. ¿No es, en tal caso, su perspectiva de perecer igual a $1/2$? “Un análisis adecuado de su estado de espíritu —dice textualmente Venn— sería una investigación psicológica más bien que lógica... Pero de acuerdo con la mejor introspección que puede hacer, yo diría que lo que realmente pasa por el espíritu, en tales casos, es algo como esto: En las situaciones y circunstancias de más vacilación, estamos acostumbrados a decidir nuestro comportamiento por la consideración de las ventajas y desventajas de cada parte, esto es, por la *frecuencia* observada o inferida con que una u otra alternativa se ha presentado. En la proporción en que éstas se hacen más equilibradas, estamos más frecuentemente equivocados en los casos individuales; esto es, se llegaría más y más cerca de lo que podría llamarse un “cara o cruz” si estamos en lo cierto o en lo erróneo. El caso en cuestión parece simplemente el caso límite en el que se ha imaginado que no hay ninguna diferencia apreciable entre las alternativas en las que basar la elección de una u otra y, en consecuencia, no hay confianza en un resultado particular... Este ejemplo sirve para presentar un punto que ha sido ya mencionado..., a saber, que todo aquello que discurre como probabilidad, es la *frecuencia estadística de un suceso*, o, si preferimos ponerlo así, la cantidad de creencia con la que cada uno de estos sucesos será considerada individualmente, pero que deja depender de esta frecuencia la conducta subsiguiente, o aquella creencia para la elección de causas” (1).

Estas ideas de Venn fueron, en realidad una reacción contra las de De Morgan, Donkin y otros contemporáneos, quienes “declaran o reconocen que nosotros somos distintamente conscientes de una cierta variación de la cantidad de nuestra creencia, y que este estado de nuestro espíritu puede ser medido y determinado con casi la misma exactitud que los sucesos externos a los que ellos se refieren,” en expresión del propio Venn (2).

El libro de Venn data de 1888. Mucho más tarde, en 1919, Richard von Mises publicó su libro “Probability, Statistics and Truth” (3). Mantiene la tesis de que el concepto de probabilidad debe ser aplicado únicamente a conjuntos de sucesos —colectivos, según su nomenclatura— que satisfacen la condición de aleatoridad y en los cuales la tendencia a converger hacia un parámetro determinado de las frecuencias relativas, es *observable*.

La definición de colectivo de von Mises es la siguiente:

(1) J.Venn: of Chance, Mac Millan, Londres, 1888.

(2) Ibid, cap III, 4.

(3) La última edición, preparada por Hilda Geiringer, data de 1965 y lleva por título: Mathematical Theory of Probability and Statistics. Acad. Press.

“ Sea $S = (a_i)$ un conjunto discreto ordenado y $K = (x_j)$ una sucesión de elementos de S . Sea G una familia de elecciones numerable. suponemos que

- 1) Para todo elemento a_i existe el límite p_i frecuencial en K ;
- 2) Para todo elemento de S se verifica $p_i = 1$;
- 3) Toda selección g perteneciente a G y aplicada a K produce una sucesión parcial de K en la cual, para todo a_i existe aún el límite frecuencial y éste es igual a p_i .

Entonces, K , o más explícitamente $K(G, S)$, es llamado *colectivo*, con respecto a G y S , y $p(a_i) = p_i$ es la probabilidad de obtener el elemento a_i en K ” (1).

Numerosas objeciones se pusieron a esta definición, como se reconoce en el libro que comentamos:” Se ha dicho que no hay pruebas de que el concepto no sea un concepto vacío, ni libre de contradicciones; que la restricción que establece nuestra noción de aleatoriedad es demasiado severa, o no suficientemente severa, etc. En los años pasados, principalmente entre 1925—40, estas cuestiones fueron ámpliamente discutidas y obtenidos resultados decisivos, probablemente desconocidos por muchos estudiantes de hoy, quienes, a menudo, sin conocer la teoría frecuencialista, del colectivo, siguen la moda corriente” (2).

El nudo Gordiano de la cuestión es que muchos autores, supuesta la equiprobabilidad de los casos posibles, confiaron en el principio de razón suficiente en vez de hacerlo en el conjunto de observaciones iniciales. Von Mises criticó agudamente esta confianza en el principio de razón suficiente, la cual se extiende a situaciones de ignorancia subjetiva, mientras que su definición, limita el concepto de probabilidad a colectivos con frecuencias relativas cuya convergencia es observable.

La posición frecuencialista sería, pues, aceptable si fuera posible establecer empíricamente una sutil separación entre conjuntos de sucesos físicos, por un lado, que aparecerían ante toda persona inteligente como los hipotéticos obtenidos por una moneda ideal y aquéllos que no obligan al decisor a usar de esta analogía. Además, había de ser un rasgo observable en todas las personas prácticamente inteligentes, el hecho de que, mientras su grados de creencia son susceptibles de ser interpretados probabilísticamente, cuando pertenecen al primer grupo, no lo son cuando pertenecen al segundo.

En realidad no se satisfacen ninguna de estas condiciones. Desde los fenómenos aleatorios, análogos a los que resultan de lanzar un dado perfecto o extraer bolas de una urna, hasta los que intervienen en la ciencia del Seguro, hay una gama cuya divisoria, en los dos grupos mencionados, es difícil de determinar. Por otra parte, gran número de personas inteligentes se sienten impulsadas a aplicar las reglas de la

(1) Op. cit., pag. 12.

(2) Ibid., pag. 40.

probabilidad a todos los sucesos aleatorios, sin distinción alguna. Las bases de la teoría frecuentista son, pues, débiles.

Adoptada la posición frecuentista, la mayor parte de las incertidumbres que afectan a la ciencia y demás dominios de aplicación de la Estadística no pueden ser medidos por probabilidades. Von Mises combatió esta afirmación en su trabajo "The Correct Use of Bayes Formula" (1), pero es evidente que un frecuentista no puede aplicar los resultados de una investigación para calcular probabilidades acerca de proposiciones inciertas sometidas a estudio. Esto supone cortarse las alas siempre que se trate de utilizar la fórmula de Bayes para el cálculo de probabilidades a posteriori, ya que se encuentra sin medios para calcular las probabilidades a priori.

"La natural inclinación que uno tiene a preguntar, ¿qué grado de convicción me autorizan a tener estos nuevos datos?, de ordinario suele ser, y ha sido, considerada — dice Savage — como una cuestión sin sentido por los frecuentistas. El frecuentista es requerido, por tanto, a tener un concepto de la evidencia, y de la reacción a la evidencia, distinto del concepto primitivo o natural que es adecuado para la aplicación del teorema de Bayes" (2).

Estas observaciones pueden dirigirse a todos los frecuentistas en general, pero la teoría frecuentista ha sido expuesta con toda una gama de matices distintos, desde Venn y von Mises hasta R.A. Fisher, Herzy Neyman, Harold Cramer y Egon Pearson.

Braithwaite, por ejemplo, en su libro "Scientific Explanation" (3), presenta un modelo conceptual en el que la probabilidad de la relación del suceso A al B es un parámetro del modelo, el cual se acepta o se rechaza de acuerdo con las observaciones llevadas a cabo sobre A y B . Estas observaciones pueden ser el resultado de un experimento, por ejemplo extracción de muestras, llevado a cabo de acuerdo con las leyes lógicas que determinan el modelo.

Toda la teoría del contraste de hipótesis de Neyman y de la escuela de estadísticos angloamericanos, no se separa mucho, en el fondo, de esta concepción de Braithwaite, aunque adopte una forma menos explícita. Para éstos, una afirmación en términos de probabilidad sólo puede ser una hipótesis estadística, siempre basada en una evidencia observada.

Para saber si una afirmación en términos de probabilidad es o no cierta, debemos realizar una investigación empírica y ésta, ordinariamente, será una investigación no limitada, de aquéllas cuyos resultados caen sólo dentro de lo probable. Esto implica negar que las afirmaciones probabilísticas pueden ser interpretadas empíricamente.

(1) Annals of Math. Statistics, n.º 13, 156, 1942,

(2) L. Savage: The Foundations of Statistics Reconsidered. Univ. of California Press, 2, 1961.

