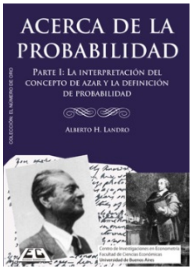
Acerca del problema de Bernoulli y la determinación del verdadero valor de una probabilidad

De los mismos autores

Acerca de la probabilidad

Parte I: la interpretación del concepto de azar y la definición de probabilidad



Elementos de econometría de los fenómenos dinámicos

Los procesos estocásticos lineales unidimensionales

Acerca del problema de Bernoulli y la determinación del verdadero valor de una probabilidad

Alberto H. Landro **Mirta L. González**

EDICIONES COOPERATIVAS

Landro, Alberto
Acerca del problema de Bernoulli y la determinación del verdadero valor de una probabilidad / Alberto Landro ; Mirta L. González. - 1a ed . - Ciudad Autónoma de Buenos Aires : Ediciones Cooperativas, 2016.
246 p. ; 23 x 16 cm.

ISBN 978-987-652-180-2

1. Probabilidades. I. González, Mirta L. II. Título
CDD 519.2

1º edición, Noviembre 2016

Tucumán 3227 (1189)
Buenos Aires – Argentina
☎ (54 011) 3528-0466 / (15) 4937 6915
http://www.edicionescoop.org.ar
✉ info@edicionescoop.org.ar

Colección: *El número de oro*
Director: *Act. Alberto H. Landro*

Imagen de tapa: *El jugador hindú* – Víctor Chab, 2005

Hecho el depósito que establece la ley 11.723

Impreso y encuadernado por: Imprenta Dorrego. Dorrego 1102, C.A.B.A.
1ª. ed. se terminó de imprimir en Noviembre 2016.

Editorial asociada a:

IMPRESO EN ARGENTINA – PRINTED IN ARGENTINA

CÁMARA ARGENTINA DEL LIBRO

a la Pepa, el Pepo y el Pipo

Invocación previa

"*Sè con tu autor (ò Libro) agradecido;*
Pues para darte vida en la memoria
En tu cuerpo su alma ha introducido"

Gerónimo Monforte y Vera
(Introducción a las "*Reglas generales de acompañar en órgano,*
clavicordio y harpa, con solo saber cantar la parte,
ò un baxo en canto figurado"
de José de Torres y Martínez Bravo (1702))

Prólogo

> "*Bisogna che il nostro dir sia inteso:*
> *dirlo chiarozzo chiarozzo,*
> *acciò che chi ode ne vada*
> *contento e illuminato*"
> (San Bernardino da Siena)

La contribución fundamental de Jakob Bernoulli a la teoría del azar consistió en proporcionar una definición estricta de probabilidad, basada en el supuesto que se podía "*aprender*" a partir de la experiencia y demostrar que este aprendizaje era cuantificable a través de un proceso de inversión que vincula a las probabilidades "a priori" –definidas a partir de un razonamiento que va de las causas a los efectos- con las probabilidades "a posteriori" –definidas a partir de un razonamiento que va de los efectos a las causas, de las frecuencias observadas de un resultado de naturaleza eventual al supuesto "*verdadero valor*" de la probabilidad de su ocurrencia. La consecuencia de esta revolucionaria propuesta fue la primera ley de los grandes números.

Este teorema fue objeto de innumerables extensiones, ampliaciones y generalizaciones que dieron origen a una sucesión aún no concluida de intentos de solución del problema de Bernoulli discutidos fervientemente durante los últimos 300 años. Una trayectoria que comenzó con las propuestas clásicas del sobrino Niklaus, de Moivre, Laplace y Poisson, continuó con la convergencia en-medida de Borel, la axiomática frecuencista de von Mises, la interpretación subjetivista de deFinetti, su teorema de representación y su vinculación con la teoría ergódica, la aproximación axiomática de Kolmogorov y su reivindicación del frecuencismo a partir de una deducción empírica de dichos axiomas y derivó en una aproximación metodológica basada en la teoría de la martingala.

La exposición realizada aquí no pretende ser una antología detallada y exhaustiva de estas propuestas de identificación de una probabilidad, sino una interpretación dirigida a procurar una revisión racional de los supuestos metafísicos en los que se basaron las numerosas propuestas a partir de las cuales se desarrolló la teoría. Su espíritu está inspirado en la recomendación que el Profesor de Finetti incluyó en su "*Filosofiadellaprobabilità*", en la cual expresa la necesidad de resolver "*...los problemas conceptuales y las controversias existentes en temas de probabilidad (...) de modo que el desarrollo de los razonamientos no se reduzca a un mero juego formal sobre expresiones matemáticas o a enunciados pseudo-filosóficos pretendidamente prácticos, pero vacíos y simplistas*".

El resultado fue este intento de tratamiento unitario sobre la teoría de la probabilidad, más desde un punto de vista conceptual que matemático, tendiente a lograr un balance adecuado entre un análisis que pretende ser riguroso y un detalle innecesario y a cuestionar toda axiomática dirigida más a redimir las conciencias matemáticas que a atender los alcances de su aplicabilidad.

Como expresión final de este mensaje introductorio, sólo queda darle la bienvenida al lector a este universo aleatorio dominado por el principio de la incertidumbre. Una incertidumbre que procuramos interpretar como libre de la servidumbre dogmática del determinismo.

AHL - MLG

PD: Cabe señalar que la aparente sobreabundancia de notas de pie de página se justifica por la necesidad de incluir resultados considerados de interés para ubicar los conceptos en su contexto y de evitar molestas interrupciones en la lectura del texto.

In perpetuam rei memoriam de
Fausto Ismael Toranzos y José Barral Souto,
Inolvidables maestros.

Índice

1.- Una introducción al "*Ars conjectandi*"

En 1713 -en forma póstuma-, como recopilación de sus trabajos inéditos sobre teoría de la probabilidad de los 20 años anteriores -y por obra de su sobrino Nicklaus-, se publicó el "*Ars conjectandi*" de Jakob Bernoulli[1]. Una obra que, por su tratamiento absolutamente novedoso de la evidencia empírica, modificó los fundamentos de la probabilidad en tal forma que debería ser considerada como el verdadero punto de partida de la teoría de la probabilidad y la culminación del proceso de formación del concepto de probabilidad.

Un proceso que tuvo su origen en el siglo XVII, en la obra de un grupo de pensadores (entre los que cabe mencionar a Joseph Glanvill, John Wilkins, Marin Mersenne, Pierre Gassendi, Hugo Grotius, John Tillotson, Robert Boyle y John Locke) quienes, a partir del deterioro del criterio de *"creencia"* en la religión, la filosofía y la ciencia, provocado fundamentalmente por las polémicas de la Reforma y la Contra-Reforma, generaron el movimiento filosófico conocido como "*escepticismo constructivo"*. Esta corriente coincidió con los escépticos en que la "*certeza absoluta"* se encontraba fuera del alcance de cualquier "*observador"*, pero admitió, sin embargo, que este observador poseía la habilidad de diseñar su comportamiento a un nivel inferior de "*certeza parcial"*. Este planteo se convirtió en el argumento habitual de la apologética de la teología natural, especialmente en Gran Bretaña[2] y tuvo su culminación en "*Analogy of religion, natural and*

[1] Si bien muchos de los desarrollos ya se encontraban en sus "*Meditationes"* (1684-1690) -según se puede concluir de su correspondencia con Leibniz-, J. Bernoulli dedicó a esta obra los dos últimos años de su vida, es decir que estaba fundamentalmente concluida en 1705. Después de su fallecimiento (a la edad de 50 años) sólo fue sometida a ligeras modificaciones pero, debido a disputas familiares, no fue publicada hasta 1713.

[2] Boyle ("*Some considerations about the reconcileableness of reason and religión"* (1675)) reconoció tres grados de certeza: "*metafísica*" (absoluta), "*física*" y "*moral*" (parciales). Consideró a la primera como derivada de axiomas que "*...nunca pueden ser sino ciertos"*, a la "*certeza física"* como "*...deducida de principios físicos"* que podían ser modificados o eliminados por Dios, y a la "*certidumbre moral"* (estado intermedio entre la certeza y la duda) como aquélla en la que "*...la conclusión se funda en alguna concurrencia de probabilidades"* (p. 182). Entendiendo por "*concurrencia de probabilidades"* no probabilidades matemáticas, sino la "*...convicción adicional que proporciona la evidencia convergente"* (Daston (1980, p. 63)).

revealed, to the constitution of nature" (1736) de Butler[3]. La racionalidad pragmática de esta "*certeza parcial"* -ubicada entre la "*certeza dogmática"* de los escolásticos y los cartesianos, y la "*duda absoluta"* de los escépticos- le permitía al observador definir "*grados de certeza"* cuya cuantificación dio origen a las probabilidades.

Este intento de cuantificación, que estuvo a cargo de matemáticos como Pascal y Fermat, se estructuró a partir de: **i)** la posibilidad de la justificación de analogías entre el comportamiento de los fenómenos y los modelos matemáticos disponibles, y **ii)** la hipótesis metafísica de que dicho comportamiento era constante y ordenado. Los filósofos racionalistas del siglo XVII[4] procuraron, de la misma forma que los teoremas eran deducidos de los axiomas y definiciones de la geometría, deducir el comportamiento de los fenómenos naturales -mediante un método de síntesis- a partir de "*leyes"* que se suponía constituían la estructura de la materia. Pero, en muchos casos la limitación de los sentidos, la falibilidad de la razón y la infinita complejidad de la naturaleza los obligaron a emplear un esquema inverso (analítico) de razonamiento empirista: desde los "*efectos evidentes"* hacia las "*causas ocultas"*. Ahora bien, dado que este método empírico, basado en un conjunto limitado de observaciones, no permitía generalizaciones a cualquier otro caso, las previsiones que se obtuvieron resultaron inevitablemente inciertas. Fueron en principio, los matemáticos del siglo XVIII quienes, mediante la cuantificación de las probabilidades de las hipótesis, intentaron dar mayor precisión a estos métodos de predicción y ampliar los alcances de la teoría de la probabilidad a la inferencia empírica[5].

---

[3] En esta obra, en la que Butler incorporó además un tipo de razonamiento inductivo y analógico concebido en términos frecuencistas, aparece por primera vez el "ejemplo del amanecer". En el que la asignación de una probabilidad a la ocurrencia del fenómeno "que mañana amanezca" sólo es posible a partir de un gran número de observaciones. Este ejemplo, utilizado luego por Richard Price en sus comentarios sobre el teorema de Bayes, se convirtió en un lugar común en la literatura sobre la inversión de la probabilidad (ver Sec. 10).

[4] Por ejemplo, Descartes ("*Principia philosophia"* (1644)).

[5] Para una cuidadosa reseña sobre la probabilidad previa a Pascal y un análisis detallado de su obra probabilística y su correspondencia con Fermat, vert Franklin (2001), Shea (2003), Hammond (2003).

A comienzos de este siglo también los filósofos empiristas trasladaron su interés de los aspectos cualitativos a los aspectos cuantitativos de la experiencia. Todo este proceso produjo como resultado una "*teoría de la inferencia inductiva*" aplicable a series de eventos que pudieran ser considerados naturalmente idénticos, para los cuales la agregación de la experiencia adquirida a partir de su observación repetida, se transformaba en formas de "*expectativa*" acerca de su comportamiento futuro (expectativa entendida como la expresión matemática del concepto de racionalidad)[6].

En este sentido, la contribución fundamental de J. Bernoulli -incluida en la "*Pars quarta, Fradens usum et applicationem procedentis doctrinæ in civilibus, moralibus et œconomicis*" del "*Ars conjectandi*"- consistió en: **i)** proporcionar la primera definición estricta de probabilidad ("*probabilitas*") en un contexto matemático como "*...grado de certeza (gradus certitudinis), que difiere de ésta de la misma forma que una parte con respecto al todo*"[7] y, en consecuencia, a partir del tratamiento cualitativo formal de la idea planteada por los mencionados apologistas ingleses de la teología natural del siglo XVII y por Arnauld y Nicole, los lógicos de Port Royal, postular que se podía "*aprender*" a partir de la experiencia, que cuanto mayor fuera la cantidad de información con que contara el observador más aproximada sería la estimación de su verdadero valor; **ii)** demostrar que este aprendizaje era cuantificable a través de un proceso de inversión de la probabilidad[8] y, en consecuencia, asumiendo ciertas hipótesis de simplicidad y regularidad, **iii)** establecer el nexo entre las probabilidades "a priori" (definidas a partir de un

---

[6] La primera versión de este tratamiento cuantitativo de la "expectativa" se halla en "*La logique, ou l'art de penser*" (1695) de Arnauld y Nicole, obra que constituye el nexo entre la tradición cualitativa de los retóricos del excepticismo constructivo y J. Bernoulli (es más, el "*Ars conjectandi*" puede ser considerado como un complemento de "*La logique*", conocida en latín como el "*Ars cogitandi*"). Posteriormente Pascal en sus "*Pensées*" (1669), utilizando el argumento hoy conocido como la "*apuesta de Pascal*", interpretó a la racionalidad como la maximización de la expectativa en condiciones de incertidumbre.

[7] En *"Die Werke von Jakob Bernoulli"*, vol. 3 (p. 239).

[8] Si bien figura en algunos cursos dictados en la École Polytechnique a fines del siglo XVIII y principios del siglo XIX (presuntamente por Jean-Baptiste J. Fourier o Jean G. Garnier), la expresión *"probabilidad inversa"* es atribuida (de acuerdo con Arne Fisher (1926)) a Augustus de Morgan (*"An essay on probabilities and their application to life contingencies and insurance offices"* (1838)).

razonamiento que va de las causas a los efectos, de la simetría de los resultados posibles al concepto de equiprobabilidad) y las probabilidades "a posteriori" (definidas a partir de un razonamiento que va de los efectos a las causas, de las estadísticas sobre la mortalidad introducidas por Graunt a la probabilidad de fallecimiento)[9].

2.- Jakob Bernoulli y el teorema de la inversión de la probabilidad

Esta demostración del principio intuitivo que la incertidumbre disminuía en la medida que se incrementaba el número de observaciones, y la cuantificación de dicho proceso de inferencia -conocida luego como la "*ley (débil) de los grandes números"*-, constituyó el primer teorema límite de la teoría de la probabilidad[10].

---

[9] Como se verá en la sección siguiente, para la elaboración de este nexo entre probabilidades "a priori" y "a posteriori" (según surge de la correspondencia intercambiada con Gottfried W. Leibniz entre octubre de 1703 y abril de 1704), Bernoulli utilizó un modelo elaborado a partir de una probabilidad "a posteriori" calculada en base a extracciones "con reposición" de una urna con bolillas de diferentes colores que representaban enfermedades terminales. Las hipótesis sobre las cuales basó su razonamiento fueron: **i)** que, de la misma forma que la urna contenía un gran número de bolillas en una proporción estable, "*...las personas estaban afectadas por un gran número de enfermedades*" y **ii)** que la "*naturaleza observa los comportamientos más simples"* (en Gerhardt (ed.)(1962), vol. 3, p. 77)).

[10] La denominación de "*ley de los grandes números*" referida a una generalización de este teorema referida al comportamiento de los promedios es muy posterior y se debe a Poisson (1837). La forma más rudimentaria de la ley de los grandes números es atribuible a Girolamo Cardano ("*Practica aritmeticæ et mensurandi singularis"* (1539)), quien demostró que el número de ocurrencias de un resultado, en una "*larga serie"* de repeticiones independientes de un fenómeno de comportamiento eventual (X), es aproximadamente igual a $n\theta$ (donde θ denota la probabilidad "a priori" de ocurrencia de dicho resultado en una repetición determinada). Cardano –quien puede ser considerado como el primer gran aleatorista- consideró que la aproximación obtenida al cabo de infinitas repeticiones estaba sujeta a las "*perturbaciones de la fortuna"* (la "*suerte"* en su nomenclatura). Su metafísica podría ser resumida en la siguiente expresión, debida a Daston (1988): "*...el cálculo no puede gobernar lo contingente"* (p. 36). Una idea similar sobre el tratamiento cualitativo del aprendizaje a partir de la experiencia puede hallarse en Halley ("*An estimate the degrees of mortality of mankind drawn from the curious tables of the births and funerals at the city of Breslau; with an attempt to ascertain the price of annuities upon lives"* (1693)), en "*De incerti æstimatione"* de Leibniz (1703-1705) y, comose viomásarriba, en "*La logique, ou l'art de penser"* de Arnauld-Nicole.

Hasta la aparición del "*Ars conjectandi*" los avances producidos en la teoría de la probabilidad no habían conseguido proporcionar una respuesta eficaz a la formalización de este proceso de inferencia. Como se mencionó en la sección precedente, los principales tratados de los autores clásicos -empleando el método de razonamiento desde las "*causas*" a los "*efectos*"- se referían exclusivamente a la resolución de problemas del tipo: dada una urna que se sabe que contiene a bolillas rojas y c bolillas azules, la probabilidad de obtener una bolilla roja al realizar una extracción al azar es $\theta = \frac{a}{a+c}$. J. Bernoulli fue el primero en tratar el esquema empírico inverso: la estimación de los valores a y c, basándose en la evidencia que proporcionaban los resultados de las sucesivas extracciones. Fue el primero en advertir a los probabilistas sobre "*...la relación entre la 'conjectandum' probabilística y la inferencia inductiva*" (Daston (1988, p. 228)), en sustituir el concepto clásico (deductivo) de probabilidad "a priori", basado en la simetría de los resultados posibles y en el concepto Humeano de equiprobabilidad [11] apropiado solamente para resolver problemas relacionados con juegos de azar, por la idea de probabilidad "a posteriori", definida como una medida del conocimiento (la expectativa) que posee el observador acerca de la veracidad de una proposición[12].

---

[11] Hume (1739): "*Dado que una indiferencia total es esencial a la aleatoriedad, no es posible suponer que unaaleatoriedad pueda ser superior a otra, en su defecto puede estar compuesta por un número de aleatoriedades iguales. Afirmar que, de alguna forma, una aleatoriedad puede ser superior a otra implica reconocer que existe algo que origina esa superioridad. Es decir, implica suponer la existencia de una causa y, por lo tanto, la negación de la hipótesis de aleatoriedad. De modo que una indiferencia perfecta y total es esencial a la aleatoriedad y una indiferencia total nunca puede ser intrínsecamente superior o inferior a otra. Esta verdad no es inherente exclusivamente a mi sistema, sino que es reconocida por todos los probabilistas*" (p. 125). Los números de página que figuran en las referencias a "*A treatise of human nature*" de Hume corresponden a la versión inglesa editada por Clarendon Press (1978).

[12] J. Bernoulli ("*Ars conjectandi*"): "*Hemos arribado al punto en el que parece que, para realizar una predicción correcta acerca de un evento cualquiera, solamente es necesario calcular exactamente el número de casos posibles y, luego, determinar cuánto más probable es que ocurra un resultado que otro. Pero aquí surge nuestra dificultad principal, ya que este procedimiento es aplicable solamente a un número reducido de fenómenos, casi exclusivamente a aquellos relacionados con los juegos de azar. Los creadores de estos juegos los diseñaron fijando el número de casos que resultarían en ganancia o pérdida y considerándolos conocidos con anticipación y, también, combinando los casos para que todos*

Utilizando notación moderna, el resultado obtenido por Bernoulli puede ser expresado de la siguiente forma: Sea $Y_n = \frac{X}{n}$ la frecuencia relativa correspondiente al resultado "bolilla roja", obtenida al cabo de una sucesión de n extracciones al azar "con reposición" de una urna cuya composición -desconocida para el observador- es de a bolillas rojas y $c$bolillas azules[13]. Entonces, dados un valor εpositivo y arbitrariamente pequeño, y un valor $t$positivo y arbitrariamente grande, se demuestra que es posible hallar un $n > n(\theta, \varepsilon, t)$ tal que se puede asegurar que, con una probabilidad mayor a $\frac{t^2-1}{t^2}$, la frecuencia relativa del resultado "bolilla roja" se encontrará a una distancia menor o igual que ε del verdadero valor de la probabilidad binomial $\theta = \frac{a}{a+c}$:

$$p(|X - n\theta| \leq \varepsilon) = \sum\nolimits_{|X-n\theta|\leq\varepsilon} b(X_i, 1, \theta) = p\left(\left|\frac{X}{n} - \theta\right| \leq \varepsilon\right) > \frac{t^2 - 1}{t^2}$$

---

fueran igualmente probables, de modo que todos los jugadores tuvieran las mismas chances de ganar. Pero ésta no es, de ninguna manera, la situación de la mayoría de los fenómenos gobernados por las leyes de la naturaleza o la voluntad del hombre (...) Ahora bien, me pregunto, qué mortal podría determinar, contando todos los casos posibles, el número de enfermedades que aflijen al cuerpo humano en cada una de sus partes y a cada edad, y decir cuánto mayor que otra es la probabilidad de una enfermedad de ser fatal. Los resultados dependen, en estos casos, de factores que son completamente oscuros y que, por la infinita complejidad de sus interrelaciones, constantemente engañan a nuestros sentidos. Existe, sin embargo, otro camino que nos conducirá a lo que estamos buscando y nos permitirá conocer, al menos 'a posteriori', lo que no podemos determinar 'a priori', es decir, adquirir conocimiento a partir de los resultados observados en numerosas situaciones similares. Debe suponerse en esta circunstancia que, en condiciones similares, la ocurrencia (o no ocurrencia) de un evento en el futuro observará el mismo comportamiento que fue observado para eventos similares, en el pasado" (p. 226).

[13] $Y_n = \frac{X}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i$ denota una sucesión de sumas parciales. Donde $X_i(1,2,\ldots,n)$ define una sucesión de variables aleatorias binomiales que asumirán los valores 1 ó 0 -con probabilidades θ y $1-\theta$, respectivamente ($X_i\!: b(1,\theta)$)- según que el resultado de la i-ésima extracción fuere "bolilla roja" o "bolilla azul", respectivamente.

A fin de simplificar la demostración de este teorema, e intentando distorsionar lo menos posible la interpretación original de Bernoulli en lo que hace a los alcances de los resultados obtenidos[14], se utilizará aquí la relación conocida como desigualdad de Bienaymé-Chebychev[15]:

$$p[|X - E(X)| \leq t\sigma(X)] \geq 1 - \frac{1}{t^2}$$

Aplicando dicha desigualdad a la variable Y_n se obtiene que[16]:

$$p\left(|Y_n - \theta| \leq t\sqrt{\frac{\theta(1-\theta)}{n}}\right) > 1 - \frac{1}{t^2}$$

y, haciendo$t\sqrt{\frac{\theta(1-\theta)}{n}} = \varepsilon$[17], será:

$$p(|Y_n - \theta| \leq \varepsilon) > 1 - \frac{\theta(1-\theta)}{n\varepsilon^2}$$

---

[14] Para quien desee recrear el escenario en el que se desenvolvía el devenir académico de los probabilistas del siglo XVII, una versión de la demostración original de J. Bernoulli puede ser hallada en Hald (1984).

[15] Bienaymé (1855), Chebychev (1867). Este resultado depende exclusivamente delvalor esperado y la varianza de una variable binomial. En su expresión más general asume la forma $p(|X| \geq \varepsilon) \leq \frac{E(|X|^s)}{\varepsilon^s}(\varepsilon > 0, s > 0)$. Debe tenerse en cuenta que en la época de la demostración de Bernoulli el concepto de varianza era desconocido.

[16] De acuerdo a las hipótesis enunciadas, Y_n es una variable binomial con distribución de probabilidades de la forma $p(Y_n = y_n) = p\left(\frac{X}{n} = \frac{x}{n}\right) = p\left(X_{(n)} = x\right) = \binom{n}{x}\theta^x(1-\theta)^{n-x}(x = 0,1,2,\dots,n)$ y medidas $E(Y_n) = E\left(\frac{X}{n}\right) = \theta$ y $\sigma^2\left(\frac{X}{n}\right) = \frac{\theta(1-\theta)}{n}$.

[17] En la versión original de Bernoulli ε no está definido como cualquier número positivo arbitrario, sino que asume un valor múltiplo de $(a + c)$. El supuesto básico de la demostración original es que $n = k(a + c)$; de modo que $n\theta = ak$ y $n\varepsilon = k$ sean enteros, única forma posible de obtener la igualdad $\frac{X}{n} = \frac{a}{a+c}$. Esta restricción no significó pérdida de rigurosidad en la demostración original, la cual admite una formulación general posterior.

Entoncesse puede escribir$\theta(1-\theta)=\left(\frac{1}{2}-k\right)\left(\frac{1}{2}+k\right)=\frac{1}{4}-k^2$. De esta forma, el producto $\theta(1-\theta)$asume su valor máximo cuando $\theta=\frac{1}{2}$. Lo que prueba que la probabilidad que figura en el primer miembro está uniformemente acotada con respecto a θy a $(1-\theta)$:

$$p(|Y_n-\theta|\leq\varepsilon)>1-\frac{1}{4n\varepsilon^2}$$

de lo que se concluye que$\lim_{n\to\infty}p(|Y_n-\theta|\leq\varepsilon)=1$[18].Es decir que, cuando $n\to\infty$, el estimador$Y_n$converge en-probabilidad al valor θ ($Y_n \xrightarrow{p} \theta$)[19]. De modo que, conocidos n e Y_n, se puede resolver la ecuación$n(Y_n,\varepsilon,t)=n$ con respecto a t, obteniendo así una aproximación al límite inferior $\frac{t^2-1}{t^2}$ (cota inferior de la "*incertidumbre residual"*) correspondiente a la probabilidad de ocurrencia del evento $|Y_n-\theta|\leq\varepsilon$y, a partir de esta expresión, determinar la probabilidad de que θesté incluida en

[18] De acuerdo con la propia terminología de Bernoulli, queda demostrado que es "*moralmente cierto"* (ver Sec. 13) que, para n suficientemente grande, $\frac{X}{n}$ no se desviará en gran medida de θ. En otros términos, establece lo que hoy se conoce como el *"principio de estabilidad de las frecuencias estadísticas"*(von Mises (1928)): si al cabo de n observaciones (para n suficientemente grande) la frecuencia relativa del resultado "bolilla roja" ha sido Y_n, existe un alto grado de probabilidad de que, al cabo de una serie (suficientemente larga) de n' observaciones, la frecuencia relativa del resultado "bolilla roja" sea $Y_{n'}\approx Y_n$. Según lo expresado en sus "*Meditationes"*, Bernoulli concluyó esta demostración entre 1688 y 1690. Ver Kolmogorov (1956).

[19] El concepto de convergencia –que originalmente se aplicó a sucesiones de números y oportunamente se extendió a sucesiones de funciones- es aplicable a sucesiones de infinitas variables aleatorias, $\{Y_n\}$ (que son sucesiones de funciones, $Y_n=Y_n(w)$, definidas en un espacio Ω). Esta convergencia –denominada *"convergencia estocástica"*- si bien posee ciertas características similares a la convergencia en el sentido del análisis, se verifica a través del operador probabilidad. Esta intervención de la probabilidad caracteriza diferentes definiciones de convergencia estocástica. En particular, se dice que una sucesión de variables aleatorias $\{Y_n\}$ converge "*en-probabilidad"* a una variable aleatoria Y, si se verifica que $\lim_{n\to\infty}p(|Y_n-Y|>\varepsilon)=0\ (\varepsilon>0)$ (esta condición suele ser denotada como $plim_{n\to\infty}Y_n=Y$). Como corolario se demuestra que, si $Y_n\xrightarrow{p}Y$ entonces, dada una función continua, g, se verifica que $g(Y_n)\xrightarrow{p}g(Y)$.

un intervalo de la forma $[Y_n - \varepsilon, Y_n + \varepsilon]$, donde θ admite cualquier representación no frecuencista.

Leibniz planteó varias críticas a estaley débil de los grandes números (interpretada como un teorema de la teoría de la estimación) que permiten reflexionar acerca de la naturaleza indiscutiblemente metafísica de los supuestos que condujeron a Bernoulli a postular la asociación entre la interpretación clásica de la probabilidad y la filosofía de la inducción. Una se refiere a la condición necesaria (puramente virtual) de contar con un número infinito de observaciones para identificar la probabilidad θ y la imposibilidad de demostrar la convergencia de una sucesión de frecuencias relativas a partir de un conjunto finito de observaciones (como se mencionó más arriba, la idealización que implica la generación de una sucesión infinita de observaciones admite sólo una convergencia estocástica basada en condiciones inherentes a la lógica de la incertidumbre)[20].

Una segunda objeción considera que, dado que dichas observaciones son asimilables a un conjunto finito de puntos, que los "*resultados ocultos*" son asimilables a una curva que pasa por entre esos puntos y que, para cada conjunto de puntos existen infinitas curvas que cumplen con dicha condición, la probabilidad "a posteriori" obtenida por inferencia es sólo una de las infinitas probabilidades posibles incluidas en el intervalo definido por el teorema.

Una tercera cuestión está relacionada con la exagerada simplificación que implicaba la utilización del esquema de urnas y con el mecanismo combinatorio que suponía la independencia estocástica de los eventos sucesivos.

Sin proponer una metodología general para valores finitos de n, la respuesta de Bernoulli con respecto a las primeras cuestiones se basó en el ambiguo supuesto de uniformidad y simplicidad de la naturaleza[21], según elcual la

[20] Como se verá más adelante, este problema de la convergencia fue tratado rigurosamente, entre otros, por von Mises, Faber, Hausdorff, Cantelli, Kolmogorov, Prokhorov, Khinchin, Doob.

[21] El supuesto pitagórico según el cual Dios había diseñado un universo racional, de acuerdo a principios matemáticos simples.

estimación de θ debía realizarse mediante la selección de la curva más simple[22], de modo que el problema de la inversión quedaba reducido exclusivamente a la selección del estimador más simple de la probabilidad "a priori"[23].

Con respecto al esquema de urnas, debe tenerse en cuenta que, contrariamente a lo que ocurrió con los probabilistas clásicos –que trataron a las loterías, los juegos de dados y las tiradas independientes de una moneda a nivel de aplicaciones prácticas exclusivamente-, como se vio en la sección precedente, Bernoulli utilizó este modelo como la metáfora más adecuada para reproducir los postulados de la filosofía natural de fines del siglo XVII e intentar demostrar una vinculación entre las causas (θ) y los efectos observados (Y_n) según la cual combinaciones aleatorias de causas inaccesibles -ocultas para el observador- podían, a partir de sucesiones infinitas de observaciones repetidas en igualdad de condiciones, producir eventualmente efectos regulares, generando una propuesta alternativa al modelo de causación racionalista, el cual, para un observador ideal capaz de acceder a un conjunto de información de tamaño infinito, la relación de causación era esencialmente deductiva y, en consecuencia, no-combinatoria[24].

---

[22] "*La determinación de una trayectoria a partir de un conjunto finito de observaciones (...) sería bastante débil e incierta si no se tuviera en cuenta que la curva a seleccionar pertenece a la clase de las curvas simples (...) ya que podemos observar que la naturaleza obedece a los comportamientos más simples*" (ver correspondencia de Bernoulli a Leibniz en Gerhardt (ed.)(1962, vol. 3, p. 11)).

[23] Como se verá en la Sec. 4, en su intento de solución del problema de la inversión, de Moivre ("*Approximatio ad summam terminorum binomii* $(a+b)^n$ *in seriem expansi*" (1733)) reemplazó este supuestode simplicidad por el de "*Orden*" resultante de un "*Diseño Original*" en el comportamiento de los fenómenos naturales. Las distintas interpretaciones matemáticas de ese "*Orden*" generaron una serie de controversias que constituyen uno de los capítulos más interesantes del desarrollo de la teoría de la probabilidad durante el siglo XVIII.

[24] Excepto para los filósofos atomistas como Charleton ("*Physologia epicuro-gassendo -harletoniana*" (1654)), Gassendi ("*Syntagma*" (1659)) y Boyle ("*Origin of forms and qualities*" (1666)), esta concepción debe haber resultado extraña a los científicos de los siglos XVII, XVIII y comienzos del XIX. Recién a fines de este último siglo, a través de los trabajos de Maxwell, Boltzman y Clausius, entre otros, esta "*explicación estadística*" comenzó a ser aceptada por los físicos. Este concepto de "*causas*" determinadas y a la vez combinatorias, regulares y a la vez estocásticamente independientes, tiene su precedente, precisamente, en la teoría corpuscular de Charleton-Gassendi-Boyle.

Fue su postura (más teológica y filosófica que matemática) de convencido militante del determinismo metafísico la que condujo a Bernoulli a la identificación de las causas ignoradas del comportamiento de los fenómenos con el parámetro θ, determinado e invariable y representativo de la naturaleza gobernada por leyes inmutables y, en consecuencia, a una interpretación de su teorema que consideraba a la probabilidad θ como conocida, limitando sus alcances a la demostración de la convergencia en-probabilidad de la frecuencia relativa Y_n a θ y, en segundo lugar, a la rigurosa determinación del número de observaciones $n(\theta, \varepsilon, t)$ que se requiere para que la probabilidad de que el valor de Y_n esté incluido en un intervalo $[\theta - \varepsilon, \theta + \varepsilon]$ sea t^2 veces mayor que la probabilidad de que no lo esté[25]. Según el ejemplo desarrollado por el mismo Bernoulli en el "*Ars conjectandi*", para una urna que contiene 30 bolillas rojas y 20 azules, un $\varepsilon = \frac{1}{50}$ y $t^2 = 1000$, se necesitarían $n \geq 25.550$ observaciones; es decir, a partir de la información proporcionada por 25.550 observaciones de los resultados del juego (extracciones "con reposición"), se podría asegurar que $p\left(\frac{29}{50} \leq \frac{X}{25.550} \leq \frac{31}{50}\right) > \frac{1000}{1001}$. De la misma forma demostró que, para $t^2 = 10.000$, se necesitarían $n \geq 31.258$ observaciones y para $t^2 = 100.000$, se necesitarían $n \geq 36.966$ observaciones. Estos resultados -cuyas magnitudes resultaban desproporcionadas para la época- indudablemente deben haber sorprendido -y decepcionado- a Bernoulli -y a sus contemporáneos-, y quizás haya sido la sospecha de la existencia de algún error en el tratamiento matemático del problema -y no razones de orden filosófico- la causa fundamental de su negativa a publicar los

[25] J. Bernoulli ("*Ars conjectandi*"): "*Este tipo de predicción requiere 'un gran número de observaciones' (...) pero, si bien todos reconocemos esto, la demostración científica de este principio no es simple (...) Existe, además, otro problema a ser tomado en consideración (...) podría ocurrir que al incrementar el número de observaciones también se incrementara la probabilidad de que la proporción registrada entre resultados favorables y no-favorables se aproximara a un 'verdadero cociente', de modo que dicha probabilidad finalmente superara cualquier grado deseado de certeza, o podría ocurrir que el problema presentara una asíntota, lo cual implicaría la existencia de un grado de certeza particular de que el 'verdadero cociente' habría sido hallado, que no podría ser superado no importa cuánto se aumentara el número de observaciones. Para que esta cuestión no sea interpretada en forma imperfecta, debe tenerse en cuenta que el cociente que representa la verdadera relación entre los números de casos, nunca puede ser obtenido con absoluta seguridad (...) El cociente al cual arribamos es sólo aproximado: deben definirse dos límites, pero esos límites pueden ser construidos de modo que se aproximen tanto como deseemos*" (p. 225).

resultados de sus investigaciones de 20 años[26].

Apartir de esta demostración Bernoulli intentó una extensión de su teorema que implicaba un planteo inverso de la forma: si se observa que la frecuencia relativa "*...converge a un valor determinado*", θ, entonces este valor definirá la "*ley*" que gobierna a dicho evento. Pero la circularidad de este esquema de razonamiento en el cual la convergencia en-probabilidad de las frecuencias relativas se verificaba porque los eventos estaban regidos por una ley determinada pero, a su vez, la convicción de que los eventos se regían por una ley determinada se fundaba en el postulado de inversión de la probabilidad según el cual las frecuencias relativas debían converger a θ, no le permitió obtener una justificación rigurosa de los alcances de su teorema como fundamento de una teoría de la inversión de la probabilidad[27].

[26] Si bien el objetivo de J. Bernoulli fue el desarrollo de una teoría general de la decisión racional en condiciones de incertidumbre, de los tres primeros capítulos de la Parte IV del "*Ars conjectandi*" se puede concluir en forma inmediata que los conceptos y problemas tratados están íntimamente relacionados con la necesidad de incrementar la importancia relativa de las consideraciones objetivas exógenas en la práctica legal de los contratos aleatorios (un movimiento que se originó entre los juristas del siglo XII en Europa continental, y que tropezó inevitablemente con el problema de la cuantificación). Estos intentos de aplicar la teoría de la probabilidad a cuestiones legales en general, y a la credibilidad de los testimonios en un juicio en particular comenzó, en realidad, con Craig ("*Theologiæ christianæ principia mathematica*", 1699) y, pasando fundamentalmente por N. Bernoulli ("*De usu artis conjectandi in jure*", 1709) y J. Bernoulli ("*Ars conjectandi*", 1713), continuaron hasta Poisson ("*Recherches sur la probabilité des jugements en matière criminelle et en matière civile*", 1837). Ver Stigler (1986).

[27] Paradójicamente, poco después de su publicación el "*Ars conjectandi*" -que logró el inmenso avance que implica trasladar el eje central del razonamiento de las expectativas a las probabilidades, es decir de los resultados posibles equiprobables a las medidas de incertidumbre- desapareció como instrumento para el estudio de la teoría matemática de la probabilidad y fue sustituido en el ámbito académico por la segunda edición del "*Essai d'analyse sur les jeux de hazard*" (1713) de Pierre de Montmort.

J. Bernoulli concluyó su tratado con el siguiente postulado, posteriormente adoptado por muchos científicos (incluyendo a Laplace) como el fundamento de la filosofía determinística[28]: "*Si algo que está destinado a ocurrir no es cierto que ocurra, no resulta claro entonces cómo puede permanecer inquebrantable el elogio de la omnisciencia y omnipotencia del Todopoderoso. Si todos los eventos fueran observados en forma continua, desde ahora y por toda la eternidad (con lo cual la probabilidad se transformaría en certeza), se concluiría que en el mundo todo ocurre por razones definidas y de conformidad con una Ley y que, por lo tanto, estamos obligados, aún en casos que parecen ser accidentales, a suponer una cierta necesidad o fatalismo(…)Dadas su posición, su velocidad y su distancia del tablero, cuando el dado abandona la mano del jugador, indudablemente, no puede caer de una forma distinta de la que, en realidad, cae. En forma similar, para una composición dada del aire y para masas, posiciones, direcciones y velocidad de los vientos, vapor y nubes dados, y dadas, también, las leyes de la mecánica que gobiernan todas estas interacciones, el clima correspondiente al día de mañana no podrá ser diferente del que debería ser en realidad. De modo que estos fenómenos se comportan con una regularidad no menor (non minus necessario) que la de los eclipses de los cuerpos celestes. Es, sin embargo, una práctica usual considerar a un eclipse como un evento regular, mientras que a la caída de un dado o al clima del día de mañana se los considera como eventos aleatorios. La razón de esto radica exclusivamente en el conocimiento insuficiente de las acciones sucesivas en la naturaleza. Y, aún en el caso en que éstas fueran conocidas, nuestro saber en matemáticas y física no está suficientemente desarrollado, por lo que, partiendo de las causas iniciales, no podemos calcular estos fenómenos, en tanto que, a partir de los principios absolutos de la astronomía, es posible precalcular y predecir los eclipses (...) La aleatoriedad depende, fundamentalmente, de nuestro conocimiento (...) el arte de la predicción está definido aquí como el arte de medir las probabilidades de los eventos tan precisamente como sea posible, a fin de lograr que nuestras decisiones y acciones sean las mejores, las más satisfactorias, fáciles y razonables. En esto consiste la sabiduría del filósofo y la prudencia del estadista*" (p. 237)[29].

[28] Laplace (1814), si bien coincidió con la posición de Bernoulli, reemplazó al "*Todopoderoso*" por "*...una inteligencia capaz de comprender a todas las fuerzas de la naturaleza*" (p. xi).

[29] En este punto el "*Ars conjectandi*" concluye abruptamente, sin hacer mención alguna a las aplicaciones de la teoría de la probabilidad a problemas civiles y económicos prometidas en el título de la "*Pars Quarta*".

3.- Nikolaus Bernoulli y el problema de los sexos

En "*Usu artis conjectandi in jure*" (1713), con motivo de la discusión que se había generado acerca de la estabilidad del cociente entre el número de nacimientos masculinos y femeninos en Londres, Nikolaus Bernoulli, a partir fundamentalmente de los resultados obtenidos por Lodewijk Huygens y John Arbuthnot y, en menor medida, por su tío Jakob, logró una demostración de la ley de los grandes números sin restricciones sobre los valores a asumir por ε, suponiendo solamente que debía ser pequeño en comparación con n.

En realidad, la primera mención del problema de la definición de un modelo para estimar la proporción de nacimientos masculinos y femeninos se debe a John Graunt ("*Natural and political observations made upon the bills of mortality*" (1662)), quien postuló que la proporción de bautismos masculinos en Londres, era mayor que 1/2 y que la misma no variaba mucho en el tiempo. En 1712, Arbuthnott ("*An argument for Divine Providence, taken from the constant regularity observed in the births of both sexes*") sugirió que la variación del número de bautismos masculinos podía ser considerada como una variable binomial e intentó diseñar un test (generalizado luego por Willem 'sGravesande, "*Démonstration mathématique du soin que Dieu prend de diriger ce qui se passé dans ce monde, tirée du nombre des garçons et des filles qui naissent journellement*" (1712)[30]) para contrastar la hipótesis de que la probabilidad de que se produjera un nacimiento masculino (θ) era igual a $\frac{1}{2}$, respecto de la alternativa $\theta \neq \frac{1}{2}$, demostrando que, para cualquier número de nacimientos, se verificaba que $p\left[M > F/\left(\theta = \frac{1}{2}\right)\right] \leq \frac{1}{2}$ (donde M y F denotan el número de nacimientos masculinos y femeninos, respectivamente) y luego utilizó este resultado para transformar la hipótesis original en la hipótesis $p(M < F) \leq \frac{1}{2}$, y la alternativa $p(M < F) > \frac{1}{2}$[31].

---

[30] Esta edición fue de circulación reducida. Los principales resultados de este trabajo fueron publicados en 1715, por el médico holandés Bernard Nieuwentyt bajo el título "*Het regt gebruik der Wereldbeschouwingen*".

[31] A partir de estos resultados William Derham publicó "*Physico-theology, or a demonstration of the being and attributes of God from his works of creation*" (1716), obra que se transformó en un clàsico del pensamiento de la Royal Society. Posteriormente, este problema fue tratado por Daniel Bernoulli ("*Mensura sortis ad fortuitam successionem rerum naturaliter contingentium applicata*" (1770-1771)), Pierre Simon de Laplace ("*Mémoire sur la probabilité*

La intención de N. Bernoulli fue comparar la distribución de las observaciones con la función de probabilidades binomial, con el objeto de determinar si la variación observada podía ser explicada por dicho esquema (comparó los datos con una distribución binomial con parámetro $\theta = \frac{18}{35}$, y aceptó, equivocadamente, la hipótesis que, para el período al cual corresponden las observaciones, la variación era binomial) y, sobre todo, de lograr una herramienta matemática que permitiera una aproximación a la probabilidad:

$$p(|Y_n - \theta| \leq \varepsilon) = p(|X - n\theta| \leq \varepsilon^*) = \sum_{|X-n\theta|\leq\varepsilon^*} \theta^x (1-\theta)^{n-x} =$$
$$= \sum_{|X-n\theta|\leq\varepsilon^*} b(x, n, \theta)$$

que evitara las enormes complicaciones que se presentaban en el cálculo de los coeficientes binomiales cuando los valores de n crecían. Los resultados fueron la obtención del límite inferior:

$$p(|X - n\theta| \leq \varepsilon^*) > 1 - max\left(\frac{b(n\theta+\varepsilon^*, n, \theta)}{b(n\theta, n, \theta)}, \frac{b(n\theta-\varepsilon^*, n, \theta)}{b(n\theta, n, \theta)}\right)$$

y la siguiente aproximación para dicho límite inferior:

$$\frac{b(n\theta, n, \theta)}{b(n\theta+\varepsilon^*, n, \theta)} = \left(\frac{n\theta+\varepsilon^*}{n(1-\theta)-\varepsilon^*+1} . \frac{n\theta+1}{n\theta} . \frac{n(1-\theta)}{n\theta}\right)^{\varepsilon^*/2}$$

---

des causes par les èvénements" (1774), "*Sur les naissances, les mariages et les morts a Paris, depuis 1771 jusqu'en 1784, et dans toute l'étendue de la France, pendant les anées 1781 et 1782"* (1786), "*Mémoire sur les approximations des formules qui sont fonctions de trés grands nombres et sur leur application aux probabilities"* (1810)) y Wilhelm Lexis ("*Das Geschlechverhältnis der Geborenen und die Wahrscheinlichkeitsrechnung"* (1876)), quien demostró que la probabilidad de los nacimientos masculinos varía temporal y geográficamente.

Tomando $n = 14.000$ como el número de nacimientos anuales en Londres, $\theta = \frac{18}{35} = 0{,}514265$ y $\varepsilon^* = n\varepsilon = 163$ (es decir $\varepsilon = 0{,}0116$), la demostración de N. Bernoulli proporciona los siguientes resultados: $\frac{b(n\theta+\varepsilon^*,n,\theta)}{b(n\theta,n,\theta)} \cong 44{,}74$ y $\frac{b(n\theta-\varepsilon^*,n,\theta)}{b(n\theta,n,\theta)} \cong 44{,}58$y, por lo tanto[32]:

$$p\left(\left|Y_{14.000} - \frac{18}{35}\right| \leq 0{,}0116\right) > 0{,}9776$$

Como corolario de su teorema se obtiene que el resultado de Jakob Bernoulli se verifica para[33]:

$$\varepsilon^*(\theta, 1-\theta, t^2) \geq \frac{ln(t^2+1)}{ln\left(\frac{(a+c)\theta+1}{a+c}\right)} \cdot \frac{(a+c)+1}{(a+c)\theta+1} - \frac{(a+c)(1-\theta)}{(a+c)\theta+1}$$

De acuerdo con esta expresión, el resultado $n \geq 25.550$ obtenido de la aplicación del teorema de J. Bernoulli para $a + c = 50$ y $t^2 = 1.000$, se transforma en $n \geq 17.350$.

En una revisión de este resultado (obtenida como una combinación de las demostraciones de Jakob y Niklaus) J.V. Uspensky (1937) concluyó que $\varepsilon^* \geq 1 + (a+c+1)ln(t^2+1)$. Resultado que le permitió demostrar que, para p, ε y $\delta > 0$, su expresión del teorema de J. Bernoulli, $p(|Y_n - p| \leq \varepsilon) > 1 - \delta$, se verifica para $n \geq \frac{1+\varepsilon}{\varepsilon} ln\left(\frac{1}{\delta}\right) + \frac{1}{\varepsilon}$. De modo que, para las condiciones fijadas en el ejemplo precedente, se obtendría que $n \geq 17.700$[34].

[32] Obsérvese que con los mismos datos, el resultado obtenido por J. Bernoulli hubiera sido $p > 0{,}1353$, lo cual pone en evidencia el avance producido por N. Bernoulli. Como se verá en la próxima sección, el resultado obtenido por de Moivre, con los mismos datos, utilizando la aproximación Normal, hubiera sido $p = 0{,}9942$.

[33] Los resultados obtenidos por N. Bernoulli se encuentran en su correspondencia con de Montmort.

[34] A.A. Markov (1924) propuso una versión "modernizada" del teorema de J. Bernoulli utilizando los resultados obtenidos por Niklaus, pero omitiendo la referencia bibliográfica. Asimismo, cabe mencionar que la versión "modernizada" de Uspensky de la demostración de Markov omite la referencia bibliográfica referida a éste y a los Bernoulli.

4.- Abraham de Moivre y el primer teorema central del límite

En 1733, de Moivre, continuando la línea teórica adoptada por N. Bernoulli, obtuvo una nueva aproximación a la distribución binomial. Este resultado, al cual de Moivre asignó gran importancia, fue publicado bajo el título de "*Approximatio ad summam terminorum binomii (a+b)^n in seriem expansi*". En realidad, estetrabajo constituye una continuación de la serie de investigaciones comenzadas en 1721 (publicadas luego en la "*Miscellanea analytica de seriebus et quadraturibus" (1730)*)[35] a partir del siguiente problema (propuesto por Alexander Cuming (1721)): Sean A y B dos jugadores empeñados en un juego consistente en una serie de n partidas, tales que las probabilidades de los jugadores de ganar cada partida sean, respectivamente, θ y $1-\theta$, para toda la serie. Al cabo de las n partidas A deberá pagar a una tercera persona C, tantas fichas como partidas haya ganado sobre la cantidad $n\theta$ (que se supone que puede asumir solamente valores enteros), y B deberá pagar a C tantas fichas como partidas haya ganado sobre la cantidad $n(1-\theta)$. ¿Cuál es el valor esperado del juego para C?

La expectativa de C será, obviamente:

$$E(C) = \sum_{x=n\theta+1}^{n} (x-n\theta)b(x,n,\theta) + \sum_{x=0}^{n\theta-1} [(n-x)-n(1-\theta)]b(x,n,\theta)$$

$$= \sum\nolimits_{x=0}^{n} |x-n\theta| b(x,n,\theta)$$

Partiendo de la igualdad $\sum_{x=0}^{n}(x-n\theta)b(x,n,\theta)=0$, entonces se puede escribir:

$$E(C) = 2\sum\nolimits_{x=n\theta}^{n}(x-n\theta)b(x,n,\theta) = 2\sum\nolimits_{y=0}^{n(1-\theta)} yb(y+n\theta,n,\theta) =$$

---

[35] El "*Approximatio*" -un pequeño trabajo de sólo 7 páginas-, traducido al inglés por el mismo de Moivre con el título "*A method of approximating the sum of the terms of the binomial (a+b)^n expanded into a series, from whence are deduced some practical rules to estimate the degree of assent which is to be given to experiments*" constituyó, en principio, un apartado de circulación privada fechado el 12-11-1733. Posteriormente fue incluido como suplemento en las ediciones sucesivas de la "*Miscellanea Analytica*" (título bajo el cual fue publicada la colección de sus investigaciones del período 1721-1730).

$$= 2\sum_{y=0}^{n(1-\theta)} y\binom{n}{y+n\theta} p^{y+n\theta} q^{n-y-n\theta}$$

Ahora bien, si se tiene en cuenta que $\frac{b(y+n\theta+1,n,\theta)}{b(y+n\theta,n,\theta)} = \frac{n-y-n\theta}{y+n\theta+1} \cdot \frac{\theta}{1-\theta}$, es decir, que:

$$(n-y-n\theta)\theta b(y+n\theta,n,\theta) = (y+n\theta+1)(1-\theta)b(y+n\theta,n,\theta)$$

sumando entre 0 y $n(1-\theta)$, se obtiene que:

$$\theta\sum_{y=0}^{n(1-\theta)} yb(y+n\theta,n,\theta) + (1-\theta)\sum_{y=0}^{n(1-\theta)} (y+1)b(y+n\theta+1,n,\theta)$$

$$= n\theta(1-\theta)\left[\sum_{y=0}^{n(1-\theta)} b(y+n\theta,n,\theta) - \sum_{y=0}^{n(1-\theta)} b(y+n\theta+1,n,\theta)\right]$$

Por lo tanto, será:

$$\sum_{y=0}^{n(1-\theta)} yb(y+n\theta,n,\theta) = n\theta(1-\theta)b(n\theta,n,\theta)$$

Es decir, queda demostrado que:

$$E(C) = 2n\theta(1-\theta)b(n\theta,n,\theta) = 2n\theta(1-\theta)\binom{n}{n\theta}\theta^{n\theta}(1-\theta)^{n\theta}$$

A fin de facilitar el cálculo de los coeficientes binomiales, de Moivre comenzó buscando una aproximación a la función $ln(n!)$ obteniendo la expresión $ln\left(\frac{n^n}{n!}\right) \approx n - \frac{1}{2}ln(n) - \frac{1}{2}ln(2\pi)$[36]. Luego, teniendo en cuenta

[36] de Moivre, basándose en esta aproximación, publicó su notable tabla de 14 dígitos de la función log(n!), para n=10(10)900, como suplemento a la "*Miscellanea analytica*" (1730). Contemporáneamente James Stirling dio a conocer su solución (conocida hoy como "*aproximación de de Moivre-Stirling*") en "*Methodus differentialis*" (1730) (ver Archibald (1926), Chow; Teicher (2003)).

que:

$$b(n\theta,n,\theta)=\frac{n!}{(n\theta)!\,\big(n(1-\theta)\big)!}\theta^{n\theta}=\frac{(n\theta)^{n\theta}}{(n\theta)!}\cdot\frac{\big(n(1-\theta)\big)^{n(1-\theta)}}{\big(n(1-\theta)\big)!}\cdot\frac{n!}{n^n}$$

concluyó en forma inmediata que:

$$ln[b(n\theta,n,\theta)]=ln\left[\frac{(n\theta)^{n\theta}}{(n\theta)!}\right]+ln\left[\frac{\big(n(1-\theta)\big)^{n(1-\theta)}}{\big(n(1-\theta)\big)!}\right]-ln\left(\frac{n!}{n^n}\right)=$$

$$=-ln\left(\sqrt{n^2\theta(1-\theta)}\right)-ln\big(\sqrt{2\pi}\big)+ln\big(\sqrt{n}\big)=$$

$$-ln\left(\sqrt{2\pi n\theta(1-\theta)}\right)$$

Es decir, que $b(n\theta,n,\theta)=\big(2\pi n\theta(1-\theta)\big)^{-1/2}$[37]. Por otra parte, a partir de la relación:

$$\frac{b(n\theta,n,\theta)}{b(n\theta+\varepsilon,n,\theta)}=\frac{(n\theta+\varepsilon)!\,(n(1-\theta)-\varepsilon)!}{(n\theta)!\,\big(n(1-\theta)\big)!}\cdot\frac{\big(n(1-\theta)\big)^{\varepsilon}}{(n\theta)^{\varepsilon}}=$$

$$=\frac{(n\theta+\varepsilon)!(n(1-\theta)-\varepsilon)!}{(n\theta)!(n(1-\theta)-1)!}\cdot\frac{\big(n(1-\theta)\big)^{\varepsilon-1}}{(n\theta)^{\varepsilon}}=$$

$$=\frac{n\theta+\varepsilon}{n\theta}\cdot\frac{\big(n(1-\theta)\big)^{\varepsilon-1}}{(n\theta)^{\varepsilon-1}}\cdot\frac{\frac{(n\theta+\varepsilon-1)!}{(n\theta)!}}{\frac{(n(1-\theta)-1)!}{(n(1-\theta)-\varepsilon)!}}=$$

$$=\frac{n\theta+\varepsilon}{n\theta}\cdot\frac{\frac{(n\theta+1)(n\theta+2)\dots(n\theta+\varepsilon-1)}{(n\theta)^{\varepsilon-1}}}{\frac{(n(1-\theta)-1)(n(1-\theta)-2)\dots(n(1-\theta)-\varepsilon+1)}{\big(n(1-\theta)\big)^{\varepsilon-1}}}=$$

[37] Nótese que, contrariamente a lo realizado por J. y N. Bernoulli -quienes intentaron una aproximación a $p(|X-n\theta|\leq\varepsilon)$-, de Moivre buscó una aproximación a $b(x,n,\theta)$.

$$= \left(1+\frac{\varepsilon}{n\theta}\right)\prod_{j=1}^{\varepsilon-1}\frac{1+\frac{j}{n\theta}}{1-\frac{j}{n(1-\theta)}}$$

obtuvo que:

$$ln\left(\frac{b(n\theta,n,\theta)}{b(n\theta+\varepsilon,n,\theta)}\right) =$$

$$= ln\left(1+\frac{\varepsilon}{n\theta}\right)+\sum_{j=1}^{\varepsilon-1}\left[ln\left(1+\frac{j}{n\theta}\right)-ln\left(1-\frac{j}{n(1-\theta)}\right)\right] =$$

$$= \left[\frac{\varepsilon}{n\theta}-\frac{1}{2}\left(\frac{\varepsilon}{n\theta}\right)^2+\frac{1}{3}\left(\frac{\varepsilon}{n\theta}\right)^3-\cdots\right]+$$

$$+\sum_{j=1}^{\varepsilon-1}\left[\frac{j}{n\theta(1-\theta)}+\frac{(\theta^3+(1-\theta)^3)}{3\big(n\theta(1-\theta)\big)^3}+\cdots\right] =$$

$$= \left[\frac{\varepsilon}{n\theta}-\frac{1}{2}\left(\frac{\varepsilon}{n\theta}\right)^2+\frac{1}{3}\left(\frac{\varepsilon}{n\theta}\right)^3-\cdots\right]+\frac{\varepsilon(\varepsilon-1)}{2n\theta(1-\theta)} \approx$$

$$\approx \frac{\varepsilon^2}{2n\theta(1-\theta)}-\frac{\varepsilon}{2n\theta(1-\theta)}+\frac{\varepsilon}{n\theta}-\frac{\varepsilon^2}{2(n\theta)^2}+\cdots$$

donde $\varepsilon = O(\sqrt{n})$, es decir, tal que $\lim_{n\to\infty}\left(\frac{|\varepsilon|}{\sqrt{n}}\right) < K$. Luego, para valores de n suficientemente grandes, la expresión anterior se redujo a $ln\left[\frac{b(n\theta,n,\theta)}{b(n\theta+\varepsilon,n,\theta)}\right] \approx \frac{\varepsilon^2}{2n\theta(1-\theta)}$, o, lo que es lo mismo, a:

$$b(n\theta+\varepsilon,n,\theta) \approx n\}b(n\theta,n,\theta)e^{-\frac{\varepsilon^2}{2n\theta(1-\theta)}} \approx$$

$$\approx \frac{1}{\sqrt{2\pi n\theta(1-\theta)}}e^{-\varepsilon^2/2n\theta(1-\theta)}$$

A partir de esteresultado concluyó, entonces, que:

$$p(|x-n\theta|\leq\varepsilon)=\sum_{|x-n\theta|\leq\varepsilon}b(x,n,\theta)=\frac{2}{\sqrt{2\pi}}\int_{0}^{\varepsilon/\sqrt{n\theta(1-\theta)}}e^{-y^2/2}dy$$

Donde $f(y)$=$\frac{1}{\sqrt{2\pi}}e^{y^2/2}dy$ denota la función de densidad[38].

En términos resumidos este teorema -conocido hoy como *"primer teorema central del límite"*[39]- podría ser enunciado de la siguiente forma: Sea $f_N(\cdot)$ la función de densidad, para una variable θ $(0<\theta<1)$ y sea una variable acotada $\frac{X-n\theta}{\sqrt{n\theta(1-\theta)}}=\frac{\sum_{(n)}U-n\theta}{\sqrt{n\theta(1-\theta\theta)}}$ (donde las U son variables del tipo $b(1,\theta)$), se verifica entonces que:

$$\lim_{n\to\infty}\left[\sqrt{n\theta(1-\theta)}b(x,n,\theta)\right]=f_{N(v)}$$

Es decir, que $\frac{\sum_n u-n\theta}{\sqrt{n\theta(1-\theta)}}$ converge en-distribuición a la variable θ

---

[38] De acuerdo con su costumbre, de Moivre omitió la demostración de este teorema. El desarrollo que figura en el texto fue elaborado estrictamente a partir de relaciones definidas en publicaciones previas (ver Hald (1984), Sheynin (1968)). Obsérvese que N. Bernoulli estuvo muy cerca de anticipar la solución de de Moivre: si hubiese utilizado series logarítmicas sobre su aproximación hubiera hallado que: $ln\left[\frac{b(n\theta,n,\theta)}{b(n\theta+\varepsilon,n,\theta)}\right]=\left[ln\left(\frac{n\theta+\varepsilon}{n(1-\theta)+\varepsilon+1}\right)+ln\left(\frac{n\theta+1}{n\theta}\right)+ln\left(\frac{n(1-\theta)}{n\theta}\right)\right]=\frac{\varepsilon}{2}\left[ln\left(1+\frac{\varepsilon}{n\theta}\right)-ln\left(1-\frac{\varepsilon-1}{n(1-\theta)}\right)+ln\left(1+\frac{1}{n\theta}\right)\right]=\frac{\varepsilon^2}{2n\theta(1-\theta)}-\frac{\theta\varepsilon}{2n\theta(1-\theta)}+\frac{\varepsilon^2}{4}.\frac{\theta^2-(1-\theta)^2}{(n\theta(1-\theta))^2}-\frac{\varepsilon^2}{2}.\frac{\theta^2}{(n\theta(1-\theta))^2}+\cdots$ Y, a partir del supuesto que $\varepsilon=O(\sqrt{n})$, hubiera obtenido inmediatamente que $\lim_{n\to\infty}b(n\theta+\varepsilon,n,\theta)=b(n\theta,n,\theta)\varepsilon^{n\theta(1-\theta)}$.

[39] El calificativo "*central*" agregado a la denominación de este teorema se debe a Pólya (1920), quien quiso, de esta manera, destacar el papel central que juega en la teoría de la probabilidad. Algunos probabilistas franceses entendieron que la inclusión del adjetivo central se debía a que el teorema describía el comportamiento de los valores centrales de la distribución de probabilidades de la variable binomial y que, por lo tanto, dicho adjetivo debía calificar al sustantivo "*límite*", no al sustantivo "*teorema*". Por esta razón, es posible hallar en la literatura también la denominación "*teorema del límite central*". La calificación de "*primero*" implica que la denominación de teorema central del límite resume una familia de teoremas que postulan sucesivas generalizaciones y cuya culminación son las denominadas "*condiciones de Feller-Lindeberg*" (ver Sec. 9).

$(\frac{\sum_n u-n\theta}{\sqrt{n\theta(1-\theta)}} \xrightarrow{d} \theta)$[40].

En particular, si el desvío ε es medido como t veces $\sqrt{n\theta(1-\theta)}$, entonces se puede obtener el valor de t que determina la probabilidad de que la variable X asuma valores comprendidos en un intervalo dado:

$$p\left(|x-n\theta| \leq t\sqrt{n\theta(1-\theta)}\right) = p\left(\left|\frac{x}{n}-\theta\right| \leq t\sqrt{\frac{\theta(1-\theta)}{n}}\right) =$$

$$= \sum_{|x-n\theta|\leq t\sqrt{n\theta(1-\theta)}} b(x,n,\theta) =$$

$$= p\left(n\theta - t\sqrt{n\theta(1-\theta)} \leq X \leq n\theta + t\sqrt{n\theta(1-\theta)}\right) =$$

$$= p\left(\theta - t\sqrt{\frac{\theta(1-\theta)}{n}} \leq Y_n \leq \theta + t\sqrt{\frac{\theta(1-\theta)}{n}}\right) \approx \frac{2}{\sqrt{2\pi}}\int_0^t e^{-u^2/2}du$$

La expresión precedente permite deducir que el grado de confiabilidad de la estimación aumenta en forma proporcional a la raíz cuadrada del número de observaciones independientes realizadas. Este resultado constituye el gran avance de de Moivre con respecto a la solución de Bernoulli: la cuantificación efectiva del aumento de confiabilidad ante un incremento de la información empírica y una justificación de su modelo implícito de causación

---

[40] Sean una sucesión de variables aleatorias independientes entre sí, $\{X_1, X_2, \ldots, X_n\}$, con funciones de ditribución $F_{X_j}(x)(j = 1,2,\ldots,n)$ y una variable X, con función de distribución $F_X(x)$, se dice que la sucesión $\{X_n\}$ converge en-distribución (o en-ley o débilmente o en el sentido de Bernoulli) a la variable X ($\{X_n\} \xrightarrow{d} X$) si se verifica que $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$. La condición necesaria y suficiente para poder asegurar que la sucesión $\{X_1, X_2, \ldots, X_n\}$ es de variables estocásticamente independientes es que $F_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = \prod_{j=1}^{n} F_{X_j}(x_j)$. Esta condición se verifica también para cualquier sucesión infinita de variables independientes.

combinatoria[41].

A fin de comprender mejor la evolución histórica posterior de los intentos de solución del problema de Bernoulli, cabe realizar algunas consideraciones sobre esta demostración de de Moivre:

i) Algunos autores, como por ejemplo, Karl Pearson (1924)(1925), interpretaron al resultado obtenido por de Moivre como el "*nacimiento*" de la función Normal. Pero, es necesario destacar que de Moivre no consideró a esta función como una función de densidad, sino simplemente como una curva útil para aproximar probabilidades binomiales[42].

ii) de Moivre intentó resolver el problema de la inversión de la probabilidad por el absurdo. Bernoulli demostró que, suponiendo ""*...una cierta ley determinada"*, la frecuencia relativa de los eventos "*...se aproximará en forma continua a dicha ley"*, a medida que el número de observaciones aumente indefinidamente. A partir de este postulado, el razonamiento de de Moivre puede ser expresado de la siguiente forma: si se observa que la frecuencia relativa "*...converge a un valor determinado"*, θ, entonces quedaría definida la "*ley"* que "*gobierna"* a dicho evento. Por otra parte, si esta ley fuera expresada por otro valor, θ', el teorema de Bernoulli dictaminaría que la frecuencia relativa debería converger a ese valor. Pero esto violaría la hipótesis de que la frecuencia relativa converge a θ.Este resultado y el supuesto que el principio de la estabilidad de las frecuencias era una prueba incontrovertible de la existencia de una inteligencia superior que regía el comportamiento de los fenómenos naturales (es decir, de la existencia de un "*Diseño original*"), lo condujeron a la convicción de haber demostrado su propia versión inversa del teorema de Bernoulli, la cual tampoco pudo resolver la circularidad del razonamiento. No obstante, este planteo constituyó un argumento de consideración contra el escepticismo radical que sostenía que las causas regulares no necesariamente tenían que producir efectos regulares. De acuerdo con suinterpretación, no sólo debía esperarse que, en el largo plazo, causas regulares produjeran efectos regulares, sino que

---

[41] $\sqrt{n}$ constituye "*...el módulo mediante el cual regulamos nuestra estimación"* ("*Doctrine of chances"* (1718, p. 252)).

[42] Si bien de Moivre no demostró que $\int_{-\infty}^{\infty} f(u)du = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-u^2/2}du = 1$, utilizó a $f(u)$ como una función de densidad.

la observación de los efectos permitía asintóticamente el descubrimiento de las causas bajo el supuesto ontológico de que dichas causas ("*leyes determinadas*") existían[43].

iii) La restricción más importante de la demostración de de Moivre radica en que la convergencia de la frecuencia relativa se justifica sólo en el límite. Es decir, al igual que Bernoulli, tampoco de Moivre logró resolver el problema de la identificación de la frecuencia relativa convergente, a partir de una sucesión finita de observaciones[44]. Sus intentos de determinar unívocamente el valor de θ a partir de un argumento matemático como el de los "*puntos de condensación"* en una sucesión finita de observaciones, terminaron en fracaso. De acuerdo con T.C. Fry (1928)[45], podría haber demostrado que la aparente convergencia de las frecuencias relativas era compatible con (y, aún, causada por) la aleatoriedad de las observaciones. Pero, obviamente, esta posibilidad no comulgaba con la concepción del determinista de Moivre, que consideraba, simplemente, que las series "*debían"* converger. Esta negación del argumento aleatorista lo condujo a un esquema de razonamiento circular que podría ser expresado de la siguiente forma: las frecuencias relativas convergen a θ porque los eventos están regidos por una ley determinada pero, a su vez, la convicción de que los eventos se rigen por una ley determinada, se funda en el postulado de la inversión de la probabilidad.

iv) A pesar del adelanto que significó en el tratamiento de los coeficientes binomiales, la aproximación Normal de de Moivre no fue utilizada en la resolución del problema planteado por N. Bernoulli acerca de la estabilidad de la proporción de nacimientos. Es más, contrariamente a lo que se hubiera

---

[43] Uno de los principales comentaristas a favor de la demostración de de Moivre fue David Hartley (*"Observations on man, his frame, his duty and his expectation"* (1749)). Su asimilación de la inducción a la probabilidad matemática lo convirtió en el principal defensor de las consideraciones de de Moivre sobre las implicancias del teorema de Bernoulli. Al igual que de Moivre, Hartley enfatizó que los resultados que proporcionaba la ley de los grandes números constituían sólo una confirmación de los efectos de la razón y postuló que su teorema representaba una evidencia más de la existencia del*"...Orden y Proporción establecidos por la Divinidad que observan los fenómenos naturales"* (p. 331).

