Lecturas Matemáticas
Volumen **37** (1) (2016), páginas 37-61
ISSN 0120-1980

Borel y Steinhaus: dos piezas claves para comprender el surgimiento de la probabilidad moderna

Borel and Steinhaus: two key players in understanding the origins of modern probability

Carmen Martínez-Adame* y Jesús Caballero

Universidad Nacional Autónoma de México

RESUMEN. Los orígenes de la teoría matemática de la probabilidad se pueden rastrear trabajos del siglo XVI. Durante el siglo XIX el desarrollo de la teoría clásica de la probabilidad tuvo un gran auge. Sin embargo, es usual que se diga que la teoría moderna de la probabilidad tiene su verdadero origen en el año 1933, con el trabajo de ANDREI KOLMOGOROV. El objetivo de este artículo es mostrar que los trabajos previos de BOREL y STEINHAUS sobre la teoría de la medida, las probabilidades numerables y la fundamentación de la probabilidad son claves para el desarrollo de la teoría moderna.

***Palabras clave*:** historia de la probabilidad, teoría de la medida, comprensión matemática.

ABSTRACT. The origins of the mathematical theory of probability can be traced back to the XVI Century and during the XIX Century classical probability theory boomed. However it is usual to say that modern probability theory has its true origin in the work of ANDREI KOLMOGOROV published in 1933. The goal of this article is to show that the previous work of BOREL and STEINHAUS on measure theory, denumerable probabilities and the foundations of probability are key to the development of the modern theory.

***Key words*:** History of probability, measure theory, mathematical understanding.

2010 AMS Mathematics Subject Classification. 01A60, 28-03, 60-03, 60A05

*Investigación realizada gracias al Programa UNAM-DGAPA-PAPIIT IN403816: La comprensión matemática.

1. Introducción

La historia de la teoría matemática de la probabilidad es amplia, compleja y abarca muchos siglos; sin embargo, en este artículo nuestro interés es presentar un estudio de un periodo relativamente breve de tiempo: 1905-1923. Mostraremos que fue en ese periodo que la probabilidad se transformó en una teoría moderna y que esta transformación tiene sus orígenes en la obra de ÉMILE BOREL, y culmina con la obra de HUGO STEINHAUS.

El interés de BOREL por la teoría de la probabilidad se debe en gran parte a los grandes tratados de BERTRAND y POINCARÉ, ambos llamados *Calcul des probabilités*, publicados en 1889 y 1896, respectivamente,[1] aunque como veremos más adelante, la teoría que desarrolla BOREL es radicalmente distinta a la que aparece en estos tratados.

Durante el siglo XIX, la teoría de la probabilidad parecía seguir el camino que habían marcado los trabajos de LEGENDRE, GAUSS y LAPLACE sobre el método de mínimos cuadrados y la distribución normal. Sin embargo, la aparición de la teoría de la medida y la integral de Lebesgue, en los albores del siglo XX, facilitó un gran cambio en la dirección de la teoría.

BOREL hace una presentación de la probabilidad muy diferente de la de sus antecesores, y consideramos que marca el inicio de la teoría moderna de la probabilidad, a pesar de los comentarios del propio BOREL, en [3, p. 126], donde dice que existían trabajos previos ya en la misma dirección:

> Les méthodes de M. LEBESGUE permettent donc d'étudier des questions de probabilité qui paraissent inaccessibles par les procé-dés d'intégration classiques. D'ailleurs, dans les cas particuliers les plus simples, il suffira de se servir de la théorie des ensembles que j'avais appelés mesurables [...] l'application de cette théorie des ensembles mesurables au calcul des probabilités a été, à ma connaissance, faite pour la première fois par M. WIMAN.[2]

WIMAN a su vez, en [23], sostiene que este es el camino que se debía seguir:

> We are quite decidedly of the opinion that if one should want to develop probability theory in the sense of the modern theory of sets, one should above all make use of the Borelian notion of content.[3]

Una vez cimentadas las bases de la nueva teoría de la probabilidad por BOREL, se escribieron muchos textos sobre el tema; sin embargo, no sería sino hasta 1923, con el texto de STEINHAUS, cuando se presente un sistema formal de axiomas vinculando ya de manera permanente a la probabilidad con la teoría de la medida.

Este es el camino que nos interesa investigar en el presente artículo, y por tanto consideramos importante presentar de manera breve el contexto en el cual estaba sumergido el mundo matemático en esa época. Creemos que es de suma relevancia tomar en cuenta el

[1] Véanse [2] y [18].
[2] El trabajo de ANDERS WIMAN al que hace referencia BOREL es [22].
[3] Traducción al inglés de JAN VON PLATO en [17, pp. 31-32].

Congreso Mundial de 1900, en el cual DAVID HILBERT planteó sus famosos 23 problemas,[4] que dictarían el rumbo de la matemática del siglo XX. En particular, el sexto problema resulta muy relevante para el tema que tratamos aquí [11, p. 454]:

> 6. Mathematical Treatment of the Axioms of Physics.
>
> The investigations on the foundations of geometry suggest the problem: To treat in the same manner, by means of axioms, those physical sciences in which mathematics plays an important part; in the first rank are the theory of probabilities and mechanics.
>
> As to the axioms of the theory of probabilities, it seems to me desirable that their logical investigation should be accompanied by a rigorous and satisfactory development of the method of mean values in mathematical physics, and in particular in the kinetic theory of gases.

Con relativa prontitud aparecieron dos respuestas a este problema.[5] La primera fue de RUDOLF LAEMMEL, en 1904, quien intentó fundamentar la probabilidad principalmente en la teoría de conjuntos, y de manera secundaria en la teoría de la medida.

La otra axiomatización la daría en 1907 un discípulo del propio HILBERT, llamado UGO BROGGI. Para esta axiomatización, BROGGI recurrió a las teorías de la medida de BOREL y de LEBESGUE. Se propuso un sistema siguiendo el modelo de la *Geometría* de HILBERT para probar consistencia, completitud e independencia entre los axiomas. Siguiendo el concepto introducido formalmente en la matemática poco antes por ZERMELO, BROGGI sostiene que se puede decidir si un elemento en un conjunto tiene la propiedad A o simplemente no la tiene: utilizando el "hecho" de que se puede escoger una $m \in M$, siendo M un conjunto cualquiera, sin decir nada sobre las propiedades de m, y propone ver a la probabilidad como la medida del conjunto de las m que cumplen alguna propiedad específica. Es importante notar, sin embargo, que sus axiomas son generales:

1. La probabilidad es una función no negativa.
2. La probabilidad de la certeza es uno.
3. Se cumple la propiedad aditiva, es decir, la suma de la probabilidad del conjunto A (de las m que cumplen cierta propiedad) y la del conjunto B (de las m que cumplen otra propiedad), siendo A y B ajenos entre sí, es la probabilidad de $A \cup B$.

BROGGI asume que la aditividad numerable sería consecuencia del tercer axioma, y afirma que la axiomatización está completa, que ninguna otra clasificación de axiomas la completará y que extensiones iguales serán equiprobables, muy posiblemente basándose en la idea de medida. Con respecto a esta axiomatización, VON PLATO asegura en [17, p. 33] que "the categoricity property becomes the same as the uniqueness of Lebesgue measure, it seems". Para el caso en que M sea numerable, comenta que se deben ordenar

[4] Aunque en realidad en su conferencia, por cuestiones de tiempo, solo presentó 10, y la lista completa de 23 fue publicada poco después en [10] y [11].

[5] Pueden consultarse en [19].

sus elementos primero, y después define la probabilidad como el límite, asumiendo que este existe, de las frecuencias relativas en cada uno de los subconjuntos de la cadena creciente que se forma al ordenar los elementos de M.[6]

Ahora bien, es en este contexto de nuevas ideas en la teoría de la probabilidad dentro del contexto delineado por HILBERT, en el cual se presenta el trabajo de BOREL.

2. El trabajo de Émile Borel

Se puede rastrear el interés desarrollado por BOREL por la teoría de la probabilidad al año 1905 aproximadamente, y KNOBLOCH describe de manera puntual este interés en [14, p. 251]:

> In addition to many textbooks, BOREL published more than fifty papers between 1905 and 1950 on the calculus of probability. They were mainly motivated or influenced by POINCARÉ, BERTRAND, REICHENBACH, and KEYNES [...] He preferred to elucidate [...] applications instead of looking for an axiomatization of probability theory. Its essential peculiarities were for him unpredictability, indeterminism, and discontinuity. Nevertheless, he was interested in a clarification of the probability concept.

BOREL, dentro de esta visión, dejó el problema de la axiomatización abierto para que años después STEINHAUS y KOLMOGOROV lo atacaran.

2.1. 1905: *Remarques sur certains questiones de probabilité*

En 1905 BOREL publica su primer artículo sobre probabilidad, titulado *Remarques sur certains questiones de probabilité* [3], en el cual considera la probabilidad como proporcional a una extensión, ya sea una longitud, un área, un volumen, etc., y por tanto la definición de probabilidad se presenta como el valor medio de una función, es decir, si $f : [x_0, x_1] \subsetneq \mathbb{R} \to \mathbb{R}$, entonces $\frac{1}{x_1 - x_0} \int_{x_0}^{x_1} f(x)dx$ es el valor buscado.

BOREL toma un número entre cero y uno y se pregunta por la probabilidad de que dicho número sea conmensurable.[7] Para contestar su pregunta considera la función $f : [0, 1] \subset \mathbb{R} \to \mathbb{R}$

$$f(x) = \begin{cases} 1 & \text{si } x \text{ es conmensurable} \\ 0 & \text{si } x \text{ es inconmensurable} \end{cases}$$

que es equivalente a la pregunta planteada cuando se toma uno en el caso favorable y cero en el caso contrario. De esta manera, el autor se convence de que la respuesta a su pregunta es cero. Hace notar que para que la integral de la función tenga sentido es necesario utilizar los métodos introducidos por LEBESGUE, puesto que en el sentido clásico sería imposible estudiar esta integral.