(3) Cambridge Univ. Press, 1953, cap. V, VI y VII

A pesar de todo, la interpretación frecuentista tiene aún hoy muchos adeptos. Estimamos que esto es debido principalmente a estas tres razones:

1^º La teoría frecuentista da a la probabilidad un sentido análogo al de “proporción” y esto es muy corriente en el lenguaje ordinario. Así decimos que hay una probabilidad $3/4$ de encontrar a un individuo en su despacho, si acude a él tres días sobre cuatro.

2^º Si partimos del Cálculo de Probabilidades, como rama específica de las matemáticas, y queremos contrastarlo con el mundo real y ver la consistencia de los axiomas, la controversia es menor que con otras doctrinas.

3^º Incluso cuando partimos de modelos matemáticos puros y queremos interpretarlos a la luz de la ciencia, según el modo que expusimos del método científico, la interpretación estadística es más fácil con la doctrina frecuentista.

Podemos, pues, concluir del modo siguiente: En sucesos engendrados por procesos en los que se observa una gran analogía con los que resultan de lanzar una moneda ideal, todos los individuos razonables tienden a aplicar, en sus decisiones, pesos que satisfacen las leyes de la teoría de la probabilidad, lo cual implica que tienden a atribuir tales pesos a los sucesos “simples” engendrados por tales eventos. Los valores numéricos de las probabilidades en cuestión, son valores límites de las frecuencias relativas. En esto parece que todos los autores están de acuerdo. Pero ¿cómo atribuir pesos a las decisiones aplicadas a sucesos no engendrados por procesos como los anteriores? ¿Satisfacen, en estos casos, los pesos de decisión las reglas de teoría de la probabilidad?

6.— INTERPRETACION SUBJETIVISTA.

Para los subjetivistas, la probabilidad representa, como para los logicistas, una relación entre una aserción y un cuerpo de evidencia, pero esta relación no es puramente lógica... Es lo que podríamos llamar una relación casi lógica y el número que le es asignado representa el *grado de creencia*. Este número no está unívocamente determinado: dada una aserción, puede asignársele cualquier probabilidad entre cero y uno, sobre un cuerpo de evidencia, de acuerdo con los grados de creencia de la persona que hace dicha asignación. En particular, si hay afirmación absoluta o negación absoluta de la evidencia, son aplicables los criterios de la lógica clásica de dos valores. En el resto de los casos, la probabilidad subjetiva postula una lógica polivalente o, mejor, una lógica de escala continua.

La persona sujeto debe ser *consistente* en el sentido lógico de la palabra: si hay una evidencia lógica de que ocurra el suceso S (entendiendo por suceso todo hecho o aserción), debe haber el máximo grado de confianza en S : si la evidencia entraña la negación de S , habrá que poner en S el mínimo grado de confianza. Pero

además de esto, que es común a la interpretación logicista de la probabilidad, la interpretación subjetivista da un paso más: los motivos de credibilidad de la persona, considerada como un todo, han de ser, no sólo consistentes, sino también *coherentes*. La teoría es subjetivista en el sentido de que una persona puede tener un grado de creencia cualquiera con respecto a cualquier afirmación y en cualquier tipo de evidencia, con tal de que sus otros grados de creencia tengan valores adecuados. Es sólo en este sentido que una teoría es subjetivista.

Tanto en la interpretación empírica como en la logicista, sólo puede asignarse un grado de probabilidad a un suceso de una cierta clase, en un conjunto de sucesos dado, o a una aserción relativa a un cuerpo de evidencia determinado. En la interpretación subjetivista nos encontramos con estas dos peculiaridades: 1° se permite cualquier grado de creencia en cualquier aserción o suceso; 2° hay restricciones impuestas a la distribución de grados de creencia entre conjuntos que están relacionados entre sí.

Las distribuciones de grados de creencia no son todas admisibles en la interpretación subjetivista. Supongamos que una persona analiza sus grados de creencia y los halla incoherentes. Entonces, deseosa de proceder con lógica y llevar a cabo sus deseos, procurará eliminar dicha incoherencia. El modo de hacerlo es un asunto totalmente personal. Algunas de sus opiniones, que luego han de transformarse en asignaciones de probabilidad, deben ser modificadas, pero esta modificación no ha de ser dictada por ninguna regla externa, ni en su esencia ni en el modo. Toda opinión que se considera acertada podrá ser mantenida, mediante la adecuada modificación de las demás.

Dada una aserción cualquiera S , si una persona deposita en ella un grado de creencia p , en la aserción contraria (la que niega S) depositaría un grado de confianza igual a $1-p$. Para medir los grados de creencia los subjetivistas han introducido, con Borel a la cabeza, el método de la apuesta, que consiste en determinar la cantidad máxima que se está dispuesto a apostar en favor de la verdad de una aserción. Decir, pues, que la probabilidad de S es $1/3$ quiere decir que se está dispuesto a apostar uno contra dos a favor de su certeza. Si fuera también $1/3$ la probabilidad de la negación de S habría dos apuestas: una de 1 contra 2 a favor de S y otra de 1 contra 2 a favor de la aserción o del hecho contrario. Sería perder en ambas condiciones y, en consecuencia, portarse *incoherente*. Por eso hemos eliminado de la teoría subjetivista los grados de creencia incoherentes.

La noción de “grado de creencia o de confianza” se encuentra, por primera vez, en el “*Ars Conjectandi*” de Bernouilli. Puesto que no podemos saber con certeza si un suceso va a ocurrir o no, debemos conformarnos con un “grado de confianza” en la proposición que afirma que ocurrirá. Este grado de confianza depende del conocimiento de que dispone cada individuo, varía de uno a otro y viene identificado a la probabilidad de un suceso. Augusto De Morgan, contra el que arremetió tan duramente John Venn, escribe en su “*Formal Logic*”, en 1847: “Por grado

de probabilidad entendemos, o debemos entender, grado de creencia... La probabilidad entonces se refiere a creencia e implica creencia, más o menos, y creencia no es más que otro nombre del conocimiento imperfecto" (1).

Toda la obra de Keynes está basada, como vimos, en el concepto de probabilidad como grado de creencia racional. La probabilidad, según él, expresa una relación entre un conjunto de proposiciones desconocido y otro conocido. Es siempre posible deducir las proposiciones desconocidas de las conocidas con un grado de certeza, desde la ignorancia absoluta a la certeza absoluta. Pero así como De Morgan afirma que los grados de creencia son medibles cuantitativamente, Keynes opina lo contrario.

Tras largas discusiones, llevadas a cabo por estos primeros pioneros, la probabilidad subjetiva cuenta hoy con una acreditada escuela de maestros entre los que cabe destacar, como principales forjadores, al inglés Frank Ramsey, al italiano Bruno de Finetti y al americano Leonard Savage.

Ramsey murió prematuramente, a los 26 años, en 1930. Su obra postuma "The Foundations of Mathematics and other Logical Essay" fué editada por Braithwaite y contiene, entre otros, el ensayo titulado "Truth and Probability", escrito en 1926. Presenta una teoría operacional de la elección, basada en el concepto de probabilidad subjetiva y en el principio de maximización de la utilidad esperada. Distingue entre utilidad de pagos y "tendencia o repugnancia a apostar", atribuyendo esto último a estímulos de gozo o aversión, respectivamente. Ramsey fué un precursor de von Neumann y Morgenstern. Según Emilio Borel (2), Ramsey pensó que el único modo de medir correctamente los grados de creencia de una persona, es examinar su patente conducta, por ejemplo la apuesta que está dispuesta a hacer sobre la posibilidad de que ocurra un suceso. La identificación de grados de creencia con la adopción de determinado tipo de conducta, ha sido corriente en algunos subjetivistas.

A Bruno de Finetti le hemos citado ya repetidamente. Desarrolló con todo detalle el concepto de probabilidad subjetiva, posteriormente a Ramsey y con independencia de él. Ha sido el gran divulgador de la teoría en diversas y profundas publicaciones, entre las que cabe destacar "La Prévision: ses lois logiques, ses sources subjectives" (3). Otro interesante y más reciente trabajo es el titulado "La probabilità e la statistica nei rapporti con l'induzione, secondo i diversi punti di vista"(4).

