Sobre la construcción del Arco del Triunfo Probabilístico.

Jesús Caballero Medrano

Tesis para obtener el título de Matemático
Tutora: Dra. Camen Matínez-Adame Isaís
Facultad de Ciencias
Universidad Nacional Autónoma de México
2011

A todos quienes estuvieron, han estado y estarán conmigo.
A quienes sacarán provecho de este trabajo.
A mi madre.

Contenido

Introducción

Probability is the most important concept in modern science, especially as nobody has the slightest notion what it means. *Bertrand Arthur William Russell* (1929)

La probabilidad tiene una gran influencia en nuestros días, desde el juego de azar más inocente hasta el modelo matemático más elaborado de pérdida o ganancia en un banco, por dar un ejemplo. Prácticamente hay cabida en todas las ciencias para ésta y seguirá siendo cada vez más importante en las ramas economico-administrativas.

Pero, ¿qué es la probabilidad? El Diccionario de la Real Academia Española (DRAE) lo define de la siguiente manera:

probabilidad.
(Del lat. probabilitas, -atis).

1. f. Verosimilitud o fundada apariencia de verdad.

2. f. Cualidad de probable, que puede suceder.

3. f. Mat. En un proceso aleatorio, razón entre el número de casos favorables y el número de casos posibles.

Pero esto no dejó ni dejaría a un matemático satisfecho. Se tiene registro que desde el siglo XV se planteó un razonamiento probabilístico en el sentido más general, que bien podría acomodarse en la tercera acepción de la definición según el DRAE. Posteriormente el pensamiento fue evolucionando y con ello el concepto de probabilidad.

Algunas de las grandes mentes europeas durante medio milenio abordaron la naciente Teoría de la Probabilidad: Pascal, Fermat, Huygens, Jakob (I) y

Daniel (I) Bernoulli, Euler, Gauss y Laplace. Este último intentó asentar la teoría en postulados, pues sólo se tenían resultados aislados. Pero todos ellos trataron con probabilidades donde la muestra era finita. Sería hasta inicios del siglo XX cuando Anders Wiman seguido de Emile Borel darían un giro importante que redefiniría para siempre a la probabilidad.

Por otro lado, se nos ha enseñado, de este lado del mundo, que *el padre fundador* de la Teoría de Probabilidad fue Andrei Kolmogorov. Este trabajo plantea el siguiente debate: ¿es adecuado presentarlo así o es la teoría actual de probabilidad el resultado de un desarrollo histórico más complejo que es necesario comprender mejor? Esto se hará desde una perspectiva occidental, a través de distintintos trabajos, que se consideraron piedras fundadoras para las ideas de Kolmogorov.

Este desarrollo que se menciona, fue producto de los avances en dos ramas distintas, que no se vislumbraba tendrían relación: la Teoría de la Medida y claro, la Teoría de la Probabilidad. Ambas teorías crecerían sin conocerse una a la otra, para encontrarse en un punto alto, para dar como resultado la *bella* Teoría de Probabilidad actual. El desarrollo bien se podría equiparar con la construcción del Arco del Triunfo que se encuentra en París. Dos pilares ajenos que se unen para formar una bella y existosa construcción final.

La historia *formal* de la teoría clásica empieza desde Cardano[1] en el siglo XVI hasta recién iniciado el siglo XX, cuando el trabajo de Wiman sobre movimientos planetarios precede al trabajo de Borel. En los capítulos siguientes se expondrán los trabajo del Émile Borel y de Hugo Steinhaus principalmente, basándose en fuentes primarias, de cuyos este último es a quien se le podría adjudicar ser el predecesor del trabajo de Kolmogorov, pues su exposición ya tiene un espíritu contemporaneo.

También se tratarán brevemente las bases históricas de la Teoría de Probabilidad clásica, los orígenes de la Teoría de la Medida y el cambio del pensamiento hacia lo infinito desde los griegos hasta las ideas de Cantor, que fueron vitales para el fortalecimiento de dicho Arco del Triunfo. Estas ideas resultan fundamentales para lograr entender y conocer de fondo la teoría actual desde nuestro punto de vista. Se mencionará uno de los orígenes de los Procesos Estocásticos y por último se enunciará con fines comparativos el tra-

[1] Al menos en el mundo europeo occidental.

bajo de Kolmogorov, que fue ampliamente aceptado y que tres decadas después fue modificado por Kolmorogv mismo para incluir unas correcciones necesarias.

La división de los capítulos de este trabajo está principalmente basada en la similitud pro-puesta entre el Arco del Triunfo parisino y la historia moderna de la Probabi-lidad. Primero se abarcan los principales exponentes de la Teoría de la Probabilidad Clásica y los precursores de la Teoría de la Medida. Dos columnas sin relación alguna aparente. Posteriormente se exponen los trabajos de Wiman, Borel y Steinhaus, quienes se han considerado los principales matemáticos que unieron estas dos ramas de la matemática obteniendo muy buenos frutos.

También a manera de ejemplo, se expone el Lema de Borel-Cantelli como una muestra la evolución no sólo en notación sino en simplicidad y en uso del análisis. Y para cerrar, se muestran los postulados de la Teoría de la Probabilidad de Kolmogorov.

El anexo incluye tanto resultados que se creyeron relevantes para ser tomados en cuenta que respaldan lo mostrado a lo largo del trabajo como demostraciones alternativas de algunos resultados.

Gran parte de la teoría matemática ha encontrado su origen o se ha visto muy influenciada por la Física, y la Probabilidad. De aquí *puede* desprenderse una serie de cuestionamientos donde pareciese que todo tiene una razón física y un comportamiento predecible, bajo ciertas circunstancias. Por lo cual, la pregunta está abierta aún: ¿Dios jugará a los dados?

Jesús Caballero Medrano
jcm@ciencias.unam.mx

Capítulo 1

Cimientos de las dos columnas

[...] Je n'a pas le temps de vous envoyer la démonstration d'une difficulté qui étonnoit fort M. [Meré]: car il a très-bon esprit, mais il n'est pas géomètre; c'est, comme vous savez, un grand défaut [...] *Pascal*

1.1 Probabilidad.

En cuanto a los cimientos de la probabilidad, que se remontan desde Cardano hasta Poisson se tratará de ser lo más fiel posible a los autores sin *contaminar* con ideas contemporáneas su trabajo, pero sí quizá vislumbrar hacia donde iba o *compararlo* con ideas modernas.

1.1.1 Cardano.

De Ludo Aleae.

Se estima que *De Ludo Aleae*[1] es el primer trabajo conocido sobre probabilidad, se cree que fue escrito en los 1500's, a pesar de que no haya sido publicado sino hasta 1663.

El libro de Cardano marca la forma en la que evolucionó el pensamiento probabilístico; todo es explicado por medio de ejemplos y un método difícil de comprender hoy en día principalmente por la ausencia de símbolos.

[1]Sobre juegos de azar. Existe al menos una traducción al inglés por Sydney Henry Goud en el libro del profesor Oystein Ore: Cardano. The Gambling Scholar

El libro está compuesto por 32 pequeños capítulos y se enuncia el *Principio Fundamental del Azar*:

" El principio más fundamental de todo el azar es estar simplemente igual de condiciones... de dinero, de situación... y del dado en sí. En medida que se quiere la equidad, si el azar está a favor de tu oponente, eres un tonto y si está al tuyo, eres injusto." [2]

Ésta es la base de su teoría con resultados dados; y es conveniente resaltar que él sólo consideraba eventos equiprobables; aún no tenía el concepto de evento que se tiene hoy.

En el capítulo noveno, introduce los conceptos: Circuito y Equidad. Con Circuito se refiere al número de posibles resultados elementales o simples, lo que en la teoría moderna/actual se conocería como tamaño de la muestra. Equidad parecería ser un concepto relacionado con el actual de esperanza y es la mitad del circuito.

Si se razona de la siguiente forma: p es la probabilidad, f la frecuencia de aparición y c el circuito: $p = \frac{f}{c}$; e la equidad: $e = \frac{c}{2}$, se observa que: $p_e = \frac{f}{e} = 2 \cdot \frac{f}{c} = 2p$. Ésta última es llamada la proporción de equidad y se relaciona con el concepto actual de esperanza de la siguiente manera:

Supóngase a dos jugadores con una suma apostada de A, por cada uno, en total $2A$. Entonces la esperanza (actual) es $E = p \cdot 2A = 2p \cdot A = p_e \cdot A$, convirtiendo a p_e en un factor natural para medir pérdida o ganancia. Aquí el motivo de su definición.

Uno de los grandes logros de Cardano es la siguiente fórmula:

$$p_n = p^n.$$

La probabilidad de obtener n éxitos en n repeticiones independientes en un experimento, sin olvidar que se basa en juego de dados. También tenemos una primera aproximación a lo que sería la *Ley de los grandes números*: cuando la probabilidad de un evento es p, para un número grande n de repeticiones, el

[2] Capítulo sexto de *De Ludo Aleae* de Cardano. También puede consultarse en *The Gambling Scholar* de Oystein Ore. pp. 189 Traducción propia.

número de veces que ocurrirá no deberá caer lejos del valor $m = np$.

La obra de Cardano permaneció en el olvido como una obra desconocida; en 1600 Galileo trabajó con tres dados pero sin hacer alusión a Cardano y fue hasta después de aparecer el trabajo de Pascal y Fermat que se reencontró.

1.1.2 Pascal - Fermat.

Cartas.

Pascal y Fermat entran en escena por una pregunta del *Chevalier du Méré*. Este último le pide a Pascal que determine cómo se debe repartir la bolsa de apuesta si un juego no ha terminado o bien no pudo terminar.

De todos los logros y/o alcances que juntos, Fermat y Pascal desarrollaron, que sobrevivieron al tiempo; se rescata lo siguiente:

Pascal señala dos enunciados sin demostración:[3]

1. Supóngase que cada jugador ha invertido una suma de dinero denotada por A. Sea $n+1$ el número de puntos en el juego y supóngase que el primer jugador ha ganado n puntos y el segundo ninguno. Si los jugadores acuerdan finalizar el juego sin jugar más, el primer jugador se debe de llevar $2A - \frac{A}{2^n}$.

2. Supóngase que la suma de dinero y el número de puntos como en (1). Supóngase que el primer jugador ha ganado un punto y el segundo ninguno. Si los jugadores dan por terminado el juego sin jugar más, el primer lugar es dueño de:

$$A + A\frac{1\cdot 3\cdot 5\cdots(2n-1)}{2\cdot 4\cdot 6\cdots 2n} = A\left(1+\frac{(2n-1)!!}{(2n)!!}\right).^{4} \tag{1.1}$$

[3]De hecho, en la correspondencia sólo se enuncian sus resultados, no se explica como llegaron a ellos ni se exponen sus demostraciones.

[4]$a!!$ se define de la siguiente manera: si a es impar como el producto de todos los impares menores o iguales a a y viceversa para los pares.

Uno bien podría quedar asombrado y desconcertado de cómo Pascal llega a singular cociente, pues no hace alusión a su método en sus cartas. Es hasta la publicación de su *Traité du triangle arithmétique* (1654) donde introduce lo que sería conocido como *El Triángulo de Pascal* a partir de que lo nombrara así De Moivre. La importancia de considerar este triángulo recae en la relevancia para la probabilidad clásica, de la combinatoria y este tratado como inicio formal de su estudio.

Triangulum Arithmeticum Pascalianum.[5]

En este tratado, Pascal define la construcción del triángulo aritmético. Deduce 19 consecuencias y un problema (como lo llama él). Si enumeramos las columnas y filas del triángulo de forma similar a las entradas de una matriz,[6] el problema es el siguiente:

Dados k y l, obtener una expresión para la celda $T_{k,l}$,[7] sin utilizar el triángulo ya armado.

Para esto se usa la consecuencia 12 que dice: *En todo triángulo aritmético, de dos celdas contiguas que están en la misma base, la superior es a la inferior como la multitud de celdas desde el superior hasta justo lo alto de la base es a la multitud de celdas desde el inferior hasta abajo inclusive.*[8]

Ésta llevada a una notación moderna sería

$$\frac{T_{r-1,n-r+2}}{T_{r,n-r+1}} = \frac{r}{n-r+1}.$$

Ahora para este problema aplicando la consecuencia varias veces se tiene:

$$\begin{gathered} k \cdot T_{k,l} = l \cdot T_{k-1,l+1} \\ (k-1) \cdot T_{k-1,l+1} = (l+1) \cdot T_{k-2,l+2} \\ \vdots \end{gathered}$$

[5]Nombrado así por primera vez por De Moivre en su *Miscellanea Analytica* (1730)

[6]Es decir, la entrada $T_{i,j}$ será la celda del renglón i y columna j.

[7]En palabras de Pascal: Dados los exponentes de los rangos perpendicular y paralelo de una celda, encontrar el número de la celda sin servirse del Triángulo Aritmético.

[8]Fragmento traducido por Jesús Basulto Santos en *La Geometría del azar. La correspondencia entre Fermat y Blaise Pascal.* Ver[2].

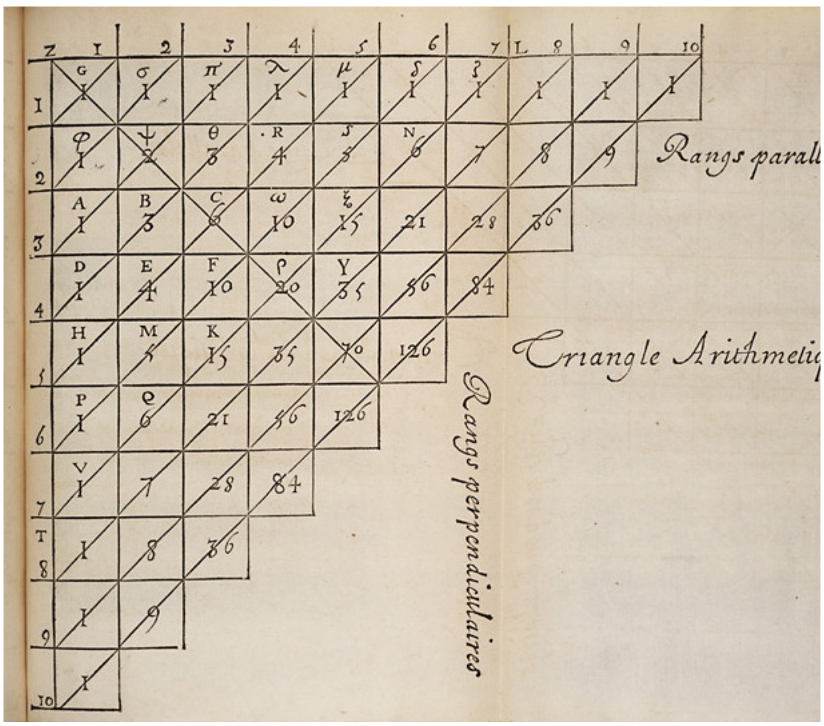

Figure 1.1: Triángulo aritmético de Pascal.

$$[k-(k-1)] \cdot T_{k-(k-1),l+k-1} = (l+k-1) \cdot T_{0,l+k}$$

Multiplicando, simplificando (teniendo en cuenta que $T_{0,k+l} = 1$) y despejando se tiene:

$$T_{k,l} = \frac{l \cdot (l+1) \cdot (l+2) \cdots (l+k-1)}{k \cdot (k-1) \cdot (k-2) \cdots 1}$$

y de manera análoga,

$$(k+1) \cdot T_{k+1,l-1} = (l-1) \cdot T_{k,l}$$
$$(k+2) \cdot T_{k+2,l-2} = (l-2) \cdot T_{k+1,l-1}$$
$$\vdots$$
$$(k+l-1) \cdot T_{k+l-1,l-(l-1)} = [l-(l-1)] \cdot T_{k+l-2,l-(l-2)}$$

Multiplicando, simplificando (teniendo en cuenta que $T_{k+l-1,1} = 1$) y despejando se tiene:

$$T_{k,l} = \frac{(k+1) \cdot (k+2) \cdots (k+l-1)}{(l-1) \cdot (l-2) \cdots 1}.$$

Lo anterior en notación moderna sería:

$$T_{k,l} = \binom{k+l-1}{l}.$$

Por otro lado, si se quiere dar una expresión para $T_{k,k+1}$, se sigue de la consecuencia 19 que dice:

$$T_{k,k+1} = 4 \cdot \frac{2k-1}{2k} T_{k-1,k}{}^{9}.$$

De donde se puede escribir como:

$$T_{k,k+1} = \left[2^2 \cdot \frac{2k-1}{2k}\right] \cdot \left[2^2 \cdot \frac{2(k-1)-1}{2(k-1)}\right] \cdot \left[2^2 \cdot \frac{2(k-2)-1}{2(k-2)}\right] \cdots$$
$$\cdots \left[2^2 \cdot \frac{2(k-(k-1))-1}{2(k-(k-1))}\right] \cdot T_{0,1}.$$

[9] En palabras de Pascal: En todo triángulo aritmético, dos celdas contiguas que están en la dividente, la inferior es a la superior tomada 4 veces como el exponente de la base de esta superior es a un número una unidad mayor (a dicho exponente).

teniendo en cuenta que $T_{0,1} = 1$ y simplificando:

$$T_{k,k+1} = 2^{2k}\frac{1\cdot 3\cdot 5\cdots(2k-1)}{2\cdot 4\cdot 6\cdots(2k)}.$$

Que es lo que enuncia Pascal en sus cartas.

Derivado de esto, he descubierto, de manera independiente, una igualdad: para toda $n \geq 1$:

$$\frac{1}{2^{2n}}\cdot\binom{2n}{n} = \frac{1\cdot 3\cdot 5\cdots(2n-1)}{2\cdot 4\cdot 6\cdots 2n} = \frac{(2n-1)!!}{(2n)!!}$$

Demostración:

$$\ln\left(\binom{2n}{n}\right) = \ln\left(\frac{(2n)!}{n!^2}\right) = \sum_{i=1}^{2n}\ln(i) - 2\sum_{i=1}^{n}\ln(i) = \sum_{i=1}^{n}\ln(2i) + \sum_{i=1}^{n}\ln(2i-1) - 2\sum_{i=1}^{n}\ln(i)$$

$$= \sum_{i=1}^{n}\ln(i) + \sum_{i=1}^{n}\ln(2) + \sum_{i=1}^{n}\ln(2i-1) - 2\sum_{i=1}^{n}\ln(i) = \sum_{i=1}^{n}\ln\left(2-\frac{1}{i}\right) + n\ln(2)$$

por otro lado,

$$\ln\left(\frac{\prod_{i=1}^{n}(2i-1)}{\prod_{i=1}^{n}(2i)}\right) = \sum_{i=1}^{n}\ln(2i-1) - \sum_{i=1}^{n}\ln(i) - n\ln(2) = \sum_{i=1}^{n}\ln\left(2-\frac{1}{i}\right) - \sum_{i=1}^{n}\ln(2)$$

además $\ln(2^{2n}) = 2n\ln(2)$, por lo cual

$$\ln\left(\frac{1}{2^{2n}}\cdot\binom{2n}{n}\right) = \sum_{i=1}^{n}\ln\left(2-\frac{1}{i}\right) + n\ln(2) - 2n\ln(2) = \sum_{i=1}^{n}\ln\left(2-\frac{1}{i}\right) - \sum_{i=1}^{n}\ln(2)$$

como se quería. [10]

Para toda $n \geq 1$:

$$\frac{1}{2^{2n}} \cdot \binom{2n}{n} = \frac{1 \cdot 3 \cdot 5 \cdots (2n-1)}{2 \cdot 4 \cdot 6 \cdots 2n}. \tag{1.2}$$

Pascal era un gran aritmético y lo resalta o bien lo deja ver en su trabajo aritmético-probabilístico, que no se limita a lo que se ha presentado aquí.

1.1.3 Jakob Bernoulli.

Para Jakob Bernoulli, la probabilidad es *the degree of certainty, and it differs from the latter as a part from the whole. Namely, if the integral and absolute certainty, which we designate by letter* α *or by unity* 1*, will be thought to consist, from example, of five probabilities, as thought of five parts, three of which favor the existence or realization of some event, whith the other ones, however, being against it, we will say that this event has* $\frac{3}{5}\alpha$*, or* $\frac{3}{5}$*, of certainty.*[11]

El nacimiento de la Ley de los Grandes Números.

Lo que se considera una de las mayores, si no es que la mayor aportación de Jakob Bernoulli es el resultado siguiente.

Let the number of fertile cases be to the numbers of sterile cases precisely or approximately as r *to* s*; or to the number of all cases as* r *to* $r+s$*, or as* r *to* t *so that this ratio is contained between the limits* $\frac{r+1}{t}$ *y* $\frac{r-1}{t}$*. It is required to show that it is possible to take such a number of experiments that it will be in any number of times (for example, in* c *times) more likely that the number of fertile observations will occur between these limits rather than beyond them, that is, that ratio of the number of fertile observations to the number of all of*

[10] También se conoce otra identidad: $\binom{2n}{n} = \sum_{i=0}^{n} \binom{n}{i}^2$

[11] un grado de certeza y difiere de la absoluta certeza como una parte difiere del todo. Si por ejemplo, la total y absoluta certeza - que se designa por a o por la unidad [12]- se supone compuesta por cinco probabilidades o partes, tres de las cuales son de la existencia o futura existencia de algún evento, las dos remanentes su no existencia, se dice que el evento tiene $\frac{3}{5}$ de certeza o bien $\frac{3}{5}\alpha$ de certeza.
Traducido de Bernoulli, J. *Ars Conjectandi.* Por Sheynin. Ver [4]

them will be not greater than $\frac{r+1}{t}$ *and not less than* $(\frac{r-1}{t})$.[13]

Donde nombra como *fertile case* (caso fértil) a los casos en el que algún evento puede ocurrir, y *sterile* (estéril) en los que el mismo evento no ocurre. Del mismo modo nombra *fertile* (fértil) a los experimentos en cuyos desarrollo aparece un caso fértil; e *infertile* o *sterile* (estéril) cuando no aparece algún fértil.

En notación moderna:

Teorema de Bernoulli. Sean r y s dos enteros positivos y sea $p = \frac{r}{r+s}$ y $t = r + s$. Para cualquier número positivo c se tiene:

$$\mathbb{P}\left(\left|\frac{s_n}{n} - p\right| \leq \frac{1}{t}\right) > \frac{c}{c+1}$$

para $n = kt$ suficientemente grande.[14] Donde s_n es el número de éxitos. Esto no termina siendo otra cosa sino *La Ley débil de los grandes números* para variables aleatorias Bernoulli.

Para más detalle o la demostración se puede consultar [18]. Más adelante se verá la importancia de no dejar pasar este resultado.

1.1.4 Laplace.

Para este momento el desarrollo de la Teoría de la Probabilidad se podría describir con una frase de *Siméon-Denis Poisson*:

Les géomètres du $XVII^{me}$ *siècle qui se sont occupés du calcul des probabilités, ne l'ont employé qu'à déterminer les chances de différents jeux de cette époque; et ce n'est que dans le siècle suivant qu'il a pris toute son extension, et qu'il est devenu une des principales branches des mathématiques, soit par le nombre et l'utilité de ses applications, soit par le genre d'analyse auquel il a*

[13] Bernoulli, J. *Ars conjectandi.* (1713)
Traducido al Inglés por Sheynin (2005). http://www.sheynin.de/

[14] Esto es para $k \geq k(r,s,c), k(s,r,c)$, donde $k(r,s,c)$ es el menor entero positivo satisfaciendo $k(r,s,t) \geq \frac{m(r+s+1)-s}{r+1}$, y m siendo el menor entero positivo que satisface $m \leq \frac{\ln(c(c-s))}{\ln((r+1)/r)}$.

donné naissance.[15]

Principios generales del cálculo de probabilidades.[16]

Entrado el siglo XIX , Pierre-Simon de Laplace enuncia en su *Essai philosophique sur les probabilités*[17] (1814) diez principios que han sido considerados de vital importancia reproducir a continuación, ya que después se hará hincapié en la presentación de la probabilidad en nuestros días y la gran diferencia de formas.

