Revista Colombiana de Estadística
Nº 15, 16 - 1987

FUNDAMENTOS DE LA TEORIA DE PROBABILIDADES

Andrei Nikolaivich Kolmogórov

Traducido por: *Ramón Matos Marengo*
Profesor Asociado
Universidad del Atlántico

INDICE

pág.

Prólogo a la Primera Edición 7

Prólogo a la Segunda Edición 9

I TEORIA ELEMENTAL DE LAS PROBABILIDADES 11

§1 Axiomas 13

§2 Relaciones con los datos de una prueba 15

§3 Terminología empleada 18

§4 Probabilidad condicional: axiomas, teoremas, corolarios. Teorema de Bayes. 20

§5 Independencia 22

§6 Probabilidades condicionales como variables aleatorias; cadenas de Markov. 28

II CAMPOS PROBABILISTICOS INFINITOS

§1 Axiomas de continuidad 31

§2 Campos probabilísticos de Borel................. 34

§3 Ejemplos de campos probabilísticos infinitos...... 37

III VARIABLES ALEATORIAS 43

§1 Funciones probabilísticas 43

§2 Definición de variables aleatorias, funciones de distribución. 45

§3 Funciones de distribución múltiple 49

§4 Probabilidad en espacios infinitos 53

§5 Variables aleatorias equivalentes; diferentes tipos de convergencia. 61

IV ESPERANZA MATEMATICA 66

§1 Integral abstracta de Lebesgue 66

§2 Esperanzas matemáticas absolutas y condicionales... 69

§3 Desigualdad de Chebishev......................... 73

§4 Algunos criterios de convergencia................ 75

§5 Diferenciación e integración de las esperanzas matemáticas según parámetro........................ 76

V PROBABILIDAD CONDICIONAL Y ESPERANZA MATEMATICA 81

§1 Probabilidad condicional.......................... 81

§2 Explicación de una paradoja de Borel.............. 86

§3 Probabilidad condicional respecto a una variable aleatoria.. 87

§4 Esperanza matemática condicional.................. 89

VI INDEPENDENCIA. LEY DE LOS GRANDES NUMEROS 95

§1 Independencia..................................... 95

§2 Variables aleatorias independientes............... 97

§3 Ley de los grandes números........................ 101

§4 Observación al concepto de esperanza matemática... 114

§5 Ley fuerte de los grandes números, convergencia de series.. 119

Anexo: Un teorema fundamental en la teoría de probabilidad.......................................

Bibliografía 131

PROLOGO A LA PRIMERA EDICION

El objetivo del trabajo expuesto es fundamentar axiomáticamente la teoría de las probabilidades. La idea central del autor fue la inclusión natural de los fundamentos de la teoría de las probabilidades, consideradas aún no hace mucho, completamente peculiares, en la serie de conceptos generales de la matemática moderna. Antes del surgimiento de la teoría de la medida y la integral de Lebesgue, este problema era casi desalentador. Después de las investigaciones de Lebesgue quedó clara la analogía entre la medida de conjuntos y la probabilidad de un suceso, como también entre la integral de funciones y la esperanza matemática de la variable aleatoria. Esta analogía permite una continuación: así, por ejemplo, muchas propiedades de las variables aleatorias independientes son completamente análogas a las correspondientes propiedades de las funciones ortogonales. Para que a partir de esta analogía se fundamentara la teoría de las probabilidades fue necesario además liberar la teoría de la integración de elementos geométricos, los cuales aún permanecen en Lebesgue. Esta liberación fue realizada por Frechet.

Los intentos de construir los fundamentos de la teoría de las probabilidades desde este punto de vista general ya existen y todo el círculo de ideas aquí expuestas han logrado popularidad en un reducido grupo de especialistas. Sin embargo no existía una presentación completa de todo el sistema libre de complicaciones innecesarias. (Seprepara en edición el libro de Frechet, ver [2]).

Quisiera, además, referirme a algunos tópicos de la exposociόn que constituyen una novedad aún para los especialistas. Estos tópicos son los siguientes: distribución de probabilidades en espacios infinitos (Cap. III, §4), diferenciación e integración de esperanzas matemáticas según parámetro (Cap. IV §5), y en particular la teoría de la probabilidad condicional y la esperanza matemática condicional (Cap. V). Conviene anotar que todos estos nuevos conceptos y problemas aparecen necesariamente en el examen de problemas físicos concretos(1).

El Cap. VI contiene un resumen de resultados obtenidos por Khintchine y el autor que se refieren a las condiciones de aplicación de las leyes débil y fuerte de los grandes números. En la lista de la bibliografía se exponen algunos nuevos trabajos que representan interés desde el punto de vista de los interrogantes de fundamentación de la teoría de las probabilidades.

Expreso sinceros agradecimientos a A.I. Khintchine, quien leyó atentamente todo el manuscrito e hizo algunas sugerencias que permitieron su perfeccionamiento.

A.M. KOLMOGOROV

Kliazma, alrededores de Moscú.
1 de mayo - 1933.

---

(1) Comparar, por ejemplo el libro A. Kolmogórov und M. Leontowitsch, Zur Berechunge der mittlerer Brownscher Frache, Physik. Zeitschr d Conjeturior, v.4 (1933) con el libro M. Leortowitsch, zur Statistik der Kortinuierlichen Systeme und des zeitlichen Verlaufes der Physikalischen Vergarge, Physik zeitchr. d. Sowjeturier, v.3, 1933, p.35-63.

PROLOGO A LA SEGUNDA EDICION

Desde la primera edición alemana de esta publicación han transcurrido cuarenta años. Sin embargo conjuntamente con Shiriaer optamos por no introducir en ella ningún cambio sustancial; sólo se hicieron pequeñas complementaciones en su presentación ; sin embargo, decidimos no introducir en ella. Modernizando algunas notaciones. Para algunos teoremas §3-5 del Cap. VI se dan demostraciones, redactadas por A.N. Shiriaev según mis trabajos de 1925-1930. En los manuales modernos estos teoremas se demuestran generalmente con ayuda del aparato de las funciones características. Mis demostraciones iniciales con medios elementales, directos, quizá conserven algún interés. Los conceptos anotados en §2 del Cap. I sobre los caminos de fundamentación de la aplicación de la teoría axiomática de probabilidades a los problemas reales, fueron desarrollados por mí en detalle en [1]. Pero aún quedaban sin explicar las causas por las cuales frecuentemente nos encontramos en la práctica con la estabilidad de las frecuencias. Una nueva interpretación del problema han sido expuestas en [2] y [3] (ver también [4]).

[1] Monografía. "*Matemática: su contenido, métodos y significado*". Ed. Académica de Ciencias URSS 1956, Cap. XI.

[2] A.N. Kolmogórov. *Tres formas para la definición del concepto cantidad de información. Problemas de transferencia de información.* Tomo I, Ed. 1 (1965).

[3] Kolmogórov. *Sobre los fundamentos lógicos de la teoría de la información y la teoría de las probabilidades. Pro-*

blemas de Transferencia de información. Tomo V, Ed. 3 (1969).

[4] Zvonkin A.K., Levin L.A., *Complejidad de los objetos finitos y fundamentos de la Teoría de la información y aleatoriedad con ayuda de la teoría de algoritmos*. USPEJI MAT. Naik, Tomo 25 Ed. 6 (1970).

Es importante señalar algunas cuestiones para cuya comprensión es preciso recomendar al lector la comparación de la exposición dada en esta publicación con otra más moderna.

1. En el §1 del Capítulo V está dada la definición de probabilidad condicional $P(A|\xi)$, donde ξ es un elemento aleatorio de algún conjunto X, o sea una relación de Ω en X. Con esta relación se puede asociar un álgebra $\mathfrak{F}^{\xi} \subseteq \mathfrak{F}$ de todas las imágenes completas de los subconjuntos del conjunto X que pertenecen a $\mathfrak{F}$. Ahora se prefiere primero definir la probabilidad condicional con respecto a cualquier σ-subálgebra $\mathfrak{F}' \subseteq \mathfrak{F}$ y después considerar que

$$P(A|\xi) = P(A|\mathfrak{F}^{\xi}).$$

2. Los resultados del §4 Capítulo III, aunque son ampliamente utilizados posteriormente, no dan distribuciones inmediatas aceptables en los espacios funcionales que representan real interés dentro de la presente exposición (página 55).

A.M. KOLMOGOROV

17 de diciembre - 1973.

CAPITULO I

TEORIA ELEMENTAL DE LAS PROBABILIDADES

Llamamos teoría elemental de las probabilidades a las parte en la cual se estudian las probabilidades solo de una cantidad finita de sucesos. Los teoremas que aquí se deducen, naturalmente, se utilizan también a las preguntas relacionadas con una cantidad infinita de sucesos aleatorios, sin embargo en el estudio de estos últimos se aplican también esencialmente nuevos principios. Por eso el único axioma de la teoría matemática de las Pribabilidades que tiene relación con una cantidad infinita de sucesos aleatorios, se introduce solo al comienzo del segundo capítulo (Axioma V).

La teoría de las probabilidades como disciplina puede y tienen que ser axiomatizada completamente de igual manera como lo estan el álgebra o la geometría. Esto quiere decir, que después de dar los nombres de los objetos de estudio y sus principales relaciones, como también los axiomas por los cuales estas relaciones deben regirse, el sucesivo material debe fundamentarse exclusivamente solo en estos axiomas, no basándose en el valor concreto común de estos objetos y sus relaciones.

Con base en lo anterior, en el §1 definiremos el concepto de campo probabilístico como un sistema de conjuntos que cumple unas condiciones determinadas. Lo que representan los elementos de estos conjuntos es totalmente indiferente para el desarrollo matemático de la teoría de las probabilidades (ver la introducción de los conceptos fundamentales de la geometría en [1] *Fundamentos de Geometría* de Hilbert, o las definiciones de grupo, anillo y cuerpo en el álgebra abstracta).

Toda teoría axiomática (abstracta) permite, como es sabido, un número infinito de interpretaciones concretas. Así, la teoría matemática de las probabilidades permite paralelamente a las interpretaciones que la originaron, muchas otras interpretaciones.

Por esto, en la aplicación de teoría matemática de las probabilidades llegamos a campos de la ciencia que no tienen ninguna relación con los conceptos de suceso y probabilidad en el sentido propio de estas palabras.

La axiomatización de la teoría de las probabilidades puede ser realizada de diferentes maneras dependiendo de la escogencia de los axiomas y de la escogencia de los conceptos fundamentalmente y de las principales relaciones. Si se persigue como objeto tanto la sencillez en el sistema de axiomas, como en las construcciones de su consecutiva teoría entonces representaremos la más racional de las axiomatizaciones de los conceptos de sucesos aleatorios y sus probabilidades.

Existen otros sistemas de axiomatización para la construcción de la teoría de las probabilidades, en los cuales el concepto de probabilidad no es tomado como fundamental sino que es definido a través de otros(1). En estas axiomatizaciones se persigue sin embargo otro fin: en lo posible acercar la teoría matemática con el aparecimiento empírico del concepto de probabilidad.

(1) Comparar con, por ejemplo, R. von Mises [1] y [2] y S.N. Bernstein [1].

§ 1. AXIOMAS [1]

Sea Ω un conjunto de elementos W, los cuales llamaremos *sucesos elementales*, $\mathcal{F}$ el conjunto de subconjuntos de Ω. Los elementos del conjunto $\mathcal{F}$ lo llamaremos *sucesos aleatorios* (o sencillamente sucesos), a Ω lo llamaremos *espacio de sucesos elementales*.

I. $\mathcal{F}$ constituye un álgebra de conjuntos[2].

NOTA: El sistema $\mathcal{F}$ de subconjunto Ω se llama álgebra, si $\Omega \in \mathcal{F}$, la unión, la intercepción y la diferencia de dos conjuntos del sistema pertenece nuevamente a este sistema. Denotaremos la intercepción de los conjuntos A y B por $A \cap B$, su unión $A \cup B$ su diferencia A B. El complemento del conjunto A lo denotamos así $\bar{A}$. $\emptyset$ nos representa el conjunto vacío. Si los conjuntos A y B no se interceptan $(A \cap B = \emptyset)$, entonces su unión $A \cup B$ la denotaremos por $A + B$ y la llamaremos suma.

---

(1) El lector que desee dar un sentido concreto a los siguientes axiomas deberá comenzar a leer §2.

(2) El sistema $\mathcal{F}$ de subconjuntos del conjunto Ω se llama álgebra, si $\Omega \in \mathcal{F}$, la unión, la intercepción y diferencia de dos conjuntos de sistema pertenecen nuevamente a este sistema. Denotaremos la intercepción de los conjuntos A y B por $A \cap B$ ó AB, su unión como $A \cup B$, su diferencia por $A \backslash B$. El conjunto complementario $\Omega \backslash A$ al conjunto A lo denotamos $\bar{A}$. Por $\emptyset$ se denota el conjunto vacío. Si los conjuntos A y B no se interceptan $(AB = \emptyset)$ entonces su unión se denotará también por $A + B$ y se llama suma. El conjunto $\mathcal{F}$ lo denotaremos en lo sucesivo con letras latinas mayúsculas. Comparar con A.M. Kolmogórov y S.V. Fomin. Elementos de teoría de funciones y análisis funcional. Editorial Nauka, M., 1968.

Los conjuntos de $\mathcal{F}$ en lo sucesivo los representaremos, con letras latinas mayúsculas.

II. Todo conjunto A de $\mathcal{F}$ se pone en correspondencia con un real no negativa $P(A)$.

Este número lo llamamos probabilidad de A.

III. Si A y B no se interceptan, entonces $P(A+B) = P(A)+P(B)$.

La terna $(\Omega,\mathcal{F},P)$ que cumpla los axiomas I-IV la llamaremos campo probabilístico.

Nuestro sistema de axiomas I-IV no es contradictorio. Esto lo muestra el siguiente ejemplo: Ω está formado por un único elemento ω, a $\mathcal{F}$ lo forman Ω y $\emptyset$, en este caso

$$P(\Omega) = 1, \quad P(\emptyset) = 0.$$

Nuestro sistema de axiomas, sin embargo, no es completo: en diferentes preguntas de la teoría de las probabilidades se examinan diferentes campos probabilísticos.

Un campo probabilístico sencillo se construye de la siguiente manera:

Tomamos cualquier conjunto finito $\Omega = \{\omega_1,\ldots,\omega_k\}$ y un conjunto cualquiera $\{p_1,\ldots,p_k\}$ de números no negativos cuya suma $p_1+p_2+\ldots+p_k = 1$. Como $\mathcal{F}$ se toma el conjunto de todos los subjuntos A de Ω y tal que si $A = \{\omega_{i_1},\ldots,\omega_{i_\lambda}\}$ se supone $P(A) = p_{i_1}+\ldots+p_{i_\lambda}$. En este caso se diría que $p_1,\ldots,p_k$ son las probabilidades de los sucesos elementales $\omega_1,\ldots,\omega_k$ o sencillamente las probabilidades elementales. Así se obtienen todos los campos probabilísticos finitos en los cuales $\mathcal{F}$ constituídos por el conjunto de todos los subconjuntos de Ω (aquí el campo probabilístico se llama finito, si el conjunto Ω es finito). Ver ejemplos capítulo II §3.

*

§2. RELACION CON LOS DATOS DE UNA PRUEBA (1)

La utilización de la teoría de las probabilidades en el mundo real de las pruebas se realizan siguiendo el siguiente esquema.

1. Suponiendo que está dado un conjunto G de condiciones que permite un número ilimitado de repeticiones.
2. Se estudia un grupo determinado de sucesos, los cuales pueden ocurrir como resultados de la presencia de las condiciones G.