[44] Tampoco indicó un método práctico para obtener un intervalo para θ, en función de los valores de n, Y_n y t.

[45] Ver Fine (1973).

podido esperar, la solución de de Moivre fue recibida con total indiferencia por sus contemporáneos, posiblemente por el carácter eminentemente matemático del trabajo, o porque debido a razones de orden filosófico-teológico imperantes en la época, el autor enfatizó casi exclusivamente los aspectos vinculados a la ley de los grandes números, es decir, a la desaparición de irregularidades en el comportamiento del fenómeno a medida que se incrementaba el número de observaciones,dejando de lado las consecuencias de su notable resultado acerca del significado de $\sqrt{n}$ como unidad de medida del grado de confiabilidad de una estimación[46]. Citando el argumento probabilístico de Arbuthnott y ‘sGravesande sobre la existencia de la Divina Providencia, a partir de la proporción entre nacimientos masculinos y femeninos y de acuerdo con los postulados de la teología natural[47], de Moivre (1718) finalizó su trabajo con la siguiente expresión: “*Se puede concluir, entonces, que si bien el azar produce irregularidades -que pueden ser infinitamente grandes-, con el transcurso del tiempo, la proporción de éstas con respecto a la recurrencia del Orden que resulta naturalmente del 'Diseño Original', no tendrá ninguna importancia (...) De esta forma se demuestra que hay, en la constitución de las cosas, ciertas Leyes de acuerdo a las cuales ocurren los Eventos.Que no es menos evidente a partir de la Observación que esas Leyes sirven a propósitos sabios, útiles y beneficiosos para preservar el Orden del Universo, para propagar las diversas especies de seres y para proporcionar a la Clase conciente tales grados de satisfacción como conviene a su Estado. Pero tales Leyes, así como el Diseño y el Propósito original de su existencia deben ser explicados desde su negación; la ‘inercia’ de la materia y la naturaleza de todos los seres creados hacen imposible que nada pueda modificar su propia*

[46] Las primeras publicaciones sobre la función Normal se produjeron por obra de los matemáticos-astrónomos de fines del siglo XVIII y comienzos del XIX, fundamentalmente Laplace y Gauss, quienes arribaron al mismo resultado en forma independiente y, presumiblemente, sin conocimiento del desarrollo de de Moivre. Puede decirse que el trabajo de de Moivre permaneció ignorado hasta su redescubrimiento por K. Pearson en la biblioteca del University College de Londres, en 1924.

[47] Arbuthnott (1712): “*Entre las innumerables huellas de la Divina Providencia que pueden hallarse en los Trabajos de la Naturaleza, cabe considerar muy especialmente el mantenimiento de un exacto Balance entre el Número de Hombres y Mujeres; esto prueba que la Especie no puede desaparecer nunca, ya que cada Varón puede tener su Mujer, y de una Edad adecuada. Esta igualdad entre Varones y Mujeres no resulta del azar sino de la Divina Providencia trabajando para un buen Fin”* (p. 186).

esencia o darse a sí mismos o a cualquier otro una determinación o una propensión original. Por lo tanto, si no nos dejamos enceguecer por el polvo de la metafísica, seremos conducidos a través de un corto y obvio sendero, al reconocimiento del gran HACEDOR y GOBERNADOR de todo, infinitamente sabio, infinitamente poderoso e infinitamente bueno" (p. 254). Esta hipótesis de "*Orden*" (impuesto por un agente externo, no-natural) a la que se refiere de Moivre -"*...un orden constante, general, magnífico, completo y hermoso*"[48]- sustituyó a la de "*simplicidad de la naturaleza*", postulada por J. Bernoulli[49].Debe tenerse en cuenta que, dado que ni éste ni de Moivre se ocuparon de obtener una confirmación empírica de sus observaciones, sus argumentos teológicos resultan bastante débiles.

5.- John Locke, George Butler, David Hume, George Leclerc-Buffon y las posiciones monista y pluralista en la interpretación de la probabilidad

A modo de reflexión sobre lainterpretación y los alcances de las demostraciones de N. Bernoulli y de de Moivre del teorema central del límite resulta conveniente realizar un breve análisis sobre las posiciones monista y pluralista en la definición de probabilidad.

Del análisis de los textos de los siglos XVII y XVIII surge que para sus autores no existían definiciones rígidas de la probabilidad (cada uno, en una misma obra, empleó indiferentemente las acepcionesclásica, frecuencista y subjetivista), sino distintos métodos de inferencia cuyas características dependían del contexto en el que debían ser utilizados. Este fenómeno se debió a que, curiosamente, en los orígenes de la teoría de la probabilidad el contraste entre las interpretaciones objetivista y subjetivista fue menos profundo que en la filosofía de la época[50].

---

[48] Süssmilch (1741, vol. 1, p. 49).

[49] De acuerdo con Poisson (1836a) un "*Orden*" interpretable solamente en términos de expectativa.

[50] Daston (1988): "*Los filósofos aún se cuestionan cómo la probabilidad pudo significar, a la vez, un grado de creencia y un número de observaciones repetidas. Christian Huygens, Gottfried Leibniz y otros probabilistas del siglo XVII identificaron ambas interpretaciones sin dudar y sin necesidad de ningún otro tipo de justificación*" (p. 126).

Fue el advenimiento de la llamada doctrina de la asociación de ideas –la cual, a partir de la vinculación de la psicología y la epistemología, intentó explicar los procesos psicológicos subyacentes al comportamiento racional- la que proporcionó el fundamento conceptual que posibilitó estas transiciones entre las interpretaciones objetivista y subjetivista[51].

Los fundamentos de este principio de "*filosofía cum psicología*" de la ciencia fueron establecidos por Locke ("*An essay on human understanding*" (1689)) quien asoció las interpretaciones cualitativa y cuantitativa de la evidencia objetiva y las vinculó a una interpretación laicista de la probabilidad (una interpretación casi exclusivamente filosófica, no necesariamente cuantitativa) basada en grados de creencia, generando, de esta forma, una relación del tipo "*experiencia=creencia*". A través de este paralelismo entre las interpretaciones objetivista y subjetivista el asociacionismo logró explicar la relación entre razonabilidad y teoría de la probabilidad postulada por los lógicos de Port Royal y los apologistas ingleses de la Royal Society. La tesis de Locke sostenía que, a partir de la repetida correlación de sensaciones que la mente transformaba en asociación de ideas, la experiencia generaba creencia y, por lo tanto, probabilidad[52]. De modo que, cuanto mayor fuera la frecuencia de la correlación observada, más fuerte resultaría la asociación mental y, por lo tanto, más intenso sería el grado de creencia, la probabilidad y, en consecuencia, la confiabilidad de las generalizaciones inductivas[53].

Además de las inductivas, Locke consideró un tipo de generalizaciones analógicas (sobre cuestiones que se encontraban más allá del alcance directo de los sentidos) las cuales, en base a la *"...conexión gradual de toda esa*

---

[51] Daston (1988): "*Desde la perspectiva del siglo XX la interpretación clásica combina un epistémico 'Ars conjectandi' y una frecuencista 'Doctrine of chances' con una arrogante – o saludable, según sea el punto de vista- omisión de las diferencias filosóficas*" (p. 189).

[52] En otros términos, la psicología asociacionista hizo de la mente un tipo de maquinaria capaz de medir automáticamente frecuencias de eventos pasados y de calcular, en consecuencia, grados de creencia sobre su recurrencia futura.

[53] Locke (1689): "*El testimonio incuestionable y la experiencia, por lo general producen confianza. El testimonio justo y la naturaleza de la cosa indiferente también producen una creencia segura. Cuando la experiencia y los testimonios se oponen, los grados de probabilidad varían en forma infinita. En lo que se refiere a los testimonios tradicionales, cuanto más lejanos, menor será su influencia*" (vol. 4, cap. 14).

variedad de cosas que vemos en el mundo", generaban una probabilidad "*más débil*" que la anterior: *"la probabilidad es la apariencia de un acuerdo sobre demostraciones falibles. Es suplir el deseo de conocimiento. Nos hace suponer que los casos son verdaderos, antes de que sepamos si verdaderamente lo son. Los ámbitos de la probabilidad son dos: la conformidad con nuestra propia experiencia, o el testimonio de la experiencia de los otros. Luego, la probabilidad suple el defecto de nuestro conocimiento y nosguía donde éste falta, siempre sabe acerca de las proposiciones respecto de las cuales no tenemos certeza sino sólo alguna inducción para considerarlas verdaderas. En resumen, sus ámbitos son:* ***i)*** *la conformidad de cualquier cosa con nuestro conocimiento, observación y experiencia y* ***ii)*** *el testimonio de otros que atestiguan sobre su observación y experiencia. En el testimonio de los otros debe considerarse.* ***i)*** *El número;* ***ii)*** *la integridad;* ***iii)*** *la habilidad de los testigos;* ***iv)*** *si el testimonio es extraído de un libro, el designio del autor;* ***v)*** *la consistencia de las partes y circunstancias de la relación y* ***vi)*** *los testimonios opuestos (...) la probabilidad es realidad o especulación"* (vol. 4, cap. 15).

Como se vio en las secciones 2 y 4, Bernoulli no sólo coincidió con Locke y los apologistas ingleses de la teología natural en que se podía aprender a partir de la experiencia sino que, utilizando el principio de la inversión de la probabilidad, planteó la idea de que ese aprendizaje era cuantificable. Posteriormente de Moivre (1730), continuando los trabajos de J. Bernoulli, concluyó que el aprendizaje aumentaba en forma proporcional a la raíz cuadrada del número de observaciones independientes realizadas. Este proceso de inversión de la probabilidad tuvo su culminación en el teorema de Bayes (*"An essay toward solving a problem in the doctrine of chance"* (1764)).

Como continuador de la obra de Locke, Butler (1736) también consideró, en el marco de la interpretación epistémica de la probabilidad, razonamientos inductivos y analógicos expresados en términos frecuencistas. Coincidió con Locke en que la probabilidad difería de la demostración porque admitía grados "*…desde el más alto de la Certeza moral al más bajo de la Presunción*", equivalentes a los "*grados de convicción*", que aumentaban o disminuían con la experiencia del observador, de modo que la agregación de observaciones sobre la ocurrencia de un fenómeno fáctico realizadas "*… por toda la eternidad por toda la humanidad (...) nos da la absoluta seguridad de* *su ocurrencia"*[54].

Por su parte, Hume (1739), si bien coincidió con el planteo de Locke de igualar probabilidad, experiencia y creencia[55], reconoció dos dimensiones netamente diferenciadas en su concepción asociacionista-empirista; una psicológica-escéptica y otra matemática-cuantitativa y logró una asimilación sin precedentes hasta entonces entre probabilidades filosóficas y matemáticas. Siguiendo los postulados de la epistemología empirista, distinguió dos fuentes para la asignación de probabilidades: una basada en la "*chance*" y otra en las causas (donde el concepto de "*chance*" implicaba la

[54] En "*Analogy of religion, natural and revealed, to the constitution of nature*" (en la que, como se mencionó en la Sec. 1, la tradición probabilístico-filosófica fundada por los apologistas ingleses de la Royal Society alcanzó su punto culminante), Butler aplicó la interpretación de la probabilidad basada en un gran número de observaciones del evento "*que mañana amanezca*". Este argumento fue adoptado posteriormente por Richard Price en su apéndice a "*An essay toward solving a problema in the doctrine of chance*" de Bayes: "*Supongamos el caso de una persona recién llegadas a este mundo que deseara comprender, a partir de su observación, el orden y el curso de los eventos que en él ocurren. Probablemente el sol sería el primer objeto que llamaría su atención; pero después de haberlo visto desaparecer la primera noche, se encontraría en una situación de ignorancia total acerca de la posibilidad de verlo nuevamente. Esta persona se encontraría, por lo tanto, en la situación de aquél que está realizando por primera vez un experimento acerca de un evento del cual nada se conoce. Cuando se produzca la segunda aparición, es decir, el primer retorno del sol, surgirá en esta persona una expectativa de un segundo retorno y podría reconocer una probabilidad de 3 a 1 para esta ocurrencia. Como ya se expuso, esta probabilidad aumentaría con el número de retornos que fuera observando. Pero ningún número finito de retornos sería suficiente para generar una certeza absoluta o física. Supóngase que ha visto el retorno del sol a intervalos regulares un millón de veces. Esta circunstancia garantizaría las siguientes conclusiones: existirá una probabilidad de la millonésima potencia de 2 a 1 de que retorne nuevamente al cabo del período habitual. Existirá una probabilidad expresada por 0,5352 de que la probabilidad de esta ocurrencia no fuera mayor que 1.600.000 a 1; y una probabilidad expresada por 0,5105 de que no fuera menor que 1.400.000 a 1*"(p. 242).

[55] Hume (1739): "*Dado que el hábito que genera la asociación surge de la frecuencia de la conjunción de objetos, debe arribar a la perfección en forma gradual y debe adquirir más fuerza con cada caso observado. Con el primer caso no poseerá ninguna fuerza o poseerá una fuerza muy pequeña; el segundo caso observado la incrementará algo; el tercero la hará aún más sensible. Es a través de este lento procedimiento que nuestro juicio ha de arribar a una seguridad total*" (p. 135).

"*negación de la causa*", y la existencia de la causa interrumpía el estado "*neutral natural*" de "*indiferencia perfecta*" entre los resultados posibles, equivalente a la garantía de equiprobabilidad en la interpretación clásica)[56] y a partir de una interpretación eminentemente descriptiva de la probabilidad y con un razonamiento tan escéptico como el que empleó en su tratamiento de la inferencia inductiva, consideró a ambas probabilidades (objetiva y subjetiva) en términos cuantitativos. Por otra parte, en contraste con el pensamiento de Locke, Hume consideró que, si bien la mente procede naturalmente de acuerdo con las probabilidades, no se puede asegurar que esta forma de proceder sea razonable. Rechazó la racionalidad de tales inferencias sobre el futuro en base a una evidencia formada por la suma de repeticiones de eventos idénticos y aceptó la psicología que las hacía inevitables (ver Daston (1988))[57].

Si bien el tratamiento filosófico-psicológico de Hume acerca de la relación entre la frecuencia de ocurrencia de un resultado, el grado de creencia y la medida de la probabilidad (quizás el aporte más importante sobre el tema en el siglo XVIII) es, sin duda, más cuantitativo que el de Locke y el de Butler, no hace ninguna referencia explícita al cálculo de probabilidades. Fue David Hartley ("*Observations on man his frame, his duty and his expectations*" (1749))[58] quien, a partir de la asimilación de las generalizaciones inductiva y

---

[56] Al igual que Hume, en general los autores posteriores interpretaron la noción de "*chance*" como sinónimo de ignorancia del observador. Según Paul Henri Thyri d'Holbach ("*Système de la nature, ou des lois du monde physique et du monde moral*" (1770)) "*El ser humano utiliza la palabra 'chance' para esconder su ignorancia de aquellas causas naturales que producen efectos observables, la cual le impide formar una idea o actuar, en la medida que no puede percibir un ordenamiento dado*" (p. 37). Una interpretación similar asumieron Laplace (quien interpretó a la "*chance*" como una noción epistemológica) y Poincaré (quien, por el contrario, la atribuyó a la inestabilidad y complejidad ded ciertos fenómenos). Ver Galavotti (2005).

[57] Fue Étienne Bonnot de Condillac ("*Essai sur l'origine des connaissance humaines*" (1746), "*Traité des sensations*" (1754)) quien planteó por primera vez reparos acerca de la validez de la relación automática "*frecuencia-creencia*".

[58] Esta obra –que combina elementos del asociacionismo de Locke con las especulaciones psicológicas referidas a las bases vibratorias de la sensación, contenidas en "*Opticks*" (1730) de Newton y con las teorías del Reverendo John Gay (*Preliminary disertation concerning the fundamental principle of virtue and morality*" (1731))- revela una influencia muy fuerte de

analógica de Locke y en base a las consideraciones de de Moivre (1738) sobre las implicaciones del teorema de J. Bernoulli, proporcionó los fundamentos matemáticos al asociacionismo.

Otra etapa importante en este proceso de cuantificación de la vinculación entre creencia, inducción y analogía y probabilidad se debe a Leclerc-Buffon quien, en su "*Essais d´arithmétique morale*" (1777), introdujo consideraciones sobre la probabilidad a partir de una jerarquía de certidumbres que denominó "*matemática*", "*física*" y "*moral*", derivadas respectivamente de argumentos demostrativos, inductivos y analógicos.

Leclerc-Buffon no consideró su propuesta como universalmente aplicable a todos los casos de inferencia de probabilidades, sino como válido solamente en aquellas circunstancias en que las causas del fenómeno en estudio fueran constantes e inaccesibles a la observación directa (debe tenerse en cuenta que la hipótesis fundamental de su razonamiento consistía en suponer que no existían fenómenos genuinamente aleatorios, que aún los "*efectos aleatoriosse regían por causas naturales*(…) *heterogéneas, necesariamente variables y tan versátiles como fuera posible imaginar. Luego, de la noción de azar se puede concluir que evidentemente no existe ningún vínculo entre sus efectos, de modo que el pasado no puede, de ninguna forma, influir sobre el futuro*" (p. 61)).

A partir de la versión de Laplace del teorema de Bayes y fundándose en el principio de uniformidad de la naturaleza, Marie-Jean Nicolas Caritat-Condorcet (introducción a la "*Mémoire sur les probabilités*" (1778) de Laplace), en su intento por cuantificar la influencia de la experiencia histórica sobre el grado de creencia, desarrolló una filosofía de la relación entre el método científico y dicho grado de creencia, basada en la probabilidad inversa. Reconoció a la experiencia histórica como "*la única regla de nuestras opiniones y de nuestras acciones*" y admitió la posibilidad de calcular, en ciertos casos, el incremento en la asignación de laprobabilidad asociada a dicha regla a partir de la acumulación de observaciones.

---

los apologistas ingleses de la Royal Society y, en particular, de la ya mencionada "*Analogy of religión, natural and revealed, to the constitution of nature*" (1736) de Butler.

Así como el desarrollo de la teoría de la asociación de ideas vinculó a las interpretaciones objetivista y subjetivista de la probabilidad, su declinación en el siglo XIX profundizó las diferencias entre las mismas hasta tornarlas incompatibles. Es así que, a partir del replanteo en términos de una "*posibilidad objetiva*" que expresaba "*…la existencia de una relación entre las cosas*" y una "*probabilidad subjetiva*" vinculada con "*…nuestra forma de juzgar o sentir, que varía de un individuo a otro*"[59] realizado por Siméon Denis Poisson (1837)[60], Antoine Augustin Cournot[61], Robert Ellis[62] y George Boole[63], la comparación de estas interpretaciones se convirtió en el tema central de las discusiones sobre la noción de probabilidad[64].

6.- Thomas Simpson y la distribución triangular de los errores de observación

De acuerdo a lo comentado en la Sec. 4, tanto Bernoulli como deMoivre (como la mayoría de los miembros de la Royal Society), en su fidelidad a la teología Newtoniana, veían en sus teoremas límite el argumento según el

---

[59] Cournot (1843) denominó "*posibilidad*" a la relación entre la probabilidad cuantitativa y la naturaleza de las cosas.

[60] Si bien Poisson reconoció profundas diferencias entre las probabilidades objetiva y subjetiva, desarrolló gran parte de su trabajo en el ámbito de la teoría de la probabilidad a partir de la interpretación clásica, pero adoptó la interpretación subjetivista en sus desarrollos sobre la inversión de la probabilidad, en los que tomó en consideración muchos de los supuestos epistemológicos enunciados por Bayes, Laplace y Condorcet.

[61] "*Recherches sur les principes mathématiques de la théorie des richesses*" (1838), "*Mémoire sur les applications du calcul des chances á la statistique judiciarie*" (1838), "*Exposition de la théorie des chances et des probabilités*" (1843).

[62]"*On foundations of the theory of probabilities*" (1849).

[63]"*An investigation of the law of thought*" (1854).

[64] Cournot, Ellis y Boole pueden ser considerados los primeros probabilistas que se apartaron definitivamente de la interpretación Bernoulliana. De acuerdo con su apreciación, la probabilidad cuantitativa no constituía una medida del grado de creencia, ni siquiera podía ser considerada como "*…la razón para creer*", sino "*…la medida de la posibilidad física*" (Cournot (1843, p. 81)).

cual la estabilidad de las frecuencias estadísticas demostraba la presencia de la "*Divina Providencia*"[65]. Uno de los autores que traicionó ese principio fue Simpson.Su trabajo se orientó al estudio de las propiedades de los errores de observación en astronomía. En este sentido, propuso una distribución triangular continua, demostró que, de acuerdo con esta distribución, el valor medio estaba menos afectado por errores no-sistemáticos que cualquier observación individual e intentó hallar su distribución en el límite.

Los problemas planteados por la astronomìa condujeron gradualmente al principio de la media aritmética -calculada a partir de observaciones realizadas en igualdad de condiciones- y a distintos métodos de estimación basados en modelos paramétricos, cuya culminación fue el método de los cuadrados mínimos. Pero la teoría de los errores de observación, que constituyó su fundamento, recién surgió con un tratamiento sistemático en el siglo XVIII, en "*Opera miscellanea sive æstimatio errorum in mixta mathesiper variationes partium trianguli plani et sphærici"* (1722) de Roger Côtes. Como continuación de este trabajo pionero, Simpson publicó "*On the advantage of taking the mean of a number of observations in practical astronomy"* (1755) y"*Miscellaneous tracts on some curious, and very interesting subjects in mechanics, physical-astronomy, and speculative mathematics"* (1757).

En estas memorias demostró en primer lugar que, supuesta una hipótesis acerca del carácter probabilístico de las observaciones sobre un fenómeno Y, el valor promedio de los valores observados era superior, como estimador ($\hat{Y}$) del verdadero valor de$Y$, a cualquier observación individual: "*Al tomar el valor promedio de un número de observaciones, disminuye en forma apreciable la probabilidad de cometer errores pequeños y prácticamente hace desaparecer toda probabilidad de cometer errores de gran magnitud. Esta última consideración, por sí misma, parece suficiente para hacer recomendable el uso de este método, no sólo a los astrónomos, sino a todos los que estén relacionados con cualquier tipo de información obtenida experimentalmente. Y cuanto mayor sea el número de observaciones realizadas, menor será la propensión a cometer errores, suponiendo que las*

[65]Pearson, K. (1925): "*Los matemáticos ingleses post-Newtonianos experimentaron una mayor influencia de la teología Newtoniana que de su matemática*" (p. 202).

observaciones se realicen en igualdad de condiciones"[66]. Esta versión de la ley de los grandes números de Simpson no aparece sustentada por ninguna demostración teórica.

Como se mencionó más arriba, la novedad conceptual introducida por Simpson consistió en dirigir la atención sobre los errores (es decir, sobre las diferencias entre las observaciones registradas y el "verdadero valor" del fenómeno)[67] -asumiendo una hipótesis específica acerca de su distribución de probabilidades- y no sobre las observaciones en sí mismas, ni sobre los objetos observados[68]. Este planteo, que le permitió -aún sin poseer una teoría de la inferencia- medir el error de estimación (es decir, la incertidumbre de la predicción en términos de una distribución de probabilidades dada), constituyó indudablemente el avance fundamental de su trabajo respecto al de de Moivre.

El principio planteado por Simpson implicaba que, una vez asignada una distribución de probabilidades a los errores de observación, la inversión de la probabilidad debía darse por añadidura, postulando quela distribución de

[66] Simpson (1755, pp. 92-93).

[67] En realidad, fue Galileo Galilei ("*Dialogo sui massimi sistemi del mondo"* (1632)) quien introdujo en la literatura el problema de los errores de observación. Si bien no obtuvo una solución cuantitativa ni analítica, arribó a ciertas conclusiones que interpretaban profundamente la esencia del problema: que dichos errores son inevitables, que se distribuyen simétricamente, que la probabilidad de cometer un error de medición aumenta a medida que disminuye la medida del error y que la mayoría de las observaciones se agrupan alrededor del "verdadero valor". Obsérvese que, en estos postulados, están contenidas las características de la distribución Normal que, con el tiempo, se constituiría en una de las funciones básicas de la teoría de la probabilidad y de los rudimentos de la teoría de la contrastación de hipótesis. Conclusiones similares recién pueden ser halladas en "*Beyträge zum Gebrauche der Mathematik und deren Anwendung"* (1765) de Johann Heinrich Lambert.

[68] Simpson supuso que la distribución de probabilidades de los errores era siempre conocida, aún cuando la *"verdadera posición"* del objeto observado pudiera considerarse desconocida (debe recordarse que Simpson era astrónomo). Esta hipótesis constituyó el elemento fundamental para la justificación de cualquier método referido a la estimación mediante la utilización de conjuntos finitos de observaciones.

probabilidades de Y condicionada por $\hat{Y}$, debía ser tal que $p(Y\ /\ \hat{Y}) \propto p(\hat{Y}\ /\ Y)$[69]. En particular, en su publicación de 1755, Simpson supuso que, dado un conjunto de n observaciones independientes, los errores contenidos en cada una de ellas ($\varepsilon_i = Y_i - \widehat{Y}_i, i = 1,2, \dots, n$) podían asumir los valores:

$$-\varepsilon, -\varepsilon + 1, \dots, -2, -1, 0, 1, 2, \dots, \varepsilon - 1, \varepsilon$$

con probabilidades proporcionales a la distribución:

$$q^{-\varepsilon}, q^{-\varepsilon+1}, \dots, q^{-2}, q^{-1}, q^0, q^1, q^2, \dots, q^{\varepsilon-1}, q^\varepsilon$$

o a la distribución:

$$q^{-\varepsilon}, 2q^{-\varepsilon+1}, \dots, (\varepsilon - 1)q^{-2}, \varepsilon q^{-1}, (\varepsilon + 1)q^0, \varepsilon q^1, (\varepsilon + 1)q^2, \dots, 2q^{\varepsilon-1}, q^\varepsilon$$

(quedando definida una distribución de probabilidades para cada $q > 0$)[70].

Simpson desarrolló solamente el caso para $q = 1$, obteniendo una distribución simétrica de los errores. Si bien la selección de las distribuciones de probabilidades de los errores fue tomada de los trabajos de de Moivre, la forma de su utilización fue totalmente novedosa; adelántandose a Lagrange y a Laplace en más de diez años, desarrolló expresiones para la probabilidad de que el valor medio de los errores del conjunto de observaciones no superara una cantidad dada, y comparó esta probabilidad con las correspondientes a las observaciones individuales.

Para la selección de la primera distribución de probabilidades se basó, en particular, en la solución de de Moivre al problema de la definición de la variable aleatoria que representa la suma de los resultados de n tiradas de

---

[69] Stigler, S. M. (1986): "*Simpson había comprendido que el concepto de distribución de los errores permitía un acceso a la medida de la incertidumbre por la puerta de servicio. Fue Laplace quien, colándose posteriormente por ella, logró abrir la puerta principal (sólo para hallar que la llave de Bayes ya se encontraba en la cerradura)*" (p. 98).

[70] Simpson observó que la suma de los términos de la última sucesión elevados a la potencia t es igual a $q^{-\varepsilon t}(1 + q + q^2 + \dots + q^\varepsilon)^{2t} = \frac{1-q^{\varepsilon t+1}}{q^{\varepsilon t}(1-q)^{2t}}$, lo que le permitió concluir que la segunda distribución era asimilable a la primera.

un dado "clásico" con k caras, numeradas de 1 a k, cuyas probabilidades están dadas por los coeficientes del desarrollo de la expresión:

$$\left(1+q+q^2+q^3+\cdots+q^{k-1}\right)^n=\frac{\left(1-q^k\right)^n}{(1-q)^n}$$

La segunda distribución la obtuvo de sumar dos variables aleatorias que representan errores independientes, cada uno con distribución de probabilidades de la forma:

$$q^{-\varepsilon/2}, q^{-(\varepsilon-1)/2}, \dots, q^0, \dots, q^{(\varepsilon-1)/2}, q^{\varepsilon/2}$$

Simpson era conciente (según lo expresó en "*On the advantage of taking the mean"*) del carácter restrictivo de las hipótesis que había asumido acerca de la distribución de los errores, pero también dejó planteada su convicción sobre la posibilidad de generalización de los resultados obtenidos. En 1757, a partir de algunas observaciones realizadas por Bayes a dicha publicación[71], estableció dos principios que caracterizaron los alcances de dicha generalización y que se constituyeron, a partir de ese momento, en las hipótesis implícitas en todos los trabajos sobre teoría del error: **i)** que las probabilidades de ocurrencia de errores de la misma magnitud, por defecto o por exceso, eran iguales; **ii)** que era posible definir límites determinados entre los cuales se podía suponer que estarían incluidos los errores[72].

A partir de este último postulado y remitiéndose a la primitiva distribución de probabilidades discreta proporcional a:

$$1,2,\dots,(\varepsilon+1),..,1$$

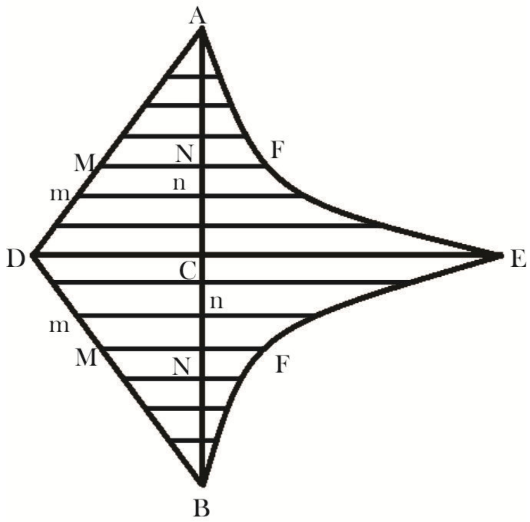
introdujo una distribución triangular de los errores (ABD) (la cual sea, posiblemente, la primera publicación de una distribución de probabilidades continua para los errores).

---

[71] Según se puede deducir de una carta enviada por Bayes a John Canton, fue probablemente "*On the advantage of taking the mean"* la obra que despertó su interés respecto a la teoría de la probabilidad.

[72] Simpson (1757) especificó -arbitrariamente- límites para su distribución de los errores "*...que dependen de las propiedades del instrumento y de la habilidad del observador"* (p. 64).

.



La curva AFEFB representa la función de probabilidades del error que se comete al utilizar el valor medio de los errores. Integrando esta función Simpson obtuvo finalmente que la probabilidad de que el valor medio de n errores no excediera una proporción dada, $1-\frac{x}{n}$, del error máximo posible (CA) estaba dada por:

$$1-\frac{2}{(2n)!}\left[x^{2n}-2n(x-1)^{2n}+\frac{2n}{1}\frac{(2n-1)}{2}(x-2)^{2n}-\right.$$

$$\left.-\frac{2n}{1}\frac{(2n-1)}{2}\frac{(2n-2)}{3}(x-3)^{2n}+\cdots\right]$$

Una circunstancia que colaboró para que Simpson alcanzara el suceso que no lograron ni los Bernoulli ni de Moivre fue, sin duda, que se independizara de los tradicionales esquemas de urnas (en los que, si bien los argumentos matemáticos necesarios eran simples, los aspectos conceptuales inherentes al problema de la inversión de la probabilidad, dadas su estructura física intrínsecamente discreta y la asimetría existente entre la composición de la urna y el mecanismo de extracción, eran muy complicados) y seleccionara otros modelos -por ejemplo, los que se derivaban de las observaciones astronómicas- en los cuales la cuestión conceptual de la inversión de la probabilidad era más fácil de aprehender.

7.- Pierre Simon Laplace y la generalización del teorema de de Moivre

Sea Y una variable aleatoria que representa el número de ocurrencias del resultado A en una sucesión de n eventos binomiales con probabilidades iniciales $p(A) = \theta$ y $p(\bar{A}) = 1 - \theta$ y sea

$$p(Y = y) = \binom{n}{y} \theta^y (1 - \theta)^{n-y} (y = 0,1,2, \dots, n)$$

su distribución de probabilidades. Laplace (1814) demostró que la probabilidad $p[(y - k < Y < y + k)/n\,, \theta]$ es aproximadamente igual a la suma de $2v + 1$ términos del desarrollo del binomio $[\theta + (1 - \theta)]^n$:

$$p[(y - k < Y < y + k)/n\,, \theta] \approx \frac{2}{\sqrt{\pi}} \int_0^v e^{-u^2} du + \frac{\sqrt{n}}{\sqrt{2\pi y(n - y)}} e^{-v^2}$$

Haciendo $y = n\theta + z$ y $v = \frac{k\sqrt{n}}{\sqrt{2y(n-y)}}$, el primer miembro de la expresión precedente asume la forma:

$$p\left(n\theta + z - v\frac{\sqrt{2y(n - y)}}{\sqrt{n}} \leq Y \leq n\theta + z + v\frac{\sqrt{2y(n - y)}}{\sqrt{n}}\right) =$$

$$= p\left(\frac{z}{n} - v\frac{\sqrt{2y(n - y)}}{n\sqrt{n}} \leq \frac{Y}{n - \theta} \leq \frac{z}{n} + v\frac{\sqrt{2y(n - y)}}{n\sqrt{n}}\right)$$

Para valores de n suficientemente grandes, z puede ser considerado despreciable respecto de $n\theta$ y, por lo tanto, $y(n - y) \approx n^2\theta(1 - \theta)$, de modo que se puede escribir[73]:

[73] Esta expresión figura sin demostración en Keynes (1921).

$$p\left(-v\sqrt{\frac{2\theta(1-\theta)}{n}} \leq \frac{Y}{n-\theta} \leq v\sqrt{\frac{2\theta(1-\theta)}{n}}\right) \approx$$
$$\approx \frac{2}{\sqrt{\pi}}\int_0^v e^{-u^2}du + \frac{e^{-v^2}}{\sqrt{2\pi n\theta(1-\theta)}}$$

A partir de este resultado, Laplace concluyó que, denotando por V al $\lim_{n\to\infty} v$ y por n_A al número de resultados A ocurridos al cabo de n repeticiones del evento binomial, teniendo en cuenta que $\frac{V\sqrt{2y(n-y)}}{n\sqrt{n}}$ es de orden $\frac{1}{\sqrt{n}}$, que lo términos de orden $\frac{1}{n}$ pueden ser considerados despreciables en las aproximaciones y sustituyendo a y por n_A y a $n-y$ por $n-n_A$, se puede escribir:

$$p\left(\frac{z}{n} - \frac{V\sqrt{2y(n-y)}}{n\sqrt{n}} \leq \frac{n_A}{n-\theta} \leq \frac{z}{n} + \frac{V\sqrt{2y(n-y)}}{n\sqrt{n}}\right) =$$

$$= p\left(\frac{n_A}{n} - \frac{V\sqrt{2n_A(n-n_A)}}{n\sqrt{n}} \leq \frac{n_A}{n-\theta} \leq \frac{n_A}{n} + \frac{V\sqrt{2n_A(n-n_A)}}{n\sqrt{n}}\right) =$$

$$= \frac{2}{\sqrt{2\pi}}\int_0^V e^{-u^2}du + \frac{\sqrt{n}e^{-V^2}}{\sqrt{2\pi n_A(n-n_A)}}$$

Un resultado que le permitió concluir que, a medida que n crece, la longitud del intervalo disminuye y que la probabilidad que θ asuma valores en dicho intervalo tiende a uno.

8.- Aleksand'r Mijailovich Lyapunov y una nueva generalización del teorema central del límite

En 1891 Pafnuty Lvovich Chebychev y en 1898 Andrei Andreevich Markov publicaron sendas generalizaciones del teorema de Laplace para sucesiones de variables no necesariamente idénticamente distribuidas, utilizando el método de los momentos.

Por su parte, Lyapunov (1901a)(1901b)(1901c) obtuvo una condición suficiente de convergencia a la función Normal para la suma de variables independientes, a partir de sus funciones características: Sea $\{X_1, X_2, \dots, X_n\}$ una sucesión de variables aleatorias independientes con valores esperados m_i, varianzas σ_i^2 y momentos de tercer orden $\mu_{3,i} (i = 1,2, \dots, n)$ finitos y sea la variable $v_{(n)} = \frac{\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} m_i}{\sqrt{\sum_{i=1}^{n} \sigma_i^2}}$. Si se verifica que:

$$\lim_{n\to\infty} \frac{\left[\sum_{i=1}^{n} \int_{-\infty}^{\infty} |X_i - m_i|^3 dF_i(x)\right]^{1/3}}{\sqrt{\sum_{i=1}^{n} \sigma_i^2}} = 0$$

entonces se puede asegurar que la variable $v_{(n)} \xrightarrow{d} N(0,1)$.

J.W. Lindeberg (1920)(1922) transformó la condición suficiente de Lyapunov en necesaria y suficiente: Sea $\{X_1, X_2, \dots, X_n\}$ una sucesión de variables aleatorias independientes con funciones de distribución $F_i(x)$, valores esperados m_i y varianzas $\sigma_i^2 < \infty$ $(i = 1,2, \dots, .n)$ y sea la variable $v_{(n)} = \frac{\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} m_i}{\sqrt{\sum_{i=1}^{n} \sigma_i^2}}$, si y sólo si se verifica que:

$$\lim_{n\to\infty} \left[\frac{1}{\sum_{j=1}^{n} \sigma_j^2} \sum_{i=1}^{n} \int_{|X-m_i|>\varepsilon\sqrt{\sum_{j=1}^{n} \sigma_j^2}} (X - m_i)^2 dF_i(x) \right] = 0$$

es decir, si y sólo si se verifica que[74]:

$$\sum_{i=1}^{n} \int_{|X-m_i|>\varepsilon\sqrt{\sum_{j=1}^{n} \sigma_j^2}} (X - m_i)^2 dF_i(x) = O\left(\sum_{j=1}^{n} \sigma_j^2\right) (\varepsilon > 0)$$

[74] Expresión conocida como la *"versión clásica de la condición de Lindeberg"*.

Se puede asegurar que se cumple la condición $\lim\max_{n\to\infty(1\leq i\leq n)} \frac{\sigma_i}{\sqrt{\sum_{j=1}^{n}\sigma_j^2}} = 0$ y, por lo tanto, que $v_{(n)} \xrightarrow{d} N(0,1)$.

Lyapunov intentó, luego, extender los resultados de su teorema al caso de variables aleatorias no-independientes. Estos intentos culminaron con la generalización de S.N. Bernstein (1926) en la que los alcances del teorema de Lyapunov se amplían a variables aleatorias "*levemente correlacionadas*" y a vectores aleatorios.

9.- Villibald Spečko (William) Feller y las distribuciones infinitamente divisibles

Las demostraciones de Lindeberg dieron origen a una serie de generalizaciones debidas a Feller, cuya inclusión bajo la denominación genérica de teorema central del límite se funda, exclusivamente, en su vinculación con la convergencia en-distribución de sumas de variables aleatorias, en este caso, a funciones de probabilidades distintas de la Normal.

Una expresión típica de estos resultados es la siguiente[75]: Para cada n fijo, sea una sucesión de variables aleatorias independientes, $\{X_{n,j}\}(1\leq j\leq n,\ n\geq 1)$, que cumplen la condición UAN ("*uniformly asymptotically negligibles*") es decir, tales que satisfacen la condición de convergencia uniforme en-probabilidad en j:

$$\max_{1\leq j\leq n} p\left(\left|X_{n,j}\right|\leq\varepsilon\right)\to 1$$

$(\,para\ todo\ \varepsilon>0\,)$ y sea $\{a_n\}(n\geq 1)$ una sucesión de constantes arbitrarias. Entonces, de acuerdo con Feller (1968): **i)**la familia de las funciones límite posibles para las sucesiones $\left\{\sum_{j=1}^{n} X_{n,j}-a_n\right\}\,(n\geq 1)$ es la de las distribuciones infinitamente divisibles, las cuales son tales quesu función característica puede ser expresada de acuerdo con la representación

[75] Para una versión detallada de estos teoremas, ver Feller (1968), Gnedenko; Kolmogorov (1954), Ibragimov; Linnik (1971), Petrov (1974), Loève (1977).

de Lévy-Khinchin de la forma[76]:

$$\ln[C_X(w)] = iw\alpha + \int_{-\infty}^{\infty} \left(e^{iwx} - 1 - \frac{iwx}{1+x^2}\right)\frac{1+x^2}{x^2} dF_X(x)$$

(dondeα denota una constante real) y **ii)** la condición necesaria y suficiente para poder asegurar la convergencia en-distribución de las sucesiones$\left\{\sum_{j=1}^{n} X_{n,j} - a_n\right\}(n \geq 1)$ es que, en todo punto de continuidad de una función de acumulación de probabilidades hasta una constante multiplicativa,$\Psi(x)$, se verifique que:

$$\Psi_n(X) = \sum_{j=1}^{n} \int_{-\infty}^{X} \frac{u^2}{1+u^2} dp\left(X_{n,j} \leq X + a_{n,j}\right) \to \Psi(\mathrm{X})$$

Donde $a_{n,j} = \int_{|X|<\tau} X dp\left(X_{n,j} < X\right)$ $(para\ algún\ \tau > 0)$ y las constantes a_n que reúnen estascondiciones son de la forma $\lim_{n\to\infty} a_n = k_n - k + o(1)$, donde k denota un número real arbitrario tal que:

$$k_n = \sum_{j=1}^{n} \left[a_{n,j} + \int_{-\infty}^{\infty} \frac{X^2}{1+X^2} dp\left(X_{n,j} \leq X + a_{n,j}\right)\right]$$

De la expresión general que figura más arriba se destacan los casos particulares, ya tratados en los apartados anteriores, de convergencia a la distribución Normal y de Poisson: **i)** las condiciones necesarias y suficientes para poder asegurar que las variables $X_{n,j}$ son UAN y que, además, se

---

[76] Las variables infinitamente divisibles fueron introducidas por de Finetti (1930). Su estudio detallado se debe a Lévy (1937). Se dice que una variable X es infinitamente divisible si, para todo $n \in \mathbb{N}$, existe un conjunto de variables aleatorias iid tal que se verifica la siguiente igualdad en-distribución: $X = X_{n,1} + X_{n,2} + \cdots + X_{n,n}$. Es decir, se verifica que $F_X = F_n * F_n * \ldots * F_n = F_n^{n*}$ o, en términos de la función característica, que $C_X = (C_n)^n$. A partir de esta definición se puede concluir que las distribuciones Normal y de Poisson son infinitamente divisibles (ver Chung (1974)). De acuerdo con el teorema de de Finetti (1928b) toda distribución de probabilidades infinitamente divisible es el límite de una distribución de Poisson compuesta (ver Lukacs (1960)).

verifica la convergencia $\sum_{j=1}^{n} X_{n,j} \xrightarrow{d} N(m,\sigma^2)$ son las siguientes:

$$\lim_{n\to\infty}\sum_{j=1}^{n} p\left(\left|X_{n,j}\right| > \varepsilon\right) = 0 \qquad (\varepsilon > 0)$$

$$\lim_{n\to\infty}\sum_{j=1}^{n} \sigma^2\left[X_{n,j} I\left(\left|X_{n,j}\right| \le \tau\right)\right] = \sigma^2 \ (\tau > 0)$$

$$\lim_{n\to\infty}\sum_{j=1}^{n} E\left[X_{n,j} I\left(\left|X_{n,j} \le \tau\right|\right)\right] = m \qquad (\tau > 0)$$

y **ii)** si se verifica que las variables $X_{n,j}$ son UAN, las condiciones necesarias y suficientes para poder asegurar que $\sum_{j=1}^{n} X_{n,j} \xrightarrow{d} \wp(\lambda)$ son las siguientes:

$$\lim_{n\to\infty}\sum_{j=1}^{n} p\left[\left(\left|X_{n,j}\right| > \varepsilon\right) \cap \left(\left|X_{n,j} - 1\right| > \varepsilon\right)\right] = 0 \qquad (0 < \varepsilon < 1)$$

$$\lim_{n\to\infty}\sum_{j=1}^{n} p\left(\left|X_{n,j} - 1\right| \le \varepsilon\right) = 0 \qquad (0 < \varepsilon < 1)$$

$$\lim_{n\to\infty}\sum_{j=1}^{n} \sigma^2\left[X_{n,j} I\left(\left|X_{n,j}\right| \le \tau\right)\right] = 0 \qquad (0 < \tau < 1)$$

$$\lim_{n\to\infty}\sum_{j=1}^{n} E\left[X_{n,j} I\left(\left|X_{n,j}\right| \le \tau\right)\right] = 0 \qquad (0 < \tau < 1)$$

En particular, B. Mandelbrot (1960)(1967) analizó una variable del tipo $\frac{1}{a_n}\sum_{j=1}^{n} X_{n,j} - b_n$ (donde $\left\{X_{n,j}\right\}$ $(1 \le j \le n, n \ge 1)$ denota una sucesión de variables aleatorias iid y a_n y b_n son constantes) para la cual se verifica una convergencia en-distribución a una subclase de la familia de las funciones infinitamente divisibles llamada de las "*distribuciones estables*"[77]. En el caso

[77] Dada una sucesión $\{X_1, X_2, \dots, X_n, \dots\}$ de variables iid con momentos naturales solamente de primer orden, su distribución límite converge a una función estable.

más general en el que las variables $X_{n,j}$ sean sólo independientes, se verifica la convergencia a una subclase de distribuciones denominadas *"auto-descomponibles"*. Si bien la función de densidad de estas variables no puede ser expresada en forma analítica, se demuestra[78] que pueden ser definidas en forma estricta a partir de una transformación bilateral de Laplace de la forma:

$$M_X(w) = \int_{-\infty}^{\infty} e^{-wx} |d[1 - F_X(x)]| = \int_{-\infty}^{\infty} e^{-wx} f_X(x) dx =$$

$$= e^{(w\varphi)^{\alpha}} + \delta w \qquad (\varphi, w > 0)$$

(donde φ denota un coeficiente de escala, α es el exponente característico de la distribución y $\delta = E(X)$). A partir de esta expresión se obtiene que[79]:

$$\lim_{x \to \infty} [1 - F_X(x)] \to x^{-\alpha} [\varphi \Gamma(1 - \alpha)]^{\alpha}$$

10.- Thomas Bayes, Richard Price y el teorema de la probabilidad de las causas

Los fracasos de los Bernoulli y de Moivre en la resolución del problema de la inversión de la probabilidad se debieron fundamentalmente, en el ámbito de la interpretación determinística, a su consideración de la probabilidad θ como una constante (de valor desconocido). El primer intento riguroso de solución se debe a Bayes ("*An essay toward solving a problema iin the doctrine of chances*" (1764)) quien consideró a θ como una variable aleatoria con una distribución de probabilidades "a priori" conocida, a partir de la cual era posible la caracterización de las propiedades y la definición de la distribución de probabilidades de la variable condicionada $(\theta \,/\, Y_n)$ a partir de un conjunto finito de observaciones.

---

[78] Ver Lévy (1925).

[79] Ver Feller (1968)(1971). De acuerdo con Nelson (1987) y en virtud de las generalizaciones comentadas en las secciones anteriores, el teorema central del límite debería ser denominado como "*teorema de de Moivre-Laplace-Lindeberg-Feller-Wiener-Lévy-Doob-Erdös-Kac-Donsker-Prokhorov*".

El "*Essay*" está dividido en dos secciones precedidas por una carta de presentación de su amigo Richard Price, en la que se encuentra el enunciado del problema propuesto por Bayes: "*Dados el número de veces que un resultado eventual ha ocurrido y el número de veces que no ha ocurrido en las mismas circunstancias, y denotando por θ a la probabilidad -desconocida para el observador- de ocurrencia de dicho resultado en una prueba individual, se desea hallar un método por el cual pudiéramos obteneralguna conclusión con respecto a dicha probabilidad*"(p.298)[80].

Antes de tratar la solución del problema Bayes presenta (*Sección I*) un desarrollo axiomático en el que figura su interpretación del concepto de independencia estocástica ("*Los eventos son independientes cuando la ocurrencia de uno de ellos no incrementa ni disminuye la probabilidad del resto*"(p.298)) [81] y una definición que, en un evidente retorno a la interpretación Huygeniana en términos de expectativa, asimila la probabilidad al valor subjetivo coherente a pagar en un juego suponiendo una ganancia unitaria[82], pero entendiendo a la subjetividad como una

[80] A pesar de la afirmación de Price en contrario -como su título parece sugerir- es posible suponer como antecedentes de este trabajo el "*Essai d'analyse sur les jeux de hazard*" (1713) de de Montmort, el "*Ars conjectandi*" (1713) de J. Bernoulli, "*The doctrine of chances*" (1718) de de Moivre y "*The nature and laws of chance*" (1740) de Simpson. Los números de página que figuran en las referencias corresponden a la reedición del "*Essay*" en Biometrika, vol. 45, 1958.

[81] De acuerdo con esta interpretación se podría concluir que, para Bayes, el concepto de independencia entre los eventos es equivalente al de independencia estocástica de orden 2.

[82] "*La probabilidad de un evento es igual al cociente entre el valor de una expectativa que debe ser calculada en función de la ocurrencia de dicho evento, y el valor de la 'cosa' esperada, en el supuesto de que haya ocurrido*" (p.298). Una definición que permite considerar a Bayes como el último representante, al menos hasta la publicación de los trabajos de de Finetti, de la interpretación de la teoría matemática de la probabilidad en términos de expectativa.

subjetividad racional (no como un grado de creencia personal)[83].

En la presentación del "*Essay*" Price interpreta el planteo de Bayes postulando que "*...en la constitución de las cosas existen leyes fijas que rigen la ocurrencia de los eventos, lo que permite concluir que el comportamiento estructural del mundo es el efecto de la sabiduría y el poder de una causa inteligente y, en consecuencia, permite confirmar el argumento sobre la existencia de la Divinidad a partir de las causas finales (...) El problema inverso resuelto en este trabajo es aplicable directamente a este propósito, demuestra con claridad y precisión para cualquier orden o recurrencia de eventos, que existen razones para suponer que tal orden o recurrencia deriva de causas o regulaciones estables de la naturaleza y no de las irregularidades del azar*" (p.297). Una proposición en la que el supuesto "*orden o recurrencia de los eventos*" sugiere una interpretación en términos de la existencia de una tendencia de la aleatoriedad a generar frecuencias estadísticamente estables y una consideración de las "*causas o regulaciones estables*" como asimilables a propensiones físicas inherentes a la aleatoriedad del sistema y, por lo tanto, generadoras también de frecuencias estables en sucesiones infinitas de pruebas repetidas en igualdad de condiciones. Esta propuesta permite concluir que Price coincide con de Moivre en su interpretación del teorema de Bernoulli, pero no acepta el supuesto implícito en su demostración respecto de la posibilidad de estimación de una causa (θ) a partir de un conjunto finito de observaciones y, por lo tanto, rechaza su versión inversa de dicho teorema.

La axiomática Bayesiana dio origen a ciertos juicios adversos referidos fundamentalmente a su posición conceptual confusa e indefinida respecto de la noción de probabilidad, debida a la utilización del concepto de subjetividad racional en un contexto objetivista. A este respecto, es necesario puntualizar que, como se mencionó en la Sec. 5, para los probabilistas de los siglos XVII y XVIII no existían definiciones rígidas de la probabilidad, sino distintos métodos de inferencia para estimar su valor.

---

[83] En la *Proposición 1* establece, posiblemente por primera vez, una regla de adición de probabilidades y en la *Proposición 2* incluye una expresión formal del concepto de equidad de un juego: "*Si una persona tiene una expectativa sobre la ocurrencia de un evento, entonces la probabilidad de ocurrencia del mismo (p) será a la probabilidad de su no-ocurrencia (q) como su pérdida (b) si el evento no ocurre, será a su ganancia (a) si ocurre*" (p.298).

En particular, con respecto a la doble interpretación frecuencista-propensionalista, debe tenerse en cuenta que, en sus orígenes, las propensiones estuvieron asociadas a condiciones repetibles capaces de generar, a partir de series suficientemente largas (pero finitas), frecuencias que podían ser consideradas como aproximaciones a las probabilidades. Lo cual permite concluir, en principio, que la proposición de Bayes-Price podría ser considerada como precursora de la teoría de las repeticiones de Popper y que lo que le impidió acceder a una definición aceptable de probabilidad objetiva no fue la (inexistente) ambigüedad en la noción de probabilidad, sino las dificultades de la teoría propensionalista para asimilar un evento individual a un elemento particular de un colectivo.

La interpretación propensionalista implica un nuevo modelo objetivista, introducido por K. Popper (1934), desarrollado en sus trabajos de 1957, 1959, 1983 y 1990 y continuado por un grupo de filósofos de la ciencia entre los que cabe destacar a D.W. Miller, J.H. Fetzer, I. Hacking y D.H. Mellor. El problema que dio origen a esta teoría consistía en decidir acerca de la posibilidad de identificar probabilidades objetivas "*singulares*" sobre la ocurrencia de eventos individuales. En principio, Popper consideró a un evento individual como un elemento particular de un colectivo de von Mises y sugirió que su "*probabilidad singular*" podía ser asimilada a su probabilidad en el colectivo considerado como una totalidadpero, luego, en sus trabajos de 1957 y 1959 abandonó esa interpretación frecuencista.

Obsérvese que la palabra propensión sugiere un tipo de enumeración ordenada, lo cual marca una diferencia con el punto de vista frecuencista que, como se vio, se caracteriza porque las probabilidades sólo pueden ser introducidas en situaciones físicas (es decir, en "*ocasiones de manifestación total*", de acuerdo con la nomenclatura de Pierce (1910)) respecto de las cuales es posible definir un colectivo. En el modelo propensionalista de Popper es absolutamente legítimo postular la existencia de probabilidades sobre conjuntos de condiciones, aunque éstas no admitan un número suficientemente grande de repeticiones, lo cual implica una ampliación indiscutiblemente significativa del conjunto de situaciones en el que resulta aplicable la teoría de la probabilidad, con respecto a la interpretación fecuencista. Las probabilidades de eventos individuales deben ser consideradas como dependientes fundamentalmente del conjunto de condiciones a las cuales está referido el evento más que del evento mismo[84].

La propuesta de Popper fue modificada por Miller (1994)(1996) quien, en su intención de resolver el problema fundacional de la identificación de probabilidades singulares, desvinculó a las propensiones de las condiciones repetibles y propuso una asociación con los estados del universo, transformando al propensionalismo de una teoría científica en una especulación metafísica.A fin de evitar esta insuficiencia del modelo, a diferencia de la propuesta original de Popper y de las modificaciones de Miller, Fetzer (1982)(1988)(1993) abandonó la idea de asociar las propensiones a un estado completo del universo y propuso vincularlas a un conjunto completo de condiciones relevantes, de modo que para "*falsar*" un valor "*conjeturado*" de una propensión se debían plantear "*conjeturas*" acerca de un conjunto de dichas condiciones relevantes. Ahora bien, dadas las dificultades insalvables que acarrea la formulación y, en consecuencia, la "*faslsación*" de las conjeturas necesarias, se puede concluir que en el modelo de Fetzer las propensiones también observan un carácter más metafísico que científico y que, por lo tanto, su concepto de probabilidad no puede ser extendido en forma general a los casos singulares para los cuales la asignación de una probabilidad de ocurrencia continúa siendo subjetiva.

Como consecuencia de estas restricciones respecto de la aproximación a las probabilidades de eventos individuales mediante la definición de clases de referencia, se puede concluir en forma inmediata que: **i)** algunas probabilidades pueden ser consideradas preferibles a otras y **ii)**el grado de preferencia respecto de las probabilidades varía en forma directa con la magnitud dela evidencia en la que se basan, pero esta relación de preferencia no implica la existencia de una probabilidad singular objetiva.

Sea un evento singular E que puede ser clasificado como un elemento de una sucesión de condiciones, $S_1, S_2, S_3, \ldots$ tales que $S_1 \subset S_2 \subset S_3 \subset \cdots$. Supóngase que se cuenta con información estadística que permite obtener

---

[84] Este planteo también podría ser considerado como una variante del logicismo dirigida a intentar el retorno a una interpretación objetivista, asociando el concepto de probabilidad al de las denominadas posibilidades potenciales de Fock, según las cuales se puede admitir que el conjunto de resultados posibles existe sólo en la mente del observador, pero no como una creación de éste, sino como algo que el observador debe admitir (ver Omelyanoskij; Fock (1972)).

buenas estimaciones $(p_1, p_2, p_3, \ldots)$ de la probabilidad objetiva sobre la ocurrencia de E con respecto a las condiciones $S_1, S_2, S_3, \ldots$. Entonces, de acuerdo con las consideraciones antertiores, se demuestra que la probabilidad p_2 es preferible a p_1, que p_3 es preferible a p_2 y así sucesivamente. En particular, si se sustituyen las condiciones S_i por la clase de referencia del conjunto de elementos, S, se puede obtener una estimación dela probabilidad asociada a la clase de referencia más restringida.Un primer problema vinculado a las posibilidades identificatorias del modelo propensionalista derivado de la aplicación de este principio, se presenta obviamente cuando no existe una única clase de referencia de máxima restricción seleccionable. Pero debe tenerse en cuenta que, aún en el caso en que dicha clase existiera, la adopción del criterio consistente en asimilar la probabilidad sobre la ocurrencia de un evento singular a su frecuencia relativa en la clase de referencia más restringida a la cual pertenece el evento, podría conducir a una decisión equivocada. Podría suceder que se conocieran circunstancias que no constituyeran datos estadísticos en una clase de referencia pero que, sin embargo, proporcionaran razones considerables para corregir la asignación de probabilidades. En ese caso, la no consideración de dicha evidencia cualitativa puede conducir a asignaciones de probabilidades con una base menos satisfactoria que la que se podría haber obtenido a partir de un análisis total de la evidencia. El procedimiento general para asignar probabilidades a eventos singulares debería entonces: **i)**asignar el evento a la clase de referencia más restringida para la cual existen datos estadísticos confiables (suponiendo que exista una clase de referencia con esas características) y calcular la frecuencia reltiva (r) de la ocurrencia del evento en dicha clase y **ii)** tomar enconsideración cualquier información de características no-estadísticas que sea relevante para laocurrencia del evento en la circunstancia en cuestión y, a laluz de esta información, corregirla frecuencia relativa. En el caso de que existiera más de una clase de referencia de restricción máxima con frecuencias relativas $r_1, r_2, r_3, \ldots$, se debería seleccionar una frecuencia relativa y corregirla utilizando la información no-estadística. Viceversa, si no existiera ninguna clase de referencia aceptable, la asignación de la probabilidad debería basarse exclusivamente en la información no-estadística.

Si bien este método de asignación de probabilidades parece razonable, esinnegable que incluye muchos elementos subjetivos y que, en consecuencia, no resulta adecuado como modelo identificatorio de una probabilidad objetiva singular, en particular en aquellos casos en los cuales no se cuenta con información estadística obtenida de sucesiones suficientemente largas de observaciones. En el caso en que se contara con una sucesión de observaciones y no existieran circunstancias ajenas a los datos estadísticos, el modelo para identificar la probabilidad (teórica) sobre la ocurrencia de un evento a partir de la frecuencia relativa (observable) será, de acuerdo con la nomenclatura de Popper (1934), "'*impermeable' a la falsación estricta*". Un vano intento de solución propuesto por Popper a esta dificultad consistió en apelar a lanoción de "*falsación metodológica*" según la cual, aunque las proposiciones sobre probabilidades no sean estrictamente "*falsables*", pueden ser utilizadas (y de hecho lo son en las ciencias experimentales) como argumentos "*falsables*" mediante la utilización de tests estadísticos. Debe tenerse en cuenta que, de acuerdo con este procedimiento, cualquier hipótesis puede ser refutada métodológicamente aún cuando desde un punto de vista estrictamente lógico no sea refutable y, por lo tanto, que este criterio de "*falsación*" no resuelve el problema de la identificación de la probabilidad[85].

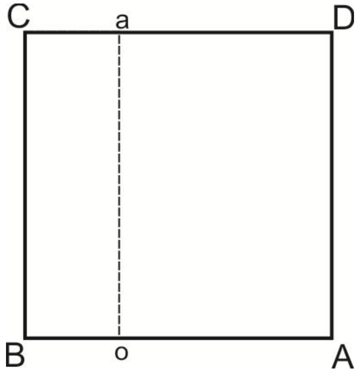
A las definiciones preliminares contenidas en el "*Essay*", mencionadas más arriba sigue una serie de proposiciones entre las que cabe mencionar la *número 1*, que contiene la demostración del principio de aditividad finita para "*eventos inconsistentes*" y la *número 3*, referida "*eventos subsecuentes*", la cual, utilizando notación moderna, puede ser expresada de la siguiente forma: dados dos eventos, E_1 y E_2, ordenados temporalmente, se verifica que $p(E_2 \,/\, E_1) = \frac{p(E_1 \cap E_2)}{p(E_1)}$ (suponiendo que ambos términos de este cociente existan). Esta proposición se continúa en la *Proposición 5* (que puede ser considerada como una versión subjetiva de la *Proposición 3*, establecida en forma independiente de ésta): $p(E_1 \,/\, E_2) = \frac{p(E_1 \cap E_2)}{p(E_2)}$, la cual, teniendo en cuenta que constituye la base de la demostración de la *Proposición 9* (que contiene la solución del problema de la inversión), puede ser considerada el núcleo del "*Essay*"[86].

---

85 Ver Festa (1999), Hacohen (2000), Kuipers (2000), ter Hark (2002)(2004).

86 En la *Proposición 5* Bayes introduce, como un nuevo ingrediente, el orden en el cual se "aprende" de la ocurrencia de los eventos. Es decir, postula la asimilación del proceso de aprendizaje a la condicionalización estocástica.

Luego de una suerte de prólogo en el que Bayes establece los axiomas de la probabilidad y postula el rol de las probabilidades como leyes de la interpretación subjetivista reguladoras de la asignación racional de probabilidades personales, sigue (*Sección II*) la solución del problema presentado en la *Sección I*, estructurada sobre la base de un ingenioso experimento que simula un modelo binomial, consistente en: **i)** (*Postulado 1*) hacer rodar una bola (b_1) sobre un plano horizontal limitado por un cuadrado ($ABCD$) (una mesa en la versión original del "*Essay*", una mesa de billar, según la frívola interpretación debida a K. Pearson y R.A. Fisher, sin duda muy apartada del austero espíritu del Reverendo Bayes), de modo que la probabilidad de que la misma se detenga en un punto particular, sea la misma para todos los puntos del plano; **ii)** (*Postulado 2*) realizar una serie de ensayos consistente en hacer rodar n veces una segunda bola (b_2), tomando en consideración la frecuencia (X_n) del evento M: "que b_2 se detenga a la derecha de b_1"[87].



Supóngase (y esto, obviamente, no significa una pérdida de generalidad) que el área del cuadrado $ABCD$ sea igual a uno y que el punto A tenga coordenadas $(0,0)$ y sea o la abcisa del punto del plano en el que se detuvo b_1. De acuerdo con estas hipótesis, Bayes postula que la variable que representa el punto en que ha de detenerse cada bola, está uniformemente distribuida en el cuadrado $ABCD$, y que o está uniformemente distribuida en el intervalo $(0,1)$. Entonces, para todos los valores de o tales que $0 < o_1 < o < o_2 < 1$, se verificará que:

[87] Este mecanismo -que, más allá de una insoslayable falta de rigurosidad en términos del estado actual del arte, describe un esquemático procedimiento de Monte Carlo- sustituyó, en la solución de Bayes, al esquema de urnas de Bernoulli.

i) *Lema 1*: $p(o_1 < o < o_2) = o_2 - o_1$

ii) *Proposición 8*:

$$p[(o_1 < o < o_2) \cap (X_n = n')] = \int_{o_1}^{o_2} \binom{n}{n'} o^{n'} (1-o)^{n-n'} do$$

iii) *Corolario:*

$$p(X_n = n') = \int_0^1 \binom{n}{n'} o^{n'} (1-o)^{n-n'} do = \frac{1}{n+1}$$

(independiente de n') y, finalmente, que: **iv)** *Proposición 9*[88]*:*

$$p[(o_1 < o < o_2)/(X_n = n')] = \frac{\int_{o_1}^{o_2} \binom{n}{n'} o^{n'} (1-o)^{n-n'} do}{\int_0^1 \binom{n}{n'} o^{n'} (1-o)^{n-n'} do} =$$

$$= \frac{(n+1)!}{n!\,(n-n')!} \int_{o_1}^{o_2} o^{n'} (1-o)^{n-n'} do$$

Una de las críticas que figura habitualmente en la literatura se refiere a la adopción por parte de Bayes de una definición no suficientemente estricta del concepto de "evento" y, consecuentemente, la posibilidad de generar una confusión con el concepto de proposición. Pero, dado que el álgebra de los eventos es isomórfica con el álgebra de las proposiciones, se puede concluir que su expresión de los resultados a veces en términos de la probabilidad de la proposición "*que el punto o esté comprendido en un intervalo dado*" (*Proposición 8*) y a veces en términos del evento "*que la probabilidad θ de ocurrencia del evento M esté comprendida en un intervalo dado*" es indiferente.

---

[88] Uno de los resultados más elegantes y profundos de la teoría de la probabilidad.

Con respecto a la condición de uniformidad física y epistemológica de los resultados (típica de la interpretación clásica) propuesta por Bayes[89], parece lógico suponer que, en principio, se debe a la influencia del entorno filosófico de la época, fundamentalmente del postulado de Hume (1739) acerca de la necesidad (metafísica) de la "*absoluta*", "*perfecta*" y "*total indiferencia*" entre los distintos resultados posibles de un fenómeno para poder asegurar su comportamiento aleatorio y el corolario (matemático) obligado que implicó la asimilación de la aleatoriedad, es decir de la ausencia de regularidad en el comportamiento de un fenómeno, a la equiprobabilidad en la distribución de dichos resultados[90].

Debe destacarse que si bien esta tan controvertida condición de uniformidad, habitualmente propuesta en términos abstractos, introdujo confusión en los aspectos metodológicos de la teoría cuantitativa de la probabilidad, fue fundamental en su génesis y su desarrollo matemático posterior y que, a su vez la teoría matemática de la probabilidad determinó cambios fundamentales en la forma de utilización de los argumentos relacionados con dicha condición que significaron pasar de una aplicación de la uniformación como argumento para la aleatorización a partir de una hipotética pérdida de información a una aleatorización en la cual la incorporación de desorden generaba, paradójicamente, un incremento de la información.