Primer principio. El primero de estos principios es la misma definición de probabilidad que es la razón entre el número de casos favorables y el de todos los casos posibles.

Segundo principio. Lo anterior supone igualmente posibles los diversos casos. Si no lo son, se determinarán primero sus posibilidades respectivas, cuya apreciación justa es uno de los puntos más delicados de la teoría del azar. Entonces, la probabilidad será la suma de las posibilidades de cada caso favorable. [18]

Tercer principio. Si los eventos son independientes unos de otros, la probabilidad de existencia de su conjunto es el producto de sus probabilidades particulares.

Cuarto principio. Cuando dos eventos dependen el uno del otro, la probabilidad del evento compuesto es el producto de la probabilidad del primero por la probabilidad de que, habiendo sucedido éste, tenga lugar el otro.

[15]Los geómetras [matemáticos] del siglo XVII que se han ocupado del cálculo de las probabilidades, sólo se han dedicado a determinar las posibilidades de diferentes juegos de la época y no es hasta el siglo siguiente que empieza a tomarse toda su extensión y convertirse en una de las principales ramas de las Matemáticas sea por su número y utilización de sus aplicaciones o bien por el género de análisis al cual da nacimiento.
Fragmeto extraido de: *Poisson.* Recherche sur la probabilité des jugements en matiére criminelle et en matière civile. Ver [37].

[16]Tomados de *Théorie Analytique des Probabilités. Volume I.* Ver [25]. Traducidos al Español por Pilar Castrillo. Ver [24].

[17]Ensayo filosófico sobre las probabilidades.

[18]Aquí corrige un error anteriormente hecho por J. D'Alambert en su artículo *Croix ou Pile de la Encyclopédie du XVIII^{eme} siècle o Dictionaire raisonné des sciences, des arts et des métiers.*

Quinto principio. Si se calculan las probabilidades *a priori* la probabilidad de un evento acaecido y la probabilidad de un evento compuesto de éste y de otro que se espera, la segunda probabilidad dividida por la primera constituirá la probabilidad del evento esperado, inferida del observado.

Sexto principio. Cada una de las causas a la que se puede atribuirse un acontecimiento observado se halla indicada con una verosimilitud tanto mayor cuanto más probable sea que ocurra el acontecimiento si se supone existente dicha causa. La probabilidad de la existencia de cualquiera de estas causas es, pues, una fracción cuyo numerador es la probabilidad del acontecimiento resultante de la causa en cuestión y cuyo denominador es la suma de las probabilidades semejantes relativas a todas las causas.

Aunque Laplace no hace referencia a Bayes, es este último quien habría expuesto este último principio con anterioridad en *Philiosophical Transactions* (1763), por esto lleva su nombre.

Séptimo principio. La probabilidad de un acontecimiento futuro es la suma de los productos de la probabilidad de cada causa, extraída del acontecimiento observado, por la probabilidad de que, en caso de que exista dicha causa, el acontecimiento futuro tenga lugar.

Laplace aquí introduce un término: Esperanza Matemática.

Le mot espérance a diverses acceptions: il exprime généralment l'advantage de celui qui attend un bien quelconque dans des suppositions qui ne sont probables. Cet avantage, dans la théorie des hasards, est le produit de la somme espérée par la probabilité de lóbtenir: c'est la somme partielle qui doit revenir lorsqu'on ne veut pas courir les risques de l'evénement, en supponsant que la répartition se fasse proportionnellement aux probabilités. Cette répartition est la seule équitable lorsqu'on fait abstraction de toutes circonstances étrangérères, parce qu'un égale degré de probabilité donne un droit égale sur la somme espérée.[19]

[19]La palabra es esperanza tiene distintas acepciones: en general expresa la ventaja del que espera un bien cualquiera dentro de suposiciones que son sólo probables. En la teoría del azar, esta ventaja es el producto de la suma esperada por la probabilidad de obtenerla: es la suma parcial que ha de ser restituida, cuando no se quieren correr los riesgos del evento, suponiendo que el reparto se haga proporcionalmente a las probabilidades. Este reparto, hecha abstracción de todas las circunstancias extrañas, es el único equitativo, ya que con

Debe hacerse notar que no es la definición actual de esperanza matemática. Es una primera aproximación a ella. Si se trata de ver con conceptos modernos, está pensando en una variable aleatoria que obtiene todo o nada; se gana o se pierde todo. Ahora enuncia el octavo principio.

Octavo principio. Cuando esta ventaja[20] depende de varios acontecimientos se le obtiene tomando la suma de los productos de la probabilidad de cada acontecimiento por el bien que se le confiere a su acaecimiento.

Noveno principio. En una serie de acontecimientos probables, de los cuales unos producen un beneficio y otros una pérdida, se obtendrá la ventaja resultante sumando los productos de la probabilidad de cada acontecimiento favorable por el beneficio que produce y restando de esta suma la de los productos de la probabilidad de cada acontecimiento desfavorable por la pérdida asignada a él. Si la segunda suma supera a la primera, el beneficio se convierte en pérdida y la esperanza se transforma en temor.

Décimo principio. El valor relativo de una suma infinitamente pequeña es igual al valor absoluto dividido por el bien total de la persona interesada.

Es importante señalar que Laplace no usa notación, todo lo deja redactado y por lo mismo ejemplifica cada uno de los principios.

1.1.5 Poisson.

Para Poisson había dos tipos de Probabilidad: la subjetiva y la objetiva, aunque él no las nombra como tal sino *chance*[21] y probabilidad. Se da un ejemplo acontinuación, pues no se encontró una definición puntual, sino una explicación. La subjetiva sería por ejemplo la probabilidad que llueva de mañana se estima en 0.10 o 10%, y la objetiva por ejemplo en un volado, sabemos que eventualmente se *alcanzará* la probabilidad $\frac{1}{2}$ de cara o sol. Es decir la objetiva se basa en conocimientos previos sólidos.

El concepto de cantidad aleatoria data del siglo XVII, teniendo sus orígenes

igual grado de probabilidad se tiene un derecho igual sobre la suma esperada. Traducción de Pilar Castrillo.

[20] La esperanza matemática. Nota propia.

[21] Que se podría traducir como oportunidad.

en Bernoulli y Huygens. Posteriormente para el siglo XVIII fue introducido el concepto para la Teoría de errores por Simpson, Lagrange, Laplace y Gauss. Sin embargo, no es hasta Poisson que lo introduce de una manera formal, dándole nombre aunque fue un término temporal dió un gran avance para la Teoría de la Probabilidad. Su definición es la siguiente:

Suppons actuellement qu'au lieu de deux événements possibles... il y en ait un nombre donné λ, dont un seul devra arriver à chaque épreuve. Ce cas est celui où l'on considère une chose A d'une nature quelconque, susceptible d'un nombre λ de valeurs, connues ou inconnes, que je représenterai par $a_1, a_2, \ldots, a_\lambda$, et parmi lesquelles une seule devra avoir lieu à chaque épreuve...[22]

Por otro lado también se le debe la primera formulación de la distribución acumulativa para una *cantidad aleatoria*:

$$F_n(x) = \mathbb{P}\{x_n < x\}.^{23}$$

Posteriormente definirá la función de densidad como la derivada de $F_n(x)$ en el caso de cantidades aleatorias continuas. Este concepto sería retomado por Davidov y Liapunov, aunque hasta el siglo XX tendría un uso esencial en la teoría.

Además trabajando con el Teorema de De Moivre-Laplace[24], Poisson pudo llegar a la distribución que hoy lleva su nombre: la probabilidad de ocurrencia de un evento F no más de n veces en $\mu = n + m$ eventos se aproxima a:

$$P \approx e^{-\omega}\left(1 + \omega + \frac{\omega^2}{2!} + \frac{\omega^3}{3!} + \ldots \frac{\omega^n}{n!}\right)$$

[22] *Supóngase dos eventos posibles (caso Bernoulli), y un número λ de pruebas al que le podrían llegar sólo un resultado a cada una. En este caso se considera una cosa A de cualquier naturaleza, suceptible de un número lambda de valores conocidos o desconocidos, que yo (Poisson) representaré por $a_1, a_2, \ldots, a_\lambda$, y para los cuales una sola deberá tener lugar en cada prueba.* Es una traducción lo más fiel posible, entendiendo la dificultad de la idea que expone Poisson y la claridad de lo que se entendería hoy en día por esto. Fragmento de [37].

[23] Donde n significa la observación n-ésima.

[24] En términos modernos: Si crece indefinidamente entonces para k en una vecindad de np, se puede aproximar por

$$\binom{n}{k} p^k q^{n-k} \approx \frac{1}{\sqrt{2\pi npq}} e^{\frac{-(k-np)^2}{2npq}} \text{ con } p + q = 1, p, q > 0$$

en el sentido que cuando n se va a infinito el cociente de ellos converge a uno.

donde ω sería la probabilidad de fallar las μ pruebas.

Pero, ¿cómo llegó a esto? Poisson orillado por la dificultad de calcular probabilidades de éxitos con la distribución Binomial debido a los factoriales enormes que pueden salir, encontró una expresión asintótica para p en

$$\mathbb{P}(S_n = k) = \frac{n!}{(n-k)!k!} p^k (1-p)^{(n-k)}$$

donde S_n denota el número de éxitos Bernoulli en n pruebas, con probabilidad de éxito p.

Esta aproximación es $p \approx \frac{\lambda}{n}$, con lo cual si se denota por $\lambda = np$ se obtiene:

$$\mathbb{P}(S_n = k) = \frac{n!}{(n-k)!k!} p^k (1-p)^k = \frac{n!}{(n-k)!k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}$$

$$\frac{n(n-1)\cdots(n-k+1)}{n^k} \frac{\lambda^k}{k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}$$

donde si n es grande y por tanto p pequeña se tiene

$$\frac{n(n-1)\cdots(n-k+1)}{n^k} \frac{\lambda^k}{k!} \approx 1, \quad \left(\frac{\lambda}{n}\right)^k \approx 1, \quad \left(1 - \frac{\lambda}{n}\right)^{n-k} \approx e^{\lambda}$$

De donde se concluye que para n grande y p pequeña

$$\mathbb{P}(S_n = k) = \frac{\lambda^k}{k!} e^{-\lambda}.$$

Su trabajo también se enfocó en La Ley de los Grandes Números, para sumas de cantidades aleatorias cuyas distribuciones no son idénticas y valores en un intervalo finito, como también para eventos *Poisson*. Trató de aplicar la Teoría de la Probabilidad al derecho, pero no tuvo eco ni función útil alguna más que organizar y tener estadísticas certeras de la situación en su país.

Posteriormente Poincaré sería un compilador de la Teoría de la Probabilidad. Quizá habría que rescatar que él ya contempla esta idea:

$$\mathbb{P}(x_1 < \xi < x_2) = \int_{x_1}^{x_2} \phi(x) dx$$

donde ξ es una variable aleatoria (concepto llegaría todavía más tarde, pero ya están las bases como se vió con Poisson) y $\phi(x)$ su función de densidad. Su libro *Calcul des Probabilités(1896)*, sería utilizado más como libro de texto, pues en realidad trabajó con cosas que no permitieron dar grandes avances en la teoría. Con esto pausamos el desarrollo de la Probabilidad, para dar paso al desarrollo de la Teoría de la Medida.

1.2 Teoría de la Medida.

1.2.1 Nociones de medida por Borel.

El origen de la teoría de la medida de Borel tiene como en casi todas las ramas de las matemáticas el origen menos esperado y éste se sitúa en el Análisis Complejo.

En 1851 Riemann puso en la mesa una pregunta: ¿la clase de funciones analíticas es idéntica a la clase de funciones de variable compleja z que son definidas al aplicar a z operaciones elementales como suma, resta, multiplicación y división, un número finito o infinito de veces? Weierstrass mostró que no son idénticas en una de sus pocas publicaciones. Consideró las funciones:

$$f(z) = \sum_{n=0}^{\infty} \frac{1}{z^n + z^{-n}}.$$

Ésta representa dos funciones para $|z| > 1$ y $|z| < 1$. Es no acotada para cualquier vecindad de un punto z_0 con $|z_0| = 1$ por lo cual esta función no puede ser analíticamente continuada dentro y fuera del disco unitario.

Para 1882, Paul Appell, alumno de Charles Hermite, construyó más ejemplos exhibiendo el mismo comportamiento que el ejemplo de Weierstrass, estos eran de la forma $\sum \frac{A_n}{(z-a_n)^{m_n}}$. Otro estudiante de Hermite, Henri Poincaré buscó más ejemplos aún. Tomó el plano complejo y lo dividió en dos: el interior y el exterior de un contorno convexo que nombró S, T y C respectivamente.

Suponiendo que C posee una tangente y radio de curvatura en cada punto, se tiene que para cada punto z en T existe un círculo con centro en z tangente a C y que cae completamente fuera de S. Poincaré definió la función f sobre T como

$$f(z) = \sum_{n=0}^{\infty} \frac{A_n}{z - b_n} \tag{1.3}$$

donde la serie $\sum_{n=0}^{\infty} |A_n|$ converge y los puntos b_n se supone que forman parte del conjunto $C \cup S$ siendo denso dondequiera en C. La función es analítica en T y tiene la propiedad, debido a la densidad de b_n en C, que su expansión en serie de potencias para cada z en T tiene su círculo tangente de convergencia. Por lo cual f no puede ser continuada de manera analítica pasando por C,

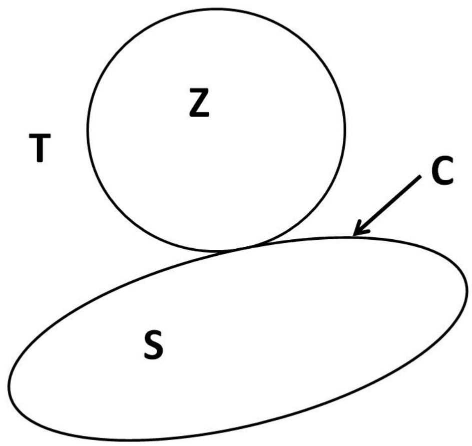

Figure 1.2: Diagrama 1.

posee un *espace lacunaire*[25] S.

Tuvieron que pasar 12 años, para que Borel, en su tesis doctoral propusiera:

Es posible para estas series, en ciertos casos, dar una definición de continuidad analítica *más allá de una linea cerrada esencialmente singular que no cree contradicción consigo misma y nociones previas.*[26]

Borel considera series de la forma $\frac{A_n}{(z-a_n)^{m_n}}$, donde m_n denota enteros acotados por un entero fijo (para eliminar el caso de Appell) y que la serie $\sum |A_n|$ converge. Se supone que la cerradura del conjunto que contiene las a_n consiste de curvas y puntos aislados solamente.

Restringiéndose al trabajo de Poincaré, así como lo hizo Borel y suponiendo que el caso en los que los puntos a_n forman un subconjunto denso de C; Borel probó bajo estas condiciones que la función (1.3) poseé varias propiedades propias de las funciones analíticas. Pero lo más importante de Borel fue descubrir que cuando $\sum |A_n|^{\frac{1}{2}}$ converge, cualquier punto en T puede ser unido con cualquier punto en S por un arco de circunferencia en el cual la serie (1.3) converge absoluta y uniformemente con $m_n = 1$ a pesar de que cruce C. Es decir, la función (1.3) puede ser continuada analíticamente a S cruzando C.

[25] Espacio poroso, traducción literal, o bien, espacio con lagunas.
[26] Borel 1894, publicado en [11]

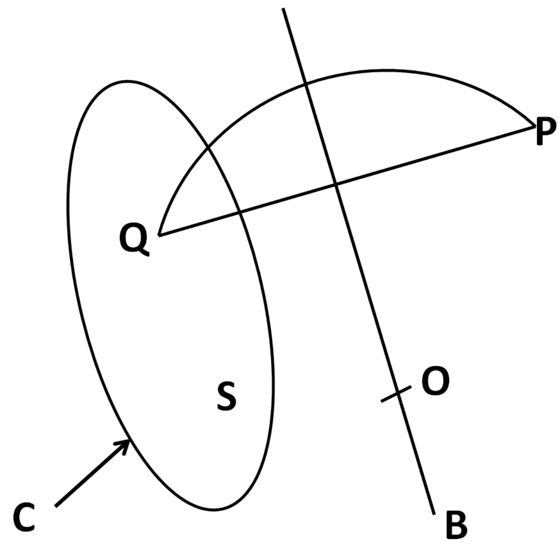

Figure 1.3: Driagrama 2.

De la demostración de Borel se puede intuir la importancia de una definición de medida que emplea un número infinito de intervalos que cubren el conjunto a medir. A grandes rasgos lo que desarrolla en su demostración es lo siguiente: [27]

Sea P un punto en la región T y Q un punto en la región S, donde S, T, C son las regiones descritas con anterioridad.[28] Sea $\overline{AB}$ cualquier segmento perpendicular y bisector del segmento $\overline{PQ}$ y sea O cualquier punto sobre $\overline{AB}$. Cada O determina un círculo que pasa por P, Q. Considérese uno de los arcos PQ. Podría pasar que el arco pase por uno de los puntos a_n de la curva C.

Supóngase lo peor: para cada n, los puntos P, Q y a_n determinan un círculo con centro O_n sobre $\overline{AB}$. Sea L la longitud de $\overline{AB}$. La hipótesis de que $\sum |A_n|^{\frac{1}{2}}$ converja implica la existencia de una $\sum u_n$ convergente de términos positivos, tal que $\sum \frac{|A_n|}{u_n}$ también converge. Por lo anterior un número N puede ser escogido de tal manera que $\sum_{n=N+1}^{\infty} u_n < \frac{1}{2}L$. Para $n > N$ constrúyase un intervalo I_n sobre $\overline{AB}$ con centro en O_n y de longitud $2u_n$. La suma total de los intervalos será $2\sum_{n=N+1}^{\infty} u_n$ por lo cual es menor a L. De este hecho, Borel fue capaz de deducir la existencia de una cantidad no-numerable de puntos de

[27]Para más detalle consultar: Borel, Emile. Sur quelques points de la théorie des fonctions. 1894

[28]No hay que olvidar que todo se encierra sobre el trabajo de Poincaré.

$\overline{AB}$ fuera de todo I_n.

Entonces existe un punto W sobre $\overline{AB}$ que no pertenece a ningún I_n y mucho menos es algún O_n $\forall n \in \{1, 2, \ldots, N\}$. Consecuentemente, el círculo que tiene a W por centro y pasa por P, Q, no contiene a_n alguno. Usando este hecho, Borel demuestra que las series (1.3) convergen absoluta y uniformemente en el círculo.

Capítulo 2

Primeros intentos por axiomatizar la probabilidad

Je le vois, mais je ne le crois pas! *Cantor*

2.1 Über eine Wahrscheinlichkeitsaufgabe bei kettenbruchenwicklungen. Wiman.

Hasta este momento la probabilidad ha sido vista simplemente para la Matemática Aplicada y en conexión con la Física Estadística. Los conceptos y métodos estaban diseñados para lo mismo; la teoría no ha sido completamente general. Pero, al cambio de siglo las cosas cambiarían: nuevas conexiones entre la matemática pura y la probabilidad se entretejerán por cambios que sufrirán las dos.

Aparece como se ha comentado, la Teoría de la Medida de Borel y Lebesgue y con ello el estudio de conjuntos de números reales, funciones reales y propiedades asintóticas de sucesiones de números reales que serían relacionadas a la identificación de un número real en su expansión decimal. Por lo cual los problemas probabilísticos de limitar el comportamiento de frecuencias relativas, por ejemplo, podría ser formulado como un problema de medida de conjuntos de números reales.

El primer intento se le debe al astrónomo Hugo Gyldén que en 1888 abrió una pregunta sobre la distribución de los enteros en lo que se conoce como la expansión de un número real dado. Es decir, ¿cómo se distribuyen los enteros a_i en la siguiente expresión (2.1)?:

$$a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cfrac{1}{a_4 + \dots}}}}\text{[1]} \tag{2.1}$$

Gyldén observó, sin dar argumento matemático alguno, que la distribución de los a_i podría ser de cierta manera regular para números irracionales escogidos de manera aleatoria. Por lo cual, se podría plantear el problema como ¿Cuál es la distribución límite de los a_i? Pero como se espera de la época, esta pregunta está motivada por algo fuera de la matemática aplicada, en este caso la Astronomía. Gyldén trabajaba en perturbaciones planetarias y necesitaba de las fracciones continuas para aproximar un número real a través de números racionales (manejables). Su intuición astrónoma le decía que la pro-babilidad que $a_n = k$ sería inversamente proporcional a k. Asume que no tienen ningún patrón que pareciera que están tomados al azar. Primero asume una distribución uniforme y da probabilidades de que rebasen un número. Posteriormente da un radio esperado donde se colocaría.

Anuncia una *completa concordancia* entre sus resultados teóricos y las mediciones y posteriormente este problema se desprende de sus raíces aplicadas y se vuelve parte de la naciente teoría probabilística. El mismo año, aproxima la distribución de los valores $a_n = 1, a_n = 2\dots$ en una fracción continua como (2.1) que representa un irracional tomado al azar.[2] Para Gyldén esto toma la cara de un sistema de urnas, y para "probar" sus enunciados con la real distribución de las a_n's toma diversos datos entre los promedios de las mediciones astronómicas, logaritmos escogidos al azar y números escogidos por una persona.[3] Esto le concuerda tanto para él que lo asume como *aproximadamente correcto.*

La medida de Émile Borel dió las bases para las intuiciones de Gyldén. El

[1]Sobre como se obtiene una fracción continua de un número real cualquiera, se puede consultar el anexo de este trabajo.
[2]Ver [35].
[3]Aunque números escogidos por una persona suelen no ser (tan) aleatorios.

primer matemático en sacar provecho de esto fue Torsten Bróden. Posteriormente su seguidor Anders Wiman trabajó en esto. Este último da finalmente una exacta determinación de la distribución límite de las a_n's conforme n crece, bajo el supuesto de una distribución uniforme sobre el intervalo unitario. Aunque no parece tan evidente a primera vista, Wiman ya usa Teoría de la Medida. Véase por qué.

Para $x \in [0, 1]$, $a_1 = k$ ocurre cuando $x \in [\frac{1}{k+1}, \frac{1}{k}]$, cuya probabilidad sería la longitud del intervalo, es decir su medida según Borel, es decir $\frac{1}{k(k+1)}$. La probabilidad que $a_2 = k$ será la suma de los intervalos ajenos correspondientes a: $a_1 = 1$ y $a_2 = k$, $a_1 = 2$ y $a_2 = k$, $a_1 = 3$ y $a_2 = k$ así sucesivamente:

$$\sum_{i=1}^{\infty} \left| \frac{k}{k+i} - \frac{k+1}{ik+i+1} \right| .$$

Similarmente para la probabilidad de $a_n = k$. La correcta ley de probabilidad de $a_n = k$ para cuando n tiende a infinito puede ser ahora determinada, según Wiman por:

$$\frac{1}{\log 2} \left(\frac{1 + \frac{1}{k}}{1 + \frac{1}{k+1}} \right) .$$

Esto fue resuelto por Brodén que responde a su vez Wiman, quien critica a Brodén de no separarse de las nociones de valor medio y valor esperado. Brodén y Wiman serían matemáticos al día en los avances continentales en teoría de funciones y despúes tendrían un contacto con Borel. Brodén haría la primera sugerencia de la historia en juntar la Teoría de la Medida y la Teoría de la Probabilidad, aunque Wiman a la par ya hacía uso de ella, de una manera un poco imperceptible. Al siguiente año fue más explícito:

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.[4]

[4]Traducción al inglés del alemán por Jan von Plato: Creating Modern Probability, pp. 31-32. Ver [35].

Totalmente una predicción, se verá más adelante el sustento de esta afirmación.