En los casos por separados estos sucesos pueden o no ocurrir en diferentes combinaciones. En el conjunto Ω se incluyen todas las posibles variantes de aparición o no de los sucesos examinados.

---

(1) El lector interesado solo en el desarrollo matemático de la teoría puede no leer este parágrafo. La siguiente presentación se basa en los axiomas §1 y no se usan las conclusiones de este parágrafo. En él nos limitaremos a mostrar en forma sencilla el origen empírico de los axiomas de la teoría de probabilidades y concientemente dejaremos a un lado la búsqueda filosófica profunda sobre el concepto probabilístico en el mundo experimental. Los supuestos necesarios para la aplicación de la teoría de probabilidades al mundo de sucesos reales en gran medida el autor lo saca de los resultados de Mises, en particular R. von Mises [1], p.21-27, parágrafo "Das Verhältnis der Theorie zur Erfahrungswet".

3. Si después de la realización de las condiciones G la variante aparecida pertenece a un conjunto A, entonces decimos que ocurrió el suceso A.

EJEMPLO. El conjunto G que consiste en el lanzamiento dos veces de una moneda. El grupo de sucesos, sobre el cual hablamos en el punto 2, consiste en que en cada lanzamiento pueden aparecer cara o sello. De aquí, se deduce que en total son posibles cuatro diferentes variantes (sucesos elementales): sello-sello, sello-cara, cara-sello, cara-cara. En calidad de suceso A se examina "La repetición". Este suceso consiste en la suma del primer y cuatro suceso elemental. Así cada suceso se puede examinar como un conjunto de sucesos elementales.

4. En condiciones conocidas, que aquí no profundizaremos, se supone que en un suceso A, el cual puede o no ocurrir como resultado de la realización de las condiciones G, se pone en correspondencia con un número real $P(A)$ que cumple con las siguientes condiciones:

A) Prácticamente se puede estar convencido que si el conjunto de condiciones G se repite un número n grande de veces y si denotamos por m el número de veces que ha ocurrido A, entonces la relación m/n se diferenciará poco de $P(A)$.

B) Si $P(A)$ es muy pequeño, entonces podemos estar seguros que en la realización de las condiciones G una sola vez el suceso A no ocurrirá.

DEDUCCION EMPIRICA DE AXIOMAS.

De ordinario, se puede suponer que el sistema F de los sucesos examinados $A, B, C, \ldots$ a los cuales se les asignan unas probabilidades determinadas, forma un álgebra de conjuntos, que contiene en calidad de elemento al conjunto Ω (Axioma I, primera parte del Axioma II - existencia de la probabilidad).

Es claro que $0 \leqslant m/n \leqslant 1$ así que la segunda parte del Axioma II resulta natural. Para el suceso Ω siempre $m = n$ por lo cual es natural colocar $P(\Omega) = 1$ (Axioma III).

Si A y B son excluyentes entre sí (o sea los conjuntos A y B no se interceptan) entonces $m = m_1 + m_2$ donde m, m_1, m_2 denotan respectivamente los números de veces que aparecen los sucesos $A+B$, A, B. De aquí se deduce

$$m/n = m_1/n + m_2/n.$$

En consecuencia, es lógico escoger

$$P(A+B) = P(A) + P(B) \quad \text{(Axioma IV)}.$$

OBSERVACION I. De la veracidad de dos afirmaciones se deduce la veracidad de la afirmación sobre su simultánea veracidad, a pesar de que en este caso el grado de veracidad es muy bajo. Sin embargo, si el número de afirmaciones es muy grande, entonces de la veracidad de cada una de estas afirmaciones por separado, es imposible deducir algo en relación a la simultaneidad de la veracidad de todas las afirmaciones. Por eso del principio A en ninguna forma se deduce que si el número de series n es grande en cada serie la relación m/n se diferenciará poco de $P(A)$.

OBSERVACION II. Al suceso imposible (al conjunto vacío $\emptyset$) corresponde en base a nuestros axiomas la probabilidad $P(\emptyset) = 0$[1] al mismo tiempo, lo contrario de $P(A) = 0$ no se deduce la imposibilidad del suceso A; de acuerdo con el principio B de la nulidad de la probabilidad se deduce que la realización de una sola vez las condiciones G el suceso A es prácticamente imposible.

Esto no quiere decir que en una larga serie de pruebas el suceso A tampoco ocurrirá.

De acuerdo con el principio A se puede solo afirmar que si $P(A) = 0$ y n es un número grande la relación m/n será pequeña.

*

§3. OBSERVACIONES TERMINOLOGICAS.

Hemos definido los objetos de nuestro estudio, sucesos aleatorios como conjuntos. Muchos de los conceptos de la teoría de conjuntos; sin embargo, se denominan en la teoría de las probabilidades de forma diferente.

Aquí presentamos una pequeña guía de estos conceptos:

En la teoría de conjuntos	*Para los sucesos aleatorios*
1. A y B no se interceptan o sea $A \cdot B = \emptyset$.	1. Los sucesos A y B no son compatibles.
2. $A \cdot B \cdots N = \emptyset$.	2. Los sucesos $A, B, \cdots, N$ no son compatibles.
3. $A \cdot B \ldots N = X$.	3. El suceso X está constituído por la realización simultánea de los sucesos $A, B, \cdots, N$.

(1) Comp. §4, formula (3).

4. $A \cup B \cup \ldots \cup N = X$.	4. El suceso X está constituído por la realización de por lo menos uno de los sucesos $A, B, \ldots, N$.
5. El complemento de un conjunto $\bar{A}$.	5. El suceso inverso $\bar{A}$ constituído por la no aparición del suceso A.
6. $A = \emptyset$.	6. A es imposible.
7. $A = \Omega$	7. A necesariamente ocurre.
8. El sistema U de conjuntos $A_1, \ldots, A_n$ forma distribución del conjunto Ω si $A_1 + A_2 + \ldots + A_n = \Omega$ (aquí se supone que cualquier par de conjuntos A_i no se interceptan).	8. La prueba U afirma que alguno de los sucesos $A_1, A_2, \ldots, A_n$ ocurre; los sucesos $A_1, A_2, \ldots, A_n$ se llaman resultados posibles de la prueba U.
9. B es subconjunto de A: $B \subseteq A$.	9. De la ocurrencia del suceso B necesariamente se desprende la ocurrencia de A.

*

§4. CONSECUENCIA INMEDIATA DE LOS AXIOMAS, PROBABILIDAD CONDICIONAL, TEOREMA DE BAYES.

De $A + \bar{A} = \Omega$ y los axiomas III y IV se deduce que

$$P(A) + P(\bar{A}) = 1 \tag{1}$$

$$P(\bar{A}) = 1 - P(A) \tag{2}$$

Como $\bar{\Omega} = \emptyset$, entonces en particular

$$P(\emptyset) = 0 \tag{3}$$

Si $A, B, ---, N$ no son compatibles, entonces del axioma IV se deduce la fórmula

$$P(A+B+\ldots+N) = P(A)+P(B)+\ldots+P(N) \tag{4}$$

(teorema de la adición).

Si $P(A) > 0$, entonces la relación

$$P(B/A) = \frac{P(AB)}{P(A)} \tag{5}$$

se llama *probabilidad condicional* del suceso B en las condiciones de A.

De (5) se deduce que

$$P(AB) = P(B/A)\cdot P(A) \tag{6}$$

La fórmula general se concluye por inducción

$$P(A_1A_2\ldots A_n) = P(A_1)P(A_2/A_1)P(A_3/A_1A_2)\ldots P(A_n/A_1\ldots A_{n-1}) \tag{7}$$

(teorema del producto).

Fácilmente se demuestran las siguientes fórmulas:

$$P(B/A) \geqslant 0 \tag{8}$$

$$P(\Omega/A) = 1 \tag{9}$$

$$P(B+C/A) = P(B/A) + P(C/A) \tag{10}$$

Comparando las fórmulas (8) y (10) con los axiomas II-IV, obtenemos que el sistema de conjunto F junto con la función $P(B/A)$ (con el conjunto A fijo) forma un campo probabilístico. En consecuencia todos los teoremas demostrados para las probabilidades $P(B)$ son ciertos para las probabilidades condicionales $P(B/A)$ (con el suceso A fijo).

Se demuestra fácilmente que

$$P(A/A) = 1 \tag{11}$$

De (6) y de la fórmula análoga $P(AB) = P(A/B)P(B)$ obtenemos la fórmula

$$P(A/B) = \frac{P(A)P(B/A)}{P(B)} \tag{12}$$

este es el contenido del teorema de Bayes.

TEOREMA (Probabilidad completa). *Supongamos que* $A_1+ \ldots +A_n = \Omega$ *y* B *es un suceso cualquiera, entonces*

$$P(B) = P(A_1)P(B/A) + P(A_2)P(B/A_2) + \ldots + P(A_n)P(B/A_n) \tag{13}$$

Demostración. De la igualdad $B = A_1B + \ldots + A_nB$ entonces, de acuerdo con (4) $P(B) = P(A_1B) + \ldots + P(A_nB)$ de acuerdo con (6) tiene lugar la igualdad

$$P(A_iB) = P(A_i)P(B/A_i) \ .$$

TEOREMA (Bayes). *Supongamos que* $A_1+A_1+\ldots+A_n = \Omega$ *y* B *es un suceso cualquiera, entonces*

$$P(A_i/B) = \frac{P(A_i)P(B/A_i)}{P(A_1)P(B/A_1) + \ldots + P(A_n)P(B/A_n)} \tag{14}$$

Demostración. De acuerdo con la fórmula (12)

$$P(A_i/B) = \frac{P(A_i)P(B/A_i)}{P(B)}$$

Para la obtención de la fórmula (14) basta reemplazar la probabilidad $P(B)$ por la expresión (13) según el teorema sobre probabilidad completa.

(Los sucesos $A_1, A_2, \ldots, A_n$ frecuentemente son llamados hipótesis y se dice que la fórmula (14) da la probabilidad $P(A_i/B)$

de la hipótesis A_i después de la ocurrencia del suceso B; $P(A_i)$ significa con este caso la probabilidad apriori de A_i).

*

§5. INDEPENDENCIA.

El concepto de independencia de dos o varias pruebas ocupa un lugar concentral en la teoría de las probabilidades. En efecto, ya vimos, que la teoría de las probabilidades desde un punto de vista matemático se puede examinar como una aplicación especial de la teoría general de las funciones aditivas de los conjuntos. Es natural preguntarse, de qué forma la teoría de las probabilidades surgió como una gran ciencia autónoma, con sus propios métodos.

Para responder a este interrogante hay que señalar la especialización que reciben en la teoría de las probabilidades, los problemas generales, relacionados con las funciones aditivas de los conjuntos.

La situación de que nuestra función aditiva de conjuntos $P = P(\cdot)$ es no negativa y cumple la condición $P(\Omega) = 1$ no condiciona aún el aparecimiento de nuevos problemas. Las magnitudes aleatorias (ver Cap. 3) desde un punto de vista matemático representan funciones medibles (en relación a F) y sus esperanzas matemáticas las constituyen las integrales abstractas de Lebesgue. Esta analogía por primera vez y completamente fué explicada en los trabajos de Fréchet[1].

(1) Comp. Fréchet [1] y [2].

La introducción de los conceptos recordados no pueden, en consecuencia, servir aún de base para el desarrollo de una teoría original y grande.

Históricamente la independencia de las pruebas y de las magnitudes aleatorias representan los conceptos matemáticos que le dieron a la teoría de las probabilidades su rasgo propio. Los trabajos clásicos de Laplace, Poisson, Chebishev, Markov, Liapunov, Mises y Bernshtein fundamentalmente están dedicados al estudio de series de magnitudes aleatorias independientes.

Si en las nuevas investigaciones (de Markov, Bernshtein y otros) frecuentemente se prescinde de la suposición de independencia total, entonces resulta, indispensable para la obtención de los resultados significativos introducir suposiciones menos fuertes. (ver en este Cap. §6 sobre las Cadenas de Markov). Llegamos, en consecuencia, a ver en los conceptos de independencia por lo menos los primeros brotes de peculiaridad de la problemática de la teoría de las probabilídades.

Esta situación será poco profundizada en este libro por que nos dedicaremos principalmente solo a las preparaciones lógicas de las investigaciones teórico probabilisticas.

A este respecto uno de los principales problemas de la filosofía de las ciencias naturales, después de la explicación de la famosa pregunta sobre la esencia del concepto de probabilidad, lo representa la explicación y precisión de las premisas, en las cuales fenómenos reales dados se pueden exminar como independientes.

Esta pregunta se sale, sin embargo, de los límites de este libro.

Pasamos a la definición de independencia. Supongamos que están dadas n pruebas $U^{(1)}, U^{(2)}, \ldots, U^{(n)}$ o sea n desarrollos.

$$\Omega = A_1^{(i)} + A_2^{(i)} + \ldots + A_{r_i}^{(i)} \qquad (i = 1, 2, \ldots, n)$$

de nuestro conjunto base Ω en suma (incompatibles) de sucesos. Entonces se puede dar $r = r_1 r_2 \ldots r_n$ de probabilidades.

$$p_{k_1 k_2 \ldots k_n} = P(A_k^{(1)} A_k^{(2)} \ldots A_{k_n}^{(n)}) \geqslant 0$$

cualquiera con la condición única[1].

$$\sum_{(k_1, k_2, \ldots, k_n)} p_{k_1 k_2 \ldots k_n} = 1 \qquad (1)$$

DEFINICION 1. *Las pruebas* $U^{(1)}, U^{(2)}, \ldots, U^{(n)}$ *se llamarán independientes si para cualquier* $k_1, k_2, \ldots, k_n$ *tiene lugar la igualdad*

$$P(A_{k_1}^{(1)} A_{k_2}^{(2)} \ldots A_{k_n}^{(n)}) = P(A_k^{(1)}) P(A_k^{(2)}) \ldots P(A_{k_n}^{(n)}) \ . \qquad (2)$$

En las r ecuaciones de (2) se tiene solo $r - (r_1 + \ldots + r_n) + (n-1)$ ecuaciones independientes[2].

---

(1) El campo probabilístico con probabilidades arbitrarias que cumpla solo las condiciones mencionadas se puede construir de la siguiente manera: el conjunto Ω lo compone r elementos $\omega_{k_1 k_2 \ldots k_n}$ y las correspondientes probabilidades serán $p_{k_1 k_2 \ldots k_n}$ y $A_k^{(i)}$ se define como el conjunto de todos los $\omega_{k_1 k_2 \ldots k_n}$ para los cuales $k_i = k$.

(2) En efecto en el caso de independencia se puede escoger arbitrariamente solo $r_1 + r_2 + \ldots + r_n$ probabilidades $p_k^{(i)} = P(A_k^{(i)})$, de tal manera que cumpla las n condiciones $\sum_k p_k^{(i)} = 1$. En consecuencia en el caso general tenemos $r-1$ grados de libertad y en el caso de independencia solo $r_1 + \ldots + r_n - n$.

TEOREMA 1. *Si las pruebas $\mathfrak{U}^{(1)},\mathfrak{U}^{(2)},\ldots,\mathfrak{U}^{(n)}$ son independientes, entonces por todo m $(m < n)$ las pruebas $\mathfrak{U}^{(i_1)},\mathfrak{U}^{(i_2)},\ldots,\mathfrak{U}^{(i_m)}$ también serán independientes*(1).

En caso de independencia tiene lugar la igualdad

$$P(A_{k_1}^{(i_1)}A_{k_2}^{(i_2)}\ldots A_{k_m}^{(i_m)}) = P(A_{k_1}^{(i_1)})P(A_{k_2}^{(i_2)})\ldots P(A_{k_m}^{(i_m)}) \qquad (3)$$

(Se supone que los i_ℓ son diferentes).