De Finetti introdujo el importante concepto de sucesos intercambiables, a los que en sus primeros trabajos llamaba equivalentes. Con esto la probabilidad subjetiva dejó de ser una curiosidad filosófica para conectarse con la inferencia esta-

(1) Formal Logic, pag. 173.

(2) E. Borel: Valeur pratique et philosophie des probabilités. Rev. philosophique, 1924.

(3) Ann. Ins. Poincaré, Vol. 7, pag. 1-68, 1937.

(4) Induzione e statistica, Rome, Istituto Mat. dell' Università, 1959.

dística clásica. Este concepto, el de consistencia y el de coherencia, forman los tres pilares de la teoría subjetivista de la probabilidad. De su estudio se han ocupado, con De Finetti, Kintchine (1), Hewit y Savage (2).

Podemos definir los sucesos intercambiables como aquellos que ocurren de modo aleatorio en una sucesión de pruebas y tales que el orden en que ocurren no afecta a las probabilidades que deseamos calcular. En "Gli eventi equivalenti e il caso degenerare" (3), De Finetti anota: "Se trata, en substancia, de lo que de ordinario se llama sucesos independientes con probabilidad constante pero desconocida". Supongamos, para aclarar este concepto, que se trata de extraer bolas de una urna cuya composición es desconocida y con reemplazamiento. Para un objetivista, los sucesos son independientes y equiprobables. En cambio, la probabilidad que un subjetivista asigna a que salga blanca en la prueba i -ésima, depende del color que haya observado en las pruebas precedentes. Se trata de una probabilidad condicional. Pues bien, estos sucesos, que son *independientes* para el objetivista, son *intercambiables* (o equivalentes) para el subjetivista.

De un modo más formal: Si asignamos el valor X_i a la prueba i y llamamos $P(X_i)$ a la probabilidad correspondiente y $P(X_i/X_{j1}, \dots, X_{jm})$ a la probabilidad condicional de X_i en el supuesto de que resulte X_j en la prueba $j, \dots, X_n$ en la prueba n , para un objetivista habrá independencia y equiprobabilidad si se verifica $P(X_i) = P(X_i/X_j, \dots, X_n)$, para todos los índices distintos ($j_1, \dots, j_n$). Para un subjetivista, habrá intercambiabilidad. Y la habrá, además, aunque no se cumpla la igualdad anterior, si se verifica una o ambas de las relaciones siguientes:

$$P(X_i) = P(X_j)$$

$$P(X_i/X_{j1}, \dots, X_{jn}) = P(X_h/X_{k1}, \dots, X_{kn}),$$

para todo conjunto de índices distintos ($j_1, \dots, j_n$) y ($k_1, \dots, k_n$), siempre que existan las probabilidades condicionantes para dichos conjuntos de índices.

Bernard O. Koopman publicó en 1940, un trabajo titulado "The Bases of Probability" (4), y casi simultáneamente, otros dos "The axioms and algebra of intuitive probability" (5) y "Intuitive probability and sequences" (6). La obra de Koopman está inspirada en la de Keynes y en la de De Finetti. Conserva la noción

(1) Sur les classes d'événements équivalents. Recueil Math. Moscou, 39, 1932.

(2) Symetric Measures on Cartesian Products. Transac. Math. Soc., 80, 1955.

(3) Istituti Italiano degli Attuari, 1952, pag. 1.

(4) Bull. of the Amer. Math. Society, 46, 1940.

(5) Ann. of Math., 41, pag. 269, 1940.

(6) Ibid.

intuitiva de grados de creencia y las probabilidades que introduce no son susceptibles de una ordenación. Para Koopman, decir que un suceso, o una aserción cualquiera S no es más probable que otro T y que T no es más probable que S , no significa que ambos tienen la misma probabilidad. Esto parece dar a entender que en nuestras opiniones falta orden. Koopman llama a las proposiciones a las que se puede asignar una probabilidad, proposiciones “contempladas” y añade que si alguna de estas proposiciones se puede verificar experimentalmente, pertenece a una subclase de proposiciones contempladas que puede denominarse “proposiciones experimentales”. La axiomática que propugna para introducir el concepto de probabilidad no cae dentro de la línea seguida por Ramsey, De Finetti y Savage.

Otro gran tratadista de la probabilidad, dentro de la nueva orientación, es, como ya hemos indicado, Leonard J. Savage. En su libro “The Foundations of Statistics” (1), publicado en 1954, dice que en vez de tomar la probabilidad como concepto básico, debemos dar varios pasos atrás y suponer que un individuo procede de acuerdo con la verificación de ciertos axiomas de modo que puede probarse que existe una función de probabilidad *personal* para cada individuo. Para Savage, todas las probabilidades son personales. En su trabajo “The Foundations of Statistics Reconsidered” (2), escribe: “La reciprocidad de ideas entre el objetivismo y el subjetivismo en Estadística, es interesante y algunas veces confuso. Ya que los frecuentistas discuten ordinariamente y creen que están en posesión de una clase de probabilidad objetiva, y ya que los personalistas afirman que la probabilidad es una cantidad subjetiva, parece natural llamar a los frecuentistas objetivistas y a los personalistas, subjetivistas. Yo mismo he sucumbido a esta tentación, pero es deplorable, pues los frecuentistas se aproximan más a lo subjetivo que nosotros los bayesianos”, Y continúa en la pag. 9: “Nosotros los bayesianos radicales, pretendemos demostrar que todo lo que tiene de atractivo la teoría frecuentista de la probabilidad, está implícito en la teoría personalista... A mi juicio, Good y yo estuvimos demasiado influidos por la tradición frecuentista —él mucho menos que yo— para hacer un remiendo perfecto”, Volveremos al pensamiento de Savage al hablar del bayesianismo.

Obligado es citar, antes de los modernos tratadistas de la Teoría de la decisión, a un autor cuyo pensamiento ha señalado un hito importante en el desarrollo de la teoría subjetivista. Nos referimos a Good, autor al que alude Savage. En 1950 publicó un pequeño tratado titulado “Probability and Weighing of Evidence”(3). De acuerdo con la clasificación que él mismo propone, y a la que ya nos hemos referido más arriba, el concepto que desarrolla de la probabilidad es axiomático, subjetivo, basado en los grados de creencia —y no en las frecuencias relativas— y no numérico. Good piensa que lo más que pueden discernir los individuos respecto a la probabilidad, es

(1) John Wiley, 1954.

(2) Univ. of California Press, 1961.

(3) Londres, Charles Griffin and Co., 1950.

que, la correspondiente a un suceso, es mayor o menor que la correspondiente a otro. Su teoría es, pues, próxima a la de Keynes y a la de Jeffreys.

Los estudios de estos distintos autores constituyen la base y lo que podríamos llamar "primera etapa de la Teoría de la decisión" y de la Estadística moderna. En esta breve reseña se han recogido las ideas más salientes de aquellos autores que han tomado parte más activa en la controversia. Los nombres de Maurice Frechet, Emile Borel, Jacob Marshak, Donald Davidson, Patrick Suppes, Frank Knight y E. L. Lehmann, deben ser añadidos a la lista anterior.

7.— POSICION SEMIPROBABILISTA.

Willian Fellner acepta la concepción básica de la escuela subjetivista, pero hace concesiones dado que "decisores razonables, cuando contemplan un movimiento en el que intervienen probabilidades, pueden desarrollar un sistema mediante el cual introducen algo de su inestable y controvertida probabilidad subjetiva como "freno" (o premio) en relación con las probabilidades del tipo Von Mises. No todos los pesos de decisión, directamente observables, de tales individuos, necesitan satisfacer las reglas específicas por las que son definidas probabilidades. Esto es importante para reconocer que una persona razonable puede atribuir una determinada pérdida de utilidad al hecho de estar expuesta a una disminución de confianza debida a controvertidos e inestables movimientos interiores." (1) Y en la pag. 4 escribe: "El pensamiento del escritor (de el mismo) acerca de problemas centrales de la Teoría de la decisión puede ser descrita con enfoque semiprobabilístico. Con esto quiero decir que en mi opinión los pesos directamente observables que individuos razonables y equilibrados atribuyen a específicas clases de perspectivas, no son necesariamente las genuinas probabilidades subjetivas de esas perspectivas, aunque estos pesos de decisión de individuos razonables y equilibrados que actúan consistentemente, guardan una comprensible relación con las probabilidades, esto es, con los pesos que satisfacen las leyes de la teoría de la probabilidad".