2.2 Los 23 problemas (del siglo XX) de Hilbert.

Hilbert quizá fue el último del conjunto de matemáticos que tuvieron un amplio alcance y dominio del estado de las matemáticas de su tiempo. En 1897 Poincaré, el único matemático contemporáneo con el que se podría *comparar* a Hilbert en conocimiento respecto a su amplitud y variedad, habría escrito una charla central que fue leída en el congreso de Zurich sobre las relaciones del análisis puro y la física matemática. Hilbert inicialmente pensó en referirse en las ideas de Poincaré para el congreso que se celebró en París en el año de 1900. Pero cambió de opinión por Minkowski; éste último lo disuadió a pensar más allá, y hablaría sobre el quehacer de los matemáticos en el futuro.

Hilbert trató entonces de delinear un plan de trabajo para los años venideros que no se limitara a una rama en particular de las matemáticas. Delinearía una lista de problemas, los cuales no debieran ser del todo inaccesibles ya que de lo contrario desistirían rápidamente. En la plática Hilbert sólo enunciaría diez problemas de los veintitrés que aparecerían en la versión escrita que se haría popular entre la comunidad matemática mundial. Estos se pueden agrupar en cuatro amplias categorías:

1. Fundamentos (Análisis, Geometría, Física) - problemas del 1 al 6 y 18.
2. Teoría de Números - problemas del 7 al 12.
3. Álgebra (Invariantes y Geometría Algebráica) - problemas del 13 al 17.
4. Análisis (Cálculo Variacional y Análisis Complejo) problemas del 19 al 23.

Es precísamente en el problema sexto: *Tratado Matemático de los axiomas de la Física* donde se deseaba principalmente axiomatizar la Mecánica y la Probabilidad, donde la matemática tiene una gran injerencia.

2.3 Algunos intentos por axiomatizar la Probabilidad después de 1900.

Dos primeras respuestas al *sexto problema de Hilbert* aparecen en los próximos años.[5] El primero sería de Rudolf Laemmel en 1904, donde basa su intento en fundar la Probabilidad en Teoría de Conjuntos y de la Medida. Aunque en realidad terminó fundamentándola en la Teoría de Conjuntos y muy rudimentáriamente en la medida.

La otra axiomatización la daría un discípulo de Hilbert: Ugo Broggi en 1907, con lo más reciente en la Teoría de la Medida del momento: la medida de Borel y la de Lebesgue. Propone un sistema siguiendo el modelo de la Geometría de Hilbert para probar: consistencia, completez e independencia entre los axiomas. Siguiendo la idea de Zermelo: se puede decidir si un elemento en el conjunto tiene la propiedad A o simplemente no la tiene; y usando el "hecho" que se puede escoger una $m \in M$, M siendo un conjunto cualquiera, sin decir nada de las propiedades de dicho m. Propone la probabilidad como la medida del conjunto de las m que cumplen alguna propiedad específica. Aunque sus axiomas son generales:

1. La probabilidad es una función no negativa.
2. La probabilidad de la certeza es uno.
3. Se cumple la aditividad[6].

Broggi asume que la aditividad numerable sería consecuencia de la finita, también afirma que la axiomatización está completa, que ninguna otra clasificación de axiomas la completará y parece dar por hecho que la probabilidad aplicará para elementos en el intervalo unitario o su $n-$generalización; más aún que extensiones iguales serán equiprobables. von Plato asegura para este

[5]Parece que se han extraviado dichos trabajos, no se encontraron en una búsqueda exhaustiva pero von Plato en [35] comenta que se puede consultarse parte de ellos en Schneider, I *Die Entwicklung der Wahrscheinlichkeitstheorie von den Anfangen bis 1933.* Einfuhrungen und texte, Wissenchaftliche Buchgesellschaft, Darmstadt. (1988.)

[6]Es decir, la probabilidad del conjunto A (de las m que cumplen la propiedad 1) y la del conjunto B (de las m que cumplen la propiedad 2), siendo A y B ajenos es la probabilidad de $C = A \cup B$.

sistema de axiomas: the categoricity property becomes the same as the uniqueness of Lebesgue measure, it seems. (sic) en [35] pp. 33.

Para el caso en que M sea numerable, comenta que se deben ordenar sus elementos primero; y define la probabilidad[7] como el límite, asumiendo que exis-ta, de las frecuencias relativas en cada uno de los subconjuntos de la cadena creciente que se forma al ordenar los elementos de M.

Esta idea de tomar como probabilidad el conjunto que forma las x con tal propiedad o medir el conjunto de pertenencia, como se verá más adelante no sería nuevo para estos exponentes. Años antes, en la física ya se usaba por Maxwell y Boltzmann.

[7] von Plato no especifíca de qué. Por el contexto podría entenderse como la probabilidad de que $x \in M$ tenga la propiedad A.

Capítulo 3

Probabilité pour Émile Borel.

God made the integers; all the rest is the work of man. *Leopold Kronecker*

3.1 Remarques sur certains questiones de probabilité.

En 1905 Émile Borel publicó *Remarques sur certains questiones de probabilité* donde principalmente muestra que la integral de Riemman no es suficiente para lo que viene.

El primer intento de Borel hacia su nueva teoría será observar la probabilidad como proporcional a una extensión: ya sea longitud, área, volumen; depende del número de las dimensiones en las que se trabaje. Por lo cual, en su artículo *Remarques sur certains questiones de probabilité* la definición de probabilidad se relaciona al valor medio: para $f : [x_0, x_1] \subsetneq \mathbb{R} \to \mathbb{R}$ se tiene:

$$\frac{1}{x_1 - x_0} \int_{x_0}^{x_1} f(x)dx$$

como valor medio y sus respectivas generalizaciones.

Si escogemos un número entre cero y uno ¿cuál es la probabilidad de que dicho número sea conmensurable? [1] Consideremos la función $f : [0, 1] \subset \mathbb{R} \to \mathbb{R}$

[1]Borel se refiere con conmensurable (con la unidad) a los racionales, como eran conocidos con anterioridad para poder definir posteriormente otra conmensurabilidad.

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

Dicha función es equivalente a la pregunta planteada, tomando uno en caso favorable y cero en caso contrario.

La respuesta es evidentemente cero.[2] Sean $f(x)$ y $F(x)$ dos funciones:

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

$$F(x) = \begin{cases} 1 & \text{si } x \text{ es irracional} \\ 0 & \text{si } x \text{ es racional} \end{cases}$$

si se tratara de calcular las integrales en el sentido exclusivamente clásico

$$\int_0^1 f(x)dx, \qquad \int_0^1 F(x)dx$$

esto no tendría el más mínimo sentido debido a que en el sentido de Darboux la integral inferior es cero y la superior es igual a uno para ambos casos.

Cualquiera que sea la convención que se quiera adoptar, si ésta no depende del valor medio de las integrales anteriores calculadas en el sentido clásico, se deberá atribuir el mismo valor medio a ambas funciones. Pero debido a que el valor medio de $f(x)$ es cero y el de $F(x)$ es uno, la anterior conclusión es absurda.

Ahora usando la (nueva) definición de integral de Lebesgue, se reconoce que ambas funciones son L-integrables mostrando que los métodos de Lebesgue permiten estudiar cuestiones de probabilidad que parecían inaccesibles por los procedimientos clásicos de integración.

Borel hace énfasis en que Wiman ha sido el primero, a su conocimiento, en aplicar la teoría de conjuntos medibles al cálculo de probabilidades.

[2]Para más detalle ver Poincaré: Calcul des probabilités p.126. Ver [36].

Si consideramos $x \in [0,1]$ y un conjunto medible $E \subseteq [0,1]$ de medida m. Por definición la probabilidad de que x esté en E es m y que no esté es $1-m$. Si $E = \mathbb{Q} \cap [0,1]$, se obtendría de inmediato que $\lambda(E) = 0$ reforzando lo enunciado anteriormente.

Ahora, Borel trata un caso poco más complicado. Se dirá que un número $\alpha \in [0,1]$ es *conmensurable de grado n con la unidad* si existen dos números primos relativos p y q tales que:

$$\left|\frac{p}{q} - \alpha\right| < \frac{1}{q^n},\ q \geq 2,\ p < q \tag{3.1}$$

¿Cuál es la probabilidad de que $x \in [0,1]$ *sea conmensurable de grado n con la unidad?*

A cada fracción irreductible $\frac{p}{q}$ se 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 la longitud total de los intervalos análogos para una q dada será $2\frac{\phi(q)}{q^n}$ donde $\phi(q)$ denota el número de primos relativos menores a q.[3]

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 (evidentemente) medible [4] y su medida e_n es inferior a

$$m_n = 2\sum_{q=2}^{\infty} \frac{\phi(q)}{q^n}$$

que sería la suma de las medidas de todos los intervalos.

Dicha serie es convergente para $n > 2$, y para $n = 2$, se sabe en efecto, que cualquier número α cumple un número infinito de maneras la desigualdad

[3]Es fácil ver que los intervalos $E_{p,q}^{(n)}$ para q fija y p variando entre los primos relativos menores a q no se intersectan y cada uno de ellos tiene longitud $\frac{2}{q^n}$.

[4]Unión numerable de medibles es medible. Se tiene la impresión que Borel sólo puede justificarlo para conjuntos ajenos.

(3.1). Tómese entonces $n \geq 3$, la serie m_n es convergente y más aún a un valor menor que uno, puesto que:

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

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

Si convenimos que para cualquier α inconmensurable que cumpla la desigualdad (3.1) para varios sistemas de valores p y q y le damos ese número por coeficiente podremos decir que la probabilidad de que x sea conmensurable de orden a la unidad es precisamente m_n; pero si no hacemos la convención anterior la probabilidad es $e_n < m_n$. Borel afirma que se puede calcular a e_n con tanta aproximación como se quiera, no pudiéndose en términos finitos, al menos de números algebraicos o de trascendentales usuales. La probabilidad de que x no sea conmensurable de grado n con la unidad es $1 - e_n$.

3.2 Les probabilités dénombrables et leurs applications artithmétiques.

Sería hasta 1909 cuando Borel retoma lo anterior escribiendo lo siguiente:

On distingue généralmente, dans le problèmes de probabilitès, deux catégories principales, suivant que le nombre des cas posibles est fini ou infini: ou la première catégorie constitue ce qu'on appelle les probabilités discontinues*, ou probabilités dans le domaine du discontinu, tandis que la seconde catégorie comprend les* probabilités continues *ou* probabilités géométriques*. Une telle clasificacition apparaît comme incomplète, lorsqu'on se reporte aux résultats acquis dans la théorie des ensembles; entre la puissance des ensembles 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.[5]

[5]Por lo general se distingue, en problemas probabilísticos, dos categorías principales, de acuerdo al número posible de casos, finito o infinito: la primera categoría constituye lo que se llama *probabilidades discontinuas*, o probabilidad en un dominio discontinuo. Mientras que la segunda comprende de las *probabilidades continuas* o *geométricas*. Esta forma de clasificación parece incompleta cuando uno se remonta a los resultados obtenidos en teoría

Borel probó que la aditividad numerable de su medida es consecuencia de su definición de medida. Y será en este artículo, *Les probabilités dénombrables et leurs applications artithmétiques (1909)* donde se puede empezar a vislumbrar la liga entre la medida de Borel y la teoría de la cual sería un parteaguas: Probabilidad. Apareció publicado en *Rendiconti del Circolo Matematico di Palermo*, una revista italiana.

Para este momento los problemas de probabilidad se encuentran divididos en dos grandes categorías basadas en el número de casos posibles: probabilidades discontinuas y probabilidades continuas o geométricas y Borel decide definir una tercera: *les probabilités dénombrables* o bien, probabilidades numerables.

3.2.1 Probabilidades numerables

Borel afirma que muchos analistas habrían pasado por alto los conjuntos numerables tomando en cuenta sólo los finitos y los continuos, sin percatarse que los continuos sólo son el camino de estudio a lo numerable.

Se dividen en tres categorías los problemas de *probabilidad numerable*:

1. El número de casos posibles es finito en cada evento, pero el número de pruebas es infinito numerable.[6]

2. El número de casos posibles es infinito numerable y el número de pruebas es finito.

3. El número de casos posibles y el número de pruebas es infinito numerable.

Dichas distinciones serán de gran utilidad en el punto de vista lógico, y no sería difícil tratarlas simultáneamente, sin embargo, llevarían más desventajas que ventajas. Aquí se puede ver algo: Borel ya intuye que la Teoría de la Pro-

de conjuntos; entre la cardinalidad de conjuntos finitos y la cardinalidad del continuo entra la cardinalidad de los conjuntos numerables. Yo propongo mostrar a grandes rasgos el interés que se liga a las cuestiones de probabilidad en la que tales conjuntos intervienen; yo les llamaré *probabilidades numerables.*

[6]Entendiendo como prueba, ensayo.

babilidad calcará de la Teoría de la Medida: la aditividad y la unión numerable de eventos posibles seguirá acotada por uno, es decir, la probabilidad total.

Primera Categoría. Primeramente se ocupará de la primera categoría, tomando como casos posibles simplemente dos, a saber, caso favorable y desfavorable.

Condiciones inciales: Se enumeran los eventos conforme a los números naturales y se toma como probabilidad p_n al caso favorable y q_n al caso desfavorable en el evento n.

Problema I. ¿Cuál es la probabilidad de que el caso favorable jamás se produzca?[7]

Si nombramos esta probabilidad A_0 entonces por el principio de probabilidades compuestas se tiene:

$$A_0 = (1 - p_1)(1 - p_2) \dots (1 - p_n) \dots = \prod_{i=1}^{\infty}(1 - p_i). \tag{3.2}$$

Se debe notar que el producto numerable es convergente. Borel lo demuestra copiando lo deseable en el producto: $\mathbb{P}(\cap_1^\infty B_i) = \prod_1^\infty \mathbb{P}(B_i)$ dado que se da $\mathbb{P}(\cup_1^\infty B_i) = \sum_1^\infty \mathbb{P}(B_i)$ para eventos independientes y ajenos en ese orden.

Más aún, tiene implícita la *continuidad de la medida de probabilidad* pues asume $\lim_{n\to\infty} \mathbb{P}\left(\cap_1^n B_i\right) = \mathbb{P}(\cap_i^\infty B_i)$.

Para demostrarlo, tómese en cuenta

$$\sum_{i=1}^{\infty} p_i = p_1 + p_2 + \dots p_n + \dots \tag{3.3}$$

y se afirma: $0 < A_0 < \infty$ si $\sum_{i=1}^{\infty} p_i < \infty$.

Para el caso en que la serie (3.3) sea divergente, Borel argumenta que si se limita a los primeros n productos en A_0 estos se irán sin remedio a cero conforme n crezca pues $(1 - p_i)$ se hará notablemente pequeño. Y habrá que señalar que contrario a las probabilidades discontinuas (discretas), en las continuas que un evento tenga probabilidad cero no significa que sea improbable

[7] Traducción de *produise.*

de ocurrir.

Para demostrar lo anterior, se dividirá en tres partes:

1. Convergencia de (3.2).

2. Si
$$\sum_{i=1}^{\infty} p_i < \infty \Rightarrow A_0 > 0 \tag{3.4}$$

3. Si
$$\sum_{i=1}^{\infty} p_i = \infty \Rightarrow A_0 = 0 \tag{3.5}$$

Para la convergencia de (3.2) primero hay que resaltar algunos puntos para los casos $p_k = 0, 1$ para alguna $k \in \mathbb{N}$. Es evidente que el producto se irá inmediatamente a cero si $p_k = 1$ y no contradice nada, pues querría decir que la prueba k sería un evento favorable, es decir, no ocurre que nunca haya casos favorables. Si $p_k = 0$ (p.a. k) no hay mayor problema pues no altera el producto. En el caso de que todas las $p_k = 0$ se tendrá la seguridad de que el producto es igual a uno, pues todo evento favorable tendrá probabilidad cero.

Por lo anterior, consideremos sólo el producto de las p_i's $\neq 0, 1$. Sin pérdida de generalidad, podemos seguir ocupando la misma A_0, sin olvidar la convención. Ahora, como el producto constituye por si mismo una sucesión estrictamente decreciente acotada que tiene a cero como ínfimo:

$$0 \leq \cdots < \prod_{i=1}^{n}(1-p_i) < \prod_{i=1}^{n-1}(1-p_i) < \prod_{i=1}^{n-2}(1-p_i) < \cdots < \prod_{i=1}^{2}(1-p_i) < (1-p_1)$$

entonces es convergente.

Para la segunda parte, (3.4), primero nótese que

$$\prod_{i=m}^{n}(1-p_i) = 1-\sum_{i=m}^{n} p_i + \sum_{m\leq i_1<i_2\leq m} p_{i_1}p_{i_2} - \sum_{m\leq i_1<i_2<i_3\leq m} p_{i_1}p_{i_2}p_{i_3} + \cdots + (-1)^{n-m-1}\prod_{i=m}^{n} p_i$$

véase que la primera suma tiene $n-m$ sumandos, la segunda $\binom{n-m}{2}$, la tercera $\binom{n-m}{3}$, así hasta la última suma tendrá $\binom{n-m}{(n-m)-1}$ por lo que en total se tendrán $Z = \sum_{i=1}^{n-m-1} \binom{m-m}{i}$ sumandos en el producto A_0.

Por otro lado, como $\sum_{i=1}^{\infty} p_i < \infty$ implica que es de Cauchy, es decir,

$$\forall \epsilon : 1 > \epsilon > 0 \, \exists \, N \in \mathbb{Z} \text{ tal que si } n, m > N \text{ entonces } \sum_{i=m}^{n} p_i < \frac{1-\epsilon}{M}, \text{donde } M = Z-1$$

además

$$\sum_{i=m}^{n} p_i \geq \begin{cases} \sum_{m \leq i_1 < i_2 \leq n} p_{i_1} p_{i_2} \\ \sum_{m \leq i_1 < i_2 < i_3 \leq n} p_{i_1} p_{i_2} p_{i_3} \\ \dots \\ (-1)^{n-m+1} \prod_{i=m}^{n} p_i \end{cases}$$

y entonces

$$\prod_{i=m}^{n} (1-p_i) \geq 1 - M \sum_{i=m}^{n} p_i > 1 - (M) \frac{1-\epsilon}{M} = \epsilon > 0$$

es decir, lo que se quería.

Para la tercera parte, (3.5), nótese que $0 \leq p_n \leq 1$ entonces $\frac{1}{p_n} \geq 1$ y $0 \leq 1 - p_n \leq 1$, por lo cual $1 - p_n \leq \frac{1-p_n}{p_n} \leq \frac{1}{p_n}$. De lo anterior se sigue:

$$\ln\left(\prod_{n=1}^{\infty} (1-p_n)\right) = \sum_{n=1}^{\infty} \ln(1-p_n) \leq \sum_{n=1}^{\infty} \ln\left(\frac{1}{p_n}\right) = -\sum_{n=1}^{\infty} \ln(p_n) \leq -\sum_{n=1}^{\infty} p_n \to -\infty$$

En conclusión:

$$\prod_{n=1}^{\infty} (1-p_n) \leq e^{-\sum_{n=1}^{\infty} p_n} \to 0.$$

Problema II. ¿Cuál es la probabilidad de que el caso favorable se produzca exactamente k veces? Denotaremos a esta probabilidad como A_k.

Si nombramos a $\omega_1$la probabilidad de que el evento favorable sólo se produzca en la primera prueba esta será

$$\omega_1 = p_1(1-p_2)(1-p_3)\cdots(1-p_j)\cdots . \tag{3.6}$$

En el caso de que (3.3) sea divergente (3.6) es cero; en el caso de la convergencia por (3.2)

$$\omega_n = \frac{p_1}{1-p_1}A_0$$

de la misma manera

$$\omega_n = (1-p_1)(1-p_2)\cdots(1-p_{n-1})p_n(1-p_{n+1})\cdots$$

sería la probabilidad de que el evento favorable se produzca exactamente en la prueba n-ésima e igual a cero en caso de que (3.3) sea divergente y

$$\omega_n = \frac{p_n}{1-p_n}A_0$$

en caso convergente.

Por el principio de probabilidad total se obtiene (en el caso de convergencia)

$$A_1 = A_0\left(\frac{p_1}{1-p_1} + \frac{p_2}{1-p_2} + \cdots + \frac{p_n}{1-p_n} + \cdots\right)$$

y la serie que multiplica a A_0 es visiblemente convergente, por lo cual si nombramos $u_n = \frac{p_n}{1-p_n}$ entonces

$$A_1 = A_0 \cdot \sum_n u_n. \tag{3.7}$$

Para extender esto último al caso divergente 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 sin querer decir imposibilidad. Borel en su artículo argumenta que se deberá tener precaución al afirmar que la probabilidad total es cero, es decir, duda de la aditividad numerable de la probabilidad siendo todas nulas y prefiere justificarlo de la forma siguiente.

Si se nombra

$$\sigma_n = (1 - p_1)(1 - p_2) \cdots (1 - p_n)(u_1 + u_2 + \cdots + u_n)$$

la probabilidad de que sólo se produzca un evento favorable en las primeras n pruebas será nula; pues teniendo en cuenta que

$$\sigma_n < e^{-(p_1+p_2+\cdots+p_n)}(u_1 + u_2 + \cdots + u_n)$$

se ve de inmediato que $\sigma_n \to 0$ con la divergencia de (3.3). *Por lo cual se concluye que en el caso divergente.*

se tendrá que

$$A_1 = 0.$$

Aquí la conclusión no es falsa, pero sí se requiere una mejor cota, véase el error:

Si se toman:

$$p_n = \frac{e^{2n}}{1 + e^{2n}}, \quad q_n = \frac{1}{1 + e^{2n}}, \quad u_n = \frac{p_n}{q_n} = e^{2n}.$$

Entonces $\sum_{n=1}^{\infty} p_n = \infty$ y

$$\sum_{k=1}^{n} u_k > u_n = e^{2n}$$

mientras que

$$\sum_{k=1}^{n} p_k < n$$

por lo cual

$$e^{-(p_1+p_2+\cdots+p_n)}(u_1 + u_2 + \cdots + u_n) > e^{-n}e^{2n} = e^n \to_{n\to\infty} \infty.$$

Es decir fracasa la cota que Borel da. El error recide en la cota; esa desigualdad no es cierta, basta con probar $p_n = \frac{1}{2}\ \forall n$. Por otro lado, ahora sí está convencido que una serie de ceros converge a cero.

Análogamente A_2 será cero si (3.3) es divergente. Para el caso convergente

$$A_2 = A_0(u_1u_2+u_1u_2+u_1u_3+\cdots+u_2u_3+u_2u_4+u_2u_5+\cdots+u_3u_4+u_3u_5+u_3u_6+\cdots)$$

que se puede escribir como

$$A_2 = A_0 \cdot \sum_{i<j} u_iu_j$$

al igual que pasará con A_k, será cero en el caso cuando (3.3) sea divergente y en el caso convergente

$$A_k = A_0 \cdot \sum_{n_i} u_{n_1}u_{n_2}u_{n_3}\cdots u_{n_k}. \tag{3.8}$$

Problema III. ¿Cuál es la probabilidad de que el evento favorable se produzca una infinidad de veces? Denotaremos esta probabilidad por A_∞

Si tomamos la suma

$$S = A_0 + A_1 + \cdots + A_n + \ldots,$$

que simboliza el evento favorable: ocurre ninguna, una, dos, n veces; se puede escribir por (3.7) y (3.8) como

$$S = A_0(1+u_1)(1+u_2)\cdots(1+u_n)\cdots,$$

producto que será convergente si (3.3) converge. Pero se debe hacer la siguiente observación:

$$1 + u_n = \frac{1}{1 - p_n}$$

por lo cual

$$\prod_n (1 + u_n) = \prod_n \left(\frac{1}{1 - p_n} \right) = \frac{1}{A_0}$$

entonces

$$S = A_0 \frac{1}{A_0} = 1.$$

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

Sin embargo, si se designa por $\phi(h)$ el rango de la prueba en la cual se produce el caso favorable h, la función $\phi(h)$ puede tener un crecimiento tan rápido como se quiera en algunos momentos. Esto como se ha dicho, es un ejemplo de que probabilidad cero no implica imposibilidad. Ahora, en caso que (3.3) sea divergente, cada una de las A_k será cero (que se puede ver análogamente a lo hecho en A_1) y se podrá ver que $S = 0$, luego $A_\infty = 1$.