DEFINICION 2. *Los sucesos $A_1,A_2,\ldots,A_n$ se llaman independientes si los desarrollos (de las pruebas)*

$$\Omega = A_k + \bar{A}_k \qquad (k = 1,2,\ldots,n)$$

son independientes.

En este caso $r_1 = r_2 = \ldots = r_n = 2$, $r = 2^n$ en consecuencia de los 2^n ecuaciones se tienen solo 2^n-n-1 ecuaciones independientes.

Para la independencia de los sucesos $A_1,A_2,\ldots,A_n$ es necesario y suficiente que se cumpla la siguiente condición(2):

---

(1) Para la demostración es suficiente establecer que de la independencia de los n desarrollos se deduce la independencia de los n-1 primeros de ellos. Además que la ecuación (2) se cumple. Entonces

$$P(A_{k_1}^{(1)}A_{k_2}^{(2)}\ldots A_{k_{n-}}^{(n-1)}) = \sum_{k_n} P(A_{k_1}^{(1)}A_{k_2}^{(2)}\ldots A_{k_n}^{(n)})$$

$$= P(A_{k_1}^{(1)})P(A_{k_2}^{(2)})\ldots P(A_{k-1}^{(n-1)}) \sum_{k_n} P(A_{k_n}^{(n)})$$

$$= P(A_{k_1}^{(1)})P(A_{k_2}^{(2)})\ldots P(A_{k_{n-1}}^{(n-1)})$$

---

(2) Comparar S.N. Bernshtein [1], p.47-57. El lector puede comprobar esto sin dificultad (por inducción).

$$P(A_{i_1}A_{i_2}\ldots A_{i_m}) = P(A_{i_1})P(A_{i_2})\ldots P(A_{i_m}), \qquad (4)$$

$$m = 1,2,\ldots,n; \qquad 1 \leqslant i_1 < i_2 < \ldots < i_m \leqslant n.$$

Todas estas ecuaciones son independientes entre sí. En particular si $n = 2$ obtenemos de (4) solo uno ($2^2-2-1 = 1$) condición para la independencia de dos sucesos A_1 y A_2

$$P(A_1A_2) = P(A_1)P(A_2) \qquad (5)$$

El sistema (2) está constituído en este caso además de (5) por tres ecuaciones:

$$P(A_1\bar{A}_2) = P(A_1)P(\bar{A}_2) \ ,$$
$$P(\bar{A}_1A_2) = P(\bar{A}_1)P(A_2) \ ,$$
$$P(\bar{A}_1\bar{A}_2) = P(\bar{A}_1)P(A_2)$$

las cuales se deducen de (5)[1].

Es bueno anotar que la independencia "par" de los sucesos $A_1, A_2, \ldots, A_n$ o sea de las relaciones

$$P(A_iA_j) = P(A_i)P(A_j) \qquad (i \neq j).$$

En el caso $n > 2$ no se desprende la independencia de estos sucesos[2] (para ello es necesario el cumplimiento de todas las igualdades (4)).

---

(1) $P(A_1\bar{A}_2) = P(A_1)-P(A_1A_2) = P(A_1)-P(A_1)P(A_2)$
$= P(A_1)\left[1-P(A_2)\right] = P(A_1)P(A_2)$, etc...

---

(2) Esto se demuestra con el siguiente ejemplo sencillo (Bernshtein): el conjunto Ω está compuesto de cuatro elementos $\omega_1, \omega_2, \omega_3, \omega_4$, las correspondientes probabilidades elementales p_1, p_2, p_3, p_4, se suponen iguales a ¼ y $A = \{\omega_1, \omega_2\}$, $B = \{\omega_1, \omega_3\}$ y

En la introducción del concepto de independencia no utilizamos la probabilidad condicional.

Nuestro objetivo fue en lo posible explicar con claridad la esencia de este concepto desde un punto de vista puramente matemático.

Su aplicación se fundamenta, sin embargo, de manera principal en las propiedades de algunas probabilidades condicionales. Si suponemos que todas las probabilidades son positivas, entonces de las ecuaciones (3) se deduce que(1)

$$P(A_{k_m}^{(i_m)} / A_{k_1}^{(i_1)} A_{k_2}^{(i_2)} \dots A_{k_{m-1}}^{(i_{m-1})}) = P(A_{k_m}^{(i_m)}) \tag{6}$$

Lo contrario, de las fórmulas (6) se deduce según el teorema del producto (fórmula (7), §4) las fórmulas (2). Tenemos entonces el siguiente teorema.

TEOREMA 2. *Si las probabilidades de todos los* $A_{k_i}^{(i)}$ *son positivas, las pruebas* $U^{(1)}, U^{(2)}, \dots, U^{(n)}$ *son independientes si y sólo si se cumple la siguiente condición: La probabilidad condicional del resultado* $A_{k_i}^{(i)}$ *respecto a hipótesis que las pruebas* $U^{(i_1)}, U^{(i_2)}, \dots, U^{(i_\ell)}$ *hayan obtenido los resultados* $A_{k_1}^{(i)}$,

---

$= \{\omega_1, \omega_4\}$. Entonces es fácil calcular que

$$P(A) = P(B) = P(C) = ½,$$

$$P(AB) = P(BC) = P(AC) = ¼ = (½)^2,$$

$$P(ABC) = ¼ \neq (½)^3.$$

---

(1) Para la demostración conviene recordar la definición de probabilidad condicional (fórmula (5), §4) y cambiar la probabilidad de la intercepción por el producto de las probabilidades según la fórmula (3).

$A_{k_2}^{(i_2)},\ldots,A_{k_\ell}^{(i_\ell)}$, *y sea igual a la probabilidad* $P(A_{k_i}^{(i)})$.

Con base en la fórmula (4) análogamente se demuestra el siguiente teorema.

TEOREMA 3. *Si todas las probabilidades* $P(A_k)$ *son positivas, entonces* $A_1,A_2,\ldots,A_n$ *son independientes sí y sólo si*

$$P(A_k/A_{k_1},A_{k_2},\ldots,A_{k_\ell}) = P(A_k) \tag{7}$$

para todo $k,k_1,k_2,\ldots,k_\ell$ *diferentes entre sí.*

En particular para $n = 2$ la condición (7) se transforma en dos ecuaciones

$$\begin{aligned} P(A_2/A_1) &= P(A_2) \\ P(A_1/A_2) &= P(A_1) \end{aligned} \tag{8}$$

Fácilmente se comprueba que solo la primera de las ecuaciones de (8) es necesaria y suficiente para la independencia de A_1 y A_2 si $P(A_1) > 0$.

*

§6. PROBABILIDAD CONDICIONAL COMO MAGNITUDES ALEATORIAS. CADENAS DE MARKOV.

Supongamos que U es un desarrollo del conjunto base Ω

$$\Omega = A_1 + A_2 + \ldots + A_m$$

y $\xi = \xi(\omega)$ una función real de los sucesos elementales ω la cual es cada conjunto A_i toma los valores x_i:

$$\xi(\omega) = \sum_{i=1}^{m} x_i I_{A_i}(\omega)$$

donde $I_{A_i}(\omega)$ es la función indicadora del conjunto A_i o sea $I_{A_i}(\omega) = 1$ si $\omega \in A_i$ y $I_{A_i}(\omega) = 0$ si $\omega \in \bar{A}_i$.

En este caso se dice que ξ es una magnitud aleatoria que toma un finito número de valores $x_1, \ldots, x_m$ y

$$M\xi = \sum_{i=1}^{m} x_i P(A_i)$$

se llama la esperanza matemática de la aleatoria ξ.

La teoría de las magnitudes aleatorias y su esperanza matemática será desarrollada en el tercer y cuarto capítulo sin limitarnos solo a las magnitudes aleatorias que puedan tomar solo un número finito de diferentes valores.

DEFINICION 1. *La magnitud aleatoria que en cada conjunto A_i toma los valores $P(B/A_i)$ la llamamos probabilidad condicional del suceso B después de la prueba dada U y la denotamos por $P(B/U)(\omega)$, o sencillamente $P(B/U)$:*

$$P(B/U) = \sum_{i=1}^{m} P(B/A_i) I_{A_i}(\omega).$$

Dos pruebas $U^{(1)}$ y $U^{(2)}$ son independientes si y sólo si

$$P(A_i^{(2)} / U^{(1)}) = P(A_i^{(2)}), \qquad i = 1,2,\ldots,m_2.$$

Dadas las pruebas $U^{(1)}, U^{(2)}, \ldots, U^{(n)}$ por $U^{(1)}U^{(2)}\ldots U^{(n)}$ denotamos la prueba que corresponde al desarrollo del conjunto Ω en los productos $A_{i_1}^{(1)} A_{i_2}^{(2)} \ldots A_{i_n}^{(n)}$.

Las pruebas $U^{(1)}, U^{(2)}, \ldots, U^{(n)}$ son independientes si y só lo si $P(A_i^{(k)}/U^{(1)}U^{(2)}\dots U^{(k-1)}) = P(A_i^{(k)})$ para cualquier k e $i^{(1)}$.

DEFINICION 2. *La sucesión de pruebas* $U^{(1)}, U^{(2)}, \dots, U^{(n)}, \dots$ *forma una cadena de Markov, si para cualquier* k *e* i

$$P(A_i^{(k)} / U^{(1)}U^{(2)}\dots U^{(k-1)}) = P(A_i^{(k)}/U^{(k-1)}).$$

Las cadenas de Markov constituyen una generalización natural de las sucesiones de pruebas independientes. Si ponemos

$$p_{i_m i_n}(m,n) = P(A_{i_n}^{(n)}/A_{i_m}^{(m)}), \qquad m < n$$

entonces la fórmula principal de teoría de las cadenas de Markov tomará la siguiente forma

$$p_{i_k i_n}(k,n) = \sum_{i_m} p_{i_k i_m}(k,m) p_{i_m i_n}(m,n) \qquad k<m<n \qquad (1)$$

denotando la matriz $\| p_{i_m i_n}(m,n) \|$ por $p(m,n)$ podemos escribir a (1) así(2)

$$p(k,n) = p(k,m)p(m,n) \qquad k<m<n \qquad (2)$$

*

(1) La necesidad de estas condiciones se deduce del Teorema 2 §5 y que ellas también son suficientes, se concluye del teorema del producto (fórmula (7) §4).

(2) Sobre el sucesivo desarrollo de la teoría de cadenas de Markov. Ver R. von Mises [1], §16 y B. Hostinsky, Méthodes générales du calcul des probabilités, Mém. Sci. Math. 52, París, 1931.

CAPITULO II

§1. ESPACIOS PROBABILISTICOS INFINITOS. AXIOMA DE CONTINUIDAD.

Denotaremos de ordinario, por $\bigcap_m A_m$ la intersección de los conjuntos A_m (número finito e infinito) y por $\bigcup_m A_m$ su unión. En el caso que los conjuntos A_m no se intersecten, a la unión $\bigcup_m A_m$ la llamaremos suma y denotaremos $\sum_m A_m$. En consecuencia

$$\bigcup_m A_m = A_1 \cup A_2 \cup A_3 \cup \ldots,$$

$$\sum_m A_m = A_1 + A_2 + A_3 + \ldots.,$$

$$\bigcap_m A_m = A_1 A_2 A_3 \ldots$$

En las sucesivas afirmaciones que examinaremos además de los Axiomas 1-4 (Cap.I §1) se cumple también el siguiente axioma de continuidad.

5. Para la sucesión decreciente

$$A_1 \supseteq A_2 \supseteq \ldots \supseteq A_n \supseteq \ldots \tag{1}$$

de sucesos de $\mathcal{F}$ tales que

$$\bigcap_n A_n = \emptyset \tag{2}$$

tenemos la igualdad

$$\lim_n P(A_n) = 0. \tag{3}$$

En todo el material sucesivo llamaremos *campo probabilís-*

tico solo a los campos probabilísticos $(\Omega, \mathfrak{F}, P)$ en el sentido examinado en el Cap.I, que además cumplen el Axioma 5. A los campos probabilísticos en el sentido del Cap.I, se les puede llamar campos probabilísticos en un sentido amplio.

Si el sistema de conjuntos F es finito, el axioma 5 se deduce de los axiomas 1-4. En efecto, en este caso solo existe un número finito de diferentes conjuntos en la sucesión (1). Supongamos que A_k es el menor conjunto entre ellos, entonces todos los conjuntos A_{k+1} coinciden con A_k y obtenemos

$$A_k = A_{k+1} = \;{}_{n} A_n = \emptyset$$

$$\lim_{n} P(A_n) = P(\emptyset) = 0.$$

Todos los ejemplos con espacios probabilísticos finitos del Cap.I cumplen, su consecuencia, también el axioma 5. Así el sistema de axiomas 1-4, no es contradictorio.

Al contrario, para los campos infinitos el axioma de continuidad 5 es independiente de los axiomas 1-4, ya que el nuevo axioma es esencial solo para los campos probabilísticos infinitos, resulta casi imposible explicar su significado empírico, como lo hicimos para los axiomas 1-4 en §2 del Cap.I.

En la descripción de algún proceso aleatorio real se pueden obtener solo espacios probabilísticos finitos.

Los campos aparecen solo como una idealización de los esquemas de los fenómenos reales. Nos limitaremos a los esquemas que cumplen el Axioma 5. Esta limitación resulta racional en las más diferentes investigaciones.

TEOREMA 1. (Generalizacion del teorema de la suma)
Si $A_1, A_2, \ldots, A_n, \ldots$ *pertenecen a* $\mathcal{F}$ *y*

$$A = \sum_n A_n \tag{4}$$

entonces

$$P(A) = \sum_n P(A_n). \tag{5}$$

Demostración. Denotemos

$$R_n = \sum_{m>n} A_m$$

es claro que

$$\bigcap_n R_n = \emptyset$$

y por el Axioma 5

$$\lim_n P(R_n) = 0 \tag{6}$$

Por otro lado, según el teorema de la suma

$$P(A) = P(A_1)+\ldots+P(A_n) + P(R_n). \tag{7}$$

De (6) y (7) se deduce (5).

Así hemos demostrado que la probabilidad $P = P(\cdot)$ es una función numerable-aditiva de conjuntos en $\mathcal{F}$.

Al contrario, los axiomas 4 y 5 tienen lugar para toda función numerable-aditiva de conjuntos en cualquier álgebra de conjuntos $\mathcal{F}$[1]. Se define el concepto de campo probabilístico de la siguiente manera.

Si Ω es un conjunto cualquiera, $\mathcal{F}$ un álgebra de subconjuntos de Ω y $P = P(\cdot)$ una función numerable-aditiva no negativa, definida en $\mathcal{F}$ que cumple con la condición $P(\Omega) = 1$.

(1) Ver Kolmogórov y Fomín.

Entonces el sistema $(\Omega,\mathcal{F},P)$ constituye un campo probabilís tico.

TEOREMA 2. (Sobre el recubrimiento). *Si* $A, A_1, A_2, \ldots, A_n, \ldots$ *pertenecen a* $\mathcal{F}$ *y*

$$A \subseteq \bigcup_n A_n \tag{8}$$

entonces

$$P(A) \leqslant \sum_n P(A_n). \tag{9}$$

Demostración. ya que

$$A = A(\bigcup A_n) = A(A_1 + A_2\bar{A}_1 + A_3\bar{A}_1\bar{A}_2 + \ldots)$$

entonces

$$P(A) = P(AA_1) + P(AA_2\bar{A}_1) + \ldots \leqslant P(A_1) + P(A_2) + \ldots$$

*

§2. CAMPOS PROBABILISTICOS DE BOREL.