Según Fellner, gran número de decisores actúan de modo distinto, cuando están bajo la influencia de grados de creencia débiles, esto es, de probabilidades cuyo valor numérico es altamente inestable en su espíritu, o bajo la influencia de firmes y estables grados de creencia. Los grados de creencia altamente inestables están influenciados por lo que Fellner llama "hipótesis elusivas". Una hipótesis, según dicho autor, es una conjetura que contribuye a la determinación de los juicios de probabilidad de un individuo, relativos a sucesos. Si, en sus actos, el decisor está dispuesto a fiarse en el efecto numérico de una hipótesis acerca de sus juicios de probabilidad sobre sucesos, estas hipótesis son *directas*: las que no satisfacen esa condición son *elusivas*.

(1) Probability and Profit, Richard D. Irving. Inc Illinois, 1965, pag. 72.

8.— POSICIONES NO PROBABILISTAS

Una posición que pudiéramos llamar no probabilista, y que es objeto hoy de numerosas controversias, es la mantenida por el profesor G.L. Shackle, de la Universidad de Liverpool, como principal pionero, y algunos otros autores que comparten más o menos sus ideas. En su libro "Decisión, determinisme et temps" (1) Shackle analiza, de una forma didáctica y filosófica a la vez, la oposición aparente entre libertad de juicio y decisión y la probabilidad de previsión.

En su opinión la incertidumbre es algo subjetivo. El estado del pensamiento consistente en el rechazo del espíritu, corresponde al sentimiento "ciertamente falso". Pero pueden aparecer otros estados más o menos alejados de éste. Estos estados pueden expresarse por los valores de una variable y estos valores hay que elegirlos entre dos aspectos que esta variable es susceptible de tomar: la forma distributiva y la forma no distributiva.

"Una variable de incertidumbre distributiva consiste en tomar el número uno y hacerle corresponder la idea de "ciertamente verdadero". Si una respuesta considerada como "ciertamente verdadera" puede considerarse como un agregado de partes, donde cada parte es indispensable para expresar la totalidad y la exactitud de la respuesta, no se puede hacer corresponder a ninguna de esas partes, aisladamente, la unidad como medida y cada una de ellas debe contentarse con la fracción conveniente. Pero, dado que el agregado de todas esas partes proporciona, por definición, una respuesta subjetivamente "completa" y "ciertamente verdadera," la lógica impone que la suma de todas las fracciones sea igual a la unidad, y esta exigencia no admite excepción sea cualquiera el número de partes y el que se quiera conceder a una cualquiera de ellas. Luego el que el estado legal de una respuesta esté más o menos alejado de la situación de "ciertamente verdadera" o "ciertamente falso", es de naturaleza distinta según que se coloque uno en el caso de una variable de incertidumbre *distributiva* o en el de una variable de incertidumbre *no distributiva*"(2).

Ahora bien, si se quiere recurrir a una variable de incertidumbre distributiva, es preciso que la lista de respuestas considerada, sea a la vez *específica y completa*, es decir, que cada respuesta halle lugar dentro de un sistema o estructura y que no haya ninguna hipótesis residual, como las definidas por la cláusula "otras respuestas". La de asignar a estas hipótesis residuales, no especificadas, la parte de frecuencia relativa o probabilidad correspondiente, es una de las afirmaciones más contundentes de Shackle. "El error que denunciarnos, dice, consiste en confundir el *conocimiento substancial* suministrado por un cuadro de frecuencias correspondiente a una *experiencia divisible*, con el hecho de asignar un grado, un estado o un juicio de distancia con respecto a la certeza, a cada una de las alternativas de un cierto número de

(1) Dunad, 1967.

(2) Ibid, pag. 84.

respuestas, en el caso de una experiencia no divisible. El error es el mismo que el que se comete al no ver la diferencia que existe entre un puñado de té seco dentro de un bote y la marca exterior colocada fuera del bote indicando "calidad extra" . . . Cada evaluación de una variable de incertidumbre muestra cuán difícil es acudir a una hipótesis particular como base de una experiencia que sólo se ha realizado en la imaginación" (1).

En consecuencia la posición que una variable de incertidumbre distributiva es capaz de representar, se sitúa entre los límites de lo "ciertamente falso" y lo "ciertamente verdadero". En la práctica, añade, "el único ejemplo de variable de incertidumbre que puede hallarse es dado por la probabilidad subjetiva a priori" (1)

Para Shackle son posibles numerosas clases de variables de incertidumbre no distributivas. Al valor extremo "ciertamente verdadero", contraponen, por ejemplo, el de "perfectamente posible". Ante lo perfectamente posible no hay sorpresa, ante lo absolutamente posible, sí, y en caso de realizarse habría el grado sumo de sorpresa. Entre esos dos extremos hay grados de sorpresa a los que corresponden *grados de posibilidad*, definidos por Shackle como "juicios subjetivos que, si son emitidos por un cierto individuo con respecto a sucesos bien precisos, lo proporcionaría, en el caso, de sucesos que acaeciesen de hecho, estos diferentes grados de sorpresa. Tales juicios pueden ser expresados en términos de sorpresa potencial, y existe, convenimos en ello, una correspondencia biunívoca entre *grados de posibilidad* y *grados de sorpresa potencial*; esta correspondencia siendo de tal naturaleza que a una sorpresa potencial nula corresponde una posibilidad perfecta y a un máximo absoluto de sorpresa potencial, una imposibilidad perfecta, siempre que estos juicios provengan de un individuo determinado" (1).

El concepto de sorpresa potencial, debido a Shackle, se ha pretendido asimilarlo al de probabilidad. El mismo Shackle advierte que la expresión "perfectamente posible" representa una idea muy alejada de la que expresa la expresión ciertamente verdadero," a la que corresponde en probabilidad el valor unitario. Se hace luego esta pregunta: ¿Puede una probabilidad, inferior a la unidad, representar la posibilidad perfecta? .

La "sorpresa potencial" es una medida de la posibilidad que tiene un individuo de alcanzar una experiencia por la imaginación. Pero esta experiencia está "inhibida o aminorada por el sentimiento de duda, es decir, por la adaptación imperfecta de una situación imaginada, o de una cadena de situaciones, a lo que se sabe de la naturaleza de los hechos y del ambiente" (1). Esta es la razón por la que Shalcke prescinde de la medida directa de la posibilidad y utiliza una *medida de descrédito*, la cual "será nula cuando la posibilidad sea perfecta".

Resumiendo, podemos decir que Shalcke distingue dos clases de variables

(1) Ibid, pag 86. y 87.

de incertidumbre, las distributivas y las no distributivas. Entre las primeras sitúa la probabilidad que, cuando es objetiva y se asienta en la observación, consiste en las frecuencias relativas, pero ésta sólo se da cuando la experiencia es *divisible*, como sucede en el cálculo actuarial. En los casos de incertidumbre no distributiva, la variable propuesta es la sorpresa potencial que, contrariamente a lo que sucede con la probabilidad, permite distinguir y expresar diversos grados de posibilidad. Con respecto a una experiencia *ni divisible ni serial*, el sujeto decisor sólo puede responder apelando a una *probabilidad subjetiva*, la cual no se apoya en ningún conocimiento, ni en ninguna base matemática ni en ningún tipo de ponderación, salvo si se trata de una cuestión puramente formal.

Estas ideas de Shackle han dado lugar a numerosas discusiones en las que no podemos entrar. Entre los principales polemistas citaremos al profesor Jürg Niehans (1) el profesor R. Darfman (2), el profesor H.S. Houthakker y M. Nicholson (3). Más recientemente, y a propósito de los axiomas propuestos por Shackle para su concepto de sorpresa potencial, han tomado parte en la controversia los profesores Cartewr (4), Pen (5) y Krelle (6) y el doctor Hamblin (7).

9.— LA INTERPRETACION AXIOMATICA.

En la interpretación axiomática, la probabilidad de un suceso se considera como un número asociado a dicho suceso, el cual posee ciertas propiedades fundamentales expresables por medio de axiomas. “La definición de probabilidad —dice Cramer— se da entonces de la siguiente forma: Una probabilidad es una cantidad numérica que satisface tales y tales axiomas” (8).