Entonces, para los casos estudiados de la primera categoría se puede concluir:

En caso que la serie (3.3) sea convergente, las probabilidades $A_0, A_1, \ldots$ son valores bien determinados no nulos y la probabilidad $A_\infty = 0$.[8] En el caso de que la serie (3.3) no converga, las probabilidades A_k son todas nulas y $A_\infty = 1$.

Es realmente fácil generalizar los anteriores 3 problemas a un número finito de casos posibles. Lo que parecería más interesante sería calcular la probabilidad de que un subconjunto finito determinado no se produzca jamás.

Segunda Categoría. Recuérdese que en esta categoría se tienen un número finito de pruebas y una cantidad numerable de posibles resultados. Si se nombran $p_1, p_2, \ldots, p_n, \ldots$ las probabilidades de los diversos eventos posibles la serie de términos positivos

[8] Cantelli en 1917 muestra que la hipótesis de independencia es no necesaria. Más adelante se tratará.

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

no sólo es convergente sino que es igual a uno.

Para Borel fue necesario aclarar que para él no es absurdo pensar si una serie de probabilidades nulas convergería, pero como no encontró circunstancias donde fuera necesario introducirla la descartó categóricamente y advierte que para usar lo siguiente se debe asegurar uno que no está en este *singular* caso. [9]

Lo que hace difícil pero a la vez interesante a los problemas de esta categoría es que las probabilidades p_i son raramente conocidas con precisión, por lo cual se deberá seguir un método indicado por Poincaré para las probabilidades *continuas*, buscando qué concluisiones generales se pueden obtener, con el mínimo de hipótesis sobre las p_i's.

Problema IV. ¿Cuál es la probabilidad de que en m pruebas sucesivas se obtengan resultados distintos en todas? Denótese con T_m dicha probabilidad.

Entonces por simple combinatoria se tiene

$$T_m = m! \sum_{n_1<n_2<\cdots<n_m} p_{n_1} p_{n_2} \cdots p_{n_m} \tag{3.9}$$

Para estimar el valor de dicha T_m se define la siguiente función compleja:

$$F(z) = (1+p_1 z)(1+p_2 z)\cdots(1+p_n z)\cdots = \prod_{n=1}^{\infty}(1+p_n z) \tag{3.10}$$

y claramente

$$F(z) = 1 + z\sum p_1 + z^2 \sum p_1 p_2 + \cdots + z^m \sum p_{n_1} p_{n_2} \cdots p_{n_m} + \cdots$$

es decir,

$$F(z) = 1 + z + \frac{T_2}{2!} z^2 + \cdots + \frac{T_m}{m!} z^m + \cdots$$

[9] Las cursivas son mías.

Considere los ceros de la función $F(z)$: $-\frac{1}{p_1}, -\frac{1}{p_2}, \ldots, -\frac{1}{p_n}, \ldots$ acomodados, sin pérdida de generalidad, de tal forma que sus módulos sean no-decrecientes. Sea r_n el módulo del cero de rango n. Se llama orden de una función entera al exponente ρ tal que la serie

$$\sum_n \frac{1}{r_n^{\rho+\epsilon}},$$

para ϵ pequeñas es convergente si $\epsilon > 0$ y es divergente si $\epsilon < 0$.

Se puede ver que si $p_n = \frac{c}{n^2}$ [10], $F(z)$ es de orden $\frac{1}{2}$. Note que $\cos(\sqrt{z})$ también es de orden $\frac{1}{2}$.[11] Como el máximo del módulo de una función entera de orden ρ sobre el círculo de radio r (y por consiguiente en su interior) es, para r suficientemente grandes, inferior a $\exp\{r\rho + \epsilon\}$ para ϵ arbitrariamente pequeña.

Por consiguiente $f(z)$ y $\cos(\sqrt{z})$ son comparables, es decir, asintóticamente sus coeficientes son muy parecidos. Es decir:

$$\frac{T_m}{m!} = \frac{1}{(2m)!}$$

luego por la aproximación de Stirling de $n!$:[12]

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

Problema V. ¿Cuál es la probabilidad de que en m pruebas se obtengan $m-1$ resultados distintos? Desígnese por V_m a dicha probabilidad.

Por combinatoria se tiene :

$$V_m = (m-1)! \sum_{n_1 < n_2 < \cdots < n_{m-1}} p_{n_1}^2 p_{n_2} \cdots p_{m-1} \tag{3.11}$$

[10] Esta c deducida de $\sum p_n = 1$

[11] Aquí es más sencillo ocupar el resultado siguiente: el orden ρ de una función $f(z)$ está dado por

$$\limsup_{r\to\infty} \frac{\log\log M(r; f)}{\log r}.$$

[12] Recuérdese que: $n! \sim n^n e^{-n}\sqrt{2\pi n}$.

luego,

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

además como $\sum p_i = 1$ y por (3.9) y (3.11) se tiene

$$\frac{V_m}{(m-1)!} = \frac{T_{m-1}}{(m-1)!} - m\frac{T_m}{m!}$$

luego

$$V_m = T_{m-1} - T_m.$$

Las cuestiones análogas se resuelven de la misma manera. El comportamiento asintótico de p_n se caracteriza por la función entera $F(z)$, la cual juega un papel importante en estas cuestiones. Ésta intervendrá en las cuestiones donde el número de casos posibles son una infinidad numerable. Por otro lado, podría tomarse un número suficientemente pequeño al 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 abordado por la probabilidad clásica y asunto resuelto.

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, pero esto no está sujeto a otra condición, puesto que dada una serie convergente $v = \sum v_n$ se podría tener

$$p_n = \frac{v_n}{v}$$

y en efecto

$$\sum p_n = 1.$$

Tercera Categoría. Aquí habrá un inifinito numerable de pruebas y de posibles resultados. A la probabilidad de que en la prueba s se obtenga el resultado n se denotará por $p_{n,s}$.

Cualquiera que sea s, es decir, en cualquier evento, la serie

$$\sum_{i=1}^{\infty} p_{i,s} = p_{1,s} + p_{2,s} + p_{3,s} + \ldots$$

no sólo es convergente sino que vale uno, lo cual ya fue tratado en la *segunda categoría*. Y se tiene lo siguiente:

$$\begin{cases} \sum_{i=1}^{\infty} p_{i,1} = p_{1,1} + p_{2,1} + p_{3,1} + \dots \\ \sum_{i=1}^{\infty} p_{i,2} = p_{1,2} + p_{2,2} + p_{3,2} + \dots \\ \sum_{i=1}^{\infty} p_{i,3} = p_{1,3} + p_{2,3} + p_{3,3} + \dots \\ \dots \end{cases} \tag{3.12}$$

Donde no necesariamente se tiene que

$$p_{n,1} + p_{n,2} + p_{n,3} + \cdots + p_{n,s} + \dots$$

converja y mucho menos a uno.

Si estas series (3.12) son convergentes, cualquiera que sea n se dirá que se está en el caso convergente, o en el caso divergente si todas divergen y en el caso mixto si unas convergen y otras no.

Limitándonos al caso convergente, se tiene:

$$c_n = \sum_{i=1}^{\infty} p_{n,i} = p_{n,1} + p_{n,2} + \cdots + p_{n,s} + \dots$$

Es claro 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 contradiría que cada suma de (3.12) converja a uno.

Para cada caso posible Borel asegura con toda razón que se podrían resolver los problemas I, II y III tratados anteriormente basta con agrupar los casos desfavorables como un solo evento y el caso favorable aparte. Desígnese para el caso de rango n:

$$B_{n,0}, B_{n,1}, \dots, B_{n,k} \text{ y } B_{n,\infty}$$

las probabilidades anteriormente designadas por:

$$A_0, A_1, \dots, A_k, \dots, A_\infty,$$

es decir, se designará por $B_{n,k}$ la probabilidad de que en el caso posible de rango n se produzca precísamente k veces en el conjunto de pruebas. Si se considera

$$v_{n,s} = \frac{p_{n,s}}{1 - p_{n,s}} \tag{3.13}$$

se tiene

$$\begin{cases} B_{n,0} = \prod(1 - p_{n,s}) \\ B_{n,1} = B_{n,0} \sum_s v_{n,s} \\ B_{n,k} = B_{n,0} \sum_{s_1 < s_2 < \cdots < s_k} v_{n,s_1} v_{n,s_2} \cdots v_{n,s_k} \\ B_{n,\infty} = 0 \end{cases} . \tag{3.14}$$

Problema VI. ¿Cuál es la probabilidad de que algún caso posible no se produzca un número finito de veces?

Una solución inmediata es resultado del razonamiento siguiente: se sabe que $B_{n,\infty} = 0$, la probabilidad de que en la caso de rango n se produzca un número infinito de veces es entonces cero. Entonces $\sum_{n=1}^{\infty} B_{n,\infty} = 0$ y la probabilidad buscada es $1 - \sum_{n=1}^{\infty} B_{n,\infty} = 1$.

No sobra justificar este hecho con un poco más de lógica: primero búsquese la probabilidad de que algún caso posible no se produzca más de una vez. La probabilidad, en el caso de rango n de que se produzca a lo más una ocasión es

$$B_{n,0} + B_{n,1}.$$

Las identidades (3.13) y (3.14) , por una tranformación sencilla

$$B_{n,0} + B_{n,1} = (1 + v_{n,1} + v_{n,2} + \cdots + v_{n,s} + \dots) \prod_{s=1}^{\infty} \frac{1}{1 + v_{n,s}}.$$

La probabilidad de que alguno de los m primeros casos se produzca a lo más una vez está dada por

$$\prod_{n=1}^{m} (B_{n,0} + B_{n,1}) = \prod_{n=1}^{m} \left[\left(1 + \sum_{s=1}^{\infty} v_{n,s} \right) \prod_{s=1}^{\infty} \frac{1}{1 + v_{n,s}} \right]. \tag{3.15}$$

Ahora consistiría en ver que pasa cuando $m \to \infty$. Nótese que

$$\frac{\sum_{s=1}^{k} v_{n,s}}{\prod_{s=1}^{k}(1+v_{n,s})}$$

se puede ver como

$$\ln\left(\sum_{s=1}^{k} v_{n,s}\right) - \sum_{s=1}^{k} \ln(1+v_{n,s})$$

cuya sucesión es decreciente y no acotada inferiormente, por lo cual

$$\frac{\sum_{s=1}^{k} v_{n,s}}{\prod_{s=1}^{k}(1+v_{n,s})} \to 0$$

Con lo cual se concluye que

$$\lim_{m\to\infty} \prod_{n=1}^{m}(B_{n,0}+B_{n,1}) = 0.$$

En las mismas condiciones

$$\forall k : \frac{\sum u_n + \sum u_{n_1}u_{n_2} + \cdots + \sum u_{n_1}u_{n_2}\cdots u_{n_k}}{\prod(1+u_n)} \to 0$$

por lo cual se puede concluir que

$$\forall k \text{ fija: } \prod_{n=1}^{m}(B_{n,0}+B_{n,1}+\cdots+B_{n,k}) \to 0 \text{ si } m \to \infty.$$

Este límite representa la probabilidad de que alguno de los casos se produzca a lo más k veces es cero sin importar que k fija se tome. Ahora, considere:

$$S_{n,k} = B_{n,0} + B_{n,1} + \cdots + B_{n,k}$$

$$R_{n,k} = B_{n,k+1} + B_{n,k+2} + \ldots$$

donde

$$S_{n,k} + R_{n,k} = \sum_{s=0}^{\infty} B_{n,s} = 1.$$

Dado un número ϵ pequeño es claro que se puden escoger números k_n tales que

$$\sum_{n=1}^{\infty} R_{n,k_n} < \epsilon.$$

Calculando la probabilidad de que el caso posible de rango n no se produzca más de k_n veces, se obtiene

$$\prod_{n=1}^{\infty} S_{n,k_n} = \prod_{n=1}^{\infty}(1 - R_{n,k_n}).$$

puesto que para α suficientemente pequeña (basta con $\alpha < \frac{1}{2}$),

$$1 - \alpha > e^{-2\alpha} > 1 - 2\alpha$$

y se concluye

$$\prod_{n=1}^{\infty}(1 - R_{n,k}) > e^{-2\sum_{n=1}^{\infty} R_{v,k_n}} > e^{-2\epsilon} > 1 - 2\epsilon.$$

Entonces se puede escoger para cada número arbitrario ϵ, números k_n de manera que la probabilidad de que el caso de rango n no se produzca más de k_n veces difiere en menos de 2ϵ de la unidad. Y la probabilidad de que un caso posible se produzca una infinidad numerable de veces es entonces inferior a 2ϵ, para cualquier ϵ, es decir, la probabilidad es nula.

La determinación efectiva de las k_n en función de ϵ depende de la rapidez de convergencia de las series

$$\sum_{s=0}^{\infty} B_{n,s}$$

la cual depende a su vez de la rapidez de convergencia de las series

$$\sum_{s=1}^{\infty} v_{n,s}$$

$$\sum_{s=1}^{\infty} p_{n,s}.$$

La naturaleza de convergencia de estas últimas series, cuando n crece, es la característica esencial de los problemas de probabilidad de la tercer categoría: su variedad es entonces muy grande, pero Borel se limitó a estas hipótesis, pues no parece haber querido desarrollar una teoría sin aplicaciones y tampoco contaba aún con más herramienta. Su gran logro fue manejar con habilidad probabilidades donde las muestras son infinitas.

3.2.2 Fracciones decimales.

Para ejemplificar lo anterior, Borel llama con el nombre de fracciones decimales de base q a las expresiones de la forma

$$\sum_{n=1}^{\infty} \frac{b_n}{q^n} \text{ donde } b_n \in \{0 \leq b_n < q\}.$$

Para calcular la probabilidad de que una fracción decimal pertenezca a un conjunto dado, conviene suponer:

1. Las cifras decimales son independientes.
2. Cada una de ellas tiene una probabilidad igual a $\frac{1}{q}$ (en el caso de base q).

Lo anterior como bien se sabe, no se cumple necesariamente en la práctica sin embargo es lo ideal y simplifica los cálculos. Además se justifican cuando se colocan no en el punto de vista lógico sino en el geométrico pues son equivalentes a lo siguiente: *El número decimal siendo representado por un punto en el segmento* $[0, 1]$*, la probabilidad de que esté en un determinado subsegmento es igual a la longitud de dicho segmento.* Con este punto de vista se relacionaría la edida de Borel con la Probabilidad.

Si se pone atención sobre una cifra determinada, a saber la cifra 3 y se observa como caso favorable donde se presente esta cifra se estaría de nuevo en el caso de la *primera categoría* y los decimales sucesivos correspondientes a la infinidad numerable de pruebas. La probabilidad del caso favorable, es aquí la misma para cada prueba y nos coloca en el caso de divergencia; la probabi-lidad 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 a $\frac{1}{3}$ en fraccion decimal, como se ha señalado la probabilidad es igual a cero no significando imposibilidad.

Como todo número puede ser transformado a base binaria, ahora cada cifra: cero o uno tendrá probabilidad $\frac{1}{2}$ en su desarrollo binario.

Tómese como caso favorable el caso donde aparece cero como cifra. Se sabe que si se tienen $2n$ pruebas, la probabilidad de que el número de casos favorables caiga entre

$$n - \lambda\sqrt{n} \text{ y } n + \lambda\sqrt{n}$$

es igual a

$$\Theta(\lambda) = \frac{2}{\sqrt{\pi}} \int_0^{\lambda} e^{-\frac{x^2}{2}} dx.$$

Esta probabilidad tiende rápidamente a la unidad conforme $\lambda \to \infty$.

Considérese una sucesión de números λ_n que aumenten indefinidamente con n, de tal manera que

$$\lim_{n\to\infty} \frac{\lambda_n}{\sqrt{n}} = 0.$$

Tómese la sucesión $\lambda_n = \log n$ por ejemplo. Ahora del conjunto de $2n$ pruebas, si se considera como favorable el caso que tenga la cifra 0 y que esté entre

$$n - \lambda_n\sqrt{n} \text{ y } n + \lambda_n\sqrt{n},$$

tendrá probabilidad

$$p_n = \Theta(\lambda_n) = \frac{2}{\sqrt{\pi}} \int_0^{\lambda_n} e^{-\frac{x^2}{2}} dx.$$

Aquí hay que añadir a una aclaración, sobre su trabajo. Se requiere 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, se 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 una n se obtiene constantemente casos favorables, puesto que en este caso, el cociente entre 0 y 1 está comprendido entre

$$\frac{n - \lambda_n \sqrt{n}}{n + \lambda_n \sqrt{n}} \text{ y } \frac{n + \lambda_n \sqrt{n}}{n - \lambda_n \sqrt{n}}$$

es decir, entre

$$\frac{1 - \frac{\lambda_n}{\sqrt{n}}}{1 + \frac{\lambda_n}{\sqrt{n}}} \text{ y } \frac{1 + \frac{\lambda_n}{\sqrt{n}}}{1 - \frac{\lambda_n}{\sqrt{n}}}.$$

Estos cocientes tienden a cero y a uno cuando n tiene a infinito; por lo cual:

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] A esto se le conocerá como la Ley de los Grandes Números de Borel o Ley Fuerte de los Grandes Números.[14]

Lo anterior puede ser *extendido* a cualquier base, incluida por supuesto la base 10. Llamó *frecuencia* de un decimal al cociente que denota cuantas veces aparece en las primeras n cifras; si la frecuencia así definida tiende a un límite conforme n aumenta indefinidamente, se dirá que la *frecuencia total* existe y su valor es igual a este límite. Lo anterior da cavida a lo siguiente:

[13] La probabilidad de que el cociente entre los números cero y uno tienda a la unidad (cuando el número de cifras consideradas aumente indefinidamente) es igual a uno. La probabilidad de que el cociente no tienda hacia algún límite o a alguno distinto a la unidad es, por consiguiente, igual a cero.
Ver [9] pp. 194

[14] Esto el mismo Borel lo señala en [9] pp. 194

La probabilidad de que la frecuencia total de una cifra determinada exista y sea igual a $\frac{1}{10}$ es uno. La probabilidad de que la frecuencia de diez cifras exista y sea igual a $\frac{1}{10}$ para cada una de ellas es tambien uno.

Cuando la condición precedente sea satisfecha por una fracción decimal, se dirá que el número igual a esta fracción es *simplemente normal* en relación a la base 10. Un número simplemente normal es entonces caracterizado por el hecho que $c_0, c_1, c_2 \ldots, c_8, c_9$ designen los número respectivos de veces que figuran las cifras $0, 1, 2, \ldots 9$ en los primeros n decimales, cada uno de los

$$\frac{c_0}{n}, \frac{c_1}{n}, \frac{c_2}{n}, \ldots \frac{c_8}{n}, \frac{c_9}{n}$$

tiene por límite $\frac{1}{10}$ cuando $n \to \infty$. Se puede decir con esta definición: la probabilidad de que un número sea simplemente normal en relación a la base 10 es igual a uno.

Todo número escrito en una base determinada, a saber 10 por ejemplo, se puede sin algún cálculo, ser visto como escrito en una base igual a una potencia de 10. Se dirá que un número es *entéramente normal* en relación a la base 10, abreviándolo como *normal* en relación a esta base, cuando este número y sus productos con las diversas potencias de 10 son *simplemente normales* en relación a todas las bases iguales a una potencia de 10: $10, 10^2, 10^3, \ldots, 10^n, \ldots$.

La probabilidad de que esta condición sea cumplida es igual a uno para cada una de estas bases; la probabilidad de que un número sea enteramente normal en relación a la base 10 es entonces también igual a la unidad. [15]

La propiedad característica de un número normal es la siguiente: *Un grupo cualquiera de p cifras consecutivas siendo consideradas, si se designa por c_n el número de veces que se encuentra en las primeras n cifras decimales, se tiene*

$$\lim_{n\to\infty} \frac{c_n}{n} = \frac{1}{10^p}.$$

Cuando un número dado es normal *en relación a todas las bases posibles*, se dirá que es *absolutamente normal*. la probabilidad de que un número sea

[15] Borel usa el Teorema de probabilidades compuestas, a pesar que es para un número infinito de casos y señala que considera no necesario argumentar de nuevo su aplicación. Acuerdo yo con él.

absolutamente normal es también igual a la unidad.

Borel muestra un método para construir números normales en relación a la base 10, para más detalle puede consultar [9] en su Nota 4.

Ahora, si en lugar de considerar el conjunto de números decimales, sólo nos limitamos al conjunto de fracciones periódicas y tomamos a los números con número de decimales finito, como si fuera periódico, por ejemplo el 0.43 lo vemos como $0.42999999\ldots$, conseguimos un conjunto con cardinalidad igual al de los racionales.

¿Se puede calcular la probabilidad de que la parte no periódica tenga precísamente i cifras, que el periodo tenga exactamente k cifras, que la parte periódica o no periódica sean formadas por las cifras anteriores?

Bajo estos supuestos, se está ahora en problema(s) de la segunda categoría: casos posibles numerables. Debe ser claro ahora porqué las probabilidades no pueden ser iguales.[16] Entonces desígnese por $p_{i,k}$ la probabilidad de que la parte no periódica se componga de i cifras y la parte periódica tenga k cifras. El número i puede tener un valor entero cualquiera desde cero y el número k un valor entero cualquiera desde uno para tener:

$$\sum_{i=0}^{\infty}\sum_{k=1}^{\infty} p_{i,k} = 1 \tag{3.16}$$

Si admitimos que las diversas cifras $0, 1, 2, 3\ldots, 9$ tengan probabilidades iguales, la probabilidad de que un número determinado cuya parte no periódica se componga de i cifras y la periódica de k, con k primo; es

$$\frac{p_{i,k}}{10^i(10^k - 10)}. \tag{3.17}$$

Las expresiones (3.16) y (3.17) son escenciales como las siguientes, se tiene:

$$\sum_{k=1}^{\infty} p_{i,k} = P_i \quad (i = 0, 1, \ldots)$$

[16]Deben sumar uno y si no son cero, la serie no converge.

$$\sum_{i=0}^{\infty} p_{i,k} = Q_k \quad (k = 1, 2, \dots)$$

de donde claramente

$$\sum_{i=0}^{\infty} P_i = \sum_{k=1}^{\infty} Q_k = 1.$$

Definimos las siguientes funciones enteras:

$$F(z) = \prod_{i=0}^{\infty}(1 + P_i z)$$
$$Q(z) = \prod_{k=0}^{\infty}(1 + Q_k z)$$
$$\phi(z) = \prod_{i=0}^{\infty}\prod_{k=1}^{\infty}(1 + p_{i,k} z)$$

y será necesario hacer suposiciones sobre el orden de las funciones. Un caso particular interesante sería interesante si se supone

$$p_{i,k} = p_i q_k$$

Habría ventaja en considerar separádamente la convergencia de las series

$$\sum_{i=0}^{\infty} p_i = 1$$
$$\sum_{k=1}^{\infty} q_k = 1$$

o lo que es lo mismo las funciones enteras

$$f(z) = \prod_{i=0}^{\infty}(1 + p_i z)$$
$$g(z) = \prod_{k=0}^{\infty}(1 + q_k z).$$

Esto último lo retomaremos con la siguiente sección de fracciones continuas.