El álgebra $\mathcal{F}$ de subconjuntos del conjunto Ω se llama álgebra de Borel si todas las sumas numerables $\sum_n A_n$ de conjuntos A_n de $\mathcal{F}$ pertenecen a $\mathcal{F}$. Las álgebras de Borel son llamadas también σ-álgebras.

De la fórmula

$$\bigcup_n A_n = A_1 + A_2\bar{A}_1 + A_3\bar{A}_1\bar{A}_2 + \ldots \tag{1}$$

se puede concluir que la σ-álgebra contiene también todos los conjuntos $\bigcup_n A_n$, compuesto de un número contable de conjuntos

A_n. De la fórmula

$$\bigcap_n A_n = \Omega \setminus \bigcup_n \bar{A}_n$$

se deduce lo mismo para la intersección de conjuntos.

El campo probabilístico $(\Omega, \mathcal{F}, P)$ se llama campo probabilístico de Borel si el álgebra $\mathcal{F}$ correspondiente es un álgebra de Borel.

En los campos de Borel la teoría de las probabilidades obtienen una libertad completa de acción sin peligro de llegar a sucesos que no tengan ninguna probabilidad. Ahora mostramos que es posible limitarse al examen solo de los campos probabilísticos de Borel. Esto será concluído del así llamado teorema sobre extensión al cual pasaremos ahora.

Dado el campo probabilístico $(\Omega, \mathcal{F}_0, P)$. Como es sabido[1], existe una σ-álgebra mínima $\mathcal{F} = G(\mathcal{F}_0)$ que contiene a $\mathcal{F}_0$.

TEOREMA (sobre extensión). *Una función numerable-aditiva no negativa de conjuntos $P = P(\cdot)$ definida en $(\Omega, \mathcal{F}_0)$ siempre puede extenderse (conservando sus propiedades no negativa y numerable-aditiva) a todos los conjuntos de $\mathcal{F} = \sigma(\mathcal{F}_0)$ y además esta extensión es única.*

El álgebra de Borel $\mathcal{F} = \sigma(\mathcal{F}_0)$ junto con la extensión de la función de conjuntos $P = P(\cdot)$ conforman un campo probabilístico $(\Omega, \mathcal{F}, P)$. Este campo es llamado campo probabilístico ampliado de Borel $(\Omega, \mathcal{F}, P)$.

---

(1) F. Hausdorff. Teoría de Conjuntos, Ed. Estatal. Moscú 1937; A.N. Kolmogórov , S.V. Fomín, Elementos de la Teoría de Funciones.

La demostración del teorema sobre extensión, la cual se relaciona a la teoría de funciones aditivas de conjuntos y que debe ser conocida en diferentes tratados, se desarrolla siguiendo el siguiente esquema.

Dado A un subconjunto cualquiera de Ω.

Definimos $P^*(A)$ como la cota inferior de las sumas

$$\sum_n P(A_n).$$

Para todos los cubrimientos

$$A \subseteq \bigcup_n A_n$$

del conjunto A con un número finito o numerable de conjuntos A_n de $\mathcal{F}_o$. Facilmente se demuestra que es la medida exterior en el sentido de Caratheodory(1). De acuerdo con el teorema sobre cubrimientos (§1), $P^*(A)$ coincide con $P(A)$ para todos los conjuntos de $\mathcal{F}_o$. Seguidamente se demuestra que todos los conjuntos de $\mathcal{F}_o$. Seguidamente se demuestra que todos los conjuntos de $\mathcal{F}_o$ son medibles en el sentido de Caratheodory. Ya que los conjuntos medibles conforman un σ-álgebra entonces todos los conjuntos de $\sigma(\mathcal{F}_o)$ son medibles.

La función de conjuntos $P^*(A)$ es numerable aditiva en $\sigma(\mathcal{F}_o)$ y en $(\mathcal{F}_o)$ podemos colocar

$$P(A) = P^*(A).$$

Con esto queda demostrada la existencia de la extensión. La unicidad de la extensión se deduce de la propiedad de mínimo del álgebra $\sigma(\mathcal{F}_o)$.

---

(1) C. Carathéodory, Vorlesrugen über reele Funktionen, Teubner, Berlin und Leipzig, 1918, p. 237-258; A.N. Kolmogórov , S.V. Fomin Cap.V, §3.

OBSERVACION. Si el conjunto (suceso) A de $\mathcal{F}_o$ puede tener sentido en calidad real y de observación (aunque sea aproximado) de sucesos, entonces de esto aún no se deduce que los conjuntos del álgebra ampliado $\sigma(\mathcal{F}_o)$ permitan la misma interpretación en calidad de observación de sucesos reales.

Puede ocurrir que el campo probabilístico $(\Omega,\mathcal{F}_o,P)$ se examina en calidad (aunque sea idealizado) de imagen de sucesos aleatorios reales al mismo tiempo que el campo probabilístico ampliado $(\Omega,\sigma(\mathcal{F}_o),P)$ es solo una construcción puramente matemática.

Los conjuntos de $\sigma(\mathcal{F}_o)$ los examinamos solo como "sucesos ideales" a los cuales no corresponde nada en el mundo externo. Sin embargo, si las afirmaciones que utilizan las probabilidades de estos sucesos ideales nos llevan a la definición de probabilidad de un suceso real de $\mathcal{F}_o$, entonces esta definición automáticamente será no contradictorio desde el punto de vista empírico.

*

§3. EJEMPLOS DE CAMPOS PROBABILISTICOS INFINITOS

1. En el primer capítulo §1 construímos diferentes campos probabilísticos finitos. Supongamos ahora que

$$\Omega = \{\omega_1,\omega_2,\ldots\}$$

en cada conjunto numerable y $\mathcal{F}$ coincide con el conjunto de todos los subconjuntos del Ω. Todos los posibles campos probabilísticos con conjuntos $\mathcal{F}$ como estos se obtienen de la siguiente forma:

Se toma una sucesión de números negativos $\{p_n\}$ que cumplan la condición

$$p_1 + p_2 + \ldots = 1$$

y para cada conjunto A se pone

$$P(A) = \sum_n{}' p_n$$

en donde la suma $\sum_n{}'$ se extiende a todos los índices n, para los cuales ω_n pertenecen a A.

Estos campos probabilísticos son de Borel.

2. Ahora se supone que Ω está constituído por la recta real $\mathbb{R}$. Inicialmente $\mathcal{F}_0$ está formada de todas las sumas finitas de los semisegmentos $[a,b) = \{\omega : a \leqslant \omega < b\}$ (aquí examinamos junto con los intervalos propios con a y b finitos, los intervalos impropios $[-\infty,b)$, $[a,+\infty)$ y $[-\infty,+\infty)$). Fácilmente se ve que $\mathcal{F}_0$ es álgebra. Según el teorema sobre la extensión es posible a cada campo probabilístico $(\Omega,\mathcal{F}_0,P)$ ampliarlo hasta un campo similar $(\Omega,\sigma(\mathcal{F}_0),P)$. El sistema de conjuntos $\mathcal{F} = \sigma(\mathcal{F}_0)$ en nuestro caso está constituído como un sistema de todos los conjuntos de Borel de la recta real.

3. Si $\Omega = \mathbb{R}$ la recta de números reales y $\mathcal{F}$ está formado por todos los conjuntos de Borel de esta recta. Para la construcción del espacio probabilístico con el álgebra de Borel dada $\mathcal{F}$ es suficiente definir en los conjuntos $A \in \mathcal{F}$ cualquier función de conjuntos numerables-aditiva, no negativa $P(A)$, que cumpla la condición $P(\Omega) = 1$.

Esta función, como es sabido[1] se define por

---

(1) Ver A. Lebesgue, Integración y búsqueda de funciones primitivas.

con sus valores

$$P\{[-\infty,x)\} = F(x) \tag{1}$$

para los intervalos especiales $[-\infty,x)$.

La función $F = F(x)$ se llama función de distribución ω. Más adelante se demostrará (Cap.III, §2) que $F(x)$ no es decreciente, continua por la izquierda y tiene los siguientes valores límites

$$\begin{aligned} \lim_{x\to-\infty} F(x) &= F(-\infty) = 0 \\ \lim_{x\to+\infty} F(x) &= F(+\infty) = 1 \end{aligned} \tag{1}$$

Lo contrario, dada una $F = F(x)$, que cumpla estas condiciones, entonces siempre se puede definir una función de conjuntos numerable-aditiva, no negativa $P(A)$ tal que $P(\Omega) = 1$[1]

4. Ahora por conjunto fundamental Ω tomaremos el espacio euclideo n-dimensional de coordenadas $\mathbb{R}^n$ o sea el conjunto de todas las n-plas ordenadas de números reales $\omega = \{x_1,x_2,\ldots,x_n\}$. El sistema $\mathcal{F}$ está formado por todos los conjuntos de Borel[2] en el espacio $\mathbb{R}^n$. En base al análisis análogo hecho en el ejemplo 2, podemos examinar el conjunto más reducido por ejemplo el sistema de todos los intervalos n-dimensionales.

Aquí como función probabilística $P(A)$ podemos tomar cualquier función de conjuntos numerable-aditiva, no negativa, definida en $\mathcal{F}$ y que cumpla la condición $P(\Omega) = 1$.

---

(1) Ver Kolmogórov y Fomín

---

(2) La definición de conjuntos de Borel en $\mathbb{R}$, ver Hausdorff Teoría de Conjuntos.

Esta función de conjuntos se define uniformemente, si damos sus valores

$$P(\Lambda_{a_1,a_2,\dots,a_n}) = F(a_1, a_2,\dots,a_n) \tag{3}$$

para los conjuntos especiales $\Lambda_{a_1,a_2,\dots,a_n}$ donde $\Lambda_{a_1,a_2,\dots,a_n}$ denota el conjunto de todos los ω para los cuales

$$x_i < a_i , \qquad i = 1,2,\dots,n.$$

No es difícil calcular que para los conjuntos

$$\Lambda_{a_1,\dots,a_n}^{b_1,\dots,b_n} = \{\omega : a_1 \leqslant x_1 < b_1,\dots,a_n \leqslant x_n < b_n\}$$

la probabilidad

$$\begin{aligned} P\left(\Lambda_{a_1,\dots,a_n}^{b_1,\dots,b_n}\right) &= F(b_1,\dots,b_n) - \sum_{i=1}^{n} F(b_1,\dots,b_{i-1},a_i,b_{i+1},\dots,b_i) \\ &+ \sum_{i<j} F(b_1\dots,b_{i-1},a_i,b_{i+1},\dots,b_{j-1},a_i,b_{j+1},\dots,b_n) - \dots \\ &+ (-1)^n F(a_1,\dots,a_n). \end{aligned} \tag{4}$$

En este caso, como función $F(a_1,\dots,a_n)$ podemos escoger cualquier función continua por la izquierda, no decreciente para todas las variables y para la cual la expresión de la derecha de (4) sea no negativa para todo $a_i \leqslant b_i$, $i = 1,\dots,n$ y que cumpla las siguientes condiciones:

$$\lim_{a_i\to-\infty} F(a_1,\dots,a_n) = F(a_1,\dots,a_{i-1},-\infty,a_{i+1},\dots,a_n) = 0, \qquad i = 1,2,\dots, n, \tag{5}$$

$$\lim_{a_1\to+\infty,\dots,a_n\to+\infty} F(a_1,\dots,a_n) = F(+\infty,\dots,+\infty) = 1.$$

La función $F(a_1,\ldots,a_n)$ se llama función de distribución de las variables $x_1,\ldots,x_n$.

El análisis de los campos probabilísticos del tipo arriba definido es suficiente para todos los problemas clásicos de la teoría de las probabilidades(1). En particular, la función probabilística en $\mathbf{R}^n$ puede ser definida de la siguiente manera: Se tima cualquier función no negativa definida en $\mathbf{R}^n$, $f = f(x_1,\ldots,x_n)$ tal que

$$\int_{-\infty}^{\infty}\cdots\int f(x_1,\ldots,x_n)dx_1\ldots dx_n = 1$$

y se pone

$$P(A) = \int\underset{A}{\cdots}\int f(x_1,\ldots,x_n)dx_1\ldots dx_n \tag{6}$$

En este caso, la función $f(x_1,\ldots,x_n)$ es la densidad de la probabilidad en el punto $(x_1,\ldots,x_n)$ (ver Cap.III, §2).

Otro tipo de función probabilística en $\mathbf{R}^n$ se obtiene de la siguiente manera: si $\{\omega_i\}$ es una sucesión de puntos de $\mathbf{R}^n$ y $\{p_i\}$ una sucesión de reales no negativos tales que $\Sigma\, p_i = 1$. Entonces, así como en el ejemplo 1, se toma $P(A) = \sum_i' p_i$ las sumas $\sum_i'$ se extienden a todos los índices i para los cuales ω_i pertenecen a A. Los dos tipos de funciones probabilísticas en $\mathbf{R}^n$ aquí recordados no agotan todas las posibilidades a pesar que ellas son suficientes en las aplicaciones de la teoría de las probabilidades.

Se puede, sin embargo, imaginarse, además de los ejemplos

---

(1) Comparar con R. von Mises [1]. Aquí se exige la existencia de las probabilidades para todos los posibles conjuntos del espacio n-dimensional.

clásicos otros problemas interesantes para la aplicación en los cuales los sucesos elementales se definen con ayuda de un número infinito de coordenadas.

Estos campos probabilísticos los examinaremos más de cerca después de introducir algunos conceptos auxiliares. (ver Cap. III §3).

*

CAPITULO III

VARIABLES ALEATORIAS

§1. FUNCIONES PROBABILISTICAS.

Supongamos que está dada la relación del conjunto Ω al conjunto X, constituída de cualquier clase de elementos, o sea definida en Ω la función unívoca $\xi = \xi(\omega)$ cuyos valores pertenecen al conjunto X. A cada subconjunto A de X le ponemos en correspondencia en calidad de preimagen en Ω al conjunto $\xi^{-1}(A)$ de todos los elementos de Ω que están relacionados con uno de los elementos de A. Supongamos, además, que $\mathcal{F}_\xi$ es un sistema de todos los subconjuntos A de X cuyas preimángenes pertenecen al álgebra de conjuntos $\mathcal{F}$.

El sistema $\mathcal{F}_\xi$ entonces es también un álgebra. Si en estas condiciones $\mathcal{F}$ es un álgebra de Borel entonces $\mathcal{F}_\xi$ también lo es. Supongamos, ahora, que

$$P_\xi(A) = P\{\xi^{-1}(A)\} . \qquad (1)$$

Esta función de conjuntos P_ξ definida en $\mathcal{F}_\xi$ cumple con respecto al álgebra $\mathcal{F}_\xi$ todos los axiomas 1-4 y en consecuencia es una función probabilística en $\mathcal{F}_\xi$. Antes de pasar a la demostración de todos estos hechos, formularemos las siguientes condiciones.

DEFINICION. *Dada la función unívoca $\xi = \xi(\omega)$ del suceso aleatorio ω. Entonces la función P_ξ definida por la fórmula (1) se llama función probabilística de ξ.*

OBSERVACION 1. En el análisis del campo probabilístico (Ω, $\mathcal{F}$, P) a la función P se llama función de probabilidad o sencillamente probabilidad, P_ξ función probabilística de ξ. En el caso en que $\xi(\omega)$ sea $=\omega$ $P_\xi(A)$ coincide con $P(A)$.

OBSERVACION 2. El suceso $\xi^{-1}(A)$ está compuesto por todos los $\xi(\omega)$ pertenecientes al conjunto (A). En consecuencia $P_\xi(A)$, es la probabilidad de que $\xi(\omega) \in A$.