[6] Esta idea de tomar como probabilidad un conjunto con cierta propiedad, o medir el conjunto de pertenencia, no era nueva; años antes, en la física ya la usaban MAXWELL y BOLTZMANN.

[7] Para BOREL, un número conmensurable (con la unidad) es un número racional. La nomenclatura nos parece apropiada, puesto que podrá definir posteriormente otro tipo de conmensurabilidad, y por tanto la conservamos.

Una vez hecha esta precisión, BOREL presenta brevemente los conceptos principales de su teoría de la medida, que será de utilidad para el problema de probabilidad que atacará.

El problema consiste en considerar $x \in [0,1]$ y un conjunto medible $E \subseteq [0,1]$ de medida m. Por definición, la probabilidad de que x esté en E es m, y de que no esté es $1-m$. Si $E = \mathbb{Q} \cap [0,1]$,[8] se obtiene de inmediato que $m = 0$.

Posteriormente BOREL trata un caso más general. Define que un número $\alpha \in [0,1]$ es *conmensurable de grado n con la unidad* si existen dos números p y q que sean primos relativos, $(p,q) = 1$, y tales que

$$\left|\frac{p}{q} - \alpha\right| < \frac{1}{q^n},\ q \geq 2,\ p < q. \tag{1}$$

BOREL se plantea la pregunta de calcular la probabilidad de que $x \in [0,1]$ sea conmensurable de grado *n* con la unidad, y es en la respuesta a esta pregunta en donde se puede ver por primera vez la técnica que ha desarrollado.

A cada fracción irreductible de la forma $\frac{p}{q}$, BOREL le asocia el intervalo

$$E_{p,q}^{(n)} = \left(\frac{p}{q} - \frac{1}{q^n}, \frac{p}{q} + \frac{1}{q^n}\right)$$

de longitud $\frac{2}{q^n}$, y hace notar que la longitud total de estos intervalos para una q dada será $2\frac{\phi(q)}{q^n}$, donde $\phi(q)$ denota el número de primos relativos menores que q. Es fácil ver que los intervalos $E_{p,q}^{(n)}$ para q fija y p variando entre los primos relativos menores que q, no se intersecan, y que cada uno de ellos tiene longitud $\frac{2}{q^n}$. Por tanto, el conjunto $E^{(n)}$ de puntos distintos contenidos en los conjuntos $E_{p,q}^{(n)}$, es decir,

$$E^{(n)} = \bigcup_{q=2}^{\infty} \bigcup_{\{p:(p,q)=1, p\leq q\}} E_{p,q}^{(n)}$$

es medible, y su medida, e_n, es a lo más $m_n = 2\sum_{q=2}^{\infty} \frac{\phi(q)}{q^n}$, que es la suma de las medidas de todos los intervalos.

Dicha serie es convergente para $n > 2$, y para $n = 2$, se sabe que cualquier número α cumple de una infinidad de maneras la desigualdad (1). Tómese entonces $n \geq 3$. La serie m_n es convergente, y más aún a un valor menor que uno, puesto que

$$\phi(q) \leq q - 1$$

y

$$2\sum_{q=2}^{\infty} \frac{q-1}{q^3} = \frac{\pi^2}{3}\left(1 - \frac{\pi}{4}\right) < 1.^9$$

[8] La notación es nuestra, y consideramos que facilita la comprensión del trabajo realizado por BOREL.

[9] En realidad $2\sum_{q=2}^{\infty} \frac{q-1}{q^3} = \frac{\pi^2}{3} - 2\zeta(3)$ y por tanto la igualdad que presenta BOREL debe entenderse como una aproximación ya que lo es que importante es que la cantidad sea menor que 1.

BOREL comenta que si convenimos que a cualquier α inconmensurable que cumpla la desigualdad (1) para varios sistemas de valores p y q, se le asigna ese número por coeficiente, entonces podremos decir que la probabilidad de que x sea conmensurable de grado n con la unidad es precisamente m_n. Sin embargo, si no hacemos la convención anterior, la probabilidad será $e_n < m_n$. En este caso, BOREL afirma que se puede calcular e_n con tanta precisión como se quiera, y finalmente concluye que la probabilidad de que x no sea conmensurable de grado *n* con la unidad es $1 - e_n$.

2.2. 1909: *Les probabilités dénombrables et leurs applications artithmétiques*

Entre 1905 y 1909 BOREL escribió varios textos relacionados con la teoría de la probabilidad, y después de 1909 escribiría muchos más. No obstante, como lo comentamos en la introducción, el estudio en el presente artículo se limita a los textos de 1905 y 1909, puesto que es en ellos en donde se observa un origen de la probabilidad moderna como resultado de la utilización de la teoría de la medida.

El artículo de 1909 comienza con la introducción del concepto de probabilidad numerable que debe ocupar un lugar entre las probabilidades discretas y las continuas:

> On distingue généralment, dans le problèmes de probabilités, deux catégories principales, suivant que le nombre des cas possibles est fini ou infini: la première catégorie constitue ce qu'on appelle les *probabilités discontinues* [...] tandis que la seconde catégorie comprend les *probabilités continues* [...] Une telle classification apparaît comme incomplète, lorsque l'on se reporte aux résultats acquis dans la théorie des ensambles; entre la puissance des ensembles finis et la puissance du continu se place la puissance des ensambles dénombrables; je me propose de montrer brièvement l'intérêt qui s'attache aux questions de probabilités dans l'énoncé desquelles interviennent de tels ensembles; je les appellerai, pour abréger, *probabilités dénombrables*. [4, p. 247]

Después de explicar al lector la importancia del caso numerable, BOREL clasifica las probabilidades numerables en tres categorías:

1. El número de casos posibles es finito en cada evento, pero el número de pruebas es numerable.[10]
2. El número de casos posibles es numerable y el número de pruebas es finito.
3. El número de casos posibles y el número de pruebas es numerable.

Esta clasificación no se basa en razones lógicas, ya que las tres categorías no se distinguen entre sí de esta manera; es una clasificación práctica para facilitar el cálculo.

El primer caso del que se ocupa BOREL es de la primera categoría, y esto lo hace en la sección 4 de [4]. Supone que solo habrá dos casos posibles, a saber, el favorable y el desfavorable, y enumera los eventos conforme a los números naturales, denotando la probabilidad del caso favorable en el evento n por p_n, y por q_n la probabilidad del caso

[10] Entendiendo *prueba* como *ensayo*.

desfavorable, también en el evento n. Establecido esto, BOREL plantea los problemas por estudiar: los primeros tres pertenecen a la primera categoría, los problemas 4 y 5 pertenecen a la segunda categoría, y el problema 6 a la última.

Problema 1. ¿Cuál es la probabilidad de que el caso favorable jamás se produzca?

Se nombra esta probabilidad A_0, y por el principio de probabilidades compuestas que el propio BOREL había definido ya en [5, p. 27], se tiene que

$$A_0 = (1-p_1)(1-p_2)\ldots(1-p_n)\cdots = \prod_{i=1}^{\infty}(1-p_i). \tag{2}$$

BOREL demuestra que el producto numerable es convergente haciendo ver que $\prod_{n=1}^{\infty}(1-p_n) \leq e^{-\sum_{n=1}^{\infty} p_n}$, y que el lado derecho de la desigualdad tiende a 0.

Algo que nos parece importante señalar en la resolución de este problema es que BOREL utiliza de manera implícita el hecho de que la medida de probabilidad es continua, puesto que asume que la probabilidad en el caso numerable se puede obtener como límite del caso finito.

Problema 2. ¿Cuál es la probabilidad de que el caso favorable se produzca exactamente k veces?

En este caso, BOREL denota con A_k la probabilidad buscada, y nombra ω_1 la probabilidad de que el evento favorable solo se produzca en la primera prueba; esta será

$$\omega_1 = p_1(1-p_2)(1-p_3)\cdots(1-p_j)\cdots \tag{3}$$

En el caso de que $\sum_{i=1}^{\infty} p_i$ sea divergente, (3) es igual a cero, y en el caso de que sea convergente se tiene $\omega_1 = \frac{p_1}{1-p_1}A_0$ por (2). De la misma manera, se tiene que ω_n, definido como $\omega_n = (1-p_1)(1-p_2)\cdots(1-p_{n-1})p_n(1-p_{n+1})\cdots$, sería la probabilidad de que el evento favorable se produzca exactamente en la prueba n-ésima, sea igual a cero en caso de que $\sum_{i=1}^{\infty} p_i$ sea divergente. En conclusión: $\omega_n = \frac{p_n}{1-p_n}A_0$ en el caso convergente.

Por el principio de probabilidad total, BOREL obtiene en el caso de convergencia que $A_1 = A_0\left(\frac{p_1}{1-p_1} + \frac{p_2}{1-p_2} + \cdots + \frac{p_n}{1-p_n} + \cdots\right)$, y como la serie que multiplica a A_0 es claramente convergente, BOREL define $u_n = \frac{p_n}{1-p_n}$ y obtiene

$$A_1 = A_0 \cdot \sum_n u_n. \tag{4}$$

Para extender esto último al caso divergente, BOREL señala que se debe proceder con cuidado, pues $\sum_n u_n$ será divergente, $A_0 = 0$ y se tendrá $\infty \cdot 0$. Por otro lado si vemos a A_1 como la suma de las ω_n, las cuales son todas nulas, se tendrá una suma no finita de ceros, que BOREL advierte que no se debe tomar a la ligera: no desea suponer, o quizá no está convencido de la aditividad numerable de probabilidades nulas.