3.2.3 Fracciones continuas

Considérese el desarrollo en fracción continua de un número irracional entre cero y uno. Será de la forma:

$$\cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \frac{1}{\ddots}}}} \tag{3.18}$$

donde las a_i's son enteros cualesquiera fijos. Este caso será un problema de la tercera categoría (de Borel), por número de pruebas y posibles resultados numerables. Designe de manera general por $p_{i,k}$ la probabilidad de que el cociente incompleto a_i sea igual al entero k. Todos los números enteros i y k pueden ser cualquier número natural distinto de cero[17], para lo cual se tendrá

$$\sum_{k=1}^{\infty} p_{i,k} = 1 \tag{3.19}$$

Se podrá hacer *a priori* sobre los $p_{i,k}$ hipótesis necesarias para que lo anterior sea satisfecho. Por otro lado, para calcular $p_{1,k}$, es decir, la probabilidad de que $a_1 = k$. De la relación (3.18), se tiene

$$\frac{1}{k+1} < x < \frac{1}{k}$$

por lo cual x debe de estar en un intervalo de magnitud

$$\frac{1}{k(k+1)}.$$

La probabilidad (geométrica) que x esté en el intervalo es igual a su longitud, entonces

$$p_{i,k} = \frac{1}{k(k+1)}.$$

[17]Si suponemos que alguno es cero, no causa contradicción alguna, simplemente sería un racional la fracción continua y estaría en el caso anterior.

En particular,

$$p_1 = \frac{1}{2}, \quad p_2 = \frac{1}{6}, \quad p_3 = \frac{1}{12}, \quad p_4 = \frac{1}{20}, \quad \ldots$$

.

Ahora para $p_{2,k}$. Supóngase $a_1 = n$, la condición para que $a_2 = k$ será por (3.18):

$$\frac{1}{n + \frac{1}{k}} < x < \frac{1}{n + \frac{1}{k+1}},$$

es decir

$$\frac{k}{nk+1} < x < \frac{k+1}{n(k+1)+1}.$$

Dado lo cual, la probabilidad $p_{2,k}$ es igual a la suma de las longitudes de los intervalos análogos correspondientes a los diversos valores de n:

$$p_{2,k} = \sum_{n=1}^{\infty} \frac{1}{(nk+n+1)(nk+1)},$$

que también se puede ver como

$$\sum_{n=1}^{\infty} \frac{1}{n^2} \left(\frac{1}{k + \frac{1}{n}} - \frac{1}{k+1+\frac{1}{n}} \right) \text{ o bien como } \frac{1}{k(k+1)} \sum_{n=1}^{\infty} \frac{1}{\left(n + \frac{1}{k}\right)\left(n + \frac{1}{k+1}\right)}.$$

De las anteriores expresiones se podría dar un valor asintótico de $p_{2,k}$ para grandes valores de k, la suma de segunda expresión se reduciría a

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$$

entonces

$$p_{2,k} = \frac{\pi^2}{6} \frac{1}{k(k+1)} (1 - \epsilon_k)$$

siendo ϵ_k un número positivo que tiende a cero cuando $k \to \infty$.

Cabe hacer notar que para grandes valores de k, $p_{2,k}$ es superior a $p_{1,k}$, entonces para valores pequeños de k, $p_{2,k}$ debe ser inferior a $p_{1,k}$ puesto que

$$\sum_{k=1}^{\infty} p_{1,k} = \sum_{k=1}^{\infty} p_{2,k} = 1$$

No es difícil demostrar que $p_{2,k} = 2\ln(2) - 1 = (4\ln(2) - 2)p_{1,1}$. [18]

Ahora de la misma manera para $p_{3,k}$. Suponiendo $a_1 = n$ y $a_2 = m$, la condición de $a_3 = k$ está dada por:

$$\frac{1}{n + \cfrac{1}{m + \cfrac{1}{k+1}}} < x < \frac{1}{n + \cfrac{1}{m + \cfrac{1}{k}}}$$

o bien

$$\frac{m(k+1)+1}{(nm+1)(k+1)+n} < x < \frac{mk+1}{(nm+1)k+1}.$$

y de igual manera que en el caso inmediato anterior

$$p_{3,k} = \frac{1}{k(k+1)} \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \frac{1}{\left(nm + 1 + \frac{n}{k+1}\right)\left(nm + 1 + \frac{n}{k}\right)}$$

puesto que el intervalo tiene longitud $\frac{1}{[(nm+1)(k+1)+n][nm+1+\frac{n}{k}]}$. Lo anterior tiene en el límite un comportamiento

$$\sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \frac{1}{(nm+1)^2} = \sum_{q=2}^{\infty} \frac{\theta(q-1)}{q^2},$$

donde $\theta(n)$ es el número de divisores entre 1 y n.

Ahora, para $a_n = k$, supóngase se reducen las fracciones continuas truncadas de rango $n-2$ y $n-1$ y se representan por

$$\frac{P_{n-2}}{Q_{n-2}} \text{ y } \frac{P_{n-1}}{Q_{n-1}}$$

[18] Con $n = 79$ se obtiene una aproximación con dos decimales.

entonces

$$\frac{P_{n-3}+kP_{n-1}}{Q_{n-2}+kQ_{n-1}} < x < \frac{P_{n-2}+(k+1)P_{n-1}}{Q_{n-2}+(k+1)Q_{n-1}}$$

La probabilidad $p_{n,k}$ será la suma de los intervalos cuyas magnitudes son

$$\frac{1}{Q_{n-1}^2}\cdot\frac{1}{\left(k+\frac{Q_{n-2}}{Q_{n-1}}\right)\left(k+1+\frac{Q_{n-2}}{Q_{n-1}}\right)}$$

y para $p_{n,k+1}$

$$\frac{1}{Q_{n-1}^2}\cdot\frac{1}{\left(k+1+\frac{Q_{n-2}}{Q_{n-1}}\right)\left(k+2+\frac{Q_{n-2}}{Q_{n-1}}\right)}$$

El cociente de longitudes entre las últimas dos es

$$\frac{k+\frac{Q_{n-2}}{Q_{n-1}}}{k+2+\frac{Q_{n-2}}{Q_{n-1}}}, \tag{3.20}$$

por lo cual la relación de $p_{n,k+1}$ a $p_{n,k}$ estará acotada por los más grandes y los más pequeños valores que toma (3.20) cualesquiera sean Q_{n-2} y Q_{n-1} siempre y cuando sean los denominadores de una misma fracción continua en la (n-2)-ésima y (n-1)-ésima reducción. Y como

$$0 < \frac{Q_{n-2}}{Q_{n-1}} < 1,$$

se tiene en conclusión que el cociente (3.20) está comprendido entre $\frac{k}{k+2}$ y $\frac{k+1}{k+3}$. Se obtiene entonces,

$$\frac{k}{k+2} < \frac{p_{n,k+1}}{p_{n,k}} < \frac{k+1}{k+3}. \tag{3.21}$$

Por recurrencia se obtiene

$$\frac{2(k-1)!}{(k+1)!} = \frac{(k-1)(k-2)\cdots 2\cdot 1}{(k+1)k\cdots 4\cdot 3} < \frac{p_{n,k}}{p_{n,1}} < \frac{k(k+1)\cdots 3\cdot 2}{(k+2)(k+1)\cdots 5\cdot 4} = \frac{3!\cdot k!}{(k+2)!}$$

es decir

$$\frac{2}{k(k+1)} < \frac{p_{n,k}}{p_{n,1}} < \frac{6}{(k+1)(k+2)} \tag{3.22}$$

y usando que $\sum_{k=1}^{\infty} p_{n,k} = 1$ se obtiene

$$2 = \sum_{k=1}^{\infty} \frac{2}{k(k+1)} < \frac{\sum_{k=1}^{\infty} p_{n,k}}{p_{n,1}} = \frac{1}{p_{n,1}} < \sum_{k=1}^{\infty} \frac{6}{(k+1)(k+2)} = 3$$

que es lo mismo a

$$\frac{1}{3} < p_{n,1} < \frac{1}{2}. \tag{3.23}$$

Multiplicando (3.22) y (3.23) se obtiene

$$\frac{2}{3k(k+1)} < p_{n,k} < \frac{3}{(k+1)(k+2)} \tag{3.24}$$

relación que nos da un intervalo de longitud $\frac{7k-4}{3k(k+1)(k+2)}$ que por sí sola muestra que es una buena aproximación y mejora conforme k crece, para $k = 14$ podría ser una buena aproximación, ya que se aproxima a 0.0093254.

La desigualdad (3.24) mostraría que la serie

$$\sum_{n=1}^{\infty} p_{n,k}$$

diverge sin remedio. Borel anuncia que no insistirá ni comentará las consecuencias de este hecho, pues para él, el caso convergente es más interesante. Sin embargo ¿qué quiere decir que diverja? Nunca deja de ser probable que el número k reaparezca en el desarrollo de la fracción, además que por lo anterior también todo número entero tiene probabilidad positiva de aparecer al menos una vez y por ende de volver a aparecer.

Por otro lado, nómbrese por $P_{n,k}$ la probabilidad de que a_n sea igual o menor a k, siendo:

$$P_{n,k} = p_{n,1} + p_{n,2} + \cdots + p_{n,k}.$$

La desigualdad (3.24) permitirían obtener cotas para $P_{n,k}$. Pero para obtener unas cotas más precisas, en virtud de (3.19):

$$1 - P_{n,k} = p_{n,k+1} + p_{n,k+2} + \dots$$

y aplicando (3.24) se obtiene

$$\frac{2}{3(k+1)} < 1 - P_{n,k} < \frac{3}{k+2}$$

es decir

$$1 - \frac{3}{k+2} < 1 - P_{n,k} < 1 - \frac{2}{3(k+1)}$$

Tómese ahora una función $\phi(n)$, valuada en los enteros, creciente de tal manera que

$$\sum \frac{1}{\phi(n)} < \infty; \tag{3.25}$$

con lo cual

$$\sum_{n=1}^{\infty} (1 - P_{n,\phi(n)}) \tag{3.26}$$

converja. Es decir, la probabilidad de que se tenga para una infinidad de valores de n:

$$a_n > \phi(n)$$

es nula. Y por otro lado se tiene que con probabilidad uno, a partir de un valor finito de n:

$$a_n > \phi(n).$$

Por otro lado, si (3.25) diverge entonces (3.26) diverge. Usando esto y con el resultado anterior de A_∞ [19] se puede concluir lo que se llamará **El teorema**

[19] Ver Problema tres de este capítulo.

de Fracciones Continuas de Borel:

Si $\sum \frac{1}{\phi(n)}$ converge entonces existe una función $\psi(n)$ tal que $\sum \frac{1}{\psi(n)}$ converge mientras $\lim_{n\to\infty} \frac{\psi(n)}{\phi(n)} = 0$ y de igual manera si $\sum \frac{1}{\phi(n)}$ diverge entonces existe una función $\psi(n)$ tal que $\sum \frac{1}{\psi(n)}$ diverge mientras $\lim_{n\to\infty} \frac{\psi(n)}{\phi(n)} = \infty$.

Con lo anterior se puede concluir lo siguiente:
la probabilidad de que $\lim_{n\to\infty} \frac{a_n}{\phi(n)} = 0$ *siendo* $\sum \frac{1}{\phi(n)}$ *convergente es uno,* pero *si la serie es divergente entonces sólo se puede asegurar que* $\overline{\lim}_n \frac{a_n}{\phi(n)} = \infty$.[20]

3.2.4 Cuestiones diversas y conclusión.

Un último ejemplo será considerar la ecuación de segundo grado con coeficientes enteros

$$ax^2 - bx + c = 0 \tag{3.27}$$

y sea $\delta = b^4 - 4ac$.

Para que la ecuación tenga raíces enteras es suficiente y necesario que δ sea una raíz exacta. Si δ es nula, las raices son iguales. Luego, se puede determinar la probabilidad p_1 que la ecuación (3.27) tenga raices racionales, p_2 que tenga raices iguales y p_3 que tenga raíces reales y quizá más.

Primero nombre a_m, b_n, c_r la probabilidad de que a, b y c hayan tomado el valor entero m, n y r respectivamente. Con lo cual se tendrá lo siguiente

$$\sum_{m\in\mathbb{Z};\ m\neq 0} a_m = 1 \tag{3.28}$$

$$\sum_{n\in\mathbb{Z}} b_n = 1 \tag{3.29}$$

$$\sum_{r\in\mathbb{Z}} c_r = 1 \tag{3.30}$$

Supóngase que los coeficientes son independientes entre ellos, por lo cual

[20]Ver [9] pp. 194-195.

$$p_1 = \sum_{\{m,n,r:\sqrt{n^2-4mr} \,\in\, \mathbb{Z}\}} a_m b_n c_r \tag{3.31}$$

y de un razonamiento similar:

$$p_2 = \sum_{\{m,n,r:n^2-4mr=0\}} a_m b_n c_r \tag{3.32}$$

y

$$p_3 = \sum_{\{m,n,r:n^2-4mr\geq 0\}} a_m b_n c_r \tag{3.33}$$

cuyas series (3.31), (3.32) y (3.33) convergen claramente pues (3.28), (3.29) y (3.30) convergen. Claramente en la práctica sería necesario poner más hipótesis sobre a, b y c o sobre a_m, b_n y c_n para conocer una aproximación real a la probabilidad deseada.

Ahora si dejamos arbitrario el grado de la ecuación: ¿Cuál será la probabilidad de obtener racionales teniendo coeficientes enteros de nuevo?

Lo anterior conduce naturalmente a pensar la cuestión siguiente: si se considera un conjunto de números, por ejempo el de los algebráicos y se pide la probabilidad de que dado un número en este conjunto pertenezca a otro (finito, infinito o un singular). Tal cuestión sólo puede ser resuelta mediante el uso de hipótesis arbitrarias; esto no disminuye, o al menos no debería, el interés de estudio. Se podría incluso decir que lo aumenta al permitir la solucíon de numerosos problemas al variar las hipótesis.

Está claro que todos los elementos analíticos, números y funciones, que pueden ser efectivamente definidos son numerables en cantidad y se podrá entonces en términos análogos a los problemas precedentes. Por *efectivamente definidos* se debe entender: definidos en un número finito de palabras y debe ser claro que la probabilidad de que la definición se extienda en muchas palabras deberá ser muy pequeña.

Y Borel cierra el artículo con las siguientes palabras:

Lorsque la théorie des probabilités dénombrables aura été développée dans le sens quie vient d'être indiqué, il sera interéssant de comparer les résultats *acquis avec ceux quón obtenient par la théorie des probabilités continues ou géométiques.*

Il existe *certainement (si ce n'est pas un abus d'employer ici le verbe* existes*) dans le continu géométrique des éléments qui ne peuvent pas être définis: tel est le sens réel de l'importante et célèbre proposition de Georg Cantor: le continu n'est pas dénombrable. Le jour où ces elements* indéfinissables *seront réellemnt mos à part et où l'on ne prétendrait point les faire intervenir plus ou moins implicitemenr, il en résulterait certainement une grande simplification dans les méthodes de lánalyse; je serai heureux si les pages précédentes pouvaient contribuer à faire pressentir l'intérêt qui s'attacherait à l'étude de telles questions.*[21]

Si pudiera tan sólo Borel ver la importancia que hoy tiene la probabilidad en toda la ciencia, cumpliría su <<debería estar feliz>>*.*

El trabajo de Borel en este artículo, que se ha estudiado aquí, sobre todo la parte de fracciones continuas, como el de Wiman tratado con anterioridad bien podrían ser vistos desde una perspectiva estocástica: como caminatas aleatorias; al menos este interés me despertó al trabajarlo. Sin embargo constituiría este esfuerzo en si mismo, un trabajo independiente y no por eso desligado de éste, pero que sale de los intereses (actuales) del presente.

3.3 El Lema de Borel: desde Cantelli.

Cantelli en 1917 publica *Sulla probabilità come limite della frequenza* y *Su due applicazioni di un teorema di G. Boole alla statistica matematica*[22] desde Roma, donde retoma parte del artículo de Borel que hemos tratado y lo amplia de la siguiente forma:

[21] Al tiempo que la Teoría de la Probabilidad numerable es desarrollada en la manera indicada, será interesante comparar los resultados adquiridos con los obtenidos en la teoría de probabilidades continuas o geométricas.
En el continuo geométrico *existen* ciertamente (si no es un abuso usar el verbo existir) algunos elementos que no pueden ser definidos: tal es el sentido real de la importante y célebre proposición de Georg Cantor: el continuo no es numerable. Un día deberá de venir, en el cual estos elementos *indefinibles* se puedan poner a un lado como no necesarios o al menos implícitamente; lo cual ciertamente traerá gran simplificación en los métodos del Análisis; Yo deberé estar feliz si las páginas anteriores pueden ayudar al estudio de estas cuestiones. (Cursivas originales).

[22] Ver [12] y [13]

Cantelli después de enunciar ciertos postulados que debe de cumplir su medida (de Probabilidad), en la búsqueda de abstraer el cálculo de probabilidades, deduce ciertas propiedades de la que rescataré una en especial pues la usaremos.

$$m(E_1 E_2 \cdots E_n) \geq 1 - \sum_{i=1}^{n} m(\bar{E}_i)^{23}$$

donde los E_i representa superficies del área unitaria dada, así lo nombra Cantelli, lo que después evolucionará a el conjunto de eventos Ω.

Posteriormente incursiona en el caso de una sucesión ilimitada de áreas parciales teniendo por lo anterior:

$$m(E_n E_{n+1} E_{n+2} \cdots E_{n+r}) \geq 1 - \sum_{i=n}^{n+r} m(\bar{E}_i)$$

Como la sucesión $m(E_n), m(E_n E_{n+1}), m(E_n E_{n+1} E_{n+2}), \ldots, m(E_n E_{n+1} \cdots E_{n+r}), \ldots$ es decreciente o al menos no creciente ésta tiende a un límite.

Si la serie

$$\sum_{i=1}^{\infty} m(\bar{E}_i)$$

es convergente dada la desigualdad anterior se tiene que

$$m(E_n E_1 E_2 \cdots) \geq 1 - \sum_{i=n}^{\infty} m(\bar{E}_i)$$

Si la serie

$$\sum_{i=n}^{\infty} m(\bar{E}_i)$$

es divergente, nos da un análisis poco concluyente. En cambio si las áreas son multiplicables (independientes en el sentido de Cantelli), se tendrá:

$$m(E_n E_1 E_2 \cdots) = \prod_{i=n}^{\infty} m(E_i) = \prod_{i=n}^{\infty} [1 - m(\bar{E}_i)]$$

y dada la divergencia de la serie desemboca en:

[23] La barra denota complemento para Cantelli

$$\prod_{i=n}^{\infty} m(E_i) = m(E_n E_{n+1} E_{n+2} \cdots) = 0$$

de lo cual se deriva en:

$$m(\bar{E}_n + \bar{E}_{n+1} + \dots) = 1$$

es decir que con probabilidad uno, una infinidad de eventos ocurrirían, en el sentido moderno.

Cómo se conoce hoy en día el Lema de Borel-Cantelli se presenta en el anexo, con unas generalizaciones y consecuencias extras.

3.4 La Probabilidad entre los trabajos de Borel y Steinhaus.

Felix Hausdorff a través de su libro *Grundzüge der Mengenlehre* publicado en 1914 muestra la probabilidad como una aplicación de la Teoría de la Medida:[24] sea P un subconjunto de M los dos medibles, entonces

$$\frac{\mu(P)}{\mu(M)}$$

puede ser definida como la probabilidad de que un punto de M pertenezca a su vez a P. Tambíen afirma que el hecho de que tenga probabilidad cero, no significa forzosamente que no esté en el conjunto al igual que Borel lo había afirmado años antes y dice que

... many theorems of the measure of point sets take on a more familiar appearance when expressed in the language of probability calculus.[25]

Para Hausdorff la probabilidad matemática no se identificaba con la medida, mientras que la medida normalizada por su puesto sí se relacionaba con

[24] Ver von Plato [35] pp. 187

[25] Cita traducida del alemán en *Grundzüge der Mengenlehre* al inglés por Jan von Plato. Muchos teoremas de la Teoría de la medida (de conjuntos de puntos) toman una apariencia más familiar cuando se expresan en el lenguaje del cálculo de probabilidades.

la probabilidad, pues cumplia (todas) las propiedades formales (o fundamentales) de una probabilidad.

Von Mises en 1919 publicó dos trabajos sobre probabilidad en los cuales critíca la definición clásica de Probabilidad: *Razón entre el número de casos favorables entre en número total de casos* igualmente probables. Ahí es justo donde pone el dedo en la llaga. Se pregunta: ¿Qué es igualmente posibles? La posibilidad no es algo medible como lo es la longitud o la temperatura, argumenta, lo cual lleva a un *círculo vicioso*.

Su sistema fundacional es el siguiente:

Hay un espacio de resultados posibles donde cada uno está representado por un número. La sucesión resultante de números es llamada, por él, un colectivo si cumple los siguientes dos postulados:

1. Los límites de frecuencias relativas en un colectivo existen.

2. Estos límtes permanecen iguales en subsucesiones formadas por la sucesión original.

Y el concepto de Probabilidad sería sólo aplicable a colectivos. El postulado segundo de von Mises, que versa sobre aleatoriedad, es una expresión de su indeterminismo. El dice que en la Física Clásica existe un algoritmo para calcular, al menos en principio, un curso futuro de eventos; esto motivado por la Física estadística. Él asumía que hay un estado finito de estados macroscópicos discernibles, con leyes de probabilidad de transición entre ellas. Los estados con las probabilidades de transición después serían conocidas como *Cadenas de Markov*; aunque no se debe olvidar que desde el trabajo de Borel ya se hacía dislumbrar el antecedente de esto, con espacio de estados infinito.

Otra crítica de von Mises a las ideas de Probabilidad, es el uso del infinito. Él, siendo uno de los fundadores del Empirismo Lógico consideraba al infinito sólo como una idealización que es aceptada en Probabilidad como en otras partes de la Matemática. Afirmaba que no se podría *alcanzar* la realidad directamente sino sólo como una idealización (bastante) útil.

Capítulo 4

Les probabilités dénombrables et leur rapport à la théorie de la mesure.

Hasta 1923 el estatus de la teoría que reina este trabajo no podría ser mejor descrita:

(...) Probability as a measure of "degree of belief" and Probability Theory as a systematic guide to "behavior under uncertainty".[1]

Después de que Borel mostrara interés en lo que el nombró probabilidades numerables y dejara un teorema que sería conocido como *La paradoja de Borel*:

La probabilidad que la frecuencia[2] *de la cifra* 0 *en el desarrollo diádico de un número tomado al azar sea igual a* $\frac{1}{2}$ *tiene por valor uno.*

Otros autores encontraron una equivalencia al teorema:

Casi todos los números α *tienen la propiedad que la frecuencia de ceros en su desarrollo diádico es igual a* $\frac{1}{2}$ donde por casi todos los números significa que la medida de Lebesgue del conjunto de números α sin la propiedad en

[1] Jan von Plato en *Creating Modern Probability*. La probabilidad como una medida del grado de creencia y la Teoría de la Probabilidad como una sistemática guía al comportamiento bajo la incertidumbre.

[2] Entendiendo por *frecuencia* el límite al infinito del cociente de veces que aparece la cifra en los primeros n números entre n.

cuestión es nula.[3]

Es por esto que Hugo Steinhaus en su artículo *Les probabilités dénombrables et leur rapport à la théorie de la mesure*[4] publicado en Fundamenta Mathematicae (1923) trata de establecer un sistema de postulados para las probabilidades numerables que permita *de una vez por todas*[5] transitar entre estas interpretaciones.

Se limitará el trabajo a problemas donde el número de casos es finito pero el número de pruebas es infinito numerable; esto cae dentro de la primera categoría de Borel y el caso simple con dos posibles resultados e igualmente probables nos lleva al Teorema de Borel.

4.1 La probabilidad y su relación con la medida. Steinhaus.

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

Supóngase 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:

1. $\mu(E) \geq 0$ para todo E perteneciente a $\bar{\mathbb{A}}$.
2.
   a) Si E_n es un conjunto compuesto de 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 sólo difieren en la entrada $i-$ésima ($i \geq n$) entonces $\mu(E_n) = \mu(E'_n)$

   c) $\mu(A) = 1$

[3] Hausdorff, *Grundzüge der Mengenlehre, Leipsig, 1914, pp. 419-422.*

[4] Las probabilidades numberables y su relación a la Teoría de la Medida.