Inmediatamente después de la *Proposición 9*, Bayes agregó un "*Scholium"* en el que intentó generalizar el resultado anterior justificando (*Proposición 10*) la analogía entre el mecanismo físico adoptado en su demostración y el comportamiento que manifiestan para el observador ciertos fenómenos de la naturaleza[91]. Postuló que el mismo razonamiento utilizado para resolver el

---

[89] En realidad, la condición de uniformidad o equiprobabilidad asimilada a la expresión "*en las mismas circunstancias*" -que, según Price, fue mencionada por Bayes en su introducción- no figura en el "*Essay*" pero, a partir del análisis de los teoremas en él contenidos, se puede concluir que se encuentra implícita en el espíritu de su solución.

[90] Ver Sec. 2.

[91] A continuación de la *Proposición 10* figuran tres *reglas* y un *Apéndice* ("*Containing an application of the foregoing rules to some particular cases*") referidos a las aplicaciones y al cálculo de las integrales que figuran en las proposiciones precedentes. La *Regla 1* se debe a Bayes, las otras dos y el *Apéndice* se deben a Price.

problema de la mesa de billar podía ser aplicado para cualquier evento cuyos resultados pudieran ser considerados simétricos "a priori" de la realización de cualquier prueba, obteniendo el siguiente resultado[92]:

$$p[(\theta_1 < \theta < \theta_2)/M^*] = \frac{\int_{\theta_1}^{\theta_2} \binom{n}{n'} \theta^{n'}(1-\theta)^{n-n'} dF(\theta)}{\int_0^1 \binom{n}{n'} \theta^{n'}(1-\theta)^{n-n'} dF(\theta)}$$

(donde M^* denota el supuesto "*el resultado E ha ocurrido* n'*veces en* n *ensayos realizados en igualdad de condiciones*").

Este planteo, basado en el ya mencionado principio de indiferencia total de Hume según el cual, a partir de unacaracterización operacionalista de un evento cuyos resultados fueran simétricos, era posible justificar el supuesto de equiprobabilidad "a priori" a partir de la condición "*...no conocemos absolutamente nada previamente a la realización de alguna observación*", condujo a Bayes a postular una distribución $dF(\theta) = d\theta$ uniforme de la variable θ.

La operacionalización postulada por Bayes queda evidenciada en el siguiente pasaje del Scholium: "*A partir de la proposición precedente queda claro que, en el caso de un evento como el que denoto por* M*, conociendo el número de veces que ocurrió y el número de veces que no ocurrió en un cierto número de ensayos repetidos, sin saber nada más acerca de él, se puede conjeturar dónde se encuentra su probabilidad y, mediante métodos usuales de cálculo de las medidas de las áreas allí mencionadas, observar la probabilidad de que dicha conjetura sea correcta. Y el postulado de que la misma regla es apropiada para ser utilizada en el caso de un evento acerca de cuya probabilidad no conocemos absolutamente nada 'a priori' de cualquier ensayo, parece*

---

[92] Según Price, Bayes tuvo dudas acerca de la validez de este postulado de generalización de su teorema (Price: "*...pero Bayes posteriormente consideró que el postulado al cual había hecho referencia podía no ser considerado unánimemente razonable y, por lo tanto, prefirió modificar la forma de expresión de la proposición que suponía que contenía la solución del problema y agregar un scholium final que explicara su forma de pensar antes que incluir en su razonamiento matemático algo que pudiera generar controversias*" (p.297)). No obstante, parece más creíble la interpretación de Dale (1991) de que Bayes resolvió en primer término el problema general para resultados simétricos y luego construyó el modelo de la mesa de billar.

surgir de la siguiente consideración: que en lo que respecta a tal evento, no tengo ninguna razón para pensar que, en un cierto número de ensayos, pueda ocurrir un número dado de veces más que otro. Pero, este es exactamente el caso del evento M. Antes de arrojar la bola W, la cual determina su probabilidad en una prueba individual, la probabilidad de que ocurra p veces y no ocurra q veces, en $p+q=n$ ensayos, está dada por el cociente entre AiB y CA, el cual es el mismo para todo p, cuando $p+q=n$ está dado. (...) Y, consecuentemente, antes de definir el punto $'o'$, o de saber el número de veces que el evento M ha ocurrido en n ensayos, no tengo ninguna razón para pensar que pueda ocurrir un número de veces más que otro. Por lo tanto, en lo que sigue, daré por descontado que la regla dada para el evento M en la Proposición 9, es también la regla a ser utilizada con relación a cualquier otro evento acerca de cuya probabilidad no sepamos nada 'a priori' de cualquier tipo de prueba. Y a tal evento lo denominaré 'evento desconocido'" (p.301).

A partir de esta conjetura Bayes obtuvo el corolario de la *Proposición 8* (conocido como el *postulado de Bayes*), según el cual la distribución de probabilidades marginal de X_n queda definida de la siguiente forma:

$$p(X_n = n') = \int_0^1 \binom{n}{n'} \theta^{n'}(1-\theta)^{n-n'} d\theta = \frac{1}{n+1} (\forall n')$$

Un resultado cuya independencia de n' fue considerada por Bayes como la justificación de la hipótesis "a priori" de la distribución uniforme de θ. Y que desmiente la consideración habitual de que la analogía propuesta por Bayes entre la mesa de billar y dichas aplicaciones se refiere exclusivamente al comportamiento "a priori" de θ.

Para que esta operacionalización supere su condición de simple conjetura intuitivamente aceptable, se requiere la demostración rigurosa de que la distribución uniforme de θ no sólo es condición necesaria, sino también suficiente para el cumplimiento del postulado de Bayes. Como corolario del teorema de representación de de Finetti (1937)[93], haciendo $n = n'$ en la expresión anterior y, dada una función de distribución $F(\theta)$, se obtiene el

[93] Ver Sec. 22.

momento de orden n de la función de densidad $dF(\theta)$, $\int_0^1 \theta^n dF(\theta) = \frac{1}{n+1}$. Lo cual permite concluir que el postulado de Bayes determina en forma unívoca la sucesión de infinitos momentos de $dF(\theta)$. Por otra parte, dado que la función $dF(\theta)$ está concentrada en un conjunto compacto, de acuerdo con el teorema de Hausdorff, se puede asegurar que está definida en forma estricta por la sucesión de sus momentos y, de acuerdo con el teorema de Murray (1930), se demuestra que la única función de densidad que satisface la sucesión de momentos que prescribe el postulado de Bayes debe ser tal que $dF(\theta) = d\theta$. Es decir que el postulado de Bayes se verifica si, y sólo si la variable θ, condicionada por la distribución binomial de la variable X_n, se distribuye uniformemente. Lo cual conduce a la conclusión que el teorema de Bayes no logró evitar el carácter metafísico de los supuestos que constituyen sus fundamentos ni, en consecuencia, obtener una solución general al problema de la identificación de la función de densidad de θ.

11.- Nuevamente Laplace y la culminación de los trabajos de Bayes y Simpson

Si bien, al igual que de Moivre y Price -y más tarde, Condorcet y Poisson-, Laplace ("*Le sublime géomètre*"[94]) intentó tratar formalmente la intuiciones metodológicas de los filósofos naturalistas mediante la definición de un conjunto de reglas matemáticas dirigidas a discutir las objeciones de los escépticos de la inducción, se diferenció de ellos en que no utilizó a la probabilidad de las causas como fundamento del principio del "*Diseño Original*"[95]. Coincidió con de Moivre y Price en que el "*orden natural*" era

[94] Gouraud (1848, p. 64).

[95] La aparición de la monumental figura de Laplace (a quien, según Todhunter (1865): "*La teoría de la probabilidad le debe más a él que a ningún otro matemático*" (art. 868)) hace oportuno definir un corte transversal en el desarrollo de este proceso de la inversión de la probabilidad, a fin de realizar un análisis estático comparativo de la situación de sus principales personajes: al momento del nacimiento de Laplace, Jakob y Johann Bernoulli, B. Pascal y P. de Fermat ya habían fallecido, de Moivre tenía 81 años, Nikolaus Bernoulli 62, Daniel Bernoulli 49, Th. Bayes 47 y Th. Simpson 39. Faltaban 28 años para el nacimiento de Gauss, 32 años para el nacimiento de Poisson y 158 años para la fundación del glorioso club San Lorenzo de Almagro.

estable, pero no en que "*causas estables*" debían producir necesariamente "*efectos estables*". Consideró una naturaleza compuesta por "*causas regulares*" (o "*constantes*") y "*causas irregulares*", postulando que estas últimas observaban un comportamiento de conjunto "*regular*", cuyos efectos simétricos en el largo plazo se anulaban. El resultado fue un nuevo modelo de causación "*...en el que era posible concebir un mundo en el cual el orden macroscópico era producido por un caos microscópico*"[96] en el que la "*composición de la urna*" podía variar sin por eso destruir la regularidad global de los efectos observados(es a partir de este modelo de causación que el teorema de Bernoulli se convierte en la ley de los grandes números).

Los orígenes de esta interpretación del proceso de causación como una interacción de causas constantes uniformes y perturbaciones simétricas, se hallan en la "*Mémoire sur la probabilité des causes par les èvénements*" (1774b), en la que Laplace realizó su primera exposición de la teoría del error en las observaciones astronómicas[97]. Después de algunas consideraciones introductorias, comenzó la segunda sección de esta memoria asimilando -como J. Bernoulli- la incertidumbre del observador respecto a la vinculación entre causas y efectos a una urna que contenía bolillas de varios colores, y estableciendo que todo problema en el ámbito de la "*teoría del azar*" pertenecía a una de las dos clases siguientes: **i)** aquélla en que el resultado del fenómeno que se analizaba era eventual, pero

---

[96] Daston, L. (1988, p. 267).

[97] La obra de Laplace sobre probabilidades comienza en 1774a con la "*Mémoire sur les suites récurro-récurrentes et sur leurs usages dans le théorie des hasards*", y abarca en su totalidad un período de cuarenta años, que culmina con la tercera edición de la "*Théorie analytique des probabilités*" (1812), su "opus magnum". En la "*Mémoire sur la probabilité des causes par les événements*"(1774b) figura la siguiente definición de probabilidad: "*La probabilidad de un evento es igual a la suma de los productos de cada caso favorable por su probabilidad, dividida por la suma de los productos de cada caso posible por su probabilidad. Si todos los casos fueran igualmente probables, la probabilidad del evento será igual al número de casos favorables, dividido por el número de casos posibles*". Si bien una definición de probabilidad expresada como un cociente entre casos "*favorables*" y "*posibles*", ya figuraba en el "*Liber de ludo aleæ*" de Cardano (1564) y en la "*Doctrine of chances*" de de Moivre (1738), éste puede ser considerado el primer caso de construcción de la definición a partir de la descomposición de un evento "compuesto" en eventos elementales no necesariamente equiprobables.

la causa que condicionaba la asignación de probabilidades sobre su ocurrencia era conocida (probabilidad "*directa"*) y **ii)** aquélla en que el resultado del fenómeno era conocido, pero su causa era desconocida (probabilidad "*indirecta"* o "*inversa"*) y dedicó su atención exclusivamente al estudio de los fenómenos de la segunda clase. Inmediatamente enunció su principio fundamental sobre la probabilidad inversa, interpretándola como una suerte de promedio: "*Si un evento puede ser producido por un número n de causas diferentes, entonces las probabilidades de estas causas, dado el evento, son, con respecto a cada una de las otras causas, como las probabilidades del evento, dadas esas causas, y la probabilidad de existencia de cada una de éstas es igual a la probabilidad del evento, dada esa causa, dividida por la suma de las probabilidades del evento, dada cada una de las causas"* (pp. 384-385)[98].

Utilizando notación moderna, la demostración por la que Laplace arribó a este principio puede ser expresada de la siguiente forma: Representando por E_1 a la ocurrencia de un evento al realizar un primer ensayo, por E_2 a la ocurrencia del mismo evento al realizar un segundo ensayo, entonces la definición de probabilidad condicionada queda expresada como $p(E_2 \,/\, E_1) = \frac{p(E_1 \cap E_2)}{p(E_1)}$. Por otra parte, a partir del teorema de la probabilidad total, denotando por $\{A_i, i = 1,2, \dots, n\}$ al conjunto de todas las causas (mutuamente excluyentes) que condicionan la asignación de probabilidades sobre la ocurrencia de E_1 y E_2, y por E a la ocurrencia del evento en alguno de los ensayos individuales, y suponiendo que los eventos E_1 y E_2 son condicionalmente independientes con respecto a cada causa A_i, concluyó que:

$$p(E_2 \,/\, E_1) = \sum_i p[(E_2 \cap A_i) \,/\, E_1] = \sum_i p(E_2 \,/\, A_i) p(A_i \,/\, E_1) =$$

$$= \sum_i p(E \,/\, A_i) p(A_i \,/\, E)$$

---

[98] Si bien en el Suplemento del "*Essay"* Price había introducido la idea de "*probabilidad de las causas"*, esta "*Mémoire"* de Laplace "*...merece ser destacada en la historia del problema de la inferencia como la primera que enunció en forma explícita el principio para estimar las probabilidades de las causas por las cuales se puede haber producido un evento"* (Todhunter I. (1865, p. 868)).

El supuesto adicional de igualdad de las probabilidades para todas las causas, le permitió a Laplace escribir las siguientes relaciones[99]:

$$p(E_1) = \frac{1}{n}\sum_i p(E_1 / A_i) = \frac{1}{n}\sum_i p(E / A_i)$$

$$p(E_1 \cap E_2) = \frac{1}{n}\sum_i p[(E_1 \cap E_2) / A_i] = \frac{1}{n}\sum_i p(E_1 / A_i)p(E_2 / A_i) =$$

$$= \frac{1}{n}\sum_i p(E / A_i)\, p(E / A_i) = \frac{1}{n}\sum_i [p(E / A_i)]^2$$

De la misma forma, la probabilidad de que el evento ocurra k veces, será:

$$p(E_1 \cap E_2 \cap \ldots \cap E_k) = \frac{1}{n}\sum_i [p(E / A_i)]^k (k = 1,2, \ldots, n)$$

Reemplazando estos resultados en las expresiones anteriores, obtiene que:

$$\sum_i p(E / A_i)\, p(A_i / E) = \frac{\sum_i [p(E / A_i)]^2}{\sum_i p(E / A_i)}$$

Y, generalizando para k ocurrencias de E, resulta que:

99 En 1774a Laplace no indica de qué manera arribó al postulado de inversión de la probabilidad. Solamente menciona que la circunstancia que lo condujo en esa dirección fue su estudio sobre los valores medios. Es en su "*Théorie analytique des probabilités*" (1812) que, a continuación del texto, figura la demostración ordenada. Obsérvese que su aproximación a la solución del problema de la inversión es tan diferente de la desarrollada por Bayes (el elegante modelo geométrico de Bayes conduce inmediatamente al caso general de probabilidades continuas, en tanto que el planteo discreto de Laplace, basado en el esquema de urnas con bolillas rojas y azules en una proporción desconocida, se generaliza al caso continuo sólo en el supuesto de un número infinito de bolillas), que permite suponer que Laplace no conocía el "*Essay*" (téngase en cuenta que, a la fecha de la publicación del "*Essay*", Laplace tenía 15 años).

$$\sum_i [p(E\ /\ A_i)]^k\, p(A_i\ /\ E) = \frac{\sum_i [p(E\ /\ A_i)]^{k+1}}{\sum_i p(E\ /\ A_i)}\,(k = 1,2,\dots,n-1)$$

Es decir, queda definido un sistema de $n-1$ ecuaciones lineales en $p(A_i\ /\ E)$ de la forma:

$$\begin{cases} p(E/A_1)p(A_1/E) + \cdots + p(E/A_n)p(A_n/E) = \dfrac{\sum_i [p(E/A_i)]^2}{\sum_i p(E/A_i)} \\ [p(E/A_1)]^2 p(A_1/E) + \cdots + [p(E/A_n)]^2 p(A_n/E) = \dfrac{\sum_i [p(E/A_i)]^3}{\sum_i p(E/A_i)} \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots. \\ [p(E/A_1)]^{n-1} p(A_1/E) + \cdots + [p(E/A_n)]^{n-1} p(A_n/E) = \dfrac{\sum_i [p(E/A_i)]^n}{\sum_i p(E/A_i)} \end{cases}$$

al que se debe agregar la ecuación:

$$p(A_1\ /\ E) + p(A_2\ /\ E) + \cdots + p(A_n\ /\ E) = 1$$

Como se ve fácilmente, la solución del sistema de n ecuaciones, $p(A_k\ /\ E) = \frac{p(E/A_k)}{\sum_i p(E/A_i)}(k = 1,2,\dots,n)$ constituye la expresión del teorema de Bayes bajo el supuesto de igualdad de las probabilidades "a priori", $p(A_i) = \frac{1}{n}(i = 1,2,\dots,n)$.

Esta restricción debe ser interpretada más como un artificio utilizado por Laplace para facilitar las demostraciones que como la aceptación del principio de la razón insuficiente (de carácter metafísico) según el cual todo lo desconocido debe interpretarse como igualmente probable[100].

[100] Su primera mención explícita acerca del principio de que la ausencia de conocimiento implica una distribución equiprobable se halla en "*Mémoire sur la probabilité des causes par les événements*" (1777).

Posteriormente, en la "*Mémoire sur les probabilités*" (1781) y en la "*Mémoire sur les approximations des formules qui sont fonctions de trés grands nombres*" (1810)[101], Laplace consideró el caso de asignación de probabilidades diferentes a las distintas causas, y propuso -mediante la "*subdivisión*" de las causas más probables- una nueva especificación, de tal forma que finalmente todas resultaran igualmente probables[102].

En el caso de equiprobabilidad, la probabilidad de que una causa dada, x, asuma un valor comprendido en el intervalo $[x_1, x_2]$ está dada por $\frac{\int_{x_1}^{x_2} f(x)dx}{\int_0^1 f(x)dx}$. Donde la función $f(x)$ denota la probabilidad de los efectos observados, cuando x varía. Supóngase ahora que los valores "a priori" de x, independientes de los efectos observados, no fueran equiprobables. Laplace sugirió, entonces, reemplazar la función $f(x)$ por la función compuesta $f(x)g(x)$, de modo que la probabilidad anterior quedara definida por la expresión $\frac{\int_{x_1}^{x_2} f(x)g(x)dx}{\int_0^1 f(x)g(x)dx}$. De esta forma, la distribución de probabilidades "a priori" de las causas se asimiló al caso más simple en que dicha distribución es uniforme pero los efectos observados son la resultante de dos componentes independientes cuyas probabilidades están definidas por las funciones $f(x)$ y $g(x)$, respectivamente. Una conclusión cuyo único efecto fue el de transformar la cuestión de definir la distribución de probabilidades "a priori" en la tarea -igualmente difícil- de estimar la ficticia función de probabilidades compuestas $f(x)g(x)$.

En la misma "*Mémoire*" Laplace generalizó el resultado anterior considerando el problema de dos eventos simples independientes (x_1 y x_2), cada uno de los cuales se suponía compuesto por un gran número de eventos simples, todos del mismo tipo. Denotando por $f_1(x_1)$ y $f_2(x_2)$ a las probabilidades de los efectos observados, obtuvo la siguiente expresión para la probabilidad de que el primer evento simple fuera más probable que el

---

[101] Esta memoria, que ejerció una gran influencia entre sus contemporáneos, fue la que produjo la difusión de las ideas Bayesianas en el mundo matemático.

[102] "*La probabilidad de la mayor parte de los eventos simples es desconocida y 'a priori' parece igualmente suceptible de asumir cualquier valor entre 0 y 1; pero, es a partir de la observación de los resultados de varios de dichos eventos, que algunos de esos valores se hacen más probables que los otros*" (1781, p. 228).

segundo:

$$p(x_1 > x_2) = \int p\{[x_2 < f_1(x_1)] \,/\, [x_1 = f_1(x_1)]\} f(f_1) df_1 =$$

$$= \frac{\int_0^1 \int_0^{x_1} f_1(x_1) f_2(x_2) dx_2 dx_1}{\int_0^1 \int_0^1 f_1(x_1) f_2(x_2) dx_2 dx_1}$$

Resultado conocido como la“*doble integral de Bayes*”[103].

Inmediatamente después de la memoria de Laplace, se publicaron algunos trabajos en los que aparecieron aplicaciones de las demostraciones de Bayes y del mismo Laplace sobre la probabilidad inversa: “*Miscellaneous treatise; containing several mathematical subjects*” de William Emerson (1766), “*Essai d'arithmétique morale*” de George Louis Leclerc-Buffon (1777)[104], “*De probabilitate causarum ab effectibus oriunda: disquisitio mathematica*” de Jean Trembley (1795-1798), “*Mémoire sur l'art d'estimer la probabilité des causes par les effects*” de Pierre Prevost y Simon Antoine Jean Lhuilier (1799), “*Traité élementaire du calcul des probabilités*” de Sylvestre François Lacroix (1816).

Como corolario delos resultados precedentes, Laplace desarrolló algunas aplicaciones de su teoría de la probabilidad de las causas a la llamada “*regla de sucesión*”. En la “*Mémoire sur les approximations des formules qui sont fonctions de trés grands nombres*”planteó el siguiente problema: Sea una urna que contiene infinitas tarjetas rojas y azules, en una proporción desconocida. Supóngase que de dicha urna se hayan realizado n extracciones al azar (con reposición), habiéndose obtenido n_A tarjetas rojas y n_B tarjetas azules $(n_A + n_B = n)$.) Cuál es la probabilidad de que al realizar una nueva extracción se obtenga una tarjeta roja?

[103] La utilización de integrales de Bayes múltiples también puede ser hallada en la “*Mémoire sur le calcul des probabilités*” (1781-1784) de Condorcet.

[104] En esta memoria -contenida en el tomo IV de su “*Histoire naturelle*” (obra que ejerció gran influencia entre sus contemporáneos)- Buffon utiliza nociones elementales de la teoría de la probabilidad para justificar su hipótesis acerca del origen de los planetas a partir de la colisión entre el sol y un cometa.

En primer lugar obtuvo que la probabilidad de que el verdadero valor de la proporción -desconocida- de tarjetas rojas y azules dentro de la urna estuviera comprendido en un intervalo $[x_1, x_2]$, estaba dada por:

$$\frac{\int_{x_1}^{x_2} \binom{n}{n_A} x^{n_A}(1-x)^{n-n_A}dx}{\int_0^1 \binom{n}{n_A} x^{n_A}(1-x)^{n-n_A}dx}$$

Como corolario de esta expresión, y utilizando la relación:

$$p(B\ /\ C) = \sum_i p[B\ /\ (A_i \cap C)]p(A_i\ /\ C)$$

Laplace concluyó que la probabilidad de que un evento que había ocurrido n_A veces en $n = n_A + n_B$ repeticiones, volviera a ocurrir a la vez siguiente, era igual a:

$$\frac{\int_{x_1}^{x_2} x^{n_A}(1-x)^{n-n_A}dx}{\int_0^1 x^{n_A}(1-x)^{n-n_A}dx}$$

Expresión que, mediante integración sucesiva por partes, conduce a la siguiente forma de la regla de sucesión: $\frac{n_A+1}{n_A+n_B+1}$.

Inmediatamente generalizó este resultado a la probabilidad de que, habiendo obtenido n_A tarjetas rojas y n_B tarjetas azules en n extracciones al azar ($n_A + n_B = n$), se obtuvieran n'_A tarjetas rojas y n'_B tarjetas azules en las siguientes n' extracciones ($n'_A + n'_B = n'$), en ese orden:

$$\frac{\int_0^1 x^{n_A+n'_A}(1-x)^{n_B+n'_B}dx}{\int_0^1 x^{n_A}(1-x)^{n_B}dx} =$$

$$= \frac{(n_B+1)(n_B+2)\ldots(n_B+n'_B)(n_A+1)(n_A+2)\ldots(n_A+n_B+1)}{(n_A+n'_A+1)(n_A+n'_A+2)\ldots(n_A+n_B+n'_A+n'_B+1)}$$

Analizó, en primer lugar, el caso para $n_A + n_B$ muy grande con respecto a $n'_A + n'_B$, y obtuvo como solución aproximada de la distribución "a posteriori", la expresión:

$$\frac{n_A^{n'_A} n_B^{n'_B}}{(n_A + n_B)^{n'_A + n'_B}}$$

Luego demostró que, para valores grandes de n'_A y n'_B, por ejemplo, para $n'_A + n'_B = n_A + n_B$, la probabilidad era aproximadamente igual a[105]:

$$\sqrt{\frac{1}{2}} \frac{n_A^{n'_A} n_B^{n'_B}}{(n_A + n_B)^{n'_A + n'_B}}$$

Laplace extendió el análisis precedente a distribuciones de probabilidades "a posteriori" para una sucesión suficientemente larga de observaciones repetidas en igualdad de condiciones, demostrando que, para $n_A + n_B \rightarrow \infty$, y para un valor w "*infinitamente menor*" que $(n_A + n_B)^{-1/3}$ e "*infinitamente mayor*" que $(n_A + n_B)^{-1/2}$, la Normalidad asintótica de la distribución Beta[106]:

$$p\left[\left(\left|x - \frac{n_A}{n_A + n_B}\right| < w\right) / \,(habiéndose\ obtenido\ n_A\ bolillas\ blancas$$

$$y\, n_B\ bolillas\ negras)\right] =$$

[105] Laplace propuso, además, una interesante extensión de la fórmula de sucesión, de acuerdo con la cual obtuvo que, dada una racha observada de n_A tarjetas rojas, la probabilidad de obtener una nueva racha de n'_A tarjetas rojas estaba dada por $\frac{n_A+1}{n_A+n'_A+2}$ (esta solución no menciona restricciones a su validez, por lo que -erróneamente- podría suponerse aplicable aún en casos en que el número de las observaciones pasadas fuera muy pequeño e incluso nulo).

[106] Posiblemente Laplace haya sido el primero en integrar la distribución Normal. Gauss (1809) se refirió a este desarrollo como "...*un elegante teorema de Laplace*" (p. 120).

$$= \frac{\int_{\frac{n_A}{n_A+n_B}-w}^{\frac{n_A}{n_A+n_B}+w} x^{n_A}(1-x)^{n_B}dx}{\int_0^1 x^{n_A}(1-x)^{n_B}dx}$$

$$= \frac{(n_A+n_B)^{3/2}}{\sqrt{2\pi n_A n_B}} \int_0^w 2\, exp\left[-\frac{(n_A+n_B)^2 y^2}{2n_A n_B}\right] dy \to 1$$

Para el caso en que el orden de los resultados de las n' siguientes extracciones resultara irrelevante, halló que la fórmula original se transformaba en:

$$\binom{n'_A+n'_B}{n'_A} \frac{\int_0^1 x^{n_A+n'_A}(1-x)^{n_B+n'_B}dx}{\int_0^1 x^{n_A}(1-x)^{n_B}dx}$$

Marie-Jean Antoine Nicolas Caritat-Condorcet -quien percibió inmediatamente la potencialidad de este tratamiento de Laplace de las probabilidades inversas- utilizó esta versión de la regla de sucesión como una herramienta para la verificación de hipótesis científicas.

En su "*Essai*" (1785) Condorcet propuso el siguiente problema: "*Sea una función f(x), que predice correctamente los valores de una cantidad observable, x, para series de n ensayos.¿Cuál es la probabilidad de que el valor observado para x, en una serie posterior de n' ensayos continúe siendo conforme a la función f(x)?"*Siguiendo el desarrollo de Laplace, asimiló este enunciado a un modelo de urnas en el cual el número desconocido de bolillas rojas representaba los valores de x que satisfacían a la función $f(x)$, y el número desconocido de bolillas azules correspondía a valores "*anómalos"* de x. Al igual que Laplace, supuso que, dado que la ignorancia del observador impedía que existiera alguna razón para preferir una hipótesis a otra, todas las posibles proporciones de bolillas rojas y azules (es decir, todas las probabilidades "a priori" de las causas) eran igualmente probables. Estableció, además que, en la medida que el intervalo a predecir aumentara con respecto al intervalo de observación, la probabilidad de que la función $f(x)$ continuara siendo acorde a las observaciones, disminuiría. Es decir, que la "*fuerza"* de las generalizaciones de regularidades pasadas a resultados futuros decrecería en la medida que el número de observaciones realizadas disminuyera, o aumentara el número de eventos futuros.

Posteriormente, en "*Réflexions sur la méthode de déterminer la probabilité des événements futurs, d'après l'observation des événements passés*" (1783), Condorcet intentó ampliar el alcance de la regla de sucesión a casos en los cuales las causas variaban en el tiempo, utilizando un esquema de urnas en el que cada evento observado correspondía a una extracción de una urna diferente, con distinta composición de bolillas rojas y azules. A partir de este modelo modificado, postuló la hipótesis de la existencia de una sucesión de efectos indistinguibles (por ejemplo, extraer una bolilla roja), producido cada uno por una causa diferente (la proporción particular de bolillas rojas y azules).

Partiendo del supuesto de que las causas de los eventos individuales observados obedecían a una ley general y que podían clasificarse como de "*probabilidades constantes*", "*probabilidades variables independientes del tiempo y del orden*" y "*probabilidades variables que dependen del tiempo y del orden*", concluyó que la única diferencia entre los casos de "*causas constantes*" y "*causas variables*" radicaba en la ignorancia de la ley que rige el comportamiento de las "*causas constantes*"[107].

De todo lo anterior se puede concluir que la diferencia fundamental entre Bayes y Laplace radica en que, en tanto el objetivo del primero consistió en la estimación de una probabilidad, la propuesta (dinámica) de Laplace estuvo dirigida a predecir el comportamiento futuro de un fenómeno. De la lectura de las memorias de ambos autores se desprende que ni Bayes ni Laplace confundieron sus resultados. Fue con el transcurso del tiempo que esta claridad conceptual se fue perdiendo.

[107] Condorcet permaneció fiel al perfecto determinismo Bernoulliano.Nunca puso en duda la hipótesis de que todos los fenómenos estuvieran gobernados por leyes constantes. En su concepción, la teoría de la probabilidad de las causas se basaba, exclusivamente, en la imperfección del conocimiento del observador la cual era una función de la extensión de la experiencia pasada y de las predicciones futuras.

12.- Siméon Denis Poisson y el problema de los juicios

En 1830 se publicó "*Sur la proportion des naissances des filles et des garçons*" de Poisson, dedicado al análisis de las proporciones de nacimientos femeninos y masculinos en distintas localidades de Francia durante un período de 10 años. En la primera sección de esta memoria, titulada "*Probabilité de la répétition d'un événement dont la chance est donée*", propone una demostración de la ley de los grandes números de Bernoulli según la cual, dada una variable aleatoria $Y\!: b(n,\theta)$, que representa el número de "éxitos" (n_E) a ocurrir en una sucesión de eventos, para valores suficientemente grandes de n[108], se verifica que:

$$p\left[N - z\sqrt{2(n+1)}\theta(1-\theta) \leq Y \leq N + z\sqrt{2(n+1)}\theta(1-\theta)\right] =$$

$$= 1 - \frac{2}{\sqrt{\pi}}\int_{z}^{\infty} e^{-v^2}dv + \frac{e^{-z^2}}{\sqrt{2\pi n\theta(1-\theta)}}$$

(donde N denota el mayor entero menor o igual a $n-\theta$ y $z\sqrt{2(n+1)}\theta(1-\theta) \ll N$).

En "*Recherches sur la probabilité des jugements en matière criminelle et en matière civile, précédées des règles générales du calcul des probabilités*"(1837)[109]demuestra que, en el caso en que la probabilidad θ sea

[108] De acuerdo con Keynes (1921), esta aproximación requiere que el valor $n\theta(1-\theta)$ sea suficientemente grande.

[109] Como su título lo indica, esta memoria -que, en cierta forma constituye una colección de sus principales aportes anteriores a la teoría de la probabilidad (entre los que cabe destacar "*Sur la probabilité des résultats moyens des observations*" (1827) y "*Suite du mémoire sur la probabilité des résultats moyens des observations, inséré dans la connaisance des temps de l'anée 1827*" (1832))- está dedicada, fundamentalmente, a la resolución del problema de cómo determinar la probabilidad de que un jurado arribe a una decisión correcta ("*La determinación de la probabilidad de que un acusado sea condenado o absuelto por mayoría de votos de los jurados, teniendo en cuenta que a cada uno de los votos le corresponde una probabilidad de*

desconocida, puede ser aproximada por la frecuencia relativa del "éxito" (Y_n) de acuerdo con la siguiente versión del teorema de Bernoulli:

$$p\left(|\theta - Y_n| \leq \frac{z}{n}\sqrt{\frac{2n_E(n-n_E)}{n}}\right) =$$

$$= 1 - \frac{2}{\pi}\int_z^\infty e^{-v^2}dv + e^{-z^2}\sqrt{\frac{n}{2\pi n_E(n-n_E)}}$$

Poisson utilizó este resultado para aproximar probabilidades de eventos futuros[110]: denotando por n^* al número de repeticiones adicionales a n y por n_E^* al número de ocurrencias del resultado "éxito" en las n^* repeticiones adicionales y suponiendo que n_E y n asuman valores suficientemente grandes y que n^* sea relativamente pequeño con respecto a n pero, a su vez, suficientemente grande, se puede escribir:

$$p\left(\left|\frac{n_E^*}{n^*} - \frac{n_E}{n}\right| \leq \frac{z}{n}\sqrt{\frac{2n_E(n-n_E)}{n}}\right) =$$

---

ser correcto y, también, la probabilidad de que el acusado haya cometido el delito. Una determinación de la probabilidad que el acusado sea condenado o absuelto siendo culpable o inocente, de acuerdo a la regla que gobierna la probabilidad de las causas o hipótesis" ("*Recherches*", p. vii). La aplicación del cálculo de probabilidades a la confiabilidad de los veredictos judiciales ya había sido presentada por John Craig ("*Theologiæ christianæ principia mathematica*" (1699)), Gerge Hooper ("*A calculation of the credibility of human testimony"* (1699)), N. Bernoulli ("*De usu artis conjectandi in jure"* (1709)), Condorcet ("*Essai sur l'application d'analyse à la probabilité des décisions rendues à la pluralité des voix"* (1785)), y tratada por Laplace en su "*Théorie analytique des probabilités"* (1812).

110 Para una discusión más detallada de esta generalización de Poisson, ver Keynes (1921), obra en la cual el teorema de Bernoulli es presentado *"...en mayor medida a partir de una interpretación algebraica que lógica"* (p. 341).

$$= 1 - \frac{2}{\sqrt{\pi}} \int_{z}^{\infty} e^{-v^2} dv + \frac{n e^{-v^2}}{\sqrt{2\pi n^* n_E (n - n_E)}}$$

A fin de intentar una versión inversa del teorema de Bernoulli, Poisson propuso la siguiente aproximación:

$$p\left(n - n_E < n(1-\theta) - r\sqrt{2n\theta(1-\theta)}\right) =$$
$$= \frac{1}{\sqrt{\pi}} \int_{r+\delta}^{\infty} e^{-v^2} dv + e^{-r^2} \frac{1 - 2\theta}{3\sqrt{2(n+1)\theta(1-\theta)}} = Q$$

(donde $\delta = \frac{-(1-2\theta)r^2}{\sqrt{2(n+1)\theta(1-\theta)}}$). Reemplazando θ por la frecuencia relativa del resultado "éxito", en el límite, la expresión anterior asume la forma:

$$p\left(1 - \theta > \frac{n - n_E}{n} + \frac{r}{n}\sqrt{\frac{2n_E(n - n_E)}{n}}\right) = Q$$

A partir de la cual se obtiene que:

$$-\frac{d}{dr}Q = p\left(\frac{n - n_E}{n} + \frac{r}{n}\sqrt{\frac{2n_E(n - n_E)}{n}} < 1 - \theta <\right.$$
$$\left. < \frac{n - n_E}{n} + \frac{r + dr}{n}\sqrt{\frac{2n_E(n - n_E)}{n}}\right)$$

En el tercer capítulo de las "*Recherches*" ("*Calcul des probabilités qui dépendant de très grands nombres*"), basándose en la consideración que el modelo de causación implícito en el teorema de Bernoulli adolecía del defecto de excesiva simplificación, Poisson propuso la primera generalización de la ley de los grandes números para sumandos binomiales no-idénticamente distribuidos (posiblemente su contribución más importante a la teoría matemática de la probabilidad y a la estadística)[111], a partir de un esquema que consideraba la existencia de múltiples causas (C_i) que influían sobre el mismo evento observado, generando probabilidades distintas sobre la ocurrencia del mismo[112]. Con este fin, recurrió a un ejemplo consistente en arrojar 2.000 monedas –en el que las "*causas desconocidas*" variaban de una moneda a otra-, en contraste con el caso al que se podría asimilar el teorema de Bernoulli, que consistía en arrojar una única moneda 2.000 veces –en el que las características de la misma le permitirían suponer la constancia de la probabilidad de cada resultado a lo largo de la serie de repeticiones-, en la convicción que este planteo constituía el tipo ideal para representar la complejidad de los fenómenos naturales y morales[113]. Su enunciado puede ser expresado de la siguiente forma: Sea la variable aleatoria $Y_n = \frac{X_{(n)}}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i$, donde las $X_i (i = 1,2, \dots, n)$ denotan variables del tipo $b(1, \theta_i)$ (pruebas "*binomiales-generalizadas*" o "*de Poisson*") en las que θ_i

---

[111] En "*Recherches*", además del teorema que se menciona en el texto, desarrolló otro tema que ha quedado indisolublemente unido al apellido Poisson: la definición de la distribución de probabilidades $p(X) = \frac{\lambda^X}{X!} e^{-\lambda} (X = 0,1,2, \dots)$ (donde $\lambda = n\theta_n$) como límite de la distribución de probabilidades binomial, cuando $\theta_n \to 0$ (esta distribución de Poisson fue denominada como "*ley de los pequeños números*" por von Bortkiewicz).

[112] Poisson distingue dos clases de causas, de acuerdo con la intensidad del vínculo causa-efecto: por un lado, aquéllas que *"producen"* necesariamente un evento ($p(E/C) = 1$) y, por otro, aquéllas que producen eventos sólo en combinación con *"causas variables"* o *"accidentales"* ($p(E/C) < 1$). Estas últimas son muy parecidas a las *"causas irregulares"* de Laplace y, al igual que éstas, Poisson las supuso simétricamente distribuidas ("... *en la aplicación del cálculo de probabilidades, tanto a fenómenos físicos como a cuestiones morales, a menudo las probabilidades varían de una prueba a otra y, frecuentemente, en forma irregular*" (p. 139)).

[113] Poisson (1837): "*Lo que ocurre en las cuestiones materiales es una imagen de lo que ocurre en las cuestiones morales, consideradas independientemente de la naturaleza de su causa, y sólo respecto de sus efectos. En el caso de los juicios criminales, por ejemplo, las posibilidades de condena y de absolución varían de una experiencia a otra, de la misma forma que las posibilidades de las dos caras de una moneda varían de una moneda a otra*" (p. 379).

representa la probabilidad de que el evento E("éxito") se presente en la i-ésima repetición, debido a la causa C_i. Aplicando la desigualdad de Bienaymé-Chebychev a la variable Y_n, se obtiene que:

$$p\left(\left|Y_n - \frac{1}{n}\sum_{i=1}^{n}\theta_i\right| \leq \varepsilon\right) > 1 - \frac{1}{n\varepsilon^2}\sum_{i=1}^{n}\theta_i(1-\theta_i) > 1 - \frac{1}{4n\varepsilon^2}$$

Por lo que queda demostrado que:

$$\lim_{n\to\infty} p\left(\left|\frac{1}{n}\sum_{i=1}^{n}X_i - \frac{1}{n}\sum_{i=1}^{n}\theta_i\right| \leq \varepsilon\right) = 1$$

Es decir, que en una sucesión infinitamente larga de pruebas, la frecuencia relativa del "éxito" converge en-probabilidad al promedio de las probabilidades de obtener éxito en cada prueba individual, aún cuando este promedio no tienda a un límite.

Poisson -como Bernoulli y de Moivre- no podía concebir un universo "gobernado" por el azar. Pero, a diferencia de ellos -que suponían la existencia de un "O*rden Providencial"*- Poisson sostenía que la tendencia de los fenómenos a exhibir regularidades estadísticas era inherente "*...al estado natural de las cosas, que subsisten por sí mismas, sin la ayuda de ninguna causa extraña y que, por el contrario, requerirían de tal causa para experimentar un cambio significativo"* (pp. 144-145)[114]. Al igual que Bernoulli, Leibniz, de Moivre, Price y Condorcet, consideró que la eventual no verificación del principio de estabilidad de las frecuencias, no significaba una refutación del "*principio de permanencia de las causas"* que gobernaban la naturaleza, sino el reconocimiento de que algunas de estas causas, $C_1, C_2, C_3, \ldots$, podrían haber sido reemplazadas por otras, haciendo que las probabilidades de ocurrencia del evento en una prueba dada, variaran.

Poisson adoptó el concepto de convergencia al promedio de las probabilidades –desarrollado posteriormente en los trabajos de Bienaymé y Lexis referidos al estudio de la homogeneidad y estabilidad en un esquema de

[114] Un principio que podría considerarse asimilable al concepto de inercia.

pruebas repetidas[115]- del "*Essai sur l'application d'analyse à la probabilité des décisions rendues à la pluralité des voix*" (1785) en el cual Condorcet, fundándose en el principio de uniformidad de la naturaleza y en un intento por cuantificar la influencia de la experiencia histórica sobre el grado de creencia analizado posteriormente por la interpretación logicista[116], desarrolló una filosofía de la relación entre dicho principio de uniformidad y el método científico. Reconoció a la experiencia histórica como "*la única regla de nuestras opiniones y nuestras acciones*" (pp. x, xi) y admitió, de acuerdo con el principio de Cournot (1838)(1843), la posibilidad de calcular, en ciertos casos, el incremento en la asignación de la probabilidad asociada a dicha regla, a partir de la acumulación de observaciones[117].

Los métodos probabilísticos propuestos por Poisson en sus "*Recherches*" se extendieron inmediatamente (e indiscriminadamente) a las ciencias sociales. Circunstancia que dio lugar ainnumerables críticas, en particular de la Académie des Sciences (el referente filosófico de la época), principalmente por obra de Charles Gouraud (1848) quien, al cabo de un comentario sobre el significado matemático y filosófico de la ley de los grandes números, criticó la pretensión de adjudicarle el rol de argumento universal para justificar la aplicación de la teoría de la probabilidad a nivel ético y moral[118]. Louis

---

[115] En particular, Bienaymé (1855) consideró que la propuesta de Poisson no difería de la de Bernoulli y cuestionó el supuesto de independencia de las observaciones (ver Heyde; Seneta (1977)).

[116] Ver Sec. 20.

[117] Continuando con el planteo de Condorcet, los probabilistas del siglo XIX basaron el concepto de esperanza matemática en la ley de los grandes números y utilizaron la paradoja de San Petersburgo como un ejemplo que confirmaba el principio que la probabilidad sólo era concebible para eventos repetibles.

[118] "*Los objetos de la naturaleza como una totalidad, tanto en el mundo moral como en el físico, están sometidos, de acuerdo con Poisson, a una ley universal que aproximadamente puede ser expresada de la siguiente forma: si se observa un número muy considerable de eventos de la misma clase que dependen de causas constantes, causas que, como ocurre en el mundo moral, varían en forma irregular, a veces en una dirección y a veces en otra, ocurre que la variación de estos efectos irregulares producidos por las causas variables sufrirá una creciente disminución en forma proporcional al número de experimentos realizados y, si los experimentos se multiplican suficientemente, las relaciones de los eventos observados se aproximarán notablemente a la estabilidad (...) Sin evaluar el peligro que pudiera acarrear la aceptación de este principio, Poisson concluyó en la posibilidad de determinar una*

Poinsot rechazó el esquema de urnas utilizado por Poisson, dado que consideró que las cuestiones planteadas en las "*ciencias morales*" eran demasiado complejas y de efectos impredecibles como para ser analizadas por cualquier clase de modelo matemático. Charles Dupin, Destutt de Tracy y Auguste Compte criticaron la falta de razonabilidad de las condiciones impuestas por Poisson. Victor Cousin ("*Cours d'histoire de la philosophie morale au dixhutième siècle*" (1841)) consideró que la pretensión de relacionar las regularidades determinadas por la ley de los grandes números y el planteo ético postulado por los filósofos era completamente injustificable. Esta controversia en el ámbito filosófico se trasladó al campo matemático y produjo una división a nivel institucional entre la Académie (representada por Lagrange y Laplace) y la École Polytechnique (representada por Cauchy).

De las secciones 11 y 12 se puede concluir finalmente, que los probabilistas Laplacianos y los probabilistas franceses y rusos que constituyeron su sucesión comenzaron asumiendo una interpretación subjetivista intentando luego, a partir de una interpretación equivocada del teorema de Bernoulli y del principio de Cournot establecer la existencia de probabilidades objetivas mediante la aplicación de la ley de los grandes números. Utilizaron, sin una base conceptual válida, los postulados del teorema de Bayes para razonar acerca de las probabilidades objetivas en casos en los que se contara con un número suficientemente grande de repeticiones.

13.- Antoine Augustin Cournot y los eventos ciertos en-probabilidad

El conocido como "*principio de Cournot*" está relacionado con la existencia de eventos distintos de Ω que también tienen probabilidad igual a uno: son los denominados "*eventos quasi-ciertos*" o "*ciertos-en-probabilidad*"[119]. La

---

probabilidad matemática asociada a cada decisión humana. Consideró que aún aquellas más complicadas, sujetas a opiniones cambiantes y caprichosas, con la ayuda de la ley de los grandes números, pueden ser analizadas y juzgadas con sumo rigor. En 'Recherches sur la probabilité des jugements' figuran expresiones numéricas obtenidas a partir del análisis de un número muy grande de juicios previos con la finalidad de determinar la probabilidad exacta de que cada acusado sea considerado culpable o inocente. Esto constituye una audacia increíble que excede la de Laplace y la de Cournot" (pp. 132-133).

[119] Markov (1912): *"Cuanto más se aproxime a la unidad la probabilidad de un evento, mayor será la razón para esperar su ocurrencia. En cuestiones prácticas, estamos obligados*

negación de un evento quasi-cierto define un evento que, no coincidiendo con el evento imposible, se comporta como tal en lo que se refiere a su probabilidad, y se denomina "*quasi-imposible*" o "*imposible-en-probabilidad*"[120]. Estos eventos poseen las siguientes propiedades: **i)** la unión de dos eventos imposibles-en-probabilidad es un evento imposible-en-probabilidad (propiedad generalizable a cualquier conjunto finito o infinito numerable de eventos); **ii)** sean E_I un evento imposible-en-probabilidad y E un evento ni cierto ni imposible, entonces se verifica que $p(E_I \cap E) = 0$ y $p(E_I \cup E) = p(E)$ y **iii)** sean E_C un evento cierto-en-probabilidad y E un evento ni cierto ni imposible, entonces se verifica que $p(E_C \cap E) = p(E)$ y $p(E_C \cup E) = 1$.

Estas definiciones derivan de los conceptos de evento "*moralmente imposible*" y "*moralmente cierto*" formulados originalmente por J. Bernoulli. La "*certeza moral probabilística*" fue ampliamente discutida en el siglo XVIII. d'Alembert (1761)(1767) planteó la necesidad de distinguir entre eventos "*metafísicamente imposibles*" y "*físicamenteimposibles*". Al respecto, Leclerc-Buffon (1777) consideró que la diferencia entre las certezas moral y física era una cuestión de grado de aproximación, según su apreciación un evento con probabilidad $\frac{9999}{10000}$ es moralmente cierto, en tanto que un evento con probabilidad mucho mayor puede ser considerado como físicamente cierto[121].

---

a considerar como ciertos aquellos eventos cuya probabilidad se aproxima a la unidad y como imposibles a aquéllos cuya probabilidad es pequeña. Consecuentemente, uno de los logros más importantes de la teoría de la probabilidad consiste en identificar aquellos eventos cuyas probabilidades son próximas a uno o a cero" (p. 67).

[120] Es decir, un evento cuya probabilidad, si bien no es efectivamente igual a cero, sólo posee un interés puramente metafísico.

[121] Leclerc-Buffon, al igual que Butler y Price, utilizó el ya mencionado ejemplo del "*amanecer*". Partiendo del supuesto que los dos resultados posibles "que mañana amanezca" y "que mañana no amanezca", eran equiprobables y teniendo en cuenta que el resultado "amanecer" se había producido sin excepción todos los días de los últimos 5781 años, concluyó que la probabilidad del evento "que mañana amanezca" era igual a $1 - \frac{1}{2^{n-1}}$, donde n denotaba el total de observaciones diarias desde la creación del mundo (debe tenerse en cuenta que, en 1777 y, de acuerdo a la cronología del Reverendo James Ussher, Obispo de Armagh, la edad atribuída al planeta tierra era de 5781 años). Ver Loveland (2001).

Cournot (1843) –un convencido determinista-, basándose en el concepto de probabilidad geométrica, planteó una formulación de la ley de los grandes números según la cual se puede asegurar que un evento moralmente imposible no ocurrirá (que es matemáticamente posible, pero físicamente imposible) y, como corolario, que es físicamente imposible que dada una sucesión suficientemente larga de repeticiones, la frecuencia relativa de un evento difiera sustancialmente de su probabilidad. Consideró que este principio, que asimilaba una probabilidad infinitesimalmente pequeña a la imposibilidad de ocurrencia de un evento, proporcionaba un significado empírico a la probabilidad clásica[122]. Pero, debe tenerse en cuenta que existe una diferencia conceptual muy importante entre la afirmación de que un evento de probabilidad infinitesimal no ocurrirá y el supuesto que, por sí sola, esa decisión de no-ocurrencia del evento proporcionará a lateoría clásica un significado empírico[123].

E. Borel (1906)(1909b)(1914)(1930), utilizando a menudo un estilo más literario que matemático o filosófico, adoptó el tratamiento de los eventos con probabilidad infinitesimal como imposibles y estableció una clasificación de los mismos más refinada que la de Buffon, según la cual una probabilidad de 10^{-6} es despreciable "*a escala humana*", una probabilidad de 10^{-15} es despreciable a "*escala terrestre*" y una probabilidad de 10^{-50} es

[122] Cournot (1843): *"...el evento físicamente imposible es, por lo tanto, aquél al que le corresponde una probabilidad infinitamente pequeña y esta sola observación es suficiente para proporcionar una sustancia objetiva fenomenal a la teoría de la probabilidad matemática"* (p.158) (la expresión *"objetiva fenomenal"* se refiere a la distinción planteada por Kant entre el *"noumenon"*, o cosa-en-sí-misma y el *"phenomenon"* u objeto de la experiencia). Ver Daston (1994).

[123] L. Boltzmann (1866)(1871)(1872)(1877)(1887) utilizó el principio de Cournot para proporcionar una explicación estadística de la segunda ley de la termodinámica (los procesos disipativos son irreversibles porque la probabilidad de un estado con entropía lejos del máximo es infinitamente pequeña). Por su parte, H. Poincaré (1890) postuló en su teorema de recurrencia, que un sistema mecánico aislado, limitado a una región acotada de su espacio de fase, retornará eventualmente a un estado arbitrariamente próximo a su estado inicial, si dicho estado no es excepcional (es decir, que los estados para los cuales no se verifica la recurrencia son excepcionales, en la medida que están contenidos en subregiones cuyo volumen total es arbitrariamente pequeño).

despreciable a "*escala cósmica*"[124].

De las consideraciones realizadas precedentemente, se deriva que el principio de Cournot admite dos modalidades que se corresponden con las versiones fuerte y débil de la ley de los grandes números: **i)** una forma fuerte o estricta que considera que un evento con probabilidad pequeña o nula no ocurrirá en la repetición siguiente y que, combinada con los postulados del teorema de Bernoulli permite concluir que la probabilidad de un evento siempre puede ser aproximada por su frecuencia relativa a partir de una serie suficientemente larga de repeticiones independientes y **ii)** una forma débil que considera que un evento con probabilidad muy próxima a cero ocurrirá muy raramente en una serie de repeticiones y que combinada con el teorema de Bernoulli, permite concluir que la probabilidad de un evento habitualmente puede ser aproximada por su frecuencia relativa a partir de una serie sufiecientemente larga de repeticiones independientes. Borel, Lévy y Kolmogorov adoptaron la forma fuerte, Chuprov y Castelnuovo la forma débil[125].

J. Hadamard (1922), de acuerdo con una posición netamente determinística, postuló que las nociones de eventos igualmente probables ("*perfectamente equivalentes*" según su nomenclatura) y de eventos muy improbables constituyen el fundamento de la teoría de la probabilidad y que el hecho que la equivalencia perfecta sea una condición matemática no verificable en la práctica, no debe impedir la aplicación del principio de los eventos muy improbables. Por su parte, Lévy (1925) coincidió en que la noción de eventos equiprobables puede ser considerada como un fundamento de la matemática de las probabilidades, pero admitió que, si se la asume como única base de razonamiento, las probabilidades obtenidas serán meramente subjetivas. Sostuvo, además, que el concepto de evento muy improbable es el único que permite atribuir un significado empírico a los resultados de la teoría matemática, que la combinación de esta noción con la ley fuerte de los grandes números permite una interpretación objetiva de la probabilidad como una constante física representada por la frecuencia relativa del evento. Como

---

[124] Las críticas de von Kriess al principio de Cournot ejercieron una notable influencia sobre los probabilistas alemanes, en particular sobre Czuber (1903) (ver Meinong (1915), Kamlah (1983)).

[125] En particular, Castelnuovo (1918) ejerció una notable influencia sobre las posiciones adoptadas por Hawlbachs (1924), Fréchet (1930a)(1930b) y, por lo tanto, sobre Kolmogorov (1933a).

se verá inmediatamente, se puede concluir que una versión propensionalista de este postulado frecuencista fue adoptada por Kolmogorov en sus "*Grundbegriffe der Wahrscheinlichkeitsrechnung*" (ver Sec. 25).

Debe tenerse en cuenta que, a partir de los trabajos de Wiman (1900)(1901) sobre fracciones continuas, los matemáticos alemanes dirigieron sus esfuerzos a demostrar que existen conjuntos de medida nula y que estos conjuntos son imposibles. Estas propuestas pusieron en evidencia el abismo que existe entre la probabilidad nula y la probabilidad infinitesimal de Cournot-Leclerc Buffon-Borel (ver F. Bernstein (1912)).

De acuerdo con una interpretación subjetivista resulta absolutamente injustificable atribuir a un evento posible el mismo grado de certidumbre subjetiva que a un evento imposible. Sean E_1 y E_2 dos eventos de probabilidad nula. De acuerdo con la notación habitual su representación probabilística sería de la forma $p(E_2) = p(E_2) = 0$. Ahora bien, esta igualdad sólo es válida a partir del supuesto restrictivo que ambos eventos son iguales en cuanto que sus probabilidades son iguales a cero.Pero, es posible definir una jerarquía entre probabilidades nulas. Una probabilidad nula puede ser infinitamente más grande o más pequeña que otra probabilidad nula. Dada una variable X, sus valores pueden ser representados como puntos sobre una recta. Si se agrega una nueva dimensión Y, la selección de un punto sobre la recta X es equivalente a la selección de toda la vertical con una abcisa determinada. Para identificar un punto de la variable (X, Y) se necesitan dos condiciones y, así sucesivamente. En el primer caso la probabilidad es 0, en el segundo la probabilidad es 0^2, en el tercero 0^3, etc. La idea intuitiva básica es que, dados dos eventos independientes de probabilidad nula, la probabilidad de su ocurrencia conjunta es $0^2 \neq 0$.

A fin de resolver estas controversias en la interpretación del teorema de Bernoulli, von Kries (1886) propuso como condición necesaria para razonar acerca de probabilidades objetivas, el principio del "*Spielraum*" según el cual las probabilidades objetivas pueden ser calculadas a partir de casos igualmente probables que cumplan las siguientes condiciones: **i)** las circunstancias que genera cada caso admiten el mismo "*Spielraum*", es decir, el mismo número de arreglos posibles y **ii)** además de éstas no existe ninguna otra circunstancia que afecte las expectativas relativas a cada caso. Con un razonamiento muy atinado von Kries concluyó que aún cuando, de acuerdo con estas condiciones, se pueda asegurar que un evento puede poseer una probabilidad objetiva, de ninguna manera se puede asignar a la ley de los grandes números los alcances que le atribuyeron Hadamard y Lévy. El teorema de Bernoulli postula solamente que un desvío significativo de la frecuencia relativa de un evento respecto de su probabilidad (suponiendo que, a partir de una interpretación determinística del comportamiento de los fenómenos fácticos, esta probabilidad exista) es simplemente tan improbable como cualquier otro evento improbable, dada una sucesión suficientemente grande de repeticiones[126].

14.- Augustus de Morgan y la corrección de los teoremas inversos de Laplace y Poisson

de Morgan (1838a) concluyó que de Moivre "*... a pesar de sus esfuerzos, no logró definir un método de inversión de la probabilidcad. Fue el Rev. T. Bayes en Phil. Trans. (ℓiii. 370) quien, a pesar de estar ahora casiolvidado, merece la consideración más destacada de los estudiosos de la probabilidad*" (p.vii).

De acuerdo con una interpretación de la probabilidad rigurosamente subjetivista –según la cual "*...por grado de probabilidad entendemos o, en realidad, deberíamos entender el 'grado de creencia' deFinettiano y 'creencia' no es sólo una denominación alternativa del conocimiento imperfecto (...) es la expresión de nuestro espíritu en un estado de conocimiento imperfecto*" (de Morgan (1847, p. 9))[127]- elaboró una propuesta basada en dos cuestiones que consideró fundamentales en el problema de la inversión: "*1) En la que se conocen las circunstancias previas y la incógnita es la probabilidad de ocurrencia de un evento y 2) En la que se conoce el evento que ha ocurrido y la incógnita es la probabilidad de que los resultados provengan de algún conjunto particular de circunstancias bajo las cuales*

[126] Al principio que indica que un evento con probabilidad de ocurrencia muy próxima a cero es imposible, von Kries (1886) lo denominó el *"error de d'Alembert"* (ver Heidelberger (2001*)*.

[127] de Finetti (1970) considera al *"grado de creencia"* como aquél que experimenta *"...un sujeto dado, en un instante dado, con un conjunto de información dado, respecto dela ocurrencia de un evento. Un concepto que se contrapone a otras concesiones que se limitan a tipos especiales de casos que atribuyen un sentido a las probabilidades objetivas, por ejemplo, casos de simetrías como el de los dados o casos 'estadísticos' de eventos repetibles"* (p. 6).

podría haber ocurrido. A la primera probabilidad la denomino directa y a la segunda inversa" (pp. 31-32). Con respecto a la segunda cuestión –es decir del "*argumento del evento a la causa*"- proporcionó una definición de causa simplemente como "*...un estado de cosas previo a la ocurrencia de un evento*" (p. 53), limitando esta definición a casos en que el número de antecedentes es finito.

Como corolario propuso una definición según la cual "*Dado un evento A observado, la probabilidad que proporciona este evento respecto de la ocurrencia de un evento B está dada por la suma de las probabilidades de cada estado precedente posible, condicionadas por A , multiplicada por la probabilidad de ocurrencia de B para cada uno de estos estados*" (pp. 60-61). Formalmente:

$$p(B\ /\ A) = \sum_i p(H_i\ /\ A)p(H_i\ /\ B)$$

(suponiendo que los eventos $(A\ /\ H_i)$ y $(B\ /\ H_i)$ sean estocásticamente independientes entre sí). A partir de esta definición, suponiendo que los eventos A y B han ocurrido m y n veces, respectivamente, planteó una versión de la regla de sucesión según la cual la probabilidad de la ocurrencia de A en la repetición siguiente está dada por $\frac{m+1}{m+n+2}$, cualesquiera fueran las probabilidades de los eventos. En general, propuso una probabilidad de que en las $m_1 + n_1$ repeticiones siguientes se produzcan m_1 ocurrencia de A y n_1 ocurrencias de B de la forma:

$$\frac{\binom{m+m_1}{m_1}\binom{n+n_1}{n_1}}{\binom{m+n+m_1+n_1+1}{m_1+n_1}}$$

En el capítulo IV ("*Use of the tables at the end of this work*") de Morgan retomó el problema de inversión postulando que, dada una sucesión suficientemente larga de n repeticiones en la que el evento E ocurrió n_E veces, la probabilidad (desconocida), θ, de la ocurrencia de E puede ser aproximada por el área:

$$p\left(\left|\theta - \frac{n_E}{n}\right| < \frac{k}{n}\right) \approx$$

$$\approx \frac{2}{\sqrt{\pi}} \int_{0}^{\frac{k}{\sqrt{\frac{2n_E(n-n_E)}{n}}}} 10^{-u^2/\alpha} du \qquad (u > 0, \alpha = ln(10))$$

En 1838b postuló una corrección a la versión inversa de Laplace y Poisson quienes habían propuesto la aproximación:

$$p(|n_E - n\theta| < l \,/\, \theta) = \int_{0}^{z} e^{-v^2} dv + \frac{e^{-z^2}}{\sqrt{2\pi uw}}$$

(donde $z = \ell\sqrt{\frac{n}{2uw}}$, con $w' = u + w$ y siendo u y v proporcionales a las probabilidades de "éxito" y "fracaso" en una repetición individual y θuna probabilidad desconocida). A partir del análisis de la aproximación del cociente $\frac{\int_a^b y^p(1-y)^q dy}{\int_0^1 y^p(1-y)^q dy}$ de acuerdo con el método de Laplace, de Morgan obtuvo que:

$$p(w - \mu\alpha < \theta < w) = \frac{1}{\sqrt{\pi}} \int_{-\mu}^{0} e^{-v^2} dv + \beta\left(e^{-\mu^2} - 1\right)$$

$$p(w - \mu\alpha < \theta < w + w + \mu\alpha) = \frac{1}{\sqrt{\pi}} \int_{-\mu}^{\mu} e^{-v^2} dv$$

$$p(w < \theta < w + \mu\alpha) = \frac{1}{\sqrt{\pi}} \int_{0}^{\mu} e^{-v^2} dv + \beta\left(1 - e^{-\mu^2}\right)$$

(donde $\alpha = \sqrt{\frac{2w(1-w)}{n}}$ y $\beta = \frac{\sqrt{2}(1-2w)}{3\sqrt{n\pi w(1-w)}}$). En la misma obra, propuso, además, una simplificación a la aproximación de Laplace, reemplazando a ℓ por $\ell + \frac{1}{2}$ y despreciando los términos del mismo orden que en la demostración de Laplace, obtuvo que:

$$p(u - \ell \leq n_E \leq u + \ell) = \frac{2}{\sqrt{\pi}} \int_{0}^{z} e^{-v^2} dv$$

15.- Pafnuty Lvovich Chebychev y la extensión de la ley de los grandes números a variables distintas de la binominal

Chebychev ("*Sur deux théorèmes relatifs aux probabilités*" (1887)) logró una importante generalización de la ley de los grandes números, al eliminar la restricción de que las variables involucradas fueran binomiales, y tomar en consideración la comparación entre una suma de n variables aleatorias cualesquiera con varianza acotada y su valor medio: Sea $\{X_1, X_2, \ldots, X_n\}$ un conjunto de variables aleatorias incorrelacionadas de a pares[128], con medidas $E(X_i) = m_i$ y $\sigma^2(X_i) = \sigma_i^2 (i = 1,2, \ldots, n)$ y tales que las varianzas cumplan con la condición de Markov, $\lim_{i\to\infty} \sigma_i^2 = 0$. Aplicando, entonces, la desigualdad de Bienaymé-Chebychev a la variable $Y_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ se obtiene que:

$$p\left(\left|\frac{1}{n}\sum_i X_i - \frac{1}{n}\sum_i m_i\right| \leq \varepsilon\right) > 1 - \frac{1}{\varepsilon^2}\left(\frac{1}{n}\sum_i X_i\right) =$$

$$= 1 - \frac{1}{(n\varepsilon)^2}\sum_i \sigma_i^2 > 1 - \frac{c}{n\varepsilon^2}$$

---

[128] Las condiciones necesarias y suficientes para poder asegurar que las variables X_i y $X_j (i \neq j)$ son incorrelacionadas son: **i)** que ambas posean momentos de segundo orden finitos y **ii)** que $E(X_i X_j) = E(X_i)E(X_j)$. En particular, si se verifica que $E(X_i X_j) = 0$, se dice que X_i y X_j son ortogonales ($X_i \perp X_j$). La condición $E(\sum_{i=1}^{n} X_i - \sum_{i=1}^{m} X_i)^2 = \sum_{i=m+1}^{n} E(X_i^2)$ (para $m < n$) (conocida como "*relación Pitagórica*") es equivalente a la condición de ortogonalidad (es decir, que la condición de ortogonalidad determina la relación entre los momentos de segundo orden de las sumas parciales, $\sum_{i=m+1}^{n} X_i$, y los de las variables individuales X_i). En particular, si $E(X_i) = 0$, para todo $i \geq 1$, se concluye en forma inmediata que la condición necesaria y suficiente para poder asegurar que las X_i son ortogonales es que sean incorrelacionadas. La condición necesaria y suficiente para poder asegurar que el conjunto $\{X_1, X_2, \ldots, X_n\}$ es de variables incorrelacionadas es que todas las variables del conjunto sean incorrelacionadas de a pares. Para este conjunto finito de variables incorrelacionadas de a pares se verifica que $\sigma^2(X_1, X_2, \ldots, X_n) = \sum_{j=1}^{n} \sigma^2(X_j)$.

Por lo que queda demostrado que[129]:

$$\lim_{n\to\infty} p\left(\left|\frac{1}{n}\sum_i X_i - \frac{1}{n}\sum_i m_i\right| \leq \varepsilon\right) = 1$$

Continuando con esta línea de razonamiento, de acuerdo con el teorema fundamental de la convergencia para variables ortogonales de Rademacher-Menchoff [130], la condición $\sum_{n=1}^{\infty}[\log n]^2 E(X_n^2) < \infty$ es necesaria y suficiente para poder asegurar que $\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{ccc} \frac{1}{n}\sum_{i=1}^{n} m_i$ [131].

16.- Andrei AndreevichMarkov y la condición de independencia

Si en la demostración precedente se retira la restricción referida a la independencia estocástica de las variables[132] se verificaque la esperanza matemática de $\mathrm{X_i}$ decrece con el incremento de la suma $X_1 + X_2 + \cdots + X_n$y resulta que:

$$\sigma^2(Y_n) = \sigma^2\left(\frac{X_{(n)}}{n}\right) = \sigma^2\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\left(\sum_i \sigma_i^2 + \sum_{i\neq j}\sum \gamma_{ij}\right) \leq$$

$$\leq \frac{1}{n^2}\left(\sum_i \sigma_i^2 + \sum_{i\neq j}\sum \sigma_i\sigma_j\right)$$

---

[129] Chebychev demostró este teorema en 1866, pero recién fue publicado -junto con su generalización del teorema central del límite- en "*Sur deux théorèmes relatifs aux probabilités*" en 1887.

[130] Rademacher (1922), Menchoff (1923).

[131] Como se verá en la Sec. 26, si $\sum_{i=1}^{n} X_i$ es una martingala, entonces la condición $\sum_{i=1}^{n} X_i^2 < \infty$ implica la convergencia casi-con-certeza.