De manera análoga, A_2 será cero si $\sum_{i=1}^{\infty} p_i$ es divergente, y para el caso convergente se tiene que

$$
\begin{aligned}
A_2 &= A_0(u_1u_2 + u_1u_2 + u_3 + \cdots + u_2u_3 + u_2u_4 + \cdots + u_3u_4 + u_3u_5 + \dots) \\
&= A_0 \sum_{i<j} u_i u_j
\end{aligned}
$$

Finalmente, concluye que A_k será cero cuando $\sum_{i=1}^{\infty} p_i$ sea divergente, y en el caso convergente se tendrá, de manera análoga, que

$$A_k = A_0 \cdot \sum_{n_i} u_{n_1} u_{n_2} u_{n_3} \cdots u_{n_k}. \tag{5}$$

Problema 3. ¿Cuál es la probabilidad de que el evento favorable se produzca una infinidad de veces?

BOREL denota esta probabilidad como A_∞ y considera la suma $S = A_0 + A_1 + \cdots + A_n + \dots$, que denota al evento favorable. Por (4) y (5) se puede reescribir a S como $S = A_0(1+u_1)(1+u_2)\cdots(1+u_n)\cdots$, producto que será convergente si $\sum_{i=1}^{\infty} p_i$ converge. Pero es importante notar que $S = A_0 \cdot \prod_n (1+u_n) = A_0 \cdot \prod_n \left(\frac{1}{1-p_n}\right) = A_0 \cdot \frac{1}{A_0} = 1$.

La probabilidad buscada es $A_\infty = 1 - S$, es decir, $A_\infty = 0$. BOREL hace hincapié sobre el hecho notable de que el resultado sea independiente de toda hipótesis sobre la frecuencia con la cual se produce la infinidad de casos favorables.

Sin embargo, BOREL añade que si se designa como $\phi(h)$ al rango de la prueba en la cual se produce el caso favorable h, entonces la función $\phi(h)$ puede tener un crecimiento tan rápido como sea deseado. Y este es un ejemplo en el cual se muestra que el hecho de que la probabilidad sea cero no es equivalente con la imposibilidad absoluta.

Ahora, en caso que $\sum_{i=1}^{\infty} p_i$ sea divergente, cada una de las A_k será cero, y se podrá ver que $S = 0$, y por tanto $A_\infty = 1$. BOREL argumenta que aunque este resultado es exacto, el razonamiento carece de rigor por las mismas razones vigentes en el caso anterior.

Con este problema BOREL concluye los casos de la primera categoría y adiciona: para estos casos se tiene que cuando la serie $\sum_{i=1}^{\infty} p_i$ es convergente, las probabilidades $A_0, A_1, \dots$ son valores bien determinados no nulos, y la probabilidad $A_\infty = 0$; a diferencia de ello, cuando la serie $\sum_{i=1}^{\infty} p_i$ no converja, las probabilidades A_k son todas nulas y $A_\infty = 1$.

Para los problemas de la segunda categoría BOREL nombra $p_1, p_2, \dots, p_n, \dots$ a las probabilidades de los diversos eventos posibles, y hace notar que la serie de términos positivos $\sum_{i=1}^{\infty} p_i$ no solo es convergente, sino que es igual a uno.

Sobre este caso BOREL aclara que para él no es absurdo pensar que en algunos casos una infinidad de probabilidades nulas podrían conducir a una suma finita; sin embargo, aunque tal hipótesis no le parece absurda desde un punto de vista lógico, la descartó categóricamente al no encontrar circunstancias en las que pareciera ventajoso incluirla. Por esta razón advierte que para usar los siguientes procedimientos se debe asegurar que no se está en este singular caso.

Por otro lado, BOREL comenta que lo que hace difíciles, pero a la vez interesantes los problemas de esta categoría, es que las probabilidades p_i son raramente conocidas con precisión. Por esta razón BOREL recomienda seguir un método indicado por POINCARÉ para las probabilidades continuas: buscar qué conclusiones generales se pueden obtener con el mínimo de hipótesis sobre las p_i.

Problema 4. ¿Cuál es la probabilidad de que en m pruebas sucesivas se obtengan resultados distintos en todas?

Sea T_m la probabilidad buscada. Se tiene que

$$T_m = m! \sum_{n_1<n_2<\ldots<n_m} p_{n_1}p_{n_2}\cdots p_{n_m}. \tag{6}$$

Para estimar el valor de T_m, BOREL define la función entera

$$F(z) = (1+p_1z)(1+p_2z)\cdots(1+p_nz)\cdots = \prod_{n=1}^{\infty}(1+p_nz), \tag{7}$$

que es comparable asintóticamente con $\cos(\sqrt{z})$, pues ambas funciones son de orden $\frac{1}{2}$, y por tanto obtiene que asintóticamente $\frac{T_m}{m!} = \frac{1}{(2m)!}$.[11] Finalmente obtiene una aproximación para T_m mediante la estimación de Stirling:

$$T_m = \frac{m^m e^{-m}\sqrt{2\pi m}}{2^{2m}m^{2m}e^{-2m}\sqrt{4\pi m}} = \frac{1}{\sqrt{2}}\left(\frac{e}{4m}\right)^m.$$

Problema 5. ¿Cuál es la probabilidad de que en m pruebas se obtengan exactamente $m-1$ resultados distintos?

Sea V_m dicha probabilidad. BOREL, utilizando combinatoria, obtiene que

$$V_m = (m-1)! \sum_{n_1<n_2<\cdots<n_{m-1}} p_{n_1}^2 p_{n_2}\cdots p_{m-1} \tag{8}$$

luego, por el método de Waring, BOREL afirma que

$$\sum p_{n_1} \sum p_{n_2}p_{n_3}\cdots p_{n_m} = \\ m \sum_{n_1<n_2<\cdots<n_m} p_{n_1}p_{n_2}\cdots p_{n_m} + \sum_{n_1<n_2<\cdots<n_{m-1}} p_{n_1}^2 p_{n_2}\cdots p_{n_{m-1}}. \tag{9}$$

Como $\sum p_i = 1$, y utilizando (6) y (8) se tiene que $\frac{V_m}{(m-1)!} = \frac{T_{m-1}}{(m-1)!} - m\frac{T_m}{m!}$, y por tanto $V_m = T_{m-1} - T_m$.

[11]Esta función, al desarrollarse, es de la forma $F(z) = 1 + z\sum p_1 + z2\sum p_1p_2 + \cdots + z^m \sum p_{n_1}p_{n_2}\cdots p_{n_m} + \cdots$, que visto con la notación de este problema sería: $F(z) = 1 + z + \frac{T_2}{2!}z^2 + \cdots + \frac{T_m}{m!}z^m + \cdots$. El orden de esta función, de acuerdo a su expresión en (7) depende del mismo orden infinitesimal de p_n, donde $p_n = \frac{c}{n^2}$ con c determinada por $\sum p_n = 1$; es decir, su orden es $\frac{1}{2}$, el mismo de $\cos(\sqrt{z})$.

Posteriormente, BOREL señala que el comportamiento asintótico de p_n se caracteriza por la función entera $F(z)$, la cual juega un papel importante en estas cuestiones. Esta intervendrá en las cuestiones en las que el número de casos posibles es una infinidad numerable. Por otro lado, el autor comenta que podría tomarse un número suficientemente pequeño en el momento de determinar la aproximación deseada en los cálculos, para despreciar las p_n que caigan por debajo del número escogido. En este caso se tendría un número finito de casos y por tanto puede ser abordado por la probabilidad clásica.

Por esta razón, el carácter esencial de un problema de esta categoría es el orden infinitesimal de p_n; este orden debe ser tal, que la serie de términos positivos $\sum p_n$ sea convergente. Sin embargo, no está sujeto a otra condición *a priori*, puesto que dada una serie convergente $v = \sum v_n$, se podría tener $p_n = \frac{v_n}{v}$, de manera que $\sum p_n = 1$. Esto muestra que hay una enorme variedad de posibles hipótesis aplicables sobre las p_n. Con esto BOREL concluye el estudio de los problemas de la segunda categoría y continúa con el estudio de la tercera categoría.

En los problemas de la tercera categoría, como se ha enunciado líneas arriba, BOREL considera una cantidad numerable tanto de pruebas como de resultados. Se denota la probabilidad de que en la prueba s se obtenga el resultado n por $p_{n,s}$ y sin importar cual sea s, es decir, en cualquier evento, $\sum_{i=1}^{\infty} p_{i,s} = p_{1,s} + p_{2,s} + p_{3,s} + \dots$ no solo es convergente, sino que vale uno. Considera ahora

$$\begin{cases} \sum_{i=1}^{\infty} p_{i,1} = p_{1,1} + p_{2,1} + p_{3,1} + \dots \\ \sum_{i=1}^{\infty} p_{i,2} = p_{1,2} + p_{2,2} + p_{3,2} + \dots \\ \sum_{i=1}^{\infty} p_{i,3} = p_{1,3} + p_{2,3} + p_{3,3} + \dots \\ \dots \end{cases} \tag{10}$$

BOREL señala que estas relaciones son compatibles con hipótesis diversas sobre la convergencia de las series y especifica que si las series en (10) son convergentes, cualquiera que sea n, se dirá que se está en el caso convergente; en el caso divergente, si todas divergen, y en el caso mixto si unas convergen y otras no. Sin embargo, en el artículo BOREL se limita a estudiar el caso de la convergencia.

En este caso nombra $c_n = \sum_{i=1}^{\infty} p_{n,i} = p_{n,1} + p_{n,2} + \dots + p_{n,s} + \dots$ y observa que la serie de términos positivos $\sum_n c_n$ es divergente, puesto que en caso contrario la serie doble $\sum_s \sum_n p_{n,s}$ también lo sería, lo que sería una contradicción con el hecho de que cada suma de (10) converja en uno.