A nosotros nos queda por demostrar las propiedades arriba mencionadas para $\mathcal{F}_\xi$ y P_ξ. Ella se desprende, sin embargo de un único hecho que es el siguiente:

LEMA. *La suma, intersección y diferencia de cualquier conjunto preimagen $\xi^{-1}(A)$ es preimagen de la respectiva suma intersección o diferencia de los conjuntos de A.*

Demostración. La demostración de este lema se recomienda como entrenamiento al lector.

Supongamos, ahora que A y B son dos conjuntos de $\mathcal{F}_\xi$, sus preimagenes A' y B' pertenecen entonces a $\mathcal{F}$. Ya que $\mathcal{F}$ es un álgebra, entonces los conjuntos $A'B'$, $A'+B'$ y A'/B' también pertenecen a $\mathcal{F}$. Pero todos estos conjuntos son preimagénes de los conjuntos AB, $A+B$ y A/B en consecuencia los últimos conjuntos pertenecen a $\mathcal{F}_\xi$. Así hemos demostrado que $\mathcal{F}_\xi$ es un álgebra. De la misma forma se demuestra que si $\mathcal{F}$ es un álgebra de Borel, entonces $\mathcal{F}_\xi$ también lo es.

Está claro que $P_\xi(X) = P\{\xi^{-1}(X)\} = P(\Omega) = 1$.

Que P_ξ siempre es no negativa es también un hecho en consecuencia queda por demostrar que P_ξ es finito-aditiva (ver al final del parágrafo 1 Cap.II).

Así, sea los conjuntos A_n y sus preimángenes $\xi^{-1}(A_n)$ entonces

$$P_\xi(\sum_n A_n) = P\{\xi^{-1}(\sum_n A_n)\}$$
$$= P\{\sum_n \xi^{-1}(A_n)\} = \sum_n P\{\xi^{-1}(A_n)\} = \sum_n P_\xi(A_n),$$

y con esto queda demostrado la finitud y actividad de P_ξ.

Por último anotaremos lo siguiente si $\xi_1 = \xi_1(\omega)$ es una función de Ω en X_1 y $\xi_2 = \xi_2(x_1)$ otra función de X_1 en X_2. Entonces la función compuesta $\xi(\omega) = \xi_2[\xi_1(\omega)]$ es una función del conjunto Ω en X_2.

Examinaremos ahora las funciones probabilísticas $P_{\xi_1}(A_1)$ y $P_{\xi_2}(A_2)$ para las funciones $\xi_1(\omega)$ y $\xi(\omega) = \xi_2[\xi_1(\omega)]$. Estas dos funciones probabilísticas están relacionadas como puede fácilmente calcularse de la siguiente manera:

$$P_\xi(A_2) = P_{\xi_1}\{\xi_2^{-1}(A_2)\}. \qquad (2)$$

*

§2. DEFINICION DE VARIABLES ALEATORIAS Y FUNCION DE DISTRIBUCION.

DEFINICION 1. *La función real unívoca $\xi = \xi(\omega)$ definida en el conjunto fundamental Ω se llama variable aleatoria si para cualquier escogencia de un número real x el conjunto $\{\xi < x\}$ de todos los ω para los cuales se cumple la desigualdad $\xi(\omega) < x$ pertenece al sistema de conjunto de F.*

Esta función $\xi(\omega)$ aplica al conjunto fundamental Ω al conjunto $\mathbb{R}$ de todos los números reales. Nuestra definición de variable aleatoria puede ser formulada así:

La función $\xi = \xi(\omega)$ es una variable aleatoria si y sólo si $\mathcal{F}_\xi$ contiene cada uno de los intervalos del tipo $(-\infty, a)$.

Ya que F_ξ es un álgebra entonces ella contiene junto con los intervalos $(-\infty, a)$ todos las posibles sumas finitas de semisegmentos $[a, b)$. Si nuestro campo probabilístico es de Borel entonces $\mathcal{F}$ y $\mathcal{F}_\xi$ son álgebras de Borel; en consecuencia en este caso $\mathcal{F}_\xi$ contiene todos los conjuntos $\mathbb{R}$ de Borel. La función probabilística P_ξ de la variable aleatoria ξ está definida por todos los conjuntos A del álgebra $\mathcal{F}_\xi$.

En particular, en el caso de que el campo probabilístico es de Borel P_ξ está definido para todos los conjuntos $\mathbb{R}$ de Borel.

DEFINICION 2. *La función*

$$F_\xi(x) = P_\xi(-\infty, x) = P\{\xi(\omega) < x\}.$$

donde $-\infty$ y $+\infty$ se permiten en calidad de valores de x se llama función de distribución de la variable aleatoria ξ. De la definición tenemos

$$F_\xi(-\infty) = 0, \quad F_\xi(+\infty) = 1 \tag{1}$$

La probabilidad de que se cumpla la desigualdad $a < \xi < b$ viene dada por la fórmula

$$P(a \leqslant \xi(\omega) < b) = F_\xi(b) - F_\xi(a). \tag{2}$$

De donde se deduce que todo $a < b$.

$$F_\xi(a) \leqslant F_\xi(b).$$

Lo cual indica que $F_\xi(x)$ es una función no decreciente.

Sea $a_1 < a_2 <\dots a_n < \dots \to b$ entonces

$$\bigcap_n \{\omega : \xi(\omega) \in [a_n, b)\} = \emptyset .$$

En consecuencia de acuerdo con los axiomas de continuidad

$$F_\xi(b) - F_\xi(a_n) = P\{\omega : \xi(\omega) \in [a_n, b)\} .$$

tiende a cero si n tiende a infinito. De aquí se ve claro que la función $F_\xi(x)$ es contínua por la izquierda.

Análogamente se puede demostrar que:

$$\lim_{x\to -\infty} F_\xi(x) = F_\xi(-\infty) = 0, \tag{3}$$

$$\lim_{x\to +\infty} F_\xi(x) = F_\xi(+\infty) = 1. \tag{4}$$

Si el campo probabilístico $(\Omega, \mathcal{F}, P)$ es de Borel, entonces los valores de la función probabilística $P_\xi(A)$ para todos los conjuntos de Borel A de $\mathbb{R}$ unívocamente se definen por los valores de la función de distribución $F_\xi(x)$ (ver Cap.II, §3, Cap. III). Ya que fundamentalmente solo nos interesan los valores de $P_\xi(A)$ entonces la función de distribución juega un papel esencial en todo el desarrollo del material siguiente.

Si la función de distribución $F_\xi(x)$ es diferenciable entonces la derivada de x

$$f_\xi(x) = \frac{dF_\xi(x)}{dx}$$

se llama densidad de probabilidad de ξ en el punto x. Si para cada x

$$F_\xi(x) = \int_{-\infty}^{x} f_\xi(y)\,dy.$$

Entonces para cada conjunto de Borel A la función probabilística de ξ puede ser escogida a través de $f_\xi(x)$ la siguiente manera:

$$P_\xi(A) = \int_A f_\xi(x)dx . \qquad (5)$$

En este caso se dice que la distribución de ξ es absolutamente continua. En el caso general por analogía escribimos

$$P_\xi(A) = \int_A dF_\xi(x). \qquad (6)$$

Todos los conceptos introducidos hasta ahora pertenecen a una generalización para el caso de las probabilidades condicionales. La función de conjunto

$$P_\xi(A/B) = P\{\xi \in A/B\}$$

Es la función probabilística condicional de ξ para la hipótesis B (se supone que $P(B) > 0$). La función no decreciente

$$F_\xi(x/B) = P\{\xi < x/B\}$$

Es la función de distribución correspondiente y por último (en caso de diferenciación de $F_\xi(x/B)$),

$$f_\xi(x/B) = \frac{dF_\xi(x/B)}{dx}$$

Es la densidad de probabilidad condicional de ξ en el punto x para la hipótesis B.

*

§3. FUNCIONES DE DISTRIBUCION MULTIDIMENSIONAL.

Sea $\xi_1,\xi_2,\ldots,\xi_n$, n variables aleatorias dadas. El punto $\xi = (\xi_1,\ldots,\xi_n)$ del espacio n dimensional $\mathbb{R}^n$ es una función del suceso elemental ω. En consecuencia por las reglas generales del §1 obtenemos el álgebra de conjuntos $\mathcal{F}_\xi = \mathcal{F}_{\xi_1,\ldots,\xi_n}$ compuestos de subconjuntos del espacio $\mathbb{R}^n$, y las funciones probabilísticas $P_\xi(A) = P_{\xi_1,\ldots,\xi_n}(A)$ definidas en $\mathcal{F}_\xi = \mathcal{F}_{\xi_1,\ldots,\xi_n}$. Esta función probabilística se llama n-dimensional de las variables $\xi = (\xi_1,\ldots,\xi_n)$.

El álgebra $\mathcal{F}_{\xi_1,\ldots,\xi_n}$ contiene (como puede deducirse de la definición de variable aleatoria) para cada escogencia i y a_i, $i = 1,2,\ldots,n$ el conjunto de todos los puntos $x = (x_1,\ldots,x_n)$ $\in \mathbb{R}^n$ para los cuales $x_i < a_i$. En consecuencia $\mathcal{F}_{\xi_1,\ldots,\xi_n}$ contiene también las intersecciones de los conjuntos mencionados, o sea el conjunto $A_{a_1,a_2,\ldots,a_n}$ de todos los puntos $x = (x_1, \ldots,x_n) \in \mathbb{R}^n$ para los cuales se cumple la desigualdad $x_i < a_i$, $i = 1,2,\ldots,n$(1)

Si llamamos semisegmento n-dimensional

$$[a_1,a_2,\ldots,a_n;\ b_1,b_2,\ldots,b_n)$$

al conjunto de todos los puntos $\mathbb{R}^n$ para los cuales se cumplen todas las desigualdades $a_i \leqslant x_i < b_i$ entonces, está claro, que cada uno de estos intervalos también pertenece al álgebra $F_{\xi_1,\ldots,\xi_n}$, ya que

$$[a_1,a_2,\ldots,a_n;\ b_1,b_2,\ldots,b_n) =$$

---

(1) x_i pueden tomar también valores infinitos $\pm\infty$

$$= \Lambda_{b_1,b_2,\dots;b_n} - \Lambda_{a_1,b_2,\dots,b_n} - \Lambda_{a_1,a_2,b_3,\dots,b_n} - \Lambda_{b_1,\dots,b_{n-1},a_n}$$

La dilatación de Borel del sistema de todos los semi-segmentos n-dimensionales está compuesto de todos los conjuntos $\mathbb{R}$ de Borel. De donde se deduce que el álgebra $F_{\xi_1,\dots,\xi_n}$ en el caso de que el campo probabilístico sea de Borel contiene todos los conjuntos de Borel del espacio $\mathbb{R}^n$.

TEOREMA. *Si el campo probabilístico es de Borel y la función $f(x_1,\dots,x_n)$ también lo es, entonces la función $\eta(\omega) = f(\xi_1(\omega),\dots,\xi_n(\omega))$ de un número finito de variables aleatorias $\xi_1 = \xi_1(\omega),\dots,\xi_n = \xi_n(\omega)$ también es una variable aleatoria.*

Para la demostración es suficiente que el conjunto de todos los puntos $(x_1,x_2,\dots,x_n)$ de $\mathbb{R}^n$ para los cuales $f(x_1,x_2,\dots,x_n) < a$ es de Borel. En particular, todas las sumas finitas y productos de variables aleatorias también son variables aleatorias.

DEFINICION. *La función*

$$F_{\xi_1,\dots,\xi_n}(x_1,\dots,x_n) = P_{\xi_1,\dots,\xi}(\Lambda_{x_1,\dots,x_n})$$

la llamamos función de distribución n-dimensional de las variables aleatorias $\xi_1,\dots,\xi_n$.

Como en el caso unidimensional se demuestra que la función de distribución n-dimensional $F_{\xi_1,\dots,\xi_n}(x_1,\dots,x_n)$ para todas las variables es continua por la izquiedda. Análogamente a las igualdades (3) y (4) de §2.

$$\lim_{x_i\to-\infty} F_{\xi_1,\dots,\xi_n}(x_1,\dots,x_n)$$

$$= F_{\xi_1,\dots,\xi_n}(x_1,\dots,x_{i-1},\ ,x_{i+1},\dots,x_n) = 0 \qquad (1)$$

$$\lim_{x_1\to+\infty,\dots,x_n\to+\infty} F_{\xi_1,\dots,\xi_n}(x_1,\dots,x_n)$$

$$= F_{\xi_1,\dots,\xi_n}(+\infty,\dots,+\infty) = 1 \tag{2}$$

La función de distribución $F_{\xi_1,\dots,\xi_n}(x_1,\dots,x_n)$ nos da directamente los valores de $P_{\xi_1,\dots,\xi_n}$ solo para los conjuntos especiales $A_{x_1,\dots,x_n}$, sin embargo, si nuestro campo probabilístico es de Borel, entonces $P_{\xi_1,\dots,\xi_n}$ unívocamente se determina para los conjuntos de Borel a través de la función de distribución $F_{\xi_1,\dots,\xi_n}(x_1,\dots,x_n)$[1].

Si existe la derivada

$$f_{\xi_1,\dots,\xi_n}(x_1,\dots,x_n) = \frac{\partial^n}{\partial x_1\dots\partial x_n} F_{\xi_1,\dots,\xi_n}(x_1,\dots,x_n)$$

esta derivada $f_{\xi_1,\dots,\xi_n}(x_1,\dots,x_n)$ la llamaremos densidad de probabilidad n-dimensional de las variables aleatorias $\xi_1,\dots,\xi_n$, en el punto $(x_1,\dots,x_n)$.

Si para cada punto $(x_1,\dots,x_n)$

$$F_{\xi_1,\dots,\xi_n}(x_1,\dots,x_n) = \int_{-\infty}^{x_1}\cdots\int_{-\infty}^{n} f_{\xi_1,\dots,\xi_n}(y_1,\dots,y_n)dy_1\dots dy_n$$

entonces la distribución de $\xi = (\xi_1,\dots,\xi_n)$ se llama absolutamente continua. Para cada conjunto de Borel $A \subseteq \mathbf{R}^n$ tiene lugar la siguiente igualdad

$$P_{\xi_1,\dots,\xi_n}(A) = \int\underset{A}{\cdots}\int f_{\xi_1,\dots,\xi_n}(y_1,\dots,y_n)dy_1\dots dy_n \tag{3}$$

Al final de este parágrafo anotaremos además una observación

---

(1) Comp. Cap. II, §3, 4,

sobre la correspondencia entre las diferentes funciones probabilísticas y funciones de distribución. Pongamos que está dada la sustitución

$$S = \begin{pmatrix} 1, & 2, & \ldots, & n \\ i_1, & i_2, & \ldots, & i_n \end{pmatrix}$$

y denotamos ρ_s la transformación interna de $\mathbb{R}^n$

$$x'_k = x_{i_k}, \qquad k = 1,2,\ldots,n,$$

entonces, queda claro que:

$$P_{\xi i_1,\ldots,\xi i_n}(A) = P_{\xi_1,\ldots,\xi_n}\{\rho_s^{-1}(A)\} \tag{4}$$

Supongamos ahora que $x' = \Pi_k(x)$ es la proyección del espacio $\mathbb{R}^n$ en el espacio $\mathbb{R}^k$, $k<n$ por la cual el punto $x = (x_1,\ldots,x_n)$ se relaciona con el punto $x' = (x_1,\ldots,x_k)$ entonces, como consecuencia de la fórmula (2) §1.

$$P_{\xi_1,\ldots,\xi_k}(A) = P_{\xi_1,\ldots,\xi_n}\ \{\Pi_k^{-1}(A)\}\ . \tag{5}$$

Para las correspondientes funciones de distribución de (4) y (5) se deduce la igualdad

$$F_{\xi i_1,\ldots,\xi i_n}(x_{i_1},\ldots,x_{i_n}) = F_{\xi_1,\ldots,\xi_n}(x_1,\ldots,x_n),$$

$$F_{\xi_1,\ldots,\xi_k}(x_1,\ldots,x_k)$$

$$= F_{\xi_1,\ldots,\xi_n}(x_1,\ldots,x_k,+\infty,\ldots,+\infty).$$

*

§4. PROBABILIDADES EN ESPACIOS INFINITOS.