[132] Debe tenerse en cuenta que, dada una sucesión de variables $\{X_1, X_2, \dots, X_n\}$ la condición necesaria y suficiente para que sean estocásticamente independientes es que $dF_{X_1,X_2,\dots,X_n}(x_1, x_2, \dots, x_n) = dF_{X_1}(x_1)dF_{X_2}(x_2)\dots dF_{X_n}(x_n)$.

Se puede, entonces, concluir fácilmente que la condición necesaria para que se cumpla la condición de Markov es que $\sigma^2 \sum X_i$ sea de orden asintótico inferior a n^2 $(\sigma^2(X_{(n)}) = o(n^2))$. Teniendo en cuenta esta restricción, Markov ("*The law of large numbers and the method of least squares*" (1898)) demostró entonces que:

$$\lim_{n\to\infty} p\left(\left|\frac{1}{n}\sum_i X_i - \frac{1}{n}\sum_i m_i\right| \leq \varepsilon\right) = 1$$

aún cuando las variables $X_i (i = 1,2, \dots)$ no fueran independientes (ver Markov (1907)(1908)(1911a)(1911b)(1912a))[133].

17.- Félix Édouard Justine ÉmileBorel y la probabilidad numerable

Los primeros intentos de vinculación de la teoría de la medida de conjuntos e integración de Lebesgue con la cuantificación de la probabilidad fueron propuestos por Borel (1905), iniciando un proceso que culminó con el sistema axiomático contenido en "*Grundbegriffe der Wahrscheinlichkeitsrechnung*" (1933b) de Kolmogorov[134]. Luego, en 1909a, mediante la formulación de su

---

[133] Se dice que una sucesión de variables aleatorias $\{Y_1, Y_2, \dots\}$ forman un proceso de Markov si se verifica que $p(Y_{n+1} < y/y_n, y_{n-1}, \dots) = p(Y_{n+1} < y / y_n)(n = 1,2, \dots)$. Si la probabilidad $p(Y_{n+1} < y / y_n)$ es invariante con respecto a n, se dice que el proceso es homogéneo. Todo proceso de Markov homogéneo con distribución inicial estacionaria es estacionario en sentido estricto. Como se verá en las secciones 24 y 25, la condición de Markov –que implica que toda la información sobre el infinito pasado del proceso está resumida en su estado inmediato anterior- es muy próxima a la condición necesaria de Kolmogorov (1928) y Khinchin (1929a)). Después de los trabajos de Niklaus y Daniel Bernoulli, Chebychev y Poisson, la generalización de la ley de los grandes números en particular y de los teoremas límite en general a variables aleatorias autocorrelacionadas entre sí, alcanzaron su culminación con los aportes de la escuela de San Petersburgo (ver Maistrov (1974)).

[134] La teoría de la probabilidad de Borel –que sólo representa una parte pequeña de su obra matemática- comienza en "*Éléments de la théorie des probabilités*" (1909), continúa con una serie de monografías resumidas en el "*Traité du calcul des probabilités et ses applications*" (1925-1939) y finaliza con "*Valeur pratique et philosophie des probabilités*" (1939). Muchas de sus propuestas alcanzaron su culminación en £omnicki (1923) (ver Knobloch (1987)).

ley fuerte de los grandes números (obtenida como una propiedad de los números reales)[135], Borel logró formalizar esta relación –en la que el infinito juega un rol esencial- entre la teoría de la medida, la interpretación geométrica de la probabilidad y el concepto de independencia estocástica en las sucesiones de eventos repetibles que constituye el núcleo de la teoría clásica de la probabilidad[136].

No obstante, si bien en su "*Éléments de la théorie des probabilités*" (1909b) Borel insiste en la interpretación geométrica de la probabilidad, en su propuesta deja de lado la medida de conjuntos y proporciona una demostración (imperfecta) basada en generalizaciones a conjuntos numerables de los teoremas de la probabilidad total y de la probabilidad compuesta. Esta decisión puede ser atribuida a razones filosóficas relacionadas con la imposibilidad de aceptar en forma absoluta su categoría de "*probabilidades numerables en el ámbito de la teoría de la medida*", la cual lo obligó a emplear argumentos que consideraban la derivación de las probabilidades como límites de sucesiones de repeticiones independientes.

Como auténtico sucesor intelectual de Bertrand (1888) y Poincaré (1896) (1902)(1905)(1908) y miembro de la escuela probabilística francesa, Borel asimiló los conceptos de indeterminismo e impredictibiidad, postulando que el indeterminismo estaba asociado a la imposibilidad de predecir con certeza. Consideró que la única aproximación al concepto de probabilidad era mediante una ecléctica interpretación pluralista que admitía una noción surgida de la interpretración de azar como sinónimo de ignorancia yuna noción frecuencista –basada en el principio de estabilidad estadística en el

---

[135] Borel fue el primero en considerar a las probabilidades de eventos compuestos como sucesiones infinitas de eventos simples. Este análisis de los problemas de distribuciones en sucesiones aritméticas atrajo la atención de especialistas en teoría de números, como Hardy y Littlewood y en teoría de conjuntos, como Hausdorff.

[136] Kolmogorov (1933) estableció que la diferencia entre la probabilidad y la teoría de la medida radica en la condición de independencia. Este postulado fundamental puede ser considerado como el punto de partida "*...hacia un consenso significativo acerca de los aspectos matemáticos de la probabilidad (...) en tanto que su interpretación filosófica transcurre hacia la alternativa de ser epistémica u objetiva"* (de Schewemaekere; Szafarz (2008, p. 2)).

comportamiento de los fenómenos[137]- y que esta interpretación determinística de la probabilidad no implicaba una definición, sino un intento de verificación de la interpretación de la probabilidad como límite de una frecuencia relativa. Una posición que lo condujo a su ley fuerte de los grandes números[138].

En su ensayo de 1924 sobre el "*Treatise on probability*" Borel adoptó una interpretación Keynesiana dela evaluación de la probabilidad a partir de un conjunto de información y la consideración de una cuasi aceptación de un grado de subjetivismo en casos relacionados con apuestas. Esta posición condujo a algunos autores a enrolarlo en un subjetivismo deFinettiano. Una interpretación inaceptable si se tiene en cuenta que no asume la posibilidad de que se verifiquen variaciones en la asignación de las probabilidadesde acuerdo con lo que hoy se conoce como "*coherencia temporal*" o "*condicionamiento Bayesiano*"[139] .

El objetivo de la introducción del concepto de "*probabilidad numerable*" fue desarrollar una teoría (como una propiedad de los números reales) que abandonara la concepción del continuo "*...como una noción de cardinalidad mayor que la del numerable*" (p. 248) y considerar "*...a la cardinalidad de los conjuntos numerables como la única que se puede conocer en forma positiva (...) que el continuo ha sido un instrumento transitorio cuya utilidad no puede considerarse como despreciable pero que, con el tiempo, será considerada exclusivamente como un argumento para el estudio de los*

[137] Borel le asignó a la teoría de la probabilidad un valor científico –aplicable especialmente a la física y a la mecánica estadística- y un valor práctico, que aplicó al cálculo de la utilidad esperada en un juego equitativo. La noción de probabilidad objetiva inspirada en la física estadística fue introducid por el filósofo Johannes von Kries. En su *"Die Principien de Wahrscheinlichkeits-Rechnung"* (1866) planteó el tratamiento de los juegos de azar como un sistema mecánico el cual, a partir de un conjunto dado de condiciones iniciales, conducía al estado final del proceso, asmilando la idea de azar al de inestabilidad del sistema.

[138] Si se considera a la ergodicidad como una noción probabilística, como se verá en la Sec. 22, la ley fuerte de los grandes números puede ser considerada como el primer teorema ergódico.

[139] *"La aceptación de un carácter subjetivo de la probabilidad implica la imposibilidad de precisar o mejorar las imprecisiones e imperfecciones de su definición; una modificación en el sistema de conocimiento (...) genera una variación en la asignación de la probabilidad que implica su sustitución por otra probabilidad mediante un simple proceso de reflexión"*(p. 2174).

conjuntos numerables, los cuales constituyenla única realidad a la cual se puede acceder" (p. 248). Distinguió tres categorías de probabilidades numerables: una relacionada con un número finito de resultados posibles en una sucesión infinita de repeticiones, otra relacionada con un número infinito de resultados en una sucesión finita de repeticiones y una tercera relacionada con un número infinito de resultados en una sucesión infinita de repeticiones.

A partir de la definición de unasucesión binomial, $E_1, E_2, \dots, E_n$, ..., tal que la probabilidad de ocurrencia de un "éxito" en la n-ésima repetición sea $p_n = \frac{1}{2}(n = 1,2, \dots)$, Borel logró una vinculación entre sus probabilidades numerables y la teoría de la medida asociándolas, en base a una interpretación geométrica, al intervalo $[0,1]$ y, en base a una interpretación lógica a las sucesiones infinitas de repeticiones ($\times_{j=1}^{\infty} \{0,1\}_j$). Una evidencia del rechazo de Borel a asimilar en forma directa su probabilidad numerable a la teoría de la medida en el intervalo unitario, radica en su esfuerzo por evitar en sus desarrollos los conceptos de aditividad y sub-aditividad numerable y en su conjetura que los conjuntos de probabilidad nula no son necesariamente vacíos. Consideraciones que permiten concluir que, en términos generales, para Borel la probabilidad numerable constituye una teoría intermedia entre la probabilidad finita o combinatoria y la probabilidad geométrica o continua[140].

Sea una sucesión binomial infinita y sea $0 \leq p_n \leq 1$ la probabilidad de obtener un "éxito" en la n-ésima repetición y $p\big(E_{(n)}\big)(n = 0,1,2, \dots)$ la probabilidad de obtener n "éxitos". De acuerdo con el teorema de la probabilidad compuesta, la probabilidad de no obtener ningún "éxito" será

---

[140] Borel (1909a): "*En los problemas de probabilidades, en general se distinguen dos categorías,de acuerdo a si el número de casos posibles es finito o infinito: la primera define las llamadas 'probabilidades discontinuas' o probabilidades en un dominio discontinuo, en tanto que la segunda categoría comprende a las 'probabilidades continuas' o 'geométricas'. Tal clasificación resulta incompleta cuando se hace referencia a los resultados obtenidos en la teoría de conjuntos. Entre la cardinalidad de los conjuntos finitos y la de los continuos existe la cardinalidad de los conjuntos numerables. Mi propósito es demostrar brevemente el interés asociado a cuestiones de probabilidad en las cuales intervienen dichos conjuntos, a las cuales, por razones de brevedad las voy a denominar como 'probabilidades numerables'. Antes de definir más precisamente a la probabilidad numerable, deseo señalar en pocas palabras las razones que me animan para recomendar su estudio. Principalmente la importancia de la noción de conjunto numerable, la cual no fue tomada en cuenta por ningún matemático y que estimo que es mayor de lo que se cree*" (pp. 247-248).

$p\big(E(0)\big) = \prod_{j=1}^{\infty}\big(1 - p_j\big)$ (suponiendo que ningún $p_j = 1$), de la misma forma, será:

$$p\big(E(1)\big) = \sum_{j=1}^{\infty} p\big(E(0)\big)\frac{p_j}{1 - p_j} = \sum_{j=1}^{\infty} p\big(E(0)\big)u_j$$

$$\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots$$

$$p\big(E(k)\big) = \sum_{1<i_1<i_2<\cdots<i_k<\infty} p\big(E(0)\big)u_{i_1}u_{i_2}\dots u_{i_k}$$

Borel demostró que: **i)** si $\sum_{j=1}^{\infty} p_j < \infty$, entonces $0 < p\big(E(k)\big) < 1$ $(k = 0,1,2,\dots)$ y $S = \sum_{k=0}^{\infty} p\big(E(k)\big) = 1$; **ii)** si $\sum_{j=1}^{\infty} p_j = \infty$, entonces $S = 0$; **iii)** si $\sum_{j=1}^{\infty} p_j < \infty$, entonces $p\big(E(\infty)\big) = 0$ (es decir que, con probabilidad 1, sólo puede ocurrir un número finito de eventos) y **iv)** si $\sum_{j=1}^{\infty} p_j = \infty$, entonces $p\big(E(\infty)\big) = 1$ (es decir que con probabilidad 1, ocurrirá un número infinito de eventos), resultado conocido como la "*ley cero-uno de Borel*"[141].

Teniendo en cuenta que el "caso cero" de la ley cero-uno fue obtenido en forma independiente de cualquier consideración geométrica y que la independencia numerable constituye una noción fundamental, se puede concluir que en Borel (1909b) la interpretación lógica superó a la interpretación geométrica.

La ley cero-uno ha sido objeto de sucesivas generalizaciones. En particular, Cantelli (1917a)(1917b) demostró que, dada una sucesión de eventos $\{E_j\}$ (no necesariamente independientes) tales que $\sum_{j=1}^{\infty} p_j < \infty$, se puede concluir que:

[141] Obsérvese que la alternativa "*uno*" no es solamente la solución del problema de la convergencia de variables aleatorias, sino una versión de la ley fuerte de los grandes números.

$$p\big(E(\infty)\big) = p\left(\lim_{j\to\infty}\bigcup_{i=j}^{\infty} E_j\right) = \lim_{j\to\infty} p\left(\bigcup_{i=j}^{\infty} E_j\right) \leq \lim_{j\to\infty}\sum_{i=j}^{\infty} p_j = 0$$

(donde $E(\infty) = limsup_{j\to\infty}\big(E_j\big) = \lim_{j\to\infty} \bigcup_{i=j}^{\infty} E_j = \bigcap_{j=1}^{\infty} \bigcup_{i=j}^{\infty} E_j$, es decir que los eventos E_j son de ocurrencia infinitamente frecuente, $[E_n(\varepsilon)\ i.f.]$). Resultado conocido como "*primer lema de Borel-Cantelli*".

Asimismo, dada una sucesión de eventos independientes $\{E_j\}$ tales que $\sum_{j=1}^{\infty} p_j = \infty$, resulta que:

$$p\big(\bar{E}(\infty)\big) = p\left(\lim_{j\to\infty}\bigcap_{i=j}^{\infty} \bar{E}_j\right) = \lim_{j\to\infty} p\left(\bigcap_{i=j}^{\infty} \bar{E}_j\right)$$

(donde $\bar{E}(\infty) = \bigcup_{j=1}^{\infty} \bigcap_{i=j}^{\infty} \bar{E}_j$). Por otra parte, para todo $n > j$, será $\bigcap_{i=j}^{\infty} E_i \subset \bigcap_{i=j}^{n} \bar{E}_i$, de modo que:

$$p\left(\bigcap_{i=j}^{\infty} \bar{E}_i\right) \leq \lim_{n\to\infty} p\left(\bigcap_{i=j}^{n} \bar{E}_j\right) = \lim_{n\to\infty}\prod_{=j}^{n}\big(1 - p_j\big)$$

Por lo tanto, será $p\big(\bigcap_{i=j}^{\infty} \bar{E}_j\big) \leq \lim_{n\to\infty} exp\big[-\sum_{i=j}^{\infty} p_j\big]$ y, dado el supuesto que $\sum_{j=1}^{\infty} p_j = \infty$, queda demostrado que $p\big(\bigcap_{i=j}^{\infty} \bar{E}_i\big) = 0$, para todo j y, por lo tanto, que $p\big(\bar{E}(\infty)\big) = 0$ o, lo que es lo mismo, que $p\big(E(\infty)\big) = 1$. Resultado conocido como "*segundo lema de Borel-Cantelli*"[142].

[142] Dada una sucesión de eventos $\{E_j\}$ tales que: **i)** $\sum_{j=1}^{\infty} p_j = +\infty$ y **ii)** $liminf_{n\to\infty} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} p(E_i \cap E_j)}{\left(\sum_{j=1}^{n} p_j\right)^2} = 1$, Erdös; Renyi (1959) demostraron que $p\big(E(\infty)\big) = 1$. Lo que permite concluir que los supuestos que constituyen el fundamento del segundo lema de Borel-Cantelli pueden ser flexibilizados mediante la sustitución de la condición de independencia por la condición de independencia de a pares de los eventos E_j (ver Renyi (1962)(1970)).

Como corolario de los lemas anteriores resulta la siguiente expresión final de la ley cero-uno de Borel: Dada una sucesión $\{E_n\}$ de eventos independientes, $p\big(E(\infty)\big)$ sólo puede asumir los valores 0 ó 1.

Generalizando los resultados precedentes, sea $\{E_j\}$ una sucesión de eventos independientes y sea $\mathcal{B}^{(n)} = \mathcal{B}(E_{n+1}, E_{n+2}, \dots)$ la menor álgebra que contiene a todos los eventos E_j (para $j \geq n+1$). Entonces, dado un evento E^*, que pertenece a $\mathcal{B}^{(n)}$ para todo n, se demuestra que $p(E^*)$ es igual a 0 ó a 1 (este resultado es conocido como la "*ley cero-uno de Kolmogorov*"). Como un resultado subsidiario de la ley cero-uno y sus generalizaciones se puede concluir, entonces, que existen eventos cuya probabilidad de ocurrencia sólo puede ser iguala 0 ó a 1.

La propiedad de independencia de las repeticiones debe ser interpretada como de "*independencia numerable*" según la cual, dada una sucesión de eventos, $E_1, E_2, \dots, E_n, \dots$, el teorema de la probabilidad compuesta se verifica aún para una sucesión infinita, $p\big(\bigcap_{j=1}^{\infty} E_j\big) = \prod_{j=1}^{\infty} p\big(E_j\big)$[143]. Borel aplicó estos resultados a desarrollos decimales y al desarrollo de fracciones continuas de un número irracional seleccionado al azar en un intervalo $[0,1]$. En particular, en lo que se refiere al desarrollo binario de un número real Y seleccionado al azar en el intervalo $[0,1]$, $Y = 0, b_1 b_2 \dots b_n \dots = \sum_{n=1}^{\infty} \frac{b_n}{2^n}$ (donde $b_n = b_n(y)$ es igual a 0 ó a 1), a partir del supuesto que la probabilidad de ocurrencia de cada dígito binario, $b_n(y)$, es igual a $\frac{1}{2}$ y que el subíndice representa el orden de las repeticiones independientes, sean: **i)** una sucesión λ_n tal que $\lim_{n\to\infty} \frac{\lambda_n}{\sqrt{n}} = 0$ y **ii)** $a_n(y)$ la variable que representa el número de resultados "1" en los primeros n dígitos binarios, $b_1(y), b_2(y), \dots, b_n(y)$. Considerando como "éxito", para cada repetición, el cumplimientode la condición $|a_{2n}(y) - n| \geq \lambda_n \sqrt{n}$ y, en consecuencia, como "fracaso", la condición $|a_{2n}(y) - n| < \lambda_n \sqrt{n}$, demostró que la probabilidad $p(E_n)$ de que ocurra un resultado "éxito" en una repetición converge a una distribución Normal del tipo:

[143] Un corolario de esta condición –considerada como esencial en la teoría de la probabilidad numerable- es el principio, que se debería denominar de "*independencia numerable limitada*", según el cual esta igualdad se verifica aún si $p\big(E_j\big) = 1$.

$$p(E_n) = \frac{2}{\sqrt{\pi}} \int_{\lambda_n}^{\infty} e^{-u^2} du$$

De modo que, si λ_n crece suficientemente rápido, la ley cero-uno permite concluir que $p\big(E(\infty)\big) = 0$. Por otra parte, si el resultado de la n-ésima repetición no es "éxito", la variable que representa la relación entre el número de resultados "0" y el número de resultados "1" al cabo de la primeras n repeticiones estará incluida en el intervalo $\left[\frac{n-\lambda_n\sqrt{n}}{n+\lambda_n\sqrt{n}}, \frac{n+\lambda_n\sqrt{n}}{n-\lambda_n\sqrt{n}}\right]$ o, lo que es lo mismo, en el intervalo $\left[\frac{1-\lambda_n/\sqrt{n}}{1+\lambda_n/\sqrt{n}}, \frac{1+\lambda_n\sqrt{n}}{1-\lambda_n\sqrt{n}}\right]$. Si existe sólo un número finito de casos no incluidos en este intervalo, es decir, si existe un número finito de ocurrencias del tipo $|a_{2n}(y) - n| \geq \lambda_n\sqrt{n}$, entonces el cociente entre el número de resultados "0" y el número de resultados "1" tenderá a 1 cuando$n \to \infty$ y dado que, como ya se demostró, $p\big(E(\infty)\big) = 0$, el conjunto $E^*(\infty)$, formado por los números Y cuyo desarrollo binario posea asintóticamente el mismo número de resultados "0" y de resultados "1", será de probabilidad unitaria. Este resultado se conoce como "*ley fuerte de los grandes números de Borel*" (denominación introducida por Khinchin (1929a))[144].

---

[144] Borel denominó como *"simplemente normales en base 10"* a aquellos números cuyo desarrollo decimal está formado por el conjunto de los dígitos, $\{0,1,2,\dots,9\}$, de probabilidad 1, a cada uno de cuyos elementos le corresponde una frecuencia relativa límite igual a $\frac{1}{10}$ y, en términos más generales, como *"simplemente normales en base q"* a aquellos números que admiten un desarrollo formado por el conjunto $x = \sum_{j=1}^{\infty} \frac{b_j}{q^j}$ $\big(b_j = 0,1,\dots,q-1\big)$, de probabilidad 1, a cada uno de cuyos elementos le corresponde una frecuencia relativa límite igual a $\lim_{n\to\infty} \frac{\text{número de ocurrencias de } i \text{ entre } b_1(x),\dots,b_n(x)}{n} = \frac{1}{q}$ $(i = 0,1,2,\dots,q-1)$. Se denominan como *"enteramente normales en base q"* a aquellos números que son simplemente normales en cada una de las bases $q, q^2, \dots, q^j, \dots$ y como *"absolutamente normales en base q"* a aquellos números que son simplemente normales para todo $q = 2,3,\dots$. Denotando por N_q^e y $N_{q^j}^a$ a los conjuntos de números enteramente normales y absolutamente normales (ambos de probabilidad 1), respectivamente, se verifica que $N = \bigcap_{j=1}^{\infty} N_{q^j}^a$ y $N = \bigcap_{q=2}^{\infty} N_q^e$, de modo que $p\left(\bigcap_{j=1}^{\infty} N_{q^j}^a\right) = \prod_{j=1}^{\infty} p\left(N_{q^j}^a\right)$ y $p\big(\bigcap_{q=2}^{\infty} N_q^e\big) = \prod_{q=2}^{\infty} p\big(N_q^e\big)$. Expresiones que representan la condición de independencia numerable limitada (ver Niven (1956), Kac (1959), Kuipers, Niederreiter (1974)).

Curiosamente Borel no plantea ninguna vinculación entre su resultado y la ley de los grandes números de Bernoulli (dela cual puede considerarse un corolario), lo cual hubiera permitido un análisis comparativo entre la convergencia "en-medida" y la convergencia "en casi todos los casos"[145]. Los trabajos inmediatamente posteriores a los de Borel pueden ser clasificados en: **i)** los orientados hacia una interpretación probabilística (que consideran a la ley fuerte de los grandes números como un refinamiento del teorema de Bernoulli) y **ii)** los orientados hacia la teoría de la medida (que consideran quela comparación no es posible, debido a que el teorema de Bernoulli no admite una interpretación en términos de teoría de la medida)[146].

Contrariamente a los postulados de la ley débil de los grandes números, que no establece una convergencia, sino la probabilidad de una aproximación, laley fuerte de los grandes números demuestra que la frecuencia relativa converge a un límite en el sentido del análisis, si se excluye un conjunto de casos con probabilidad nula. Como corolario resulta que, contrariamente a lo que ocurre en la versión débil, en la que los límites ε y δ permiten determinar el valor de n, el único resultado de la versión fuerte consiste en la demostración de la existencia de un número límite (expresado como una propiedad de los reales, según la cual la medida del conjunto de decimales binarios con una frecuencia límite de resultados "1" diferente de $\frac{1}{2}$ es igual a cero)[147]: $\mathrm{p}\left[\lim_{\mathrm{n}\to\infty}\left(\mathrm{Y_n}-\frac{1}{2}\right)=0\right]=1$ (es decir, que $X-\frac{n}{2}=o(n)$ o que $\{X_j\}$ converge casi-con-certeza a $\frac{1}{2}$, $\{X_j\}\xrightarrow{ccc}\frac{1}{2}$)[148].

---

[145] Este análisis comparativo fue desarrollado por Cantelli (1917a), Pólya (1920), Slutsky (1925), Fréchet (1930).

[146] Es posible reconocer un antecedente de la ley fuerte de los grandes números de Borel como corolario de la teoría de las probabilidades numerables en van Vleck (1908).

[147] Esta propiedad de los números reales fue denominada por Steinhaus (1923) como *"la paradoja de Borel"*.

[148] En términos generales, dada una sucesión de constantes $\{a_n\}$, se dice que una sucesión $\{X_i\}$ converge casi-con-certeza con respecto a una sucesión de constantes $\{b_n\}$ positivas y tales que $\lim_{n\to\infty} b_n=\infty$ (o, lo que es lo mismo, que cumple con la ley fuerte de los grandes números con respecto a la sucesión $\{b_n\}$), si se verifica que $p\left[\lim_{n\to\infty}\left(\frac{\sum_{j=1}^{n}X_j}{b_n}-a_n\right)=\right.$

18.- Georg Faber y la demostración rigurosa de la propuesta de Borel

Como continuación de los resultados de Borel, cabe mencionar los trabajos de Faber, Bernstein, Hausdorff y Rademacher, quienes modificaron los alcances de la relación entre la teoría de la medida y la probabilidad. En particular, Faber (1910) –a partir de la demostración que, dada una función continua $f(x)$, el conjunto de puntos x para el cual f no tiene derivada, tiene medida de Lebesgue nula- denotando por $Y(1)$ e $Y(0)$ la frecuencia de los resultados "1" y "0", respectivamente, en una sucesión de n dígitos binarios de x, demostró que, si $supinf\left(\frac{Y(1)}{Y(0)}\right) < 1 - \varepsilon$ ó $infsup\left(\frac{Y(1)}{Y(0)}\right) > 1 + \varepsilon$[149], no existe derivada y que, por lo tanto, el conjunto de puntos x para los cuales $\lim\left(\frac{Y(1)}{Y(0)}\right) = 1$ tiene medida uno. Esta puede considerarse la primera demostración analítica rigurosa del resultado de Borel mediante argumentos de teoría de la medida. Por otra parte, considerando que la relación entre la probabilidad numerable y la medida de Lebesgue era un tema inconcluso, planteó la cuestión acerca de la continuidad de la medida de Lebesgue y la probabilidad y, viceversa, la continuidad de ésta respecto de aquélla; es decir, la cuestión acerca de la posibilidad de asegurar que la probabilidad de un punto incluido en un conjunto de medida nula siempres

$0\Big] = 1$ (debe tenerse en cuenta que, cuanto menor sea el orden de magnitud de b_n, más precisa será la convergencia). Sustituyendo en esta expresión la condición de convergencia casi-con-certeza por la de convergencia en-probabilidad, se concluye en forma inmediata que $\xrightarrow{ccc}\Rightarrow\xrightarrow{p}$ con respecto a la sucesión $\{b_n\}$, pero que la implicación inversa no tiene por qué verificarse. En otros términos, se demuestra que si una sucesión $\{X_i\}$ cumple con la ley fuerte de los grandes números, entonces también cumple con la ley débil de los grandes números. En particular, dada una sucesión $\{X_1, X_2, \ldots, X_n\}$ de variables aleatorias independientes, se demuestra que, para la sucesión $\sum_{j=1}^{n} X_j$, las convergencias en-distribución, en-probabilidad y casi-con-certeza coinciden.

[149] Calcular el $supinf\left(\frac{Y(1)}{Y(0)}\right)$ implica: **i)** tomar, para cada número natural k, la mayor cota inferior a_k para $k = n+1, n+2, \ldots$ y **ii)** para $k \to \infty$, tomar la menor cota superior de la sucesión $\{a_k\}$. Viceversa, calcular el $infsup\left(\frac{Y(1)}{Y(0)}\right)$ implica: **i)** tomar, para cada número natural k, la menor cota superior b_k para $k = n+1, n+2, \ldots$ y **ii)** para $k \to \infty$, tomar la mayor cota inferior de la sucesión $\{b_k\}$.

es igual a cero y, viceversa, que un conjunto puede ser considerado de medida nula, si la probabilidad de un punto cualquiera de dicho conjunto siempre es igual a cero[150].

Por su parte, F. Bernstein (1912) propuso una definición (arbitraria) de probabilidad como medida de Lebesgue y Rademacher (1922) concluyó en la compatibilidad natural de las nociones de probabilidad y medida[151].

19.- Felix Hausdorff y el refinamiento de la ley fuerte de los grandes números de Borel

Hausdorff (1914), contrariamente a la posición de Borel, trató a la ley fuerte de los grandes números y al resultado sobre fracciones continuas exclusivamente como ejemplos de la teoría de la medida:Sea el conjunto de los números irracionales x, cuyo desarrollo binario está formado por los siguientes primeros n dígitos: $b_1, b_2, \dots, b_n$ que, en consecuencia, están incluidos en el intervalo:

$$\frac{b_1}{2}+\frac{b_2}{2^2}+\cdots+\frac{b_n}{2^n}<x<\frac{b_1}{2}+\frac{b_2}{2^2}+\cdots+\frac{b_n+1}{2^n}$$

de medida $\frac{1}{2^n}$ (debe tenerse en cuenta que la no consideración de los números racionales, que definen un conjunto de medida nula, permite asimilar la

[150] Ver Doob (1989), von Plato (1994).

[151] Rademacher (1922) introdujo un sistema de funciones ortogonales, $\{R_i, i \geq 0\}$, de la forma $R_0(w)=\begin{cases}1 & si \;\; 0\leq w\leq \frac{1}{2}\\ -1 & si \;\; \frac{1}{2}<w\leq 1\end{cases}$ y $R_i(w)=R_0\left(2^i w\right)(i\geq 1, 0\leq w\leq 1)$, asimilables a variables aleatorias iid y tales que $p(R_i=1)=p(R_i=-1)=\frac{1}{2}$. Se demuestra fácilmente que la transformación $B_i=\frac{1-R_i}{2}(i>0)$, para cada $w\epsilon[0,1]$, $0.B_0(w), B_1(w), \dots$ define el desarrollo binario de w. Esta vinculación entre las funciones R_i y el desarrollo binario de un número real en el intervalo $[0,1]$ conduce a una versión inversa del lema de Borel-Cantelli según el cual, si los B_i representan eventos independientes, entonces $\sum_{i=1}^{n} p(B_i)=\infty$ implica que la ocurrencia infinitamente a menudo de B_i es igual a uno (para un estudio detallado de las funciones de Rademacher, ver Kolmogorov (1929), Lévy (1931), Erdös (1942), Khinchin (1950)).

medida a la longitud del intervalo). El conjunto de números x cuyo desarrollo cuenta con k ceros y $n-k$ unos en los primeros n números de su desarrollo binario se puede considerar formado por $\binom{n}{k}$ "intervalos" disjuntos de a pares, todos de medida $\frac{1}{2^n}$, de modo que su medida será $\binom{n}{k}\frac{1}{2^n}$.La medida del conjunto $E_n(\varepsilon)$de los números x para los cuales $\left|\frac{k}{n}-\frac{1}{2}\right| \geq \varepsilon$ está dada por:

$$p_n(\varepsilon) = \sum_{\left|\frac{k}{n}-\frac{1}{2}\right| \geq \varepsilon} \binom{n}{k}\frac{1}{2^n}$$

A partir de la estimación de los momentos de orden par de la variable $\frac{Y_n(x)}{n}-\frac{1}{2}$ (donde $Y_n(x)$ denota la variable que representa el número de ceros en la sucesión $\{b_1, b_2, \dots, b_n\}$ correspondiente al desarrollo binario de x), Hausdorff demostró que el momento de cuarto orden satisface la relación[152]:

$$\sum_{k=0}^{n}\left(\frac{k}{n}-\frac{1}{2}\right)^4\binom{n}{k}\frac{1}{2^n} < \frac{3}{16}\frac{1}{n^2}$$

y, por lo tanto, la relación:

$$\varepsilon^4 \sum_{\left|\frac{k}{n}-\frac{1}{2}\right| \geq \varepsilon} \binom{n}{k}\frac{1}{2^n} < \sum_{\left|\frac{k}{n}-\frac{1}{2}\right| \geq \varepsilon}\left(\frac{k}{n}-\frac{1}{2}\right)^4\binom{n}{k}\frac{1}{2^n} < \frac{3}{16}\frac{1}{n^2}$$

152 El primer miembro de la desigualdad puede ser denotado como $E\left[\left(\frac{x_1+x_2+\cdots+x_n}{n}-\frac{1}{2}\right)^4\right]$, donde las x_i son variables aleatorias iid con distribución $b\left(1,\frac{1}{2}\right)$. En general, Hausdorff demostró cómo obtener estimaciones de los momentos de la forma $\sum_{j=0}^{n}(2j-n)^{2s}\binom{n}{j}\frac{1}{2^j} < A_{2s}n^s$ y, por lo tanto, de la forma $\sum_{j=0}^{n}(2j-n)^{2s}\binom{n}{j}\frac{1}{2^j} < \frac{A_{2s}}{n^s}\frac{1}{2^s}$ y, en particular, $\varepsilon^{2s}\sum_{\left|\frac{j}{n}-\frac{1}{2}\right| \geq \varepsilon}\binom{n}{j}\frac{1}{2^j} < \frac{A_{2s}}{n^s}\frac{1}{2^{2s}}$.

Luego, se puede concluir que $p_n(\varepsilon)$ satisface una desigualdad del tipo de Bienaymé-Chebychev, pero para momentos de cuarto orden, $p_n(\varepsilon) < \frac{3}{16}\frac{1}{\varepsilon^4}\frac{1}{n^2}$ y, en consecuencia, que $\sum_{n=1}^{\infty} p_n(\varepsilon) < \infty\ (\varepsilon > 0)$.

Ahora bien, sea:

$$E(\varepsilon) = limsup\, E_n(\varepsilon) = \bigcap_{n=1}^{\infty}\bigcup_{n=N}^{\infty} E_n(\varepsilon)$$

el conjunto de números irracionales x en el intervalo $(0,1)$ para los cuales $\text{limsup}_{n\to\infty}\left|\frac{y_n(x)}{n} - \frac{1}{2}\right| \geq \varepsilon$ (los conjuntos $\bigcup_{n=N}^{\infty} E_n(\varepsilon)$ decrecen a medida que N crece y, obviamente, $E(\varepsilon) \subset \bigcup_{n=N}^{\infty} E_n(\varepsilon)$) y sea $A_\infty(\varepsilon) = P\big(E(\varepsilon)\big) = medida\ de\ E(\varepsilon)$. Luego, para cada N y, como corolario de las condiciones de monotonicidad y sub-aditividad, será:

$$A_\infty(\varepsilon) \leq P\left(\bigcup_{n=N}^{\infty} E_n(\varepsilon)\right) \leq \sum_{n=N}^{\infty} P\big(E_n(\varepsilon)\big) = \sum_{n=N}^{\infty} p_n(\varepsilon)$$

Y, teniendo en cuenta que $\sum_{n=N}^{\infty} p_n(\varepsilon) < \infty$ (para todo $\varepsilon > 0$), se puede concluir que $A_\infty(\varepsilon) = 0$ (para todo $\varepsilon > 0$). En particular, $\bigcup_{j=1}^{\infty} E\left(\frac{1}{j}\right)$ define una unión numerable de conjuntos de medida nula que, en consecuencia, es de medida nula. Pero, dado que $\bigcup_{j=1}^{\infty} E\left(\frac{1}{j}\right)$ es un conjunto de x tal que:

$$\lim_{n\to\infty} sup\left|\frac{y_n(x)}{n} - \frac{1}{2}\right| > 0$$

Si $\sum p_n(\varepsilon) < \infty$ para todo $\varepsilon > 0$, entonces:

$$p\left\{\lim_{n\to\infty} sup\left|\frac{y_n(x)}{n} - \frac{1}{2}\right| = 0\right\} = 1$$

o, lo que es lo mismo, $p\left\{\lim_{n\to\infty}\left|\frac{y_n(x)}{n} - \frac{1}{2}\right| = 0\right\} = 1$. Resultado conocido como la "*ley fuerte de los grandes números de Hausdorff*"[153].

Como se vio en la Sec. 17, el teorema de Borel establece que el conjunto de x de probabilidad 1 es:

$$\lim_{n\to\infty}\left(\frac{y_n(x)}{n}-\frac{1}{2}\right)=0$$

Ahora bien, el teorema de Hausdorff demuestra que el conjunto de x de probabilidad 1 es:

$$\lim_{n\to\infty}\left(\frac{y_n(x)}{n}-\frac{1}{2}\right)n^{\theta}=0\left(\theta<\frac{1}{2}\right)$$

Lo que permite concluir que el resultado de Hausdorff constituye un refinamiento de la ley fuerte de los grandes números de Borel dirigido a precisar, en la mayor medida de lo posible, la tasa para la cual se puede asegurar que $\frac{y_n(x)}{n}$ se aproxima a $\frac{1}{2}$ con probabilidad igual a 1.

Las demostraciones de Borel, Faber y Hausdorff fueron generalizadas por Cantelli (1916a)(1916b)(1917a) [154] .Posteriormente Kolmogorov (1931) (1933b) y Prokhorov (1956) establecieron las condiciones necesarias y suficientes para asegurar la validez de la ley de los grandes números. Un proceso cuya culminación fue la ley del logaritmo iterado de Khinchin.

20.- Richard Martin Elder von Mises y el "Kollektiv" Fechneriano

Las aproximaciones a una probabilidad puramente matemática expuesta en las secciones precedentes no fue universalmente aceptada, posiblemente por la idea que esta interpretación constituía una estructura vacía que limitaba el ámbito de aplicación del concepto de probabilidad respecto de una interpretación que combinara rigurosamente contextos matemáticos y no-matemáticos.

---

[153] Este resultado fundamental le permitió obtener una expresión de la ley fuerte de los grandes números sin recurrir al teorema central del límite.

[154] Los resultados obtenidos por Cantelli fueron de gran importancia en la medida que impulsaron el desarrollo de los diferentes conceptos de convergencia estocástica.

La demostración de Poisson y los trabajos de Cournot (1838)(1843), Ellis (1849), Venn (1866), el principio de asociación de ideas de Locke –consistente en igualar experiencia, creencia y probabilidad- y la reacción del empirismo británico contra el racionalismo continental de Laplace, condujeron a intentar una interpretación que conciliase la teoría de la probabilidad con los principios del indeterminismo fundado por Theodor Fechner a mediados del siglo XIX.

Fechner –a quien se reconoce como el creador de la psico-física[155] y el primer "*indeterminista universal*"[156]-, en una revisión de los principios de la mecánica clásica y, en particular, de la concepción determinística de la causalidad que rigió el pensamiento científico en el siglo XIX, desarrolló una variante del indeterminismo –objetivo y parcial- que lo llevó a postular que la naturaleza posee una cierta cantidad de aleatoriedad (indeterminación) que "*...depende de la libertad*" y no "*...de nuestra ignorancia de las condiciones*"[157], y determinó la declinación del paradigma Laplaciano (que había dominado el panorama de la ciencia durante más de medio siglo) y la reformulación de la interpretación frecuencista de la probabilidad desarrollada por los autores de los siglos XVII y XVIII[158].

Los antecedentes del pensamiento de Fechner y, en general, de los filósofos y fisiólogos alemanes de la segunda mitad del siglo XIX, se encuentran en las ideas en biología y en filosofía de la historia imperantes en Alemania en la primera mitad del siglo XIX y en su reacción contra las especulaciones

---

[155] "*...la teoría exacta de la relación funcional o de la dependencia entre el cuerpo y el alma o, en términos más generales, entre lo corporal y lo espiritual, entre los mundos físico y psíquico*" ("*Elemente der Psychophysic*" (1860, p. 8)).

[156] Heidelberger (1987, p. 117)).

[157] Fechner (1860, p. 53).

[158] Kamlah (1987):"*Cuando los axiomas de la teoría de la probabildad fueron derivados de la definición de la probabilidad como frecuencia relativa, y cuando se concluyó que los métodos para estimar cantidades físicas y distribuciones de probabilidades convertían en superfluo el principio de Bayes, entonces la teoría Laplaciana de la probabilidad murió*" (p.98).

metafísicas del tardío idealismo alemán en filosofía de la ciencia[159]. La principal consecuencia de esta reacción, en lo que hace a la teoría del azar, fue la aplicación de métodos de investigación empíricos a la psicología (debe tenerse en cuenta que, en este contexto, esta disciplina, además de asuntos como las emociones, las percepciones y las voliciones, abarcaba la formación del concepto de inferencia lógica, de modo que la probabilidad quedaba incluida en su ámbito).

En el capítulo "*Über das Causalgesetz*" de "*Beriche über die Verhandlungen der Königlich sächsischen Gesellschaft der Wissen schaften zu Leipzig*" (1849a), a fin de demostrar la compatibilidad de las leyes causales con "*...la libertad y la indeterminación*"[160], Fechner introdujo la ya mencionada noción de "*indeterminación por novedad*" basada en el siguiente principio: En la evolución de los fenómenos dinámicos se van generando nuevas condiciones iniciales las cuales, de acuerdo con una ley general, conducen a efectos no ocurridos previamente y, por lo tanto, no permiten la recurrencia de comportamientos idénticos; pero, no obstante, cada conjunto de dichas condiciones conserva alguna similitud con los anteriores por loque, si bien todo fenómeno fáctico posee cierta dosis de aleatoriedad objetiva, su comportamiento no es completamente libre[161].

---

159 De autores como Christian Hermann Weisse, Immanuel Hermann Fichte y Hermann Lotze.

160 A partir de una clase especial de neo-Spinozismo según el cual lo espiritual y lo material constituyen dos aspectos que pueden generar los mismos resultados y, por lo tanto, la libertad del espíritu también debe manifestarse en el mundo físico, Fechner (*"Zend-Advesta, oder über die Dinge des Himmels und des Jenseits"* (1851)) concluyó que, si bien lo físico y lo psíquico están estrictamente coordinados, no existe ninguna relación causal entre ellos, que las leyes que los vinculan más que de interacción o sucesión, son de coexistencia.

161 Entre las influencias de los idealistas alemanes merece una mención particular la ejercida sobre Fechner por K.F.Burdach, que se evidencia en el siguiente párrafo de su *"Die persönliche Besonderheit" (1842): "La naturaleza con sus leyes y sus fuerzas es eternamente inmutable: los elementos son siempre los mismos en todo lugar; descansando sobre el mismo fundamento aparecen fenómenos similares en todo lugar y en todo tiempo y lo que ha agitado las emociones de una persona para quien el sol brilló por primera vez, también agitará el corazón de quien viva para ver cómo el sol alumbrará la tierra en el fin de los tiempos. Pero la sensación es que siempre volverá a ocurrir algo similar, no algo idéntico. La mayoría de los elementos admite una multiplicidad de combinaciones y, a través*

A partir de 1853 y hasta el final de su vida, la obra de Fechner –contenida fundamentalmente en "*Die mathematische Behandlung organischer Gestalten Processe*" (1849) y en "*Kollektivmasslehre*" (1897)[162]- estuvo dirigida a desarrollar un lenguaje matemático-probabilístico-estadístico para la interpretación de la aleatoriedad en los fenómenos fácticos. Para ello se basó en la llamada "*estadística moral*" del astrónomo belga Adolphe Quetelet[163], en los estudios de Moritz Wilhelm Drobisch[164] sobre la relación entre dicha estadística moral y la libertad y en la teoría de los errores de observación no-sistemáticos desarrollada por los matemáticos Gauss, Encke, Bessel y Hauber.

Fechner tomó de Quetelet la idea de la analogía entre el modelo de Laplace sobre mecánica celeste y una física de la sociedad humana, es decir, de la asimilación de las leyes gravitacionales a las causas constantes que gobiernan la sociedad, en la que, al igual que las perturbaciones que afectan a los fenómenos físicos, las diferencias entre los individuos que la componen (las cuales no deben ser interpretadas como desvíos aleatorios del verdadero valor, sino como "*...la forma libre de reacción del individuo ante un estímulo*"[165]) se distribuyen de acuerdo con una función de probabilidades[166] y se vuelven despreciables si se considera un número

de la coincidencia de tales conexiones, surge un número infinito de series de circunstancias. La fuerza creativa del mundo es infinita en su esencia y, por lo tanto, es infinita en sus manifestaciones: inagotable en sus combinaciones, la naturaleza produce sólo novedades por toda la eternidad, no tolera la uniformidad ni admite que una ley gobierne con rigidez estricta, sino que atempera su severidad mediante la interferencia de otra ley" (p. 232).

[162] Fechner no llegó a concluir esta obra. A instancia de la Royal Saxony Academy of Sciences, fue completada por Gottlob Friederich Lipps.

[163] *"Sur l'homme et le développement de ses facultés"* (1835), *"Sur la statistique morale et les príncipes qui doinent en former la base"* (1848).

[164] *"Bertäge zur Orientierung über Herbart's System de Philosophie"* (1834), *"Die moralische Statistik un die menschliche Willensfeiheit"* (1867).

[165] Heidelberger (1987, p. 112).

[166] Una propuesta revolucionaria incluida en el "*Kollektivmasslehre*" con respecto a los postulados imperantes en la teoría de la aproximación estadística del siglo XIX, fue el

suficientemente grande de individuos. Estos postulados condujeron a Fechner al concepto de "*objeto colectivo*" ("*Kollektivgegenstand*") al que definió como una sucesión de individuos homogéneos que varían aleatoriamente con respecto a un atributo común[167] (en particular, un atributo cuantificable)[168], regido por leyes de naturaleza probabilística. En su expresión más simple el "*Kollektiv*" puede ser considerado como una sucesión de resultados obtenidos de una serie de observaciones repetidas en igualdad de condiciones ("*serie colectiva*" ("*Kollektivereihe*")), cada una de las cuales admite sólo dos alternativas posibles. Es decir, matemáticamente, el colectivo puede se considerado como aproximable por una serie binaria infinita[169].

---

principio de la habitualidad de las distribuciones asimétricas en los fenómenos fácticos. Debe tenerse en cuenta que, si bien en 1838 Bessel ("*Untersuchungen über die Wahrscheinlichkeit der Beobachtungsfehler*") había puntualizado que era posible hallar distribuciones de probabilidades de errores aleatorios asimétricas y que, por lo tanto, la distribución Normal no podía ser considerada válida en todos los casos, fue sólo después de la publicación de *"The gramar of science"* (1892), *"Asymetrical frequency curves"* (1893) y *"Contribution to the mathematical theory of evolution"* (1893-96) de K. Pearson, que la distribución Normal perdió su preeminencia.

167 "*Zür experimentalen Ästhetik*" (1871). Este concepto de colectivo es similar al introducido por el estadístico Gustav Rümelin en *"Zür Theorie der Statistik"*, en 1863, con el único –y fundamental- agregado de la condición de variación aleatoria.

168 La traducción literal de Kollektivmasslehre es "*doctrina de los objetos colectivos cuantificables*" y el atributo cuantificable puede ser interpretado como un punto k-dimensional en el espacio de los números reales.

169 Fechner (1897): *"Condición I: Existencia de límites.Sea A un conjunto arbitrario de puntos del espacio definido por un atributo común. Entonces, el límite de la frecuencia relativa de A en la sucesión de elementos existe: $\frac{N_A}{N} = W(A)$ (...) si se verifica la segunda condición, W_A representa la probabilidad de A en el colectivo. Nótese que la probabilidad está definida en relación a un colectivo. Luego, las probabilidades son números comprendidos entre 0 y 1 que cumplen la condición de aditividad finita. Los valores de probabilidad W_A para todo punto del conjunto A, definen la 'distribución' de un colectivo. Nótese, además, que la distribución es una función definida sobre los subconjuntos del espacio caracterizado por el atributo común. Este es un concepto más general que el de las distribuciones de probabilidades de la teoría clásica. Las cuales constituyen un caso particular cuando A consta de un punto"* (p. 5). *"Condición II: Irregularidad en el ordenamiento. Sean A y B dos conjuntos disjuntos, con probabilidades no-nulas W_A y W_B en un colectivo K. Si se descartan todos los elementos de K cuyo atributo común no*

El concepto de "*Kollektivgegenstand*" experimentó un resurgimiento con el florecimiento del empirismo desarrollado por el Círculo de Viena en la obra, entre otros, del filósofo y psicólogo G.F. Lipps[170] y los astrónomos Georg Helm[171] y Heinrich Bruns[172], quienes fueron los primeros en postular que la teoría de los objetos colectivos no era sino una versión de la teoría del azar, con una probabilidad definida como una frecuencia relativa (inevitablemente condicionada por la definición del colectivo) y tuvo su culminación en la reformulación realizada por Reichenbach[173] y von Mises del concepto de probabilidad[174] "*... a fin de reemplazar o complementar la rígida estructura causal de la teoría clásica*" (von Mises (1928, p. 427)).

Como un intento de extensión de esta definición frecuencista de la probabilidad surgen las propuestas de la interpretación logicista, en la que la probabilidad es considerada como una rama de la lógica, como una extensión de la lógica deductiva al ámbito inductivo y de la interpretación subjetivista, en la que se la considera asociada a grados de creencia de observadores individuales.

---

pertenece a A ni a B y se construye una sub-sucesión formada sin utilizar las diferencias en los atributos comunes, entonces los límites $W_{A'}$ y $W_{B'}$ de la sub-sucesión existen y son tales que $\frac{W_{A'}}{W_{B'}}=\frac{W_A}{W_B}$" (p. 58).

170 "*Über Fechner's collectivmasslehre und vie Vertheilungsgesetze der Collectivgegenstände*" (1898), *"Die Theorie der collectivgegenstände"* (1901), *"Die Betimmung der Abhängigkeit zwischen der Merkmalen eines Gegenstándes"* (1805).
171 *"Die Wahrscheinlichkeitslehre als Theorie der Kollektivbegriffe"* (1902).

172 *"Über die Darstellung von Fehlengesetzen"* (1897), *"Zür collective-Masslehre"* (1898), *"Warscheinlichkeitsrechnung und Kollektivmasslehre"* (1905), *"Das Gruppenschema für zafällige Ereignisse"* (1906).

173 "*Axiomatik der Wahrscheinlichkeitsrechnung*" (1932), "*Wahrscheinlichkeitslehre*" (1935).

174 *"Über die Grundbegriffe der Kollektivmasslehre"* (1912), *"Fundametalsatze der Wahrscheinlichkeitsrechnung"* (1919), *"Grundlagen der Wahrscheinlichkeitsrechnung"* (1919), *"Wahrscheinlichkeit, Statistik und Wahrheit"* (1928). En esta última obra von Mises realiza una exposición de su teoría en términos no matemáticos, un notable análisis de sus consecuencias filosóficas y una recopilación de sus últimas reflexiones sobre la teoría de la probabilidad.

La propuesta de la denominada interpretación logicista condujo a un modelo en el que la noción general de probabilidad (la cual se traduce en un grado de creencia racional o idea similar acerca de la ocurrencia de un fenómeno determinado) es función –exclusivamente- de un cierto estado de conocimiento definido por un conjunto de argumentos, intrínsecos y extrínsecos a dicho fenómeno, que el observador adquiere a partir de la percepción de una relación lógica entre proposiciones[175]. Una probabilidad, $p(A \,/\, B)$, concebida como una relación (indefinida) entre una proposición, A, y un cuerpo de conocimiento, B, entre una *"...afirmación y otra afirmación (o conjunto de afirmaciones)que representa la evidencia"* (Kyburg; Smokler (1980, p. 13), condicionada por la verdad de dicha evidencia. En la que el evento A puede, en consecuencia, ser representado mediante un subconjunto $A \subset \Omega$ tal que $A = \{w/S(w)\ \ es\ verdadera\}$, de modo que a cada evento le corresponde un único conjunto A y viceversa, es decir, a cada $S(w)$ del espacio proposicional le corresponde un conjunto A en el espacio de eventos.

La interpretación logicista se basó en los aportes realizados por Augustus de Morgan, John Venn, Harold Jeffrey y, en particular, por John Maynard Keynes, continuados por los miembros del Círculo de Viena, Bernard Bolzano, Ludwig Wittgenstein, Frederich Waismann y, en particular, Rudolf Carnap y Karl Popper.

El punto de partida de la aproximación de Keynesiana consistió, precisamente, en definir una teoría del vínculo parcial como una generalización de la teoría del vínculo total de la lógica deductiva y considerar a la probabilidad como una evaluación de ese vínculo parcial, de modo que carece de sentido considerar la probabilidad de una hipótesis, que sólo es admisible la consideración de su probabilidad condicionada por una cierta evidencia vinculada parcialmente con ella. Luego, dado un conjunto h de proposiciones y una conclusión consistente en un conjunto de proposiciones a, si h implica parcialmente a a en un grado α entonces, identificando los grados de vínculo parcial con los grados de creencia racional, Keynes concluyó que, dado h, existirá un grado α de creencia racional en a, es decir, una relación de probabilidad de grado α entre a y h.

[175] Ramsey (1926): *"De acuerdo con esta interpretación, la teoría de la probabilidad es considerada como una rama de la lógica, la lógica de la creencia parcial y del argumento no-concluyente"* (p. 157).

Obsérvese que Keynes asimiló su modelo probabilístico a un grado de creencia racional no simplemente a un grado de creencia individual. Es decir, consideró a las probabilidades como valores fijados objetivamente por el observador, los cuales son asimilables a relaciones lógicas conocidas por intuición pero utilizando un concepto Platónico del término "objetivo", es decir, no referido a "cosas" del mundo material, sino a "algo" en un supuesto mundo Platónico formado por ideas abstractas, similar al prostulado por los filósofos de Cambridge, que incluía ideas objetivas, cualidades éticas (con la idea de la virtud ocupando un lugar prominente) y entes matemáticos.

De acuerdo con la corriente empirista y en contraste con los principio en los que se basa la interpretación logicista, en la interpretación frecuencista de von Mises la teoría de la probabilidad es considerada como una rama de la matemática y de la física teórica (comparable a la mecánica clásica)[176].

Con respecto a los fenómenos repetibles que constituyen la razón de ser de esta interpretación, von Mises reconoce tres categorías: **i)** los juegos de azar, **ii)** los fenómenos relacionados con estadísticas biológicas y **iii)** ciertos fenómenos vinculados con la física de campo. Como se mencionó más arriba, asociado a cada uno de estos fenómenos es posible definir un conjunto de atributos considerados "a priori" como posibles y que forman lo que von Mises denominó el "*espacio de atributos*" (ω)[177] (estrictamente hablando el conjunto ω está formado por atributos elementales, de modo que es posible particionarlo en subconjuntos tales que cada uno defina un atributo o resultado posible).

En cuanto a los colectivos, von Mises (1928) considera la necesidad de distinguir entre "*colectivos empíricos*" (que pueden ser hallados en el mundo real que, por lo tanto, son observables y que, en consecuencia, están formados por un número finito de elementos)[178] y "*colectivos matemáticos*" (formados

---

[176] Von Mises (1928): *"La nueva idea, surgida alrededor de 1919 (como una extensión anticipada por A.A. Cournot en Francia, John Venn en Inglaterra y Georg Helm en Alemania) consistió esencialmente en considerar a la teoría de la probabilidad como una ciencia del mismo orden que la geometría y la mecánica teórica"* (p. vvi).

[177] Esta denominación fue sustituida con el tiempo por la menos afortunada de "*espacio muestral*".

[178] Que dan origen a lo que algunos autores han denominado "*frecuencismo finito*".

por una sucesión infinita de elementos)[179].Basándose en una circunstancia que se verifica habitualmente en la física y considerando que las sucesiones infinitas son abstracciones matemáticas o idealizaciones necesarias para obtener un representación matemática admisible de la realidad, von Mises estableció el principio (muy discutible) según el cual un colectivo matemáticoinfinito puede ser expresado en términos analíticos por un colectivo empírico finito (convirtiendo a la teoría de la probabilidad en una teoría de convergencia de sucesiones)[180]. Debe recordarse que von Mises fue un empirista y que su análisis sebasó siempre en una filosofía operacionalista según la cual los principios teóricos de una ciencia deben ser definidos en términos de fenómenos observables con las características de un colectivo empírico. Según esta interpretación operacionalista la naturaleza de los fenómenos repetibles es tal que: **i)** es posible, por abstracción, obtener ciertos conceptos matemáticos que permiten formular las leyes empíricas que rigen su comportamiento; **ii)** recurriendo nuevamente a la abstracción y, a partir de dichas leyes empíricas, es posible definir los axiomas de la teoría matemática asociada a dicho comportamiento y **iii)** a partir de esta teoría matemática es posible descubrir consecuencias que permiten la explicación y predicción de otros fenómenos repetibles[181].

De acuerdo con von Mises (1928), los colectivos empíricos obedecen a dos principios fundamentales: **i)** la "*ley de estabilidad de las frecuencias estadísticas*"[182] y **ii)** la "*ley de irregularidad*", que postula la aleatoriedad absoluta en la proyección de los atributos sobre los elementos del colectivo. A partir de los cuales, mediante un proceso de abstracción (o idealización), definió los axiomas de convergencia y de aleatoriedad ("*Regellosigkeitsaxiom*") aplicables a colectivos matemáticos de la forma

---

[179] Que dan origen al llamado "*frecuencismo hipotético*".

[180] Este principio se basa en el concepto que*"Una sucesión aleatoria infinita contiene más información que todos los sistemas lógicos humanos finitos tomados en conjunto"* (Casti (1990, p.309)).

[181] *"…von Mises intentó crear su deseada disciplina matemática, pero su teoría de los 'colectivos' resultó una confusa, aunque sugestiva mezcla de contextos matemáticos y no-matemáticos"* (Doob (1989, p. 815)).

[182] Esta denominación se debe a Keynes (1921). von Mises (1928) definió a este principio como el *"fenómeno primario"* (*"Urphänomen"*) de la teoría de la probabilidad.

$C = \{\omega_1, \omega_2, \dots, \omega_n, \dots\}$ (donde $\omega_n \in \Omega$ $(n = 1,2, \dots)$)[183].

Sea $E \subseteq \Omega$ un atributo asociado a un colectivo. Supóngase que en los primeros n elementos del colectivo el atributo E se haya presentado $n(E)$ veces, de modo que la frecuencia relativa del atributo E al cabo de las n observaciones sea $\frac{n(E)}{n}$. La ley de estabilidad de las frecuencias estadísticas establece que, a medida que n crece, la fecuencia relativa se aproximará indefinidamente a un valor fijo: la probabilidad θ[184]. Estos axiomas de convergencia y aleatoriedad son el fundamento de la "*ley de los grandes números de von Mises*".

Fry (1928) y Cantelli (1935) plantearon una objeción formal al "*Regellosigkeitsaxiom*" (la parte más original, más controvertida y más interesante de la teoría de los colectivos) sosteniendo que sólo una formulación matemática precisa del axioma de aleatoriedad podría garantizar la consistencia de ambos axiomas introducidos intuitivamente por von Mises y, en consecuencia, de la teoría. A este respecto, debe tenerse en cuenta que von Mises, adoptando una postura que podría denominarse "definicionista", parte de la tesis de que todos los conceptos de una ciencia matemática deben ser introducidos mediante definiciones explícitas. Este criterio presenta el inconveniente que, excepto en el caso en que ciertas nociones básicas sean asumidas como ideas primitivas, puede conducir a una regresión infinita, dado que la definición de un concepto siempre debe ser expresada en términos de otros conceptos[185]. En particular, como en este caso la teoría de la

---

183 Una axiomática intrínsecamente descriptiva y fundamentalmente empírica, como contraposición a la axiomática de Kolmogorov, exclusivamente matemática por naturaleza.

184 De acuerdo con von Mises (1928), la ley de estabilidad de las frecuencias se verifica en todos los juegos de azar. En particular, se refiere al caso del caballero de Méré. En su interpretación, el problema de de Méré involucra a dos colectivos, C_1 y C_2. Los elementos de C_1 consisten en 4 tiradas de un dado y el atributo (E_1) es obtener por lo menos una vez el resultado "6". Los elementos de C_2 consisten en 24 tiradas de 2 dados y el atributo (E_2) es obtener por lo menos 2 veces el resultado "6". de Méré concluyó, a través de la observación empírica que la frecuencia relativa de E_1 en C_1 convergía a un valor mayor que $\frac{1}{2}$, en tanto que la frecuencia relativa de E_2 en C_2 convergía a un valor menor que $\frac{1}{2}$.

185 Por ejemplo, Cramér (1946), de acuerdo con un moderno tratamiento impuesto por Hilbert (1899) en la geometría, asume a la probabilidad como una noción primitiva y la

probabilidad es interpretada como una rama de una ciencia fáctica, von Mises justifica la necesidad de introducir estas nociones básicas mediante definiciones operacionalistas en términos de elementos observables[186].

El operacionalismo-positivista de las ideas de von Mises[187] –que, como se mencionó en páginas anteriore, lo identifica con la filosofía del Círculo de Viena y la Unidad por la Evolución de la Ciencia- se debe fundamentalmente a la influencia de la obra de Ernst Mach (en particular de *"The science of mechanics: A critical and historical account of its development"* (1833)*)*. Su desarrollo de la teoría de la probabilidad siguió el mismo esquema que el desarrollo que Mach hizo de la mecánica: introdujo la ley de estabilidad de las frecuencias estadísticas (que supuso válida a partir de la observación) y sobre ella basó su definición de probabilidad (la definición de un concepto teórico, la probabilidad, en términos de un concepto observable, la frecuencia relativa), pero no proporcionó ninguna vinculación entre observación y teoría más allá de la muy controvertida utilización de los límites de una sucesión finita[188]y su justificación de la idealización del límite de la frecuencia relativa (que contradice el principio metodológico que postula que las propiedades de un objeto idealizado

---

caracteriza mediante los axiomas de la teoría: *"Algunos autores tratan de introducir un sistema de axiomas basado directamente sobre las propiedades de los cocientes de frecuencias. El máximo exponente de esta escuela es von Mises (...) quien define a la probabilidad como el 'límite de la frecuencia'* $\frac{\vartheta}{n}$ *de dicho evento, cuando* n *tiende a infinito. La existencia de este límite, en un sentido estrictamente matemático, es postulado como el primer axioma de la teoría. Si bien, indudablemente, una definición de este tipo a primera vista parece muy atractiva, incluye ciertas dificultades matemáticas que la despojan de su aparente simplicidad. Por otra parte, la definición de probabilidad propuesta de esta forma incluiría una mezcla de elementos empíricos y teóricos que las teorías axiomáticas modernas intentan evitar. Sería, por ejemplo, comparable a la definición de un punto geométrico como el límite de una figura de dimensiones infinitamente decrecientes, método que no es utilizado habitualmente en la geometría moderna"* (p. 150).

[186] von Mises, en el prefacio de "*Wahrscheinlichkeit,statistik und Wahrheit*", proporciona una muestra clara de su criterio operacionalista: *"La frecuencia relativa de la repetición es la 'medida' de la probabilidad, de la misma forma que la longitud de una columna de mercurio es la 'medida' de la temperatura"* (p. viii).

[187] Contenido en su *"Kleines Lehrbuch des Positivismus"* (1939).

[188] La ya comentada objeción acerca de la representación de los colectivos matemáticos infinitos por sucesiones finitas de los colectivos empíricos (ver Khinchin (1921)(1961)).

deberían aproximar a las propiedades de un objeto real) apartir de la aplicación del concepto de límite en la física teórica[189].

El operacionalismo, que tuvo su mayor difusión en la década de 1920, fue posteriormente muy criticado por los filósofos de la ciencia. La alternativa que prevaleció consideró que los términos teóricos de una ciencia natural podían ser introducidos como primitivos no-definidos y relacionados con la experiencia no en forma directa mediante una definición operacionalista, sino en forma indirecta[190]. En concordancia con esta corriente de pensamiento formalista, muchos probabilistas abandonaron el supuesto de que los conceptos teóricos debían ser definidos directamente en términos de elementos observables[191] y, tomando en consideración la propuesta de Fry-Cantelli, intentaron justificar rigurosamente la axiomática de von Mises utilizando argumentos de la teoría algorítmica de la complejidad y de la teoría ergódica[192].

Las primeras ideas sobre complejidad (y, en consecuencia, sobre aleatoriedad como complejidad máxima) fueron propuestas por Leibniz en su "*Discours de métaphysique*" (1686)[193]. En esta obra, en lo que podría considerarse un cuestionamiento al principio determinístico de la razón insuficiente, Leibniz planteó la distinción entre fenómenos formalizables, es

---

[189] von Mises (1928): *"Los resultados de una teoría basada en la noción de colectivo infinito pueden ser aplicados a sucesiones finitas de observaciones de una forma que no es definible lógicamente pero que, no obstante, es suficientemente exacta en la práctica. La relación teoría-observación es, en este caso, esencialmente la misma que en todas las demás ciencias físicas"* (p. 85).

[190] Este planteo formalista concuerda con el tratamiento matemático de la probabilidad basado en la axiomática de Kolmogorov.

[191] Para un análisis de las consideraciones críticas a la definición de Mach y un comentario detallado acerca de la relación Mach-von Mises, ver Gillies (1972a)(1973).

[192] La propuesta de von Mises, su posición crítica respecto de la dinámica clásica y la formulación de un esquema probabilístico para el tratamiento de problemas de física estadística, fueron las que motivaron el desarrollo del concepto de "*sistema dinámico abstracto*"y de la teoría ergódica (ver Sec. 22).

[193] En "*Mathematischen Schriften*" (1686).

decir explicables por una ley que permite describir sistemas de relación y, en consecuencia, construir modelos sobre su comportamiento y fenómenos irregulares. Como una aproximación a la propuesta de Leibniz, en 1928 Hilbert-Ackermann, a partir de una posición formalista, plantearon el"*Entscheindungsproblem*" (o problema de la decisión) que consistía, precisamente, en decidir sobre la posibilidad de definir un sistema formal coherente y sintácticamente completo que permitiera determinar la veracidad o falsedad de cualquier proposición. Es decir, un sistema tal que, dado un problema perfectamente definido, siempre fuera posible hallar un algoritmo capaz de resolverlo[194] (debe tenerse en cuenta que el concepto matemático riguroso de "algoritmo" como conjunto finito de reglas ordenadas que se puede ejecutar en un número determinado de pasos, fue introducido recién en la década de 1930 como culminación de las investigaciones sobre los fundamentos de la lógica matemática yque una definición operativa de la noción de algoritmo fue propuesta por Turing recién en 1937)[195]. El teorema de Gödel y los conceptos de la teoría de la información y de la complejidad algorítmica (en particular, de los postulados del teorema de Waerden), condujeron a la demostración de la imposibilidad de resolución del "*Entscheindungsproblem*" y a la convalidación de la propuesta de Leibniz de asimilación del concepto de aleatoriedad de una sucesión a la ausencia de una ley que permita explicar su estructura (es decir, a la ausencia absoluta de regularidades no-triviales

---

[194] Se dice que S satisface la propiedad fundamental de los sistemas formales si el conjunto de sus teoremas es (computablemente) numerable. Esta propiedad es una consecuencia de la condición que los axiomas del sistema formen un conjunto computable y que el conjunto de las demostraciones también sea computable. La propiedad de computabilidad también suele ser mencionada en la literatura como de "*calculabilidad*" o "*calculabilidad efectiva*"y, por extensión, como de "*decidibilidad*" (ver Franzén (2005), Berto (2009)). Se dice que un conjunto es "*decidible*" si existe un algoritmo que permite, para todo valor de un argumento x, decidir si x pertenece o no a dicho conjunto. Se dice que una propiedad es decidible si existe un algortimo capaz de "*decidir*" si un objeto x posee o no dicha propiedad.