Para cada caso posible, BOREL asegura, con toda razón, que se podrían resolver los problemas 1, 2 y 3 tratados anteriormente, bastando agrupar los casos desfavorables como un solo evento, y el caso favorable aparte. Habiendo hecho este análisis, BOREL introduce el último problema que debe tratarse en esta sección:

Se denotan para el caso posible de rango n por $B_{n,0}, B_{n,1}, \dots, B_{n,k}$ y $B_{n,\infty}$, las probabilidades anteriormente denotadas como $A_0, A_1, \dots, A_k, \dots, A_\infty$, es decir, se designa como $B_{n,k}$ la probabilidad de que el caso posible de rango n se produzca precísamente k veces en el conjunto de pruebas.

BOREL establece $v_{n,s} = \frac{p_{n,s}}{1-p_{n,s}}$, y se tiene entonces que [12]

$$\begin{cases} B_{n,0} = \prod(1-p_{n,s}) \\ B_{n,1} = B_{n,0}\sum_s v_{n,s} \\ B_{n,k} = B_{n,0}\sum_{s_1<s_2<\cdots<s_k} v_{n,s_1}v_{n,s_2}\cdots v_{n,s_k} \\ B_{n,\infty} = 0. \end{cases} \tag{11}$$

Problema 6. ¿Cuál es la probabilidad de que cada caso posible solo se produzca un número finito de veces?

Una solución inmediata se obtiene a partir del razonamiento siguiente: se sabe que $B_{n,\infty} = 0$, y por tanto la probabilidad de que el caso de rango n se produzca un número infinito de veces es 0. Entonces $\sum_{n=1}^{\infty} B_{n,\infty} = 0$ y la probabilidad buscada es $1 - \sum_{n=1}^{\infty} B_{n,\infty} = 1$. BOREL justifica este hecho más a fondo: primero calcula la probabilidad de que algún caso posible no se produzca más de una vez, después la probabilidad de que algunos de los primeros casos no se produzcan más de k veces, para después concluir que la probabilidad buscada es inferior a 2ε para cualquier ε tomada, es decir, la probabilidad es nula.

La naturaleza de la convergencia de las series utilizadas en este análisis cuando n crece, dice BOREL, es la característica esencial de los problemas de la tercera categoría. Por esta razón hay una gran variedad de ellos; sin embargo, BOREL se limita a las hipótesis tales que el problema tenga aplicaciones, ya que su objetivo no era desarrollar una teoría completamente general sin aplicaciones. No obstante, consideramos un gran logro el manejo de probabilidades donde las muestras son infinitas.

Finalmente, para concluir esta sección, BOREL dice que va a mostrar cuál puede ser la utilidad de las nociones que acaba de introducir, y para cumplir con este objetivo incluye tres capítulos más en su artículo: uno dedicado a fracciones decimales, otro dedicado a fracciones continuas (que nos recuerda el trabajo hecho por WIMAN en 1900-1901 en [22] y [23]) y el último dedicado a cuestiones diversas. Por razones de espacio nosotros presentaremos un análisis del capítulo dedicado a las fracciones decimales, que creemos muestra claramente el vínculo que deseamos estudiar entre probabilidad y teoría de la medida.

2.21. Fracciones decimales

Para ejemplificar lo anterior, BOREL considera fracciones decimales comprendidas en el intervalo $[0,1]$, es decir expresiones de la forma $\sum_{n=1}^{\infty} \frac{a_n}{10^n}$, tomando a a_n como enteros no negativos menores que 10, y llama *fracciones decimales de base q* a las expresiones de la forma $\sum_{n=1}^{\infty} \frac{b_n}{q^n}$, en donde las b_n son enteros no negativos menores que q.

El objetivo que se plantea el autor es estudiar la probabilidad de que una fracción decimal pertenezca a un conjunto dado, suponiendo que

[12] En estas igualdades solo renombra $A_0, A_1, \ldots, A_\infty$, descritos en (2), (4) y (5)... adecuados para este problema.

1. las cifras decimales que la componen sean independientes, y que
2. cada una de ellas tiene una probabilidad igual a $\frac{1}{q}$ (en el caso de base q) de tomar cada uno de los valores posibles: $0, 1, \ldots, q-1$.

BOREL argumenta que aunque se sabe que no siempre se cumplen estas hipó-tesis, desde el punto de vista geométrico se justifican fácilmente, pues son equivalentes a calcular la probabilidad de que un número decimal, representado por un punto en el segmento $[0, 1]$, se encuentre en un subsegmento del mismo intervalo; y esta probabilidad es igual a la longitud de este subsegmento.

BOREL a continuación toma la cifra 3, por considerar una cifra determinada, y dice que si se toman como casos favorables aquellos en los que se presente esta cifra, se estaría de nuevo en el caso de la primera categoría de problemas, siendo los decimales sucesivos aquellos que corresponden a la infinidad numerable de pruebas. La probabilidad del caso favorable es aquí la misma para cada prueba, y nos sitúa en el caso de divergencia; la probabilidad de que la cifra 3 se repita infinitamente es, entonces, igual a la unidad. La probabilidad de que todas las cifras sean 3 es nula; sin embargo, esto es lo que ocurre cuando transformamos $\frac{1}{3}$ en fracción decimal: como se ha señalado, la probabilidad es igual a cero, no significando imposibilidad.

Para simplificar el problema, BOREL considera el sistema base 2, en donde cada cifra, cero o uno, tendrá probabilidad $\frac{1}{2}$. BOREL propone tomar como caso favorable el caso en el que aparece cero como cifra. Se sabe que si se tienen $2n$ pruebas, la probabilidad de que el número de casos favorables caiga entre $n - \lambda\sqrt{n}$ y $n + \lambda\sqrt{n}$ es igual a $\Theta(\lambda) = \frac{2}{\sqrt{\pi}} \int_0^{\lambda} e^{-\frac{x^2}{2}} dx$. Esta probabilidad tiende rápidamente a 1, conforme λ tiende a ∞.

Después BOREL considera una sucesión de números λ_n que aumenten indefinidamente con n, de tal manera que $\lim_{n\to\infty} \frac{\lambda_n}{\sqrt{n}} = 0$. Toma como ejemplo $\lambda_n = \log n$. Ahora, del conjunto de $2n$ pruebas, si se considera como favorable el caso que tenga la cifra 0 y que esté entre $n - \lambda_n\sqrt{n}$ y $n + \lambda_n\sqrt{n}$, tendrá probabilidad $p_n = \Theta(\lambda_n) = \frac{2}{\sqrt{\pi}} \int_0^{\lambda_n} e^{-\frac{x^2}{2}} dx$.

Consideramos que es importante hacer notar que BOREL debió requerir que

$$\sum_n \left| p_n - \frac{2}{\sqrt{\pi}} \int_{\lambda_n}^{\infty} e^{-\lambda^2} d\lambda \right| < \infty,$$

puesto que, de no converger $\sum_n \int_{\lambda_n}^{\infty} e^{-\frac{\lambda^2}{2}} d\lambda$, no necesariamente converge $\sum_n p_n$.

Ahora bien, a continuación BOREL define $q_n = 1 - p_n$, y la probabilidad de que el caso desfavorable se presente una infinidad de veces es nula. En otros términos, con probabilidad uno, a partir de alguna n se obtienen constantemente casos favorables, puesto que en este caso, el cociente entre 0 y 1 está comprendido entre $\frac{n-\lambda_n\sqrt{n}}{n+\lambda_n\sqrt{n}}$ y $\frac{n+\lambda_n\sqrt{n}}{n-\lambda_n\sqrt{n}}$, es decir, entre $\frac{1-\frac{\lambda_n}{\sqrt{n}}}{1+\frac{\lambda_n}{\sqrt{n}}}$ y $\frac{1+\frac{\lambda_n}{\sqrt{n}}}{1-\frac{\lambda_n}{\sqrt{n}}}$. Estos cocientes tienden a cero y a uno cuando n tiende a infinito, por lo cual se tiene que

> La probabilité pour que le rapport entre le nombre des 0 et des 1 tende vers l'unité (lorsque le nombre des chiffres considérés augmente indéfiniment) est

> égale à un. La probabilité pour que se rapport ne tende vers aucune limite, ou tende vers une limite différente de l'unité, est, par suite, égale à zéro.[13] [4, p. 260]

Posteriormente BOREL extiende lo anterior a cualquier base, estudia las fracciones decimales periódicas y finalmente cierra ese artículo con las siguientes palabras:

> Lorsque la théorie des probabilités dénombrables aura été développée dans le sens qui vient d'être indiqué, il sera interéssant de comparer les résultats acquis avec ceux que l'on obtenient par la théorie des probabilités continues ou géométriques.
>
> Il *existe* certainement (si ce n'est pas un abus d'employer ici le verbe exister) dans le continu géométrique des éléments qui ne peuvent pas être définis: tel est le sens réel de l'importante et célèbre proposition de M. GEORG CANTOR: le continu n'est pas dénombrable. Le jour où ces éléments indéfinissables seraient réellement mis à part et où on ne prétendrait point les faire intervenir plus ou moins implicitement, il en résulterait certainement une grande simplification dans les méthodes de l'Analyse; je serais heureux si les pages précédentes pouvaient contribuer à faire pressentir l'intérêt qui s'attacherait à l'étude de telles questions. [4, p. 271]

3. El trabajo de Hugo Steinhaus

Después de este trabajo llevado a cabo por BOREL, y antes de la publicación del trabajo de STEINHAUS, en 1923, [21], consideramos que es importante mencionar el trabajo de FELIX HAUSDORFF. Este fue publicado entre los trabajos de BOREL y de STEINHAUS, y aunque dista de ser el único trabajo de probabilidad publicado en esa época, dada la dirección que hemos decidido seguir en el presente texto, consideramos que un estudio de los demás textos sobre probabilidad (aplicada, fundacional, etc.) queda fuera de nuestra línea de estudio.