Con esta interpretación la teoría de la probabilidad pasa a ser un capítulo de la matemática general, aunque con contenido propio. Corrientemente al hablar de interpretación axiomática nos referimos a la que adopta una base frecuencial, como sucede con las axiomáticas hoy más en boga, pero la axiomatización, hemos de recalcarlo, no es privativa de las probabilidades de origen empírico. La axiomatización es posible en las diversas clases de probabilidades que venimos considerando y puede llevarse a cabo de infinitos modos, tantos como hombres haya dispuestos a elaborar una estructura axiomática que reúna las condiciones precisas de consistencia e independencia.

(1) Reflections on Shackle, Probability and our Uncertainty about Uncertainty, *Metroeconomica*, XI, pag. 74.

(2) Review of *Economics and Statistics*, XXXVII n.º 3

(3) A note on M. Gould's Discussion of Potential Surprise, *Ecn. Jour.*, LXVIII, 272.

(4) *Uncertainty and Business Decisions*, Liverpool Univ. Press 1951.

(5) *Uncertainty in Economic and other Reflections*. De *Economic* 1955 n.º 11.

(6) *Econometria* XXV 1957.

(7) The Modul “Probability” *Mind* LXVIII, 270, 1959.

(8) H. Cramer: *Elementos de la teoría de probabilidades*. Aguilar S.A. Pag. 14

Reseñaremos brevemente algunas de las axiomáticas más importantes y más características.

9.1.— *Axiomática de DeFinetti*. “Consideremos —dice Bruno de Finetti— un suceso bien determinado y supongamos que no sabemos de antemano si ocurrirá o no. La duda, acerca de si va o no a ocurrir, a la que estamos sometidos, se presta a establecer comparaciones y, consecuentemente, graduaciones” (1). Si tenemos, pues, los siguientes conocimientos de los sucesos, podemos elaborar una teoría de la probabilidad totalmente rigurosa:

I.— Un suceso incierto es igualmente probable, más probable o menos probable que otro.

II.— Un suceso incierto siempre se nos presenta más probable que el suceso imposible.

III.— Cuando juzgamos un suceso S más probable que otro S' , el cual a su vez, es juzgado más probable que otro S'' , el suceso S puede aparecer más probable que el S'' .

IV.— Dadas las desigualdades,

S_1 es más probable que S_2 ,

S'_1 es más probable que S'_2 ,

puede afirmarse esta otra.

$S_1 \cup S'_1$ es más probable que $S_2 \cup S'_2$,

con tal que S_1 sea incompatible con S'_1 y S_2 con S'_2 .

Partiendo de este conjunto de axiomas, puramente cualitativo, se llega a una medida cuantitativa de la probabilidad y luego al teorema de la probabilidad total que permite la construcción de todo el Cálculo de probabilidades.

Puede darse también una definición directa, en términos numéricos, del grado de probabilidad atribuido por una persona a un suceso dado, pero el método axiomático, indicado antes, tiene la ventaja, en el sentir de De Finetti, de permitir un análisis más profundo y detallado, al empezar con nociones cualitativas hasta llegar al concepto de probabilidad y eliminar el concepto de moneda. Con todo el método directo es más claro.

9.2.— *Axiomática de Koopman*: Sean a, b, h, k proposiciones experimentales. El símbolo $(\sim)$ representa en el lenguaje de Koopman la negación, $(.)$ la conjunción o producto lógico, (C) la implicación, $(=)$ la equivalencia y $(\propto)$ la ordenación parcial. El símbolo $a/h \propto b/k$, se lee “ a en la hipótesis h no es más probable que b en la hipótesis k ”

(1) B. de Finetti: *La previsión, ses lois . . .*, cap. I.

Las proposiciones experimentales forman el contenido básico de la teoría de Koopman y, en su opinión, “cubren todos los casos clásicos de la probabilidad matemática y muchos más (1) razón por la cual su teoría se restringe a ese tipo de proposiciones. Los axiomas que consignamos a continuación, se refieren a un concepto cualitativo (ordinal) de la probabilidad, la cual es concebida como una relación lógica entre proposiciones, esto es “no más probable que” “y no menos probable que”.

Los axiomas propuestos por Koopman son los siguientes, expresados en su mismo lenguaje:

V. Axioma de verificación: Si $K \subset b$, entonces $a/h \propto b/K$.

I. Axioma de implicación: Si $a/h \propto b/K$ y $h \subset a$, entonces $X \subset b$.

R. Axioma de reflexividad: Si $H = K$ y $a h = b K$, entonces $a/h \propto b/K$.

T. Axioma de transitividad: Si $a/h \propto b/K$ y $b/K \propto a/l$, entonces $a/h \propto e/l$.

A. Axioma de antisimetría: Si $a/h \propto b/K$, entonces $a/k \not\propto b/K$.

C. Axiomas de composición: Sea $0 \neq a_1 \subset b_1 \subset e_1$ y $0 \neq a_2 \subset b_2 \subset e_2$

C_1 Si $a_1/b_1 \propto a_2/b_2$ y $b_1/c_1 \propto b_2/c_2$, entonces $a_1/e_1 \propto a_2/e_2$.

C_2 Si $a_1/b_1 \propto b_2/c_2$ y $b_1/c_1 \propto a_2/b_2$ entonces $a_1/c_1 \propto a_2/c_2$.

A estos axiomas de composición corresponden otros (en número de cuatro) de descomposición. Añade luego Koopman otros dos: uno que llama de hipótesis alternativa y otro de subdivisión. El mismo demuestra que estos son consecuencia de los demás, en un sistema completamente ordenado por la relación ($\propto$).

Tomando como base estos axiomas deduce de ellos todos los teoremas de la probabilidad.

La primera axiomática que hemos reseñado tiene un carácter subjetivo. Del mismo carácter goza la propuesta por Ramsey en “Truth and Probability”. La de Koopman es de naturaleza lógica. Veamos ahora las axiomáticas con base frecuencial. Antes es preciso hablar de la matemática precisa para esta axiomatización.

9.3.— *La matemática de la axiomática frecuencionalista.* Para elaborar una axiomática frecuencionalista “tendremos que decidimos —dice Willian Feller— por una clase de conjuntos mediante los cuales son definidas las probabilidades y una clase de funciones que sean aceptables como variables aleatorias” (2). Ambos conceptos están relacionados por lo que Loève denomina “indicador de un conjunto.” Indicador de un conjunto A es una función que toma el valor uno en todos los puntos de A y

(1) The bases of Probability, Bull of the Amer. Math Society, 46, 1940.

(2) W. Feller: An Introduction to Probability Theory and its Applications. Vol II, pag 102.

el valor cero en todos los puntos de su complemento A' . Suele representarse por I_A . En símbolos:

$$I_A(x) = 1, \text{ si } x \in A; \quad I_A(x) = 0, \text{ en los demás casos.}$$

La noción más general de función es demasiado amplia para los fines de la Estadística. Basta con el concepto de función de Euler modernizada. Si partiendo de funciones continuas podemos pasar a otra función continua como límite de una sucesión, de suerte que constituyan un conjunto tal que:

a) toda función que es continua pertenece al conjunto; b) se verifican las condiciones

$$(i) \quad f_i(x) \in B, \quad i = 1, 2, \dots, n$$

$$(ii) \quad \lim_{n \rightarrow \infty} f_n(x) = f(x)$$

$$(iii) \quad f(x) \in B,$$

entonces tenemos definida una clase de funciones que es cerrada desde el punto de vista del límite. Una tal clase —dice Feller— existe evidentemente y la clase de todas estas funciones es única. La menor de todas ellas es asimismo una familia cerrada y obviamente es la menor de tales clases. La prudencia nos exige ceñir nuestras consideraciones a esta clase mínima.

“La menor clase cerrada de funciones que contiene todas las funciones continuas, es llamada la clase de Baire y será designada por B . Las funciones de la clase B son llamadas funciones de Baire”. (1).

Es fácil ver que toda función de funciones de Baire es aún una función de Baire. Ninguna operación realizada con funciones de Baire conducirá a un resultado que no pertenezca a la clase. Esta clase es, pues, suficiente para elaborar toda la teoría y al ser mínima, no es posible elegir otra más simplificada.