En el parágrafo 3 del Cap. II, vimos como se construye en diferentes campos probabilísticos aplicados en la teoría de probabilidad.

Podemos, sin embargo, imaginar interesantes problemas por su aplicabilidad en los cuales los sucesos elementales están definidos con la ayuda de un número infinito de coordenadas. Así podemos escoger un conjunto $\mathcal{N}$ de índices ν de cualquier potencia $\mathfrak{n}$. El conjunto de todos los sistemas

$$\omega = \{ x_\nu \}$$

de números reales x_ν, en donde ν recorre todo el conjunto $\mathcal{N}$ lo llamaremos espacio $\mathbf{R}^{\mathcal{N}}$ (para la definición del elemento ω del espacio $\mathbf{R}^{\mathcal{N}}$ debemos poner en correspondencia a cada elemento del conjunto $\mathcal{N}$ un número real x_ν o lo que es lo mismo, definir una función real unívoca x_ν definida en el conjunto $\mathcal{N}$ para todo elemento ν[1]).

Si el conjunto $\mathcal{N}$ está compuesto de los n números nuevos naturales 1.2...n entonces $\mathbf{R}^{\mathcal{N}}$ es el espacio $\mathbf{R}^n$ ordinario. Si escogemos en calidad de conjunto $\mathcal{N}$ al conjunto de todos los números reales $\mathbf{R}^1$ entonces el correspondiente espacio $\mathbf{R}^{\mathcal{N}}$ (es igual) $= \mathbf{R}^{\mathbf{R}^1}$ está compuesto de todas las funciones reales

$$\omega = \{ x_t \}$$

de argumento real t, $-\infty < t < \infty$.

El conjunto $\mathbf{R}^{\mathcal{N}}$ para cualquier conjunto $\mathcal{N}$ lo podemos tomar ahora como conjunto fundamental Ω. Sea $\omega = \{x_\nu\}$ elemento de Ω,

(1) Comparar con F. Hausdorff. Teoría de Conjuntos.

por $\Pi_{\nu_1,\ldots,\nu_n}(\omega)$ denotaremos al punto $(x_{\nu_1},x_{\nu_2},\ldots,x_{\nu_n})$ n-dimencional del espacio $\mathbb{R}^n$. El subconjunto A de Ω lo llamaremos conjuntos cilindricos si el se puede representar de la forma

$$A = \Pi^{-1}_{\nu_1,\ldots,\nu_n}(A')$$

donde A' es un subconjunto $\mathbb{R}^n$. La clase de todos los conjuntos cilíndricos coincide en consecuencia con la clase de todos los conjuntos que pueden ser definidos por la relación del siguiente tipo

$$f(x_{\nu_1},x_{\nu_2},\ldots,x_{\nu_n}) = 0 \tag{1}$$

Para definir cualquier conjunto cilindrico $\Pi^{-1}_{\nu_1,\ldots,\nu_n}(A')$ por esta relación es suficiente tomar como f, la función que A' sea igual a cero y fuera de A' igual a la unidad.

Los conjuntos cilindricos se llaman conjuntos cilindricos de Borel si respectivamente los conjuntos A' son de Borel. Todos los conjuntos cilindricos[1] de Borel del espacio $\mathbb{R}^N$ constituyen un álgebra, la que en lo sucesivo denotaremos por $\mathcal{F}^N$.

---

(1) De los argumentos anteriores se sigue que los conjuntos cilíndricos de Borel son aquellos que pueden ser definidos con las relaciones de Borel (1). Sean ahora A y B dos conjuntos cilíndricos definidos por las relaciones

$$f(x_{\nu_1},\ldots,x_{\nu_n}) = 0, \quad g(x_{\lambda_1},\ldots,x_{\lambda_m}) = 0$$

Entonces los conjuntos $A+B$, AB y $A\setminus B$ se pueden definir respectivamente así

$$f\cdot g = 0, \quad f^2+g^2 = 0, \quad f^2+h(g) = 0,$$

donde $h(x) = 0$ para $x \neq 0$ y $h(0) = 1$. Si f y g son funciones de Borel, entonces de Borel son también las funciones fg, f^2+g^2 y $f^2+h(g)$. En consecuencia, $A+B$, AB y $A\setminus B$ son conjuntos cilíndricos de Borel. Con esto se demuestra que $\mathcal{F}^N$ es un álgebra.

La dilatación de Borel del álgebra $\mathcal{F}^{\mathcal{N}}$ la denotaremos como siempre $\sigma(\mathcal{F}^{\mathcal{N}})$. El conjunto del sistema $\sigma(\mathcal{F}^{\mathcal{N}})$ lo llamaremos conjunto de Borel del espacio $\mathbb{R}^{\mathcal{N}}$.

Seguidamente daremos el método para la construcción y manipulación con las funciones probabilísticas en $\mathcal{F}^{\mathcal{N}}$ y en consecuencia según el teorema sobre la prolongación también en $\sigma(\mathcal{F}^{\mathcal{N}})$.

En el caso de numerabilidad del conjunto $\mathcal{N}$ obtenemos de esta manera un campo probabilístico suficiente para todos nuestros objetivos.

Dominamos en consecuencia todas las preguntas relacionadas con la numerabilidad de la sucesión de variables aleatorias. Si el conjunto $\mathcal{N}$ es no numerable entonces muchos subconjuntos de $\mathbb{R}^{\mathcal{N}}$ quedan fuera de sistema $\sigma(\mathcal{F}^{\mathcal{N}})$. Por ejemplo; el conjunto de todos los elementos ω para los cuales x_ν en cualquier escogencia del índice ν es menor que una magnitud constante c, o sea el conjunto $\{\omega : x_\nu < c,\ \nu \in \mathcal{N}\}$ no pertenece al sistema $\sigma(\mathcal{F}^{\mathcal{N}})$ en el caso del no numerabilidad de $\mathcal{N}$.

En consecuencia es bueno procurar obtener en lo posible la conversión de cualquier problema a la forma en la cual el espacio de todos los sucesos elementales ω tiene solo un número de coordenadas.

Supongamos que en $\mathcal{F}^{\mathcal{N}}$ está definida una función de probabilidad P. Cada coordenada x_ν del suceso elemental ω puede examinarse como variable aleatoria. En consecuencia todo grupo finito $(x_{\nu_1}, x_{\nu_2}, \ldots, x_{\nu_n})$ de estas coordenadas tienen una función probabilística n-dimensional $P_{\nu_1,\nu_2,\ldots,\nu_n}(A)$ y una función de distribución correspondiente $\mathcal{F}_{\nu_1,\nu_2,\ldots,\nu_n}(a_1,a_2,\ldots,a_n)$. Es claro que para todo conjunto cilíndrico de Borel

$A = \Pi^{-1}_{\nu_1,\nu_2,\ldots,\nu_n}(A')$ tiene lugar la siguiente igualdad

$$P(A) = P_{\nu_1,\nu_2,\ldots,\nu_n}(A')$$

además A' es un conjunto de Borel de $\mathbb{R}^n$. De esta manera la función de probabilidad P es una función unívocamente definida en el álgebra $\mathcal{F}^N$ de todos los conjuntos cilíndricos a través de los valores de todas las funciones probabilísticas finita para todos los conjuntos de Borel correspondientes al espacio $\mathbb{R}^n$. Sin embargo, para los conjuntos de Borel los valores de la función probabilística $P_{\nu_1,\nu_2,\ldots,\nu_n}$ univocamente se define a través de las correspondientes funciones de distribución. En consecuencia, hemos demostrado el siguiente teorema.

TEOREMA. *El conjunto de todas las funciones de distribución finitas $F_{\nu_1,\nu_2,\ldots,\nu_n}$ unívocamente define la función de probabilidad P para todos los conjuntos de $\mathcal{F}^N$. La función de probabilidad P (según el teorema sobre prolongación) se define unívocamente en $\sigma(\mathcal{F}^N)$ por los valores de la función de distribución $F_{\nu_1,\nu_2,\ldots,\nu_n}$*

Ahora, podemos preguntarnos bajo que condiciones a priori el sistema de funciones de distribución dado $F_{\nu_1,\nu_2,\ldots,\nu_n}$ define la probabilidad en $\mathcal{F}^N$ (y en consecuencia en $\sigma(\mathcal{F}^N)$).

Primeramente anotamos que toda función de distribución $F_{\nu_1,\nu_2,\ldots,\nu_n}$ debe cumplir las condiciones dadas en el Cap.II (§3,III) que, por supuesto están contenidas en el concepto mismo de función de distribución. Además, como consecuencia de las fórmulas (13) y (14) de §2 tienen lugar las siguientes relaciones:

$$F_{\nu_{i_1},\nu_{i_2},\ldots,\nu_{i_n}}(x_{i_1},x_{i_2},\ldots,x_{i_n}) = F_{\nu_1,\nu_2,\ldots,\nu_n}(x_1,x_2,\ldots,x_n) \tag{2}$$

$$F_{\nu_1,\nu_2,\ldots,\nu_k}(x_1,x_2,\ldots,x_k) = F_{\nu_1,\nu_2,\ldots,\nu_n}(x_1,x_2,\ldots,x_k,+\infty,\ldots+\infty) \qquad (3)$$

donde $k < n$ y $\begin{bmatrix} 1, 2, \ldots, \\ i_1, i_2, \ldots, i_n \end{bmatrix}$ es una sustitución cualquiera.

Estas condiciones necesarías, sin embargo, son también suficientes como se evidencia del siguiente teorema:

TEOREMA FUNDAMENTAL. *Todo sistema de función de distribución $F_{\nu_1,\nu_2,\ldots,\nu_n}$ que cumpla las condiciones (2) y (3) define una función de probabilidad P en $\mathcal{F}^N$ la cual cumple todos los axiomas 1-5.*

Esta función de probabilidad P puede ser prolongada (según el teorema sobre prolongación) en $\sigma(\mathcal{F}^N)$.

Demostración. Supongamos que está dadas las funciones de distribución $F_{\nu_1,\nu_2,\ldots,\nu_n}$ que cumplen las condiciones generales del Cap.II (§3, III) y las condiciones (2) y (3). Toda función de distribución $F_{\nu_1,\nu_2,\ldots,\nu_n}$ unívocamente define una función de probabilidad $P_{\nu_1,\nu_2,\ldots,\nu_n}$ para todos los conjuntos de Borel de $\mathbf{R}^n$ (ver §3).

En lo sucesivo examinaremos solo conjuntos de Borel de $\mathbf{R}^n$ y conjuntos cilíndricos de Borel en Ω.

Para todo conjunto cilíndrico $A = \Pi^{-1}_{\nu_1,\nu_2,\ldots,\nu_n}(A')$ denotemos

$$P(A) = P_{\nu_1,\nu_2,\ldots,\nu_n}(A') \qquad (4)$$

Ya que un mismo conjunto cilíndrico A puede ser definido a través de diferentes conjuntos, A' entonces, es necesario inicial mente demostrar que la fórmula (4) da siempre un mismo valor para $P(A)$.

Sea $(x_{\nu_1}, x_{\nu_2}, \ldots, x_{\nu_n})$ sea un sistema finito de variables aleatorias x_ν partiendo de las funciones probabilísticas $P_{\nu_1, \nu_2, \ldots, \nu_n}$ de estas variables aleatorias, podemos según las reglas definir la función probabilística (§2) $P_{\nu_{i_1} \nu_{i_2}, \ldots, i_k}$ de cada subsistema $(x_{\nu_{i_1}}, x_{\nu_{i_2}}, \ldots, x_{\nu_{i_k}})$. Las igualdades (2) y (3) tienen como consecuencia que esta función de probabilidad definida según el §2 coincide con la función dada a priori $P_{\nu_{i_1}, \nu_{i_2}, \ldots, \nu_{i_k}}$. Supongamos ahora que el conjunto cilíndrico A se define a través de $A = \Pi^{-1}_{\nu_{i_1}, \nu_{i_2}, \ldots, \nu_{i_k}}(A')$ y simultáneamente por $A = \Pi^{-1}_{\nu_{j_1}, \nu_{j_2}, \ldots, \nu_{j_m}}(A'')$ además todas las variables aleatorias x_{ν_i} y x_{ν_j} pertenecen al sistema $(x_{\nu_1}, x_{\nu_2}, \ldots, x_{\nu_n})$ lo cual, claramente, no es ninguna restricción. La condición $(x_{\nu_{i_1}}, x_{\nu_{i_2}}, \ldots, x_{\nu_{i_k}}) \in A'$ y $(x_{\nu_{j_1}}, x_{\nu_{j_2}}, \ldots, x_{\nu_{j_m}}) \in A''$ son equivalentes. En consecuencia

$$P_{\nu_{i_1} \nu_{i_2}, \ldots, \nu_{i_k}}(A') = P_{\nu_1, \nu_2, \ldots, \nu_n}\{x_{\nu_{i_1}}, x_{\nu_{i_2}}, \ldots, x_{\nu_{i_k}}) \in A'\}$$

$$= P_{\nu_1, \nu_2, \ldots, \nu_n}\{(x_{\nu_{j_1}}, x_{\nu_{j_2}}, \ldots, x_{\nu_{j_m}}) \in A''\} = P_{\nu_{j_1}, \nu_{j_2}, \ldots, \nu_{j_m}}(A'')$$

lo cual demuestra nuestra afirmación respecto a la univalencia de la definición $P(A)$.

Demostremos ahora, que el campo probabilístico $(\Omega, \mathcal{F}^{\mathcal{N}}, P)$ cumple todos los axiomas 1-5. El axioma 1 cumple solo que $\mathcal{F}^{\mathcal{N}}$ debe ser un álgebra; esto (excepto la axigencia de que $\Omega \in \mathcal{F}^{\mathcal{N}}$) fue demostrado arriba. Seguidamente, para cualquier $\nu \in \mathcal{N}$

$$\Omega = \Pi^{-1}_{\nu}(\mathbb{R}^1) \ ; \quad P(\Omega) = P_\nu(\mathbb{R}^1) = 1$$

lo cual demuestra la aplicabilidad de los axiomas 1 y 3.

Finalmente de la definición de $P(A)$ directamente se deduce que $P(A)$ es no negativa (axioma 2).