En 1914 HAUSDORFF publicó su *Grundzüge der Mengenlehre* [9], que es un texto dedicado a la teoría de conjuntos, y en él, en el último capítulo, HAUSDORFF presenta la probabilidad como una aplicación de la teoría de la medida, y dice que muchos resultados sobre la medida de conjuntos de puntos se vuelven mucho más familiares cuando se expresan en el lenguaje del cálculo de las probabilidades. HAUSDORFF no identifica la probabilidad matemática con la medida; lo que hace es señalar que una medida normalizada cumple con todas las propiedades formales que debe cumplir la probabilidad matemática.

Lo que HAUSDORFF hace en [9, pp. 416-17] es considerar dos conjuntos $\mu-$medibles M y P tales que $P \subseteq M$, y establece que $\frac{\mu(P)}{\mu(M)}$ puede ser definida como la probabilidad de que un punto de M pertenezca a P. Para HAUSDORFF la medida se define como una probabilidad, y al igual que BOREL, afirma que *probabilidad cero* no significa necesariamente *imposibilidad*.

[13] A este resultado se le conocerá como la *ley de los grandes números de* BOREL o *ley fuerte de los grandes números*.

Posteriormente, para mostrar cómo se aplica la teoría de la medida a cuestiones probabilísticas, HAUSDORFF retoma el teorema de BOREL, que presenta de la siguiente manera: el conjunto de reales en el intervalo unitario, para los cuales el límite de la frecuencia relativa de 1 es $\frac{1}{2}$, tiene medida 1 [9, p. 419]. Dada la interpretación de HAUSDORFF de medida como probabilidad, obtiene que la probabilidad de que haya asintóticamente tantos unos como ceros en la expresión diádica de un número es 1.

Dada la importancia que tuvo el libro de HAUSDORFF, nos parece claro que esta visión, presente desde BOREL, en la que se vinculan ya de manera estrecha la probabilidad y la teoría de la medida, es la que debe haber proliferado en el mundo matemático de esa época, del cual HUGO STEINHAUS claramente formaba ya parte.

3.1. 1923: *Les probabilités dénombrables et leur rapport à la théorie de la mesure*

HUGO STEINHAUS publicó su artículo Les probabilités dénombrables et leur rapport à la théorie de la mesure en *Fundamenta Mathematicae* en 1923. El artículo comienza retomando el teorema de BOREL,[14] como ya lo había hecho también HAUSDORFF, y STEINHAUS afirma que su objetivo en el artículo es establecer una teoría axiomática de la probabilidad numerable en términos de teoría de la medida que permitiera pasar libremente entre las dos interpretaciones [21, p. 286]:

> Le but de cette Note est d'établir un système des postulats pour les probabilités dénombrables qui permettra d'une fois pour toutes de passer d'une interprétation à l'autre dans les recherches de ce genre.

STEINHAUS se limitará a estudiar problemas en los que el número de casos es finito, pero el número de pruebas es numerable; es decir, problemas dentro de la primera categoría de BOREL, y en particular tomará el caso más simple con únicamente dos posibles resultados, e igualmente probables, que lo llevará al teorema de BOREL, que es el resultado que aparentemente ha hecho que STEINHAUS se interese por este tema.

STEINHAUS divide su artículo en cuatro secciones: en la primera presenta su fundamentación axiomática, en la segunda se encuentran aplicaciones de la misma, la tercera sección retoma un teorema de LAPLACE, y finalmente la última sección está dedicada a generalizaciones de su teoría.

Para comenzar, STEINHAUS considera una serie de eventos independientes con solo dos casos posibles, rojo (ρ) y negro (ν), y supone que son igualmente probables. Una partida (ω) será, por definición, una sucesión determinada de ν y ρ, A será el conjunto de todos las ω posibles; E_i, los subconjuntos de A, y $\mathbb{M}$, la clase de todos los E_i.

A continuación STEINHAUS pide suponer que es posible dar una clase $\bar{\mathbb{A}}$ de los E_i (que son parte de $\mathbb{M}$) y una función de conjuntos $\mu(\cdot)$ definida para todos los E_i de $\bar{\mathbb{A}}$, de manera que se tenga lo siguiente:[15]

[14] STEINHAUS se refiere a este resultado como "la paradoja de BOREL", y lo enuncia diciendo que la probabilidad de que la frecuencia del dígito 0 en el desarrollo diádico de un número tomado al azar sea igual a $\frac{1}{2}$ es 1.

[15] Al igual que en el resto de nuestro texto, hemos tratado de conservar la notación original. En este caso solo hemos hecho modificaciones mínimas para facilitar la comprensión.

1. $\mu(E) \geq 0$ para todo E perteneciente a $\bar{\mathbb{A}}$.
2. *a*) Si E_n es un conjunto compuesto por todos los ω con las primeras n entradas iguales dadas, entonces E_n pertenece a $\bar{\mathbb{A}}$.
   b) Si E_n y E'_n solo difieren en la entrada i-ésima ($i \geq n$), entonces $\mu(E_n) = \mu(E'_n)$
   c) $\mu(A) = 1$
3. Si E_i pertenece a $\bar{\mathbb{A}}$ para toda i, y $E_i \cap E_j = \emptyset$ para $i \neq j$, entonces $\sum_{i=1}^{n} E_i$ y $\sum_{i=1}^{\infty} E_i$ pertenecen a $\bar{\mathbb{A}}$, y

$$\mu\left(\sum_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} \mu(E_i) \text{ y } \mu\left(\sum_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \mu(E_i)$$

4. Si $E_2 \subset E_1$ y E_1, E_2 pertenecen a $\bar{\mathbb{A}}$, entonces $E_1 - E_2$ pertenece a $\bar{\mathbb{A}}$.
5. Si E pertenece a $\bar{\mathbb{A}}$ y $\mu(E) = 0$, entonces toda parte[16] E' de E pertenece a $\bar{\mathbb{A}}$.

Después de enunciar los cinco postulados, STEINHAUS presenta algunos comentarios sobre los mismos:

Lo primero que afirma es que el sistema de postulados conduce a una clase $\bar{\mathbb{A}}$ compuesta de conjuntos de partes para las cuales "el problema de la determinación de una probabilidad general es posible" [21, p. 288]. La función $\mu(\cdot)$ es la probabilidad buscada, y los postulados (1), (2c), (3) y (4) muestran que $0 \leq \mu(E) \leq 1$.

El postulado (2a) exige que las partidas finitas consideradas en la teoría clásica de probabilidad tengan una probabilidad definida. El (2b) expresa la equiprobabilidad (que se tiene por hipótesis) entre el rojo y el negro, mientras que el (2c) determina la constante $\mu(A)$ conforme a la probabilidad clásica, e impide tomar a μ como la función idénticamente cero.

Por otro lado, (3) es el principio bien conocido que establece que la probabilidad de una suma de eventos mutuamente excluyentes[17] es igual a la suma de probabilidades correspondientes. En (4) se atribuye una probabilidad a un conjunto resultado de la operación resta, de acuerdo a (3), si ambos conjuntos tienen probabilidades determinadas. Con (5), en virtud de $(1), (3)$ y (4), se establece que $\mu(E') = 0$, si $\mu(E) = 0$ y E' es parte de E. Si un evento total tiene probabilidad nula, entonces un evento más particular o subevento tendrá también probabilidad nula.

STEINHAUS afirma que estos postulados concuerdan con los principios de la probabilidad clásica, y (2a) y (3) generalizan al caso numerable.

Posteriormente considera $\mathbb{A}$ como la intersección de todas las clases $\bar{\mathbb{A}}$ que satisfacen los cinco postulados. Se verifica que $\mathbb{A}$ contiene todos los E_i y todos los conjuntos que se producen por uniones y restas finitas entre ellos. Se puede ver también que se puede trabajar con todos los E de $\mathbb{A}$ tales que $\mu(\cdot)$ y $\mathbb{A}$ satisfagan $(1) - (5)$; $\mu(\cdot)$ sería la probabilidad

[16]Entiéndase *subconjunto*.
[17]STEINHAUS usa la palabra *incompatibles*.

ordinaria para partidas finitas y sus combinaciones. Esto hace notar que la introducción de $\mathbb{A}$ como una suerte de $\bar{\mathbb{A}}$ mínima eleva la indeterminación que descansa en torno a $\bar{\mathbb{A}}$, sin ser única. Sólo faltaría demostrar que esta $\mathbb{A}$ es una $\bar{\mathbb{A}}$.

STEINHAUS hace corresponder a toda partida finita ω (a saber, $\nu\rho\rho\rho\rho\ldots\nu\rho$) con un número real $(0{,}01111\ldots01)$ que en desarrollo diádico comienza por cero, y coloca cero o uno en la i-ésima cifra de acuerdo a si el $i-$ésimo lugar de ω es negro (ν) o rojo (ρ), respectivamente. Esta convención: hacer corresponder a todo conjunto E de partidas infinitas de conjuntos de puntos en $[0,1]$[18] a los conjuntos E_i, correspondiéndoles intervalos de longitud $\frac{1}{2^n}$ y en particular a A el intervalo $[0,1]$. STEINHAUS designa con las mismas letras de los conjuntos de partidas los conjuntos imágenes, sin temor a introducir confusión, y pide considerar los postulados como válidos para los conjuntos lineales de puntos.