Las funciones en las que estamos particularmente interesados son las llamadas *aditivas*. Sea F una función que asigna a cada intervalo I un valor finito $F(I)$. Esta función se llama *aditiva* si para cada partición del intervalo I en subintervalos disjuntos $I_1, \dots, I_n$, se verifica: $F(I) = F(I_1) + \dots + F(I_n)$. La restricción a valores finitos es excesiva si se quiere elaborar una teoría general de la probabilidad. Entonces se introduce el concepto de clase *completamente aditiva* o σ -aditiva, de funciones en que el intervalo I queda dividido en un infinito numerable de subintervalos disjuntos $I_1, I_2, \dots$, de modo que $F(I) = \sum F(I_k)$.

(1) W. Feller: *Ibid*, pag. 104.

Si hacemos $A_n = I_1 \cup I_2 \cup \dots \cup I_n$, para n suficientemente grande, A_n adquiere un valor prácticamente igual a I , es decir, $F(A_n)$ debe tender a $F(I)$, que es equivalente a lo que se postula en las clases completamente aditivas, por lo que éstas se consideran suficientes para elaborar la teoría, como queda comprobado por la experiencia.

Además, por tratarse de funciones que pretenden definir probabilidades, habrán de ser no negativas y “normales” por la condición $F(-\infty, \infty) = 1$. Esta última condición exige dividir la longitud en el espacio E_1 , el área en el E_2 , etc, en intervalos unitarios y estudiarlos por separado.

Dada una función escalar U que toma los valores $a_1, \dots, a_n$, en los intervalos $I_1, \dots, I_n$ respectivamente, definimos, con Feller, la función llamada “esperanza de u ” así:

$$E(u) = a_1 \cdot F(I_1) + \dots + a_n \cdot F(I_n) \quad (A)$$

Esta función es

(i) *Aditiva*, para formas lineales: $E(c_1 u + c_2 u_2) = c_1 E(u_1) + c_2 E(u_2)$.

(ii) *Positiva*: $u \geq 0 \quad E(u) \geq 0$

(iii) *Normada*: $E(I) = 1$

En la teoría de la integral de Riemann, para cada función u , continua en el intervalos $(0,1)$, existe una sucesión de funciones escalares u_n tales que $u_n \rightarrow u$ en $(0,1)$, Entonces es, por definición, $E(u) = \lim E(u_n)$. Resulta, por lo tanto, innecesaria la convergencia uniforme —seguimos citando a Feller— y puede usarse la misma definición para $E(u)$ siempre que $u_n \rightarrow u$ punto a punto. De este modo es posible extender $E(u)$ a todas las funciones de Baire acotadas, y la extensión es única. Cuando se trata de funciones no acotadas, son inevitables las integrales divergentes. . . Entonces no se produce ninguna perturbación porque la teoría de Lebesgue considera únicamente la integrabilidad absoluta. En términos poco precisos, podemos decir que a partir de la definición (A), para esperanzas de funciones simples, es posible definir $E(u)$ para funciones Baire generales, por aproximaciones sucesivas y paso al límite. El número $E(u)$, así definido, es la integral de Stieltjes—Lebesgue de u con respecto a $F''(1)$.

Estas dos propiedades de $E(u)$ son de especial interés:

I.— Sea U_n integrable y tal que $U_n \rightarrow u$ punto a punto. Si existe un integrable U tal que $|U_n| < U$ para todo n , es u integrable y $E(U_n) \rightarrow E(u)$.

(1) Ibid, pag. 108.

II.—Para todo integrable u y todo $\epsilon > 0$ es posible encontrar una función escalar v tal que $E(|u - v|) < \epsilon$

Vemos, pues, que tenemos a nuestra disposición dos modos distintos de expresar el concepto de la función esperanza. Supuesto que el indicador I_A es una función de Baire, podemos escribir:

$$E(I_A u) = \int_A u(x) dF(x).$$

La notación del primer miembro es preferible, según Feller, en aquellos casos en que la función de intervalo F permanece fija y se acentúa la dependencia con respecto a la función de punto u . La “notación integral” es preferible cuando sucede lo contrario.

Los conjuntos utilizados en las axiomáticas frecuentistas son los de Borel.

Un conjunto A en el espacio E_r cuyo indicador I_A es una función de Baire, se llama conjunto de Borel. Los intervalos son, pues, conjuntos de Borel.

Una σ — álgebra es una familia α de subconjuntos de un conjunto E dado, tales que

(i) $S_i A \in \alpha$ también $A' = E - A \in \alpha$

(ii) $S_i \{A_n\}$ es una sucesión numerable de conjuntos que $\in \alpha$ también $\cup A_n$ y $\cap A_n$, pertenecen a α .

La mínima σ — álgebra que contiene todos los conjuntos abiertos de un espacio topológico E , es llamada *álgebra de Borel*. Para $E = E_r$, ésta coincide con con la familia de Borel.

Luego veremos que pueden obtenerse aproximaciones de los conjuntos de Borel por medio de intervalos.

9.4.— *Axiomática de Kolmogoroff — Cramér — Frechet*. La axiomática propuesta por Kolmogoroff (1) es la que suelen adoptar la mayor parte de autores modernos, con modificaciones más o menos substanciales.

Es como sigue:

I.— A cada suceso $A \in \alpha$ le corresponde un número real $P(A)$, su probabilidad, tal que $0 \leq P(A)$.

II.—Si $A_1, A_2, \dots$, son incompatibles

(1) A. N. Kolmogoroff: *Grandbegriffe der Wahrscheinlichkeitsrechnung*. Traducción al Inglés por N. Morrison, 1950

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_1^{\infty} P(A_i)$$

III. — $P(E) = 1$

Llamaremos al espacio E , con la σ -álgebra $\mathcal{a}$ de sucesos sobre los que se ha definido la función de probabilidad P , *espacio de probabilidad o espacio probabilístico y lo designaremos por $(E, \mathcal{a}, P)$* .

Los axiomas I y III coinciden con el 1º y 2º de Kolmogoroff. El II sintetiza el 3º y 4º de Kolmogoroff, Kolmogoroff añade un 5º axioma que llama de continuidad y que dice así:

Consideremos una sucesión decreciente de conjuntos

$$A_1 \supset A_2 \supset A_3 \supset \dots \supset A_n \supset \dots,$$

cuya intersección se aproxima al conjunto vacío cuando $n \rightarrow \infty$ entonces

$$\lim_{n \rightarrow \infty} P(A_n) = 0$$

Puede verse fácilmente que este axioma es equivalente al 4º de Kolmogoroff y al II anterior, es decir, que una familia de conjuntos completamente aditiva posee la anterior continuidad y recíprocamente.

Cramér propone tres axiomas. “Los axiomas 1 y 2 son equivalentes a los establecidos por Kolmogoroff —según su propia confesión—. Sin embargo —añade— los axiomas de Kolmogoroff son aplicables a variables aleatorias definidas en espacios de carácter más general que los aquí considerados” (1).

El axioma 3 dice: “Si $x_2, \dots, x_n$ son variables aleatorias, toda variable aleatoria combinada de $(x_2, \dots, x_n)$ es así mismo una variable aleatoria.

El punto de partida de todas estas axiomáticas es un conjunto E llamado *espacio muestral de sucesos elementales*. Los subconjuntos de E son llamados por algunos autores “sucesos”. En el caso de un espacio muestral discreto, el conjunto $\mathcal{a}$ de todos los subconjuntos de E , forman una familia de Borel que contiene E . En el caso más general, $\mathcal{a}$ es alguna familia de subconjuntos de E y salvo que $\mathcal{a}$ sea la mínima familia de Borel, que representamos por $\mathcal{a}$ ésta se obtiene, a partir de $\mathcal{a}$, por medio del “teorema de extensión” de Banach.

La idea general está perfectamente trazada en estas palabras de Cramér: “Siguiendo a Kolmogoroff, tomamos como punto de partida la observación de que la probabilidad $P(A)$ puede ser considerada como una función aditiva de conjunto, del conjunto A . Nos contentamos en efecto, con postular, de modo especial, la exis-

(1) S. Ríos: *Métodos Matemáticos de Estadística*. pág. 178.

tencia de una función de este tipo, definida por una *cierta familia* de conjuntos A en el espacio en el que nuestra variable puntual X está restringida y tal que $P(A)$ designa la probabilidad de que $X \in A$ " (1).