[195] En la literatura se utiliza el término "algoritmo" también para caracterizar conjuntos de reglas que involucran un número infinito de pasos (ver Knuth (1985), Davies (2000), Landro (2010)).

en su trayectoria)[196].

Los aportes de Kolmogorov sobre aleatoriedad y complejidad tuvieron su culminación en los trabajos sobre teoría ergódica de la década de 1960 y estuvieron dirigidos fundamentalmente a descubrir determinismo en eventos aleatorios[197] y a definir la estructura de "*objetos discretos*" a partir de la ley de los grandes números o de la noción de complejidad algorítmica.De acuerdo con los postulados de la teoría de la información de Shannon (1948), la complejidad (entropía) de un objeto aleatorio, $X\colon p(x)$, está dada por $H = -\sum p(x) log[p(x)]$. Expresión que representa el número promedio de "bits" suficientes para su descripción. En 1965 Kolmogorov propuso[198] una definición de complejidad de un objeto finito como la longitud del programa binario más corto, $K(x) = \min_{U(p)=x} \ell(p)$, para su descripción[199]. Lo que permite concluir que $E[K(x)] \approx H$. Esta definición conduce a una interpretación de la teoría algorítmica de la información como una noción de complejidad computacional y la relaciona con el concepto de medida de probabilidad universal. Denotando por p_U a la medida de la probabilidad de que un autómata universal (U) imprima a X, se demuestra que $\log\left(\frac{1}{p_U(x)}\right) \approx\approx K(x)$ (para todo x)[200]. Más adelante, en

[196] Lo que condujo a de Finetti (1989, p. 244) a afirmar que "*...los frecuencistas, incluso Reichenbach, no hacen predicciones, sino profecías*".

[197] La aparente redundancia en la calificación de "aleatorio" se debe a la necesidad de distinguir a este evento de los eventos "ciertos" o "imposibles" definidos por Kolmogorov (1931).

[198] En forma independiente de Solomonov (1964) y Chaitin (1966)(1969).

[199] Thom (1975), a partir del ya mencionado principio postulado por Kolmogorov, de que la complejidad algorítmica está dirigida a reducir la aleatoriedad contenida en la información, planteó que si la reproducción de un algoritmo requiere un programa de la misma longitud que dicho sistema, el algoritmo no reduce la aleatoriedad de la información.

[200] El siguiente pasaje de Kolmogorov (1983a) resulta fundamental para una interpretación más detallada de la relación entre la noción de complejidad y la teoría de la probabilidad: *"La teoría de la información debe preceder a la teoría de la probabilidad y no basarse en ésta. De acuerdo con su verdadera esencia, los fundamentos de la teoría de la información poseen un carácter combinatorio finito. Las aplicaciones de la teoría de la probabilidad pueden ser*

la misma obra, Kolmogorov propuso una nueva aproximación a la definición de complejidad algorítmica, dirigida a incorporar la cantidad de información requerida para especificar un "*objeto*" X, a partir de una definición matemática precisa del concepto de función computable, proporcionada por la teoría algorítmicamediante la introducción de la máquina de Turing[201]. La condición necesaria y suficiente para que una función sea computable es que se pueda obtener un "output" de la aplicación de un programa (P) a una máquina de Turing universal (M_U)[202]. A partir de los trabajos de Kolmogorov y Turing se puede concluir que, dado un sistema algorítmico formal (F) existe una máquina de Turing (M_U) tal que los postulados de los teoremas de Fcoinciden con los "outputs" producidos por M_U y, viceversa, dada una máquina de Turing, es posible definir un sistema formal tal que los "outputs" de M_U coincidan con los resultados de los teoremas de F. Este isomorfismo entre máquinas y sistemas algorítmicos le permitió a Turing trasladar el problema de la computabilidad a un problema expresable en términos de máquinas. Como corolario de este resultado se demuestra que, si para reproducir una sucesión binaria de observaciones es necesario un programa de la misma longitud que la sucesión, las observaciones son al azar y se comprueba que el conjunto de sucesiones de complejidad menor que la longitud del programa que las reproduce forma un subconjunto infinitesimal del conjunto de sucesiones binarias.

---

expresadas sobre una base uniforme. Su contenido siempre está vinculado a cuestiones referidas a consecuencias de hipótesis acerca de la imposibilidad de reducir de una forma u otra la complejidad de la descripción de los objetos en cuestión. Obviamente, esta aproximación no impide el desarrollo de la teoría de la probabilidad como una rama de la matemática, en particular, como un caso especial de la teoría general de la medida. Los conceptos de teoría de la información aplicados a sucesiones infinitas generan investigaciones muy interesantes, las cuales, no siendo indispensables como base de la teoría de la probabilidad, pueden asumir cierto valor en el estudio de la parte algorítmica de la matemática" (p. 1).

[201] Los tabajos de Turing, junto con los de Church, constituyen la base de lo que hoy se conoce como "*tesis de Turing-Church*", según la cual todo algoritmo es implementable en una máquina de Turing mediante la utilización de un programa adecuado.

[202] Se denomina "*máquina de Turing universal*" a aquélla capaz de simular cualquier cálculo a resolver por una máquina de Turing, es decir una máquina capaz de calcular todas las unciones recursivas.

A partir de estos resultados Chaitin demostró que ningún programa puede reproducir una sucesión más compleja que sí mismo y que, en consecuencia, dada una sucesión suficientemente larga (n) de observaciones, no se puede demostrar que tenga una complejidad mayor que n o, en otros términos, que existe un nivel n tal que no se puede demostrar que ninguna sucesión de longitud mayor que n sea aleatoria.Es decir que, también desde el punto de vista de la teoría la complejidad algorítmica, se demuestra que la aleatoriedad de una sucesión de observaciones es indemostrable[203].

Otras críticas (elementales)al axioma de aleatoriedad se refieren a que cada colectivo definido por una sucesión binaria posee una infinidad de ceros y unos yque, por lo tanto, admite una sub-sucesión formada exclusivamente por ceros y otra formada exclusivamente por unos (un argumento que, para von Mises, demuestra que los colectivos no forman la clase de objetos matemáticos explícitamente definidos que permiten la definición individual de sus elementos) y a que, en consecuencia, excluye sucesiones que son lógicamente posibles. Quien contribuyó en mayor medida a la refutación de estas objeciones y a una formulación lógicamente consistente de la teoría, fue Wald (1937) con su demostración que, para todo conjunto numerable de funciones capaces de realizar la selección de sub-sucesiones, existen colectivos y que la probabilidad de todo conjunto $L \subset M$ (donde M denota el espacio muestral de los eventos) está dada por el límite de la frecuencia relativa, si ese límite existe. Con respecto a esta condición de existencia,su teorema fundamental postula que, dados un conjunto M, un conjunto J de subconjuntos de M medibles en el sentido de Jordan (con una medida normalizada μ) y un sistema numerable de funciones de selección S, entonces existen infinitos colectivos para los cuales $L \in J$ y la probabilidad está dada por la medida $\mu(L)$. Pero, planteó además la siguiente restricción: si M es continuo, todo sub-conjunto de M, medible en el sentido de Lebesgue y no-medible en el sentido de Peano-Jordan, no posee una medida (μ) aproximable por una medida finita-aditiva. Es decir que la frecuencia relativa sólo puede ser considerada como medida de la probabilidad de un evento si puede ser aproximada por una medida finita-aditiva. Una restricción aplicable tanto a la ley fuerte de los grandes números como a la ley del logaritmo iterado.

---

[203] Martin-Löf (1966), siguiento los lineamientos del teorema de invariancia de Kolmogorov (1965), desarrolló los conocidos como tests de aleatoriedad, los cuales componen una teoría formal de las sucesiones aleatorias finitas e infinitas (ver Zvinkin; Levin (1970), Cover; Gacs; Gray (1989)).

Ville (1939)[204] demostró que, a partir de la aproximación por la teoría de la medida de Kolmogorov[205], considerando eventos simples, independientes, con probabilidad de "éxito" θ, se verifica infinitamente a menudo que $\left|\frac{Y}{n} - \theta\right| > \frac{1}{\sqrt{n}}$, con probabilidad igual a uno[206]. Concluyó que, para todo θ y todo conjunto numerable de funciones de selección (S) y denotando por π a la condición que la desigualdad anterior se verifique sólo un número finito de veces, entonces existe una sucesión de colectivos cada uno con una frecuencia relativa límite igual a θ, tal que el límite de la frecuencia relativa de los colectivos para los cuales se verifica la condiciónπ, es igual a 1, excepto que el límite de las sucesión de colectivos sea igual a 0. En otros términos, demostró que los resultados a obtener a partir de la aditividad numerable no están incluidos en el ámbito de la interpretación frecuencista y concluyó finalmente que no es teóricamente legítimo definir a una probabilidad como límite en el sentido del análisis[207].

Una forma diferente de tratar la teoría de los colectivos de von Mises es a partir de la interpretación de una sucesión como un proceso dinámico caótico. La característica más importante de un proceso dinámico es su conjunto atractor, es decir, el subconjunto de su espacio de estadosen el cual evoluciona su trayectoria cuando $t \to \infty$. De acuerdo con los postulados del teorema de Levin-Schnorr-Chaitin la condición necesaria y suficiente para que una sucesión binaria pueda ser considerada aleatoria con respecto a la distribución $b\left(n, \frac{1}{2}\right)$ es que sea caótica con respecto a dicha distribución. Luego, se puede concluir que la aceptación de la existencia de atractores caóticos y la interpretación de la aleatoriedad como límite del grado de

---

204 En la que puede considerarse la última contribución a la teoría de von Mises durante un período muy largo.

205 La *"teoría clásica modernizada"* (p. 67) según la nomenclatura de Ville.

206 El valor $\frac{1}{\sqrt{n}}$ representa el primer límite inferior de los desvíos de la frecuencia relativa para la ley fuerte de los grandes números, luego optimizado por la ley del logaritmo iterado de Khinchin, como se verá en la Sec. 24.

207 Ville propuso, además, una caracterización de la aleatoriedad utilizando el concepto de martingala (ver Sec. 26). Por su parte, Schnorr (1971) demostró que los tests basados en martingalas son equivalentes a los tests universales de Martin-Löf.

complejidad de un atractor pueden ser consideradas como los únicos argumentos capaces de reducir (pero no de eliminar) la distancia entre la afirmación sobre la aleatoriedad de una sucesión y su demostrabilidad[208].

Además de la imposibilidad de obtener una formulación matemática rigurosa del "*Regellosigkeitsaxiom*", la defInición frecuencista es pasible de otras objeciones, en particular en lo referido a lo restringido de sus aplicaciones, dado que en muchos casos prácticos es imposible definir un colectivo empírico[209]. von Mises (a partir de una posición monista de la definición de probabilidad)consideró que esta circunstancia, lejos de ser una crítica, constituía un argumento a favor de su teoría, en la que sólo se pueden considerar probabilidades en un sentido matemático o cuantitativo si existe un conjunto suficientemente grande de eventos repetibles[210]: "*Primero el colectivo-entonces la probabilidad*" (von Mises (1928, p. 18)).

En defensa de los argumentos de von Mises, debe tenerse en cuenta que, al momento de la publicación de su teoría, el único método existente para evaluar probabilidades estaba vinculado con el principio de indiferencia y que este principio conduce a paradojas irresolubles. Recién en los años 1930-1931 aparecieron los primeros trabajos sobre una nueva aproximación subjetivista, en la que las probabilidades podían ser medidas a partir de grados de creencia personal y cuyos axiomas derivaban de la "*condición de coherencia*". Estos nuevos resultados no invalidaban los de von Mises y permitían demostrar que era posible extender el ámbito de las probabilidades cuantitativas y del cálculo matemático a muchos casos que no admitían la definición de colectivo.

---

[208] Un postulado de aleatoriedad basado en una condición de existencia completamente diferente a la propuesta por la teoría de conjuntos.

[209] Keynes (1921): *"Pienso que parte de la debilidad de la teoría frecuencista deriva de los restringidos límites de su aplicabilidad"* (p. 96).

[210] von Mises (1928): *"Nuestra teoría de la probabilidad no tiene nada que ver con preguntas tales como: ¿existe alguna probabilidad de que alguna vez Alemania se encuentre involucrada en una guerra con Liberia?"* (p. 9).

21.- Hans Reichenbach y las modificaciones a la interpretación de von Mises

Entre las numerosas modificaciones a que se sometió a la interpretación frecuencista, la más importante fue, indudablemente, la debida a Reichenbach (1932)(1937) quien procuró obtener una definición de probabilidad por una vía axiomática y justificar el significado intuitivo de la misma, generado por las "*complicadas e inexplicables regularidades*"[211]. Con respecto a la primera cuestión Reichenbach intentó una solución basada exclusivamente en la teoría de conjuntos y en las operaciones lógicas, obteniendo una definición de probabilidad (puramente formal) expresada como una relación entre dos clases de proposiciones.

En lo que hace al papel que desempeña la intuición en el análisis filosófico de la probabilidad objetiva, tomando como punto de partida el supuesto (contrario a la propuesta de von Mises) de que la interpretación frecuencista de la probabilidad podía ser generalizable a todas las aplicaciones del término probable, Reichenbah intentó ampliar sus alcances a eventos no-repetibles, mediante la definición de lo que denominó "*clases de referencia*" formadas por eventos similares al analizado, y consideró a la teoría de la probabilidad como la disciplina que calcula probabilidades desconocidas de "*colectivos derivados*" a partir de probabilidades conocidas de "*colectivos originarios*". Pero esta generalización tropezó con la dificultad insalvable que significa la imposibilidad de la determinación de reglas de selección objetivas universalmente aceptadas de los eventos que deben integrar dichas clases de referencia.

Con respecto a la interpretación analítica de la convergencia, Reichenbach postuló que la frecuencia límite no poseía nada de empírico, sino que consitutía un concepto matemático definible en el ámbito de las llamadas "*sucesiones de probabilidad*", las cuales se caracterizan porque la frecuencia con la que se presenta cualquier resultado posible en las primeras

[211] Se realizaron numerosos esfuerzos dirigidos a describir la naturaleza de las estabilidades y a comprender su significado. En realidad, la primera modificación de la definición frecuencista de von Mises se debe a Arthur Copeland (1928)(1931)(1937), quien generalizó el concepto de *"número normal"* propuesto por Borel (1909a). Las definiciones alternativas de la noción de colectivo de Reichenbach, Copeland y Popper se encuentran resumidas en Ville (1939).

n pruebas tiende a valores límite bien precisos cuando n aumenta indefinidamente. Este resultado lo condujo a establecer que las únicas sucesiones de pruebas alas cuales resultaba aplicable el concepto intuitivo de probabilidad eran las que pertenecían a la clase de "*series necesariamente protegidas en el futuro*" (Popper (1934)) y a asimilar, en consecuencia, el tratamiento del problema de la inferencia inductiva al estudio de tales series (a postular la equivalencia entre el problema de determinar una probabilidad y el de expresar conclusiones sobre el futuro en base a información sobre el pasado). Su justificación en el ámbito del empirismo lógico del principio inductivo puede ser expresado de la siguiente forma: dado el conjunto de los n primeros términos de una sucesión de eventos, a partir de los cuales se obtiene, para el resultado E, una frecuencia relativa $h_E^{(n)} = \frac{n_E}{n}$ y, suponiendo que no se conoce nada acerca de la probabilidad de la tendencia de dicha frecuencia a evolucionar hacia un valor límite prefijado θ, entonces a medida que aumente la sucesión, dicha frecuencia tenderá a un límite λ, tal que $p(\lambda - \varepsilon \leq \theta \leq \lambda + \varepsilon) \to 1$, donde ε denota un infinitésimo positivo tan pequeño como se desee. Este principio constituye la que podría denominarse como "*ley de los grandes números de Reichenbach*".

Este nuevo enunciado del principio frecuencista se distingue del de von Mises porque constituye un juicio sobre el conjunto de los eventos, no sobre cada uno de ellos en particular: contrariamente a la formulación de von Mises del principio de inducción, que supone que todos los eventos observados deben ser iguales entre sí y que dicho postulado es extensible a todos los eventos futuros de la misma sucesión, Reichenbach, a partir de la hipótesis más general de que la frecuencia asume un valor $h_E^{(n)}$ y suponiendo la permanencia de la misma en toda la sucesión, elabora su juicio sobre el conjunto de eventos futuros[212].

De todo lo anterior se deduce que, tanto la definición de von Mises como la de Reichenbach, a fin de evitar cualquier tipo de regularidad en las sucesiones de eventos que constituyen su fundamento, intentaron proporcionar a la relación de probabilidad un contenido estrictamente matemático mediante la imposición de complicadas condiciones que,

[212] Las propuestas de Reichenbach, en la medida que no pudieron justificar la posibilidad de enunciar proposiciones fundadas sobre el comportamiento de eventos futuros, no lograron resolver las consideraciones de Hume (1739) sobre la no-logicidad del principio de inducción.

inevitablemente, restringían el concepto de aleatoriedad total y demostraban que (como se expresó en páginas anteriores) era imposible dar un carácter matemáticamente preciso a la noción de "*irregularidad absoluta*"[213].

Luego se puede concluir que, si bien una de las principales objeciones de von Mises a la interpretación clásica de la probabilidad (y, en consecuencia, a las leyes de los grandes números de Bernoulli, Poisson y Chebychev)[214]era la de ser una teoría matemática que no trataba con eventos concretos[215], las modificaciones a las que fue sometido el planteo frecuencista eliminaron sus fundamentos empíricos, transformándolo también en una extravagante teoría puramente matemática que, en vez de tratar con resultados favorables y resultados posibles, trata con límites que son entidades matemáticas abstractas y en la cual las demostraciones de los teoremas se obtienen exclusivamente mediante la utilización de métodos lógico-matemáticos.

---

[213] Borel (1939): *"La objeción esencial que creo se puede hacer a la teoría del colectivo y a cualquier otra teoría análoga es la siguiente: al espíritu humano no le es posible imitar en forma perfecta a la aleatoriedad, es decir, sustituir el método empírico que consiste en realizar una serie indefinida de pruebas repetidas (por ejemplo, de tiradas a 'cara' o 'cruz') por un mecanismo racional. Existe una diferencia fundamental entre el cálculo de probabilidades y la geometría; cuando sustituimos la recta empírica e imperfecta por la recta ideal de los geómetras, conservamos las propiedades realmente fundamentales y eliminamos sólo las accidentales; en ese caso, es posible sustituir la definición empírica por una definición axiomática extremadmanete simple, a partir de la cual se podrán deducir todas las propiedades de las figuras formadas por rectas, no importa cuán complejas fueren. Por el contrario, cualquiera sea el procedimiento mediante el cual se puede definir una serie de elementos, reducidos a las cifras 0 y 1 a fin de simplificar el ejemplo, esta serie no poseerá todas las propiedades de una serie 'definida' por el azar, salvo que haya sido obtenida por sorteos sucesivos o métodos análogos; de no ser así, existirá algún defecto que permitirá a un observador sagaz afirmar que la serie fue construida racionalmente y no al azar"* (p. 2169). (ver Gnedenko (1950), Khinchin (1952)(1961)).

[214]"*Las deducciones matemáticas de Bernoulli, Poisson y Tschebysheff, en la medida que se basan en una definición de probabilidad que no tiene nada que ver con la frecuencia de ocurrencia de los eventos de una sucesión de observaciones, no pueden ser utilizadas para realizar predicciones relativas a los resultados de tales sucesiones. Por lo tanto, no poseen ningún tipo de vinculación con la regla empírica general formulada por Poisson en la introducción de su libro*" (1928, p. 109).

[215] "*...a menos que aceptemos que está afectada por un círculo vicioso, la definición clásica de la probabilidad implica la reducción de todas las distribuciones al caso más simple de las distribuciones uniformes*" (1928, p. 68).

Como se verá en lasseciones 24 y 25, curiosamente fueron Kolmogorov y Khinchin, quienes, en la década de 1930, produjeron la resurrección de la interpretación frecuencista.

22.- Frank Plumpton Ramsey, Bruno de Finetti, el teorema de representación y la teoría ergódica

Como se vio en la Sec. 17, Borel (1939), contrariamentea la posición monista de von Mises, demostró que existen operaciones mentales de naturaleza especial que son totalmente diferentes de la simple observación de las frecuencias, descalificó la afirmaciónde que el cálculo "a posteriori" de la probabilidad de ocurrencia de un fenómeno sólo podía ser justificado mediante la repetición de observaciones realizadas en igualdad de condiciones y el correspondiente análisis estadístico de los resultados, y concluyó que, precisamente, la noción de probabilidad de ocurrencia de un fenómeno único es la que constituye el fundamento de la teoría de la probabilidad.

La falencia de la interpretación frecuencista relacionada con la imposibilidad deproporcionar probabilidades objetivas de ocurrencia de fenómenos singulares ("*probabilidades singulares*") condujo a Keynes (1921) y a Popper (1935)(1957) a proponer nuevas interpretaciones objetivas de la probabilidad basadas, respectivamente, en relaciones lógicas en un ámbito formado por ideas abstractas y en la teoría de las propensiones, de carácter puramente metafísico (ver secciones 10 y 20).

La incapacidad de las definiciones logicista y propensionalista para identificar el supuesto verdadero valor de la probabilidad de ocurrencia de un evento dio origenal modelo subjetivista fundado en una concepción aleatorista generada en el supuesto ontológico de azar absoluto, en el marcode una filosofía empiricista-pragmatista-relativista[216], que postula la validez de

[216] El 'probabilismo' de de Finetti puede ser considerado como el verdadero heredero de la tradición filosófica Humeana. Para un análisis de la capacidad matemática y del contenido filosófico del pragmatismo-relativismo en la aproximación a la probabilidad, ver Galavotti (1989)(1991)(2005).

evaluaciones de probabilidades no coincidentes entre sí[217].

El modelo subjetivista carece de sentido en el dominio de la lógica de lo cierto, ya que no permite caracterizar como cierto o imposible algo que para el observador aparece como posible pero eventual y que permanecerá en esa categoría sin importar en qué forma ni en qué medida varíe su conjunto de información [218]. Las explicaciones de dicho observador se justifican solamente asumiendo "*...como instrumento fundamental del pensamiento científico, en lugar de la lógica ordinaria, rígida, fría, una lógica viva, elástica psicológica*"[219], a la que de Finetti denominó la "*lógica de lo probable*"[220], en la que sus posibilidades se reducen a ponderar el rango de los resultados posibles, asignando a cada uno de ellosuna probabilidad, es decir una evaluación cuantitativa de su grado de creencia racional –o idea

---

[217] La interpretación subjetivista no despertó en sus comienzos ningún interés en los probabilistas. Las primeras menciones acerca de las ideas de de Finetti se encuentran en Borel (1924) y en Lévy (1925). Puede afirmarse que la aceptación plena de la originalidad de la propuesta deFinettiana se logró a partir de la comprensión de los alcances de la posición pragmatista-relativista contenida en "*Probabilismo*" (1931b). En esta obra de Finetti reconsidera una de las conjeturas fundamentales planteadas por Poincaré en "*La scienze et l'hypothèse*" (1902), según la cual "*El método de las ciencias físicas se basa en la inducción, a partir de la cual esperamos la repetición de un fenómeno cuando se reproducen las circunstancias que habían generado su primera ocurrencia. Pero nunca sucederá la reproducción conjunta de 'todas' estas circunstancias. Siempre alguna de ellas no estará ¿Podemos asegurar que éstas no tienen importancia? (...) La respuesta a esta pregunta podrá ser verosímil pero nunca podrá ser considerada como rigurosamente cierta*" (pp. 8-9, referencia bibliográfica correspondiente a la versión italiana "*La scienza e l'ipotesi*", Bonpiani (2003)).

[218] No obstante, en parte de su obra dedicada al tratamiento de los procesos continuos en el dominio del tiempo, de Finetti trató de justificar la prevalencia de leyes estadísticas en el ámbito de las explicaciones del comportamiento de los fenómenos dinámicos en términos de mecánica clásica: "*..no existen leyes: las previsiones no pueden ser ciertas, sólo más o menos verdaderas o probables; las denominadas leyes de la naturaleza sólo son expresiones de regularidades estadísticas (...) si refutamos el determinismo, debemos aceptar (...) que las previsiones no pueden ser ciertas, sino más o menos probables*" (1931a, p. 64).

[219] de Finetti (1989, p. 21).

[220] de Finetti (1937, p. 152).

similar- acerca de su ocurrencia (suponiendo que cualquier grado de incertidumbre puede ser representado en forma numérica). Y que "... *nodebería ser considerada como una teoría auxiliar para aquellas ramas de la ciencia que aún no han descubierto mecanismos determinísticos que 'deben' existir; por el contrario, debe ser considerada como la premisa lógica del razonamiento inductivo en su totalidad. Así como la lógica ordinaria bivalente es el instrumento necesario de todo razonamiento en aquellos casos en los cuales es pertinente solamente la ocurrencia o no ocurrencia de un evento, así la lógica de lo probable, la lógica de una escalacontinua de valores, es el instrumento necesario de todo razonamiento afectado, en forma explícita o implícita, por un grado de duda, un juicio de certeza o de imposibilidad práctica o, en fin, una estimación de la verosimilitud de un evento cualquiera. Todo lo que no se reduce a una simple constatación, a una verdad histórica aislada, todo lo que nos guía hacia el futuro, hasta la creencia que al salir de este cuarto veremos como siempre las mismas calles en los mismos sitios, todo esto constituye un juiciobasado, aún inconciente e indistintamente, sobre los principios del cálculo de probabilidades. Este cálculo constituye, por lo tanto, el fundamento de la mayor parte de nuestro pensamiento por el cual podemos repetir con Poincaré: 'Sin esto la ciencia sería imposible'*" (de Finetti (1989, pp. 144-145)).

Esta interpretación de la teoría de la probabilidad como lógica de lo incierto constituye el núcleo de la interpretación subjetivista de F. Ramsey y B. de Finetti[221]. De acuerdo con estos autores, la teoría de la probabilidad constituye una generalización de la lógica formal en la medida que las leyes de la probabilidad –como las de la lógica- son "*leyes de no-contradicción (...) no restringen la libertad de nadie para atribuir un valor cualquiera a la probabilidad de ocurrencia de un evento cualquiera. Sólo se debe evitar que, al evaluar más de una probabilidad, dichas asignaciones sean contradictorias entre sí*"[222].

[221] de Finetti (1981): "*El cálculo de la probabilidad es la lógica de lo probable. Así como la lógica formal enseña a deducir la verdad o la falsedad de ciertas consecuencias a partir de la verdad o falsedad de ciertas premisas, el cálculo de la probabilidad enseña a deducir la mayor o menor verosimilitud o la probabilidad de ciertas consecuencias de la mayor o menor verosimilitud o probabilidad de ciertas premisas*" (p. 262).

[222] de Finetti (1972, p. 72).

Más allá de algunos curiosos antecedentes (ver Landro (2010)), se puede considerar que la teoría subjetiva de la probabilidad fue introducida en forma independiente por Ramsey (1926) y de Finetti (1930a) utilizando tratamientos que difieren tanto en sus detalles formales como en su aproximación conceptual. Ramsey planteó su propuesta –de carácter netamente antilogicista- a partir de una crítica detallada de la interpretación Keynesiana, en tanto que la obra de de Finetti –de carácter netamente antifrecuencista- se basa en C. Czuber (1903), quien proporcionó una de las mejores exposiciones sobre las paradojas de la probabilidad geométrica (ver Keynes (1921)) y concluyó que no existe ninguna necesidad de suponer el cumplimiento de la condición de la razón insuficiente (ver de Finetti (1931a)).

De acuerdo con su visión biconceptual, Ramsey -que considera que el significado de la probabilidad en lógica (tanto inductiva como deductiva) puede ser diferente de su significado en estadística- [223] propuso dos interpretaciones de la probabilidad, una frecuencista en términos de cocientes de clases y otra como grado de creencia parcial[224] y planteó una forma de relación entre ambas: *"...existen muchas conexiones entre creencias parciales y frecuencias. Por ejemplo, las frecuencias observadas a menudo conducen a los correspondientes grados de creencia y las creencias*

---

[223] Ramsey (1926): *"En este ensayo la teoría de la probabilidad es asumida como una rama de la lógica, la lógica de la creencia parcial y del argumento no-concluyente; pero esto no implica sugerir que éste es el único ni siquiera el más importante aspecto en esta materia. La probabilidad es de fundamental importancia no sólo en lógica, sino también en estadística y en física y no podemos asegurar de antemano que su interpretación, tan útil en lógica, será también apropiada en física. Es claro que la diferencia de opinión entre los estadísticos que, en su mayoría, adoptan la teoría frecuencista de la probabilidad y los lógicos que, en general, la rechazan, hace que parezca que las dos escuelas estuvieran discutiendo cosas diferentes y que la palabra ´probabilidad´ fuera usada en un sentido por los estadísticos y en otro sentido por los lógicos. Nuestras conclusiones son que el significado de la probabilidad en lógica no deber ser tomado como prejuzgando su significado en física"* (p. 157).

[224] Ramsey (1926): "…*al comienzo de este ensayo hemos visto que el cálculo de probabilidades puede ser interpretado en términos de cocientes de clases; ahora hemos hallado que también puede ser interpretado como un cálculo de la creencia parcial consistente. Por lo tanto, resulta natural que supongamos la existencia de una íntima conexión entre estas dos interpretaciones que permita una explicación de la posibilidad de aplicación del mismo cálculo matemático a dos conjuntos de fenómenos diferentes*" (p. 163).

parciales, de acuerdo al teorema de Bernoulli, conducen a la expectativa de las correspondientes frecuencias. Pero ninguna de éstas es la conexión que deseamos; una creencia parcial, en general, no puede ser relacionada únicamente con la frecuencia real, la conexión siempre se construye tomando la proposición en cuestión como una instancia de una función proposicional. Cuál función proposicional elegir constituye una decisión en cierta forma arbitraria que influirá considerablemente sobre la frecuencia correspondiente. La pretensión de algunos exponentes de la teoría frecuencista de que en una proposición frecuencista la creencia parcial significa creencia total no es sostenible. Pero hemos hallado que la idea de creencia parcial involucra una referencia a una frecuencia hipotética o ideal: suponiendo que los éxitos sean aditivos, el grado de creencia m/n es el tipo de creencia que conduce a la acción óptima si en n repeticiones se obtiene que en m veces la proposición es verdades o, en forma resumida, podemos decir que es la clase de creencia más apropiada para un número de ocasiones hipotéticas idénticas en una proporción m/n en la cual la proposición es verdadera. Es esta conexión entre la creencia parcial y la frecuencia la que nos permite utilizar el cálculo de las frecuencias como un cálculo de la creencia parcial consistente. Y, en este sentido, podemos decir que las dos interpretaciones constituyen los aspectos objetivo y subjetivo del mismo íntimo significado, así como la lógica formal puede ser interpretada objetivamente como un cuerpo de tautologías y subjetivamente como las leyes del pensamiento consistente"(pp. 187-188*).*

Teniendo en cuenta el ya mencionado problema de la selección de la clase de referencia, se podría concluir que la afirmación de Ramsey acerca de que "*…las dos interpretaciones constituyen los aspectos objetivo y subjetivo del mismo íntimo significado*" parece demasiado simplista. Esta sospecha se ve confirmada por la afirmación de Keynes (1921) que indica que basar la asignación de la probabilidad sobre la ocurrencia de un fenómeno singular en una frecuencia relativa estadística exclusivamente, a menudo conduce a resultados poco satisfactorios, dado que la no consideración de información no-estadística relevante puede hacer que la evaluación de la probabilidad difiera significativamente del resultado proporcionado por el cálculo frecuencista[225].

[225] Para un análisis comparativo de las propuestas de Ramsey y de Finetti, ver Galavotti (1991)(1994)(2005), Dokic; Engel (2001).

La obra de de Finetti comprende dos tipos de aproximaciones al concepto de probabilidad: una –cuantitativa-, que considera a la probabilidad "*...como una medida cuya función consiste en transformar nuestro grado de incertidumbre en un número*" (1931c, p. 302), basada en un cociente de apuestas y otra, que considera que "*...así como en las ciencias experimentales se sustituye el mundo de las sensaciones por un mundo ficticio en el cual las medidas son un valor determinable exactamente, en el cálculo de probabilidades se sustituye un estado de ánimo vago e inaferrable por aquél de un individuo ficticio que no conoce incertidumbres en la adjudicación de los grados de creencia*" (1989, pp. 50-51)), basada en una teoría axiomática de probabilidad cualitativa.Ya en 1935 había manifestrado su preferencia por una axiomatización puramente cualitativa debido "... *a su ventaja para permitir un análisis más profundo, en la medida que parte de nociones puramente cualitativas y no necesita introducir la noción de 'dinero' que es totalmente extraña al cálculo de la probabilidad, pero que es necesaria para poder hablar de apuestas*" (1989, p. 76). Posteriormente, a partir de la década de 1960 prefirió analizar la probabilidad basándose en las "*reglas de penalización*"[226]

El objetivo fnal del modelo subjetivista consistió en precisar la relación entre las condiciones de coherencia y la probabilidad cuantitativa axiomatizada por Kolmogorov[227], en construir una teoría unificada de la realidad y del valor esperado (a la que denominó "*teoría de la previsión*"), caracterizada como un operador líneal sobre funciones de cantidades.A partir de la condición de equidad –que implica la indiferencia entre ser uno u otro jugador en un juego de dos contrincantes que no admite empates, entre pagar o cobrar p para cobrar o pagar 1 en caso de verificarse el resultado eventual E- se dice que la asignación de la probabilidad p es "*coherente*" en cuanto no coloca a

---

[226] de Finetti (1989): "*A la lógica matemática y a la crítica positiva del mundo empírico se agrega el probabilismo, que corrige e integra las anteriores en aquellos aspectos que le resultaban inaceptables, en los cuales una cosa cualquiera parecía que se debía considerar dotada de un valor absoluto, que trasciende y es independiente del valor psicológico que posee sobre mí*" (p. 69).

[227] Debe tenerse en cuenta que, si bien, como se verá en la Sec. 25, Kolmogorov consideró a sus axiomas como el fundamento matemático de la interpretación frecuencista-propensionalista de la probabilidad, paradójicamente su fomalización los hizo suceptibles de una interpretación subjetivista.

ninguno de los dos jugadores en la situación de ganar con seguridad. Es decir, si p es una evaluación coherente en el sentido de Ramsey-de Finetti de la probabilidad de ocurrencia de un evento E para un individuo, dado que, como se mencionó más arriba, el precio es entendido en este caso como una magnitud lineal, la evaluación de la probabilidad de no ocurrencia de E (es decir, de la ocurrencia de $\bar{E}$) para dicho individuo, debe ser $p(\bar{E}) = p(1-E) = 1 - p(E) = 1 - p$.En caso contrario sería posibleque existiera una apuesta en su contra que lo colocaría en la posición de perder con prescindencia de la ocurrencia o no de E: si se verificara que $q < 1 - p$, el individuo-evaluador podría realizar, contemporáneamente, dos apuestas, una de p sobre la ocurrencia de E y otra de q sobre la ocurrencia de $\bar{E}$, a cambio de recibir (por cualquiera de las dos apuestas) una cantidad igual a 1. Dado que, en forma cierta, el individuo ganará una (y sólo una) de las apuestas, se encontrará en situación de recibir, concerteza, una cantidad igual a 1 habiendorealizado un gasto de $p + q < 1$, obteniendo, en consecuencia, una ganancia segura igual a $1 - (p + q) > 0$. Si, por el contrario, se verificara que $q > 1 - p$, el individuo-evaluador podría aceptar contemporáneamente dos apuestas, una de p sobre la ocurrencia de E y otra de q sobre la ocurrencia de $\bar{E}$, obligándose a pagar a su contrincante una cantidad igual a 1, en caso de ocurrencia de E. Dado que el contrincante ganará con certeza una de las dos apuestas, se hallará en la situación de pagar 1, habiendo recibido una cantidad igual a $p + q > 1$, obteniendo, en consecuencia, una ganancia segura igual a $(p + q) - 1 > 0$. Obsérvese que, como se emncioné más arriba, la insistencia de de Finetti en considerar a la coherencia como condición necesaria en la evaluación de la probabilidad, se basa en la necesidad de preservar la relación de dicha evaluación con la probabilidad cuantitativa axiomatizada por Kolmogorov[228].

Obviamente, el valor de la expectativa de recibir un importe monetario disminuirá si está subordinado a una cláusula restrictiva según la cual dicha expectativa será considerada válida a condición de que se verifique la ocurrencia de un evento E. Es decir que la expectativa de recibir una cantidad ϑ condicionada a la ocurrencia de un evento E, tendrá un valor (ϑ') menor

[228] "*Es natural denominar a un individuo como incoherente si un conjunto de apuestas que él considera equitativas, proporciona una ganancia segura a uno de los dos contendientes*" (1981, p. 263). Para un análisis crítico sobre la coherencia como la condición que caracteriza la racionalidad en Ramsey y de Finetti, ver Mura (1995).

que ϑ (en la medida que la ocurrencia de E se suponga más probable, la cantidad ϑ' se aproximará más a ϑ y, viceversa, cuanto menos probable se suponga la ocurrencia de E, ϑ' se aproximará más a cero). A partir de estas consideraciones se puede concluir que, en general, un indicador adecuado de la probabilidad de ocurrencia de E podría ser la relación entre ϑ' y ϑ. Como corolario de este razonamiento surge que, haciendo $\vartheta = 1$, la probabilidad de ocurrencia de E, la cual asumirá valores en el intervalo $p(E) = \begin{cases} 1 & si \;\; E \\ 0 & si \;\; \bar{E} \end{cases}$, quedará caracterizada como el precio equitativo de la expectativa de recibir 1, en caso de ocurrencia de E.Si se considera, entonces, a Y como una variable aleatoria que puede asumir los valores 1 ó 0 según que E se verifique o no se verifique, $p(E)$ puede ser interpretada como la "*previsión*" o valor esperado de E ($\wp(E)$), equivalente al precio de una oferta aleatoria igual a E.

Si bien, en este caso la probabilidad coincide con su previsión, el concepto de previsión es más general que el de probabilidad[229]. En general, la "*previsión*" de una variable aleatoria X -que puede asumir los valores $x_i (i = 1,2, \dots, n)$- está dada por el precio equitativo del importe aleatorio, $X = \sum_{i=1}^{n} x_i E_i$, donde E_i representa el evento $(X = x_i)$. De acuerdo con esta expresión y teniendo en cuenta la condición de linealidad de los precios, se puede deducir la definición de $\wp(X)$ a partir de su interpretación como precio equitativo[230]:

---

[229] de Finetti (1995): *"En muchos textos –especialmente en aquellos más antiguos- la 'previsión'es denominada 'esperanza matemática'. La esperanza puede ser positiva o negativa.Cuando es negativa, el término 'esperanza' parece impropio: por, ejemplo, la expresión 'esperanza matemática de fallecimiento' es a la vez, ridícula y fuera de lugar. Por eso prefiero la palabra 'previsión' (...) ya que tiene la misma inicial '℘' que 'probabilidad', de modo que al escribir $\wp(X)$ se puede interpretar (en general) como la previsión de X y (en particular) como la probabilidad de X, en los casos en que X fuese un evento"* (p.18).

[230] De acuerdo con este razonamiento, las operaciones lógicas entre eventos pueden ser definidas como operaciones aritméticas de la siguiente forma: $E_1 \cup E_2 = E_1 + E_2 - E_1 E_2 = Max(E_1, E_2)$,$E_1 \cap E_2 = E_1 E_2 = min(E_1, E_2)$ y $\bar{E} = 1 - E$ Estas operaciones lógicas son aplicables también a variables aleatorias, $X \cup Y = Max(X, Y)$, $X \cap Y = min(X, Y)$ y $\bar{X} = 1 - X$. De la misma forma, los conceptos de frecuencia absoluta (N_n) y frecuencia relativa (f_n) pueden ser caracterizados en términos de operaciones aritméticas entre eventos de la siguiente forma: $N_n = E_1 + E_2 + \dots + E_n$ y $f_n = \frac{N_n}{n} = \frac{1}{n}(E_1 + E_2 + \dots + E_n)$.

$$\wp(X) = p\left(\sum_{i=1}^{n} x_i E_i\right) = \sum_{i=1}^{n} p(x_i E_i) = \sum_{i=1}^{n} x_i p(E_i)$$

(obsérvese que, a partir de esta interpretación de la probabilidad como un caso particular de la previsión, es posible considerar al concepto de previsión como primitivo y al de probabilidad como derivado)[231].

El valor ϑ' constituye lo que habitualmente se denomina una apuesta, por lo que la relación $\frac{\vartheta\prime}{\vartheta}$ define un cociente de apuestas: de acuerdo con la definición dada más arriba, la equidad de este cociente, para un observador dado, está dada por su indiferencia en apostar a favor de la ocurrencia de E o apostar contra la ocurrencia de E o no apostar. Operar a favor o en contra de la ocurrencia de E determina la "*dirección de la apuesta*"[232].
Supónganse dos eventos incompatibles, E_1 y E_2, que definen una clase completa (uno y sólo uno de ellos debe ocurrir, es decir, E_2 es el complemento deE_1,$E_2 = \overline{E_1}$). Sean p_1 y p_2 las evaluaciones realizadas por un individuo de las respectivas probabilidades y S_2 y S_2 las cantidades –desconocidas- a recibir, respectivamente, por el mismoen caso de ocurrencia de E_1 o E_2 ($S_2, S_2 \neq 0$).Las ganancias a obtener por el individuo-evaluador en los casos de ocurrencia de E_1 y E_2, estarán dadas, respectivamente, por las diferencias entre las cantidades a cobrar y las cantidades apostadas en cada caso:

$$G_1 = (1 - p_1)S_1 - p_2 S_2$$

---

[231] En general. Dada una variable continua $\{X, \Omega(X), F_X(x)\}$, será $\wp(X) = E(X) = \int_{\Omega(X)} x dF_X(x)$ y, dada una función $h(X)$ medible en el sentido de Borel, será $E[h(X)] = \int_{\Omega(X)} h(x) dF_X(x)$.

[232] Esta interpretación de la probabilidad como relación entre valores de expectativas se remonta, en realidad, a Bayes (1764) quien definió la probabilidad de un evento como *"...la relación entre el valor al cual debe ser calculada una expectativa que depende de la ocurrencia de un evento y el valor que se espera que aquélla asuma una vez que el evento se ha verificado"* (p. 143) (si bien del texto no resulta claro qué entendía Bayes por *"valor"*, la interpretación más aceptable parecería ser la de precio justo).

$$G_2 = -p_1 S_1 + (1 - p_2) S_2$$

Si el determinante correspondiente a este sistema de ecuaciones lineales:

$$\begin{vmatrix} 1 - p_1 & -p_2 \\ -p_2 & 1 - p_1 \end{vmatrix} = 1 - (p_1 + p_2)$$

fuera no-nulo, entonces sería posible hallar valores de S_1 y S_2 tales que G_1 y G_2 resultaran ambas positivas pero, según se expresó más arriba, esto significaría que la evaluación del valor de p realizada por el individuo-evaluador (debido a influencias psicológicas ajenas a lo que constituye estrictamente su grado de creencia en la ocurrencia del evento) sería tal que lo colocaría en la posición de perder con certeza[233]. En este caso se dice que la evaluación realizada por dicho individuo es incoherente. Luego, la "*condición de coherencia*", como se mencionó en páginas anteriores, implica que se verifique la relación $p_1 + p_2 = 1$. De estos resultados se puede concluir que la condición de coherencia constituye una restricción, que de alguna forma, permite asegurar que los grados de creencia son racionales (al menos hasta el punto de satisfacer dicha restricción).

Supóngase que E fuera un evento cierto. La única ganancia posible sería $G_1 = [1 - p(E)]S_1$. Si $1 - p(E) \neq 0$, entonces sería posible hallar un valor de S_1 tal que $G_1 > 0$. Luego, para asegurar la coherencia, debe ser $p(E) = p = 1$. De la misma forma, si E fuera un evento imposible, la única ganancia posible sería $G_2 = [1 - p(E)]S_2 = p(\bar{E})S_2$. En este caso, si $p(\bar{E}) \neq 0$, sería posible hallar un valor de S_2 tal que $G_2 > 0$. Por lo tanto, para asegurar el cumplimiento de la condición de coherencia, se debe verificar que $p(E) = p = 0$. Estos resultados permiten concluir que, para un evento posible (es decir, ni cierto ni imposible), debe ser $0 < p(E) < 1$[234].

---

[233] Se dice, entonces, que el individuo-evaluador sufrió un "dutch book" por parte de su contrincante. El origen de esta denominación de "apuesta holandesa" (habitual en la literatura inglesa) es desconocido. Para un análisis histórico-crítico de este tipo de apuesta, ver Festa (1996).

[234] Una definición de la probabilidad $p(E)$ en términos de "*utilidad prevista*" ya había sido propuesta por Ramsey (1926). Posteriormente la interpretación subjetiva asumió el concepto de maximización de la esperanza moral.

Por otra parte, dados los eventos E_1 y E_2, si se satisface la condición de coherencia, se verificará que $p_1 G_1 + p_2 G_2 = 0$ para cualquier valor de las cantidades G_1 y G_2. De modo que G_1 y G_2 no pueden ser ambas positivas. De esto se puede concluir que la condición $p_1 + p_1 = 1$ no es sólo necesaria, sino también suficiente para la existencia de coherencia[235] (y, en consecuencia, que el cumplimiento de la condición de coherencia implica que un conjunto de probabilidades definidas como cocientes de apuestas satisfaga la axiomática de Kolmogorov)[236].

Generalizando la condición de coherencia, sea $\varepsilon = \{E_1, E_2, \dots, E_n\}$ un conjunto de eventos incompatibles que definen una clase completa. Sean $p_1, p_2, \dots, p_n$ las evaluaciones realizadas por un individuo de las respecivas probabilidades de ocurrencia y sean $S_1, S_2, \dots, S_n$ las cantidades a recibir por dicho individuo en caso de ocurrencia de $E_1, E_2, \dots, E_n$, respectivamente. Las ganancias a obtener por el individuo-evaluador en caso de ocurrencia de $E_i (i = 1,2, \dots, n)$ están dadas por $G_i = S_i - \sum_{j=1}^{n} p_j S_j \; (i = 1,2, \dots, n)$. Si se interpreta a las S_i como incógnitas, queda definido un sistema de ecuaciones lineales cuyo determinante es de la forma:

$$\begin{vmatrix} 1 - p_1 & -p_2 & \dots & -p_n \\ -p_1 & 1 - p_2 & \dots & -p_n \\ \dots & \dots & \dots & \dots \\ -p_1 & -p_2 & \dots & 1 - p_n \end{vmatrix} = 1 - \sum_{j=1}^{n} p_j S_j$$

Para que este determinante se anule y, en consecuencia, no existan valores de S_j que hagan que todas las cantidades G_j sean positivas, debe verificarse

---

[235] Este concepto de coherencia es equivalente al *"principio de la cosa cierta"* (*"sure-thing principle"*) de Savage (1954) según el cual, dadas dos ofertas aleatorias (A y B) que, ante la ocurrenciadel evento E, produzcan la misma ganancia (G), si las ofertas A y B fueran modificadas de tal modo que produjeran la misma ganancia G^* ($\neq G$), se puede asegurar que las relaciones de preferencia entre A y B permanecerán inalteradas. Debe tenerse en cuenta que este principio de la cosa cierta es condición necesaria para la existencia de una función de utilidad lineal que preserve las relaciones de preferencia, pero no es condición necesaria ni suficiente para la caracterización de la probabilidad subjetiva como cociente equitativo de apuestas.

[236] Ramsey (1926) denominó a esta condición como de *"consistencia"*. En general, la literatura adoptó la denominación de coherencia ya que la consistencia posee una connotación diferente en el ámbito de la lógica deductiva.

necesariamente que $p_1 + p_2 + \cdots + p_n = 1$. Asimismo, si se cumple esta condición, se verificará que $\sum_{j=1}^{n} p_j G_j = 0$. Esto demuestra que, dada una clase completa de eventos incompatibles, la condición necesaria y suficiente para la existencia de coherencia es que la suma de sus probabilidades sea igual a la unidad, $\sum_{j=1}^{n} p_j = 1$ (obsérvese que la aditividad numerable no es condición necesaria para que una probabilidad sea coherente). En consecuencia, se denominará probabilidad de ocurrencia de un evento E, para un individuo dado, al número real $p(E) = p$ que representa la medida de su grado de creencia en E y que, interpretado como una apuesta de valor p sobre la ocurrencia de E, cumple la condición de coherencia.

En 1955 Shimony introdujo el concepto de "*coherencia estricta*", que requiere que las evaluaciones de las apuestas no sólo impidan al competidor la posibilidad de obtener una ganancia cierta, sino que sean tales que no le permitan perder, si existe para él la posibilidad de una ganancia (suponiendo un número finito de alternativas elementales). Esta condición -que, obviamente, es más restrictiva que la de la coherencia simple- implica la posibilidad de perder, pero con probabilidad nula, es decir, implica la equivalencia entre la probabilidad nula y la imposibilidad lógica y, en consecuencia, la equivalencia éntrela probabilidad unitaria y la certeza lógica[237].

A partir del concepto más general de "*probabilidad cualitativa comparativa*" ($\wp$) como representación de los grados de creencia que un individuo-evaluador asigna a la ocurrencia de un evento dado, la axiomática subjetivista considera las propiedades de una relación binaria estricta del tipo "es más probable que" o a su versión más débil "es, al menos, tan probable como", referidas a conjuntos de proposiciones o eventos: Dados dos eventos, E_1 y E_2, ni ciertos ni imposibles: **i)** $\wp(E_1) \succ \wp(E_2)$representa la relación "el individuo evaluador considera que el evento E_1 es más probable que el evento E_2"; **ii)** $\wp(E_1) \approx \wp(E_2)$representa la relación de "*equiprobabilidad comparativa*", "el individuo evaluador considera que no existe una diferencia significativa entre las probabilidades de ocurrencia de los eventos E_1 y E_2"; **iii)** $\wp(E_1) \succcurlyeq \wp(E_2)$representa la relación "el individuo evaluador considera que el evento E_1 es al menos tan probable como el evento E_2".

---

[237] Debe tenerse en cuenta que, como se mencionó en la Sec. 13, resulta injustificable atribuir a un evento posible el mismo grado de incertidumbre que a un evento imposible.

Como consecuencia inmediata de estas definiciones, sepuede concluir que:**i)** si no se verifican las relaciones $\wp(E_1) \succ \wp(E_2)$ ni $\wp(E_2) \succ \wp(E_1)$, entonces será $\wp(E_1) \approx \wp(E_2)$; **ii)** si se verifica la relación $\wp(E_1) \succ \wp(E_2)$ o $\wp(E_1) \approx \wp(E_2)$, entonces será $\wp(E_1) \succcurlyeq \wp(E_2)$; **iii)** si se verifican las relaciones $\wp(E_1) \succcurlyeq \wp(E_2)$ y $\wp(E_2) \succcurlyeq \wp(E_1)$, entonces será $\wp(E_1) \approx \wp(E_2)$ y **iv)** si se verifica la relación $\wp(E_1) \succcurlyeq \wp(E_2)$ y no se verifica la relación $\wp(E_2) \succcurlyeq \wp(E_1)$, entonces será $\wp(E_1) \succ \wp(E_2)$.

Una extensión inmediata de estas relaciones permite deducir que, cuando no se verifica la relación $\wp(E_1) \succcurlyeq \wp(E_2)$ ni la relación $\wp(E_2) \succcurlyeq \wp(E_1)$, se puede considerar que los eventos E_1 y E_2 son no-comparables en el sentido de Keynes.

El conjunto sobre el cual están definidas las relaciones binarias $\succ$ y $\succcurlyeq$ define un álgebra $\Im$ de subconjuntos $E_1, E_2, \ldots$ de un dominio Ω. Según la notación utilizada precedentemente $E_i \in \Omega$ define un evento y, para todo $E_i \in \Omega$, se verifica que $\emptyset \subseteq E_i \subseteq \Omega$ (donde $\emptyset$ denota el evento vacío o imposible).El término evento tal como es interpretado en este contexto subjetivista, se refiere a un caso singular que, para un individuo que en ciertas circunstancias no puede asegurar su ocurrencia en forma cierta, es aleatorio (de modo que en el ámbito de la interpretación subjetivista, la aleatoriedad no es considerada como una propiedad de los eventos, sino del conocimiento que el individuo posee de dichos eventos). Esta noción de evento no coincide con el concepto abstracto general de proposición de la interpretación clásica, la cual –de acuerdo con el sentido que le adjudican los lógicos- posee incondicionalmente la propiedad de ser verdadera o falsa. En la interpretación subjetivista la identificación entre eventos y proposiciones se logra a partir de una lógica trivalente.

De acuerdo con de Finetti (1937), se dice que $\wp(\Omega)$ define una estructura de probabilidad cualitativa si, para cualquier conjunto de eventos $\{E_1, E_2, E_3, \ldots\}$, se verifican los siguientes axiomas:**1) de asimetría**: si se verifica que $\wp(E_i) \succ \wp(E_j)(i \neq j)$, entonces no se verifica que $\wp(E_j) \succ \wp(E_i)$; **2) de no-trivialidad**: el evento cierto es estrictamente más probable que el evento imposible, $\wp(\Omega) \succ \wp(\emptyset)$; **3) de no-negatividad**: cualquier evento es débilmente más probable que el evento imposible, $\wp(E) \succcurlyeq \wp(\emptyset)$; **4) de monotonicidad**: si se verifica que $E_i \supseteq E_j$, entonces se verifica que $\wp(E_i) \succcurlyeq \wp(j)$; **5) de inclusión y monotonicidad**: si se verifica que $E_i \supseteq E_j$

y que $\wp(E_j) \succ \wp(E_h)(i \neq j \neq h)$ o que $\wp(E_i) \succ \wp(E_j)$ y que $E_j \supseteq E_h$, entonces será $\wp(E_i) \succ \wp(E_h)$; **6) de transitividad**: si se verifica que $\wp(E_i) \succcurlyeq \wp(E_j)$ y $\wp(E_j) \succcurlyeq \wp(E_h)$, entonces será $\wp(E_i) \succcurlyeq \wp(E_h)$; **7) de aditividad simple**: si se verifica que $\wp(E_i \wedge E_h) = \emptyset$ y que $\wp(E_j \wedge E_h) = \emptyset$, entonces, si $\wp(E_i) \succcurlyeq \wp(E_j)$, será $\wp(E_i \vee E_j) \succcurlyeq \wp(E_j \vee E_h)$ y viceversa; **8) de complementariedad**: si se verifica que $\wp(E_i) \succcurlyeq \wp(E_j)$, entonces no se verifica que $\wp(\overline{E_i}) \succcurlyeq \wp(\overline{E_j})$.

De los axiomas **1)** y **2)** se puede concluir que la relación $\succ$ determina un cierto ordenamiento (débil) de los eventos en $\Im$. En particular, Savage (1954) considera a una relación $\succcurlyeq$ como capaz de generar una estructura de probabilidades cualitativa cuando satisface los axiomas de ordenamiento débil, no-trivialidad, no-negatividad y aditividad. Asimismo, Como corolario de los axiomas precedentes se obtiene fácilmente que: **i)** si se verifican las relaciones $\wp(E_i) \succ \emptyset$ y $\wp(E_i \wedge E_j) = \emptyset$, entonces será $\wp(E_i) \prec \wp(E_i \vee E_j)$; **ii)** si se verifican las relaciones $\wp(E_i) \succcurlyeq \wp(E_j)$, $\wp(E_h) \succcurlyeq \wp(E_k)(i \neq j \neq h \neq k)$y $\wp(E_i \wedge E_h) = \emptyset$, entonces será $\wp(E_i \vee E_h) \succcurlyeq \wp(E_j \vee E_k)$; **iii)** si se verifican las relaciones $\wp(E_i \wedge E_j) \succcurlyeq \wp(E_h \wedge E_k)$ y $\wp(E_h \wedge E_k) = \emptyset$, entonces será $\wp(E_i) \succcurlyeq \wp(E_h)$ o $\wp(E_j) \succcurlyeq \wp(E_k)$ y **iv)** si se verifican las relaciones $\wp(E_i) \succcurlyeq \wp(\sim E_i)$ y $\wp(E_j) \preccurlyeq \wp(\sim E_j)$, entonces será $\wp(E_i) \succcurlyeq \wp(E_j)$.

El espíritu con que fue establecido este sistema de axiomas está relacionado con la necesidad, ya comentada, de fijar las condiciones necesarias para garantizar la existencia de una función real (p) del grado de incertidumbre sobre Ω tal que: **i)** $p(\Omega) = 1$; **ii)** dados dos eventos cualesquiera E_1 y E_2 (siendo $\wp(E_1), \wp(E_2) \in \Im$), se verifique que $p(E_1) > p(E_2)$ si, y sólo si $\wp(E_1) \succ \wp(E_2)$ y **iii)** dados dos eventos E_1 y $E_2 \in \Omega$ tales que $p(E_1) \geq 0$, $p(E_2) \geq 0$y $(E_1 \wedge E_2) = \emptyset$, se verifique que $p(E_1 \vee E_2) = p(E_1) + p(E_2)$.

Generalizando esta última condición, se puede concluir que, para garantizar la existencia de la función p es necesario que, dados dos conjuntos de eventos $E_1 = \{E_{11}, E_{12}, \dots, E_{1n}\}$ y $E_2 = \{E_{21}, E_{22}, \dots, E_{2n}\}$, tales que $\wp(E_1) \succ \wp(E_2)$, se verifique la relación $\sum_{i \in E_1} p(E_{1i}) > \sum_{i \in E_2} p(E_{2i})$. A partir de esta condición, conocida como de aditividad finita, se puede concluir que la cuestión central que dio origen al sistema de axiomas de de Finetti fue la definición de la forma a asumir por la relación "más probable que"[238].

Uno de los principales problemas que presenta la aproximación deFinettiana en términos de cocientes de apuestas radica en la justificación de la extensión de la aditividad finita a la aditividad numerable. Teniendo en cuenta que, desde un punto de vista conceptual, la extensión de la propiedad de aditividad a conjuntos infinitos de eventos incompatibles constituye un ejemplo del riesgo que traen aparejadas las conclusiones que se obtienen de un sistema axiomático en el cual, basándose exclusivamente en conveniencias matemáticas, se han incluido axiomas sin una justificación en términos del concepto de probabilidad, de Finetti, compartiendo los reparos de von Mises y Kolmogorov, concluyó en la conveniencia de restringir el análisis a la aditividad finita[239].

---

[238] de Finetti (1949) obtuvo este resultado (conocido como "*teorema de extensión*") casi contemporáneamente y en forma independiente de la demostración más restringida de Horn; Tarsky (1948). En de Finetti (1955) figura un notable análisis de la estructura de la aditividad finita desarrollado en forma independiente de los trabajos de Bochner (1939) y Sobczyk; Hammer (1944a)(1944b). Bochner, en base a un coeficiente de divisibilidad apropiado, proporciona una descomposición de las probabilidades finitamente aditivas en: **i)** una componente discreta (masas concentradas en puntos de un conjunto numerable); **ii)** una componente aglutinada (concentrada en ultrafiltros que forman una clase numerable) y **iii)** una componente continua (infinitamente divisible) que puede ser descompuesta, con respecto a una medida finitamente aditiva dada, en dos subcomponentes: uno condensado y otro difuso (absolutamente continuo). La memoria de de Finetti contiene, además, una demostración del teorema de Radon-Nikodým para medidas finitamente aditivas, basada en una extensión de la noción de función de concentración de Lorenz-Gini. Esta extensión de de Finetti es utilizada como una medida de robustez de los métodos estadísticos Bayesianos y como un argumento para la demostración de las versiones exactas del teorema de Radon-Nikodým (ver Cifarelli; Regazzini (1987), Berti; Regazzini; Rigo (1992), Fortini; Ruggieri (1994)).

[239] de Finetti (1970): "*El supuesto de aditividad numerable (...) es, en la actualidad habitualmente aceptado; si bien no debe su origen a los axiomas de Kolmogorov (1933), tuvo su sistematización en dichos axiomas. Su éxito se debe en gran medida a la conveniencia matemática de convertir el cálculo de probabilidades en un mero traslado de la moderna teoría de la medida (...) Nadie ha proporcionado una justificación real de la aditividad numerable (que vaya más allá de su consideración simplemente como una mera 'extensión natural' de la aditividad finita)*" (p. 119).

Esta prevención respecto de la aditividad numerable constituye uno de los argumentos fundamentales que justifican la actitud constructivista-intuicionista de de Finetti opuesta a cualquier planteo de tipo axiomático cuantitativo abstracto[240] Sus críticas ser refieren, precisamente, a aquel aspecto de la axiomática que los matemáticos clásicos más apreciaban: la posibilidad del desarrollo de una disciplina que partiera de una noción abstracta de probabilidad[241]. Por definición, los axiomas son proposiciones arbitrarias seleccionadas libremente como "*verdades de partida*"[242] con la única obvia condición de no ser contradictorias, de modo que todas las conclusiones que se puedan derivar de ellas son consideradas como válidas con respecto al marco de referencia –abstracto- tomado como punto de partida[243]. Como lo ha demostrado la historia, este planteo y la imposibilidad de la aplicación estricta de las condiciones inherentes a la concepción de la probabilidad en términos de teoría de la medida, traen aparejados el peligro de

[240] El intuicionismo constructivista de comienzos del siglo XX postula que la matemática es una representación de las actividades humanas (a las que denomina "*intuiciones*") y que, por lo tanto, no tiene sentido considerar las implicaciones de axiomas arbitrarios basados en idealizaciones que no representan ninguna aplicabilidad específica y que pueden conducir a resultados inconsistentes. Luego, como las actividades humanas son de extensión necesariamente finita, los axiomas vinculados a la representación intuicionista sólo expresan requerimientos que involucran un número finito de operaciones (ver Heyting (1956), Bencerraf; Putnam (1983), Scozzafava (1984), Hill (1994)). A modo de resumen resulta significativa la siguiente expresión de Lad (1996): "*...el punto de vista subjetivo operacional se ubica en el contexto de una aproximación operacional y positivista a la ciencia, una aproximación analítica a la filosofía y una aproximación constructivista, finitísta e intuicionista a la matemática*" (p. 19).

[241] de Finetti (1952): "*De acuerdo con el punto de vista axiomático, se introducen axiomas para la probabilidad (por ejemplo, mediante ciertas ecuaciones lineales sobre un cierto campo), sin que nadie pueda imaginar, si no lo sabe por sí mismo, qué cosa significa el término 'probabilidad'. Se podría, por ejemplo, interpretar a la probabilidad como medida y pensar que las palabras 'medida' y 'probabilidad' son sinónimos*" (p. 686).

[242] de Finetti (1995, p. 111).

[243] Poincaré (1902): "*Para calcular una probabilidad y para dar sentido a este cálculo debemos admitir, en principio, una hipótesis o convención que siempre introduce cierta dosis de arbitrariedad*" (p. 15).

pretender que dichas conclusiones sean consideradas válidas aún en aplicaciones concretas para las cuales parezca cómodo adoptar un modelo abstracto (Borel (1914): "*Como han propuesto muchos autores, de la misma forma que se define la circunferencia en geometría ola masa en mecánica racional, se puede dar una definición puramente abstracta de probabilidad. A partir de esta definición se pueden deducir consecuencias lógicas absolutamente rigurosas.Pero, en la medida que se desee aplicar estas consecuencias a cualquier fenómeno real, se debe sustituir la probabilidad concreta de dicho fenómeno por la probabilidad abstracta. Entonces la incertidumbre inherente a toda medida concreta, retorna*" (p. 18))[244].

de Finetti (1931a) sostiene que las proposiciones contenidas en los axiomas no sólo deben ser formalmente consistentes, sino también intrínsecamente necesarias con respecto a una comprensión completa del concepto de probabilidad y estableció que la interpretación subjetivista no pretende reducir el conjunto de las probabilidades admisibles a una familia de probabilidades aditivamente numerables[245].

Los argumentos de de Finetti (1931a)(1931b)(1970) a favor de la aditividad finita (y en contra de la aditividad numerable) se basan en que, en la teoría subjetivista, las probabilidades están definidas por cocientes de apuestas individuales y que estas apuestas sólo tienen significado sobre un conjunto finito de eventos[246], que la idea de un conjunto infinito de apuestas resulta

---

[244] Borel continuó su discusión sobre la axiomatización y la aplicación de la matemática a la teoría de la probabilidad en los fascículos primero ("*Principes et formules classiques du calcul des probabilités*") y último ("*Valeur pratique et philosophie des probabilités*") de su "*Traité du calcul des probabilités et de ses applications*" (1929-1935).

[245] Ya en sus memorias de 1930, en ocasión de la discusión con Fréchet (1930a)(1930b) acerca de la aplicación del operador límite en la teoría de la probabilidad, de Finetti desarrolló observaciones críticas sobre la formulación usual del teorema de Cantelli (1917a)(1917b) y estableció la "*innecesidad*" de la aditividad numerable o completa.

[246] de Finetti (2006): "*...a partir de un conjunto infinito de eventos no se puede deducir ninguna conclusión práctica, es decir, ninguna afirmación verificable en un tiempo finito (...) y aún considerando series de series de pruebas o series de series de series de pruebas y así sucesivamente, no se obtiene lógicamente nada si no se considera una 'infinitid' (...) y no se obtiene ninguna posibilidad de afirmar cualquier cosa que sea verificable en un tiempo finito*" (p. 118).

bastante artificiosa y que su caracterización sólo podría ser justificable en términos asintóticos considerando el límite de una sucesión finita de n apuestas. Pero esta aproximación no podría ser considerada como equivalente a un número infinito de apuestas en un punto del tiempo (este caso infinito podría ser concebible sólo si se lo considerara como "*...un caso límite no-posible de los casos posibles*")[247]. Adoptar la aditividad numerable implica reconocer laimposibilidad de asignar una distribución uniforme sobre un conjunto numerable, como por ejemplo, $\{1,2,\dots,n\dots\}$ y aceptar que lo único admisible es, de acuerdo con la nomenclatura de de Finetti (1970), un conjunto (inaceptable) de "*particiones extremadamente desbalanceadas*". Supóngase que $p(i) = p > 0\ (i = 1,2,\dots,n)$, entonces se verificará que $\sum_{i=1}^{\infty} p(i) = \infty$ cuando, de acuerdo con la axiomática, debe ser $\sum_{i=1}^{\infty} p(i) \leq 1$. Asimismo, si se supone que $p(i) = 0\ (i = 1,2,\dots,n\dots)$, entonces se verificará que $\sum_{i=1}^{\infty} p(i) = 0$ cuando debe ser $\sum_{i=1}^{\infty} p(i) = p(\Omega) = 1$[248].

El caso típico de aditividad numerable es el que surge del comportamiento en el límite de las sucesiones de variables aleatorias. Las primeras propuestas de de Finetti en este sentido se refieren a la igualdad de la probabilidad y el límite de la frecuencia relativa en la ley fuerte de los grandes números de Cantelli. De su discusión con Fréchet y su interpretación de la probabilidad a partir delateoría de la medida, su conclusión es que la aditividad numerable no es una condición necesaria en su axiomática. Que, si bien la condición de coherencia no requiere la aditividad numerable, no la contradice. Esta conclusión lo conduce al planteo de la "*paradoja de la no-conglomerabilidaden la selección al azar de un número natural*" (1930c).