STEINHAUS afirma que partiendo de esta interpretación, sus cinco postulados no son más que una repetición de los postulados utilizados por SIERPIŃSKI para definir los conjuntos L-medibles y su medida, con una única diferencia que se encuentra en el segundo postulado. El artículo en el cual SIERPIŃSKI presenta los postulados, a los que STEINHAUS hace alusión, es “Sur une définition axiomatique des ensambles mesurables (L)” [20], y es justo esta identificación la que permite ver la unificación del concepto de *probabilidad* con el concepto de *medida*. STEINHAUS muestra esto haciendo ver que la clase de conjuntos Lebesgue-medibles es precisamente la clase $\bar{\mathbb{A}}$, y con este resultado concluye la primera sección de [21].

La segunda sección del artículo de STEINHAUS, como ya habíamos anticipado, está dedicada a las aplicaciones del principio que acaba de demostrar.

STEINHAUS toma la probabilidad de una partida determinada como la medida de un conjunto compuesto por un solo punto, es decir, cero. De esta manera cero puede representar la probabilidad de un evento posible, y por tanto STEINHAUS propone nombrar a estos eventos como *casi-imposibles*, y a los eventos con probabilidad uno como *casi-certezas*. También hace notar, como ya lo había hecho BOREL desde 1905, que una casi-certeza no implica una certeza.

STEINAHUS considera $\phi(n)$ una función positiva, creciente, y C una constante (independiente de n) que pueda tener los valores $+\infty$ y $-\infty$, y se pregunta por la probabilidad que para $\phi(n)$ y C dadas se tenga

$$\limsup_{n\to\infty} \frac{\nu(n)}{\phi(n)} = C \tag{12}$$

donde $\nu(n)$ designa el número de entradas negras que se presentan en las primeras n pruebas.

STEINHAUS afirma que esta probabilidad no puede ser más que uno o cero debido al teorema que ha demostrado. Basta buscar la medida de un conjunto E, que evidentemente es el conjunto de puntos de $[0,1]$, cuyo desarrollo diádico satisface (12), con $\nu(n)$ el

[18] STEINHAUS usa la notación $< 01 >$.

número de ceros que aparecen en las primeras n cifras. STEINHAUS toma $\alpha, \beta \in [0,1]$, $\alpha, \beta \neq \frac{1}{2^q}$ con $q \in \mathbb{Z}^+$. Supone que α y β solo difieren en sus primeras q cifras, por lo cual el límite (12) sería el mismo. Entonces, o bien $\alpha, \beta \in E$, o bien $\alpha, \beta \notin E$, es decir, pertenecerían a su complemento. Observa que la medida de E en $[\frac{n}{2^q}, \frac{n+1}{2^q}]$ para $n \in \{0, 1, \ldots 2^q - 2, 2^q - 1\}$ es siempre igual, y supone que E está en $[\frac{i}{2^q}, \frac{i+1}{2^q}]$ y tiene medida $m(E)$. Para insertarlo en $[\frac{k}{2^q}, \frac{k+1}{2^q}]$ para $k \neq i$, solo hay que considerar la traslación del conjunto E por la cantidad $(-1)^m 2^{-q}$, es decir, el conjunto $E + (-1)^m 2^{-q}$, donde $m = i - k$. Como solo se *cambia* un número finito de dígitos, el nuevo conjunto sigue cumpliendo (12), además de que tiene la misma medida, pues solo se desplaza un número real. Entonces, la "densidad" de E es constante, por lo cual solo puede ocurrir que $m(E) = 0$ o $m(E) = 1$.

Posteriormente STEINHAUS se pregunta: ¿Cómo se determinarían $\phi(n)$ y C de manera que la probabilidad de (12) sea 1? Por el teorema de Bernoulli, afirma STEINHAUS, se sabe que para $\varepsilon > 0$ y $1 > \eta > 0$ se tiene que para $n > N_{\varepsilon,\eta}$

$$\left| \frac{\nu(n)}{n} - \frac{1}{2} \right| < \varepsilon \tag{13}$$

con una probabilidad que sobrepasa a $1 - \nu$. De esta manera, el conjunto $E \subseteq [0,1]$, que está conformado por puntos cuyo desarrollo diádico cumple (13) a partir de un cierto n, tiene una medida mayor que $1 - \eta$.

Ahora, como se había visto anteriormente, dados ϕ y C, la probabilidad de que se cumpla (12) es cero o uno, y en este caso, como la probabilidad es mayor que cero, entonces debe ser uno. Por lo anterior, la medida del conjunto de puntos que no tienen esta propiedad es cero, y el autor lo denota como $D(\varepsilon)$. Entonces, $\sum_{m=1}^{\infty} D\left(\frac{1}{m}\right)$ tendrá también medida cero. Los puntos ω que no pertenecen a $\sum_{m=1}^{\infty} D\left(\frac{1}{m}\right)$ satisfacen todas las desigualdades

$$\left| \frac{\nu(n)}{n} - \frac{1}{2} \right| < \frac{1}{m}$$

para $n > Q_{\omega,m}$ y tomando m tan grande como se quiera, se tiene que

$$\lim_{n\to\infty} \frac{\nu(n)}{n} = \frac{1}{2} \tag{14}$$

y por tanto la medida del conjunto respectivo es uno.

STEINHAUS señala que esto es precisamente el teorema de BOREL en su forma aritmética, y que por tanto sus postulados y el resultado sobre la equivalencia permiten inmediatamente pasar a la interpretación original.

3.11. Un teorema de Laplace

En la sección 3 de [21] STEINHAUS retoma un teorema que le atribuye a LAPLACE y aclara que el teorema ya ha sido demostrado rigurosamente en varias ocasiones, ellas en el *Cours d'analyse* de CAMILLE JORDAN.[19] Es por tanto evidente que el interés por

[19] Véase [12, tomo II, pp. 187-190].

incluir este teorema en el presente artículo recae sobre los métodos utilizados para su demostración.

STEINHAUS enuncia el teorema de la siguiente manera: Sea $\nu(n)$ el número de veces que aparece el color negro en las primeras n pruebas y $f(t,n)$ la probabilidad que se tenga

$$\left|\frac{\nu(n)}{n} - \frac{1}{2}\right| \leq \frac{t}{\sqrt{2n}} \tag{15}$$

entonces se tiene que $\lim_{n\to\infty} f(t,n) = \frac{2}{\sqrt{\pi}}\int_0^t e^{-x^2}dx$ con t constante.

De hecho, este teorema se remonta al trabajo de JAKOB BERNOULLI en su *Ars conjectandi* de 1713 [1], y en particular a un resultado conocido hoy como la *ley de los grandes números de Bernoulli*, y abordado años más tarde por DE MOIVRE en la tercera edición de *The Doctrine of Chances* [16]. En términos actuales, lo que BERNOULLI mostró equivale a mostrar que $\lim_{n\to\infty n} \mathbb{P}(|h_n - p| \leq \varepsilon) = 1\ \forall\varepsilon > 0$, donde h_n representa la frecuencia relativa de un evento en particular con probabilidad de ocurrencia $p = a/b$ con $a, b \in \mathbb{Z}^+$ en una serie de n ensayos idénticos e independientes.

BERNOULLI da una estimación de n, para la cual $\mathbb{P}(|h_n - p| \leq \varepsilon) \geq \eta$, con $0 < \eta < 1$ y $\varepsilon = \frac{1}{b}$, donde b puede ser arbitrariamente grande. De Moivre va más allá: concluye que la probabilidad

$$\begin{aligned}
&\mathbb{P}(\left|Z - \left[\frac{n}{2}\right]\right| \leq t) \\
&\qquad = \sum_{i=-t}^{t} \mathbb{P}\left(Z = \left[\frac{1}{2}\right] + i\right) \approx 2\frac{2}{\sqrt{2\pi n}}\sum_{i=0}^{t} e^{-2\frac{i^2}{n}} \approx \frac{4}{\sqrt{2\pi n}}\int_0^t e^{-2\frac{x^2}{n}}\,dx.
\end{aligned}$$

En estas ecuaciones, DE MOIVRE utiliza Z para denotar la h_n de BERNOULLI y obtiene las aproximaciones deseadas basándose en que $\frac{\binom{n}{[\frac{n}{2}]}}{2^n} \approx \frac{2}{\sqrt{2\pi n}}$. Cabe hacer la importante aclaración que DE MOIVRE no lo ve explícitamente en términos de la exponencial o de la integral. DE MOIVRE no calcula las integrales de forma directa, las lleva a series y aproxima los valores de ε para que la probabilidad sea prácticamente igual a uno, cuando $p = \frac{1}{2}$, y deja un método incompleto pero viable para cuando $p \neq \frac{1}{2}$.

Por otro lado, LAPLACE, para 1774 había ya diseñado métodos de aproximación útiles para cálculos de probabilidades con grandes números de observaciones, pero no fue hasta 1810 que pudo ajustar su método en [15], *Mémorie sur les approximations des formules qui sont fonctions de très grands nombres et sur leur application aux probabilités.*

Inspirado por las inclinaciones de las órbitas de los planetas, LAPLACE llegó a que la probabilidad de que la suma de n variables con distribución rectangular o uniforme, como le diríamos hoy en día, entre 0 y h, caiga entre a y b con $0 \leq a < b \leq nh$ es

$$\begin{aligned}
\frac{1}{1\cdot 2\cdot 3\cdots n\cdot h^n}[b^n - n(a-h)^n + \frac{n(n-1)}{1\cdot 2}(b-2h)^n - \cdots \\
-a^n + n(a-h)^n - \frac{n(n-1)}{1\cdot 2}(a-2h)^n + \cdots]
\end{aligned}$$

que en notación contemporánea se reescribiría como

$$\mathbb{P}(a \leq \sum X_i \leq b) = \frac{1}{h^n n!}\left(\sum_{i=0}^{N}\binom{n}{i}(-1)^i(b-hi)^n - \sum_{i=0}^{N}\binom{n}{i}(-1)^i(a-hi)^n\right),$$

tomando N como el mínimo entre n y $[\frac{b}{h}]$.