Feller, después de definir "una medida de probabilidad P sobre una σ -álgebra α en un espacio E " mediante unas condiciones análogas a las anteriores, añade: "Naturalmente esta definición provee únicamente el fundamento para una estructura general, y en casos individuales es necesario elegir una σ -álgebra adecuada y construir sobre ella la medida de probabilidad". (2).

Por su parte Halmos escribe: "La probabilidad es una rama de las matemáticas. La probabilidad numérica es una función de medida, esto es, una función P aditiva, finita, no negativa y numerable, de elementos en una σ -álgebra de Boole" (3).

9.5.— *Controversia Von Mises – Cramér.* Ahora —dice Von Mises— una tal descripción de los instrumentos matemáticos utilizados en el cálculo de la probabilidad nos parece sólo parte de la historia. Distribuciones de masa, distribuciones de densidad y carga eléctrica, son asimismo funciones aditivas de conjunto. Si no hay nada específico de la probabilidad. ¿por qué definimos la "independencia" para distribuciones de probabilidad y no para distribuciones de masa? ¿Por qué consideramos variables aleatorias, convoluciones, cadenas y otros conceptos y problemas, específicos del Cálculo de probabilidades? (4).

Von Mises niega que la probabilidad sea una ciencia matemática, del mismo modo que "la hidrodinámica no es una rama de la teoría de ecuaciones diferenciales en derivadas parciales." "Nuestra meta —dice— al presentar la probabilidad como ciencia matemática es incorporar, a las hipótesis básicas, las relaciones básicas idealizadas con la realidad, como se hace en Mecánica y otras ciencias".

Estas hipótesis básicas son las siguientes:

$$\text{I.} \quad \lim_{n \rightarrow \infty} \frac{n_i}{n} = P_i, \quad i = 1, 2, \dots, K,$$

donde n_i es el número de resultados, correspondientes a un valor dado del espacio muestral, entre los n primeros de la sucesión, con la hipótesis adicional:

$$\sum_{i=1}^{\infty} P_i = 1$$

(1) H. Cramér: *Random Variables and Probability Distributions*, Cambridge, 1931.

(2) *An Introd. To Prob. Theory*, pag 112.

(3) *The foundations of probability*. Amer. Math. Monthly, 51, pag. 493.

(4) *Math. Theory of Prob.* pag. 44.

II.—“En la teoría de la probabilidad trataremos (no exclusivamente) con sucesiones particulares que satisfacen la hipótesis I y para las cuales puede suponerse que el límite frecuencial de un resultado a_i no depende del punto de elección. Si esto se verifica, el límite será llamado también probabilidad, o, más completamente la probabilidad $P(a_i)$ de hallar el resultado a_i en la sucesión bajo consideración”.

Cramer ha criticado duramente estas hipótesis que, según él, “contienen una mezcla de elementos empíricos y teóricos lo que, usualmente se evita en las modernas teorías axiomáticas. Ello sería comparable, p. ejem., a definir un punto geométrico, como límite de una mancha de tiza de dimensiones indefinidamente decrecientes, lo cual no se hace usualmente en la geometría axiomática”. (1)

Von Mises, a su vez, replica del siguiente modo: “La mezcla de elementos empíricos y teóricos es, en nuestra opinión inevitable en una ciencia matemática. Cuando en la teoría de la elasticidad introducimos los conceptos de tensión y fuerza, no podemos contentarnos con exponer que son tensores simétricos de segundo orden. Tenemos que poner en las hipótesis básicas la mecánica del continuo. La ley de Hooke, etc. cada una de ellas contiene una mezcla de elementos teóricos y empíricos. La teoría de la elasticidad no es análisis de tensores. Nuestra definición es menos comparable a una mancha de tiza que la definición de velocidad como límite del cociente incremental del espacio por el tiempo o a la definición de la densidad específica como límite de la masa partido por el volumen. Todas estas definiciones tienen también algo de la “mancha de tiza”: la transición de la observación a los conceptos teóricos, no puede ser completamente matematizada. No es una conclusión lógica, sino más bien una elección, con lo que uno cree “alzarse”, frente a nuevas observaciones” (2).

Para poner un poco de luz en la discusión, importa deslindar bien los terrenos. En la aproximación frecuencial, “siempre que se utiliza el término probabilidad, se refiere al límite de una frecuencia, y esta frecuencia debe ser aproximadamente “verificada” al menos conceptualmente”. (2). Este es el punto de vista de Von Mises, Tornier y Wald. En la definición de la probabilidad como medida de un conjunto no hay ninguna incorporación a la teoría de las frecuencias relativas, sino que éstas aparecen como mera interpretación de la misma y sin que sea necesario apelar a la verificabilidad, tal como propugnan Kolmogoroff, Cramer y Frechet como más representativos.

Samuel Wilks escribe: “En la actual asignación de probabilidades a sucesos, es uno habitualmente guiado por una hipótesis basada en lo que se cree “abocará” la frecuencia relativa de los sucesos en una serie de pruebas con base experimental. En la formulación de Kolmogoroff, el punto fundamental está en que la teoría ma-

(1) Met. Mat. de Estadística, pág. 174.

(2) Math Theory of Prob., pág. 45

temática empieza después de la asignación de probabilidades" (1)

Esta observación de Wilks, estimamos que es de particular importancia. La axiomatización supone abstracción o conceptualización y esto implica una sustitución de objetos empíricos por ideales. La teoría empieza después de realizar esta sustitución. Lo que no vemos muy claro es el paso al límite de unos y otros, como quieren los frecuentistas puros.

Von Mises se escuda en estas dos afirmaciones. a) "Mantenemos el simple hecho que si la parte matemáticamente deductiva de la teoría se desarrolla a partir de los axiomas especificando meramente la naturaleza de las variables básicas y funciones implicadas, entonces las relaciones con la experiencia no pueden ser deducidas matemáticamente" Y añade: "No se puede sacar lo que no se puso". b) "En las teorías de probabilidad como medida, en las que todas las deducciones están basadas en la definición de medida, la relación frecuencial es a menudo introducida luego, como una más o menos vaga idea o reflexión tardía" (2).

Pone luego el siguiente ejemplo: Uno de los lemas de Borel-Cantelli dice que si una misma prueba con probabilidad $1/2$ se repite indefinidamente la probabilidad P de que el suceso ocurra un número finito de veces tiende a cero". Pues bien, dice Von Mises, "en la teoría frecuentista, la probabilidad de que un suceso ocurra indefinidamente no tiene lugar. Tal probabilidad no admite verificación, ni aun conceptualmente". Pero, en cambio —añade— la anterior probabilidad P es la medida de un *bien definido* conjunto y, para Kolmogoroff, este conjunto, como todo conjunto de Borel, admite una probabilidad".

Otro punto vulnerable de la axiomática de Cramér, por donde le ataca nuevamente Von Mises, es el siguiente: "De los trabajos de Tornier, Eald y Cope'and, se deduce que la probabilidad, en sentido comparable, no puede asignarse a todos los conjuntos medibles y, ni siquiera, a todos los conjuntos de Borel. El postulado enunciado por Cramér de que" toda probabilidad asignada a un suceso determinado debe, en principio, estar sujeta a comprobación "está en contradicción con los axiomas que asignan probabilidades a todos los conjuntos de Borel" (3).

9.6. — *Axiomática de Tornier*. El concepto básico de Tornier (4) es lo que llama *conjunto o espacio de sucesiones* (folgenmenge), que designa por B , y define como "el conjunto de todas las infinitas sucesiones que, lógicamente, podrían obtenerse por la regla experimental dada". Cada sucesión particular infinita es un punto de B . Todos los subconjuntos de B a los que se asignan valores, junto con estos valores, forman la familia B . Un subconjunto E de B que contiene todas las sucesiones de B y sólo aquellas que tienen la misma "longitud" inicial n , es llamado

(1) *Mathematical Statistics*, pag. 11.

(2) *Ibid*, pag 46.

(3) *Ibid*, pag 47.