---

247 de Finetti (1995, p. 111).

248 En desacuerdo con la posición de de Finetti acerca de la inadmisibilidad de la asignación de una distribución uniforme a un conjunto numerable de resultados posibles, Adams (1965)(1966) demostró que la introducción de un elemento subjetivo que conduce a la adopción de una distribución asimétrica como sustituto de la distribución uniforme que prescribe la aplicación del principio de indiferencia de la interpretación logicista, permite concluir que la defensa de la distribución uniforme por parte de de Finetti es más una defensa de la interpretación logicista que de la interpretación subjetivista. Con relación a esta cuestión de la aditividad finita y su relación con la aditividad numerable y con la condición de coherencia, se ha publicado una gran cantidad de trabajos entre los que merecen ser destacados Heath; Sudderth (1978), Dubins (1975), Hill (1980), Cohen (1992), Schervisch; Seidenfeld; Kadane (1984), Hill; Lane (1985) y Gillies (2000).

El núcleo de la teoría deFinettiana y la conciliación de su filosofía neo-Bayesiana con su matemática están contenidos en la noción de intercambiabilidad y el teorema de representación[249]. Se dice que una sucesión $\{E_1, E_2, \dots, E_n, \dots\}$ es de eventos intercambiables si, para cualquier subconjunto finito $\{E_1, E_2, \dots, E_n\}$, la distibución de probabilidades conjunta es la misma[250].

La finalidad de la condición de intercambiabilidad fue eliminar la idea de probabilidad objetiva con valores constantes desconocidos y proponer una asignación de probabilidades variable de acuerdo con la regla de condicionamiento Bayesiano según la cual, dadas una "hipótesis" H, con una probabilidad "a priori" $p(H)$ y una "evidencia" E, de acuerdo con los postulados del teorema de Bayes, la probabilidad "a posteriori" de H, condicionada por la evidencia E, será[251]:

$$p(H\ /\ E) = \frac{p(E\ /\ H)}{p(E)} p(H) =$$

$$= verosimilitud \times probabilidad\ a\ priori\ de\ H$$

Sea un conjunto de sucesiones de repeticiones de un evento, tales que para cada sucesión de longitud n, la probabilidad de ocurrencia de$Y$ "éxitos" es la misma. Dada una probabilidad p, desconocida, de ocurrencia de un "éxito"

---

[249] La teoría de la probabilidad de de Finetti se basa en gran medida en los trabajos de Castelnuovo (1918) y Lévy (1925)(1929), los cuales le proporcionaron los argumentos apropiados para obtener los primeros resultados sobre la clase de los eventos intercambiables y las leyes probabilísticas para procesos estocásticos continuos en el dominio del tiempo.

[250] La condición de intercambiabilidad había sido estudiada previamente por Ramsey en un trabajo no publicado del final de la década de 1920, incluido posteriormente en "*Notes on philosophy, probability and mathematics*" (1991). de Finetti introdujo la noción de "*eventos equivalentes*" en el Congreso Internacional de Matemática de Boloña, de 1928. La denominación de "*eventos intercambiables*" se debe a Fréchet (1939).

[251] Obsérvese que, como se comentó al comienzo de esta sección, si bien la intercambiabilidad le quita rigidez a la condición de independencia, no rechaza la estabilidad estadística de los fenómenos.

individual[252] y, suponiendo que estos eventos se rigen por la ley de la razón insuficiente, es decir, suponiendo que la variable p se distribuye de acuerdo con una función uniforme, la "*regla de sucesión de Laplace*" permite calcular las probabilidades "a posteriori" mediante el esquema de condicionamiento, según el cual la probabilidad de ocurrencia de un nuevo "éxito", condicionada por Y "éxitos" en $n$repeticiones, está dada por $\frac{Y+1}{n+2}$[253].

De acuerdo con de Finetti (1930), dada una sucesión de eventos independientes con probabilidad de ocurrencia de un "éxito" individual constante, p, la probabilidad de que en una sucesión de n repeticiones ocurran Y "éxitos" es $p^Y(1-p)^{n-Y}$ y, teniendo en cuenta que existen $\binom{n}{Y}$ de estas sucesiones, la probabilidad de ocurrencia de la frecuencia relativa $Y_{(n)} = \frac{Y}{n}$ será $\omega_Y^{(n)} = \binom{n}{Y} p^Y(1-p)^{n-Y}$. Las condiciones de independencia y de equiprobabilidad son necesarias y suficientes para este caso particular de intercambiabilidad.

En términos más generales, dadas m urnas que contienen bolillas rojas y azules, tales que laprobabilidad de obtener una bolilla roja efectuando una extracción al azar de la i-ésima urna sea p_i y suponiendo que la probabilidad "a priori" de seleccionar la i-ésima urna sea π_i, la probabilidad de obtener Y bolillas rojas en n extracciones al azar con reposición está definida por una "*mezcla de esquemas Bernoullianos*" de la forma:

$$\binom{n}{Y} \sum_{i=1}^{m} \pi_i p_i^Y (1-p_i)^{n-Y}$$

Supóngase, ahora, que el problema consiste en seleccionar un número $p \in [0,1]$ de acuerdo con una función de densidad $f(p)$, entonces la probabilidad intercambiable anterior asumirá la forma:

---

[252] La probabilidad desconocida, p, representa una variable aleatoria tal que, condicionados por sus posibles valores, los eventos son estocásticamente independientes. Es decir, tales que $p(E_1 \cap E_2 \cap \ldots \cap E_n / p) = \prod_{j=1}^{n} p(E_j / p)$. Como antecedente a las definiciones de independencia e independencia condicionada cabe mencionar también a Haag (1924) y van Deuren (1934).

[253] Ver Sec. 11.

$$\binom{n}{y}\int_0^1 f(p)p^Y(1-p)^{n-Y}dp$$

Viceversa, de acuerdo con el "*teorema de representación*", dada una función de probabilidades intercambiable para todo n, existe una única función de distribución $F(p)$ tal que:

$$\omega_Y^{(n)} = \binom{n}{Y}\int_0^1 p^Y(1-p)^{n-Y}dF(p)$$

(un resultado que conduce a la solución del problema de la inducción de Hume).

de Finetti basó su demostración del teorema de representación en el método de las funciones características de probabilidades intercambiables[254]. De acuerdo con su notación original, se verifica que:

$$\omega_Y^{(n)} = \binom{n}{Y}\int \xi^Y(1-\xi)^{n-Y}d\Phi =$$

$$= \binom{n}{Y}\int (Y-n\xi)^{Y-1}(1-\xi)^{n-Y-1}\Phi(\xi)d\xi$$

Lo que permite concluir que, a partir de la definición de la función característica, es posible determinar la distribución límite de la frecuencia relativa de una sucesión intercambiable de eventos y que, en consecuencia, todas las probabilidades $\omega_Y^{(n)}$ quedan determinadas por la función característica límite. El objeto de de Finetti "*...era demostrar cómo y en qué sentido era posible realizar una inferencia de las probabilidades a partir de la información que proporcionan las frecuencias, de acuerdo con su aproximación subjetivista*" (von Plato (1994, p. 248)). Es decir, calcular la distribución de probabilidades de la variable frecuencia relativa, $Y_{(n)} = \frac{Y}{n}$, para una función de probabilidades intercambiable. En "*Funzione*

[254] Su demostración se basó en las llamadas "*fórmulas de Lévy*", contenidas en el "*Calcul des probabilités*" (1925), las cuales constituyen el nexo entre las distribuciones de probabilidades y las correspondientes funciones características.

caratteristica di un fenómeno aleatorio” (1928b) demuestra que, si se cumple la condición de intercambiabilidad, la distribución límite de la variable frecuencia relativa queda determinada, es decir que, de acuerdo con la forma inversa del teorema de Bernoulli, se verifica la ley de los grandes números y que, en este sentido, la distribución de probabilidades para eventos intercambiables, se aproxima formalmente (no conceptualmente) al caso de eventos independientes con la frecuencia relativa como representativa de la probabilidad “a priori” de ocurrencia de un “éxito”.Este resultado, curiosamente, parecería revelar la aceptación, en la concepción subjetivista, de la existencia de unlímite de las frecuencias relativas. Pero, “*...obviamente, expresando el razonamiento inductivo de esta forma es imposible ‘demostrar’ la validez del principio de inducción y, por lo tanto, del principio según el cual el valor de la probabilidad debería ser próximo a la frecuencia observada (...) este principio es admisible ‘sólo en casos particulares’ (...) si la probabilidad es puramente subjetiva, nada puede obligar a estimarla como próxima a la frecuencia. A partir de este supuesto lo máximo que se puede aceptar es que una asignación de este tipo, por una razón de simple coherencia, es una consecuencia de la asignación inicial cuando ésta satisface ciertas condiciones naturales, conocidas como ‘condiciones de intercambiabilidad’*” (de Finetti (1989, pp. 100-101)).

La interpretación habitual de lacondición de intercambiabilidad como un esquema con una probabilidad de “éxito” constante desconocida carece de sentido en una interpretación subjetivista en la medida que las condiciones de independencia y de constancia de la probabilidad no se verifican, ya que: **i)** pueden ocurrir los mismos resultados a partir de diferentes hipótesis; **ii)** la probabilidad de los resultados de las extracciones depende de las probabilidades de las hipótesis; **iii)** la probabilidad de cada una de estas hipótesis depende de las frecuencias observadas y, teniendo en cuenta que no se conoce la composición de la urna, **iv)** la probabilidad de las distintas composiciones puede variar. En conclusión, la noción de intercambiabilidad reemplaza el esquema de eventos independientes con probabilidad constante pero desconocida y, de acuerdo con una interpretación subjetivista, es equivalente al concepto de independencia condicionada por una partición de hipótesis[255].

A partir de la definición de intercambiabilidad ilimitada[256], el teorema de representación constituye una justificación exclusivamente formal de dicho esquema y demuestra que, "*...suponiendo que esta expresión tuviera sentido"*, sería equivalente al caso "*…deun fenómeno aleatorio 'con eventos independientes de probabilidad constante pero desconocida, p', en el que la distribución límite es interpretada como la ley (distribución) de probabilidades de la probabilidad desconocida p*" (1930a, p. 89)[257]. Un resultado que obviamente reduce el contenido metafísico de lainterpretación objetivista a favor de una interpretación subjetivista[258].

Como se verá en la Sec.24, Khinchin (1932b) planteó una interpretación alternativa de la condición de intercambiabilidad que le permitió obtener una versión más simple de la ley fuerte de los grandes números. A partir de estos resultados, de Finetti (1933b) propuso la siguiente forma general de la

[255] de Finetti (1995): *"...mediante la noción de intercambiabilidad procuré recuperar aquello que se entiende como 'probabilidad desconocida' y demostré que se puede lograr, porque realizar una mezcla de 'probabilidades desconocidas' respecto de las cuales todas las pruebas sean independientes y de igual probabilidad equivale a satisfacer la condición de intercambiabilidad"* (p.155).

[256] Paradójicamente, la condición necesaria de infinitud de las sucesiones de variables intercambiable parece contradecir las consideraciones de de Finetti (1931)(1970) con respecto al axioma de aditividad numerable. Los conceptos de "*intercambiabilidad ilimitada*" y "*limitada*" (utiizados por de Finetti (1937)(1938)(1969)) están asociados al cumplimiento o no de la condición de repetibilidad infinita del feómeno.

[257] de Finetti (1930b) define un "*fenómeno aleatorio*" como aquél formado por un número arbitrario de repeticiones, tal que el orden de ocurrencia de sus resultados es completamente aleatorio e interpreta a un "*evento*" como un hecho único. De modo que utiliza el término "*fenómeno*" como un concepto general en el cual cada repetición constituye un "*evento*".

[258] Debe tenerse en cuenta que, a partir de una aproximación subjetivista, la condición de intercambiabilidad y sus generalizaciones no están vinculadas con la independencia causal, sino con la independencia estocástica, la cual implica *"...no solamente la ausencia de cualquier relación causal, sino también la ausencia de cualquier influencia sobre nuestro juicio respecto de la asignación de la probabilidad"* (de Finetti (1934, p. 2000). Para un análisis del teorema de representación en el contexto de la justificación del razonamiento inductivo, ver Mondadori (1979), Mondadori; Morini (1982), Fürst (1982).

probabilidad de ocurrencia de un evento E intercambiable:

$$p(E) = \int p_\xi(E) d\Phi(\xi)$$

(donde $p_\xi(E)$ denota la probabilidad de ocurrencia de un evento E independiente) [259]. En 1933c demostró que, dada una sucesión $\{Y_2, Y_2, \dots, Y_n, \dots\}$ de variables aleatorias intercambiables, la variable $Y_{(n)} = \frac{1}{n}\sum Y_i$ se distribuye de acuerdo con una función $\Phi_n(\xi)$ tal que $\lim_{n\to\infty} \Phi_n(\xi) = \Phi(\xi)$, es decir, tal que cumple con una ley de los grandes números cuya única condición es la existencia de una distribución límite de las frecuencias relativas. Una ley de los grandes números obtenida a partir de la definición de una sucesión intercambiable de eventos, que es más general que la obtenida a partir de la condición de independencia, en la medida que no supone la existencia de un límite único para las frecuencias relativas.

En el ámbito del análisis dinámico, esta condición de existencia de una distribución límite de las frecuencias relativas es equivalente a la condición de estacionariedad. Las probabilidades que reúnen esta condición se denominan ergódicas y su mezcla genera, precisamente, probabilidades estacionarias, de las cuales las probabilidades intercambiables son un caso particular[260]. Teniendo en cuenta que, a partir del modelo objetivista, el concepto de probabilidad ergódica constituye la mayor generalización de la propiedad de independencia estocástica, se puede considerar, entonces, a la representación de probabilidades estacionarias mediante mezclas únicas de probabilidades ergódicas como la expresión más general del principio de reducción a la intercambiabilidad.

La teoría ergódica comienza con el intento de Boltzmann (1868) de representar la distribución de probabilidades de un proceso estocástico en términos de promedios en el dominio del tiempo. Dados un proceso estocástico $\{X(t)\}$, continuo en el dominio de los estados y continuo en el

---

[259] De acuerdo con el *"postulado de suficiencia"* y el consecuente teorema de Johnson (1932), la función $d\Phi(\xi)$ es la función simétrica de Dirichlet.

[260] Fue Khinchin (1932a)(1932b)(1932c)(1932d) quien demostró que, en la medida que la intercambiabilidad es un caso particular de la condición de estacionariedad, la ley de los grandes números de de Finetti constituye un resultado particular del caso estacionario.

dominio del tiempo y una función $g(X)$ en el dominio $\Omega(X)$ de los estados, su valor esperado está definido por $\int_{\Omega} g(x)f_X(x)dx$, donde $f_X(x)$ denota una función de densidad. Asimismo, sea una región $A\epsilon\Omega(X)$, sustituyendo la función g por una función indicador de la permanencia del proceso en la regiónA, I_A,se verificará entonces que $\int_{\Omega} I_A(x)f_X(x)dx = p(A)$. Sea, por otra parte, $T(t,X)$ la ley que determina la trayectoria del proceso $\{X(t)\}$, de modo que, si $X(0) = x$, entonces $X(t) = T(t,X)$ y supóngase que esta trasformación sea medible y "measure preserving", es decir tal que $M[T^{-1}(t,A)] = M(A)$ (donde $M(\cdot)$ denota la clase de todas las medidas de probabilidad). Entonces, de acuerdo con el teorema ergódico individual de Birkhoff (1931), el promedio en el dominio del tiempo de la función g está dado por $m[g(x)] = \lim_{t\to\infty} \frac{1}{t}\int_0^t g[T(y,x)]dy$. Sustituyendo la función g por una función indicador I_A, este promedio es asimilable a una ley de los grandes números y puede ser interpretado como el tiempo promedio a permanecer en la región Apor el proceso que se encontraba en el estado x en el momento 0. El problema fundamental de la teoría ergódica consiste, entonces, en demostrar que el promedio del proceso en el dominio del tiempo existe (condición de estacionariedad), que es único (es decir, independiente del estado inicial del proceso) y que el promedio en el dominio de los estados puede ser calculado como un promedio en el dominio del tiempo:

$$\int_{\Omega} g(x)f_X(x)dx = \lim_{t\to\infty} \frac{1}{t}\int_0^t g[T(y,x)]dy$$

(es decir, demostrar que la probabilidad $p(A)$ puede ser expresada objetivamente de forma unívoca como el promedio del tiempo a permanecer por el proceso en la región A).

Sea, en particular, un espacio Ω formado por infinitas sucesiones binomiales $E = \{E_1, E_2, \dots, E_n, \dots\}$, sea $x_n(E)$ el n-ésimo elemento de una sucesión de E y sea una transformación T similar a la mencionada precedentemente, tal que $x_n(TE) = x_{n+1}(E)$, de modo que $x_{n+1}(E) = x_1(T^nE)$. Esta transformación representa la repetición de la prueba de la sucesión binomial $E = \{E_1, E_2, \dots, E_n, \dots\}$ y su aplicación proporciona el resultado a obtener en la repetición siguiente, es decir, la sucesión $\{E, TE, T^2E, \dots\}$ define una realización o trayectoria del proceso[261]. Se dice que una probabilidad $p(\cdot)$sobre Ω es una medida estacionaria si, para todo conjunto $A \subset \Omega$, se verifica que $p(T^{-n}A) = p(A)(n = 1,2, \dots)$, es decir, si las funciones de probabilidades sobre sucesiones finitas de eventos permanecen invariantes en el tiempo[262].Supóngase que, en particular, $E_i = 1$ y $-E_i = 0$ $(i = 1,2, \dots)$ representen los resultados posibles de un fenómeno binomial, la propiedad de estacionariedad es condición necesaria y suficiente para poder asegurar que el límite de la frecuencia relativa del resultado E, $\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} E_i$, existe y que, dada una función g integrable, el límite del promedio en el dominio del tiempo, $\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} g(T^i E) = E(g) = m(g)$, también existe y es independiente de E. Lo cual demuestra que todas las realizaciones del proceso poseerán las mismas propiedades asintóticas con probabilidad igual a 1. La propiedad de ergodicidad implica, entonces, que un proceso ergódico posee en su totalidad las mismas propiedades asintóticas, es decir que no puede ser separado en partes y garantiza la unicidad de los límites de las frecuencias relativas del resultado E[263].

Como corolario de esta afirmación se puede concluir que, como se mencionó en la Sec. 17, las leyes de los grandes números soncasos especiales del teorema ergódico[264], que, en particular, el teorema de Bernoulli es el primer teorema ergódico y que el fundamento en el que se basa la interpretación frecuencista de la probabilidad no es la condición de independencia, sino la propiedad más general de ergodicidad de la sucesión de repeticiones generadas por el fenómeno aleatorio. No obstante, debe tenerse en cuenta

---

[261] Se dice que una sucesión $E = \{E_1, E_2, \dots\}$ estrictamente estacionaria es ergódica, si puede ser representada por una transformación ergódica.

[262] Si, como en este caso, se supone un proceso discreto en el dominio del tiempo y discreto en el dominio de las variables, la asignación de probabilidades como promedios en el dominio del tiempo se asimila a una interpretación frecuencista.

[263] La mayoría de los teoremas de la teoría de los procesos estocásticos están dirigidos a demostrar que, a partir de ciertas apropiadas hipótesis analíticas, algunas propiedades referidas a sucesiones finitas de observaciones se verifican casi-con-certeza.

[264] En particular la ley débil de los grandes números corresponde al llamado *“teorema ergódico en el valor medio”* y la ley fuerte al llamado *“teorema ergódico individual”*.

que con respecto a los fundamentos de la interpretación ergódica de la ley de los grandes números, los intentos de definición esencialmente objetivistas no logran caracterizar a la intercambiabilidad como una condición que trasciende a la independencia estocástica y sugieren que es una propiedad formal y, en apariencia, completamente desvinculada de cualquier aproximación al concepto de probabilidad. Alternativamente a esta visión objetivista el teorema de representación admite una interpretación según la cual la condición de intercambiabilidad (es decir de estacionariedad) es necesaria y suficiente para poder asegurar la existencia del límite, desconocido, de la frecuencia relativa y que, en términos de una aproximación subjetivista este desconocimiento de la frecuencia relativa puede ser caracterizado por una mezcla sobre sus posibles valores en la que las ponderaciones de las distintas hipótesis surgen como asignaciones personales de los supuestos verdaderos valores de las probabilidades mediante el proceso de condicionamiento Bayesiano y permiten obtener una evaluación intersubjetiva de la probabilidad.

En resumen, si bien a partir de una aproximación subjetivista, la propiedad de intercambiabilidad permite resolver algunas controversias referidas a las relaciones entre probabilidad y frecuencia y al teorema de Bayes, el teorema de representación por sí mismo no logra justificar los postulados de la ley de los grandes números, no por una insuficiencia del método sino por las exigencias esenciales de la interpretación subjetivista, que condiciona dicho teorema a una asignación coherente de la probabilidad inicial. Debe tenerse en cuenta que, como ya se mencionó, de Finetti (1930) denominó "*evento*" a un hecho único definido en forma estricta y utilizó una noción de "*fenómeno aleatorio*", más general, que considera a un evento como la consecuencia de "*…cada prueba repetida bajo condiciones de contorno homogéneas respecto de dicho fenómeno*". En otros términos, un fenómeno aleatorio deFinettiano es aquél cuyas repeticiones generan una sucesión de eventos intercambiables. Pero, dadoque desde una aproximación subjetivista, los eventos son singulares en tiempo y espacio, la hipótesis que las repeticiones de un evento forman parte del mismo fenómeno es sólo una conjetura, lo cual conduce a la inevitable conclusión que la condición de intercambiabilidad admite una representación de características exclusivamente subjetivas. Por otra parte, de Finetti (1934) consideró que la ya mencionada interpretación de la intercambiabilidad como asimilable a la independencia estocástica condijonada por una probabilidad constante pero desconocida, carece de sentido, que la probabilidad varía de una prueba a otra de acuerdo con la experiencia adquirida por el observador y que la que es desconocida y con una distribución que varía en base a los resultados de las realizaciones del evento es la frecuencia límite.

El teorema de descomposición ergódica[265] considera la relación entre los conceptos generales de ergodicidad y estacionariedad y demuestra que toda sucesión estacionaria admite una representación integral en función de probabilidades estacionarias o, lo que es lo mismo, una única descomposición ponderada en partes ergódicas (es decir, con idénticas propiedades asintóticas para cada una de las partes, pero diferentes para las distintas partes), cada una de las cuales puede ser definida como un subconjunto del conjunto de infinitas sucesiones y cuyas ponderaciones están dadas por los tiempos promedio de permanencia del proceso en cada una de las partes. En el caso de eventos simples, las propiedades asintóticas están caracterizadas por los postulados de la ley de los grandes números dentro de cada parte, pero son distintas para las diferentes partes y la probabilidad total está dada por la mezcla ponderada de dichos límites, donde las ponderaciones están definidas por las medidas de las particiones.

Como corolario de los resultados anteriores, se obtiene que las sucesiones de eventos intercambiables son estacionarias, pero con propiedades asintóticas variables debido al condicionamiento de las probabilidades y que las sucesiones de eventos independientes están asociadas a una medida ergódica. Luego, dado que las medidas Bernoullianas satisfacen la primera ley de los grandes números y que la propiedad de ergodicidad implica un comportamiento límite idéntico para todas las sucesiones, se puede concluir que las probabilidades Bernoullianas son ergódicas y, por lo tanto que el teorema de representación clásico es un caso particular del teorema de descomposición ergódica, en el que la probabilidad p define una componente ergódica formada por el conjunto de sucesiones cuya frecuencia relativa converge a p[266].

---

[265] Debido a Koopman (1931), Birkhoff (1931) y von Neumann (1932a)(1932b)(1932c).

[266] La demostración rigurosa de que el teorema de representación constituye un caso particular del teorema ergódico se debe a Ryll-Nardzewski (1957). Ver, además, Freedman (1963) y Dynkin (1978).

Ahora bien, si el proceso es tal que admite una descomposición por la existencia de una partición ergódica que depende exclusivamente de las propiedades de la ley que determina la trayectoria del proceso, las ponderaciones de la descomposición asumen un significado físico, lo cual indujo a algunos autores a postular que su distribución no dependía de las características subjetivas del condicionamiento Bayesiano. Pero, dadas la imposibilidad de determinar la ergodicidad de una sucesión estacionaria y la naturaleza axiomática de las premisas en las que se basa el teorema de descomposición y en contradicción con estepostulado, se puede concluir que su aceptación reviste un carácter exclusivamente subjetivo.

Por otra parte, el hecho que la propensión sea una propiedad física del diseño del experimento basada en el supuesto de un comportamiento estructural dado del fenómeno y que su interpretación del concepto deprobabilidad está referida a eventos individuales y la consideración de una definición de las propensiones basada en las propiedades dinámicas de un fenómeno que posee una "verdadera" trayectoria y que, por lo tanto, admite una explicación en términos de mecánica clásica, permite afirmar que es posible obtener una justificación analítica de la interpretación propensionalista de la teoría ergódica. La cual conduce a la conclusión que la teoría ergódica está afectada por las mismas características metafísicas del modelo propensionalista en el cual la asignación de probabilidades es inevitablemente subjetiva.

23.- Jules Henri Poincaré, Marian Wilhelm Theofil von Smoluchovski, Frederich Ferdinand Hopf y el método de las funciones arbitrarias

Como una alternativa a la descomposición ergódica,para justificar la existencia de una definición objetiva de probabilidad única en el ámbito de los fenómenos dinámicos que admiten leyes determinísticas que determinan su trayectoria, Poincaré (1896) introdujo el método de las funciones arbitrarias, el cual fue generalizado en sus aspectos matemáticos por Fréchet (1952), quien demostró que la condición suficiente para su aplicación es contar con un continuo de trayectorias del proceso y una función de densidad sobre este conjunto la cual, bajo ciertas condiciones inherentes al mismo, converge a una única distribución final[267].

[267] Borel (1906b) fue uno de los primeros en adoptar el método de Poincaré. Una interpretación alternativa del método de las funciones arbitrarias figura en la tesis doctoral de Reichenbach (1915).

Poincaré planteó una noción de aleatoriedad como inestabilidad de un sistema no como consecuencia de la ignorancia del observador, sino de las condiciones que afectan su ocurrencia. Por su parte,von Smoluchovski (1918) definió la aleatoriedad como una relación causal en la cual el efecto y se asume como una función f de una variable causa x, de modo que $y = f(x)$, donde el efecto y depende de las variaciones infinitesimales (inestabilidades) de la causa x,suponiendo que f garantiza la unicidad de la distribución de probabilidades de $y$motivada en la inestabilidad de x. Su objeción al método de las funciones arbitrarias –referida a la insuficiente explicación que este método proporciona sobre el origen de la aleatoriedad, es decir, sobre el origen de la inestabilidad aleatoria de x- se resume en los siguientes conceptos: "*Nos parece que un resultado extremadamente importante también para los filósofos –aunque sólo pueda ser demostrado en el ámbito restringido de la física matemática- es que el concepto de probabilidad, en el sentido usual de probabilidad de eventos que observan una ley de frecuencia, posea un contenido estrictamente objetivo; que se pueda precisar el concepto y el origen de la aleatoriedad aún negando el determinismo y que la ley de los grandes números no sea un principio místico, ni un resultadopuramente empírico, sino que sea una consecuencia matemática simple de la forma particular de representar el vínculo causal en este tipo de casos*" (p. 262). Este análisis sobre los fundamentos filosóficos de la probabilidad física objetiva tuvo su culminación, como resultado de la síntesis del método de las funciones arbitrarias y la teoría ergódica, en la obra de Khinchin (1933) y Hopf (1934)(1936).

El objetivo de Hopf (1934) consistióprecisamente en "*...determinar el verdadero origen de las leyes de probabilidades*" (p. 51). Propuso una nueva interpretación de las definiciones de Poincaré y von Smoluchwski fundada en la conjetura que sostiene: **i)** que la justificación analítica del método de las funciones arbitrarias se encuentra exclusivamente en la teoría ergódica en la cual, como sevio, el supuesto de existencia de una distribución inicial es reemplazado por el de continuidad absoluta y esa noción de independencia física permite formular una "*ley física de los grandes números*" basada en la idea que la importancia de la teoría ergódica radica en el supuesto de la estabilidad de las frecuencias; **ii)** quela condición de independencia es considerada como un resultado derivado de dicha propiedad de continuidad[268]; **iii)** que en ciertos casos la naturaleza ergódica de un fenómeno dinámico puede ser determinada deductivamente y, en consecuencia, **iv)** que el método de las funciones arbitrarias permite, en dichos casos, definir probabilidades objetivas[269].El fundamento teórico de la conjetura de Hopf fue proporcionado por Sinai (1973)(1976), quien demostró que si el dominio de los estados de un proceso estocástico es sometido a una partición finita en macro-estados, la sucesión de macro-estados compone un proceso Bernoulliano, es decir, una sucesión en la que los macro-estados son estocásticamente independientes y su probabilidades están determinadas exclusivamente por la ley que rige la trayectoria del proceso.

Ahora bien, debe tenerse en cuenta que, como se mencionó al comienzo de esta sección, la conjetura de Hopf ylos resultados obtenidos por Sinai se refieren exclusivamente a fenómenos cuyos comportamientos admiten una explicación en términos de mecánica clásica. Por otra parte, si como ocurre en los fenómenos fácticos, la condición de continuidad absoluta no se verifica entonces, dada la inevitable presencia de factores aleatorios, una (única) realización del proceso no puede ser considerada como una consecuencia necesaria de su representación determinística. Debe tenerse en cuenta además que, de acuerdo con Khinchin (1954), una condición necesaria para la convergencia de las frecuencias relativas es que los valores esperados de las funciones en el dominio de los estados sean independientes de la distribucióninicial y que esta condición no se cumple para distribuciones singulares, de modo que para el caso de trayectorias únicas, los postulados de los teoremas ergódicos y, en consecuencia, de las leyes de los grandes números no se verifican.

---

[268] La explicación de la independencia estadística está contenida en su *"Teorema de independencia: Dados un evento A, estadísticamente regular respecto de un mecanismo y un evento A' que posee las mismas propiedades respecto de otro mecanismo y denotando por $L(A)$ y $L(A')$ los límites de las frecuencias relativas de A y A', respectivamente, entonces el evento simultáneo $A \times A'$ siempre es estadísticamente regular respecto del mecanismo producto y el límite de su frecuencia relativa es tal que $L(A \times A') = L(A)L(A')$"* (p. 76).

[269] No obstante, debe tenerse en cuenta que la interpretación subjetivista de la condición de continuidad absoluta conduce a la conjetura de Savage (1973) según la cual las funciones de probabilidades sólo pueden ser derivadas de otras funciones de probabilidades.

En realidad, el reemplazo de la hipótesis ergódica de la dinámica clásica por una formulación probabilísticay la introducción de una dosis de aleatorismo en su teoría física que, en principio, resulta incompatible con el determinismo inherente a la mecánica clásica, se inició con Bernstein (1912). En sus "*Ausschaltung der Ergodenhypothese aus der physikalischen Statistik*" (1920), "*Über die gegenwärtige Krise der Mechanik*" (1931b), "*Über kausale und statistiche Gesetzmässigkeit in der Physik"* (1930) y "*Wahrscheinlichkeitsrechnung ind ihre Anwendung in der Statistik und theoretischen Physik*" (1931a), von Mises postula que no es possible imaginar una descripción del comportamiento de los fenómenos que no incluya elementos de la teoría de la probabilidad; que, en tanto que las representaciones en términos de mecánica clásica son algorítmicas, las sucesiones de eventos que componen la trayectoria de un proceso dinámico no admiten ninguna representación algorítmica y que, por lo tanto, son incompatibles en un paradigma determinístico[270]. Estas ideas ejercieron una influencia decisivaen la transición de la ergodicidad de los procesos dinámicos a la ergodicidad como concepto puramente probabilístico y a la formulación de la teoría de los sistemas dinámicos abstractos (cuya estructura definitiva se debe fundamentalmente a Koopman (1931), von Neumann (1932a)(1932b)(1932c), Birkhoff (1931)(1932), Birkhoff; Koopman (1932) y Kolmogorov (1933b).

24.- Alexander Iacolevich Khinchin y la ley del logaritmo iterado

Bernstein (1912) y Hausdorff (1913)(1914) obtuvieron una demostración del teorema de Borel utilizando exclusivamente argumentos de la teoría de la medida[271]**.** En particular, Hausdorff analizó el comportamiento asintótico de los valores absolutos de los desvíos de la frecuencia relativa del resultado "1"

---

[270] Esta primera postulación acerca de la contradicción entre la mecánica clásica y la representación probabilística de un proceso dinámico se resume en el siguiente párrafo: "*En la física estadística permanece una profunda contradicción que aún no ha sido resuelta: que a partir de un cierto punto de vista, se interpreta a la evolución de los eventos como completamente determinada mediante las ecuaciones de la física (las ecuaciones diferenciales de la dinámica del sistema), aunque se piensa que se pueden postular diferentes supuestos acerca de la evolución de los mismos desde un punto de vista completamente distinto*" (von Mises (1920, p. 227)).

[271] Ver Doob (1989)(1994), von Plato (1994).

de la "*ley cero-uno*" respecto de $\frac{1}{2}\left(\left|\frac{Y(1)}{n}-\frac{1}{2}\right|\right)$ y demostró que, con probabilidad igual a 1, la cota superior es de la forma $X-np=O\left(n^{\varepsilon+\frac{1}{2}}\right)$ y Hardy; Littlewood (1914) demostraron que, con probabilidad igual a 1, $X-np=O\left(\sqrt{nln(n)}\right)$ excepto para un conjunto de medida nula[272] y Cantelli (1917a) generalizó el resultado de Borel para todo $0<\theta<1$.

Khinchin (1923) modificó el resultado de Hardy-Littlewood, obteniendo una cota superior igual a $\sqrt{nlog[log(n)]}$, expresión conocida como "*ley del logaritmo iterado*"[273]. En 1924 postuló una formulación probabilística a esta interpretación geométrica según la cual, dada una sucesión de variables aleatorias $X_i\colon b(1,\theta)$, existe una función:

$$f(n)=\sqrt{2n\theta(1-\theta)log[log(n)]}$$

que caracteriza la tasa de convergencia, tal que, para todo ε, δ, un n_0 y una variable $X_{(n)}=\sum_{i=1}^{n}X_i$, se verifica que $p\left(\left|\frac{X_{(n)}-\frac{n}{2}}{f(n)}\right|<1+\varepsilon\right)=1-\delta$ y $p\left(\left|\frac{X_{(n)}-\frac{n}{2}}{f(n)}\right|>1-\varepsilon\right)=1-\delta$[274].

Como una continuación inmediata del teorema de Khinchin se demuestra que, dada una sucesión $\{X_i\}(i\geq 1)$ de variables independientes con distribución Normal del tipo $N(0,\sigma^2)$, a partir del teorema de Azuma (1967)[275] se obtiene que:

---

[272] Como en su trabajo Hardy-Littlewood no utilizaron nomenclatura probabilística, asimilaron esta condición a la expresión "*la proposición casi siempre es verdadera*". Bernstein consideró a este conjunto de medida nula como *"sin significado físico"*.

[273] Para muchos autores, este es uno de los resultados más hermosos de la teoría de la probabilidad.

[274] Ver Kolmogorov (1929b), Feller (1943) y Lévy (1937-1954).

[275] Ver Chover (1966), Stout (1974).

$$\lim_{n\to\infty} sup \frac{\left|\sum_{i=1}^{n} X_i\right|}{\sqrt{2\sigma^2 \log[\log(\sigma^2)]}} \leq 1$$

converge casi-con-certeza.

Este resultado es un caso particular del teorema de Hartman-Winter (1941) según el cual, dada una sucesión $\{X_i\}(i \geq 1)$ de variables centradas iid con varianza finita, se demuestra que:

$$\lim_{n\to\infty} sup \frac{\sum_{i=1}^{n} X_i}{\sqrt{2nE\left(X_i^2\right) \log[\log(n)]}} = 1$$

converge casi-con-certeza.

A partir del resultado de Khinchin, Feller (1943)(1946) demostró que, dada una sucesión de variables independientes con distribución $N(0, \sigma^2)$, se verifica que:

$$\limsup_{n\to\infty} \frac{\sum_{i=1}^{n} X_i}{\sqrt{2\sigma^2 \log[\log(\sigma^2)]}} = 1$$

converge casi-con-certeza[276]

En 1924 Khinchin obtuvo una versión de la ley de los grandes números que, a diferencia de los teoremas anteriores, no establece ningún supuesto acerca de la existencia de las varianzas de las variables aleatorias involucradas: Sea $\{X_1, X_2, \ldots, X_n\}$ una sucesión de variables aleatorias iid, con función característica $C_j(w) = C_X(w)(j = 1,2, \ldots, n)$ y con valor esperado $E\left(X_j\right) = m \; (j = 1,2, \ldots, n)$. La función característica de la variable $Y_n = \frac{1}{n}\sum_i X_i = \frac{X_{(n)}}{n}$ será, entonces, de la forma:

$$C_{Y_n}(w) = E\left(e^{iwY_n}\right) = E\left[exp\left(iw\frac{X_{(n)}}{n}\right)\right] =$$

[276] Las generalizaciones del teorema de Khinchin culminan con la ley del logaritmo iterado de Kolmogorov (1929b) y Lévy (1937-1954).

$$= E\left\{exp\left[iw\frac{1}{n}\sum_{j=1}^{n}X_j\right]\right\} = \left[C_X\left(\frac{w}{n}\right)\right]^n$$

Como se supone que m existe, se puede escribir $C_X(w) = 1 + iwm + o(w)$. De modo que resulta $C_{Y_n}(w) = \left[1 + \frac{iw}{n} + o\left(\frac{w}{n}\right)\right]^n$. Por lo tanto, será:

$$\lim_{n\to\infty} C_{Y_n}(w) = \lim_{n\to\infty}\left[C_X\left(\frac{w}{n}\right)\right]^n = e^{iwm}$$

Este resultado demuestra que la sucesión de funciones características $\{C_{Y_n}(w)\}$ es convergente a e^{iwm}, la cual es la función característica correspondiente a una variable aleatoria Z que puede asumir solamente el valor m, con probabilidad $p(Z = m) = 1$. Luego, de acuerdo con el teorema de Lévy-Cramér, se puede concluir que la sucesión de funciones de distribución $F_{Y_n}(y)$ converge a la función de distribución de la variable Z o, lo que en este caso es lo mismo, que la sucesión de variables $\{Y_n\}$ converge en-probabilidad al valor m.Y, en consecuencia, queda demostrado que, dada una sucesión de variables iid, la existencia de m es condición necesaria y suficiente para el cumplimiento de la ley débil de los grandes números.

En1932a Khinchin propuso una interpretación alternativa de la condición de intercambiabilidad de de Finetti, considerándola equivalente a la condición que la probabilidad $\omega_Y^{(n)}$ sea la misma para Y éxitos (no necesariamente consecutivos). Esta condición y la demostración mediante la aplicación de la desigualdad de Chebychev que las probabilidades $\omega_Y^{(n)}$ definen los momentos de la distribución límite de la variable frecuencia relativa, le permitieron obtener una versión (más simple que la de de Finetti) del teorema de representación y de la ley fuerte de los grandes números para eventos intercambiables. En "*Remarques sur les suites d'événements obeissant à la loi des grands nombres*" (1932b) demostró que, si la frecuencia relativa $(Y_n = \frac{1}{n}\sum_i X_i)$, en n repeticiones de una sucesión de eventos, es tal que:

$$\lim_{n\to\infty} p(|Y_n - Y_{n+h}| > \varepsilon) = 0\ (h > 0, \varepsilon > 0)$$

Entonces dicha sucesión obedece a una ley fuerte de los grandes números.

25.- Andrei Nikolaevich Kolmogorov y la reivindicación de von Mises

Los "*Grundbegriffe der Wahrscheinlichkeitsrechnung*" (1933b) de Kolmogorov constituyó un hito que se transformó en el símbolo de la teoría moderna de la probabilidad[277]. Las dos propuestas fundamentales contenidas en esta obra son la teoría de las probabilidades condicionadas para conjuntos infinitos de eventos elementales y su aplicación a casos en los cualesal evento condicionante le corresponde una probabilidad nula[278](como en la teoría del movimiento Browniano) y la teoría general de los procesos estocásticos[279].

Kolmogorov comenzó su análisis definiendo el ya mencionado principio de ergodicidad como aquél que asegura que la probabilidad de que un proceso dinámico ingrese a un conjunto de estados A, es asintóticamente independiente del estado inicial, cuando t aumenta indefinidamente:

$$\lim_{t\to\infty}[p(t_0,x,t,A)-p(t_0,y,t,A)]=0$$

(donde x e y denotan los estados del proceso al momento inicial t_0), es decir, un principio que permite asegurar que las probabilidades de transición son asintóticamente iguales para todas las trayectorias. A partir de una interpretación frecuencista de estas probabilidades se puede suponer que, de acuerdo con esta propiedad, todas las sucesiones finitas de realizaciones del proceso observarán el mismo comportamiento asintótico. Lo que permite concluir que, como se expresó en la Sec. 22, la propiedad de ergodicidad conduce a la ley de los grandes números.

---

[277] La obra probabilística de Kolmogorov se inscribe en la denominada escuela de Moscú, la cual se inició con Chebychev y Markov y fue continuada, entre otros, por Bernstein, Slutsky y Kolmogorov.

[278] En este caso las probabilidades están definidas como variables aleatorias y tratadas de acuerdo con los postulados del teorema de Radon-Nikodým (ver Révész (1968)).

[279] Más allá de los trabajos de Bachelier (1900)(1912), Einstein (1905)(1906)(1907)(1908) y Wiener (1921), puede considerarse que el estudio sistemático de los procesos estocásticos continuos en el dominio del tiempo comenzó con de Finetti (1929a)(1929b) y con la teoría de los procesos Markovianos de Kolmogorov (1931)(1932a)(1932b)(1936). Un estudio previo de las cadenas de Markov puede hallarse en Hadamard (1927)(1928), Hostinsky (1928)(1929)(1930) y von Mises (1928)(1930).

En 1956, Kolmogorov analizó la vinculación entre la teoría de algoritmos, la definición de Church (1940) de aleatoriedad de una sucesión y la teoría de la probabilidad.A partir de esta relación y del principio de ergodicidad propuso un retorno a la interpretación de von Mises[280]. Su propuesta metodológica considera que la probabilidad es un coeficiente aplicable a eventos repetibles, "*...una constante $\pi = p(A/S)$ (objetivamente determinada por la conexión entreun conjunto de condiciones S y el evento A) tal que las frecuencias v se aproximan 'hablando en términos generales' a π a medida que el número de observaciones aumenta*" (1963, p. 231). Este postulado lo condujo a una doble interpretaciónBernoulliana de la ley de los grandes números: **i)** que, dada una probabilidad p de que la frecuencia relativa se aparte de la probabilidad π de ocurrencia del evento, en más de ε, permite determinar el número de observaciones en igualdad de condiciones (n) necesario para poder asegurar que p es menor que una cota determinada (α) y **ii)** que, dadas sucesiones de n términos, la frecuencia se encontrará en un entorno de π de longitud ε en una proporción $1-\alpha$.

A partir de esta propuesta planteó el siguiente conjunto de condiciones necesarias para una interpretación frecuencista de la probabilidad: **i)** que exista un conjunto$S$ de condiciones indefinidamente repetibles; **ii)** que exista un conjunto $E = \{x_1, x_2, \dots\}$ de posibles modalidades de presentación de un evento respecto del conjunto S; **iii)** que el evento Ase considerará como ocurrido, si se presenta una modalidad de E que pertenece al conjunto A y**iv)** que se puede asociar al evento A un número, $p(A)$, tal que: **a)** si las condiciones S se repiten un número (n) de veces suficientemente grande, entonces se puede asegurar que es "*prácticamente cierto*" que la frecuencia relativa de ocurrencia de A $\left(\frac{m}{n}\right)$ diferirá muy poco de $p(A)$ y **b)** si $p(A)$ es muy pequeña, se puede asegurar que es "*prácticamente cierto*" que A no ocurrirá en una realización simple delas condiciones S.

Como corolario de las condiciones **iv-a)** y **iv-b)**, Kolmogorov establece: **i)** que, dados dos eventos individualmente "*prácticamente ciertos*", el evento conjunto también es "*prácticamente cierto*", pero con un grado de certeza menor y, en consecuencia, dado un número suficientemente grande de eventos individualmente "*prácticamente ciertos*", no se puede asegurar que el evento conjunto sea "*prácticamente cierto*" y de la misma forma, **ii)** que

[280] Esta publicación se encuentra traducida al inglés en Kolmogorov (1963a).

$p(E) = 0$ no implica que E sea un evento imposible, sino que su no ocurrencia es "*prácticamente cierta*" en una repetición individual y que, en una sucesión suficientemente larga de repeticiones, la frecuencia de E será próxima a cero.

Obsérvese que si bien, en principio, las condiciones **iv-a)** y **iv-b)** son reglas de inferencia de las probabilidades a las frecuencias, es decir reglas de predicción de propiedades de eventos observables, a partir de modelos probabilísticos (reglas de inferencia para la construcción de tests de hipótesis), en lo que se podría interpretar como una tendencia hacia un inductivismo Kolmogoroviano, también pueden ser interpretadas como implicaciones en el sentido inverso como reglas de inferencia de las frecuencias a las probabilidades: tomando como punto de partida a las frecuencias, concluir que, para sucesiones de observaciones suficientemente largas, se puede considerar como "*prácticamente cierto*" que las probabilidades asumirán valores en un entorno arbitrariamente próximo a dichas frecuencias.

Todas estas condiciones ("*...respecto de las cuales no deseo profundizar ahora*" (Kolmogorov (1933, p. 3)) están vinculadas con la aplicabilidad de la ley de los grandes números. En particular, las condiciones **iv-a)** y **iv-b)** son, en sí mismas, versionesfinitas de la ley de los grandes números, basadas en el ya mencionado principio de Cournot, en las que las probabilidades no son interpretadas como límites de frecuencias relativas, sino vinculadas a través de una vaga relación con frecuencias relativas obtenidas a partir de una sucesión finita de repeticiones, asociada a la noción de "*certeza práctica*"[281]. En esta definición de "*certeza práctica*" como resultado del límite de las frecuencias, contrariamente a su posición en la generalización de 1931 de la ley fuerte de los grandes números de Borel[282], Kolmogorov revela su

[281] de Finetti (2006): "*En la previsión científica ya no es posible pensar en una certeza absoluta, existe solamente una cierta probabilidad que, como máximo, puede llegar a ser tan grande que merezca la denominación de 'certeza práctica'. Y podemos seguramente comprender por primera vez el gran valor del análisis propuesto por Hume hace mucho tiempo (demasiado adelantado a su época como para ser comprendido), sometido a la idea de causa y la gran pobreza de las tentativas de poner a cubierto de la profanación que significaba dicho análisis, un concepto que se prefería embalsamar y exponer bajo el cristal de las aprioristiquerías*" (pp. 77-78).

[282] Ver Sec. 17.

reconocimiento de la debilidad de la interpretación frecuencista de la probabilidad[283]. El teorema conocido como la "*paradoja de Borel*" demuestra que,dado un conjunto de sucesiones binarias infinitas, $X = \{X_1, X_2, \dots\} \epsilon \{0,1\}^N$, formadas por repeticiones independientes con probabilidades de ocurrencia individual del resultado "1" igual a $\frac{1}{2}$ y, por lo tanto con una probabilidad $\left(\frac{1}{2}\right)^n$ para cada sucesión de longitud n, queda definida en forma unívoca una probabilidad θ en el espacio $\{0,1\}^N$, que es el espacio de las frecuencias relativas. Este espacio puede ser interpretado como el intervalo $[0,1]$ y, en consecuencia, θ como la medida de Lebesgue de $[0,1]$ la cual, para el conjunto definido más arriba con $\frac{1}{2}$ como límite de los resultados "1" es igual a uno. Ahora bien, dado que el experimento consistente en seleccionar una sucesión del conjunto $\{0,1\}^N$ requeriría un período infinito, la interpretación frecuencista de esta medida no es posible desde el punto de vista de la probabilidad[284]. Para solucionar esta circunstancia Kolmogorov propuso adoptar una versión finita de la ley de los grandes números,la cual postula que, dados una probabilidad δ y un infinitésimo $\varepsilon > 0$, se puede asegurar que al cabo de n repeticiones, la frecuencia relativa $Y_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ cumple la condición $|Y_n - \theta| \leq \varepsilon$ con una probabilidad $1 - \delta$, a partir de la consideración del concepto de "*certeza práctica*".

Comose mencionó en la Sec. 20, los axiomas a cumplir por los colectivos empíricos de von Mises son la existencia de la ley de estabilidad de las frecuencias estadísticas y la ley de irregularidad absolutay, como corolario, que el límite de todas las sub-sucesiones coincida conel límite de la sucesión original[285]. Ahora bien, las consideraciones anteriores permiten concluir que

---

[283] En la propuesta de Kolmogorov (1963b) la frecuencia límite es una idealización teórica, un concepto de interés puramente matemático: "*La noción de probabilidad como límite de la frecuencia relativa no contribuye en nada a la aplicabilidad de los resultados de la teoría de la probabilidad a problemas prácticos reales que sólo admiten un tratamiento a partir de un número finito de repeticiones*" (p. 369).

[284]Ver Sec. 17.

[285] Un criterio similar puede hallarse en Fisher (1956), para quien la "*probabilidad estadística*" estaba basada en la noción de población homogénea (una noción de connotaciones innegablemente subjetivas).

Kolmogorov (1963b) pretendió abarcar un ámbito de aplicación de la teoría de la probabilidad más amplio que el de von Mises basándose, a partir de la propuesta de Church (1940), en una aproximación diferente a la aleatoriedad absoluta en la proyección de los atributos sobre los elementos de un colectivo finito. En el ámbito de la teoría algorítmica, propuso una "*clase de selección de algoritmos*" y el requerimiento de una invariancia aproximada de la frecuencia relativa con respecto a dicha clase[286]. Definió una serie binaria finita como aleatoria cuando posee el grado de complejidad máxima y la teoría de la probabilidad asociada de acuerdo con la teoría de la medida, complementada por la teoría de las sucesiones aleatorias que permite su aplicabilidad[287].Dada una sucesión binaria $Y = \{Y_1, Y_2, \dots, Y_n\}$ de longitud n, su complejidad respecto de un algoritmo A está definida como $K_A(Y/n) = min[\ell(p)]$, donde $\ell(p)$ representa la medida del "input" p para el cual se verifica que $A(p,n) = Y$. Como se comentó en la Sec. 20, de acuerdo con Kolmogorov (1965) se demuestra que existe por lo menos un algoritmo universal "*asintóticamente óptimo*", U, tal que $K_U(Y/n) \leq K_A(Y/n) + c$, para todo algoritmo A. En particular, se dice que una sucesión Y es aleatoria si $K_U(Y/n) \cong n$ (en otros términos, si la información que contiene no puede ser comprimida).

A partir de estas premisas y a fin de justificar los postulados de su ley finita de los grandes números, Kolmogorov (1933) propuso una "*deducción empírica de la axiomática*"[288]. Esta axiomática, contenida en los "*Grundbegriffe*", asume la noción de probabilidad como primitiva y está basada en la teoría de

---

[286] Ver Kolmogorov (1965)(1968)(1983a)(1984).

[287] La propuesta de Kolmogorov fue continuada por Martin-Löf (1966)(1969)(1971), quien demostró además que el límite de las frecuencias relativas de sucesiones aleatorias existe, lo cual hace que el axioma de estabilidad pueda ser considerado como un corolario del de aleatoriedad.

[288] Kolmogorov (1986): *"El valor conjunto de la teoría de la probabilidad depende del establecimiento de regularidades estrictas resultantes de los efectos combinados de fenómenos aleatorios masivos. La verdadera noción de matemática de la probabilidad no hubiera sido útil si no hubiera sido estructurada como la frecuencia de un cierto resultado obtenida de la experimentación repetida. Por esta razón los trabajos de Pascal y Fermat pueden ser considerados como la prehistoria y la ley de los grandes números de Bernoulli coo el comienzo de la historia de la probabilidad"* (p. i).

la medida y la teoría métrica de funciones[289] y postula que,dado un espacio muestralΩ formado por un conjunto de eventos elementales, $\{E_1, E_2, \ldots\}$:**1)** los eventos forman una σ-algebra s, es decir, una clase cerrada respecto de las operaciones de unión, intersección y negación de conjuntos numerables de eventos y del límite de sucesiones de eventos;**2)** $s \supset \Omega$; **3)** asociado a cada evento $E_i \epsilon s$, existe un número real no-negativo, $p(E_i)$, al que se denominará "*probabilidad de ocurrencia del evento*E_i"; **4)** la probabilidad de que ocurra al menos uno de los eventos incluidos en el espacio muestral es igual a 1, $p(\Omega) = 1$ y **5)** siE_i y $E_j$son eventos incompatibles (es decir, tales que $E_i \cap E_j = \emptyset$, $i \neq j$), entonces será $p\left(E_i \cup E_j\right) = p(E_i) + p\left(E_j\right)$ (expresión conocida como "*axioma de aditividad*")[290].

Desde un punto de vista sintáctico, la diferencia entre este sistema axiomático y el de de Finetti radica fundamentalmente en que, en tanto éste identifica a la probabilidad con una medida en un espacio de conjuntos que proporciona una referencia a una teoría de las variables aleatorias, sus distribuciones de probabilidades y sus expectativas, la axiomática deFinettiana proporciona una teoría unificada de la probabilidad y la expectativa (denominada, como se vio en la Sec. 22, como "*previsión*") caracterizada como un operador lineal sobre funciones de cantidades[291].

Se puede concluir fácilmente que su interpretación de las leyes de probabilidad como aproximaciones a distribuciones empíricas a través de la noción de certeza práctica constituye un intento insuficiente de obtener su pretendida "*deducción empírica de los axiomas*" basados en la teoría de la medida. En particular: **i)** la asimilación de las frecuencias relativas a las

---

[289]Kolmogorov (1933): *"la teoría de la probabilidad, como una disciplina matemática, puede y debería ser desarrollada a partir de axiomas, exactamente de la misma forma que la geometría ty el álgebra. Esto significa que, después que hayamos definido los elementos a ser estudiados y sus relaciones básicas, y de haber establecido los axiomas que los gobiernan, toda la exposición posterior debe estar basada exclusivamente en dichos axiomas, independientes del significado ususal concretode los elementos y sus relaciones"* (p. 1).

[290] La primera versión de su axiomática data de 1929. Otros antecedentes en los intentos de axiomatización de la probabilidad que merecen ser mencionados son: Borel (1909a), Bohlmann (1909), Bernshtein (1917). En la axiomática de Bernshtein el conjunto de eventos es considerado como un álgebra de Boole y se basa en la comparación cualitativa de eventos aleatorios de acuerdo con la medida de sus probabilidades.

[291]Para un análisis detallado de las axiomatizaciones, ver Roper; Leblanc (1999).

probabilidades no está suficientemente justificada; **ii)** dado que, obviamente, las frecuencias relativas obtenidas a partir de sucesiones finitas de observaciones siempre asumen números racionales, la aproximación de Kolmogorov genera una injustificable restricción al conjunto de valores a asumir por las probabilidades; **iii)**una justificación completa del sistema axiomático debería considerar una demostración de por qué las probabilidades deben asumir valores reales únicos[292].

Cabe recordar que, en un período previo a la reivindicación de la teoría frecuencista de von Mises, Kolmogorov (1928)(1933b) había establecido las condiciones matemáticas necesarias y suficientes para poder asegurar la validez de la ley de los grandes números para sucesiones de variables aleatorias independientes. Demostró que, dadas una sucesión $\{X_1, X_2, \dots, X_n\}$ de variables aleatorias, con valor esperado nulo y la variable:

$$X_i^* = \begin{cases} X_i & si \quad |X_i| \leq b_i \\ 0 & si \quad |X_i| > b_i \end{cases}$$

la convergencia de las series: **i)** $\sum_{i=1}^{\infty} p(|X_i| > b_i) < \infty$; **ii)** $\frac{1}{b_n} \sum_{i=1}^{\infty} \sigma^2(X_n^*) < \infty$ y **iii)** $\sum_{i=1}^{\infty} E(X_i^*) < \infty$ (donde $\{b_1, b_2, \dots, b_n\}$ denota una sucesión de constantes positivas tales que $\lim_{n \to \infty} b_n = \infty$), es condición necesaria y suficiente para poder asegurar que la variable:

$$Y_n = \frac{1}{b_n} \sum_{i=1}^{n} X_i \xrightarrow{ccc} 0$$

(este resultado es conocido como el "*teorema de las tres series*" o el "*criterio de Kolmogorov-Khinchin*")[293]. Brown (1971) generalizó este teorema de

[292] Una solución a esta última restricción fue proporcionada por de Finetti (1931), quien demostró que la condición necesaria y suficiente para que un sistema de probabilidades sea coherente, es que posea la propiedad de unicidad. Debe tenerse en cuenta que, si bien tanto Kolmogorov como Khinchin conocían la obra de de Finetti, nunca hicieron ninguna referencia a la interpretación subjetivista de la probabilidad.

[293] Los trabajos fundamentales de Kolmogorov sobre las condiciones que permiten asegurar la validez de las leyes débil y fuerte de los grandes números comprenden el período 1927-1929. A fines de 1927 ya había completado sus resultados sobre las condiciones necesarias y suficientes para la validez de la ley débil de los grandes números debidas a J.

Kolmogorov eliminando la condición de independencia de las variables X_iy reemplazando las condiciones **i)**, **ii)** y **iii)** por la condición que las tres series converjan casi-con-certeza[294].

A partir de la definición de función característica puede obtenerse la siguiente expresión alternativa del teorema de las tres series: Sean: **i)** una sucesión $\{X_1, X_2, \dots, X_n\}$ de variables aleatorias independientes; **ii)** la sucesión $\{C_j(w)\}(j = 1,2, \dots, n)$ de las funciones características de las variables X_j y **iii)** una sucesión $\{b_1, b_2, \dots, b_n\}$ de constantes reales tales que la sucesión $\sum_{j=1}^{\infty}(X_j - b_j)$ converge casi-con-certeza. Entonces $\prod_{j=1}^{\infty}|C_j(w)|$ converge uniformemente para todo intervalo finito de w y viceversa, si $\prod_{j=1}^{\infty}|C_j(w)|$ converge en un intervalo de w de medida de Lebesgue positiva, entonces se puede asegurar que existe una sucesión $\{b_1, b_2, \dots, b_n\}$ de constantes reales tales que la sucesión $\sum_{j=1}^{\infty}(X_j - b_j)$ converge casi-con-certeza.

Bajo estos mismos supuestos y como una extensión del ya mencionado resultado de Khinchin (1924), Kolmogorov (1930) demostró que las condiciones **i)** $\lim_{n\to\infty} \sigma_n^2 = \lim_{n\to\infty} \sigma^2(\sum_{i=1}^{n} X_i) = \infty$ y **ii)** $b_n = o\left(\frac{\sigma_n}{\sqrt{\ln[\ln(\sigma_n^2)]}}\right)$ son necesarias y suficientes para el cumplimiento de una versión de la ley del logaritmo iterado de la forma[295]:

$$\lim_{n\to\infty} sup\left\{\frac{\sum_{i=1}^{n} X_i}{\sqrt{2\sigma_n^2 \ln[\ln(\sigma_n^2)]}}\right\} = 1$$

Bernoulli y continuadas por Chebychev y Markov. El teorema de las tres series es el resultado más importante en el ámbito de la convergencia de sumas parciales y de sumas parciales de cuadrados de variable aleatorias independientes.

[294] Esta generalización significó un avance importante si se tiene en cuenta la tesis generalmente aceptada que *"...la dependencia es un fenómeno endémico"* (Heyde (1985)).

[295] Debe tenerse en cuenta que, si las variables X_i son uniformemente acotadas, la primera condición implica la segunda.

En 1931 demostró que, dada una sucesión de variables $Y_n = \sum_{i=1}^{n} X_i - k_n$ (donde $\{k_1, k_2, \dots\}$ denota una sucesión de constantes) es posible hallar un valor N tal que $p(\max_{n\geq N}|Y_n| \leq \varepsilon) > 1 - \vartheta$.Es decir, es posible hallar un valor N tal que, para por lo menos un $n \geq N$, la probabilidad de ocurrencia del evento $|Y_n| < \varepsilon$ sea mayor que $1 - \vartheta$. En otros términosque ,dadauna sucesión de variables $\{X_i\}$,se verifica que $p(\lim_{n\to\infty} Y_n = 0) = 1$; es decir, que $X_i \overset{ccc}{\longrightarrow} 0$. O, lo que es lo mismo,que la sucesión $\{X_i\}$ cumple con la ley fuerte de los grandes números[296].

Posteriormente, en1933 demostró que, dada una sucesión de variables aleatorias independientes $\{X_1, X_2, \dots, X_n\}$, con medidas $E(X_i) = m_i < \infty$ y $\sigma^2(X_i) = \sigma_i^2 < \infty$ $(i = 1,2, \dots, n)$, si se reemplaza la condición de Markov, $\lim_{n\to\infty} \frac{1}{n^2} \sum_j \sigma_j^2 = \lim_{n\to\infty} \sigma^2\,(Y_n) = 0$,por la más restringidacondición de Kolmogorov $\sum_{j=1}^{n} \frac{1}{j^2} \sigma_j^2 < \infty$[297]y, basándose en la desigualdad[298]:

$$p\left\{\max_{i\leq j\leq n} |Y_n - E(Y_n)| \leq \varepsilon\right\} > 1 - \frac{1}{\varepsilon^2}\sigma^2(Y_n)$$

se obtiene la ley de los grandes números de Kolmogorov:

$$p\left\{\lim_{n\to\infty} \frac{1}{n}\left[\sum_{i=1}^{n}(X_i - m_i)\right] = 0\right\} = 1$$

---

296 Este teorema admite dos demostraciones alternativas, una debida a Doob (1953), en el ámbito de la teoría de la martingala (ver Sec. 26) y una segunda basada en el concepto de estacionariedad. La generalización de este teorema es la conocida como ley fuerte de los grandes números de Marcinkiewicz (ver Marcinkiewicz; Zygmund (1937)).

297 Esta condición sobre las sumas parciales Y_n, implica que la varianza $\sigma^2(Y_n) = \sigma^2(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \sigma_i^2$ debe ser, a lo sumo, de orden n ($\sigma^2(Y_n) = O(n)$). Y, como $\sigma^2(Y_n) = O(n) \Rightarrow \sigma^2(Y_n) = o(n^2)$, se concluye inmediatamente que esta condición de Kolmogorov implica la condición de Markov.

298 Esta desigualdad -conocida como "*desigualdad de Kolmogorov*"- es una generalización de la desigualdad de Chebychev. Bernstein (1937) extendió sus alcances como una aplicación de la desigualdad de una submartingala (ver Sec. 26).

Es decir, demostró que, dada una sucesión de variables aleatorias independientes, la condición de que sus varianzas sean uniformemente acotadas es suficiente para poder asegurar que cumple con la ley fuerte de los grandes números. Asimismo, demostró que, dada una sucesión de variables aleatorias independientes, las condiciones de ser incorrelacionadas de cuarto orden y con momentos de cuarto orden finitos y uniformemente acotados, también son suficentes para poder asegurar que cumple con la ley fuerte de los grandes números.

Esta ley fuerte de los grandes números de Kolmogorov para variables iid admite una interpretación en función del teorema ergódico de Birkhoff-Khinchin según el cual, dada una sucesión $\{X_t\}$ estacionaria y ergódica, entonces $\frac{1}{n}\sum_{t=1}^{n} X_t$ converge en-probabilidad y casi-con–certeza a $E(X)$.

Finalmente estableció que, dada una sucesión $\{X_1, X_2, \dots, X_n\}$ de variables aleatorias iid, con función de distribución $F_X(x)$, la existencia de $E(X)$ es condición necesaria y suficiente para poder asegurar que:

$$p\left[\lim_{n\to\infty}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - E(X)\right) = 0\right] = 1$$

Lo que permite concluir que las condiciones necesarias y suficientes para poder asegurar la validez de la ley fuerte de los grandes números, para variables aleatorias idénticamente distribuidasson las mismas que las requeridas para el cumplimiento de la ley débil de los grandes números[299].

En forma similar Hsu; Robbins (1947), a partir del concepto de "*convergencia completa*" [300], demostraron que, dada una sucesión de variables aleatorias centradas iid, $\{X_n\}$, con función de distribución $F_X(x)$, la

[299] Khinchin (1932c) demostró que la paradoja probabilística del teorema ergódico de Birkhoff para probabilidades estacionarias, proporciona condiciones necesarias y suficientes más generales.

[300] De acuerdo con Hsu; Robins, se dice que una sucesión $\{X_i\}$ converge completamente a una constante c si se verifica que $\sum_{n=1}^{\infty} p(|X_j - c| > \varepsilon) < \infty$ para todo $\varepsilon > 0$ (Dugué (1954)(1957) denominó a esta modalidad como *"convergencia-casi-completa")*.

condición $\sigma^2(X_n) < \infty$ es suficiente para poder asegurar que la sucesión $\left\{Y_n = \frac{1}{n}\sum_{j=1}^{n} X_j\right\}$ es tal que:

$$\lim_{n\to\infty}\left[\sum_{n=1}^{\infty} p(|Y_n| > \varepsilon)\right] = 0$$

para todo $\varepsilon > 0$. Es decir que la sucesión $\{X_n\}$ converge completamente a cero[301].

En 1937 Kolmogorov demostró que el caso de un proceso estocástico continuo en el dominio del tiempo era reducible al caso discreto y que, bajo condiciones más réstringidas se verifica que $\frac{1}{t}\int_0^t X(T^s w)ds$ converge en-probabilidad y casi-con-certeza a $E\big(X(w)\big)$ (donde T^s denota un semi-grupo de transformaciones ergódicas y $E(|X(w)|) < \infty)$[302].

A su vez, para un proceso $\{X(t)\}(t \geq 0)$ estacionario y ergódico, continuo en el dominio del tiempo, Khinchin (1938) demostró que $\frac{1}{t}\int_0^t X(u)du$ converge en-probabilidad y casi-con-certeza a $E\big(X(t)\big)$.

En 1929 Kolmogorov obtuvo su versión de la ley del logaritmo iterado según la cual, dada una sucesión $\{X_1, X_2, \dots\}$ de variables aleatorias independientes tales que $E(X_n) = 0$ y $|X_n| \leq o\left(\sqrt{M_n}\right)$ (para $n \geq 1$), donde:

$$M_n = \frac{\sum_{j=1}^{n} \sigma^2(X_j)}{\ln\left[\ln\left(\sum_{j=1}^{n} \sigma^2(X_j)\right)\right]}$$

se verifica que:

[301] Erdös (1949) demostró la implicación inversa.