[5] Así lo enuncia Steinhaus. (1923)

3. Si E_i pertenece a $\bar{\mathbb{A}}$ para toda i y $E_i \cap E_j =$ para $i \neq j$[6] entonces $\sum_{i=1}^{n} E_i$ y $\sum_{i=1}^{\infty} E_i$ pertenecen a $\bar{\mathbb{A}}$ y

$$\mu\left(\sum_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} \mu(E_i) \text{ y } \mu\left(\sum_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \mu(E_i)^{7}$$

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 $\mu(E) = 0$, toda parte[8] de E' de E pertenece a $\bar{\mathbb{A}}$.

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. La función $\mu(\cdot)$ es la probabilidad buscada y nos aseguran los postulados $(1), (2)\ c), (3)$ y 4 que $0 \leq \mu(E) \leq 1$.

El postulado $2.a)$ exige que las partidas finitas consideradas en la teoría clásica de Probabilidad tengan una probabilidad definida; $2.b)$ expresa la equiprobabilidad (que se tiene de hipótesis) entre el rojo y negro y $2.c)$ determina la constante $\mu(A)$ en virtud del uso de la Probabilidad clásica.

(3) es un principio bien conocido : la probabilidad de una suma/unión de eventos mutuamente excluyentes[9] 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) 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[10] tendrá también probabilidad nula.

Estos postulados van de acuerdo a los postulados de la Probabilidad clásica, $2.a)$ y 3) generalizan sólo al caso numerable. Considérese entonces ahora $\mathbb{A}$ 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

[6] Aquí aunque se trata de conservar la notacin original, se ha creido necesario cambiar un cero por y incluir el simbolo $\cap$.

[7] Suma de conjuntos equivaldría a union sin importar si finita o numerable.

[8] Subconjunto.

[9] Steinhaus usa la palabra *incompatibles*.

[10] Steinhaus usa la expresión *todo evento más especial*.

trabajar con todos los E de $\mathbb{A}$ tales que $\mu(\cdot)$ y $\mathbb{A}$ satisfagan $(1)-(5)$; $\mu(\cdot)$ sería la probabilidad 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 entorno a $\bar{\mathbb{A}}$, ésta no siendo única. Sólo faltaría demostrar que esta $\mathbb{A}$ es una $\bar{\mathbb{A}}$.

Ahora se hará corresponder a toda partida finita ω (por ejemplo $\nu\rho\rho\nu\rho\nu\rho\rho\rho\ldots\nu\rho$) con un número real $(0.011010111\ldots01)$ cuyo número en desarrollo diádico comienza por cero y en la $n-$ésima cifra se pondrá cero o uno si el $n-$ésimo lugar de ω es negro o rojo.

Este arreglo hace corresponder a todo conjunto E de partidas infinitas sobre conjuntos puntuales del intervalo $[0,1]$[11] a los conjuntos E_i, correspondiéndoles intervalos de longuitud $\frac{1}{2^n}$ y especialmente $\mathbb{A}$ a el intervalo $[0,1]$. Desígnese por las mismas letras de los conjuntos de partidas a los conjuntos imágenes sin temor a confusión y tómese los postulados, por lo tanto, como válidos a los conjuntos de puntos lineales.

En esta interpretación los cinco postulados no son más que una repetición de los postulados de Sierpiński para definir los conjuntos L-medibles y su medida. Sólo difieren en sus segundos postulados en la redacción, pues son equivalentes como lo hacen notar el mismo Sierpinski y Steinhaus.[12] Sin embargo la independencia entre ellos es cuestión aparte.

Por otro lado, si se nombra por un momento como $3.a)$ y $3.b)$ las dos partes de 3) donde la primera obedece a la suma de un número finito de términos y la segunda a una de una cantidad númerable de términos; se podría bien preguntar si $1), 2)$ y$3.a)$ dan como consecuencia $3)-b)$. U. Broggi en su tesis doctoral[13] demuestra, en su disertación sobre los (posibles) axiomas de la Probabilidad, que en efecto sería consecuencia. Para senñalar el error de Broggi, basta con citar un contraejemplo: la generalización que presenta Banach de la medida de Lebesgue para todos los sub-conjuntos del [0,1], el cual demuestra la existencia de una función que satisface $1), 2)$ y $3.a)$ mas no $3)-b)$.[14]

[11]Steinhaus aún usa la notacin $< 01 >$.

[12]Para más detalle se puede consultar [46] [47].

[13]Broggi, U. (1907). Die Axiome der Wahrscheinlichkeitsrechnung. Ph.D. thesis, Universität G̈ttingen. Fragmentos publicados en Schneider (1988)

[14]Para más detalle se puede consular [1]

4.2 Aplicación

La probabilidad de una partida determinada sería entonces la medida del conjunto compuesto por un sólo punto, es decir, cero. Steinhaus propone nombrar a estos eventos como *casi-posible(s)* y a los eventos con probabilidad uno *casi-certeza(s).* Y recalca al igual que ya lo había hecho notar Borel desde 1905 que una *casi-certeza* no implica una certeza total, ni un *casi-posible* implica imposibilidad.

Sea $\phi(n)$ una función positiva, creciente y C una constante (independiente de n) que pueda tener los valores $+\infty$ y $-\infty$. ¿Cuál será la probabilidad que para $\phi(n)$ y C dadas se tenga

$$\limsup_{n\to\infty} \frac{\nu(n)}{\phi(n)} = C \ ? \tag{4.1}$$

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

Esta probabilidad no puede ser uno ni cero:

Como $\frac{\nu(n)}{\phi(n)}$ es una función del número *desarrollado* ω, se puede considerar como una $f_n(\omega)$, las cuales son medibles.[15] El límite superior de una sucesión de funciones medibles es medible y puede ser considerado como $g(\omega) = \limsup \frac{\nu(n)}{\phi(n)}$, por lo cual el conjunto de raíces de $g(\omega) = C$ es medible.

Este conjunto E es evidentemente el conjunto de puntos de $[0,1]$ cuyo desarrollo diádico satisface (4.1) con $\nu(n)$ siendo el número de ceros que aparecen en las primeras n cifras. Siendo $\alpha, \beta \in [0,1]$, $\alpha, \beta \neq \frac{1}{2^q}$ con $q \in \mathbb{Z}^+$. Supóngase que α y β sólo difieren en sus primeras q cifras, por lo cual el límite (4.1) sería el mismo. Entonces o bien $\alpha, \beta \in E$ o bien $\alpha, \beta \notin E$ es decir, pertenecen a su complemento.

Obsérvese 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. Supóngase 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 $k \neq i$, sólo habría que considerar el conjunto $E + (-1)^m 2^{-q}$ donde $m = i - k$. Como sólo se *cambian* un número finito de dígitos, el nuevo conjunto sigue cumpliendo (4.1) además de que sigue mi-

[15] L-medibles a menos que se señale lo contrario.

diendo lo mismo, pues sólo se desplaza un número real. Entonces la densidad de E es constante. Por lo cual sólo puede ocurrir que $m(E) = 0$ o $m(E) = 1$.

Con esta interpretación de $\nu(n)$, el límite superior es en sí una función de ω, el número que se desarrolla en binario. Ya se había notado que esta función es medible. Un teorema de Burstin nos asegura que la función es contante casi dondequiera.

¿Cómo se determinarían $\phi(n)$ y C de manera que la probabilidad de (4.1) sea 1? Sería lo hecho pero en "sentido inverso". Por el *Teorema de Bernoulli* sabemos que para $\epsilon > 0$ y $1 > \eta > 0$ dadas se tiene que para $n > N_{\epsilon,\eta}$

$$\left|\frac{\nu(n)}{n} - \frac{1}{2}\right| < \epsilon \tag{4.2}$$

con una probabilidad que sobrepasa a $1 - \nu$. Entonces el conjunto E de puntos cuyo desarrollo diádico tiene la propiedad de cumplir (4.2) a partir de un cierto n dentro de $[0,1]$ tiene una medida más grande que $1 - \eta$. Como se vió dados ϕ y C la probabilidad de que se cumpla el límite (4.1) es cero o uno, y en este caso la probabilidad es mayor que cero, entonces debe ser uno. Por lo anterior, el conjunto de puntos sin la propiedad es cero, el cual denótese por $D(\epsilon)$, 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 $n > Q_{\omega,m}$; m siendo grande si queremos que los puntos que satisfaga

$$\lim_{n\to\infty} \frac{\nu(n)}{n} = \frac{1}{2} \tag{4.3}$$

y la medida de su conjunto es uno en $[0,1]$.

Esto es precisamente *El Teorema de Borel* en su forma aritmética y con el postulado y la equivalencia señalada en su momento, permiten ir a la interpretación original. En realidad lo que Borel demuestra es para cualquier base q en lugar de 2. Se puede ver fácilmente que lo trabajado se puede extender a cualquier base sin problema. Una leve idea:

Los números ω que cumplen

$$\lim_{n\to\infty} \frac{\nu_p(n)}{n} = \frac{1}{q} \quad \text{para} \;\; p = 0, 1, 2, \ldots q-1$$

donde $V_p(n)$ significa lo análogo, el número de veces que aparece p en las primeras n cifras dentro la expresión q-ádica o q-naria de ω. Si $E(q)$ denota ahora el conjunto de medida cero de números que no cumplen la propiedad en cuestión para una base dada q, $\sum_{q=2}^{\infty} E(q)$ también tiene medida nula. Por lo cual casi todos los números tienen la propiedad para todas las bases q y la frecuencia relativa de todas las cifras es igual. Justamente estos números son los que Borel bautizó como *absolutamente normales.* Quien encontró por primera vez un número tal que cumpliría lo señalado fue dado por Sierpiński en 1917.

Por otro lado, sean c_n una sucesión numérica. ¿Cuál será la probabilidad que

$$\sum_{n=1}^{\infty} \pm c_n \tag{4.4}$$

converja si los signos son aleatorios, o se sacan pelotas de una urna que tienen el signo que llevará cada c_n?

La probabilidad que la serie converja será la medida del conjunto de puntos $\omega \in [0, 1]$ tales que la serie

$$\sum_{n=1}^{\infty} c_n f_n(\omega) \tag{4.5}$$

sea convergente, donde

$$f_n(\omega) = \begin{cases} (-1)^k & \text{si } \frac{k}{2^n} \leq \omega < \frac{k+1}{2^n} \\ -1 & \text{si } \omega = 1 \end{cases}$$

con $k \in \{0, 1, 2, \ldots, 2^n - 1\}$ y $n \in \mathbb{Z}^+$.

Hágase corresponder a cada símbolo positivo o negativo de la serie a una sucesión de números ceros y unos. El símbolo $+$ con 0 y $-$ con 1, de tal manera que se podría construir la expresión diádica de un conjunto entre cero y uno. Sea P el conjunto que resulta, es decir, el conjunto de todas las expresiones diádicas *construidas.* Por lo anterior, la medida del conjunto buscado, a saber

S, sería la misma que P. ¿Por qué P es medible? Porque tiene la misma constitución que el pasado conjunto E, que se construía de expresiones diádicas arbitrarias. Por lo cual se puede considerar la equivalencia entre:

$$\sum_{n=1}^{\infty} c_n f_n(\omega) \text{ y } \sum_{n=1}^{\infty} \pm c_n$$

Un teorema de H. Rademacher (1922) asegura que la convergencia de

$$\sum_{n=1}^{\infty} c_n^2$$

basta para asegurar la convergencia casi dondequiera de

$$\sum_{n=1}^{\infty} c_n f_n(\omega).$$

Lo cual hace dividir el conjunto de sucesiones c_n en dos: las cuales hacen a la serie casi-*segura* o casi-*imposible* convergente. Recuérdese que la densidad de un conjunto tal que satisface una condición, en este caso será que (4.5) sea convergente, no dependería del intervalo $[\frac{i}{2^q}, \frac{i+1}{2^q}]$ con $i \in \{0, 1, 2, .., 2^q - 1\}$ y tampoco de la q en cuestión pues ha sido tomada arbitrariamente. Entonces necesariamente como en el caso anterior, la probabilidad debe ser uno o cero.

4.3 Un teorema de Laplace.

Steinhaus demuestra el siguiente teorema bien conocido de Laplace; ahora con Teoría de la Medida. Si se recuerda que se ha nombrado por $\nu(n)$ el número de veces que el color negro en las primeras n pruebas y por $f(t, n)$ la probabilidad que se tenga

$$\left| \frac{\nu(n)}{n} - \frac{1}{2} \right| \leq \frac{t}{\sqrt{2n}} \tag{4.6}$$

se tiene

$$\lim_{n \to \infty} f(t, n) = \frac{2}{\sqrt{\pi}} \int_0^t e^{-x^2} dx$$

con $t \in \mathbb{R}$ fija.

Si ahora se designa por $F(t,n)$ el conjunto de puntos en los cuales las cantidades $\nu(n)$ de ceros del desarrollo diádico satisface (4.6) y $\mu(F(t,n))$ su medida[16] como para cualquier otro conjunto. Entonces se tiene la desigualdad siguiente:

$$\mu\left(\liminf_{n\to\infty} F(t,n)\right) \leq \lim_{n\to\infty} \mu\left(F(t,n)\right) = \lim_{n\to\infty} f(t,n) = \frac{2}{\sqrt{\pi}} \int_0^t e^{-x^2} dx < 1. \tag{4.7}$$

Aquí Steinhaus ya usa el concepto de Probabilidad casi indistintamente con el de la medida, midiendo el conjunto que tiene la propiedad de la cual se calcula su probabilidad; que se puede observar en la primera igualdad de (4.7). Pero el conjunto $S(t')$ de cuyos puntos cumplen:

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

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

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

y es aquí donde entra un razonamiento análogo usado anteriormente,[17] por lo cual

$$|S(t') = 0|$$

De lo que se desprende $\mu\left(\bigcup_{k=1}^{\infty} S(k)\right) = 0$ y fuera de $\bigcup_{k=1}^{\infty} S(k)$ (para casi todo) se tendrá, por la definición propia de $S(k)$

$$\limsup_{n\to\infty} \sqrt{2n}\mu\left(\frac{\nu(n)}{n} - \frac{1}{2}\right) = +\infty. \tag{4.8}$$

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

[17]Cuando se quería saber la probabilidad del límite $\limsup_{n\to\infty} \frac{\nu(n)}{\phi(n)} = C$.

Ahora considérese el límite superior de $\{F(t,n)\}$:

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

Y constrúyase el conjunto

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

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 de que se tiene para casi todo

$$\liminf_{n\to\infty} \sqrt{2n}\left|\frac{\nu(n)}{n} - \frac{1}{2}\right| \leq t';$$

t' siendo un número positivo arbitrario, por lo cual

$$\liminf_{n\to\infty} \sqrt{2n}\left|\frac{\nu(n)}{n} - \frac{1}{2}\right| = 0.$$

Por otro lado, considérese la siguiente desigualdad

$$\forall k, n \in \mathbb{Z}^+ : \quad \sum_{p=0}^{n} \binom{n}{p}(2p-n)^{2k} \leq 2^n(2k-1)!!n^k \tag{4.9}$$

donde $(2k-1)!! = 1 \cdot 3 \cdot 5 \cdot (2k-3)(2k-1)$.[18]

Ahora, la condición necesaria y suficiente para que casi todas las n primeras cifras del desarrollo diádico de un número α tenga p ceros y $n-p$ unos es que α esté situada en un intervalo perteneciente a una 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 una n fija, considerando todos los intervalos I_ϵ tales que los p correspondientes cumplan:

$$\left|\frac{p}{n} - \frac{1}{2}\right| > \epsilon \text{ con } \epsilon > 0 \tag{4.10}$$

y

$$\sum_{p'} \binom{n}{p}\frac{1}{2^n}$$

[18] La demostración se encuentra en el anexo y es una distinta a la que propone Steinhaus.

será entonces la longitud de los intervalos ajenos para p fija de longitud $\binom{n}{p}\frac{1}{2^n}$ y que satisfagan (4.10). Para k natural:

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

y por (4.9)

$$\sum_{p'}\binom{n}{p}\frac{1}{2^n}<\frac{1}{2^n(2n\epsilon)^{2k}}\sum_{p=0}^{n}\binom{n}{p}(2p-n)^{2k}<\frac{1}{(2n\epsilon^2)^k}\cdot\frac{(2k-1)!!}{2^k}=\frac{k!}{(2n\epsilon^2)^k}<\frac{k^ke^{-k}\sqrt{2\pi k}e^2}{(2n\epsilon^2)^k}$$

donde el primer término es la suma de longitudes de los I_ϵ y para la última desigualdad se propone tomar una η positiva y menor que la unidad tal que

$$2n\epsilon^2>(1+\eta)\ln n \text{ para } n\geq 3$$

$$k=[(1+\eta)\ln n],$$ [19]

$[\cdot]$ simbolizando la función parte entera. Con lo anterior se tiene

$$I_\epsilon<e^{\frac{1}{12n}}\frac{[(1+\eta)\ln n]^{[(1+\eta)\ln n]}e^{-[(1+\eta)\ln n]}\sqrt{2\pi[(1+\eta)\ln n]}}{((1+\eta)\ln n)^{[(1+\eta)\ln n]}}$$

$$<e^{\frac{1}{12n}}e^{1-(1+\eta)\ln n}\sqrt{2\pi(1+\eta)\ln n}<2e^{1+\frac{1}{12n}}\sqrt{\pi}\frac{\sqrt{\ln n}}{n^{1+\eta}}.$$

Si se nombra por Z_n el conjunto compuesto por todas los I_ϵ para n fija se tiene entonces

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

El conjunto límite, a saber, L de la sucesión de $\{Z_n\}$ tiene medida cero puesto que la serie

$$\sum_{n=3}^{\infty}e^{\frac{1}{12n}}\frac{\sqrt{\ln n}}{n^{1+\eta}}$$

[19]Steinhaus propone para ambas la igualdad. Me parece que es un error pues no siempre es posible escoger una *eta* que cumpla la igualdad para n y ϵ dados fijos.

converge y por lo tanto $\sum_{n=1}^{\infty}\mu(Z_n)$ también.

Si α no pertenece a L, no pertenece a ningun Z_n a partir de una $n > N(\alpha)$; lo cual se traducirá en negar (4.10):

$$\left|\frac{\nu(n)}{n} - \frac{1}{2}\right| \leq \epsilon$$

que es equivalente a

$$\sqrt{\frac{2n}{\ln n}}\left|\frac{\nu(n)}{n} - \frac{1}{2}\right| \leq \sqrt{1+\eta}$$

para una ϵ dada y para $n > N(\alpha)$ lo cual implica

$$\limsup_{n\to\infty}\sqrt{\frac{2n}{\ln n}}\left|\frac{\nu(n)}{n} - \frac{1}{2}\right| \leq \sqrt{1+\eta}. \tag{4.11}$$

Como L tiene medida nula, (4.11) aplica para casi toda α además como η puede ser arbitrariamente chica

$$\limsup_{n\to\infty}\sqrt{\frac{2n}{\ln n}}\left|\frac{\nu(n)}{n} - \frac{1}{2}\right| \leq 1 \tag{4.12}$$

para casi toda. Esta última (4.12) puede ser también interpretada como: casi seguramente la diferencia absoluta entre $\frac{1}{2}$ y el cociente del número de entradas negras entre el total n de entradas no sobrepasará $\sqrt{\frac{\ln n}{2n}}$ para partidas (realmente) grandes, regresando a la idea original.

4.4 Teoremas.

Ahora nos enfocaremos en el caso en el que la cifra $\rho_i -$ ésima sea cero, donde las ρ_i's constituyen en sí una $\{\rho_n\}$ sucesión. ¿Cuál sería su medida?

Véase los siguiente teoremas:

1. Sea $A \subseteq [0,1]$, medible y $\{\rho_n\}_{n\mathbb{N}}$ una serie de números naturales distintos entre ellos. Sea $B \subseteq [0,1]$ el conjunto de números cuyas cifras ρ_n-ésima se llamarán a_n en la expansión diádica que forman

$$\alpha = 0.a_1 a_2 a_3 \dots a_n \dots,$$

un número en A. Entonces B resulta ser medible y más aún su medida es la misma que la de A.

Lo anterior no es más que el enunciado del trabajo en la primera subsección del trabajo de Steinhaus.

2. Sean $A \subseteq [0,1]$ y $\{\rho_n\}$ la misma sucesión que en teorema anterior y B un conjunto compuesto sólo por los puntos con la propiedad expuesta en el Teorema I. La medida exterior de B no sobrepasa la de A.

   Tómese un conjunto medible $M(A)$ tal que $A \subseteq M(A)$ y la $m_e(A) = \mu(M(A))$[20] y B' sólo los números que cumplan la propiedad del Teorema I en relación al conjunto medible $M(A)$. Por el mismo Teorema I, se tiene que B' es medible y que $\mu(B) = \mu(M(A))$.

   Puesto que $B \subseteq B'$, entonces $m_e(B) \leq m_e(A)$.

3. Ahora, si se tienen dos conjuntos $A, B \subseteq [0,1]$ y dos sucesiones $\{\rho_n\}_{n\in\mathbb{N}}$ y $\{\sigma_n\}_{n\in\mathbb{N}}$ del mismo género que en el Teorema I y II; y si las cifras de índices ρ_i de un número de B da un número de A y las cifras de σ_i dan también un número de A, entonces las medidas de A y B son iguales.

   Su demostración es usar el Teorema II.

   Recuérdese como Borel describe a los números *normales*: aquellos en los que el $\lim_{n\to\infty} \frac{\nu(n)}{n} = \frac{1}{2}$. Ahora llámese a α *normal respecto a* β $(\alpha - R - \beta)$ si las cifras del desarrollo de α que ocupan el mismo rango que los ceros en el desarrollo de β, satisfacen también $\lim_{n\to\infty} \frac{\nu(n)}{n} = \frac{1}{2}$. Por ejemplo:

   Si

$$\alpha = 0.0100110111011001\dots$$

[20]Recuérdese que $m_e(\cdot)$ es la medida exterior de un conjunto, propiedad que cualquier conjunto acotado tiene.

y

$$\beta = 0.0001010011010100\ldots$$

se construye un número $\gamma = 0.010010101\ldots$, esto es, si en β hay un cero, se ve que dígito se tiene en el mismo rango en α, y se añade a γ. Posteriormente se verifica la normalidad de γ.

De lo anterior se puede demostrar: Todo número normal α es normal respecto a *casi todo* número β.

Demostración.

Sea α un número normal fijo y sea $k_1(\nu)$ el rango del $\nu-$ésimo cero en el desarrollo de α. [21] Considérese las cifras de un número β, tómese las cifras $k_1(\nu)-$ésimas para $\nu \in \mathbb{Z}^+$. Desígnase por $N_1(\nu)$ el número de ceros que hay entre estas últimas cifras, se tiene para casi toda β

$$\lim_{\nu\to\infty} \left| \frac{N_1(\nu)}{\nu} - \frac{1}{2} \right| = 0 \tag{4.13}$$

Ahora, sea $k_2(\rho)$ el rango del $\rho-$ésimo uno en el desarrollo de α. Si $N_2(\rho)$ significa el número de ceros que hay en el número β en los rangos $k_2(\rho)-$ésimos, se tendrá también que para casi toda β

$$\lim_{\rho\to\infty} \left| \frac{N_2(\rho)}{\rho} - \frac{1}{2} \right| = 0 \tag{4.14}$$

Ahora sólo se consideran las cifras de α que coinciden en rango a los ceros de β. De los $\nu(n)$ ceros inicialmente sólo se toman los $N_1(\nu)$ y de los ρ unos sólo se quedan $N_2(\rho)$. De la definición de normal y de (4.13) y (4.14) se tiene

$$\lim_{n\to\infty} \frac{N_1(v(n))}{N_1(v(n)) + N_2(\rho(n)} = \frac{1}{2} \tag{4.15}$$

puesto que $N_1 = \frac{\nu}{2} + o(\nu)$ donde $\nu(n) = \frac{n}{2} + o(n)$ y $N_2(\rho) = \frac{\rho}{2} + o(\rho)$ donde

$\rho(n) = \frac{n}{2} + o(n)$.

[21] Por ejemplo si $\alpha = 0.10110\ldots$ entonces $k_1(1) = 2, k_1(2) = 5$.