(4) E. Tornier: *Wahrscheinlich Keitsrechnung und allgemeine Integrations theorie*. Ann. Arbor, 1944.

conjunto básico de orden n . B contiene, naturalmente, muchos conjuntos básicos. Dos conjuntos básicos son disjuntos o, por el contrario, uno está totalmente contenido en el otro.

Estos conceptos son expresados formalmente, por Tornier, en los siguientes axiomas:

I.— Los Conjuntos de B forman una familia.

II.— Todos los conjuntos básicos de B , incluidos B y el conjunto vacío ϕ , pertenecen a B .

III.— a) A todo conjunto A de B asignamos un número entre cero y uno, representado por $J(A)$ y llamado su contenido. En particular, $J(\phi) = 0, J(B) = 1$.

b) Si $A = A_1 \cup A_2 \cup \dots$, donde los A_i son disjuntos y $A \in B, A_i \in B$, es

$$J(A) = \sum_1^n J(A_n)$$

IV.— Sea M un conjunto de B . Sean A_n y $B_n, n = 1, 2, \dots$, sucesiones de conjuntos de B , tales que para todo n ,

$$B_n \supset M \supset A_n,$$

y tales que para todo ϵ positivo y suficientemente pequeño, puede hallarse un n tal que $J(B_n) - J(A_n) < \epsilon$,

entonces, $M \in B$ y $J(M) = \lim_{n \rightarrow \infty} J(A_n) = \lim_{n \rightarrow \infty} J(B_n)$

V.— B es la familia mínima que satisface los 4 primeros axiomas.

A diferencia de los axiomas de Kolmogoroff, los de Tornier señalan el material básico (axioma II) de construcción. Además especifican la condición mínima de amplitud (axioma V) y la de completitud (axioma IV). Esta teoría se inspira en la de conjuntos de Jordán.

9.7.— *Axiomática de Renyi*. La diferencia esencial entre la axiomática de Renyi (1) y las del tipo Kolmogoroff, radica en que la de Renyi toma como base las probabilidades condicionadas, en vez de las absolutas. La probabilidad de un suceso A está siempre condicionada a la realización de otro B . Así al lanzar una moneda la probabilidad $1/2$ de obtener cara está condicionada a la simetría y homogeneidad de la moneda. Las pruebas sucesivas suponen que no habrá desgastes en la misma, ni otras alteraciones que cambien de naturaleza el experimento. La supuesta constancia de estas circunstancias (sucesos) es causa de que se les ignore en los

(1) A. Renyi: *Wahrscheinlich Keitsrechnung*. Berlin, 1962.

cálculos corrientes, pero esto no presupone negar su existencia, del mismo modo que se considera fijo el peso de un cuerpo en cualquier punto de una región aunque esto no sea cierto.

Para simplificar la escritura representaremos por $A \cap B$ la intersección de los sucesos A y B , y por $\sum_n A_n$ la unión de la sucesión $\{A_n\}$.

Dados dos sucesos A y B , la probabilidad de A , dado B , es un número $P(A/B)$ que satisface los siguientes axiomas:

$$\text{I.} - 0 \leq P(A/B) \leq 1 \quad \text{y} \quad P(A/A) = 1.$$

II. - Si los sucesos $\{A_n\}$ son excluyentes dado B , se verifica,

$$P\left(\sum_n A_n/B\right) = \sum_n P(A_n/B).$$

$$\text{III.} - P(C/AB) \cdot P(A/B) = P(AC/B).$$

Los dos primeros axiomas se enuncian con un sólo suceso condicionante y el tercero, con dos.

La justificación del axioma I es evidente ya que el límite frecuencial está siempre entre cero y uno, y si nos ceñimos al suceso de obtener cara, en el lanzamiento de una moneda, de modo que el suceso condicionante sea "obtener cara", el resultado será necesariamente cara y el límite tomará el valor uno.

Para explicar el segundo axioma, supongamos que los sucesos $A_1, A_2, \dots, A_n$, relativos al lanzamiento de una moneda, son tales que, dado B se excluyen dos a dos. Entonces es fácil ver que la frecuencia relativa correspondiente a la suma de sucesos es igual a las sumas parciales de las frecuencias de cada uno. Es conveniente suponer que esta aditividad se cumple tanto en el caso de un número finito de sucesos, como en el de un infinito numerable. La aditividad en el segundo caso "no viene impuesta —dice el profesor Ríos— con la misma fuerza que en el primero, pero como tampoco la intuición es contraria, se introduce por resultar más sencillo el desarrollo de la teoría" (1).

Para interpretar el axioma III, consideremos, de nuevo, n lanzamientos de una moneda, en condiciones análogas, esto es, n resultados del suceso B . Sea n el número de caras, es decir, que A ocurre m veces, y de éstas m veces, ocurre r veces C . Entonces, r es el número de veces que ocurre $A \cap C$.

Ahora bien, un resultado trivial es el siguiente:

$$\frac{r}{m} \cdot \frac{m}{n} = \frac{r}{n}$$

(1) S. Ríos: Métodos estadísticos, pág. 38

Pero cuando $n \rightarrow \infty$, $m \rightarrow \infty$, siempre que $P(A/B) \neq 0$ y entonces tenemos:

$$\lim_{m \rightarrow \infty} \frac{r}{m} = P(C/AB); \quad \lim_{n \rightarrow \infty} \frac{m}{n} = P(A/B); \quad \lim_{n \rightarrow \infty} \frac{r}{n} = P(AC/B).$$

Luego el tercer axioma resulta de obtener unos límites razonables en una expresión trivial.

Sucede a menudo que un suceso B ocurre como parte o todo del suceso condicionante. Si el suceso B es el espacio total E , hay economía en la notación escribiendo $P(A)$ en vez de $P(A/E)$ y $P(A/C)$ por $P(A/CE)$. Omitir el suceso condicionante al expresar la probabilidad de un suceso elemental, equivale a considerar un efecto al margen de su causa. Es lo que se hace con las llamadas "probabilidades absolutas", base de toda la inferencia estadística clásica, pero la inferencia Bayesiana toma como base probabilidades condicionadas.

Es fácil observar que los sucesos observados A deberán pertenecer a una σ - álgebra de Borel y los conjuntos condicionantes B , a una subclase de la familia de conjuntos de Borel. Una discusión completa de estos axiomas puede verse en la obra citada de Renyi.

9.8. - Extensión de esta axiomática a los grados de creencia. La probabilidad matemática considerada anteriormente, era sugerida por las propiedades de los sucesos y las frecuencias relativas de los mismos. De los sucesos podemos pasar a proposiciones, que son colecciones de sucesos por cuanto enuncian la ley de su obtención, y de las probabilidades de los sucesos, a las de las proposiciones. Estas probabilidades, es difícil que admitan una interpretación frecuencial, ya que resulta muy artificial imaginar repeticiones en cada una de las cuales una proposición es cierta o falsa. Si A es una proposición, $P(A/B)$ representa la creencia o grado de convicción de uno, de que A sea cierta, dado B .

En Estadística se usa la palabra hipótesis en vez de proposición.

Del mismo modo que la longitud y la masa tienen propiedades matemáticas comunes así los límites frecuenciales y los grados de creencia poseen caracteres análogos que permiten identificarlos con probabilidades. Los conjuntos A de ordinario, representan sucesos, cuando se trata de frecuencias relativas, y proposiciones o hipótesis, cuando se refieren a grados de creencia. Las hipótesis, como hemos indicado, son al fin y al cabo, colecciones de sucesos, y no es dar un gran salto, el paso de unos a otros y considerar ciertos conjuntos del espacio muestral como hipótesis.

La justificación de los axiomas anteriores aplicados a las proposiciones o hipótesis se basan en el método de la "apuesta" al que ya nos hemos referido. No volveremos sobre las ideas aquí expuestas y, principalmente a la interpretación dada por Savage. En particular, ya se discutió la duda presentada por Keynes y otros autores sobre la posibilidad de medir numéricamente los grados de creencia.

Como anotación final, a las diversas axiomáticas reseñadas, queremos señalar que los resultados a los que conducen todas en las aplicaciones, son los mismos.

Hemos pretendido en este artículo poner de manifiesto las modernas corrientes sobre el concepto de probabilidad. Este concepto es único, pero son distintas las interpretaciones del paso del mundo real al modelo matemático.