[302] Como se verá en la Sec. 26, Doob (1949) obtuvo una demostración de la ley fuerte de los grandes números de Kolmogorov en la que $\frac{1}{n}\sum_{t=1}^{n} X_t$ representa una martingala inversa. Otras demostraciones se deben a Etemandi (1981) y Grimmet;Stirzaker (1982).

$$\lim_{n\to\infty} sup \frac{\sum_{j=1}^{n} X_j}{\sqrt{2\sum_{j=1}^{n} \sigma^2(X_j) \ln[\ln(\sum_{j=2}^{n} \sigma^2(X_j))]}} = 1$$

Como se mencionó en la Sec. 24, Hartmann;Winter (1941) generalizaron la ley del logaritmo iterado para toda sucesión $\{X_1, X_2, ...\}$ de variables iid, suponiendo solamente que $E(X_i) = 0$ y $E(X_i^2) < \infty$, obteniendo el siguiente resultado:

$$\lim_{n\to\infty} sup \frac{\sum_{j=1}^{n} X_j}{\sqrt{2E(X_1^2) \log[\log(n)]}} = 1$$

Por su parte, Prokhorov (1956) estableció que, dada una sucesión $\{X_1, X_2, ..., X_n\}$ de variables aleatorias independientes acotadas tales que $|X_i| = o\left(\frac{i}{\ln \ln(i)}\right)$, la existencia de un $\varepsilon > 0$ arbitrario tal que:

$$\sum_{j=1}^{\infty} exp\left[-\frac{\varepsilon}{\frac{1}{2^{2r}}\sum_{i=2^r+1}^{2^{r+1}} \sigma^2(X_i)}\right] < \infty$$

es condición necesaria y suficiente para poder asegurar la validez de la ley fuerte de los grandes números.

Strassen (1964)(1965)(1966), a partir de su demostración de los principios de invariancia casi-con-certeza, obtuvo una versión de la ley del logaritmo iterado funcionalpara toda sucesión $\{X_1, X_2, ...\}$ de variables iid centradas, con momentos de segundo orden finitos, de la forma[303]:

$$\lim_{n\to\infty} sup \frac{\sum_{j=1}^{n} X_j}{\sqrt{2n \ln[\ln(n)]}} < \infty$$

y la expresión inversa de la ley de Hartman-Winter del logaritmo iterado según la cual, dada una sucesión $\{X_i\}(i \geq 1)$ de variables iid tales que:

[303] Strassen (1964) también obtuvo una versión funcional de la ley del logaritmo iterado.

$$p\left[\limsup_{n\to\infty}\frac{|\sum X_i|}{\sqrt{2n\log[\log(n)]}}<\infty\right]>0$$

se verifica que $E(X_i)=0$ y $E(X_i^2)<\infty$ (ver Feller (1968), Heyde (1968a)(1968b), Steiger; Zaremba (1972))[304].

26.- Joseph Leo Doob y la teoría de la martingala

La teoría de la martingala, introducida por Doob (1953), proporcionó una alternativa metodológica al estudio de los teoremas límite y,en particular, representó una aproximación novedosa a la ley de los grandes números.

La martingala "*con respecto al conjunto de información* $\{Y_{t-1},Y_{t-2},\dots\}$" o "*adaptada a la sucesión* $\{Y_{t-1},Y_{t-2},\dots\}$" o "*con respecto a su filtración natural*" es una clase de proceso estocástico que se caracteriza por cumplir las siguientes condiciones: **i)** $E(|Y_t|)<\infty(t=1,2,\dots)$ y **ii)** $E(Y_t/Y_{t-1},Y_{t-2},\dots)=Y_{t-1}$[305]. Esta última condición de que la mejor predicción para el período t es, simplemente, el estado del proceso en el momento $t-1\ (t=2,3,\dots)$ (conocida como la "*propiedad de la martingala*") obviamente también puede ser expresada de la siguiente forma[306]:

$$E(Y_t/Y_{t-1},Y_{t-2},\dots)=E(\triangle Y_t/Y_{t-1},Y_{t-2},\dots)=E(\varepsilon_t/Y_{t-1},Y_{t-2},\dots)$$

---

[304] Basu (1973) extendió estos resultados a las martingalas (ver Sec. 26).

[305] El proceso denominado martingala (introducido en la literatura probabilística por Ville (1939), pero apreciado, en cuanto a su utilidad en las aplicaciones, recién después de la publicación de los trabajos de Doob) es asimilable a un proceso que representa la fortuna de un jugador en un estado dado del juego y que, en virtud del criterio del juego equitativo, el valor esperado de dicha fortuna al cabo de una nueva partida es igual al monto presente de su fortuna y no está afectado por su historia pasada. La denominación "martingala" se relaciona con la ciudad de Martigues (Provence) cuyos habitantes eran famosos por ejecutar un sistema de juego que consistía en duplicar la apuesta precedente cada vez que el jugador perdía, retornando a la apuesta inicial después de cada triunfo (ver Weaver (1982), Gingerenzer; Swijtink; Porter; Daston; Beatty; Kruger (1989)).

[306] Contrariamente a lo que ocurre con el RW y los procesos con incrementos independientes, la martingala admite una cota superior en los momentos de primer orden.

Si se verifica que $E(|Y_t|^s) < \infty$ (para algún $s > 1$), se dice que el proceso $\{Y_t\}$ es una "*martingala de orden s*". En particular, para $s = 2$, queda definida una "*martingala de segundo orden*" (o "*martingala de cuadrado integrable*")[307].

Al proceso $\{\Delta\, Y_t\} = \{Y_t - Y_{t-1}\} = \{\varepsilon_t\}$ se lo conoce como "*diferencia de una matingala con respecto al conjunto de información* $\{Y_{t-1}, Y_{t-2}, \ldots\}$". A partir de esta definiciónse puede concluir, entonces, que el error medio cuadrático mínimo de la predicción de un incremento futuro de una martingala es nulo.

Sea $\{Y_t\}$ una martingala con respecto a su filtración natural, entonces:

$$E(\Delta\, Y_t/Y_{t-1}\,, Y_{t-2}, \ldots) = E(Y_t/Y_{t-1}\,, Y_{t-2}, \ldots) - Y_{t-1} = 0$$

Como corolario de esta propiedad se obtiene inmediatamente que:

$$E(Y_t/Y_{t-1}\,, Y_{t-2}, \ldots) = E[E(Y_{t+1}/Y_t\,, Y_{t-1}, \ldots)/Y_{t-1}\,, Y_{t-2}] =$$

$$= E(Y_{t+1}/Y_{t-1}\,, Y_{t-2}, \ldots) =$$

$$= E[E(Y_{t+2}/Y_{t+1}\,, Y_t, \ldots)/Y_{t-1}\,, Y_{t-2}] =$$

$$= E(Y_{t+2}/Y_{t-1}\,, Y_{t-2}, \ldots) = \cdots = E(Y_{t+k}/Y_{t-1}\,, Y_{t-2}, \ldots)$$

Lo que permite concluir quela invariancia de la esperanza matemática es condición necesaria pero no suficiente para poder asegurar que el proceso $\{Y_t\}$ es una martingala.

En términos generales, se dice que el proceso $\{Y_t\}$ es una martingala "*con respecto a la filtración* $\Omega_t(Y)$"(o "*adaptada a la sucesión* $\{\Omega_t(Y)\}$"$(t \geq 1)$ o, simplemente, "*martingala-* Ω_t ") si se cumplen las siguientes condiciones: **i)** $E(|Y_t|) < \infty$ y **ii)** $E\big(Y_t/\Omega_{t-1}(Y)\,, \Omega_{t-2}(Y), \ldots, \Omega_0(Y)\big) = Y_{t-1}$. Donde $\Omega_t(Y) = \big\{\{Y_t\}, \{Z_t\}, \{W_t\}, \ldots\big\}\ (t = 1,2, \ldots)$ denota una

[307] La diferencia de una martingala con momentos de segundo orden finitos define una sub-clase de sucesiones ortogonales.

sucesión creciente, es decir, tal que $\Omega_1(Y) \subset \Omega_2(Y) \subset \Omega_3(Y) \subset \cdots$ [308].

De la misma forma, se verifica que:

$$E\big(Y_{t+s}/\Omega_0(Y), \Omega_1(Y), \dots, \Omega_t(Y)\big) = Y_t (s = 0,1,2,\dots)$$

Lo cual implica que la información disponible al momento$t$ coincide con el estado del proceso al momento$t$. Como corolario de esta condición se puede concluir en forma inmediata que todo incremento, $Y_{t+s} - Y_t$, es independiente de $\Omega_t(Y)$, es decir que el conjunto de información $\Omega_t(Y)$ no sirve para predecir el valor Y_{t+s}.

Se dice que el proceso $\{Y_t\}$ es una "*martingala inversa*" con respecto alafiltración $\Omega_t(Y)$ si se verifica que: **i)** $E(|Y_t|) < \infty$ y **ii)** $E\big(Y_t/\Omega_{t+1}(Y), \Omega_{t+2}(Y), \dots, \Omega_{t+n}(Y)\big) = Y_{t+n}$, donde la sucesión $\{\Omega_t(Y)\}$ es decreciente (es decir, tal que $\Omega_{t+1}(Y) \supset \Omega_{t+2}(Y) \supset \Omega_{t+3}(Y) \supset \cdots$). De esta definición se puede concluir que la condición necesaria y suficiente para que el proceso $\{Y_t\}$ sea una martingala inversa con respecto ala filtración $\Omega_t(Y)$es que $\{Y_{t-i+1}\}$ $(i = 1,2,\dots,n)$ sea una martingala. Esto demuestra que la teoría relacionada con las martingalas inversas es simétrica con respecto a la teoría de las martingalas finitas. Debe tenerse en cuenta que esta simetría no es generalizable en forma automática a la aplicación de los teoremas en el límite.

Sea $\{\Delta\, Y_t\}$ un proceso diferencia de una martingala con respecto a su filtración natural, tal que $E(\varepsilon_t^2) < \infty$ (es decir, con varianza finita) y sea un proceso $\{C_t\}(t = 1,2,\dots)$ predecible con respecto a $\{Y_{t-1}, Y_{t-2}, \dots\}$ y tal que $E(C_t^2) < \infty$, entonces, de acuerdo con la desigualdad de Cauchy-Schwarz, se verifica que el proceso $\{Z_t^*\} = \{[C \cdot\Delta\, Y]_t\} = \{[C \cdot \varepsilon]_t\} = \sum_{\tau=1}^{t} C_\tau \varepsilon_\tau$ es tal que:

$$E(|Z_t^*|) \leq \sum_{\tau=1}^{t} E(|C_\tau \varepsilon_\tau|) \leq \sum_{\tau=1}^{t} \sqrt{E(C_\tau^2)E(\varepsilon_\tau^2)} < \infty$$

---

[308] Debe tenerse en cuenta que $\Omega_t(Y)$ constituye la estructura causal de $\{Y_t\}$ y que, por lo tanto, el proceso $\{Y_t\}$ puede ser considerado como una función de $\{\Omega_t(Y)\}$.

Si, además, se tiene en cuenta que $\{Z_t^*\}$ y $\{C_t\}$ son procesos predecibles con respecto al conjunto de información $\{\triangle Y_t, \triangle Y_{t-1}, \ldots\} = \{\varepsilon_t, \varepsilon_{t-1}, \ldots\}$, se puede concluir que:

$$E(Z_t^* - Z_{t-1}^* / \triangle Y_{t-1}, \triangle Y_{t-2}, \ldots) = E(\triangle Z_t^* / \triangle Y_{t-1}, Y_{t-2}, \ldots) =$$

$$= E(C_t \triangle Y_t / \triangle Y_{t-1}, \triangle Y_{t-2}, \ldots) = C_t E(\triangle Y_t / \triangle Y_{t-1}, \triangle Y_{t-2}, \ldots)$$

Es decir, que el proceso transformado $\{\triangle Z_t^*\}$ es una diferencia de una martingala y $\{Z_t\}$ es una martingala con respecto a la filtración $\{Y_{t-1}, Y_{t-2}, \ldots\}$[309].

A partir de los resultados anteriores se pueden obtener las siguientes conclusiones:

1) Dado un proceso diferencia de una martingala con respecto a su filtración natural, $\{\triangle Y_t\} = \{\varepsilon_t\}$ tal que: **i)** $\varepsilon_{(n)} = \frac{1}{n}\sum \varepsilon_t$; **ii)** $E(\varepsilon_t^2) = \sigma^2(\varepsilon_t)$; **iii)** $\lim_{n\to\infty} \sum_{t=1}^{n} \sigma^2(\varepsilon_t) = \sigma_\varepsilon^2$; **iv)** $E(|\varepsilon_t|^s) < \infty$ $(s = 3,4, \ldots ; t = 1,2, \ldots)$y **v)** $\frac{1}{n}\sum_{t=1}^{n} \varepsilon_t^2 \overset{p}{\to} \sigma_\varepsilon^2$, a partir de una extensión del teorema central del límite debida a White (1984), se demuestra que $\sqrt{n}\varepsilon_{(n)} \overset{d}{\to} N(0, \sigma_\varepsilon^2)$.

2) Dada una martingala de segundo orden, de acuerdo con la condición de Kolmogorov[310], es posible hallar una sucesión $\{c_t\}$ de constantes positivas $(c_t \uparrow \infty)$ tal que $\sum_{t=1}^{\infty} \frac{1}{c_t^2} E(\varepsilon_t^2) < \infty$. Lo cual implica que $\{\varepsilon_t\}$ se comporta como una sucesión de variables incorrelacionadas y, en consecuencia, que se verifica el teorema de las tres series[311], según el cual $\frac{1}{c_n}\sum_{t=1}^{n} \varepsilon_t \overset{ccc}{\longrightarrow} 0$. Es decir que cumple con la ley fuerte de los grandes números:

[309] La transformación C_t para martingalas continuas en el dominio del tiempo es la integral estocástica de Itô.

[310] Ver Sec. 25.

[311] Ver Sec. 25.

$p\left[\lim_{n\to\infty}\left(\frac{1}{n}\sum_{t=1}^{n}\varepsilon_t\right)=0\right]=1$[312].

3) Sea una martingala de primer orden tal que su diferencia $\{\Delta Y_t\}=\{\varepsilon_t\}$ sea de varianza finita. De acuerdo con la desigualdad de Markov[313] para todo $\zeta>0$, se verifica que $p(|\sum_{t=1}^{n}\varepsilon_t|\geq\zeta)\leq\frac{1}{\zeta^2}E[(\sum_{t=1}^{n}\varepsilon_t)^2]$. Luego, teniendo en cuenta que:

$$E(\varepsilon_r\varepsilon_t)=E\{\varepsilon_t E[\varepsilon_\tau/\Omega_{t-1}(Y),\Omega_{t-2}(Y),\ldots]\}=$$

$$=E\left\{\varepsilon_\tau\left(E\left[\sum_{j=1}^{t}\varepsilon_j/\Omega_{t-1}(Y),\Omega_{t-2}(Y),\ldots\right]-\sum_{j=1}^{t-1}\varepsilon_j\right)\right\}=0\quad(t>0)$$

resulta que:

$$E\left[\left(\sum_{t=1}^{n}\varepsilon_t\right)^2\right]=\sum_{t=1}^{n}E(\varepsilon_t^2)+2\sum_{t>\tau}\sum E(\varepsilon_\tau\varepsilon_t)=\sum_{t=1}^{n}E(\varepsilon_t^2)$$

Se puede concluir, entonces, que $p(|\sum_{t=1}^{n}\varepsilon_t|\geq n\varepsilon)\leq\frac{1}{(n\varepsilon)^2}\sum_{t=1}^{n}E(\varepsilon_t^2)$. Lo cual implica que$\{\varepsilon_t\}$satisface la ley débil de los grandes números, $\frac{1}{n}\sum_{t=1}^{n}\varepsilon_t\overset{p}{\to}0$.

4) Todo proceso diferencia de una martingala estrictamente estacionario cumple con la ley fuerte de los grandes números:

$p\left\{\lim_{n\to\infty}\left(\frac{1}{n}\sum_{t=1}^{n}\left[Y_t-E\left(Y_t/\Omega_{t-1}(Y),\Omega_{t-2}(Y),\ldots,\Omega_1(Y)\right)\right]\right)=0\right\}=1.$

5) Sea una martingala de primer orden tal que su diferencia no sea estacionaria de segundo orden. Las condiciones necesarias y suficientes para que satisfaga la ley débil de los grandes números son: **i)** que $\lim_{n\to\infty}\sum_{t=1}^{n}p(|\varepsilon_t|\geq n)=0$; **ii)** que $\lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}E\left(\varepsilon_t^{(n)}\right)=$

312 Ver Bauer (1981).

313 Ver Sec. 16.

0 (donde $\varepsilon_t^{(n)} = \varepsilon_t$ ó 0, según que $|\varepsilon_t| < n$ o que $|\varepsilon_t| \geq n$) y **iii)** que $\lim_{n\to\infty} \frac{1}{n^2} \sum_{t=1}^{n} \left\{ E\left[\left(\varepsilon_t^{(n)} \right)^2 \right] - \left[E\left(\varepsilon_t^{(n)} \right) \right]^2 \right\} = 0$ (ver Loève, M. (1977)).

6) Todo proceso diferencia de una martingala tal que: **i)** $E(\varepsilon_t^2) < \infty (t = 1,2, \dots)$; **ii)** satisface la condición de Lindeberg (1922)[314]:

$$\lim_{n\to\infty} \left\{ \frac{1}{\sigma_{(n)}^2} \sum_{t=1}^{n} E\left[\varepsilon_t^2 I_{(|\varepsilon| > \varepsilon\sigma_{(n)})} \right] \right\} = 0$$

(donde: $\varepsilon > 0$ y$\sigma_{(n)}^2 = \sigma^2(\sum_{t=1}^{n} \varepsilon_t)$) y **iii)**

$$E\{E[\varepsilon_t^2 / \Omega_{t-1}(Y), \Omega_{t-2}(Y), \dots]\} = \sigma_t^2 (t = 1,2, \dots)$$

se puede asegurar que satisface los postulados del teorema central del límite[315]:

$$\lim_{n\to\infty} p\left[\frac{1}{\sigma_{(n)}} \sum_{t=1}^{n} \varepsilon_t \leq u \right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} e^{y^2/2} dy = F_N(u)$$

---

[314] Ver Nelson (1987).

[315] Los primeros resultados relacionados con esta generalización del teorema central del límite (debidos a Bernstein (1926)(1939) y Lévy (1935)(1937)) suponían la restricción de que la variable condicionada $[\varepsilon_t^2 / \Omega_{t-1}(Y), \Omega_{t-2}(Y, \dots)]$ fuera tal que $\sum_{\tau=1}^{t} E[\varepsilon_\tau^2 / \Omega_{t-1}(Y), \Omega_{t-2}(Y), \dots] = cte.$ Doob (1953) desarrolló una demostración del resultado de Lévy (1937), utilizando el concepto de función característica. Billingsley (1961) e Ibragimov (1963) obtuvieron, en forma independiente, una demostración del teorema central del límite para martingalas con diferencias estacionarias y ergódicas (es decir, con diferencias con varianza condicionada asintóticamente constante, $\frac{1}{\sigma_n^2} \sum_{\tau=1}^{t} E[\varepsilon_\tau^2 / \Omega_{t-1}(Y), \Omega_{t-2}(Y), \dots] = \frac{V_n^2}{E(V_n^2)} = \frac{V_n^2}{E(Y_n^2)} \xrightarrow{p} 1$). Posteriores gneralizaciones fueron obtenidas por Rosén (1967a)(1967b), Dvoretzky (1969)(1971)(1972), Loynes (1969) (1970), Bergström (1970), Brown (1971), Gänsler; Strobel; Stute (1978).

7) Sea una martingala de cuadrado integrable tal que su diferencia sea un proceso estacionario de segundo orden, ergódico. Entonces, se verifica que:

$$\frac{1}{n}\sum\nolimits_{t=1}^{n} \varepsilon_t = \sigma_\varepsilon \zeta(n)\sqrt{\frac{2}{n} ln[ln(n)]}$$

(donde $\text{limsup}_{n\to\infty} \zeta(n) = 1$ y $\text{liminf}_{n\to\infty} \zeta(n) = -1$). Este resultado (debido a Stout, W.F. (1970)(1974)) contiene como caso particular a la ley clásica del logaritmo iterado para sumas de variables aleatorias iid (debida a Hartman; Winter (1941))[316].

8) Sea $\{\varepsilon_t\}$ un proceso diferencia de una martingala con respecto a la filtración Ω_t y sea una transformación definida por una sucesión $\{v_t\}(t \geq 1)$. El proceso $Z_n = \sum_{t=1}^{n} v_t \varepsilon_t$ define una martingala transformada. De acuerdo con Burkholder (1966), dada un diferencia de una martingala, $\{\varepsilon_t\}$, tal que $sup[E(\sum_{t=1}^{n} \varepsilon_t)] < \infty$ y una transformación tal que $sup(|v_t|) < \infty$ casi-con-certeza, la martingala transformada $Z_n = \sum_{t=1}^{n} v_t \varepsilon_t$ converge casi-con-certeza.

Si se verifica que $E(Y_t - Y_{t-1}/Y_{t-1}, Y_{t-2}, \ldots) > 0$, es decir, si $E(Y_t/Y_{t-1}, Y_{t-2}, \ldots) > Y_{t-1}$, entonces, se dice que el proceso $\{Y_t\}$ es una "*submartingala*" y, si $E(Y_t/Y_{t-1}, Y_{t-2}, \ldots) < Y_{t-1}$, entonces se dice que el proceso $\{Y_t\}$ es una "*supermantingala*".En términos más generales, se dice que el proceso $\{Y_t\}$ es una "*supermartingala con respecto al conjunto de información* $\{\Omega_t(Y)\}$ ", si se verifica que: **i)** $E\left(Y_t^{(-)}\right) < \infty \left(Y_t^{(-)} = min(Y_t, 0)\right)$ y **ii)** $E\big(Y_{t+1}/\Omega_{t-1}(Y), \Omega_{t-2}(Y), \ldots, \Omega_1(Y)\big) \leq Y_t$ .De la misma forma, si se verifica que: **i)** $E\left(Y_t^{(+)}\right) < \infty \left(Y_t^{(+)} = max(Y_t, 0)\right)$ y **ii)** $E\big(Y_{t+1}/\Omega_{t-1}(Y), \Omega_{t-2}(Y), \ldots, \Omega_1(Y)\big) \geq Y_t$, se dice que $\{Y_t\}$ es una "*submartingala con respecto ala filtración*$\{\Omega_t(Y)\}$"[317].

[316]Ver Sec. 25.

[317] Por analogía con lo expresado en la nota 305 con respecto a la martingala, la submartingala y la supermartingala son asimilables a juegos no equitativos, *"favorables"* o *"desfavorables"*.

Como corolario de estas definiciones se puede concluir que: **i)** la condición necesaria y suficiente para que $\{Y_t\}$ sea una supermartingala con respecto a la filtración $\{\Omega_t(Y)\}$ es que $\{-Y_t\}$ sea una submartingala con respecto a la filtración $\{\Omega_t(Y)\}$; **ii)** la condición necesaria y suficiente para que $\{Y_t\}$ sea una martingala con respecto a la filtación $\{\Omega_t(Y)\}$ es que $\{Y_t\}$ sea, simultáneamente, una submartingala y una supermartingala con respecto a la filtración $\{\Omega_t(Y)\}$.

A partir de estas definiciones, se obtienen las siguientes propiedades de las submartingalas:

1) Sean: **i)** $\{Y_t\}$ una submartingala con respecto a la filtración $\{\Omega_t(Y)\}$ y **ii)** una función $f(\cdot)$ no-decreciente y convexa [318] y tal que $E(|f(\sum_{t=1}^{n} \Delta Y_t)|) = E(|f(\sum_{t=1}^{n} \varepsilon_t)|) < \infty \; (n = 1,2,\dots)$ entonces, a partir de una extensión de la desigualdad de Jensen, según la cual $E(|f(\sum_{t=1}^{n} \varepsilon_t)|) \geq f[E(|\sum_{t=1}^{n} \varepsilon_t|)]$, se puede concluir que el proceso $\{f(\sum_{t=1}^{n} \varepsilon_t)\}$ es una submartingala[319].

2) Sea $\{Y_t\}$ una submartingala con respecto a la filtración $\{\Omega_t(Y)\}$, entonces el proceso $\{Max(\sum_{t=1}^{n} \varepsilon_t, 0)\}$ también es una submartingala con respecto a la filtración $\{\Omega_t(Y)\}$.

3) Sea $\{Y_t\}$ una martingala con respecto a la filtración $\{\Omega_t(Y)\}$, tal que $E(|\sum_{t=1}^{n} \Delta Y_t|^s) = E(|\sum_{t=1}^{n} \varepsilon_t|^s) < \infty (n = 1,2,\dots; s \geq 1)$, entonces el proceso $\{|\sum_{t=1}^{n} \varepsilon_t|\}$ es una submartingala con respecto a la filtración $\{\Omega_t(Y)\}$.

Con un razonamiento similar al empleado en las martingalas de parámetro discreto, se dice que un proceso $\{Y(t)\}(t \geq 0)$ es una martingala continua en el dominio del tiempocon respecto a su filtración natural $\{Y(\tau)\}(\tau < t)$, si se verifica que:

---

318 Es decir, tal que $f[\lambda Y_1 + (1-\lambda)Y_2] \leq \lambda f(Y_1) + (1-\lambda) f(Y_2)$ (donde Y_1 e Y_2 denotan dos valores cualesquiera de su dominio y $\lambda \in [0,1]$).

319 Dadas una variable $\{Y, \Omega(Y) = \{Y_1, Y_2, \dots\}, p(Y_j)\}$ y una función convexa f, como una generalización de la desigualdad de Minkowski, resulta la desigualdad de Jensen, según el cual $f[E(Y)] = f[\sum_i Y_i p(Y_i)] \leq \sum_i f(Y_i) p(Y_i) = E[f(Y)]$.

$$E\big[Y(t_n)/Y(t_{m,}),Y(t_\ell),\dots\big] = Y(t_m)(t_n \geq t_m \geq t_\ell \geq \cdots)$$

y, en general, se dice que $\{Y(t)\}$ es una martingala continua en el dominio del tiempo con respecto a la filtración $\Omega(Y(t))$, si se verifica que:

$$E\big[Y(t_n)/\Omega(\mathrm{Y(t_m)}),\Omega(\mathrm{Y(t_\ell)}),\dots\big] = Y(t_m)$$

Sea $\{Y(t)\}(t \geq 0)$ un proceso estocástico de parámetro continuo. Si se verifica que: **i)** $E\big[Y^{(+)}(t)\big] < \infty$ (donde $Y^{(+)} = Max(0,Y(t))$) y **ii)** $E\big[Y(t_n/\Omega(Y(t_m)),\Omega(Y(t_\ell)),\dots)\big] \geq Y(t_m)$, se dice que $\{Y(t)\}$es una submartingala continua en el dominio del tiempo con respecto a la filtración $\Omega(Y(t))$.

De la misma forma, se puede concluir que: **i)** $\{Y(t)\}$ es una supermartingala continua en el dominio del tiempo con respecto a la filtración $\Omega(Y(t))$si se verifica que $\{-Y(t)\}$es una submartingala continua en el dominio del tiempocon respecto a la filtración $\Omega(Y(t))$ y **ii)** $\{Y_t\}$es una martingala continua en el dominio del tiempo con respecto a la filtración $\Omega(Y(t))$, si es a la vez una supermartingala y una submartingala con respecto a dicha filtración.

El fundamento de la teoría límite de las martingalas se desarrolló a partir de la obra de Bernstein (1926)(1939)(1940)(1941) y Lévy (1935)(1937)(1954), como una generalización de los teoremas límite para sumas de variables aleatorias independientes. Esta tendencia cambió radicalmente con la publicación de losteoremas de convergencia de la martingala, debidos a Doob, (1953)[320], según los cuales: **1)** dada una submartingala con respecto a la filtración $\Omega_t(Y)$, entonces existe una variable Y tal que: **i)** $\boldsymbol{Y_t} \xrightarrow{ccc} Y$ y **ii)** $E(Y) \leq \liminf_{t\to\infty} E(|Y_t|) < \infty$. Asimismo, si la submartingala $\{Y_t \,/\, \Omega_t(Y)\}$ tiene varianza finita, entonces $Y_t \xrightarrow{m_2} Y$ (debe tenerse en cuenta que éste es un teorema de existencia que no permite obtener ninguna conclusión acerca de las características de la variable límite, excepto que

[320]Esta obra ejerció gran influencia sobre los desarrollos relacionados con los procesos estocásticos en los siguientes 30 años. En los últimos años del siglo XX se produjo un resurgimiento de la teoría límite de la martingala como una generalización de los resultados referidos a la suma de variables independientes.

posee momentos de primero y segundo orden finitos)[321] y **2)** dada una martingala con respecto a la filtración $\Omega_t(Y)$, tal que $E[sup(\varepsilon_t)] < \infty$, entonces se puede asegurar que $sup(\sum_{t=1}^{n} \varepsilon_t) < \infty$ y, en consecuencia, que $\sum_{t=1}^{n} \varepsilon_t$ converge casi-con-certeza (ver Shiraev (1981)).

27.- Donald L. McLeish, Donald Wilfrid KaoAndrews y la teoría de la mixingala

La martingala pertenece a una clase más general de procesos (propuesta por McLeish (1975) como un híbrido obtenido de la combinación de las martingalas y las sucesiones mixing) conocida como "*mixingala*". Dado un proceso $\{Y_t\}$ con $E(Y_t) = 0 (t = 1,2, \dots)$, se dice que es una "*$L^1$-mixingala con respecto a una filtración* $\Omega_{(t)}(Y) = \{\Omega_t(Y), \Omega_{t-1}(Y), \dots, \Omega_1(Y)\}$", si se verifica que $E\left(\left|E\left[Y_t \,/\, \Omega_{C,(t-j)}(Y)\right]\right|\right) \leq c_t \zeta_j \ (t = 1,2, \dots)$ (donde $\{c_t\}$ $(t = 1,2, \dots)$ y $\{\zeta_j\}$ $(j = 0,1,2, \dots)$ denotan sendas sucesiones de constantes determinísticas tales que: **i)** $c_t \geq 0 \ (t = 1,2, \dots)$; **ii)** $\zeta_j \geq 0 \ (j = 0,1,2, \dots)$ y **iii)** $\lim_{j \to \infty} \zeta_j = 0$). En otros términos, todo proceso $\{Y_t\}$ con valor esperado nulo, tal que su predictor $E\left(Y_t \,/\, \Omega_{C,(t-j)}(Y)\right)$ converge en valor absoluto a su valor esperado no-condicionado ($E(Y_t) = 0$), es una L^1-mixingala[322].

A partir de la interpretación de la mixingala como una martingala asintótica que, como el proceso diferencia de una martingala, satisface el teorema de convergencia, Andrews (1988) desarrolló una ley de los grandes números para L^1-mixingalas, a partir de la condición que el proceso seauniformemente integrable.Se dice que un proceso $\{Y_t\}$ es uniformemente integrable: **i)** si, para todo $\varepsilon > 0$ y arbitrariamente pequeño, existe una constante $c$tal que $E\left[|Y_t| \delta_{(|Y_t| \geq c)}\right] < \varepsilon \ (t = 1,2, \dots)$,

---

[321] Este resultado conduce a una demostración del teorema de las tres series condicionado (ver Sec. 25).

[322] En realidad, hubo diferentes definiciones del concepto de "mixing". En términos generales se dice que una sucesión de variables aleatorias cumple las condiciones mixing, si dos variables cualesquiera muy separadas entre sí pueden ser consideradas como aproximadamente independientes. Las condiciones que reemplazan a la condición de independencia se deben a Blum; Hanson; Koopman (1963). Para un análisis más profundo de estas condiciones, ver también Bártfai; Révész (1967).

donde:

$$\delta_{(|Y_t|\geq c)} = \begin{cases} 1 & si \quad |Y_t| \geq c \\ 0 & en\ cualquier\ otro\ caso \end{cases}$$

y **ii)** si existe un $s > 0$ y un $K < \infty$ tales que $E(|Y_t|^s) < K \ (t = 1,2,\dots)$. Por otra parte, si existe un $s > 0$ y un $K < \infty$ tales que $E(|Y_t|^s) < K \ (t = 1,2,\dots)$ y una sucesión $\{h_j\} \ (j = 0, \pm 1, \pm 2, \dots)$ tal que $\sum_{j=-\infty}^{\infty} |h_j| < \infty$, entonces se puede asegurar que el proceso $Z_t = \sum_{j=-\infty}^{\infty} h_j Y_{t-j}$ es uniformemente integrable.

Sea $\{Y_t\}$ un proceso uniformemente integrable, si existe una sucesión tal que $\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} c_t < \infty$, entonces se verifica que $\overline{Y_{(n)}} = \frac{1}{n} \sum_{t=1}^{n} Y_t \xrightarrow{p} 0$ y, por lo tanto, que $\{Y_t\}$ es una L^1-mixingala y que ésta satisface la ley débil de los grandes números. Esta condición, necesaria para asegurar la ergodicidad de $\{Y_t\}$, es alternativa a la condición $\sum_{j=0}^{\infty} |\gamma_j(Y)| < \infty$ (ver McLeish (1975)(1977), Hall; Heyde (1980)).

Como corolario de lascondiciones de la mixingala, se obtiene que:

1) Dado un proceso "*diferencia de una mixingala*"$\{\Delta Y_t\}$, entonces, para $c_t = E(Y_t)$ y

$$\xi_j = \begin{cases} 1 & para \quad j = 0 \\ 0 & para \quad j = 1,2,\dots \end{cases}$$

se puede concluir que$\{\Delta Y_t\}$es una L^1-mixingala con respeto a la filtración natural $\{\Omega_t(Y)\} = \{Y_t, Y_{t-1}, \dots, Y_1\}$, es decir, se puede concluir que:

$$E\left(\left|E\left(Y_t \,/\, \Omega_{C,(t-j)}(Y)\right)\right|\right) \leq E(|Y_t|)(t = 1,2,\dots)$$

Si, además, se verifica que $E(|Y_t|^s) < K$ (para algún $s > 1$ y $K < \infty$), entonces se puede asegurar que el proceso $\{Y_t\}$ es uniformemente integrable y puede ser interpretado como una L^1-mixingala con $c_t = K$, de modo que $\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} c_t = K < \infty$ y, por lo tanto, se puede concluir que se cumple la ley fuerte de los grandes números, $\overline{Y_{(n)}} = \frac{1}{n} \sum_{t=1}^{n} Y_t$

$\xrightarrow{p} 0$[323].

2) Dado un proceso $\{Y_t\}$ (de variables centradas) que admite una representación de la forma $Y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$, donde: **i)** $\{\varepsilon_t\}$denota un proceso diferencia de una martingala con $E(|\varepsilon_t|) < K$ (para $t = 1,2, \ldots$ y $K < \infty$) y **ii)** la sucesión de ponderaciones$\{\psi_j\}$es tal que $\sum_{j=0}^{\infty} |\psi_j| < \infty$, entonces se verifica que:

$$E\{|E(Y_t/\varepsilon_{t-j}, \varepsilon_{t-j-1}, \ldots)|\} = E\left(\left|\sum_{h=j}^{\infty} \psi_h \varepsilon_{t-h}\right|\right) \leq E\left(\sum_{h=j}^{\infty} |\psi_h \varepsilon_{t-h}|\right)$$

y, dado que: **i)** la sucesión $\{\psi_j\}$ es absolutamente sumable y **ii)** $E(|\varepsilon_{t-j}|) < K$, se puede escribir:

$$E\left(\sum_{h=j}^{\infty} |\psi_h \varepsilon_{t-h}|\right) = \sum_{h=j}^{\infty} |\psi_h| E(\varepsilon_{t-h}) \leq K \sum_{h=j}^{\infty} |\psi_h|$$

Luego, para: **i)** $c_j = K$ y **ii)** $\xi_j = \sum_{h=j}^{\infty} |\psi_h| < \infty$, se verifica que $\lim_{j\to\infty} \xi_j = 0$. De modo que se puede concluir que $\{Y_t\}$:L^1-mixingala. Si además se verifica que$E(|\varepsilon_t|^s) < K < \infty$ $(s = 2,3, \ldots)$, se puede asegurar que $\{Y_t\}$ es un proceso uniformemente integrable y que satisface la ley fuerte de los grandes números, $\overline{Y_{(n)}} \xrightarrow{p} 0$.

28.- Conclusiones: A modo de resumen

1.- Más allá de la indiscutible importancia que sus resultados representaron para el desarrollo de la teoría del azar, ni los Bernoulli ni de Moivre lograron resolver el problema de la identificación del supuesto verdadero valor de una probabilidad (θ), en la medida que no consiguieron solucionar la cuestión central del mismo, la inversión de la probabilidad a partir de la información proporcionada por una sucesión finita de observaciones (Y_n). Es decir, no

[323] Philipp (1967) fue el primero en demostrar que las leyes del logaritmo iterado para procesos estrictamente estacionarios satisfacen las condiciones mixing. Este trabajo fue continuado por Iosifescu (1968) y Reznik (1968).

lograron pasar de la probabilidad directa, $p(Y_n/\theta)$ de los tratadistas clásicos a la probabilidad inversa (o indirecta), $p(\theta/Y_n)$ .Su fracaso se debió fundamentalmente a la imposibilidad, en el contexto de su interpretación determinística, de considerar a θ como una variable aleatoria. Debe tenerse en cuenta que tanto los Bernoulli como de Moivre –como la mayoría de los científicos de la época-, en su fidelidad a la teología Newtoniana, veían en sus leyes (débiles) de los grandes números el argumento según el cual la estabilidad de las frecuencias demostraba la presencia de la "*Divina Providencia*". En ese marco filosófico θ sólo podía ser interpretada como una constante (de valor desconocido) y la frecuencia relativa como una variable aleatoria.

2.- Recién en 1764 apareció la solución rigurosa (aunque parcial) propuesta por Thomas Bayes al problema de la inversión, contenida en su famoso "*An essay toward solving a problem in the doctrine of chances*", que consideraba a θ como una variable aleatoria con una distribución de probabilidades "a priori" conocida, a partir de la cual era posible la caracterización de las propiedades y la definición de la distribución de probabilidades de la variable θ, condicionada por un conjunto finito de observaciones,(θ/Y_n).

Las hipótesis sobre las que se basó la demostración de Bayes dieron origen a críticas referidas fundamentalmente: **i)**a la utilización de un concepto de subjetividad racional en un contexto objetivista, que lo condujo a una interpretación supuestamente confusa e indefinida de la noción de probabilidad; **ii)**a la conjetura sobre la existencia de causas estables generadoras del orden de ocurrencia de los eventos y su asimilación a propensiones físicas inherentes a la aleatoriedad del sistema y **iii)**a las características exclusivamente intuitivas de la justificación de la condición de equiprobabilidad "a priori" de la variable θ a partir del supuesto de ignorancia absoluta acerca de su comportamiento previamente a la realización de alguna observación.

Con respecto a la confusión en la interpretación de la noción de probabilidad y su asimilación a propensiones físicas, debe tenerse en cuenta que, para los probabilistas de los siglos XVII y XVIII no existían definiciones rígidas de probabilidad, sino distintos métodos de inferencia de sus valores cuyas características dependían del contexto en el que debía ser utilizada y que lo que le impidió a Bayes acceder a una definición aceptable de probabilidad objetiva no fue la presunta ambigüedad generada por la doble interpretación frecuencista-propensionalista, sino las limitaciones de la teoría propensionalista para asimilar un evento individual a un elemento particular de un colectivo.

Es innegable que la interpretación propensionalista incluye muchos elementos subjetivos y que, en consecuencia, no resulta adecuado como modelo identificatorio de una probabilidad objetiva singular, en particular en aquellos casos en los cuales no se cuenta con información estadística obtenida de sucesiones suficientemente largas de observaciones. En el caso en que se contara con una sucesión de observaciones y no existieran circunstancias ajenas a los datos estadísticos, el modelo para identificar la probabilidad (teórica) sobre la ocurrencia de un evento a partir de la frecuencia relativa (observable) será, de acuerdo con la nomenclatura de Popper, "*impermeable a la falsación estricta*". Un vano intento de solución propuesto por Popper a esta dificultad consistió en apelar a lanoción de "*falsación metodológica*" según la cual, aunque las proposiciones sobre probabilidades no sean estrictamente "*falsables*", pueden ser utilizadas (y de hecho lo son en las ciencias experimentales) como argumentos "*falsables*" mediante la utilización de tests estadísticos. Debe tenerse en cuenta que, de acuerdo con este procedimiento, cualquier hipótesis puede ser refutada metodológicamente aún cuando desde un punto de vista estrictamente lógico no sea refutable y, por lo tanto, que este criterio de "*falsación*" no resuelve el problema de la identificación de la probabilidad.

En lo que se refiere al tercer comentario, cabe destacar que, contrariamente a lo que suele afirmar la literatura sobre el particular, el postulado de uniformidad de la distribución no pretende ser una caracterización de la distribución de probabilidades de θ, sino de la variable observable Y_n y que la operacionalización realizada por Bayes de la conjetura de Hume y los teoremas de de Finetti y Murray permiten concluir que la distribución uniforme de la variable θ es condición necesaria y suficiente para el cumplimiento del postulado de Bayes.Lo que permite concluir que la operacionalización de Bayes no logra demostrar la asimilación del supuesto de ignorancia absoluta al de uniformidad de la distribución $dF(\theta)$, sino que permite solamente una definición condicionda por la distribución binomial de la variable X_n.Es decir que, más allá de las justificaciones mencionadas, el teorema de Bayes no logró evitar su fundamentación en supuestos metafísicos ni, en consecuencia, obtener una solución general al problema de la identificación de la función de densidad de θ.

3.-Poisson, basándose en la consideración que el modelo de causación implícito en la ley de los grandes números de Bernoulli adolecía del defecto de excesiva simplificación, propuso la primera generalización a partir de un esquema basado en el supuesto de la existencia de múltiples causas que influían sobre el evento observado, generando probabilidades distintas sobre la ocurrencia del mismo.

Como Bernoulli y de Moivre, Poisson no podía concebir un universo "*gobernado*" por el azar. Pero, a diferencia de ellos -que suponían la existencia de un "O*rden Providencial*"- Poisson sostenía que la tendencia de los fenómenos a exhibir regularidades estadísticas era inherente "*...al estado natural de las cosas*" y, al igual que Bernoulli, Leibniz, de Moivre, Price y Condorcet, consideró que la eventual no verificación del principio de estabilidad de las frecuencias, no significaba una refutación del "*principio de permanencia de las causas*" que gobernaban la naturaleza, sino el reconocimiento de que algunas de estas causas, podrían haber sido reemplazadas por otras, lo que hacía que las probabilidades de ocurrencia del evento en una prueba dada, variaran. Este supuesto lo condujo a adoptar el concepto de convergencia al promedio de las probabilidades,desarrollado posteriormente en los trabajos de Bienaymé y Lexis referidos al estudio de la homogeneidad y estabilidad en un esquema de pruebas repetidas propuesto por Condorcet quien, fundándose en el principio de uniformidad de la naturaleza y en un intento por cuantificar la influencia de la experiencia histórica sobre el grado de creencia de la interpretación logicista, desarrolló una filosofía de la relación entre dicho principio de uniformidad y el método científico. A partir de esta conjetura Poisson reconoció, de acuerdo con el principio de Cournot, la posibilidad de calcular, en casos de fenómenos repetibles, la variación en la asignación de la probabilidad en función de la acumulación de observaciones, considerando a la experiencia histórica como la única regla capaz de influir en dicho proceso.

4.- Los primeros intentos de vinculación de la teoría de la medida, el concepto de independencia estocástica en las sucesiones de eventos repetibles y una interpretación geométrica de la probabilidad se deben a Borel. Como auténtico sucesor intelectual de Bertrand y Poincaré y como miembro de la escuela probabilística francesa, Borel asimiló los conceptos de indeterminismo e impredictibilidad, postulando que el indeterminismo estaba asociado a la imposibilidad de predecir con certeza. Consideró que la única aproximación al concepto de probabilidad era mediante una ecléctica interpretación pluralista que admitía una noción surgida de la interpretración de azar como sinónimo de ignorancia y una noción frecuencista –basada en el principio de estabilidad estadística en el comportamiento de los fenómenos- y que esta interpretación determinística de la probabilidad no implicaba una definición, sino un intento de verificación de la interpretación de la probabilidad como límite de una frecuencia relativa. Una posición que lo condujo a su ley fuerte de los grandes números.

Contrariamente a los postulados de la ley débil de los grandes números, que no establece una convergencia, sino la probabilidad de una aproximación, la ley fuerte de los grandes números demuestra que la frecuencia relativa converge a un límite en el sentido del análisis, si se excluye un conjunto de resultados con probabilidad nula.

5.- La demostración de Poisson y los trabajos de Cournot, Ellis, Venn, el principio de asociación de ideas de Locke –consistente en igualar experiencia, creencia y probabilidad- y la reacción del empirismo británico contra el racionalismo continental de Laplace, condujeron a intentar una interpretación que conciliase la teoría de la probabilidad con los principios del indeterminismo fundado por Fechner a mediados del siglo XIX, basado en el concepto de "*Kollektiv*" considerado como una sucesión de resultados obtenidos de una serie de observaciones repetidas en igualdad de condiciones. La noción de "*Kollektivereihe*" tuvo su culminación en la reformulación realizada por Reichenbach y von Mises del concepto de probabilidad definida como una frecuencia relativa.

De acuerdo con su operacionalismo-positivista, von Mises consideró la necesidad de distinguir entre "*colectivos empíricos*" (que pueden ser hallados en el mundo real que, por lo tanto, son observables y que, en consecuencia, están formados por un número finito de observaciones) y "*colectivos matemáticos*" (formados por una sucesión infinita de observaciones). Basándose en una circunstancia que se verifica habitualmente en la física y considerando que las sucesiones infinitas son abstracciones matemáticas o idealizaciones de la realidad empírica necesarias para obtener una representación matemática admisible de la realidad, von Mises estableció el principio (muy discutible) según el cual un colectivo matemáticoinfinito puede ser expresado en términos analíticos por un colectivo empírico finito (convirtiendo a la teoría de la probabilidad en una teoría de convergencia de sucesiones).

En consonancia con esta posición empirista, el análisis de von Mises se basó en una filosofía operacionalista, según la cual los principios teóricos de una ciencia debían ser definidos en términos de fenómenos observables con las características de un colectivo empírico. Una interpretación que supone que la naturaleza de los fenómenos repetibles es tal que: **i)** es posible, por abstracción, obtener ciertos conceptos matemáticos que permiten formular las leyes empíricas que rigen su comportamiento; **ii)** recurriendo nuevamente a la abstracción y, a partir de dichas leyes empíricas, es posible definir los axiomas de la teoría matemática asociada a dicho comportamiento y **iii)** a partir de esta teoría matemática es posible descubrir consecuencias que permiten la explicación y predicción de otros fenómenos repetibles.

En el marco de este paradigma filosófico conjeturó que los colectivos empíricos obedecen a dos principios fundamentales: **i)** la "*ley de estabilidad de las frecuencias estadísticas*" y **ii)** la "*ley de irregularidad*", que postula la aleatoriedad absoluta en la proyección de los atributos sobre los elementos del colectivo. A partir de los cuales, mediante un proceso de abstracción (o idealización), definió los axiomas de convergencia y de aleatoriedad ("*Regellosigkeitsaxiom*") aplicables a colectivos matemáticos.

La necesidad de refutar las críticas filosóficas sobre el operacionalismo derivaron en un intento de justificar rigurosamente la axiomática de von Mises utilizando argumentos de la teoría algorítmica de la complejidad y la teoría ergódica.El teorema de Gödel y los conceptos de la teoría de la información condujeron a la demostración de la imposibilidad de resolución del "*Entscheindungsproblem*" y de la convalidación de la propuesta de Leibniz de asimilación del concepto de aleatoriedad de una serie a la ausencia de una ley que permitiera explicar su estructura.Es decir, a la demostración de laausencia absoluta de rregularidades no-triviales en su trayectoria y, en consecuencia, de la imposibilidad de obtener una formulación matemática rigurosa del "*Regellosigkeitsaxiom*".

Todo lo anterior permite concluir que, tanto la interpretación de von Mises como las modificaciones de Reichenbach, a fin de evitar cualquier tipo de regularidad en las sucesiones de eventos que constituyen su fundamento, asignaron a la definición de probabilidad un contenido estrictamente matemático mediante la imposición de complicadas condiciones que, inevitablemente, restringían el concepto de aleatoriedad total y demostraban que era imposible dar un carácter matemáticamente preciso a la noción de "*irregularidad absoluta*".

Estas restricciones, contradiciendo las objeciones de von Mises a la interpretación clásica como una teoría matemática que no trataba con eventos concretos, si bien agregaron algún rigor matemático al planteo frecuencista, eliminaron sus fundamentos empíricos, transformándolo también en una teoría puramente matemática que, en vez de tratar con resultados favorables y resultados posibles, trata con límites que son entidades matemáticas abstractas y en la cual las demostraciones de los teoremas se obtienen exclusivamente mediante la utilización de métodos lógico-matemáticos, convirtiéndolo en una extravagante teoría semi-matemática que, curiosamente fue más tarde reivindicada por Kolmogorov y Khinchin.

6.- La falencia de la interpretación frecuencista respecto de la imposibilidad de proporcionar probabilidades objetivas de ocurrencia de fenómenos singulares ("*probabilidades singulares*") condujo a Keynes y a Popper a proponer nuevas interpretaciones de la probabilidad basadas, respectivamente, en relaciones lógicas en un ámbito formado por ideas abstractas y en la teoría de las propensiones, de carácter indiscutiblemente metafísico.

Esta incapacidad de las definiciones logicista y propensionalista para identificar el supuesto verdadero valor de la probabilidad de ocurrencia de un evento dio origen al modelo subjetivista de deFinetti-Ramsey fundado en una concepción aleatorista generada en el supuesto ontológico de azar absoluto, en el marco de una filosofía empiricista-pragmatista-relativista que postula la validez de evaluaciones de probabilidades no coincidentes entre sí.

El núcleo de la teoría deFinettiana y la conciliación de su filosofía neo-Bayesiana con su matemática están contenidos en la noción de intercambiabilidad y en el teorema de representación. Los postulados de este teorema son equivalentes, en el ámbito del análisis dinámico, a la propiedad de ergodicidad. Lo cual implica que un proceso ergódico garantiza la unicidad de los límites de sus frecuencias relativas y, como corolario permite concluir que las leyes de los grandes números son casos especiales del teorema ergódico.

No obstante, debe tenerse en cuenta que, si bien a partir de una aproximación subjetivista, la propiedad de intercambiabilidad permite resolver algunas controversias referidas a las relaciones entre probabilidad y frecuencia y al teorema de Bayes, el teorema de representación por sí mismo no logra justificar los postulados de la ley de los grandes números, no por una insuficiencia del método sino por las exigencias esenciales de la interpretación subjetivista, que condiciona dicho teorema a una asignación coherente de la probabilidad inicial.

Luego, se puede concluir que la propuesta de identificación del verdadero valor de una probabilidad a partir de los postulados de una ley de los grandes números basada en la propiedad de intercambiabilidad, en los postulados del teorema de representación y en su interpretación en el ámbito de la teoría ergódica, es inconsistente dado que: **i)** a partir de una interpretación subjetivista del teorema de representación es posible concluir que la proposición acerca de la convergencia de la frecuencia relativa a un “verdadero valor” de la probabilidad, en los casos de fenómenos Bernoullianos, no es más que una ilusión metafísica, que en realidad es atribuible a un comportamiento asintótico de las evaluaciones personales de las probabilidades iniciales de acuerdo con un esquema de condicionamiento Bayesiano, que conduce a una asignación intersubjetiva fácilmente confundible con la probabilidad objetiva; **ii)** dado el carácter axiomático de las premisas en que se basan los resultados de Koopman, von Neumann, Birkhoff, Khinchin y Sinai referidos a los resultados del teorema de partición ergódica y al método de las funciones arbitrarias (según el cual los fenómenos admiten una explicación determinística en términos de mecánica clásica y satisfacen la condición de continuidad absoluta), su aceptación es de carácter exclusivamente subjetivo y **iii)** la justificación analítica de la interpretación propensionalista de la teoría ergódica implica su afectación por las mismas característica metafísicas del modelo propensionalista en el cual la asignación de probabilidades es inevitablemente subjetiva.

7.- Las dos propuestas fundamentales contenidos en los “*Grundbegriffe der Wahrscheinlichkeitsrechnung*” de Kolmogorov son: **i)** la teoría de las probabilidades condicionadas para conjuntos infinitos de eventos elementales y su aplicación a casos en los cuales al evento condicionante le corresponde una probabilidad nula y **ii)** la teoría general de los procesos estocásticos.

Su análisis, en principio, se basa en la condición de ergodicidad, que permite asegurar que las probabilidades de transición de un proceso ergódico son asintóticamente iguales para todas sus trayectorias. A partir de una interpretación frecuencista de estas probabilidades se puede suponer que, de acuerdo con esta propiedad, todas las sucesiones finitas de realizaciones del proceso observarán el mismo comportamiento asintótico. Lo que permite concluir nuevamente que la propiedad de ergodicidad conduce a la ley de los grandes números.

Posteriormente, su interpretación de la vinculación entre la teoría de algoritmos, la definición de Church de aleatoriedad de una sucesión, el principio de ergodicidad y la teoría de la probabilidad, condujo a Kolmogorov a una reivindicación de la interpretación de von Mises y a la introducción de la definición de "*prácticamente cierto*", atribuible a todo evento cuya probabilidad fuera arbitrariamente próxima a su frecuencia relativa (obtenida a partir de una sucesión finita suficientemente larga de observaciones). Una definición que revela su reconocimiento de la debilidad de la interpretación frecuencistay la necesidad de considerar una versión finita de la ley de los grandes números y su justificación mediante una deducción empírica de la axiomática contenida en los "*Grundbegriffe*".

Ahora bien, dado: **i)** que la asimilación de las frecuencias relativas a las probabilidades no está suficientemente justificada; **ii)** que, obviamente, las frecuencias relativas obtenidas a partir de sucesiones finitas de observaciones siempre asumen números racionales y que, en consecuencia, la aproximación de Kolmogorov genera una injustificable restricción al conjunto de valores a asumir por las probabilidades y **iii)** queuna justificación completa del sistema axiomático implicaría una demostración de por qué las probabilidades deben asumir valores reales únicos, se puede concluir que la interpretación de las leyes de probabilidades como aproximaciones a distribuciones empíricas a través de la noción de certeza práctica y de una versión finita de la ley de los grandes números, constituye, también, un intento infructuoso de identificación de una probabilidad.

8.- Si bien la mayoría de los teoremas de la teoría de los procesos estocásticos y, en particular los teoremas de convergencia de la martingala debidos a Doob y de la mixingala de McLeish-Andrews, están dirigidos a demostrar que, a partir de ciertas hipótesis analíticas, algunas de las propiedades refereidas a la convergencia de las sucesiones finitas de observaciones se verifican casi-con-certeza y, por lo tanto, constituyen una alternativa metodológica al estudio de la ley de los grandes números, también adolecen del defecto de estar basados en una estructura abstracta que no admite una representación capaz de combinar contextos matemáticos y no-matemáticos.

9.- Finalmente, generalizando los comentarios anteriores se puede afirmar que la aproximación puramente matemática introducida por Borel y generalizada por Faber, Hausorff, Cantelli, Kolmogorov, Prokhorov y Khinchin y la teoría de la martingala de Doob-McLeish-Andrews colocan a la probabilidad a un nivel injustificable de abstracción y conducen a la conclusión que la interpretación basada en la teoría de la medida constituye una estructura vacía que limita el ámbito de aplicación del concepto de probabilidad respecto de cualquier interpretación obtenida de la combinación de contextos matemáticos y no matemáticos. Que los fundamentos de la teoría de la medida sirven más para redimir a las conciencias matemáticas que para proporcionar una metodología apropiada para la aplicación de la teoría de la probabilidad.

Bibliografía

- Adams, E. (1965): "The logic of conditionals". Inquiry, vol. 8 (166-197).
- Adams, E. (1966): "Probability and the logic of conditionals". En Hintikka; Suppes (eds.).
- Aleksandrov, A.D.; Kolmogorov, A.N.; Laurent'ev, M. (eds.) (1963): "*Mathematics. Its contents.Methods and meaning*"- MIT Press (original en ruso, 1956).
- Anderson, A.R.; Belnap, N.D.; Dunn, J.M. (eds.)(1992): "*Entailment: The logic of relevance and necessity*". Princeton University Press.
- Antretter, G. (1989): "*Von der Ergodenhypothese zu stochatischen Prozessen: Die Entfaltung der Theorie der Markovchen Ketten und Prozessen von dem Hintegrund statistisch-machanischer Probleme*". Schrifliche Hausarbeit aus dem fach Mathematik, Universität München.
- Arbuthnot, J. (1710-1712): "An argument for Divine Providence, taken from the regularity observ'd in the birth of both sexes". Philosophical Transactions of the Royal Society, vol. 27 (186-190).
- Archibald, R.C. (1926): "A rare pamphlet of de Moivre and some of his discoveries". Isis, vol. 8, n° 4 (671-684).
- Arnauld, A.; Nicole, P. (1695): "*La logique ou l'art de penser*". Pierre Clair & FranCois Girbal.
- Atkinson, A.C.; Fienberg, S.E. (eds.)(1985): "*A celebration of statistics*". The ISI Centenary Volume.Springer.
- Azuma, K. (1967): "Weighted sums of certain dependent random variables". Tohoku Mathematical Journal, vol. 19 (357-367).
- Bachelier, L. (1900): "*Théorie de la spéculation*". Gauthier-Villars. Reeditado en Annales Scientifiques de l'École Normale Supérieur, vol. 17 (21-86).
- Bachelier, L. (1912): "*Calcul des probabilités*". Gauthier-Villars.
- Barnard, G.A. (1958): "Thomas Bayes: Abiographical note". Bometrika, vol. 45 (293-315).
- Bártfai, P.; Révész, P. (1967): "On a zero-one law". Zeitschrift f¨r Wahrscheinlichkeitstheorie, vol. 7 (43-47).
- Basu, A.K. (1973): "A note on Strassen's version of the law of the iterated logarithm". Proceedings of the American Mathematical Society, vol. 41 (596-601).
- Bauer, H. (1981): "*Probability theory and elements of measure theory*". Academic Press.
- Bayes, Th. (1764): "An essay toward solving a problem in the doctrine of chances". Philosophical Transactions of the Royal Society, vol. 53 (370-418). Reeditado en Barnard (1958) y en Pearson; Kendall (1970).
- Bencerraf, P.; Putnam, H. (1983): "*The philosophy of mathematics: Select readings*". Cambridge University Press.

- Bergström, H. (1970): "Comparison method for distribution functions of sums of independent random variables. Theory of Probability and Applications, vol. 15 (430-457).
- Bernardo, J.M.; De Groot, M.H.; Lindley, D.V.; Smith, A.F.M. (eds.)(1980): "Bayesian statistics". Proceedings of the 1st. International Meeting, Universidad de Valencia.
- Bernoulli, D. (1738): "Specimen theoriæ novæ de mensura sortis". Commentarii Academiæ Scientiarum Imperialis Petropolitanæ, vol. 5 (175-192). Traducción al inglés como"Exposition of a new theory on the measurement of risk". Econometrica, vol. 22 (23-36), 1954.
- Bernoulli, D. (1770-1771): "*Mensura sortis ad fortuitam successionem rerum naturaliter contingentium applicata*". Reeditado en "*Die Werke von Daniel Bernoulli*", Birhäuser (1982)
- Bernoulli, J.(1713a): "*Meditationes*". En "*Die Werke von Jakob Bernoulli*", Basel Naturforschende Gesellschaft, 1969-75.
- Bernoulli, J.(1713b): "*Ars conjectandi*". En "Die Werke von Jakob Bernoulli", Basel Naturforschende Gesellschaft, 1969-75.Traducción al inglés por Harvard University, Department of Statistics.Tecnical Report nº 2, 1966.
- Bernoulli, J. (1969-1975): "*Die Werke von Jakob Bernoulli*". Basel Naturforschende Gesellschaft.
- Bernoulli, N. (1709): "De usu artis conjectandi in jure". *En "Die Werke von Jakob Bernoulli*", vol. 3 (287-326).
- Bernshtein, S.(1917): "An essay on the axiomatic foundation of the theory of probability". Proceedings Khar'kov Mathematical Society (original en ruso).
- Bernstein, F. (1912): "Über eine Anwendung der Mengenlehre auf ein aus der Theorie der säkularen störungen herrührendes Problem". Mathematische Annalen, vol. 71 (417-439).
- Bernstein, S.N. (1926): "Sur l'extension du théorème limite du calcul des probabilities aux sommes des quantitésdépendants". Mathematische Annalen, vol. 97 (1-59).
- Bernstein, S.N. (1937): "On some modifications of Chebychev's inequality". Dokl.Akad.Nauk.. SSSR, vol. 17 (275-278) (original en ruso).
- Bernstein, S.N. (1939): "Quelques remarques sur le théorème limite de Liapunoff". Dokl. Akad. Nauk SSSR, Serie Matemática, vol. 24 (3-8). Original en ruso.
- Bernstein, S.N. (1940): "New applications of almost independent quantities". Izv. Akad. Nauk SSSR, Serie Matemática, vol. 24 (137-150). Originalen ruso.
- Bernstein, S.N. (1941): "On sums of dependent quantities having a join regression which is almost zero". Dokl. Akad. Nauk SSSR, Serie Matemática, vol. 32 (303-307). Original en ruso.
- Berti, P.; Regazzini, E.; Rigo, P. (1992): "Finitely additive Radon-Nikodým theorem and concentration of a probability with respect to a probability". Proceedings of the American Mathematical Society, vol. 114 (106-108).

- Berto, F. (2009): "*Tutti pazzi per Gödel. La guida completa al teorema di incompletezza*". Laterza.
- Bertrand, J.(1888): "*Calcul des probabilities"*. Gauthier-Villars.
- Bessel, F.W. (1838): "Untersuchungen über Wahrscheinlichkeit der Berobachtungsfehler". Astronomische Nachrichten, vol. 15 (358-369). Reeditado en Engelman (ed.)(1876).
- Bienaymé, I.J. (1855): "Communication sur un principe que M. Poisson avait cru decouvrir et qu'il avait appelé loi des grandes nombres". Séances et Travaux de l'Académie des Sciences Morales et Politiques, Serie 3, vol. 11 (379-389).
- Billingsley, P. (1961): "The Lindeberg-Lévy theorem for martingales". Proceedings of the American Mathematical Society, vol. 12 (788-792).
- Birkhoff, G.B. (1931): "Proof of the ergodic theorem". Proceedings of the National Academy of Sciences, vol. 17 (656-660).
- Birkhoff, G.B. (1932): "Probability and physical systems". Bulletin of the American Mathematical Society, vol. 28 (361-379).
- Birkhoff, G.D.; Koopman, B.O. (1932): "Recent contributions to the ergodic theory". Proceedings of the National Academy of Sciences, vol. 18 (279-282).
- Blaug, J.C.; McLure, M. (eds.)(1999): "*The international library of critical writings in economics*". Elgar.
- Blum, J.R.; Hanson, D.L.; Koopman, L. (1963): "On the strong law of large numbers for a class of stochastic processes". Zeitschrift für Wahrscheinlichkeitstheorie", vol. 2 (1-11).
- Bochner, S. (1939): "Additive set functions on groups". Annals of Mathematics, vol. 40 (769-799).
- Bohlmann, G. (1909): "Die Grundbegriffe der Wahrscheinlichkeitsrechnung in ihrer Anwendung auf die Lebensversicherung". Atti del IV Congresso Internazionale dei Matematici.
- Boltzmann, L. (1866): "Über die mechanische Bedeutung des zweiten Hauptsatzes der Wärmetheorie". Wissenschafliche Abhandlungen, vol. 1 (9-33).
- Boltzmann, L. (1871): "Über das Wärmeglechgewicht swischen mehratomigen Gasmolekülen". Wissenschafliche Abhandlungen, vol. 1 (237-258).
- Boltzmann, L. (1872): "Weitere Studien über das wärmegleichgewicht unter Gasmolekülen". Wissenschafliche Abhandlungen, vol. 1 (316-402).
- Boltzmann, L. (1877a): "Bemerkungen über einige Probleme der mechanischen Wärmetheorie". Wissenschafliche Abhandlungen, vol. 2 (112-148).
- Boltzmann, L. (1877b): "Über die Beziehung zwischen dem zweiten Hauptsatze der mechanischen Wärmetheorie und der Wahrscheinlichkeitsrechnung respective den Sätzen über das Warmegleichgewicht". Wissenschafliche Abhandlungen, vol. 2 (164-223).
- Boltzmann, L. (1887): "Über die mechanischen Analogien des zweiten Hauptsatzes der Thermodynamik". Wissenschafliche Abhandlungen, vol. 3 (258-271).

- Boole, G. (1854): "*An investigation of the laws of thought*", Londres. Reeditado en Nueva York (1958).
- Bonnot de Condillac, E. (1746): "*Essai sur l'origine des connaisance humaines*". En "*Oeuvres*", París (1795).
- Bonnot de Condillac, E. (1754): "*Traité des sensations*". En "*Oeuvres*", París (1795).
- Borel, E. (1906a): "Sur les principles de la théorie cinétique des gaz". En "*Oeuvres de Emile Borel*".
- Borel, E. (1906b): "La valeur pratique du calcul des probabilités". En "*Oeuvres de Emile Borel*".
- Borel, E. (1909a): "Les probabilités dénombrables et leurs applications arithmétiques". Rendiconti del Circolo Matematico di Palermo, vol. 27 (247-271).
- Borel, E. (1909b): "*Éléments de la théorie des probabilités*". Hermann.Traducción al inglés, "*Elements of the theory of probability*", Prentice-Hall, 1965.
- Borel, E. (1912): "Sur un problème de probabilités relatif aux fractions continues". En "*Oeuvres de Emile Borel*", vol. 2 (1085-1091), 1939.
- Borel, E. (1914): "*Le hasard"*. Presses Universitaires de France.
- Borel, E. (1924): "A propos d'un traité de probabilités". "*Oeuvres de Emile Borel*", vol. 4 (2169-2184), 1939.
- Borel, E. (1925): "*Principes et formules classiques du calcul des probabilités"*. Gauthier-Villars.
- Borel, E. (1930): "Sur les probabilités universellment négligeables". Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences, vol. 190 (537-559).
- Borel, E. (1929-1935): "*Traité du calcul des probabilités et des ses applications, valeur pratique et philosophie des probabilités"*. Gauthier-Villars.
- Borel, E. (1939): "*Valeur pratique et philosophie des probabilities*". Gauthier-Villars.
- Boyle, R. (1666): "*Origin of forms and quantities*". En "*The works of the Honorable Robert Boyle*".Londres (1772).
- Boyle, R. (1675): "*Some considerations about the reconcileableness of reason and religion*". En "*The works of the Honourable Robert Boyle*", Londres (1772).
- Brown, B.M. :(1971) "A generalized three series theorem". Proceedings of the American Mathematical Society, vol. 28 (572-577).
- Bruns, H. (1897):"Über die Darstellung von Fehlergesetzen". Astronomische Nachrichten, vol. 43 (329-349).
- Bruns, H. (1898):"Sur collective-masslehre". Philosophi Studien, vol. 14 (339-375).
- Bruns, H. (1905):"Wahrscheinleichkeitsrechnung und Kollektvmasslehere".En "Beiträge zur Quotenrechnung".Berichte der Könglich Sächsischen Gesellschaft der Wissenschaften, vol. 58 (216-222).