Esto implica que

$$N_1(\nu) = \frac{n}{4} + o(\frac{n}{2} + o(n)) = \frac{n}{4} + o(n)$$

$$N_2 = \frac{n}{4} + o(n)$$

Y (4.15) conduce al teorema propuesto; $N_1(\nu)$ es el número de ceros tomados en cuenta, $N_1 + N_2$ el número de cifras totales tomados en cuenta que quedan de las n primeras cifras de α dando como resultado

$$\lim_{n\to\infty} \frac{N_1}{N} = \frac{1}{2} \tag{4.16}$$

para casi toda β, donde N_1 designa el número de ceros tomados en cuenta en los N primeras cifras no eliminadas de α.

Para demostrar su inverso, es decir: *Si α es normal en relación a casi todo número β entonces α es normal.*

Supóngase lo contrario, esto equivale a la existencia de una serie n_k creciente de números naturales tales que

$$\begin{aligned} \nu(n_k) &= zn_k + o(n_k) \\ \rho(n_k) &= (1-z)n_k + o(n_k) \end{aligned} \tag{4.17}$$

con $z \neq \frac{1}{2}$ con las mismas $v(n)$ y ρ que se han venido usando. Ahora tomamos los $N_1(\nu(n_k))$ ceros de los cifras que coinciden en rango con las cifras de β, como se hizo anteriormente, y los $N_2(\rho) = \frac{\rho}{2} + o(\rho)$

con (4.17) se tiene

$$N_1(\nu(n_k)) = \frac{z}{2}n_k + o(zn_k + o(n_k)) + \frac{1}{2}on(n_k)$$

$$N_2(\rho(n_k)) = \frac{1-z}{2}n_k + o((1-z)n_k + o(n_k)) + \frac{1}{2}o(n_k)$$

de donde se deduce

$$N_1(\nu(n_k)) = \frac{z}{2}n_k + o(n_k), \quad N_2(\rho(n_k)) = \frac{1-z}{2}n_k + o(n_k)$$

y

$$\lim_{k\to\infty} \frac{N_1(\rho(n_k))}{N_1(\nu(n_k)) + N_2(\rho(n_k))} = z \neq \frac{1}{2}$$

para casi toda β, lo que contradice el teorema.

Ahora, en un caso más general no suponiendo que rojo y negro tienen la misma probabilidad pero que sí $p + q = 1; p, q > 0$.

Recuérdese los postulados al inicio de esta sección sobre el trabajo de *Steinhaus*, se debe redefinir b) de la siguiente forma:
b) Si E_n y E'_n sólo difieren en la $i-$ésima entrada ($i \geq n$) entonces $\mu(E_n) : \mu(E'_n) :: p : q$.

El razonamiento se vuelve análogo al anterior, de donde se rescata lo siguiente, se tiene el teorema de Bernoulli

$$\left|\frac{\nu(n)}{n} - q\right| < \epsilon \text{ para } n > N(\epsilon\eta)$$

con una probabilidad mayor a $1 - \eta$, más aún

$$\lim_{n\to\infty} \frac{v(n)}{n} = q, \lim_{n\to\infty} \frac{\rho(n)}{n} = p$$

para casi todo número α con $\nu(n), \rho(n)$ teniendo el mismo significado.[22]

También el Teorema de Laplace, afirma que la probabilidad $f(t, n)$ de la desigualdad siguiente (para $t \geq 0$)

$$\left|\frac{v(n)}{n} - q\right| \leq \frac{t\sqrt{pq}}{\sqrt{n}},$$

tiende a

$$\frac{1}{\sqrt{\pi}} \int_0^t e^{-x^2} dx$$

[22] Recuérdese que α es la representación binaria de a.

conforme $n \to \infty$ y si se deja t fija:

$$\limsup_{n\to\infty} \sqrt{\frac{n}{2pq}} \left| \frac{v(n)}{n} - q \right| = \infty$$

$$\liminf_{n\to\infty} \sqrt{\frac{n}{2pq}} \left| \frac{v(n)}{n} - q \right| = 0$$

lo cual se podría generalizar con algunas técnicas del cálculo para obtener algo equivalente a lo final cuando se trabajó con equiprobabilidad.

Se demostrará el siguiente teorema útil para lo que seguirá. Si E es un conjunto de partidas con una probabilidad definida $\mu(E)$ y si se considera una sucesión de indices diferentes $i_1, i_2, \ldots i_n$ y el conjunto E' de partidas que se obtienen al hacer corresponder toda partida $a \in E$ todas las partidas a' cuyas $i_1, i_2, \ldots, i_n-$ésima entrada son iguales a la $1^{er}, 2^{da} \ldots n-$ésima entrada de a. Entonces la probabilidad $\mu(E')$ es igual a $\mu(E)$.

La demostración sólo consiste en aplicar el Teorema I de esta subsección percatándose que se debe cambiar en los postulados la definición de E_n. Para E en este caso se mantendrá igual pero para E' en lugar de las primeras n entradas iguales, serán las $i_1, i_2, \ldots, i_n-$ésima entradas iguales. Y convenserse que la probabilidad aquí de nuevo es sólo un accesorio de la Teoría de la Medida.

Capítulo 5

Después de Borel y Steinhaus.

Probability as a measure of degree of belief and Probability Theory as a systematic guide to behaviour under uncertainty. Anónimo

Como se vió, muchas personas quisieron aplicar la poderosa Teoría de Integración de Lebesgue a la probabilidad. Pero, ¿por qué ninguna funcionó a la primera? ¿Por qué sería hasta 1931? Una posible respuesta sería que el trabajo aplicado de la Teoría de la Probabilidad por lo ge-neral no estaba en un nivel que sea necesario un uso propio y formal de la Teoría de la Medida: cuando se manejan conjuntos infinitos o sucesiones de eventos.

De Finetti publicó en los comienzos de 1930[1] su subjetivista programa sobre la fundación de la Probabilidad, como resultado del trabajo Bayesiano que habría desarrollado en respuesta al indeterminismo en el desarrollo científico en los últimos años de 1920. En la filosofía Bayesiana, la posibilidad y la probabilidad tienen posiciones diferentes, donde la probabilidad no es derivada de nociones, sino es parte del sistema del conocimiento.

La probabilidad, en sí, es interpretada como un concepto primitivo parte del comportamiento humano bajo la incertidumbre. Su base natural es cualitativa, pero puede también medirse numéricamente, a través de radios que uno puede ir aceptando para la ocurrencia del evento que en sí, es desconocido. Con esto como base, Bruno de Finetti creó la Probabilidad cualitativa y muestra como las propiedades numéricas usuales de la probabilidad puden ser vistas como resultado de la noción de *apuestas coherentes*.

[1]Republicados en *Scritti (1926-1930)*, Cedam, Pádua .

Se define $E \preceq E'$ como el evento E es almenos 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'. Los axiomas son los siguientes cuatro:

1. $E \preceq E'$ o $E' \preceq E$ para cualesquiera dos eventos E, E'.
2. Si un 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 sólo si $E_1 \preceq E_2$. Específicamente $E_1 \approxeq E_2$ si y sólo si $E_1 + E \approxeq E_2 + E$.

De Finetti considera la idea de apuestas coherentes como un pensamiento conveniente, parcialmente arbitrario y un camino simplificado para introducir la probabilidad *numérica*. Si se asume 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 si E ocurre o no, respectivamente. Si se apuesta varias veces, de manera finita, la apuesta sería

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

Donde la coherencia entraría al requerirle al banquero no maneje el esquema de pérdida segura, esto es, que G sea positiva para algúna suma de S_i's, bajo todos los posibles resultados de las E_i's. De Finetti prueba que (su) coherencia implica que los números p_i's asociados a los eventos E_i's satisfacen los axiomas de aditividad finita de la Probabilidad clásica.

5.1 La axiomatización de Kolmogorov

Casi a la par, aparece en 1933 *Grundbegriffe der Wahrscheinlichkeitsrechnung*,[2] que en realidad más que un reporte de investigación es una exposición en el

[2] Fundamentos de la Teoría de la Probabilidad.

cual trata de llenar un vacío que existía. Así como la Teoría de la Medida e Integración se habrían desprendido de la Geometría, se quería una Probabilidad en una completa presentación de un sistema libre de complicaciones superfluas, como lo afirmó Kolmogorov.[3] Esto lo consiguió en sólo 62 páginas escritas en ruso, seguido por dos grandes volúmenes por Fréchet en 1937-38; obra que quedó como un pie de página al logro del propio matemático ruso.

Kolmogorov empieza con cinco axiomas, considerando un conjunto E y $\mathcal{F}$ de subconjunto de E que llamó y llamaremos eventos aleatorios:

1. $\mathcal{F}$ es un álgebra de conjuntos
2. $\mathcal{F}$ contiene al conjunto E.
3. Para cada conjunto A de $\mathcal{F}$ se le asigna un número real no negativo $\mathbb{P}(A)$. A éste se le nombra la probabilidad del evento A.
4. $\mathbb{P}(E) = 1$
5. Si A y B son disjuntos 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 sería el axioma de continuidad y dados los cinco primeros, sería equivalente a la aditividad numerable de la probabilidad.[4]

Se debe señalar por su peso histórico que Kolmogorov está pensando en álgebra como se le conoce hoy en día pero define una álgebra de Borel, lo que hoy sería conocida como una simple $\sigma-$álgebra y aún tiene también lo que hoy se conoce como $\sigma-$álgebra de Borel, pero aún sin nombre en particular.

[3] Pensamiento que fue plasmado en la obra de Fréchet años después.

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

También el libro ya muestra Teoremas importantes de convergencia, entre ellos el planteado originalmente por Bernoulli y llevado a su forma general por Slutsky para 1925; como también Cadenas de Markov discretas, ya definidas.

5.1.1 La reacción y después.

Las reacciones en los primeros años después de la publicación de Kolmogorov se pueden resumir a dos comentarios, uno positivo y uno no tanto:[5]

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. Willy Feller, publicado en el *Zentralblatt*.[6]

La otra, no tan positiva es de Henry L. Rietz publicado en el *Bulletin of the American Mathematical Society*:

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 Émile Borel, first in analysis by his notion of measure, and then in probability theory by his theory of denumerable probabilities[7]

Esta última refleja la realidad. El crédito no es sólo de Kolmogorov. Como se afirmó al principio de este capítulo, sólo es un compendio, el merito está en que por fin fue presentado sin huecos y bien ordenado. Aunque para el imaginario colectivo matemático la real axiomatización se debe a él, no se debe

[5]Fréchet *Les mathématiques et le concret.* pp. 153-154

[6]Este fragmento, a su vez, fue publicado en [42]

[7]*Idem.*

olvidar que esto fue posible gracias a Borel y al trabajo de Steinhaus, donde ya se ve (bien) formulada la axiomatización que *retomaría* Kolmogorov.

Antes del siguiente comentario quizá debería de tenerse en mente que en estos años estalla la Segunda Guerra Mundial, para tener un escenario mundial. Por un lado, también, es recibido con deficiencias el trabajo de Kolmogorov, pues no es inmediata la aplicación para la no-tan-naciente Teoría de los Procesos Estocásticos y por otra, no se conoce muy pronto. Por el otro, la Escuela Francesa quiere proponer su propia axiomatización, con el rechazo del propio Borel hacia el trabajo del matemático ruso, además de que la Escuela Italiana también sigue en su intento haciendo caso omiso (intencional o no) al mismo trabajo.

William Feller, que trabajaba con Harald Cramér fue el mayor entusiasta adherente al *Grundbegriffe* y se reflejó en el apoyo que le dió Cramér al proponerlo en sus libros de texto y posteriormente en los años 50's lo tradujera al Inglés. Finalmente, sería Doob quien daría sobre el trabajo de Kolmogorov el tratamiento definitivo de los Procesos Estocásticos en el marco de la Teoría de la Medida.

Antes de entrar a la Probabilidad de hoy, hay que resaltar unas cosas más sobre la exposición de Kolmogorov respecto a los axiomas. El autor trabaja sobre un álgebra aunque después afirme que es posible extender todo a una $\sigma-$álgebra. ¿Por qué lo hace así? Una posible explicación sería porque sigue reinando el mundo finito en su mente. Aunque sabe que existe el plano infinito, él sigue pensando en la idea de que las cosas no sólo se deben definir en un número finito de palabras, sino deben tener bases sólidas en caso finito; a pesar de que hace resaltar que el sistema (1)-(5) de axiomas, sólo dejan una teoría consistente pero incompleta.Ya después se modificará esto y se definirá la probabilidad sobre una $\sigma-$álgebra como se verá a continuación.

Para la decada de 1950, Kolmogorov mismo retoma la corriente frecuentista y escribe que en 1933 no enunció como la probabilidad es aplicada porque en realidad no lo sabía.[8] Su publicación contiene un acercamiento frecuentista a problemas aplicados de Probabilidad.

[8]Kolmogoeov, A. (1963) The theory of probability, in A.D. Aleksandrov, A. Kolmogorov and M. Lavrent'ev (eds.) *Mathematics, Its Contents, Methods and Meaning*, vol. 2. 229-264, MIT Press (Original ruso de 1956).

Kolmogorov entonces retoma las ideas de von Mises y concluye que su Teoría de la Probabilidad con medida sería aplicable a una secuencia de resultados empiricos en el sentido de von Mises, sólo si esa secuencia puede ser probada como aleatoria en el sentido modificado finito. Pero las secuencias de von Mises no son finitas, sin embargo todo experimento lo es. Esto desarrolla más teoría que queda fuera de los alcances e intereses de este trabajo. Sin embargo, es importante recalcar, que Kolmogorov mismo, daba a entender que no estaba bien fundamentada la teoría. Aún había situaciones que resolver.

5.2 La probabilidad axiomatizada hoy en día.

Un trabajo que exhibe la historia (o al menos intenta mostrar sus mayores exponentes) no puede terminar sin mostrar el final. El fin de esta sección es mostrar como se tiene ahora a la probabilidad.

La probabilidad se muestra ahora como una medida. Sea Ω un espacio de estados:

Una medida de probabilidad definida sobre una $\sigma-$álgebra $\mathcal{A}$ de Ω es una función $\mathbb{P} : A \to [0, 1]$ *que satisface:*

1. $\mathbb{P}(\omega) = 1$
2. Para toda sucesión numerable $(A_n)_{n\geq 1}$ de elementos de $\mathcal{A}$, disjuntos dos a dos, se tiene:

$$\mathbb{P}\left(\cup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n)$$

Donde a (2) se le conoce como la $\sigma-$aditividad y $\mathbb{P}(A)$ es llamada la probabilidad del evento A.

La independencia de eventos se define:

1. Sean A y B dos eventos. A y B son independientes si $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$
2. Sea $(A_i)_{i\in I}$ una colección (posiblemente infinita) de eventos, ésta es una colección de eventos independientes si para todo subconjunto finito $J \subseteq I$ se tiene

$$\mathbb{P}(\cap_{i\in J}A_i) = \prod_{i\in J}\mathbb{P}(A_i).$$

Por otro lado, la probabilidad condicional de A dado B será

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)}$$

siempre y cuando $\mathbb{P}(B) > 0$.

Ahora, sea $\mathbb{P}$ una probabilidad sobre el conjunto finito Ω; ésta se le llamda uniforme si $\mathbb{P}(\{\omega\})$ no depende de $\omega \in \Omega$. De aquí no es difícil demostrar que se tiene:

$$\mathbb{P}(A) = \frac{\#(A)}{\#(\Omega)}, \tag{5.1}$$

es decir, se tiene la definición clásica de probabilidad. Con lo cual, en primera instancia podría afirmarse que se está en buen camino.

Sea X una variable aleatoria, es decir que si $(\Omega, \mathcal{A}, \mathbb{P})$ es un espacio de probabilidad y $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ un espacio medible; $X : \Omega \to \mathbb{R}$ es $\mathcal{A}, \mathcal{B}(\mathbb{R})-$medible. Y la distribución de X se puede definir como:

$$\forall B \in \mathbb{R} : \qquad \mathbb{P}^X(B) = \mathbb{P}(\{\omega : X(\omega) \in B\}) = \mathbb{P}(X^{-1}(B)) = \mathbb{P}(X \in B)$$

Y la función de distribución inducida por una probabilidad $\mathbb{P}$ sobre $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ será la función

$$F(x) = \mathbb{P}((-\infty, x])$$

,

pudiéndose demostrar su equivalencia con los anteriores preceptos de esta función:

1. F es no-decreciente
2. F es continua por la derecha

3. $\lim_{x\to-\infty} F(x) = 0$ y $\lim_{x\to+\infty} F(x) = 1$

Si X es una v.a.[9] simple, es decir toma sólo un número finito de valores se puede escribir de la forma:

$$X = \sum_{i=1}^{n} a_i \mathbb{I}_{\{A_i\}}$$

donde $a_i \in \mathbb{R}, A_i \in \mathcal{A}$, $i \leq i \leq n$ y $\mathbb{I}_{\{\cdot\}}$ es la función indicadora usual. Y su esperanza será el número:

$$\mathbb{E}\{X\} = \sum_{i=1}^{n} a_i \mathbb{P}(A_i) = \int X(\omega)\mathbb{P}(d\omega) = \int X d\mathbb{P}$$

Defínase para toda v.a. positiva $E\{X\} = \sup_{0\leq Y\leq X}\{\mathbb{E}\{\mathbb{Y}\}\}$ donde Y son v.a.'s simples, y defínase $X^+ = \max\{0, X\}$ y $X^- = -\min\{0, X\}$.

Considérese la siguiente expresión:

$$\mathbb{E}\{X\} = \mathbb{E}\{X^+\} + \mathbb{E}\{X^-\}$$

se dice que X tiene esperanza finita, si la $\mathbb{E}\{X^+\}, \mathbb{E}\{X^-\} < \infty$, o bien que admite esperanza si alguno de los dos no diverge a ∞.

No debe ser difícil encontrar toda la similitud con la Teoría de Integración de Lebesgue, de donde pareciese que se calca todo, es decir, se tiene entre otras cosas que se cumple:

1. El Teorema de Convergencia Monótona
2. El Lema de Fatou
3. El Teorema de Convergencia Dominada de Lebesgue.

Un teorema importante es el siguiente:

Sea X una v.a. sobre $(\Omega, \mathcal{A}, \mathbb{P})$ con valores en $(E, \mathcal{E})$ y distribución $\mathbb{P}^X$. Sea $h : (E, \mathcal{E}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ medible entonces:

1. $h(X) \in \mathcal{L}(\Omega, \mathcal{A}, \mathbb{P})$ si y sólo si $h(X) \in L(E, \mathcal{E}, \mathbb{P}^X)$

[9] Variable aleatoria, abreviatura usual.

2. Si h es positiva o bien cumple la condición (1) inmediata anterior entonces:

$$\mathbb{E}\{h(X)\} = \int (x)\mathbb{P}^X(dx).$$

y como corolario otro enfoque con la teoría clásica:

Supóngase X es una v.a. con densidad f, es decir, tal que $F(x) = \mathbb{P}(X \leq x)$ y $F(x) = \int_{-\infty}^{x} f(u)du$ para $x \in (-\infty, \infty)$; si $\mathbb{E}\{|h(X)|\} < \infty$ o si h es positiva entonces

$$\mathbb{E}\{X\} = \int h(x)f(x)dx.$$

La independencia, ahora, de sub-$\sigma-$álgebras $(\mathcal{A})_{i\in I}$ de $\mathcal{A}$ se define si $\forall J \subseteq I$ finito y $\forall A_i \in \mathcal{A}_i$ se tiene que

$$\mathbb{P}(\cap_{i\in J} A_i) = \prod_{i\in J} \mathbb{P}(A_i),$$

es decir, de manera totalmente análoga a como se define para eventos y se ha definido por siglos; y se dirá que las $(X_i)_{i\in I}$ v.a.'s con valores en $(E_i, \mathcal{E}_i)$ son independientes si las σ-álgebras generadas por $X_i^{-1}(\mathcal{E}_i)$ son independientes. donde $E_i \subseteq \mathbb{R}$ y $\mathcal{E}_i \subseteq \mathcal{B}(\mathbb{R})$. De donde se desprenden como equivalencias:

1. $\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(X \in B)$
2. $\mathbb{E}\{h(X)g(Y)\} = \mathbb{E}\{h(X)\}\mathbb{E}\{g(Y)\}$

Se cerrará este trabajo mostrando dos teoremas que fueron enunciados en el marco de la axiomatización de la Teoría de la Probabilidad con el cual se unieron dos importantes ramas de la Matemática: El Lema de Borel- Cantelli y la Ley Cero-uno de Kolmogorov.

Para mostrar el *Lema Borel-Cantelli* como se tiene hoy en día es necesario introducir una definición y enunciar un teorema que no se demostrará:

Sea $\{A_n\}$ una sucesión de eventos en $\Im$. Definimos $\limsup_n A_n = \cap_{n=1}^{\infty}(\cup_{m\geq n A_m}) = \lim_{n\to\infty}(\cup_{m\geq n} A_m)$

Este evento puede ser interpretado probabilísticamente como

$$\limsup_n A_n = \{A_n \text{ ocurre para un número infinito de } n.\}$$

Como simple notación. $\limsup_n A_n = \{A_n \ \ i.o.\}$[10]

Teorema 5.2.1 (*Lema de Borell-Cantelli*)**.** *Sean* $\{A_n\}$ *eventos en* $(\Omega, \Im, \mathbb{P})$.

1. *Si* $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$ *entonces* $\mathbb{P}(A_n \ i.o.) = 0$

2. *Si* $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$ *y además* $\{A_n\}$ *son independientes entonces* $\mathbb{P}(\limsup_n A_n) = 1$ [11].

Para la Ley de Kolmogorov, es necesario una definición:

Sean X_n v.a.'s definidas en un espacio de probabilidad $(\Omega, \mathcal{A}, \mathbb{P})$, defínase la $\sigma-$álgebra $\mathcal{C}_\infty = \cap_{n=1}^{\infty} \mathcal{C}_n$. Ésta es conocida en inglés como *tail* $\sigma-$álgebra.

Teorema 5.2.2 (*Ley Cero-uno de Kolmogorov*)**.** *Sea* X_n *v.a.'s independientes definidas en un* $(\Omega, \mathcal{A}, \mathbb{P})$ *y* $\mathcal{C}_\infty$ *su tail* $\sigma-$*álgebra. Si* $C \in \mathcal{C}_\infty$ *entonces* $\mathbb{P}(C) = 0, 1$.

[10] i.o. = infinitely often

[11] También se suele enunciar de la siguiente manera: Si A_n son independientes y si $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$ entonces $\mathbb{P}(\{A_n \ \ i.o.\}) = 1$

Anexos. A

Resultados varios y propios.

A.1 Wiman. La medida acaricia la probabilidad.

1. Si se tiene el número real x, primero se toma $[x] = a = 0$, es decir, la parte entera; ahora, $x = [x] + x_1$ p.a. $x_1 \in [0, 1]$, después tómese $[\frac{1}{x_1}] = a_1$ donde $\frac{1}{x_1} = a_1 + x_2$ con $x_2 \in [0, 1]$ y así $[\frac{1}{x_i} = a_i]$ y $\frac{1}{x_i} = a_i + x_{i+1}$.

   Así se construiría

$$a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cfrac{1}{a_4 + \dots}}}} \tag{A.1}$$

A.2 El Lema de Borel Cantelli hoy en día.

Esta sección tiene fines comparativos: hoy en día como se presenta el Lema de Borel Cantelli y algunas generalizaciones, teniendo en mente como lo presentó Borel y el trato que le dió Cantelli al mismo, antes de ser conocido con este nombre.