El autor aclara que se deben rechazar los términos que sean negativos bajo la potencia, y posteriormente retoma el trabajo de DE MOIVRE haciendo el cambio de variable $t = i\omega$, definiendo así la tan aclamada función característica.

Con esto obtiene que la probabilidad de que la suma de n observaciones esté contenida entre los límites $nq \pm h$ es $\frac{2}{\sqrt{\pi}}\sqrt{\frac{k}{2k'}}\int e^{-\frac{k}{k'}r2}dr$, en donde, en términos modernos, k' es la desviación estándar.

Este es el camino que STEINHAUS retoma; sin embargo, ahora haciendo uso de la teoría de la medida. STEINHAUS designa cpmo $F(t,n)$ el conjunto de puntos en los cuales la cantidad $\nu(n)$ de ceros en su desarrollo diádico satisface (15), y nosotros designaremos como $\mu(F(t,n))$ su medida.[20] Se tiene entonces la siguiente desigualdad:[21]

$$\begin{aligned}\mu\left(\liminf_{n\to\infty} F(t,n)\right) &\leq \lim_{n\to\infty}\mu\left(F(t,n)\right)\\ &= \lim_{n\to\infty} f(t,n) = \frac{2}{\sqrt{\pi}}\int_0^t e^{-x^2}dx < 1. \quad (16)\end{aligned}$$

Aquí STEINHAUS ya usa el concepto de probabilidad casi indistintamente con el de medida, midiendo el conjunto que tiene la propiedad de la cual se calcula su probabilidad, como se observa en (16).

STEINHAUS designa como $S(t')$ el conjunto cuyos puntos cumplen que

$$S(t') = \left\{x : \limsup_{n\to\infty}\sqrt{2n}\left|\frac{\nu(n)}{n} - \frac{1}{2}\right| \leq t' < t\right\}$$

y observa que, por definición, está contenido en $\liminf_{n\to\infty} F(n,t)$, por lo cual

$$\mu\left(S(t')\right) \leq \mu\left(\liminf_{n\to\infty} F(n,t)\right) < 1.$$

Ahora, utilizando un razonamiento análogo a cuando se quería calcular la probabilidad de que $\limsup_{n\to\infty}\frac{\nu(n)}{\phi(n)} = C$, se obtiene que la medida de $S(t')$ es cero.

[20] STEINHAUS usa $|\cdot|$ para denotar la medida de un conjunto, pero consideramos mejor utilizar la notación actual.

[21] La primera desigualdad es un resultado bien conocido de teoría de la medida, y la primera igualdad es la deficinión de la medida de $F(t,n)$, es decir, su probabilidad de ocurrencia.

A partir de esto se obtiene que $\mu\left(\bigcup_{k=1}^{\infty} S(k)\right) = 0$, y fuera de $\bigcup_{k=1}^{\infty} S(k)$ se tendrá que casi dondequiera, por la definición de $S(k)$,

$$\limsup_{n\to\infty} \sqrt{2n}\left|\frac{\nu(n)}{n} - \frac{1}{2}\right| = +\infty. \tag{17}$$

Luego, STEINHAUS considera el límite superior de $\{F(t,n)\}$ y afirma que

$$\mu\left(\limsup_{n\to\infty} F(t,n)\right) \geq \lim_{n\to\infty} \mu(F(t,n)) = \lim_{n\to\infty} f(t,n) = \frac{2}{\sqrt{\pi}}\int_0^t e^{-x^2}dx > 0.$$

Define el conjunto $I(t') = \left\{x : \liminf_{n\to\infty} \sqrt{2n}\left|\frac{\nu(n)}{n} - \frac{1}{2}\right| \leq t'\right\}$ y si $t' > t$, entonces $\limsup_{n\to\infty} F(t,n) \subseteq I(t')$ y $0 < \mu(\limsup_{n\to\infty} F(t,n)) \leq (I(t'))$. Además, $\liminf_{n\to\infty} \sqrt{2n}\left|\frac{\nu(n)}{n} - \frac{1}{2}\right| \leq t'$ casi siempre, y como t' es un número positivo arbitrario, entonces $\liminf_{n\to\infty} \sqrt{2n}\left|\frac{\nu(n)}{n} - \frac{1}{2}\right| = 0$.

Por otro lado, STEINHAUS demuestra la desigualdad

$$\forall k, n \in \mathbb{Z}^+ : \sum_{p=0}^{n} \binom{n}{p}(2p-n)^{2k} \leq 2^n(2k-1)!!n^k \tag{18}$$

donde $(2k-1)!! = 1 \cdot 3 \cdot 5 \cdot (2k-3)(2k-1)$, puesto que se apoyará en ella para llegar al resultado deseado.[22]

Ahora bien, continúa STEINHAUS, la condición necesaria y suficiente para que entre las n primeras cifras del desarrollo diádico de un número α se encuentren p ceros y $n-p$ unos es que α esté situado en un intervalo perteneciente a cierta colección compuesta de $\binom{n}{p}$ intervalos ajenos, cuyas longitudes son $\frac{1}{2^n}$. Claramente la suma de las longitudes será $\binom{n}{p}\frac{1}{2^n}$.

Para n fija, considera todos los intervalos I_ε tales que las p correspondientes cumplan la siguiente desigualdad:

$$\left|\frac{p}{n} - \frac{1}{2}\right| > \varepsilon \text{ con } \varepsilon > 0 \tag{19}$$

y observa que $\sum_{p'} \binom{n}{p}\frac{1}{2^n}$ será la longitud de los intervalos ajenos para p fija de longitud $\binom{n}{p}\frac{1}{2^n}$ y que satisfagan (19).

Luego, para k natural se tendrá que

$$\sum_{p=0}^{n} \binom{n}{p}\left(\frac{p}{n} - \frac{1}{2}\right)^{2k}\frac{1}{2^n} \geq \sum_{p'}\left(\frac{p}{n} - \frac{1}{2}\right)^{2k}\frac{1}{2^n} > \sum_{p'}\binom{n}{p}\frac{\varepsilon^{2k}}{2^n}$$

[22] STEINHAUS procede por inducción para demostrar esta desigualdad; nosotros omitiremos la demostración por ser larga y no aportar nada al entendimiento del desarrollo de la teoría de la probabilidad.

y por (18) STEINHAUS obtiene que

$$\sum_{p'} \binom{n}{p} \frac{1}{2^n} < \frac{1}{2^n (2n\varepsilon)^{2k}} \sum_{p=0}^{n} \binom{n}{p} (2p-n)^{2k}$$

$$< \frac{1}{(2n\varepsilon^2)^k} \cdot \frac{(2k-1)!!}{2^k} = \frac{k!}{(2n\varepsilon^2)^k} < \frac{k^k e^{-k}\sqrt{2\pi k} e^2}{(2n\varepsilon^2)^k}$$

donde el primer término es la suma de longitudes de los I_ε y para la última desigualdad se propone tomar $0 < \eta < 1$ tal que $2n\varepsilon^2 > (1+\eta)\ln n$ para $n \geq 3$ y $k = [(1+\eta)\ln n]$.[23]

Luego, es posible obtener que $I_\varepsilon < 2e^{1+\frac{1}{12n}} \sqrt{\pi} \frac{\sqrt{\ln n}}{n^{1+\eta}}$, y si se nombra Z_n al conjunto compuesto por todas los I_ε para n, fija se tiene que

$$\mu(Z_n) < 2e^{1+\frac{1}{12n}} \sqrt{\pi} \frac{\sqrt{\ln n}}{n^{1+\eta}} \text{ para } n \geq 3.$$

El conjunto límite, L, de la sucesión de $\{Z_n\}$ tiene medida cero, puesto que la serie $\sum_{n=1}^{\infty} \mu(Z_n)$ converge.

Si α no pertenece a L, no pertenece a ningún Z_n a partir de una $n > N(\alpha)$, lo cual se traducirá en negar (19), es decir, $\left| \frac{\nu(n)}{n} - \frac{1}{2} \right| \leq \varepsilon$, que a su vez es equivalente a que $\sqrt{\frac{2n}{\ln n}} \left| \frac{\nu(n)}{n} - \frac{1}{2} \right| \leq \sqrt{1+\eta}$ para una ε dada y para $n > N(\alpha)$, lo cual implica

$$\limsup_{n\to\infty} \sqrt{\frac{2n}{\ln n}} \left| \frac{\nu(n)}{n} - \frac{1}{2} \right| \leq \sqrt{1+\eta}. \tag{20}$$

Como L tiene medida nula, (20) aplica para casi toda α; además, como η puede ser arbitrariamente chica, entonces STEINHAUS obtiene que

$$\limsup_{n\to\infty} \sqrt{\frac{2n}{\ln n}} \left| \frac{\nu(n)}{n} - \frac{1}{2} \right| \leq 1 \tag{21}$$

casi siempre.

Para terminar la sección, STEINHAUS hace notar que (21) puede ser también interpretada de la siguiente manera: casi seguramente la diferencia absoluta entre $\frac{1}{2}$ y el cociente del número de entradas negras y el número total n de entradas no sobrepasará $\sqrt{\frac{\ln n}{2n}}$ para las partidas realmente grandes, y concluye diciendo "L'expérience confirme aussi ce resultat" [21, p. 302].

Finalmente, en la última sección de [21], STEINHAUS enuncia algunos teoremas que ejemplifican casos más generales de los tratados en la segunda sección, que decidimos no incluir en este trabajo.