- Bruns, (1897): "Das Gruppenschema für zufällige Ereignisse". Abhandlungen der Könglich Sächsischen Gesellschaft der Wissenschaften, vol. 29 (577-628).
- Brush, S. (1965-1966-1972): "*Kinetik Theorie*". Pergamon Press.
- Buffon, G.L-. (1777): "Essais d'arithmètique morale". Histoire Naturelle, Supplément.o vol. 4 (46-148).
- Burdach, K.F. (1842): "Die personalische Bersonderheit". Blicke ins Leben, vol. 2 (232-271), Leopold Voss.
- Burkholder, D.L. (1966): "Martingale transforms". Annals of Mathematical Statistics, vol. 37 (1494-1504).
- Butler, J. (1736): "*The analogy of religion, natural and revealed, to the constitution and course of nature*". Londres.
- Calude, C. (2000): "Who is afraid of randomness?". http:www.cs.auckland.ac.nz/staff-cgi-bin/mjd/secondcgi.pl.
- Cantelli, F.P. (1916a): "Sulla legge dei grandi numeri". Atti della Reale Accademia dei Lincei. Memorie della Classe di Scienze Fisiche, Matematiche e Naturali, vol. 11 (330-349).
- Cantelli, F.P. (1916b): "La tendenza ad un limite nel senso del calcolo delle probabilità". Rendiconti del Circolo Matematico di Palermo, vol. 41 (191-201).
- Cantelli, F.P. (1917a): "Sulla probabilità come limite della frequenza". Atti della Reale Accademia dei Lincei, Classe di Scienze Fisiche, Mathematiche e Naturali, Rendiconti, vol. 26 (39-45).
- Cantelli, F.P. (1917b): "Su due applicazioni d'un teorema de G. Boole alla statistica matematica". Atti della Reale Accademia Nazionale dei Lincei, Classe di Scienze Fisiche, Mathematiche e Naturali, Rendiconti., vol. 26 (295-302).
- Cantelli, F.P. (1935): "Consideration sur le convergence dans le calcul des probabilités". Annales Institut Henri Poincaré, vol. 5 (1-50).
- Cardano, G. (1539): "Pratica aritmeticæ et mensurandi singularis". En "Opera omnia" (1663). Traducción al inglés por Stuttgart-Bad Cannstatt (1966).
- Cardano, G. (1564): "*Liber de ludo aleæ*". Treaducción al inglés como "*The book on games of chance*", Nueva York (1961).
- Castelnuvo, G. (1918): "*Calcolo delle probabilità*". Zanichelli.
- Casti, J.L. (1990): "Chaos, Gödel and truth". En Casti; Karlqvist (eds.).
- Casti, J.L.; Karlqvist, E. (eds.)(1990): "*Beyond belief: Randomness, prediction and explanation in science*". CRC Press.
- Chaitin, G.J. (1966): "On the length of programs for computing finite binary sequences". Journal Association Computing Machinery, vol. 13 (145-159).
- Chaitin, G.J. (1969): "On the length of programs for computing finite binary sequences: Statistical considerations". Journal Association Computing Machinery, vol. 16 (145-159).
- Chaitin, G. (1975): "A theory of program size formally identical to information theory". Journal ACM, vol. 22 (329-340).

- Chaitin, G. (1987a): "*Algorithmic information theory*". Cambridge University Press.
- Chaitin, G.J. (1987b): *"Information, randomness and incompleteness: Papers on algorithmic information theory"*. World Scientific.
- Chaitin, G. (1988): "Randomness in arithmetic". Scientific American, vol. 259 (80-85).
- Charleton, W. (1654): "*Physiologia Epicuro-Gassendo-Charletoniana*", Londres. Reeditado Nueva York-Londres (1966).
- Chebychev, P.L. (1867): "Des valeurs moyennes".Journal de Mathématiques pures et appliqués, vol. 12 (177-184).
- Chebychev, P.L. (1887): "Sur deux théorèms relatives aux probabilités". Acta Mathematica, vol. 14 (98-116).
- Chover, J. (1966): "A law of the iterated logarithm for stable summands". Proceedings of the American Mathematical Society, vol. 17 (441-443).
- Chow, Y.S.; Teicher, H. (2003): "*Probability theory: Independence, interchangeability, martingales*". Springer.
- Chung, K.L. (1974): "Note on some strong laws of large numbers". American Journal of Mathematics, vol. 69 (189-192).
- Chung, K.L. (1974): "*A course in probability theory*". Academic Press.
- Church, A. (1936a): "A note on the Entscheindungsproblem". Journal of Symbolic Logic, vol. 1 (40-41).
- Church, A. (1936b): "An unsolvable problem of elementary number theory". American Journal of Mathematics, vol. 58 (345-363).
- Church, A. (1940): "On the concept of a random sequence". Bulletin of the American Mathematical Society, vol. 46 (130-135).
- Cifarelli, D.M.; Regazzini, E. (1982): "Some considerations about mathematical statistics". En Koch, G.; Spizzichino, F. (eds.).
- Cifarelli, D.M.; Regazzini, E. (1996): "de Finetti's contribution to probability and statistics". Statistical Science, vol. 11 (253-282).
- Cohen, G.D.; Thirring, W. (eds.)(1973): "*The Boltzmann equation*". Springer.
- Cohen, L.J. (1992): "Relevant implications andconditional assertion". En Anderson; Belnap; Dunn (eds.).
- Condorcet, M.J.A.N.C. (1783): "Réflexions sur la méthode de déterminer la probabilité des événements futurs d'aprés l'observation des événements passés". Histoire de l'Académie Royale des Sciences (539-553).
- Condorcet, M.J.A.N.C. (1784): "Mémoire sur le calcul des probabilités". Mémoires de l'Académie Royale des Sciences, Año 1781 707-720).
- Condorcet, M.J.A.N.C. (1785): "*Essai sur l'application de l'analyse à la probabilité des décisions rendues A la pluralité des voix*". París.
- Copeland, A.H. (1928): "Admissible numbers in the theory of probability". American Journal of Mathematics, vol. 53 (153-162).

- Copeland, A.H. (1937): "Consistency of the conditions determining kollektivs". Transactions of the American Mathematical Society, vol. 42 (333-357).
- Côtes, R. (1722): "*Æstimatio errorum in mixta mathesi, per variationes partium trianguli plani et sphærici*". En "*Opera miscellanea*". Cambridge.
- Cournot, A.A. (1838a): "*Recherceh̄es sur les principes mathématiques de la théorie des richesses*". París.
- Cournot, A.A. (1838b): "Mémoire sur les applications du calcul des chances à la statistique judiciaire". Journal de Mathématiques pures et appliquées". Mémoires de l'Académie Royale des Sciences, vol. 3 (257-334).
- Cournot, A.A. (1843): "*Exposition de la théorie des chances et des probabilités*". Hachette.
Cousin, V. (1841): "*Cours d'histoire de la philosophie morale audixhuitième siècle*". París.
- Cover, T.M.; Gacs, P.; Gray, R.M. (1989): "Kolmogorov's contributions to information theory and algorithmic complexity". Annals of Probability, vol. 17 (840-865).
- Craig, J. (1699): "*Theologiæ christianæ principia mathematica*", Londres. Traducción al inglés como "*Craig's rules of historical evidence*". En "*History and theory*", Beiheft 4, The Hage (1964).
- Cramér, H. (1946). "*Mathematical methods of statistics*".Princeton University Press.
- Czuber, E. (1903): "*Wahrscheinlichkeitsrechnung ind ihre Anwendung aufFehlersaugsleichung, Statistik und Lebensverscherung*". Teubner.
- Dale, A.I. (1982): "Bayes or Laplace? An examination of the origin and early applications of Bayes' theorem". Archive of History of Exact Sciences, vol. 27 (23-47).
- Dale, A.I. (1991): "*A history of inverse probability: Fron Thomas Bayes to Karl Pearson*". Springer.
- d'Alembert, J.B.L.R. (1761): "*Réflexions sur le calcul des probabilités*". Opusucules Mathématiques, vol. 2 (287-288). Reeditado en "*Oeuvres de d'Alembert*", Berlín (1821-1822).
- d'Alembert, J.B.L.R. (1767): "Doutes et questions sur le calcul des probabilités". Mélanges de Littérature d'Histoire et de Philosophie, vol. 5 (277-278).
- Daston, L.J. (1980): "Probabilistic expectation and rationality in classical probability theory". Historia Mathematica, vol. 7 (234-260).
- Daston, L.J. (1988): "*Classical Probability in the enlightenment"*. Princeton University Press.
- Daston, L. (1994): "How probabilities came to be objective and subjective". Historia Mathematica, 21 (235-254).
- Davies, M. (2000): "*The universal computer: The road from Leibniz to Turing*". W.W.Norton & Co.

- de Finetti, B. (1928a): "Sulle probabilità numerabili e geometriche". Rendiconti del Reale Instituto Lombardo di Scienze e Lettere, vol. 61 (817-824).
- de Finetti, B. (1928b): "Funzione caratteristica di un fenómeno aleatorio". Congresso Internazionale dei Matematici, Bologna. Reeditado en Atti della Reale Accademia Nazionale dei Lincei, Rendiconti, vol. 4 (86-133).
- de Finetti, B. (1929a): "Sulle funzioni ad incremento aleatorio". Rendiconti della Reale Accademia Nazionale dei Lincei, 1930, vol. 10 (163-168).
- de Finetti, B, (1929b): "Sulla possibilitá di valori eccezionali per una legge di incrementi aleatori". Rendiconti della Reale Accademia Nazionale dei Lincei, vol. 10 (325-329).
- de Finetti, B. (1930b): "Problemi determinati e indeterminati nel calcolo delle probabilità". Rendiconti della Accademia Nazionale dei Lincei, vol. 12 (367-373).
- de Finetti, B. (1930c): "Sulla proprietà conglomerativa della probabilità". Rendiconti del Reale Istituto Lombardo di Scienze e Lettere, vol. 63 (414-418).
- de Finetti, B. (1931a): "Le leggi differenziali e la rinunzia al determinismo". Rendiconti del Seminario Matematico della Reale Universitá de Roma, vol. 7 (64-74).
- de Finetti, B. (1931b): "Probabilismo". Logos, Biblioteca de Filosofía (163-219). Trad. Al inglés en Erkenntnis (1989), vol. 31 (169-223).
- de Finetti, B. (1931c): "Sul significato soggettivo della probabilità". Fundamenta Mathematicæ, vol. 17 (198-329).
- de Finetti, B. (1933a): "Classi di numeri aleatori equivalenti". Atti della Reale Accademia Nazionale dei Lincei, Rendiconti, vol. 18 (107-110).
- de Finetti, B. (1933b): "La legge dei grandi numeri nei casi dei numeri aleatori equivalenti". Atti della Reale Accademia Nazionale dei Lincei, Rendiconti, vol. 18 (203-207).
- de Finetti, B. (1933c): "Sulla legge di distribuzione dei valori in una successione di numeri aleatori equivalenti". Atti della Reale Accademia Nazionale dei Lincei, Rendiconti, vol. 18 (279-284).
- de Finetti, B. (1934): "Indipendenza stocastica ed equivalenza stocastica". Atti della Societá Italiana per il Progreso delle Scienze (XXII Reunione), vol. 2 (199-202).
-de Finetti, B. (1937): "La prevision: ses lois logiques, ses sources subjectives". Annales de l'Institut Henri Poincaré, vol. 7 (1-68).
-de Finetti, B. (1938): "Sur la condition d'equivalence partielle". Actualités Scientifiques et Industrielles, vol. 739 (5-18).
- de Finetti, B. (1949): "Sull'impostazione assiomatica del calcolo delle probabilità". Annali Triestini, vol. 19 (19-81).
- de Finetti, B. (1952): "Sulla preferibilità". Giornale degli Economisti e Annali di Economia, vol. 11 (685-709).

- de Finetti, B. (1955): “La struttura delle distribuzioni in un insieme astratto qualsiasi”. Giornale dell’Istituto Italiano degli Attuari, vol. 18 (12-28). Traducción al inglés en Probability, Induction and Statistics, 1972(129-140).
-de Finetti, B. (1969): “Sulla proseguibilità di processi aleatori scambiabili”. Rendiconti dell’Istituto di Matematica dell’Universitá de Trieste, vol. 1 (53-67).
- de Finetti, B. (1970): “*Teoria delle probabilità*”. Torino. Traducciónal inglés como “*Theory of probability*”. Cambridge University Press.
- de Finetti, B. (1972): “*Probability, induction and statistics: The art of guessing*”. Wiley.
- de Finetti, B. (1981): “*Scritti(1926-1930)*”. CEDAM.
- de Finetti, B. (1989): “*La lógica dell’incerto*”. Il Saggiatore.
- de Finetti, B. (1995): “*Filosofia della probabilità*”. Il Saggiatore. Traducción al inglés como “*Philosophical lectures on probability*”, Springer (2008).
- de Finetti, B. (2006): “*L’invenzione della verità*”. Raffaello Cortina.
- de Moivre, A. (1733): “Approximatio ad summam terminorum binomii $\overline{a+b}|^{n}$ in seriem expansi”. Reeditado en Archibald, R.C. (1926).
- de Moivre, A. (1718): “*The doctrine of chances: or a method of calculating the probabilities of events in play*”.Woodfall. Reeditado por Chelsea, 1967.
- de Moivre, A. (1730): “*Miscellanea analytica des seriebus et quadraturis*”. Tonso & Watts.
- de Montmort, P.R. (1713): “*Essai d’analyse sur les jeux de hazard*”. Quillau.Reeditado por Chelsea, 1980.
- de Morgan, A. (1838a): “*An essay on probabilities and their application to life contingencies and insurance offices*”. Longman, Orme, Brown, Green & Longmans.
- de Morgan, A. (1838b): “On a question in the theory of probabilities”. Transactions of the Cambridge Philosophical Society, vol. 6 (423-430).
- de Morgan, A. (1847): “*Formal logic: or the calculus of inference, necessary and probable*”. Taylor & Walton.
- Derham, W. (1716): “*Physico-theology or a demonstration of the being and attributes of God from his works of creation*”, Londres. Reeditado, Nueva York (1977).
- de Scheemaekere, X.; Szafarz, A. (2008): “The special status of mathematical probability: A historical sketch”. Centre Emile Bernheim., Research Institute in Management Sciences (1-14).
- d’Holbach, P.H.T. (1770): “*Système de la nature ou des lois du monde physique et du monde moral*”, Rey. Edición en inglés como “*The system of nature*”, Franklin (1970).
- Dokic, J.; Engel, P. (2001): “*Verité et success*”. Presses Iniversitaires de France. Edición en inglés como “*Frank Ramsey, truth and success*”. Routledge (2003).
- Doob, J.L. (1949): “Heuristic approach to the Kolmogorv-Smirnov theorems” Annals of Mathematical Statistics, vol. 20 (393-403).

- Doob, J.L. (1953): "*Stochastic processes*". Wiley.
- Doob, J.L. (1989): "Kolmogorov's early work on convergence theory and foundations". Annals of probability, vol. 17 (815-821).
- Doob, J.L. (1994): "The development of rigor in mathematical probability 1900-1950". En Pier (ed.).
- Drobisch, M.W. (1834): "*Beiträge zur Orientierung über Herbart's System der Philosophie*". Voss.
- Drobisch, M.W. (1867): "*Die moralische Statistik und die Menschliche Willensfreiheit*". Voss.
- Dubins, L.E. (1975): "Finitely additive conditional probabilities, conglomerability and disintegration". Annals of Probability, vol. 3 (89-99).
- Dugué, D. (1954): "Eléménts limite stochastique". Proceedings 28th Session ISI, Roma (60-71).
- Dugué, D. (1957): "Sur la convergence stochastique au sens de Césaro et sur des différences importantes entre la convergence presque certain et les convergences en probabilité et presque completes". Sankhya, vol. 18 (127-138).
- Dvoretzky, A. (1971): "The central limit problem for dependent random variables". Actes du Congres International des Mathématiciens, Niza, vol. 2 (565-570).
- Dvoretzky, A. (1972): "Asymptotic normality for sums of dependent random variables". Proceedings 6th Berkeley Symposium on Mathematical Statistics, vol. 40 (513-535).
- Dynkin, E.B. (1978): "Sufficient statistics and extreme points". Annals of Probability, vol. 32 (272-284).
- Einstein, A. (1905): "Über die vonder molekularkinetischen Theorie der Wärme geforderte Bewegung von in Flüssigkeiten suspendierten Teilchen". Annalen der Physik, vol. 17 (549-560).
- Einstein, A. (1906): "Zur Theorie der Brownschen Bewegung". Annalen der Physik, vol. 19 (371-381).
- Einstein, A. (1907): "Theoritische Bemerkungen über die Brownsche Bewegung". Zeitschrift für Elektrochemie, vol. 13 (41-42).
- Einstein, A. (1908): "Elementare Theorie der Beownschen Bewegung". Zeitschrift für Elektrochemie, vol. 14 (235-239).
- Ellis, R.L. (1849): "On the foundations of the theory of probabilities". Transactions of the Cambridge Philosophical Society, vol. 8 (1-6).
- Emerson, W. (1766): "*A miscellaneous treatise; containing several mathematical subjects*". Nourse.
- Engelman, R. (ed.)(1876): "*Abhandlungen von Friederich Bessel*". Q. Engelman.
- Erdös, P. (1942): "On the law of iterated logarithm". Annals of Mathematics, vol. 43 (419-436).
- Erdös, P. (1949): "On a theoremof Hsu and Robbins". Annals of Mathematical Statistics, vol. 20 (286-291).

- Erdös, P.; Rényi, A. (1959): "On Cantor's series with convergent $\sum(1/q^n)$". Acta Univ. Sci. Budapestiensis, Sec. Mathematica, vol. 2 (93-109).
- Etemandi, N. (1981): "An elementary proof of the strong law of large numbers". Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, vol. 55 (119-122).
- Faber, G. (1910): "Über stetige Funktionen". Mathematische Annalen, vol. 69 (372-443).
- Fechner, G.T. (1849a): "*Beriche über die Verhandlungen der Königlich sächsisten Gesellschaft der Wissen scahften zu Leipzig*". Weidmann.
- Fechner, G.T. (1849b): "DieMathematische Behandlung organischer Gestalten uns Processe". Offentliche Sitzung Am 18.
- Fechner, G.T. (1851): "*Zend-Avesta, oder über die des Himmels und des Jenseits*". Leopold Voss. Tercera edición Leopold Voss (1906).
- Fechner, G.T. (1860): "*Elemente der Psychophysik*". Breitkopf & Härtel. Edición en inglés por Holt-Rinehart-Winston (1966).
- Fechner, G.T. (1871): "Zur experimentalen Ästhetik". Abhandlungen der Königlich sächsischen Gesellschaft der Wissenschaften, vol. 9 (555-635).
- Fechner, G.T. (1897): "*Kollektivemasslehre*". Q. Engelmann.
- Feller, W. (1943): "The general form of the so-called law of the iterated logarithm". Transactions of the American Mathematical Society, vol. 54 (373-402).
- Feller, W. (1946): "The law of the iterated logarithm for identically distributed random variables". Annals of Mathematics, vol. 47 (631-638).
- Feller, W. (1968): "*An introduction to probability theory and its applications*". Wiley.
- Festa, R. (1996): "*Cambiare opinione. Temi e problem di epistemologia bayesiana*". CLUEB.
- Festa, R. (1999): "Bayesian confirmation". En Galavotti, M.C.; Pagnini, A. (eds.)
- Fetzer, J.H. (1982): "Probabilistic explanation". PSA, vol. 2 (194-207).
- Fetzer, J.H. (1988): "Probabilistic metaphysics". En Fetzer (ed.), "*Probability and causality: Essaysin honour of Wesley C. Salmon*", Dodrecht-Reidel.
- Fetzer, J.H. (1993): "Peirce and propensities". En Moore (ed.).
- Fine, T.L. (1973): "*Theories of probability: an examination of foundations"*. Academic Press.
- Fisher, R.A. (1926): "*The mathematical theory of probability and its applications to frequency curves and statistical methods*". Oliver & Boyd.
- Fisher, R.A. (1956): "*Statistical methods and scientific inference*". Oliver & Boyd.
- Fortini, S.; Ruggieri, F. (1994): "On defining neighbourhoods of measures through the concentration function". Sankhya, Serie A, vol. 56 (444-457).
- Franklin, J.W. (2001): "*The science of conjecture: Evidence and probability before Pascal*". The John Hopkins University.
- Franzén, T. (2005): "*Gödel's theorem. An incomplete guide to its use an abuse*".Peters.

- Fréchet, M. (1930a): "Sur l'extension du théorème des probabilités totals au cas d'une suite infinite d'événements". Rendiconti del Reale Istituto Lombardo di Scienze et Lettere, vol. 63 (1059-1062).
- Fréchet, M. (1930b): "Sur le convergence en probabilité". Metron, vol. 8 (1-48).
- Fréchet, M. (1939): "*Les probabilities asociées à un systéme d'événements compatibles*". Herman.
- Fréchet, M. (1952): "Méthode des fonctions arbitraires, théorie des événements en chaine dans le cas d'un nombre fini d'états possible". Gauthier-Villars.
- Freedman, D.A. (1962): "Invariance under mixing wich generalize de Finetti's theorem". Annals of Mathematics and Statistics, vol. 33 (916-923).
- Freedman, D.A. (1963): "Invariance under mixing which generalize de Finetti's theorem: Continuous time parameter". Annals of Mathematics and Statistics, vol. 34 (1194-1216).
- Freeman, P.R.; Smith, A.F.M. (eds.(1994): "*Aspects on uncertainty: Atribute to D.V. Lindley*". Wiley.
- Fry, T.C. (1928): "*Probability and its engineering uses*", Van Nostrand.
- Fürst, D. (1982): "de Finetti: A scientist man". En Koch; Spizzichino (eds.).
- Galavotti, M.C. (1989): "Anti-reaslism in the philosophy of probability: Bruno de Finetti's subjectivism". Erkenntnis, vol. 31 (239-261).
- Galavotti, M.C. (1991a): "Sulla filosofiadella probabilità di Bruno de Finetti". Rivista di Filosofia, vol. 83 (121-138).
- Galavotti, M.C. (1991b): "The notion of subjective probability in the work of Ramsey and de Finetti". Theoria, vol. 57 (239-259).
- Galavotti, M.C. (1994): "F.P. Ramsey and the notion of 'chance'". Proceedings of the 17th International Wittgenstein-Symposium (330-340).
- Galavotti, M.C. (2005): "*Philosophical introduction to probability*". CSLI publications, LELAND Stanford Junior University.
- Galavotti, M.C.; Pagnini, A. (eds.)(1999): "*Experience, reality and scientific explanation*". Kluwer.
- Galilei, G. (1632): "*Dialogo sui massimi sistemi del mondo*". En "*Opere*", editado por A. Favaro, Florencia (1968).
- Gänsler, P.; Strobel, J.; Sute, W. (1978): "On central limit theorems or martingale triangular arrays". Acta Mathematica Academyof Sciences of Hungary, vol. 31 (205-216).
- Gassendi, P. (1659): "*Syntagma*". Londres.
- Gauss, C.F. (1809): "*Theoria motus corporum cœlestiorum sectionibus conicis solum ambientium*". Pertes et Bresser. Traducción al inglés como "*Theory of motion of the heavenly bodies movingabout the sun in conics sections*". Little-Brown.Reeditado por Dover (1963).
- Gerhardt, C.I. (ed.)(1962): "*G. Leibniz Mathematische Schriften*", G. Olms.

- Gigerenzer, G.; Swijtink, Z.; Porter, T.; Daston, L.; Beatty, J.; Krüger, L. (1989): "*The empire of chance*". Cambridge University Press.
- Gillies, D. A.(1972a): "Operationalism" Synthèse, vol. 25 (1-24).
- Gillies, D.A. (1972b): "The subjective theory of probability". British Journal for the Philosophy of Science, vol. 29 (138-157).
- Gillies, D.A. (1973): "*An objective theory of probability*". Methuen.
- Gillies, D.A. (1990): "Bayesianism versus falsificationism. Review of Howson and Urbach 1989". Ratio, New Series III, vol. 1 (82-98).
- Gillies, D.A. (2000): "*Philosophical theories of probability*". Routledge.
- Gnedenko, B.V. (1950): "*The theory of probability*". Original en ruso.Traducción al inglés por Chelsea (1962).
- Gnedenko, B.V.; Kolmogorov, A.N. (1954): "*Limit distributions for sums of independent random variables*". Addison-Wesley.
- Gouraud, M.C.Ch. (1848): "*Histoire du calcul des probabilityés, depuis ses origines jusq'à nos jours*". Durand.
- Graunt, J. (1662): "*Natural and political observations mentioned in a following index and made upon the bills of mortality*". Londres.
'sGravesande, G.J. (1712): "*Démonstration mathématique du soin que dieu prend de diriger ce qui se passé dans ce monde, tirée du nombre des garçons et des filles qui naissent journellement*". En "*Oeuvres philosophiques et matématiques de Mr. G.J.'sGravesande*". Marc Michel Rey (1774).
- Grimmet, G.; Stirzaker, D. (1982): "*Probability and random processes*". Clarendon Press.
- Haag, J. (1924): "Sur un problème général des probabilités et ses diverses applications". Proceedings of the International Congress of Mathematics, University of Toronto.
- Hacking, I. (1965): "*Logic statistical inference*". Cambridge University Press.
- Hacohen, M.H. (2000) : « *Karl Popper : The formative years 1902-1945* ». Cambridge University Press.
- Hadamard, J. (1922): "Les principes du calcul des probabilités". Revue de Métaphysique et de Morale, vol. 39 (114-126).
- Hadamard, J. (1927): "Sur le battage des cartes". Comptes Rendus, vol. 185 (5-9).
- Hadamard, J. (1928): "Sur le battage des cartes et ses relations avecla mécanique statistique". Atti del Congreso Internazionale dei Matematici, vol. 5 (133-139).
- Hald, A. (1984): "A. De Moivre: '*De mensura sortis*' or '*On the measurement of chance*'". International Statistical Review, vol. 52 (229-236).
- Hall, P.;Heyde, C.C. (1980): "*Martingale limit theory and its application*". Academic Press.

- Halley, E. (1693): "An estimate on the degrees of mortality of mankind, drawn from curious tables of the birth and funerals at the city of Breslaw, with an attempt to ascertain the price of annuities upon lives" Philosophical Transactions of the Royal Society, vol. 17 (596-610).
- Hammond, N. (ed.)(2003): "*The Cambridge companion to Pascal*". Cambridge University Press.
- Hardy, G.H.; Littlewood, J.E.; Pólya, G. (1934): "*Inequalities*". Cambridge University Press.
- Hartley, D. (1749): "*Observations on man, his frame, his duty and his expectations*". Londres.
- Hartmann, P.; Winter, A. (1941): "On the law of iterated logarithm". American Journal of Mathematics, vol. 63 (169-176).
- Hausdorff, F. (1914): "*Grundzüge der Mengenlehre*",Veit. Reeditado por Chelsea (1955).
- Heath, D.; Sudderth, W. (1978): "On finitely additive priors, coherence andextended admissibility". Annals of Statistics, vol 6 (335-345).
- Heidelberger, M. (1987): "Fechner's indeterminism: From freedom to laws of chance". En Krüger; Daston; Heidelberger (eds.).
- Heidelberger, M. (2001): "Origins of the logical theory of probability: von Kries, Wittgenstein, Waismann". International Studies in the Philosophy of Science, vol. 15 (177-188).
- Helm, G. (1902): "Die Wahscheinlichkeitslehre als Theorie der Kollektivbegriffe". Annalen der Naturphilosophie, vol. 1 (364-381).
- Heyde, C.C. (1968a): "On almost sure convergence for sums of independent random variables". SankhyA, Serie A, vol. 30 (353-358).
- Heyde, C.C. (1968b): "On the converse to the iterated logarithm law". Journal of Applied Probability, vol. 5 (210-215).
- Heyde, C.C. (1985): "On some probabilistic developments of significance to statistics: Martingales, long range dependence, fractals, and random fields". En Atkinson; Fienberg (eds.).
- Heyde, C.C.; Seneta, E. (1977): "*I.J. Bienaymé: statistical theory anticipated*". Springer.
- Heyting, A. (1956): "*Intuitionism: An introduction*". North-Holland.
- Hilbert, D. (1899): "*Grundlagen der Geometrie*". Teubner.
- Hilbert, D.; Ackermann, W. (1928): "*Grundzüge der theoretischen Logik*". Springer.
- Hill, B.M. (1980a): "Conglomerability and countable additivity". Sankhya, Serie A, vol. 47 (366-379).
- Hill, B.M. (1980): "On some statistical paradoxes and non-conglomerability". En Bernardo, J.M.; De Groot, M.H.; Lindley, D.V.; Smith, A.F.M. (eds.).

- Hill, B.M. (1994): "On steinian shrinkage estimators: the finite/infinite problem and formalism in probability and statistics". En Freeman, P.R.; Smith, A.F.M. (eds.).
- Hille, B.M.; Lane, D. (1985): "Conglomerability and countable additivity". Sankhya, Serie A, vol. 47 (366-379).
- Hintikka, J.; Suppes, P. (eds.)(1966): "*Aspects of inductive logic*". North-Holland (265-316).
- Hooper, G. (1699): "A calculation of the credibility of human testimony". Philosophical Transactions of the Royal Society, vol. 21 (359-365).
- Hopf, E. (1932): "On the time average theorems in dynamics". Proceedings of the National Academy of Sciences, vol. 18 (93-100).
- Hopf, E. (1934): "On causality, statistics and probability". Journal of Mathematics and Physics, vol. 13 (51-102).
- Hopf, E. (1936): "Über die Bedeutung der willkürlichen Funktionen für die Wahrscheinlichkeitstheorie". Vereinigung, vol. 46 (179-195).
- Hopf, E. (1937): "*Ergodentheorie*". Springer.
- Horn, A.; Tarsky, A. (1948): "Measures on Boolean algebras". Transactions of the American Mathematical Society, vol. 64 (467-497).
- Hostinsky, B. (1928): "Sur les probabilités relatives aux transformations répétées". Comptes Rendus, vol. 186 (59-61).
- Hostinsky, B. (1929): "Sur la théorie générale des phénomenes de la diffusion". Comptes Rendus du I Congrés des Mathématiciens des Pays Slaves (341-347).
- Hostinsky, B. (1930): "Diskussion über Wahrscheinlichkeit". Erkenntnis, vol. 1 (184-185).
- Hostinsky, B. (1931): "*Méthodes générales du calcul des probabilités*". Gauthier-Villars.
- Howard, R.A. (1971): "*Dynamics probabilities systems*". Wiley.
-Hsu, P.L.; Robbins, H. (1947): "Complete convergence and the law of large numbers". Proceedings of the National Academy of Sciences, vol. 33 (25-31).
- Hume, D. (1739): "*A treatise of human nature, being an attempt to introduce the experimental method of reasoning into moral subjects, and dialogues concerning natural religion*". Green & Grose.Versión en inglés en Clarendon Press, 1978.
- Ibragimov, I.A. (1963): "A central limit theorem for a class of dependent random variables". Theory of Probability and Applications, vol. 8 (83-89).
- Ibragimov, V.Z.; Linnik, Y.V. (1971): "*Independent and stationary sequences of random variables*". Nordhoff.
- Iosifescu, M. (1968): "La loi du logarithme itéré pour une clase de variables aléatoires dépendant". Teor. Verojonost i Premenen, vol. 13 (315-325).
- Johnson, W.E. (1932): "Probability: The deductive and inductive problems". Mind, vol. 49 (409-423).
- Kac, M. (1959): "*Statistical independence in probability analysis and number theory*". Wiley.

- Kamlah, A. (1983): "Probability as a quasi-theoretical concept: J.V. Kries'sophisticated account after a century". Erkenntnis, vol. 19 (239-251).
- Kamlah, A. (1987): "The declive of the Laplacian theory of probability: A study of Stumpf, von Kries and Meinong". En Krüger, Daston, Heidelberger (eds.).
- Keynes, J.M. (1921): "*A treatise on probability*". MacMillan.
- Khinchin, A.Y. (1923): "Über dyadische Brüche". Mathematische Zeitschrift, vol. 18 (109-116).
- Khinchin, A.Y. (1924): "Über einen Satz der Wahrscheinlichkeitsrechnung". Fundamenta Mathematicae, vol. 6 (9-20).
- Khinchin, A.Y. (1929a): "Sur la lois des grandes nombres". Comptes Rendus Académie des Sciences, vol. 188.
- Khinchin, A.Y. (1929b): "von Mises theory of probability and the principles of physical statistics". Versión en ruso, Uspkhy Fisicheski Nauk, vol. 9 (141-166).
- Khinchin, A.Y. (1932a): "Sur les classes d'événements equivalents". Matemati□eskii Sbornik, vol. 39 (40-43).
- Khinchin, A.Y. (1932b): ""Remarques sur les suites d'événements obeisant à la loi des grands nombres". Matemati□eskii Sbornik,, vol. 39 (115-119).
- Khinchin, A.Y. (1932c): "Zur Birkhoffs Lösung der Ergodenproblems". Mathematische Annalen, vol. 107 (485-488).
- Khinchin, A.Y. (1932d): "Sulle sucessioni stazionarie di eventi". Giornale dell'Istituto Italiano degli Attuari, vol. 3 (267-272).
- Khinchin, A.Y. (1933): "Zur mathematischen Begründung der statistichen Mechanik". Zeitschrift für Angewandte Mathematik und Mechanik, vol. 13 (101-103).
- Khinchin, A.Y. (1938): "The correlation theory of stationary processes". Uspekhi Matematicheskikh Nauk, vol. 5 (42-51) (original en ruso).
- Khinchin, A.Y. (1950): "Über dyadische brüche". Mathematische Zeitschrift, vol. 18 (109-116).
- Khinchin, A.Y. (1961a): "R. Mises frequentist theory and contemporary ideas in probability theory". Original en ruso Voprosy Filosofii, vol. 1 (92-102), vol. 2 (77-89).
- Khinchin, A.Y.; Kolmogorov, A.N. (1925): "Über Konvergenz von reihen deren Glieder durch den Zufall bestimmt warden". Matemati□eskii Sbornik,vol. 32 (668-676).
- Knobloch, E. (1987): "Emile Borel as a probabilist". En Krüger; Daston; Heidelberger (eds.).
- Knuth, D.E. (1985): "Algoritmic thinking and mathematical thinking". American Mathematical Monthly, vol. 92 (170-181).
- Koch, G.; Spizzichino, F. (eds.)(1982): "*Exchangeability in probability and statistics*". North-Holland.

- Kolmogorov, A.N. (1928): “Über die Summen durch den Zufall bestimmter unabhängiger Grösse”. Mathematische Annalen, vol. 99 (309-319).
- Kolmogorov, A.N. (1929a): “The general theory of measure and the calculus of probability”. Communist Academy, Section of Natural and Exact Sciences. Collected Works of the Mathematics Section, vol. 1 (8-21) (original en ruso).
- Kolmogorov, A.N. (1929b): “Über das Gesetz des iterierten Logarithmus”. Mathematische Annalen, vol. 101, (126-135).
-Kolmogorov, A.N. (1930): “Sur la loi forte des grandes nombres”. Comptes Rendus Académie des Sciences, vol. 19 (910-912).
- Kolmogorov, A.N. (1931): “Über die analitischen Methoden in der Wahrscheinlichkeitsechnung”. Mathematische Annalen, vol. 104 (415-458).
- Kolmogorov, A.N. (1932a): “Sulla forma generale di un proceso stochastico omogeneo (Un problema di Bruno de Finetti)”. Atti della Reale Accademia Nazionale dei Lincei, vol. 15 (805-808).
- Kolmogorov, A.N. (1932b): “Ancora sulla forma generale di un proceso stochastico omogeneo”. Atti della Reale Accademia Nazionale dei Lincei, vol. 15 (866-869).
- Kolmogorov, A.N. (1933a): “Sulla determinazione empirica di una legge di distribuzione”. Giornale Istituto Italiano degli Attuari, vol. 4 (83-91).
- Kolmogorov, A.N. (1933b): “*Grundbegriffe der Wahrscheinlichkeitsrechnung”*. Springer.Traducción al inglés "*Foundations of the theory of probability*".Chelsea, 1933.
- Kolmogorov, A.N. (1936): “Zur Theorie der markoffschen Ketten”. Mathematische Annalen, vol. 112 (155-160).
- Kolmogorov, A.N. (1937): “Ein vereinfachter Beweis der Birkoff-Khintchineschen Ergodensatzes”. MathemtiCeskii Sbornik, vol. 2 (367-368). Traducido al inglés como “A simplified proof of the Birkhoff-Khinchin ergodic theorem”. Uspekhi Matematicheskikh Nauk, vol. 5 (52-56).
- Kolmogorov, A.N. (1963a): “The theory of probability”. En Aleksandrov; Kolmogorov; Laurent’ev (eds.) (229-264). MIT Press (original en ruso, 1956).
- Kolmogorov, A.N. (1963b): “On tables of random numbers”. Sankhya, vol. 25 (369-376).
- Kolmogorov, A.N. (1965): “Three approaches to the quantitative definition of information”. Problems of Information Transmission, vol. 1 (1-7).
- Kolmogorov, A.N. (1968): “Logical basis of information theory and probability theory”. IEEE Transactions on Information Theory, vol. 14 (662-664).
- Kolmogorov, A.N. (1983a): “Combinatorial basis of information theory and probability theory”. Russian Mathematical Surveys, vol. 38 (29-40).
- Kolmogorov, A.N. (1983b): “On logical foundations of probability theory”. Probability Theory and Mathematical Statistics, Lecture Notes in Mathematics, vol. 1021 (1-5).

- Kolmogorov, A.N. (1986): "Prefacio a '*The law of large numbers*' de J. Bernoulli". Nauka (original en ruso).
- Koopman, B. (1931): "Hamiltonian systems and transformations in Hilbert space". Proceedings of the National Academy of Sciences, vol. 17 (315-318).
- Krüger, L.; Daston, L.J.; Heidelberger, M. (eds.) (1987): "*The probabilistic revolution"*. Vol. I. MIT Press.
- Krüger, L.; Gigerenzer, G.; Morgan, M.S. (eds.) (1990): "*The probabilistic revolution"*. Vol. II, MIT Press.
- - Kuipers, L.; Niederreiter, H. (1974): "*Uniform distribution of sequences*". Wiley.
- Kuipers, T.A. (2000): "*From instrumentalism to constructive empiricism*". Kluwer.
- Kyburg, H.E.; Smokler, H.E. (1980): "*Studies in subjective probability*", Wiley.
- Lacroix, S.F. (1816): "*Traité élémentaire du calcul des probabilités*". Courcier.
- Lad, F. (1996): "*Operational subjetive statistical methods: A mathematical, philosophical and historical introduction*". Wiley.
- Lambert, J.H. (1765): "*Beyträge zum Gebrauche der Mathematik und deren Anwendung*". Berlín.
- Landro, A.H. (23010): "Acerca del Regellosigkeitsaxiom de von Mises". Cuadernos del CIMBAGE, vol. 12 (1-21).
- Laplace, P.S. (1774a): "Mémoire sur les suites récurro-récurrentes et sur leurs usages dans la théorie des hasards". Mémoires de l'Académie Royale des Sciences, vol. 8 (5-24).
- Laplace, P.S. (1774b): "Mémoire sur la probabilité des causes par les événements". Mémoires de l'Académie Royale des Sciences, vol. 6 (621-656).
- Laplace, P.S. (1781): "Mémoire sur les probabilités". Mémoires de l'Académie des Sciences (227-232).
- Laplace, P.S. (1786): "Sur les naissances, les mariages et les morts a Paris, depuis 1771 jusqu'en 1781, et dans tout l'étendue de la France, pendant les anées 1781 et 1782". Mémoires de l'Académie Royale des Sciences, año 1783 (693-702).
- Laplace, P.S. (1802): "Mémoire sur les approximations des formules qui sont fonctions de tres grands nombres et sur leur application aux probabilités". Mémoires de l'Académie Royale des Sciences, año 1809 (353-415; 559-565).
- Laplace, P.S. (1812): "*Théorie analytique des probabilités*". Courcier.
- Laplace, P.S. (1814): "*Essai philosophique sur les probabilités*". En "*Oeuvres completes*". Académie des Sciences. Traducción al inglés como "*A philosophical essay on probabilities*". Nueva York (1951).
- Leibniz, G.W. (1686): "*Mathematische Schriften*", Gerhardt. Reeditado por Hildsheim (1855).
- Lévy, P. (1925): "*Calcul des probabilités*". Gauthier-Villars.
- Lévy, P. (1929): "Sur les lois de probabilité dont dépendent les quotients complets et incomplets d'une fraction continue". Bulletin de la Societé Mathématique de France, vol. 57 (178-194).

- Lévy, P. (1931): "Sur un théorème de M. Khinchine". Bulletin des Sciences Mathématiques, vol. 55 (145-160).
- Lévy, P. (1935): "Proprietés asymptotiques des sommes de variables aléatoires enchainées". Bulletin des Sciences Mthématiques, vol. 59 (Serie 2)(84-96/109-128).
- Lévy, P. (1937,1954): "*Théory de l'addition des variables aléatoires*". Gauthier-Villars.
- Lexis, W. (1876): "Das Gesachlechtsverhältnis der Geborenen und die Wahrscheinlichkeitsrechnung". Jahrbücher für Nationalökonomie und Statistik, vol. 27 (209-245).
- Lindeberg, J.W. (1922): "Eine neue Harleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung". Mathematische Zeitschrift, vol. 15 (211-225).
- Lipps, G.F. (1898): "ÜberFechner Collectivmasslehre un vie Vertheilungsetze der Collectivgegestände". Philosophische Studien, vol. 13 (579-612).
- Lipps, G.E. (1901): "Die Theorie der collectivgegestände". Philosophische Studien, vol. 17 (78-183).
- Lipps, G.E. (1905): "Die Betimmung der Abhängigkeit zwischen den Merkmalen eines Gegestándes". Berchte der Könglich Säsischen Gesellschaft der Wissenschaften, vol. 57 (1-32).
- Locke, J. (1689): "*An essay concerning human understanding*", Londres. Reeditado por Dover (1959).
- Loève, M. (1977): "*Probability theory*". Springer.
- Lomnicki, A. (1923): "Noveaux fondements du calcul des probabilités". Fundamenta Mathematica, vol. 4 (34-71).
- Loveland, J. (2001): "Buffon, the certainty of sunrise, and the probabilistic reduction ad absurdum". Archive of History of Exact Sciences, vol. 55 (89-110).
- Loynes, R.M. (1969): "The central limit theorem for backwards matingales". Zeitschrift Wahrscheinlichkeits Verwandte Gebiete. Vol. 13 (1-8).
- Loynes, R.M. (1970): "An invariance prtinciple for reversed matingales". Proceedings of the American Mathematical Society, vol. 25 (56-64).
- Lyapunov, A.M. (1901a): "Nouvelle forme du théorème sur la limit de probabilité". Notes Academy of Sciences Physics-Mathematics de Saint Petersburg, Serie 8, vol. 12 (1-24).
- Lyapunov, A.M. (1901b): "Sur un théorème du calcul des probabilités". Comptes Rendus Academie des Sciences, vol. 132 (126-128).
- Lyapunov, A.M. (1901c): "Une proposition générale du calcul des probabilités". Comptes Rendus Académie des Sciences, vol. 132 (814-815).
- Mach, E. (1833): "*Die Mechanik in Ihrer Entwicklung*". Leipzig. Traducción al inglés como"*The science of mechanics: Acritical and historical account of its development*", La Salle (1960). Traducción al castellano como "*Desarrollo histórico-crítico de la mecánica*", Espasa-Calpe (1960).

- Maistrov, L.E. (1974): "*Probability theory: A historical sketch*". Original en ruso (1967). Edición en ingles, Academic Press.
- Mandelbrot, B. (1960): "The Pareto-Lévy law and the distribution of income". International Economic Review, vol. 1 (79-106). Reeditado en Wood; McLure (eds.)(1999) y en Blaug (2002).
- Mandelbrot, B.; Taylor, H. (1967): "On the distribution of stock prices differences". Operations Research, vol. 15 (1057-1062).
- Marcinkiewicz, J.; Zygmund, A. (1937): "Sur les fonctions indépendantes". Fundamenta Mathematica, vol. 29 (60-90).
- Markov, A.A. (1898): "Sur les racines de l'equation$\frac{e^{X^2}\partial^m e^{X^2}}{\partial X^m} = 0$",
- Markov, A.A. (1907): "Extension of the law of large numbers to dependent variables". Physical-Mathematical Society, Kazan University, serie 2, vol. 15 (135-156).
- Markov, A.A. (1908): "Extension of limit theorems of probability theory to a sum of variables connected in a chain". En Howard (ed.)(1971)(552-576). Original en ruso enAcademy Sciences St. Petersburg, Serie 8, vol. 22. Traducción al alemán como "Ausdehnung der Satze über die Granzwerte in der Wahrscheinlichkeitsrechnung auf eine Summe verketter Grossen". En Markov (1912) (212-298).
- Markov, A.A. (1910): "On connected variables not forming a genuine chain". Izvestiya Akademii Nauk., poFiziko-Matematisches Komu Otdeleniyu, vol. 25 (original en ruso).
- Markov, A.A. (1911b): "On a case of trials connected into a compound chain". Izvestiya Akademii Nauk., SPb, vol. 5 (171-186) (original en ruso).
- Markov, A.A. (1912a): "On trials connected in a chain from unobservable events". Izvestiya Akademii Nauk., vol. 6 (551-572)(original en ruso).
- Markov, A.A. (1912b): "*Wahrscheinlichkeitsrechnung*". Teuber.
- Markov, A.A. (1924): "*The calculus of probabilities*". Moscú.
- Markov, A.A. (1951): "*Selected works*". Academia de Ciencias USRR.
- Martin-Löf, P. (1966): "The definition of random sequences". Information and Control, vol. 9 (602-619).
- Martin-Löf, P. (1969): "The literatura on von Mises collectives revisited". Theoria, vol. 35 (12-37).
- Martin-Löf, P. (1971): "Complexity oscillations in infinite binary sequences". Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, vol. 19 (225-230).
- McLeish, D.L. (1975): "A maximal inequality and dependent strong laws". Annals of Probabiity, vol. 3 (829-839).
- McLeish, D.L. (1977): "On the invariance principle for non-stationary mixingales". Annals of Probability, vol. 5 (616-621).
- Meinong, A. (1915): "*Über Möglichkeit und Eahrscheinlichkeit*". Anbrosius Barth.
- Mellor, D.H. (1971): "*The matter of chance*". Cambridge University Press.

- Menchoff, D. (1923): “Sur les seriesdes fonctions ortogonales, I”. Fundamenta Mathematica, vol. 4 (82-105).
- Miller, D.W. (1994): “*Critical rationalism. A restatement and defence*”. Open Court.
- Miller, D.W. (1996): “Propensities and indeterminism”. En O’Hear (ed.).
- Mondadori, M. (1979): “Induzione e statistica”. En “*Enciclopedia*” (Einaudi), vol. 7 (384-430).
- Mondadori, M.; Morini, S. (1982): “Probabilità” En “*Enciclopedia*” (Einaudi), vol. 15 (498-510).
- Moore, E.C. (ed.)(1993): “*Charles Peirce and the philosophy of science: Papers from the Harvard Sesquicentennial Conference*”. Tuscalosa: University of Alabama Press.
- Morini, S. (1992): “Bruno de Finetti: L’origine de son subjectivism”. Cahiers du Centre d’Analyse et des Mathématiques Sociales. Accesible e www.rescogitans.it.
- Mura, A. (1995): “Probabilità soggetiva e non contraddittorietà”. En de Finetti (1995).
- Murray, F.H. (1930): “Note on a scholium of Bayes”. Bulletin of the American Mathematical Society, vol. 36 (129-132).
- Nelson, E. (1987): “*Radically elementary probability theory*”, Princeton University Press.
- Newton, I. (1730): “*Optiks*”. Reeditado en Nueva York (1952).
- Nieuwentyt, B. (1715): “*Het Regt gebruik der wereldeschouwingen*”. Traducido al inglés como “*The religious philosopher: or, the right use of contemplating the works of the Creator*” , Londres.
- Niven, I. (1956): “*Irrational numbers*”. Wiley.
- O’Hear, A. (ed.)(1996): “*Karl Popper: Philosophy and problems*”. Cambridge University Press.
- Omelyanoskij, M.E.; Fock, V.A. (eds.)(1972): “*L’interpretazione materialistica della mecanica quantistica*”. Feltrinelli.
- Pearson, K. (1892): “*The grammar of science*”. Walter Scott.
- Pearson, K. (1893a): “Asymetrical frequency curves”. Nature, vol. 48 (20-35).
- Pearson, K. (1893b): “Contribution to the mathematical theory of evolution, I”. Proceedings of the Royal Society, Serie A, vol. 54 (211-242).
- Pearson, K. (1895): “Contribution to the mathematical theory of evolution, II”. Proceedings of the Royal Society, Serie A, vol. 186 (343-414).
- Pearson, K. (1896): “Contribution to the mathematical theory of evolution, III”. Proceedings of the Royal Society, Serie A, vol. 187 (174-227).
- Pearson, K. (1924): “Historical note on the origin of the normal curve of errors”. Biometrika, vol. 16 (5402-404).
- Pearson, K. (1925): “James Bernoulli theorem”. Biometrika, vol. 17 (201-210).

- Pearson, K.; Kendall, M. (eds.)(1970): "*Studies in the history of statistics and probability*", vol. I. Charles Grffin.
- Peirce, C.S. (1910): "*Notes on the doctrine of chances*". Reeditado en "*Essays in the philosophy of science*". The American Heritage Series, Bobbs-Merril (1957).
- Petrov, V.V. (1974): "*Sums of independent random variables*". Springer.
- Philipp, W. (1967): "Das Gesetz von iterierten Logarithmus fúr stark mischende stationare Prozese". Zeitschrift Wahrsecheinlichkeitstheorie und Verwandte Gebiete, vol. 8 (204-209).
- Pier, J.P. (ed.)(1994): "*Development of mathematics 1900-1950*". Birkhäuser.
- Poincaré, H. (1896): "*Calcul des probabilités"*. Gauthier-Villars.
- Poincaré, H. (1902): "*La science et l'hypothèse"*. Flammarion.
- Poincaré, H. (1905): "*La valeur de la science"*. Flammarion.
- Poincaré, H. (1908): "*Science et method"*. Flammarion.
- Poisson, S.D. (1827): "Sur la probabilité des résultat moyens des observations". Connaisance des temps pour l'anne 1827 (273-302).
- Poisson, S.D. (1830): "Mémoire sur la proportion des naissances des filles et des garçons". Mémoire de l'Académie des Sciences, vol. 9 (239-308).
- Poisson, S.D. (1832): "Suite du mémoir sur la probabilité des résultat moyens des observations, inséré dans la Connaisance des Temps del'anée 1827". Connaisance des temps pour l'année 1832 (3-22).
- Poisson, S.D. (1836a): "Note sur le calcul des probabilités". Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences, vol. 2 (395-400).
- Poisson, S.D. (1836b): "Note sur la loi des grandes nombres". Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences, vol. 2 (377-380).
- Poisson, S.D. (1837): "*Recherches sur la probabilité des jugements en matière criminelle et en matière civile, précedées des règles générales du calcul des probabilités*". Bachelier.
- Pólya, G. (1920): "Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung und das Momentenproblem". Mathematische Zeitschrift, vol. 8 (171-181).
- Popper, K.R. (1934): "*Logik der Forschung*". Springer. Traducción al inglés como "*The logic of scientific discovery*", Hutchinson (1972)
- Popper, K.R. (1957): "Probability magic or knowledge out of ignorance". Dialectica, vol. 11 (354-374).
- Popper, K.R. (1959): "The propensity interpretation of probability". British Journal for the Philosophy of Science, vol. 10 (25-42).
- Prevost, P.; Lhuilier, S.A.J. (1799): "Mémoire sur l'art d'estimer la probabilité des causes par les effects". Mémoire de l'Académie Royale des Sciences et Belles Lettres, Classe de Philosophie Spéculative, año 1796 (3-25).

- Prokhorov, Y.V. (1956): "Convergence of random processes and limit theorems of probability theory". Theory of Probability and its Applications. vol. 1 (157-214) (original en ruso: Teor. Verojatnost Primenenija, vol. 1 (177-258), 1956).
- Quetelet, A. (1835): "*Sur l'homme et le développement de ses facultés*". Bachelier.
- Quetelet, A. (1848): "Sur la statistique morale et les principes qui doivent en former la base". Mémoires de l'Académie Royale des Sciences et Belles Lettres de la Belgique, vol. 21.
- Rademacher, H. (1922): "Einige Satze über Reihen von allgemeinen orthogonal Funktion". Mathematische Annalen, vol. 87 (112-138).
- Ramsey, F. (1926): "*Truth and probability*". Harvard University Press.
- Ramsey, F. (1991): "Notes on philosophy, probability and mathematics". En Galavotti, M.C. (ed.). Bibliopolis.
- Rao, M.M. (1984): "*Probability theory with applications*". Academic Press.
- Reichenbach, H. (1915): "*Der Begriff de Wahrscheinlichkeit für die mathematische Darstellung der Wirklichkeit*". Leipzig.
- Reichenbach, H. (1932): "Axiomatik der Wahrscheinlichkeitsrechnung". Mathematische Zeitschrift, vol. 34 (568-619).
- Reichenbach, H. (1935): "*Wahrscheinlichkeitslehre*". Sijthoff's.
- Rényi, A. (1962): "Wahrscheinlichkeitsrechnung deutscher Verlag der Wissenschafteng"
- Rényi, A. (1970): "*Probability theory*". North-Holland.
- Révész, P. (1968): "*The laws of large numbers*". Academic Press.
- Reznik, M.K. (1968): "The law of the iterated logarithm for some classes of stationary processes". Theory of Probabilityand its Applications, vol. 8 (606-621).
- Roeper, P.; Leblanc, H. (1999): "*Probability theory and probability logic*". University of Toronte Press.
- Rosén, B. (1967a): "On asymptotic normality in sums of dependent random variables". Zeitschrift für Wahrscheinlichkeits Theorie und Verwandte Gebiete, vol. 7 (95-102).
- Rosén, B. (1967b): "On the central limit theorem for sums of dependent random variables". Zeitschrift für Wahrscheinlichkeits Theorie und Verwandte Gebiete, vol. 7 (103-115).
- Rümelin, G. (1863): "*Zur Theorie der Statistik*". En Reden und Aufsátse, Friburgo (1875).
- Ryll.Nardzewski, C. (1957): "On stationary sequences of random variables and the de Finetti equivalences". Colloquia Mathematica, vol. 4 (149-156).
- Savage, L.J. (1954): "*The foundations of statistic*". Wiley.
- Savage, L.J. (1973): "Probability in science: A personality account". En Suppes (ed.).

- Schervisch, M.J.; Seidenfeld, T.; Kadane, J.B.: "The extent of non-conglomerability of finitely additive probabilities". Zeitschrift für Wahrscheinlichkeits Theorie und Verwandte Gebiete, vol. 66 (205-226).
- Schnorr, C.P. (1971): "Zufälligskeit und Wahrscheinlichkeit: Eine algorithmische Behandlung der Wahrscheinlichkeitstheorie". Lectures Notes in Mathematics, vol. 218, Springer.
- Schafer, G. (1996): "The significance of Jakob Bernoulli's 'Ars conjectandi' for the philosophy of probability today". Journal of Econometrics, vol. 75 (15-32).
- Scozzafava, R. (1984): "A survey of some common misunderstanding concerning the role and meaning of finitely additive probabilities in statistical inference". Statistica, vol. 44 (21-45).
- Shannon, C.E. (1948): "A mathematical theory of communication". Bell Systems Technical Journal, vol. 27 (379-423).
- Shea, W. (2003): "*Designing experiments and games of chance*". Reidel.
- Sheynin, O.B. (1968): "On the early history of the law of large numbers". Biometrika, vol.55, n1 3 (459-467).
- Shimony, A. (1955): "Coherence and the axioms of confirmation". Journal of Symbolic Logic, vol. 20 (1-27).
- Shiryaev, A.N. (1981): "Martingales: Recent developments, results and applications". International Statistical Review, vol. 49 (199-233).
- Shiryaev, A.N. (1989): "Kolmogorov: life and creative activities". Annals of Probability, vol. 17 (866-944).
- Simpson, Th. (1740): "The nature of laws of chance", Edward Cave. Reeditado en 1792.
- Simpson, Th. (1755): "On the advantage of taking the mean of a number of observations in practical astronomy. A letter to the Righ Honorable George Earl of Macclesfield, President of the Royal Society". Philosophical Transactions of the Royal Society, vol. 49 (82-93).
- Simpson, Th. (1757): "*Miscellanous tracts on some curious and very interesting subjets in mechanics, physicalastronomy and speculative mathematics*". J. Nourse.
- Sinai, Y. (1973): "Ergodic theory". En Cohen; Thirring (eds.).
- Sinai, Y. (1976): "*Introduction to ergodic theory*". Princeton University Press.
- Slutsky, E. (1925): "Über stochastisch Asymptoten und Grezenzwerte". Metron, vol. 5 (3-89).
- Sobczyk, A.; Hammer, P.C. (1944a): "A descomposition of additive set functions". Duke Mathematical Journal, vol. 11 (839-846).
- Sobczyk, A.; Hammer, P.C. (1944b): "The ranges of additive set functions". Duke Mathematical Journal, vol. 11 (847-851).
- Solomonoff, R.J. (1964): "A formal theory of inductive inference I". Information and Control, vol. 7 (1-22).

- Solomonoff, R.J. (1964): "A formal theory of inductive inference II". Information and Control, vol. 7 (224-254).
- Steiger, W.L.; Zaremba, S.K. (1972): "The converse of the Hartman-Winter theorem". Zeitschrift Wahrsecheinlichkeitstheorie und Verwandte Gebiete, vol. 22 (193-194).
- Steinhaus, H. (1923): "Les probabilités dénombrables et leur rapport a la théorie de la mesure". Fundamenta Mathematica, vol. 4 (286-310).
- Stigler, S.M. (1985): "John Craig and the probability of history: From the death of Christ to the birth of Laplace". Technical Report n1 165, Department of Statistics, University of Chicago.
- Stigler, S.M. (1986a): "*The history of statistics: The measurement of uncertainty before 1900*". Harvard University Press.
- Stigler, S.M. (1986b): "Laplace's 1774 memoire on inverse probability". Statistical Science, vol. 1 (359-378).
- Stirling, J. (1730): "*Methodus differentialis: sive tractatus de summation et interpolation serierum infinitarium*". Bouyer.
- Stout, W.F. (1970a): "The Hartman-White law of the iterated logarithm for martingales". Annnals of Mathematical Statistics, vol. 41 (2158-2160).
- Stout, W.F. (1970b): "A martingale analogue of Kolmogorov's law of the iterated logarithm". Zeitschrift Wahrscheinlichkeitstheorie und Verwandte Gebiete, vol. 15 (279-290).
- Stout, W.F. (1974): "*Almost sure convergence*". Academic Press.
- Strassen, V. (1964): "An invariance principle for the law of the iterated logarithm". Zeitschrift Wahrscheinlichkeitstheorie und Verwandte Gebiete, vol. 3 (211-226).
- Strassen, V. (1965): "Almost sure behavior of sums of independent random variables and martingales": Proceedings 5th Berkely Symposium of Mathematicas Statistics and Probability, vol. 2 (315-344).
- Strassen V. (1966): "A converse to the law of the iterated logarithm". Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete, vol. 4 (265-268).
- Suppes, P. (ed.)(1973): "*Logic, methodology and philosophy of science*". North-Holland.
- Süsmilch, J.P. (1741): "*Die göttliche Ordnung in der Veränderungen des menslichen Geschlechs*". Reeditado como "*L'ordre divin aux origins de la démographie*". Institut National d'Études Démographiques (1979)
- ter Hark, M. (2002): "Between autobiography and reality: Popper's inductive years". Studies in History and Philosophy of Science, vol. 33 (79-103).
- ter Hark, M. (2004): "*Popper, Otto Selz and the rise of evolutionary epistemology*". Cambridge University Press.
- Thom, R. (1975): "*Structural stability and morphogenesis*". Reading.
- Todhunter, I. (1865): "*A history of the mathematical theory of probability from the time of Pascal to that of Laplace*". Cambridge

- Trembley, J. (1795-1798): “De probabilitate causarum ab effectibus oriunda: Disquisitio mathematica”. Commentationes Societatis Regiæ Scientiarum Gottingensis, Commentationes Mathematica, vol. 13 (64-119).
- Turing, A. (1937): “On computable numbers, with an application to the Entscheindungsproblem”. Proceedings London Mathematical Society, Serie 2, vol. 42 (230-265).
- Uspensky, J.V. (1937): “*Introduction to mathematical probability*”, McGraw-Hill. Traducido como “*Matemática de las probabilidades*”, Nigar (1954).
- van Deuren, P. (1934): “*Leçons sur le calcul des probabilités*”. Gauthier-Villars.
- van Vleck, E. (1908): “On non-mensurable sets of points, with an example”. Transactions of the American Mathematical Society, vol. 9 (238-244).
- Venn, J. (1866): “*The logic of chance*”. MacMillan.
- Ville, J. (1939): “*Étude critique de la notion de collectif*”. Gauthier-Villars.
- von Kries, J. (1886): “*Die Principien der Wahrscheinlichkeits-rechnung*”. Mohr.
- von Mises, R. (1912): “Über die Grundbegriffe der Kollektivmasslehre”. Jahresberichte der Deutschen Mathematiker Vereinigung, vol. 21 (9-20).
- von Mises, R. (1919a): “Fundamentalsatze der Wahrscheinlichkeitsrechnung”. Mathematische Zeitschrift, vol. 4 (1-97).
- von Mises, R. (1919b): “Grundlagen der Wahrscheinlichkeitsrechnung”. Mathematische Zeitschrift, vol. 4 (1-97).
Mathematische Zeitschrift, vol. 5 (52-99).
- von Mises, R. (1920): “Ausschaltung der Ergodenhypothese aus der physikalischen Statistik”. Physikalische Zeitschrift, vol. 21 (225-232;256-262).
- von Mises, R. (1920): “*Wahrscheinlichkeit, Statistik und Wahrheit*”, Springer. Traducción al inglés como“*Probability, statistics and truth*”. MacMillan (1928).
- von Mises, R. (1930): “Über kausale und statistiche Gesetzmässigkeit in der Physik”. Die Naturwissenschaften, vol. 18 (145-153).
-von Mises, R. (1931a): “*Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und theoretischen Physik”*. Rosenberg.
- von Mises, R. (1931b): “Über die gegenwärtige Krise der Mechanik”. Zeitschrift für Angewandte Mathematik und Mechanik, vol. 1 (425-431).
- von Mises, R. (1939): “*Kleines lehrbuch des Positivismus*”. Den Haag.
- von Mises, R. (1964a): “*Mathematical theory of probability and statistics*”. Academic Press.
- von Mises, R. (1964b): “*Selected papers of Richard von Mises*”. American Mathematical Society.
- von Neumann, J. (1932a): “A proof of the quasi-ergodic hypothesis”. Proceedings of the National Academy of Sciences, vol. 18 (70-82).
- von Neumann, J. (1932b): “Zur Operationenmethode in der klassichen Mechanik”. Annals of Mathematics, vol. 33 (587-642).

- von Neumann, J. (1932c): "Physical applications of the ergodic hypothesis". Proceedings of the National Academy of Sciences, vol. 18 (263-266).
- von Plato, J. (1990): "Probabilistic physics: The classical way". En Krüger; Gigerenzer; Morgan (1990).
- von Plato, J. (1994): "*Creating modern probability*". Cambridge University Press.
- von Smoluchowski, M. (1918): "Über den Begriff des Zufalls und den Ursprung der Wahrscheinlichkeitsgesetze in der Physik". Die Naturwissenschaften, vol. 6 (253-263).
- Wald, A. (1937): "Die Widerspruchsfrenheit des Kollektivbegriffes in der Wahrscheinlichkeits-rechnung". Ergebnisse eines Mathematischen Kolloquiums, vol. 8 (38-72).
- Weaver, W. (1982): "*Lady luck*". Dover.
- White, H. (1984): "*Asymptotic theory for econometricians*". Academis Press.
- Wiener, N. (1921): "The average af an analytic functional and the Brownian movement". Proceedings of the National Academy of Sciences, vol. 7 (294-298).
- Wiman, A. (1900): "Über eine Wahrscheinlichkeitsaufgabe bei kettebrucentwicklungen". Öfverseghr af Kongliga Svenska Vatenskaps-Akademiens Förhandlingar Femtiondesjunde Ärgängen, vol. 57 (829-841).
- Wiman, A. (1901): "Bemerkung über eine von Glydén Aufgeworfene Wahrscheinlichkeitsfrage". HÅkan Ohlssons Boktrykeri.
- Wood, J.C.; McLure, M. (eds.)(1999): "*Vilfredo Pareto critical assessments*". Routledge.
- Zvinkin, A.K.; Levin, L.A. (1970): "The complexity of finite objects and the development of the concepts of information and randomness by means of the theory of algorithms". Russian Mathematical Surveys, vl. 25 (83-124).

Índice Temático

A

B

C

"*Explicit hoc totum;*
Pro Christo da mihi potum"

(de un escriba anónimo de comienzos
del siglo XV)

COLECCIÓN: EL NÚMERO DE ORO
Director: *Act. Alberto Landro*

Álgebra para estudiantes de Ciencias Económicas
Alejandro E. García Venturini – Axel Kicillof

Análisis Matemático I para estudiantes de Ciencias Económicas
Alejandro E. García Venturini – Axel Kicillof

Análisis Matemático II para estudiantes de Ciencias Económicas
Alejandro E. García Venturini – Axel Kicillof

Los matemáticos que hicieron la historia
Alejandro E. García Venturini

Análisis de Series de Tiempo, univariadas y multivariadas
Heriberto Urbisaia – Juana Brufman

Decisión Estadística Bayesiana, a modo de introducción
Emma Fernández Loureiro de Pérez

Estadística no Paramétrica, a modo de introducción
Emma Fernández Loureiro de Pérez

Teoría de los Conjuntos Borrosos, a modo de introducción
Emma Fernández Loureiro de Pérez

Estadística: Herramientas de Inferencia
Gabriela Kurincic

Estadística: Probabilidades y Distribuciones
Gabriela Kurincic

Los Métodos Cuantitativos en las Ciencias Sociales
Alejandro E. García Venturini – Federico Castelli

Aplicaciones del Análisis Matemático a la Economía
Blanca R. Vitale

Modelos para el Análisis de Series de Tiempo
Juan Carlos Abril

Análisis Matemático I para estudiantes de Ingeniería
Alejandro E. García Venturini – Mónica Scardigli

Cálculo Financiero
Juan R. Garnica Hervás - Esteban O. Thomasz - Romina P. Garófalo

Elementos de Econometría de los fenómenos dinámicos
Alberto H. Landro – Mirta L. González

Acerca de la probabilidad – Parte I
Alberto H. Landro – Mirta L. González

Análisis Matemático II para estudiantes de Ingeniería
Alejandro E. García Venturini

Algebra ya Geometría analítica I para estudiantes de Ingeniería
Alejandro E. García Venturini

Acerca del problema de Bernoulli y la determinación del verdadero valor de una probabilidad
Alberto H. Landro – Mirta L. González