Para probar el teorema, es conveniente tener en mente otro teorema:

Teorema A.2.1. *Sean* $\{X_n\}$ *v.a.*

(a) Si todas las X_n *son positivas entonces* $\mathbb{E}\{\sum_{n=1}^{\infty} X_n\} = \sum_{n=1}^{\infty} \mathbb{E}\{X_n\}$ *siendo ambos lados finitos o divergentes.*

(b) Si $\sum_{n=1}^{\infty} \mathbb{E}\{|X_n|\} < \infty$ *entonces* $\sum_{n=1}^{\infty} X_n$ *a.s. y la suma de estas series es integrable, más aún, se tiene la parte* (1)*.*

Ahora sí, se enuncia y demuestra el teorema:

Teorema A.2.2 (Lema de Borel-Cantelli)**.** *Sean* $\{A_n\}$ *eventos en* $(\Omega, \Im, \mathbb{P})$*.*

(a) Si $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$ *entonces* $\mathbb{P}(A_n \ i.o.) = 0$

(b) Si $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$ *y además* $\{A_n\}$ *son independientes entonces* $\mathbb{P}(\limsup_n A_n) = 1$ [1]

Demostración.

(a) Sea $a_n = \mathbb{P}(A_n) = \mathbb{E}(1_{A_n})$. Se sabe que

$$\sum_{n=1}^{\infty} a_n < \infty \Rightarrow \sum_{n=1}^{\infty} 1_{A_n} \quad a.s.^2$$

Por otro lado:

$$\sum_{n=1}^{\infty} 1_{A_n}(\omega) = \infty \iff \omega \in \limsup_n A_n$$

de lo que se sigue de inmediato (1).

Aquí está una muestra clarísima como ayudó **La integral de Lebesgue** a simplificar las demostraciones (las maravillas que puede hacer: recuérdese la demostración de Borel).

[1] También se suele enunciar de la siguiente manera: Si A_n son independientes y si $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$ entonces $\mathbb{P}(\{A_n \ i.o.\}) = 1$

[2] Por el teorema anterior.

(b) Basta con demostrar que $\mathbb{P}(\liminf_n A_n^c) = 0$. Para esto veamos que por la independencia de eventos:

$$\begin{aligned}
\mathbb{P}(\liminf_n A_n^c) &= \mathbb{P}(\cap_{k=n}^{\infty} A_k^c) \\
&= \lim_{N\to\infty} \mathbb{P}(\cap_{k=n}^{N} A_k^c) \\
&= \lim_{N\to\infty} \prod_{k=n}^{N} \mathbb{P}(A_k^c) \\
&= \lim_{N\to\infty} \prod_{k=n}^{N} (1 - \mathbb{P}(A_k)) \\
&\leq \lim_{N\to\infty} \prod_{k=n}^{N} \exp -\mathbb{P}(A_k) \\
&= \lim_{N\to\infty} \exp -\sum_{k=n}^{N} \mathbb{P}(A_k) \\
&= 0
\end{aligned}$$

puesto que la serie diverge por hipótesis.

Ahora como

$$\mathbb{P}(\{A_n \quad i.o.\}) = 1 - \mathbb{P}(\{A_n \quad i.o.\}^c) = 1 - \mathbb{P}(\liminf_n A_n^c) = 1$$

Es preciso comentar, que Arquímides ya habra demostrado que si se tiene una figura geométrica y se divide infinitamente, se podrá conseguir subfiguras tan chicas como se quieran. Es un precedente del primer inciso de **Lema B-C.**

A.2.1 Generalizaciones y consecuencias de el Lema B-C.

Antes de seguir es necesario definir Esperanza condicional y Martingalas.

Definición A.2.3. *Sea* $(\Omega, \mathcal{A}, \mathbb{P})$ *un espacio de probabilidad. Sea* X *una v.a. real y* $\mathcal{B}$ *una sub*$-\sigma-$*álgebra tal que* $\mathcal{B} \subseteq \mathcal{A}$. *Entonces una esperanza condicional de* X *dado* $\mathcal{B}$ *es cualquier función* $\mathcal{B}-$*medible* $\mathbb{E}(X|\mathcal{B} : \Omega \to \mathbb{R}$ *que satisfaga:*

$$\int_B \mathbb{E}(X|\mathcal{B})(\omega) d\mathbb{P}(\omega) = \int_B X(\omega) d\mathbb{P}(\omega)$$

para cada $B \in \mathcal{B}$.

De nuevo, sean $(\Omega, \mathcal{A}, \mathbb{P})$ un espacio de probabilidad y sean $(\mathcal{A}_n)_{n\geq 0}$ una sucesión de $\sigma-$álgebras tales que $\mathcal{A}_\backslash \subset \mathcal{A}_{\backslash+\infty} \subset \mathcal{A}$ para toda $n \geq 0$.

Definición A.2.4. *Una* $(X_n)_{n\geq 0}$ *una sucesión de v.a. es llamada martingala o una* $\mathcal{A}_n-$*martingala si para cada* n*:*

(a) $\mathbb{E}(|X_n| < \infty)$

(b) X_n *es* $\mathcal{A}_n-$*medible*

(c) $\mathbb{E}(X_n|\mathcal{A}_\Updownarrow) = X_m$ *a.s. para toda* $m \leq n$.

Teorema A.2.5. *Sea* $\{E_n\}$ *una sucesión de eventos independientes por pares tal que*

$$\sum_{n=1}^{\infty} \mathbb{P}(E_n) = \infty$$

entonces $\mathbb{P}(\limsup_n E_n) = 1$

A pesar de que este teorema último nos permite considerar solamente independencia por pares, sigue siendo una condición muy fuerte.

Es inmediato reconocer que es una generalización del *Lema de Borel-Cantelli* original y apartir de la demostración de este tenemos que si $\{E_n\}$ es una sucesión de eventos independientes por pares, entonces

$$\mathbb{P}(\limsup_n E_n) = 0 \iff \sum \mathbb{P}(E_n) < \infty$$

Teorema A.2.6. *Sea* $\{E_n\}$ *una sucesión de eventos tal que* $\sum_{n=1} \mathbb{P}(E_n) = \infty$ *y*

$$\liminf_n \frac{\sum_{j,k\leq n} \mathbb{P}(E_j \cap E_k)}{(\sum_{k\leq n} \mathbb{P}(E_k))^2} \leq 1$$

entonces $\mathbb{P}(\limsup_n E_n) = 1$

Teorema A.2.7. *Si* $\{X_n\}$ *es martingala y*

$$E\{sup_{n\geq 0}\{X_{n+1} - X_n\}\} < \infty$$

con $X_0 = 0$

entonces $\lim_{n\to\infty} X_n$ *existe y es finito para casi toda* $\omega \in \Omega$ *y* $\limsup_n X_n(\omega) < \infty$

Corolario A.2.8. *Sean* $Y_1, Y_2, \ldots$ *una sucesión de v.a. uniformemente acotadas y sea* $p_j = \mathbb{E}[Y_j | Y_1, Y_2, \ldots, Y_{j-1}]$. *Entonces la serie* $\sum_1^\infty (= \infty)$ *casi seguramente (c.s.) en el conjunto de* $\omega \in \Omega$ *donde* $\sum_1^\infty p_j(\omega)$ *converge (diverge).*

Corolario A.2.9 (Lévy)**.** *Sea* $\{E_n\}$ *una sucesión de eventos. Si*

$$\sum_{n\geq 1} \mathbb{P}(E_n | \Im_{n-1}) < \infty (= \infty) \quad c.s.$$

entonces $\mathbb{P}(\limsup E_n) = 0 \ (= \infty)$.

Considerando $\Im_n = \sigma\{E_1, E_2, \ldots, E_n\}$

La conclusión se sigue a partir que $\limsup_n E_n = \{\omega : \sum I_{E_n}(\omega) = \infty\}$.

El resultado de Lévy reduce las condiciones de independencia a la convergencia de las probabilidades condicionales además de refinar la convergencia de la serie de probabilidades, al tenerse que

$$\sum_{n\geq 1} \mathbb{P}(E_n) < \infty \Rightarrow \sum_{n\geq 1} \mathbb{P}(E_n | \Im_{n-1}) < \infty \quad c.s.$$

Es importante y fácil de notar que si los eventos son independientes, se reduce a la versión conocida actual del *Lema de Borel-Cantelli*.

Definición A.2.10. *La sucesión* $\{X_n\}$ *es llamada* $* - mixing$ *si existe* $N \geq 1$ *y una sucesión* $f_m \downarrow 0$ *tal que* $\forall m \geq N$ *se tiene que*

$$|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| \leq f_m \mathbb{P}(A)\mathbb{P}(B)$$

para $A \in \sigma \langle X_1, X_2 \dots, X_n \rangle$ *y* $B \in \sigma \langle X_{n+m} \rangle$

Teorema A.2.11. *Sea* $\{A_n\}$ *una sucesión de eventos tal que* $\{I(A_n)\}$ *es* $* - mixing$ *y* $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$*. Entonces*

$$\mathbb{P}(\limsup_n A_n) = 1$$

Teorema A.2.12 (Chandra y Ghosal (1993))**.** *Supóngase* $\{A_n\}$ *sucesión de eventos que satisface*

$$\mathbb{P}(A_i \cap A_j) - \mathbb{P}(A_i)\mathbb{P}(A_j) \leq q(j-i)\mathbb{P}(A_j)$$

donde $q(m) \geq 0$ $\forall m \geq 1$ *y*

$$\sum_{m=1}^{\infty} \frac{q(m)}{\sum_{j=1}^{m} \mathbb{P}(A_j)} < \infty$$

.

Si $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$ *entonces* $\mathbb{P}(\limsup_n A_n) = 1$.

Corolario A.2.13. *Considere que se tiene*

$$\mathbb{P}(A_i \cap A_j) - \mathbb{P}(A_i)\mathbb{P}(A_j) \geq q(j-i)\mathbb{P}(A_j) \quad i < j$$

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$$

y $q(m)$ *es decreciente.*

(a) Si $\sum_{m=1}^{\infty} q(m) < \infty$ *entonces* $\mathbb{P}(\limsup_n A_n) = 1$

(b) Si $\limsup_n n^{\alpha}\mathbb{P}(A_n) > 0$ *y* $\sum_{m=1}^{\infty} q(m)m^{\alpha-1} < \infty$ *para algún* $0 \leq \alpha < 1$ *entonces* $\mathbb{P}(\limsup_n n\mathbb{P}(A_n)) = 1$

(c) *Si* $\limsup_n n\mathbb{P}(A_n) > 0$ *y* $\sum_{m=1}^{\infty} \frac{q(m)}{\ln m} < \infty$ *entonces se tiene* $\mathbb{P}(\limsup A_n)$

Definición A.2.14. *Sea* $\{E_n\}$ *una sucesión de eventos. Definimos* $\phi_1 = 0$ *y*

$$\phi_n = \sup_{F \in \Im_{n-1}} |\mathbb{P}(E_n|\Im_n) - \mathbb{P}(E_n)| \quad para\ n \geq 2$$

donde $\Im_n = \sigma(E_1, E_2, \ldots, E_n)$.

Es evidente que si $\{E_n\}$ son eventos independientes $\phi_1 = 0$ además de que
$0 \leq \phi_n \leq 1$.

Teorema A.2.15 (Teorema de Iosifescu y Theodorescu (1969))**.** *Supongamos que* $\sum_{n=1}^{\infty} \phi_n < \infty$ *y que* $\sum_{n=1}^{\infty} \mathbb{P}(E_n) = \infty$ *entonces*

$$\mathbb{P}(\limsup_n E_n) = 1$$

Este teorema es un refinamiento de la parte de divergencia del *Lema de Borel-Cantelli*, pues no pide que sean los eventos independientes entre sí, sino que vayan siendo "suficientemente independientes" con respecto a sus $n-1$ anteriores.

Al igual como anteriormente se ha señalado, si $\{E_n\}$ son independientes se reduce al **Lema** en estudio.

El siguiente teorema gracias a Serfling (1975) generaliza el resultado del Teorema anterior de una manera (a mi parecer) más bonita y elegante

Teorema A.2.16. de Serfling. *Para la sucesión de eventos* $\{E_n\}$ *arbitraria tal que*

$$\sum_{n=1}^{\infty} \mathbb{E}|p_n - \mathbb{P}(E_n)| < \infty$$

y que $\sum_{n=1}^{\infty} \mathbb{P}(E_n) = \infty$ *entonces*

$$\mathbb{P}(\limsup_n E_n) = 1$$

donde $p_1 = \mathbb{P}(E_1)$ *y* $p_n = \mathbb{P}(E_n|\Im_{n-1})$ *para* $n \geq 2$.

A.3 Steinhaus.

2. La desigualdad del trabajo de Steinhaus.

$$\sum_{p=0}^{n} \binom{n}{p} (2p-n)^{2k} \leq 2^n (2k-1)!! n^k \tag{A.2}$$

Demostración. Por inducción.

Para $n = 1$.

$$1 = \binom{1}{0}(2(0)-1)^{2k} + \binom{1}{1}(2-1)^{2k} \leq 2^1(2k-1)!!1^k = (2k-1)!!$$

Supóngase válido para una $n \geq 1$ fija.

Ahora se debe mostrar

$$\sum_{p=0}^{n+1} \binom{n+1}{p} (2p-(n+1))^{2k} \leq 2^{n+1}(2k-1)!!(n+1)^k \tag{A.3}$$

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

$$2(n+1)^{2k} + \sum_{p=1}^{n} \left(\binom{n}{p} + \binom{n}{p-1} \right) (2p-(n+1))^{2k} =$$

$$2(n+1)^{2k} + \sum_{p=1}^{n} \binom{n}{p} (2p-(n+1))^{2k} + \sum_{p=1}^{n} \binom{n}{p-1} (2p-(n+1))^{2k} =$$

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

$$2(n+1)^{2k}+(n-1)^{2k}+\sum_{p=1}^{n-1}\binom{n}{p}\left((2p-n-1))^{2k}+(2p-n+1))^{2k}\right)=$$

$$\sum_{p=0}^{n}\binom{n}{p}\left((2p-n-1))^{2k}+(2p-n+1))^{2k}\right)=$$

$$\sum_{p=0}^{n}\binom{n}{p}\left(\sum_{i=1}^{2k}\binom{2k}{i}(2p-n)^{2k-i}\left((-1)^i+1\right)\right)=\sum_{p=0}^{n}\binom{n}{p}\left(2\sum_{j=1}^{k}\binom{2k}{2j}(2p-n)^{2(k-j)}\right)=$$

$$2(\sum_{p=0}^{n}\binom{n}{p}\binom{2k}{0}(2p-n^{2k})+\sum_{p=0}^{n}\binom{n}{p}\binom{2k}{2}(2p-n^{2(k-1)})+\sum_{p=0}^{n}\binom{n}{p}\binom{2k}{4}(2p-n^{2(k-2)})+\dots$$

$$\dots+\sum_{p=0}^{n}\binom{n}{p}\binom{2k}{2(k-1)}(2p-n^{2k})+\sum_{p=0}^{n}\binom{n}{p}\binom{2k}{2(k-1)}(2p-n^{2k})+2^n\binom{2k}{2k})\leq$$

$$2\left(\binom{2k}{0}2^n(2k-1)!!n^k+\binom{2k}{2}2^n(2k-3)!!n^{k-1}+\dots+\binom{2k}{0}2^n(0)!!n+\binom{2k}{2k}2^n\right)$$

Ahora simplemente al convencerse que del hecho: $2^jj!\leq(2j)!$ implica que

$$\binom{2k}{2j}(2(k-j)-1)!!\leq\binom{k}{j}(2k-1)!!$$

con lo cual

$$\sum_{p=0}^{n+1}\binom{n+1}{p}(2p-(n+1))^{2k}\leq 2^{n+1}(2k-1)!!\sum_{j=0}^{k}n^{k-j}=2^{n+1}(2k-1)!!(n+1)^k$$

es decir

$$\sum_{p=0}^{n+1}\binom{n+1}{p}(2p-(n+1))^{2k}\leq 2^{n+1}(2k-1)!!(n+1)^k$$

Bibliografía.

[1] Banach, Stefan. *Sur le problème de la mesure* Fundamenta mathematica IV, Varsovie, 7-33 (1923)

[2] Basulto Santos, Jesús; Camúñez Ruiz, José Antonio. *La geometría del azar*. La correspondencia entre Fermat y Blaise Pascal. Nivola libros y ediciones. España (2007).

[3] Barone, Jack; Novikoff, Albert. *A History of the Axiomatic Formulation of Probability from Borel to Kolmogorov: Part I.* Archive for History of Exact Sciences, Vol. 18, no. 2. 123-190 (1978)

[4] Bernoulli, Jakob. *On the Law of Large Numbers*: Translated into English by Oscar Sheynin. Berlin 2005. www.sheynin.de (Abril 9, 2011)

[5] Blanco Castañeda, Liliana. (2004) *Probabilidad* Universidad Nacional de Colombia, Sede Bogotá, Fac. Ciencias

[6] Boas, Ralph Philip. *Entire functions.* New York : Academic (1954).

[7] Borel, Émile. *Le calcul des probabilités et la méthode des majorités.* L'Anée psychologique. Vol 14. (1907).

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

[9] Borel, Émile *Leçons sur la theorie de fonctions.* (Principles de la théorie des ensembles en vue des applicarions à la théorie des functions). Gauthier-Villars, Imprimeur-editoeur. Libraire du bureau des longitudes, de l'École Polytechnique. (1950)

[10] Borel, Émile *Remarques sur certaines questions de probabilité.* Bulletin de la S.M.F., tome 33 (1905), p. 123-128. http://www.numdam.org/ (20 Noviembre 2010)

[11] Borel, Émile *Sur quelques points de la théorie des functions.* Paris, École Norm. Ann. (3) 12 pp. 9-55 (1985)

[12] Cantelli, F. *Sulla probabilità come limite della frequenza.* Classe di Scienze Fisiche, Mathematiche e Naturali. Rendiconti, 26 (5), 3945. Accademia dei Lincei Roma (1917)

[13] Cantelli, F. *Su due applicazioni di un teorema di G. Boole alla statistica matematica.* Classe di Scienze Fisiche, Mathematiche e Naturali. Rendiconti, 26 (5), 295-302. (1917)

[14] Corry, Leo, *David Hilbert and the Axiomatization of Physics (1894-1905)*, Archive for History of Exact Sciences 51, 83-198. (1997)

[15] de Finetti, B. *Scritti (1926-1930)*, Cedam, Pádua .

[16] Fréchet, M. *Les mathematiques et le concret.* Paris : Presses Univ. de France. (1955)

[17] García Álvarez, Miguel Ángel. *Introducción a la teoría de la probabilidad II. Segundo Curso.* Fondo de Cultura Econónica. México (2005)

[18] Hald, Anders *History of Probabilitiy and Statistics and Their Applications before 1750.* John Wiley & Sons, Inc. (2003)

[19] Hausdorf, F. *Grundzüge der Mengenlehre* (1914) Reimpreso por American Mathematical Society en inglés. (1949)

[20] Hawkins, Thomas *Lebesgue's theory of integration: It's origins and development.* Chelsea Publishing Company- The Bronz, New York, N.Y. (1975)

[21] Holland, Anthony S. B. *Introduction to the theory of entire functions.* Academic. New York. (1973)

[22] Jacod,Jean; Protter Philip. *Probability Essentials* Springer-Verlag, Alemania. (2004)

[23] Kolmogorov, A. The theory of probability, in A.D. Aleksandrov, A. Kolmogorov and M. Lavrent'ev (eds.) *Mathematics, Its Contents, Methods and Meaning*, vol. 2. 229-264, MIT Press (1963) (Original ruso de 1956).

[24] Laplace, Pierre Simon. *Ensayo filosófico sobre las probabilidades.* Espasa - Calpe Argentina, S.A. Traducción y notas por Alfredo B. Besio y José Banfi.

[25] Laplace, Pierre Simon. *Théorie analytique des probabilités.* Volumen I Éditions Jacques Gabay(1837) Reimpresión 1995.

[26] Lebesgue, Henri. *Intégrale, Longueur, Aire.* Tesis doctoral. Annali di Matematica.(1902)

[27] Lebesgue, Henri. *Lecons sur l'integration et la recherche des fonctions primitives.* Deuxième édition. Gauthier-Villars et Cie., Editeurs. Parś. (1928)

[28] von Mises, R. *Fundamentalsätze der Wahscheinlichkeitsrechnung.* Mathematische Zeitschrift 4. 1-97 (1919)

[29] von Mises, R. *Grundlagen der Wahscheinlichkeitsrechnung.* Mathematische Zeitschrift 5. 52-99 (1919)

[30] von Mises, R. On foundations of Probability and Statistics. The Annals of Mathematical Statistics. Vol. 12, No. 2, Jun. (1941)

—bibitemStanley Stanley W. Nash. *An Extension of the Borel-Cantelli Lemma.* The Annals of Mathematical Statistics, Vol. 25, No. 1 (Mar., 1954), pp. 165-167

[31] Todhunter, Isaac. *A history of mathematical theory of probability.* Macmillan and Co. (1865)

[32] Pascal, Blaise. *Les lettres de Blaise Pascal : accompagnes de lettres de ses correspondants.* (1922).

[33] Pascal, Blaise Obras Matemáticas (Selección de textos). Mathema. Facultad de Ciencias, Universidad Nacional Autónoma de México. (1995).

[34] Ortiz, José R *El concepto de infinito.* Boletín de la Asociación Matemática Venezolana. No. 2 pp. 59-81 (1994).

[35] von Plato, Jan *Creating Modern Probability.* Cambridge Studies in Probability, Induction and Decision Theory, Cambridge University Press. (1994). Reimpresión 2006.

[36] Poincaré, Henri. *Calcul des probabilités.* Éditions Jacques Gabay(1912). 2 Ed. Reimpresión 1987.

[37] Poisson, Siméon-Denis *Recherches sur la Probabilité des Jugements en Matiére Criminelle et en matiére civile. Précédées des règles générales du calcul des probabilités.* Éditions Jacques Gabay(1837) Reimpresión 2003.

[38] Ramirez Ramirez, Lilia Leticia, *Generalizaciones de la ley fuerte de los grandes números y Lema de Borel-Cantelli* Tesis Licenciatura (Actuaria) UNAM, Facultad de Ciencias. Mxico. (1998)

[39] Rubel, Lee A. *Entire and meromorphic functions.* Springer, New York. (1995)

[40] Santibañez Romellón, Jorge, *Sobre la historia de axiomatizacion de la probabilidad: Corriente Borel-Kolmogorov* Tesis Licenciatura (Matemático) UNAM, Facultad de Ciencias. Mxico. (1977)

[41] Schneider, I *Die Entwicklung der Wahrscheinlichkeitstheorie von den Anfangen bis 1933.* Einfuhrungen und texte, Wissenchaftliche Buchgesellschaft, Darmstadt. (1988.)

[42] Shafer, Gleenn; Vovk, Vladimir *The origins and legacy of Kolmogorov's Grundbegriffe.* The Game-Theoretic Probability and Finance Project (2006) http://www.probabilityandfinance.com (Enero 28, 2011)

[43] Sheynin, Oscar B. *Theory of Probability. A Historical Essa.* Second revised and enlarged edition. Berlin 2009. www.sheynin.de (Abril 9, 2011)

[44] Sheynin, Oscar B. *Theory of Probability and Statistics As Exemplified in Short Dictums.* Second revised and enlarged edition. Berlin 2009. www.sheynin.de (Abril 9, 2011)

[45] Sierpinski *Sur les définitions axiomatiques des ensembles mesurables (B)* bull. Intern. Acad. Sci. Cracovie A 1918, p. 29-34

[46] Sierpinski *Sur les définitions axiomatiques des ensembles mesurables (L)* bull. Intern. Acad. Sci. Cracovie A 1918, p. 173-178

[47] Steinhaus, Hugo *Les probabilités dénombrables et leur rapport à la théorie de la mesure.* Fundamenta Mathematica IV. (1923)

[48] Wiman, A. 1900. *Ueber eine Wahrscheinlichkeitsaufgabe bei Kettenbruchenwicklungen*, fversigt af Kongl. Svenska Vetenskaps-Akademiens Frhandlingar, 7, 829-841

[49] Wiman, A. 1901 *Bemerkungen ber eine von Gyldn aufgeworfene Wahrscheinlichkeitsfrage*, Acta Universitatis Lundensis, 1-19.