Un poco más complicado es la demostración de aplicabilidad del axioma 4, para este objetivo examinaremos los dos conjuntos cilíndricos

$$A = \Pi^{-1}_{\nu_{i_1},\nu_{i_2},\ldots,\nu_{i_k}}(A')$$

y

$$B = \Pi^{-1}_{\nu_{j_1},\ \nu_{j_2},\ldots,\nu_{j_m}}(A'').$$

Bajo estas condiciones suponemos que todas las variables x_{ν_i} y x_{ν_j} pertenecen a su sistema volumétrico finito $(x_{\nu_1}, x_{\nu_2},\ldots,x_{\nu_n})$ si los conjuntos A y B son entonces las relaciones

$$(x_{\nu_{i_1}},\ldots,x_{\nu_{i_k}}) \in A'$$

$$(x_{\nu_{j_1}},\ldots,x_{\nu_{j_m}}) \in B'$$

son incompatibles en consecuencia

$$P(A+B) = P_{\nu_1,\nu_2,\ldots,\nu_n}\{(x_{\nu_{i_1}},x_{\nu_{i_2}},\ldots,x_{\nu_{i_k}}) \in A'$$

$$ó \quad (x_{\nu_{j_1}},x_{\nu_{j_2}},\ldots,x_{\nu_{j_m}}) \in B'\}$$

$$= P_{\nu_1,\nu_2,\ldots,\nu_n}\{(x_{\nu_{i_1}},x_{\nu_{i_2}},\ldots,x_{\nu_{i_k}}) \in A'\}$$

$$+ P_{\nu_1,\nu_2,\ldots,\nu_n}\{(x_{\nu_{j_1}},x_{\nu_{j_2}},\ldots,x_{\nu_{j_m}}) \in B'\}$$

$$= P(A) + P(B).$$

Queda por comprobar el axioma 5. Pongamos

$$A_1 \supseteq A_2 \supseteq \ldots \supseteq A_n \supseteq \ldots$$

es una sucesión decreciente de conjuntos cilíndricos que cumple con la condición $\lim_n P(A_n) = L > 0$.

Demostraremos que la intersección de todos los conjuntos A_n no es vacía. Se puede sin ninguna restricción suponer que en la definición de los n primeros conjuntos cilíndricos A_k entran solo los n primeras coo denadas x_{ν_k} de la sucesión $x_{\nu_1}, x_{\nu_2}, \ldots, x_{\nu_k}, \ldots$ o sea que $A_n = \Pi^{-1}_{\nu_1,\nu_2,\ldots,\nu_n}(B_n)$.

Pongamos para mayor sencillez $P_n(B) = P_{\nu_1,\nu_2,\ldots,\nu_n}(B)$. Entonces es claro que: $P_n(B_n) = P(A_n) \geqslant L > 0$. En cada conjunto B_n se puede encontrar un conjunto cerrado acotado U_n tal que $P_n(B_n \setminus U_n) \leqslant \frac{\varepsilon}{2^n}$ y de esta desigualdad, para el conjunto $V_n = \Pi^{-1}_{\nu_1,\nu_2,\ldots,\nu_n}(U_n)$. Se deduce la desigualdad

$$P(B_n \setminus U_n) \leqslant \frac{\varepsilon}{2^n} \tag{5}$$

Supongamos ahora que $W_n = V_1 V_2 \ldots V_n$, de (5) se deduce que: $P(A_n \setminus W_n) < \varepsilon$, ya que $W_n \subseteq V_n \subseteq A_n$. Entonces

$$P(W_n) \geqslant P(A_n) - \varepsilon \geqslant L - \varepsilon .$$

Si ε es suficientemente pequeña, entonces $P(W_n) > 0$. El conjunto W_n no es vacío. Escogeremos ahora, en cada conjunto W_n un punto $\omega^{(n)}$ con coordenadas $x_\nu^{(n)}$. Todo punto $\omega^{(n+p)}$, $p = 0,1,2,\ldots$ pertenece al conjunto V_n en consecuencia

$$(x_{\nu_1}^{(n+p)}, x_{\nu_2}^{(n+p)}, \ldots, x_{\nu_n}^{(n+p)}) = \Pi^{-1}_{\nu_1,\nu_2,\ldots,\nu_n}(\omega^{(n+p)}) \in U_n$$

ya que los conjuntos U_n son acotados entonces de la sucesión $\{\omega^{(n)}\}$ se puede escoger (por el método de la diagonal) una subsucesión

$$\omega^{(n_1)},\omega^{(n_2)},\ldots,\omega^{(n_i)},\ldots$$

Para las cuales las correspondientes coordenadas $x_{\nu_k}^{(n_i)}$ tiende para cualquier k a un fin determinado x_k. Supongamos finalmente que ω es un punto del conjunto Ω con coordenadas

$$x_{\nu_k} = x_k ,$$
$$x_{\nu} = 0 \qquad (\nu \neq \nu_k, \quad k = 1,2,\ldots).$$

El punto $(x_1,x_2,\ldots,x_k)$ como fin de la sucesión $(x^{(n_i)},x^{(n_i)},\ldots,x_k^{(n_i)})$, $i = 1,2,\ldots$. pertenece al conjunto U_k. En consecuencia ω pertenece al conjunto

$$A_k \subseteq V_k = \Pi^{-1}_{\nu_1,\nu_2,\ldots,\nu_k}(U_k)$$

para cualquier k, y en consecuencia a la intersección

$$A = \bigcap_k A_k.$$

El teorema queda demostrado.

*

§5. VARIABLES ALEATORIAS EQUIVALENTES, DIFERENTES TIPOS DE CONVERGENCIA.

Desde este parágrafo examinaremos exclusivamente campos probabilísticos de Borel (Ω,F,P). Esto no es, como se explicó en el parágrafo 2 del Cap.II una restricción esencial para nuestros análisis.

DEFINICION 1. *Dos variables aleatorias ξ y η se llaman equi* *valentes si la probabilidad de la relación $\xi \neq \eta$. Esta claro que dos variables aleatorias diferentes tienen una misma función de probabilidad.*

En consecuencia las funciones de distribución F_ξ y F_η también coinciden. En muchas preguntas de las teorías de las probabilidades es posible cambiar una variable aleatoria por su variable equivalente.

Supongamos ahora que

$$\xi_1, \xi_2, \ldots, \xi_n, \ldots \tag{1}$$

Es una sucesión de variables aleatorias. Exiaminemos el conjunto A de todos los sucesos elementales ω para los cuales la sucesión (1) converge. Si por $A_{np}^{(m)}$ denotamos al conjunto de todo los ω para los cuales se cumplen las desigualdades

$$|\xi_{n+k} - \xi_n| < \frac{1}{m} \qquad k = 1, 2, \ldots, p$$

entonces directamente obtenemos que:

$$A = \bigcap_m \bigcup_n \bigcap_p \dot{A}_{np}^{(m)} \tag{2}$$

De acuerdo con el parágrafo (3) el conjunto $A_{np}^{(m)}$ pertenece siempre al σ-álgebra de conjunto $\mathfrak{F}$. La relación (2) nos muestra que el conjunto A también pertenece a $\mathfrak{F}$. En consecuencia siempre tiene sentido hablar sobre probabilidad de convergencia de sucesión de convergencia de variables aleatorias. Supongamos ahora que la probabilidad $P(A)$ del conjunto $A = 1$ (a la unidad). Entonces nosotros afirmamos que la sucesión (1) converge con probabilidad 1 hacia una variable aleatoria ξ la cual se define unívocamente con precisión de equivalencia.

Para la construcción de esta variable aleatoria supongamos que: $\xi = \lim_{n} \xi_n$ en A y $\xi = 0$ fuera de A. Nos quede por demostrar ξ es una variable aleatoria, o sea que el conjunto $A(x)$ de elementos ω para los cuales $\xi < x$ pertenecen al algebra F. Pero

$$A(x) = A \bigcup_{n} \bigcap_{p} \{\omega : \xi_{n+p} < x\}$$

En el caso $x < 0$ entonces

$$A(x) = A \bigcup_{n} \bigcap_{p} \{\omega : \xi_{n+p} < x\} + \bar{A} ,$$

de donde directamente se desprende nuestra afirmación.

DEFINICION 2. *Si la probabilidad de convergencia de la sucesión $\xi_1, \xi_2, \ldots$, así a la variable $\xi = 1$ entonces, nosotros decimos que ésta sucesión casi siempre converge a ξ. Sin embargo, para la teoría de las probabilidades también son importantes otros tipos de convergencia.*

DEFINICION 3. *La sucesión de variables aleatorias $\xi_1, \xi_2, \ldots$ converge por probabilidad a la variable aleatoria ξ, si para todo $\varepsilon > 0$ o la probabilidad*

$$p\{|\xi_n - \varepsilon| > \varepsilon\} \to 0$$

cuando $n \to \infty$ (1).

1. Si la sucesión (1) converge por probabilidad simultáneamente a ξ y ξ' entonces ξ y ξ' son equivalentes.

En efecto

---

(1) Este concepto aparece en lo fundamental desde Bernoulli, sin embargo en su complet generalización fue introducida por Slutsky [1].

$$P\{|\xi-\xi'|>\frac{1}{m}\}\leqslant P\{|\xi-\xi_n|>\frac{1}{2m}\}+P\{|\xi'-\xi_n|>\frac{1}{2m}\}.$$

2. Si la sucesión (1) casi siempre converge a ξ entonces ella converge a ξ también por probabilidad.

Sea A el conjunto de convergencia de la sucesión (1). Entonces

$$1=P(A)\leqslant \lim_n \{|\xi_{n+p}-\xi|<\varepsilon,\quad p=1,2,\dots\}$$
$$\leqslant \lim_n \{|\xi_n-\xi|<\varepsilon\}$$

de donde se deduce la convergencia por probabilidad.

3. Para la convergencia por probabilidad de la sucesión (1) es necesario y suficiente que se cumpla la siguiente condición:

para todo $\varepsilon>0$ existe un n, tal que para cada $p>0$ tiene lugar la siguiente antiguedad

$$P\{|\xi_{n+p}-\xi_n|>\varepsilon\}<\varepsilon .$$

Supongamos ahora que $F(x),F_1(x),F_2(x),\dots$ son las funciones de distribución de las variables aleatorias $\xi,\xi_1,\xi_2,\dots$, Si la sucesión $\xi_1,\xi_2,\dots$ converge por distribución a ξ entonces, la función de distribución $F(x)$ unívocamente se define por valores de las funciones $F_n(x)$, ya que tiene lugar el siguiente teorema.

TEOREMA. *Si la sucesión $\xi_1,\xi_2,\dots$ converge por probabilidad a ξ, entonces la sucesión de las respectivas funciones de distribución $F_n(x)$ converge la función de distribución $F(x)$ de la variable aleatoria ξ en cada punto de continuidad de $F(x)$.*

Demostración. El hecho de que con $F(x)$ se define a través de $\{F_n(x)\}$ se deduce de que $F(x)$ en monótona contínua por la izquierda y unívocamente se define por los valores en los puntos de continuidad[1]. Para la demostración de la convergencia supondremos que x es un punto de continuidad de $F(x)$. Sea $x' < x$. Entonces en el caso de que ξ sea menor que x', $\xi_n \geqslant x$ es necesario $|\xi_n - \xi| > x - x'$.

En consecuencia:

$$\lim_n P\{\xi < x', \xi_n \geqslant x\} = 0$$

$$F(x') = P\{\xi < x'\} \leqslant P\{\xi_n < x\} + P\{\xi < x', \xi_n \geqslant x\}$$

$$= F_n(x) + P\{\xi < x', \xi_n \geqslant x\},$$

$$F(x') \leqslant \liminf_n F_n(x) + \lim_n P\{\xi < x', \xi_n \geqslant x\},$$

$$F(x') \leqslant \liminf_n F_n(x) \tag{3}$$

Análogamente se demuestra que $x'' > x$ entonces

$$F(x'') \geqslant \limsup_n F_n(x) \tag{4}$$

ya que $F(x')$ y $F(x'')$ cuando $x' \uparrow x$ y $x'' \downarrow x$ tiende $F(x)$ entonces de (3) y (4) se deduce que:

$$\lim_n F_n(x) = F(x).$$

El teorema queda demostrado.

---

(1) En efecto, ella tiene a lo sumo un conjunto contable de puntos de discontinuidad (Lebesgue. Integración..., p.70) por eso los puntos de continuidad son siempre densos y los valores de la función $F(x)$ en el punto de discontinuidad se define como límite de sus valores de los puntos continuos situados a la izquierda.

CAPITULO IV

ESPERANZA MATEMATICA

§1. INTEGRAL ABSTRACTA DE LEBESGUE

Sea ξ una variable aleatoria y A un conjunto de $\mathfrak{F}$[1]. Para un número positivo λ construyamos la suma

$$S_\lambda = \sum_{k=-\infty}^{+\infty} k\lambda P[\{k\lambda \leqslant \xi < (k+1)\lambda\} \cap A]. \qquad (1)$$

Si esta serie converge absolutamente para cualquier λ, entonces cuando $\lambda \to 0$, S_λ tiende a un límite definido, el cual por definición es la integral

$$\int_A \xi(\omega)\, P(d\omega)\ . \qquad (2)$$

Esta forma abstracta del concepto de integral fue introducida por Fréchet[2]. En particular éste concepto resulta necesario para la teoría de las probabilidades. (En los siguientes pará grafos el lector verá que la definición corriente de esperanza matemática condicional de la variable $\mathfrak{F}$ con hipótesis A co rresponde exactamente a la definición de la integral (2) dife renciándose sólo en un factor constante).

---

(1) Como fue mencionado en el §5 del capítulo tercero nosotros examinamos en este y en los siguientes capítulo solo campos pro babilísticos de Borel.

---

(2) Fréchet, Sur l'integrale d'une fonctionnel étendué á un ensemble abstrait, Bull. Soc. Math. France, V.43 (1915), p.248

Daremos aquí una lista breve de las más importantes propiedades de las integrales de forma (2). El lector encontrará sus demostraciones en cada texto de teoría de funciones de va viable real, a pesar que ellas en su mayoría están hechas para el caso cuando P es una medida de Lebesgue de los conjuntos en $\mathbb{R}^n$.

La transformación de estas demostraciones al caso general no representa un nuevo problema matemático, una gran cantidad de ellas exactamente las mismas[1].

1. Si la variable aleatoria ξ es integrable en A, entonces ella es integrable en cualquier subconjunto A' de A, que pertenece a $\mathcal{F}$.

2. Si ξ es integrable en A y A y divide en un número discreto de conjuntos disyuntos A_n de $\mathcal{F}$, entonces,

$$\int_A \xi(\omega)P(d\omega) = \sum_n \int_{A_n} \xi(\omega)P(d\omega) .$$

3. Si ξ integrable entonces $|\xi|$ también lo es y además

$$\left| \int_A \xi(\omega)P(d\omega) \right| \leq \int_A |\xi(\omega)|P(d\omega) .$$

4. Si para cada ω se cumple la desigualdad $0 \leq \eta(\omega) \leq \xi(\omega)$, entonces si $\xi(\omega)$ es integrable la variable $\eta(\omega)$[2] también lo es, además,

---

(1) Ver Kolmogórov y Fomin.

---

(2) Aquí se supone que $\eta(\omega)$ es una variable aleatoria, o sea en la terminología de la teoría general de la integración ella es medible con respecto a $\mathcal{F}$.

$$\int_A \eta(\omega)P(d\omega) < \int_A \xi(\omega)P(d\omega).$$

5. Si $m \leqslant \xi(\omega) \leqslant M$ en donde m, M son constantes, entonces

$$mP(A) \leqslant \int_A \xi(\omega)P(d\omega) \leqslant M\cdot P(A).$$

6. Si $\xi(\omega)$ y $\eta(\omega)$ son integrables y K y L son dos constantes reales, entonces la variable aleatoria $K\xi(\omega) + L\eta(\omega)$ también es integrable y

$$\int_A |K\xi(\omega)+L\eta(\omega)|P(d\omega) = K\int_A \xi(\omega)P(d\omega)+L\int_A \eta(\omega)P(d\omega).$$

7. Si la serie

$$\sum_n \int_A |\xi_n(\omega)|P(d\omega)$$

converge, entonces la serie

$$\sum_n \xi_n(\omega) = \xi(\omega)$$

converge en cada punto del conjunto A, exactamente, hasta algún conjunto B tal que $P(B) = 0$.