[23] En donde $[\cdot]$ denota la función parte entera. STEINHAUS propone tomar η de modo que $2n\varepsilon^2 = (1+\eta)\ln n$ para $n \geq 3$, pero es suficiente pedir la desigualdad.

4. Algunos comentarios finales

Para concluir, consideramos que es importante llevar nuestro trabajo hasta 1933, con la axiomatización que se presenta hoy en día como *la* axiomatización de la teoría de la probabilidad.

BRUNO DE FINETTI publicó en sus *Scritti* [7] un programa sobre la fundamentación de la probabilidad, y creó la probabilidad cualitativa, en donde muestra cómo las propiedades numéricas usuales de la probabilidad pueden ser vistas como resultado de la noción de "apuestas coherentes".

Se define $E \preceq E'$ como el evento E es al menos tan probable como E', $E \preceq E'$ y $E' \preceq E$ significa $E \approxeq E'$, que se leería E y E' son idénticamente probables. $E \prec E'$ simboliza $E \preceq E'$, pero no $E \approxeq E'$, y $E+E'$ es el evento E o el evento E'. Posteriormente DE FINETTI presenta cuatro axiomas:

1. $E \preceq E'$ o $E' \preceq E$ para cualesquiera dos eventos E, E'.
2. Si cierto evento A es certeza, B es imposible y E ninguno de los dos, entonces $A \prec E \prec B$.
3. Si $E \preceq E'$ y $E' \preceq E''$, entonces $E \preceq E''$: transitividad.
4. Si E_1 y E_2 son incompatibles con E, $E + E_1 \preceq E + E_2$, si y solo si $E_1 \preceq E_2$. Específicamente, $E_1 \approxeq E_2$ si y solo si $E_1 + E \approxeq E_2 + E$.

DE FINETTI considera la idea de apuestas coherentes como un camino simplificado para introducir la probabilidad numérica. Si se supone un "banquero" que tiene que aceptar las apuestas para un evento E para cualquier suma S escogida por quien apuesta, el banquero tiene el derecho de escoger la "probabilidad" p del evento E. La apuesta se definiría como la cantidad aleatoria $G = (|E| - p)S$, donde $|E| = 1, 0$, dependiendo de si E ocurre o no, respectivamente. Si se apuesta varias veces, de manera finita, la apuesta sería

$$G = (|E_1| - p_1)S_1 + (|E_2| - p_2)S_2 + \ldots (|E_n| - p_n)S_n.$$

Aquí, la coherencia se presentaría al requerirle al banquero que no maneje el esquema de pérdida segura, esto es, que G sea positiva para alguna suma de S_i, bajo todos los posibles resultados de las E_i. DE FINETTI prueba que su coherencia implica que los números p_i asociados a los eventos E_i satisfacen los axiomas de aditividad finita de la probabilidad clásica.

Finalmente, en 1933 ANDREI KOLMOGÓROV publica su *Grundbegriffe der Wahrscheinlichkeitsrechnung* [13], obra que culmina la búsqueda de una fundamentación para la teoría de la probabilidad. Su objetivo es claro: "El propósito de esta monografía es dar una fundamentación axiomática a la teoría de la probabilidad".

KOLMOGÓROV considera un conjunto E y un conjunto $\mathcal{F}$ de subconjuntos de E, a cuyos elementos llama *eventos aleatorios*, y enuncia cinco axiomas:

1. $\mathcal{F}$ es un álgebra de conjuntos.

2. $\mathcal{F}$ contiene al conjunto E.
3. A cada conjunto A de $\mathcal{F}$ se le asigna un número real no negativo $\mathbb{P}(A)$. A este se le nombra la probabilidad del evento A.
4. $\mathbb{P}(E) = 1$.
5. Si A y B no se intersecan, entonces $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.

   Posteriormente añade un sexto axioma, quizá redundante para una $\mathcal{F}$ finita, pero independiente para una infinita:
6. Si $... \subseteq A_2 \subseteq A_1$ es una sucesión decreciente de eventos de $\mathcal{F}$ con $\cap_{i=1}^{\infty} A_n = \emptyset$, entonces $\lim_{n\to\infty} \mathbb{P}(A_n) = 0$.

Este último es el axioma de continuidad, y dados los cinco primeros, es equivalente a la aditividad numerable de la probabilidad.[24]

Las reacciones en los primeros años después de la publicación se pueden observar en los siguientes dos comentarios. El primero es de WILLIAM FELLER, quien publicó una reseña de la obra de KOLMOGÓROV en *Zentralblatt für Mathematik* en 1934.

> The calculus of probabilities is constructed axiomatically, with no gaps and in the greatest generality, and for the first time systematically integrated, fully and naturally, with abstract measure theory. The axiom system is certainly the simplest imaginable... The great generality is noteworthy; probabilities in infinite dimensional spaces of arbitrary cardinally are dealt with... The presentation is very precise, but rather terse, directed to the reader who is not unfamiliar with the material. Measure theory is assumed.

El segundo comentario fue hecho por MAURICE FRÉCHET en el Congreso Mundial de 1936 en Oslo.[25] Es con este comentario que quisiéramos concluir, pues creemos que refleja fielmente los hechos como los hemos estudiado en el presente texto:

> The foundations of probability theory have changed little. But they have been enriched by particulars about the additivity of the probabilities of denumerable sets of incompatible events. This revision was completely and systematically exposited for the first time by A. KOLMOGOROV. But on this point we mention, on the one hand, that one would find the elements of such an exposition scattered about in the previous publications of many of the authors who have written about the calculus of probabilities in recent years. And on the other hand, if this revision was needed it was because of the revolution brought about by ÉMILE BOREL, first in analysis by his notion of measure, and then in probability theory by his theory of denumerable probabilities. [8, pp. 153-154]

[24]Estos seis axiomas podrían reducirse a una función conjuntista no negativa, aditiva en el sentido de FRÉCHET con $\mathbb{P}(E) = 1$.

[25]El texto aparece en [8].

Referencias

[1] J. Bernoulli, *Ars conjectandi*, 1713.

[2] J. Bertrand, *Calcul des probabilités*, Gauthier-Villars, París, 1889.

[3] É. Borel, *Remarques sur certaines questions de probabilité*, Bulletin de la Société Mathématique de France **33** (1905), 123-128.

[4] É. Borel, *Les probabilités denombrables et leurs applications arithmétiques*, Rendiconti del Circolo Matematico di Palermo 27 (1909), 247-271.

[5] É. Borel, *Éléments de la theorie des probabilités*, Hermann, 1909.

[6] É. Borel, *Leçons sur la theorie de fonctions*, Gauthier-Villars, Paris, 1950.

[7] B. De Finetti, *Scritti (1926-1930)*, Cedam, Padova, 1981.

[8] M. Fréchet, *Les mathématiques et le concret*, Presses Universitaires de France, París, 1955.

[9] Hausdorff, F. *Grundzüge der Mengenlehre*, De Gruyter, Leipzig, 1914.

[10] D. Hilbert, *Mathematische probleme*, Göttinger Nachrichten (1900), 253-297.

[11] D. Hilbert, *Mathematical problems*, Bulletin of the American Mathematical Society, **10** (1902), no. 8, 437-479.

[12] C. Jordan, *Cours d'Analyse de l'Ecole Polytecnique*, Gauthier-Vilars, Paris, 1883.

[13] A. Kolmogorov, *The theory of probability*, en A. D. Aleksandrov, A. Kolmogorov and M. Lavrent'ev (eds.), *Mathematics, Its Contents, Methods and Meaning*, 2, pp. 229-264, MIT Press, 1963.

[14] E. Knobloch, *Emile Borel as a probabilist*, en *The probabilist revolution* Vol. 1, Cambridge Massachusetts, 1987, pp. 215-233.

[15] P. S. Laplace, *Mémorie sur les approximations des formules qui sont fonctions de très grands nombres et sur leur application aux probabilitès*, Mémoires de l'Académie Royale des Sciences de Paris, año 1809, 1810, 353-415.

[16] A. De Moivre, *The Doctrine of Chances*, Londres, 1738.

[17] J. Von Plato, *Creating modern probability*, Cambridge Studies in Probability, Induction and Decision Theory, Cambridge University Press, 1994.

[18] H. Poincaré, *Calcul des Probabilités*, Gauthier-Villars, París, 1896.

[19] I. Schneider, *Die Entwicklung der Wahrscheinlichkeitstheorie von den Anfängen bis 1933: Einführungen und Texte*, Darmstadt Wissenschaftliche Buchgesellschaft, 1988.

[20] W. Sierpiński, *Sur les définitions axiomatiques des ensembles mesurables (L)* bull. Intern. Acad. Sci. Cracovie A, 1918, pp. 173-178.

[21] H. Steinhaus, *Les probabilités dénombrables et leur rapport à la théorie de la mesure*, Fundamenta Mathematica **IV** (1923), 286-310.

[22] A. Wiman, *Ueber eine Wahrscheinlichkeitsaufgabe bei Kettenbruchentwicklungen*, Kongliga Vetenskaps-Akademiens Furhandlingar, 1900, 829-841.

[23] A. Wiman, *Bemerkungen über eine von Gyldén aufgeworfene Wahrscheinlichkeitsfrage*, Hakan Ohlssons Boktryckeri, Lund, 1901.

Recibido en agosto de 2015. Aceptado para publicación en marzo de 2016.

CARMEN MARTÍNEZ-ADAME
DEPARTAMENTO DE MATEMÁTICAS, FACULTAD DE CIENCIAS
UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
MÉXICO D.F.
e-mail:cmai@ciencias.unam.mx

JESÚS CABALLERO
DEPARTAMENTO DE MATEMÁTICAS, FACULTAD DE CIENCIAS
UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
MÉXICO D.F.
e-mail: jcm@ciencias.unam.mx