Si en el exterior del conjunto $A\setminus B$ hacemos $\xi(\omega) = 0$ entonces

$$\int_A \xi(\omega)P(d\omega) = \sum_n \int_A \xi_n(\omega)P(d\omega).$$

8. Si $\xi(\omega)$ y $\eta(\omega)$ son equivalentes, $P\{\xi(\omega) \neq \eta(\omega)\} = 0$ entonces para cada conjunto A de $\mathcal{F}$

$$\int_A \xi(\omega)P(d\omega) = \int_A \eta(\omega)P(d\omega). \tag{3}$$

9. Si tiene lugar la desigualdad (3) para cada conjunto A de F, entonces $\xi(\omega)$ y $\eta(\omega)$ son equivalentes.

De la definición integral recordada más arriba se obtienen además las siguientes propiedades las cuales no aparecen en la teoría de Lebesgue común.

10. Sean P_1 y P_2 dos funciones probabilísticas, definidas en una misma σ-álgebra F, $P = P_1+P_2$ y $\xi = \xi(\omega)$ integrable en el conjunto A respecto a P_1 y P_2. Entonces

$$\int_A \xi(\omega)P(d\omega) = \int_A \xi(\omega)P_1(d\omega) + \int_A \xi(\omega)P_2(d\omega).$$

11. Toda variable aleatoria acotada es integrable.

*

§2. ESPERANZA MATEMATICA ABSOLUTA Y CONDICIONAL.

Sea $\xi = \xi(\omega)$ una variable aleatoria integrable. La integral $M\xi = \int_\Omega \xi(\omega)P(d\omega)$ se llama en la teoría de las probabilidades esperanza matemática de la variable. De las propiedades 3, 4, 5, 6, 7 y 9 se desprende que

1. $|M\xi| \leqslant M|\xi|$;

2. $M\eta \leqslant M\xi$, si $0 \leqslant \xi(\omega) \leqslant \eta(\omega)$;

3. $\inf_\omega \xi(\omega) \leqslant \xi \leqslant \sup_\omega \xi(\omega)$;

4. $M(K\xi + L\eta) = KM\xi + LM\eta$;

5. $M(\sum_n \xi_n) = \sum_n M\xi_n$ si la serie $\sum_n M|\xi_n|$ converge;

6. Si ξ y η son equivalentes, entonces $M\xi = M\eta$;

7. Todo variable acotada tiene esperanza matemática.

Por definición la integral

$$M\xi = \lim_{\lambda\to 0} \sum_{k=-\infty}^{+\infty} k\lambda P\{k\lambda \leqslant \xi(\omega) \leqslant (k+1)\lambda\}$$

$$= \lim_{\lambda\to 0} \sum_{k=-\infty}^{+\infty} k\lambda [F_\xi((k+1)\lambda) - F_\xi(k\lambda)] .$$

La segunda fila no es otra cosa que la definición de la integral de Stieltjes $\int_{-\infty}^{\infty} xF_\xi(dx)$. Por eso

$$M\xi = \int_{-\infty}^{\infty} xF_\xi(dx) \tag{1}$$

La fórmula (1) puede, en consecuencia servir también como definición de la esperanza matemática M_ξ.

Sea $\xi = \xi(\omega)$ una función de sucesos elementales ω, y η una variable aleatoria, definida como una función unívoca de ξ: $\eta = \eta(\xi)$. Entonces

$$P\{k\lambda \leqslant \eta < (k+1)\lambda\} = P_\xi\{x : k\ \lambda \leqslant \eta(x) < (k+1)\lambda\}$$

donde P_ξ es una función probabilística de ξ. De donde por la definición de la integral tenemos que

$$\int_\Omega \eta(\xi(\omega))P(d\omega) = \int_X \eta(x)P_\xi(dx)$$

en consecuencia

$$M\eta = \int_X \eta(x)P_\xi(dx) \tag{2}$$

donde X es el conjunto de todos los valores posibles de B. En particular, cuando $\xi = \xi(\omega)$ es una variable aleatoria entonces

$$M\eta = \int_{\Omega} \eta(\xi(\omega))P(d\omega) = \int_{R} \eta(x)P_{\xi}(dx) = \int_{-\infty}^{\infty} \eta(x)F_{\xi}(dx). \qquad (3)$$

La última integral en la fórmula (3) constituye, en el caso de la continuidad de la función $\eta = \eta(x)$, una integral de Stiltey

$$\int_{-\infty}^{\infty} \eta(x)F_{\xi}(dx)$$

puede existir también en el caso de no existir la esperanza matemática $M\eta$. Para la existencia de $M\eta$ es necesario y suficiente que la integral $\int_{-\infty}^{\infty}|\eta(x)|F_{\xi}(dx)$ sea finita[1].

Si $\xi = (\xi_1,\xi_2,\ldots,\xi_n)$ es un punto aleatorio del espacio $\mathbf{R}^n$, entonces según el corolario (2)

$$M\eta = \int_{\mathbf{R}^n}\!\!\cdots\!\int \eta(x_1,x_2,\ldots,x_n)P_{\xi_1,\xi_2,\ldots,\xi_n}(dx_1,\ldots,dx_n) \qquad (4)$$

Hemos visto que si $P(B) > 0$, entonces la probabilidad condición $P(\cdot|B)$ tiene todas las propiedades de la función probabilística (cuando B es un conjunto fijo).

La respectiva integral

$$M(\xi \mid B) = \int_{\Omega} \xi(\omega)P(d\omega \mid B). \qquad (5)$$

La llamaremos esperanza matemática condicional de la variable aleatoria $\xi = \xi(\omega)$ respecto al suceso B. Ya que

(1) Comp. V. Glivenko, Sur les valeurs probables de fonctions, Rend. Accad. Lincei, V.8 (1928), p.480-483.

$$P(\bar{B} \mid B) = 0$$

$$\int_{\bar{B}} \xi(\omega) P(d\omega \mid B) = 0$$

Entonces de (3) se desprende la igualdad

$$\begin{aligned} M(\xi \mid B) &= \int_{\Omega} \xi(\omega) P(d\omega \mid B) \\ &= \int_{B} \xi(\omega) P(d\omega \mid B) + \int_{\bar{B}} \xi(\omega) P(d\omega \mid B) \\ &= \int_{B} \xi(\omega) P(d\omega \mid B). \end{aligned}$$

Recordamos que en el caso cuando $A \subseteq B$ y $P(B) > 0$

$$P(A \mid B) = \frac{P(AB)}{P(B)} = \frac{P(A)}{P(A)}$$

obtenemos

$$M(\xi \mid B) = \frac{1}{P(B)} \int_{B} \xi(\omega) P(d\omega) \tag{6}$$

$$P(B) M(\xi \mid B) = \int_{B} \xi(\omega) P(d\omega) \tag{7}$$

de (6) y de la igualdad

$$\int_{A+B} \xi(\omega) P(d\omega) = \int_{A} \xi(\omega) P(d\omega) + \int_{B} \xi(\omega) P(d\omega)$$

se deduce finalmente que

$$M(\xi \mid A+B) = \frac{P(A) M(\xi \mid A) + P(B) M(\xi \mid B)}{P(A+B)} \tag{8}$$

En particular, si $0 < P(A) < 1$, entonces tiene lugar la siguiente fórmula:

$$M_\xi = P(A)M(\xi \mid A) + P(\bar{A})M(\xi \mid \bar{A}) \tag{9}$$

*

§3. DESIGUALDAD DE CHEBISHEV

Sea $f = f(x)$ una función no negativa de argumento real x, la cual para $x \geqslant a$ tiene valores no menores que b en donde $b > 0$. Entonces para cualquier variable aleatoria $\xi = \xi(\omega)$

$$P\{\xi(\omega) \geqslant a\} \leqslant \frac{Mf(\xi)}{b} \tag{1}$$

(Se supone la existencia de la esperanza matemática $Mf(\xi)$.
En efecto

$$Mf(\xi) = \int_\Omega f(\xi(\omega))P(d\omega) \geqslant \int_{\{\omega:\xi(\omega)\geqslant a\}} f(\xi(\omega))P(d\omega) \geqslant bP\{\xi(\omega) \geqslant a\}$$

De donde se desprende (1).

Por ejemplo para todo número positivo c

$$P\{\xi(\omega) \geqslant a\} \leqslant \frac{Me^{c\xi}}{e^{ca}} \tag{2}$$

supondremos ahora que $f = f(x)$ es no negativa, por y además para todo x positivo es una función no decreciente. Entonces para toda variable aleatoria $\xi = \xi(\omega)$ y para toda constante $a > 0$ tendremos la siguiente desigualdad

$$P\{|\xi(\omega)| \geqslant a\} \leqslant \frac{Mf(\xi)}{f(a)} \tag{3}$$

En particular

$$P\{|\xi - M\xi| \geqslant a\} \leqslant \frac{Mf(\xi - M\xi)}{f(a)} \tag{4}$$

Un caso de particular importancia se presenta cuando $f(x) = x^2$. En este caso de (3) obtenemos la desigualdad de Chebishev

$$P\{|\xi(\omega)| \geqslant a\} \leqslant \frac{M\xi^2}{a^2} \tag{5}$$

De (4) obtenemos también que

$$P\{|\xi - M\xi| \geqslant a\} \leqslant \frac{M(\xi - M\xi)^2}{a^2} = \frac{D\xi}{a^2} \tag{6}$$

La magnitud $D\xi = M(\xi - M\xi)^2$

La llamaremos dispersión (varianza) de la variable ξ. Es fácil calcular $D\xi = M\xi^2 - (M\xi)^2$.

La función $f(x)$ es acotada

$$|f(x)| \leqslant K .$$

Entonces $P\{|\xi(\omega)| \geqslant a\}$ se puede evaluar también por debajo. En efecto

$$\begin{aligned} Mf(\xi) &= \int_\Omega f(\xi(\omega))P(d\omega) \\ &= \int_{\{\omega:|\xi(\omega)|<a\}} f(\xi(\omega))P(d\omega) + \int_{\{\omega:|\xi(\omega)|\geqslant a\}} f(\xi(\omega))P(d\omega) \\ &\leqslant f(a)P\{|\xi(\omega)|<a\} + KP\{|\xi(\omega)|\geqslant a\} \leqslant f(a) + KP\{|\xi(\omega)| \geqslant a\} \end{aligned}$$

y en consecuencia

$$P\{|\xi(\omega)| \geqslant a\} \geqslant \frac{Mf(\xi) - f(a)}{K} . \tag{7}$$

Si en lugar del acotamento de la función exigiéramos el acotamiento de la variable aleatoria $\xi = \xi(\omega)$

$$|f(\omega)| \leqslant M .$$

Entonces $f(\xi(\omega)) \leqslant f(M)$, y en lugar de (7) obtenemos la fórmula

$$P\{|\xi(\omega)| \geqslant a\} \geqslant \frac{Mf(\xi)-f(a)}{f(M)}\,. \tag{8}$$

En el caso cuando $f(x) = x^2$ de (8) encontramos como

$$\{|\xi\ (\omega)|\ \geqslant a\} \geqslant \frac{M\xi^2-a^2}{M^2} \tag{9}$$

*

§4. ALGUNOS CRITERIOS DE CONVERGENCIA

Sea

$$\xi_1, \xi_2, \ldots, \xi_n, \ldots \tag{1}$$

una sucesión de variables aleatorias y $f = f(x)$ una función por no negativa y monótona creciente para todo $x > 0$.[1] Entonces tienen lugar las siguientes afirmaciones:

1. Para la convergencia (1) por probabilidad es suficiente las siguientes condiciones:

Para todo $\varepsilon > 0$ existe un n tal que para cualquier $p > 0$ se cumple

$$Mf(\xi_{n+p} - \xi_n) < \xi. \tag{2}$$

2. Para la convergencia por probabilidad de la sucesión (1) a la variable aleatoria ξ es suficiente la condición

$$\lim_{n\to+\infty} Mf(\xi_n - \xi) = 0. \tag{3}$$

(1) En consecuencia $f(x) > 0$, si $x \neq 0$.

3. Si $f(x)$ es acotada y contínua $f(0) = 0$, entonces las condiciones 1, 2, son además necesarias.

4. Si $f(x)$ es continua $f(0) = 0$ y todas las variables aleatorias $\xi, \xi_1, \xi_2, \ldots$ son acotadas en conjunto entonces las condiciones 1, 2, son además necesarias.

De 2 y 4 se desprende que

5. Pasa la convergencia para la probabilidad de la sucesión (1) así a ξ es suficiente la condición

$$\lim_{n\to\infty} (\xi_n - \xi)^2 = 0 \tag{4}$$

Si además $\xi, \xi_1, \xi_2, \ldots$ son acotados en conjunto, entonces esta condición es también necesaria.

Las demostraciones de las afirmaciones 1, 4, se pueden ver en los trabajos de Slutsky [1] y Fréchet [1]. Sin embargo estos teoremas se desprenden directamente de las fórmulas (3) y (8) del parágrafo anterior.

*

§5. DIFERENCIACION E INTEGRACION DE LAS ESPERANZAS MATEMATICAS POR PARAMETROS.

Supongamos que a cada suceso elemental (ω) se le pone en correspondencia una función real definida $\xi_t(\omega)$ de la variable real t.

DEFINICION. *Diremos que $\xi = \{\xi_t(\omega)\}$, en donde $-\infty < t < \infty$ es una función aleatoria, si para todo valor de t fijo la variable $\xi_t(\omega)$ es una variable aleatoria.*

Surge entonces la pregunta en que condiciones el símbolo de la esperanza matemática se puede trasladar con los símbolos de integración y diferenciación.

Los siguientes dos teoremas pueden, sin agotar todos los problemas, dar para muchos casos sencillos una respuesta satisfactoria a esta pregunta.

TEOREMA 1. *Si para cualquier t la esperanza matemática $M\xi_t(\omega)$ es finita, $\xi_t(\omega)$ es diferenciable (por t para todo (ω)) y la derivada*

$$\xi_t'(\omega) = \frac{d\xi_t(\omega)}{dt}$$

El valor absoluto siempre es menor que una constante, entonces

$$\frac{d}{dt} M\xi_t(\omega) = M\xi_t'(\omega).$$

TEOREMA 2. *Si el valor absoluto de $\xi_t(\omega)$ siempre es menor que una constante K e integrable por t en el sentido de Riemann, entonces*

$$\int_a^b M\xi_t(\omega)\,dt = M\left[\int_a^b \xi_t(\omega)\,dt\right].$$

Si solo $M\xi_t(\omega)$ es integrable en el sentido de Riemann.

Demostración del Teorema 1. Inicialmente anotaremos que $\xi_t'(\omega)$ como límite de las variables aleatorias

$$\frac{\xi_{t+h}(\omega)-\xi_t(\omega)}{h}, \qquad h = 1, \frac{1}{2}, \ldots, \frac{1}{n}, \ldots$$

constituye una variable aleatoria. Ya que $\xi_t'(\omega)$ es acotada entonces existe la esperanza matemática $M\xi_t'(\omega)$ (propiedad 7 de las esperanzas matemáticas §2).

Escogeremos ahora una constante t y denotaremos por A al suceso

$$\left\{\omega : \left| \frac{\xi_{t+h}(\omega)-\xi_t(\omega)}{h} - \xi'_t(\omega) \right| > \varepsilon\right\}$$

La probabilidad $P(A)$ tiende a cero cuando $h \to 0$ para todo $\varepsilon > 0$ ya que siempre

$$\left| \frac{\xi_{t+h}(\omega) - \xi_t(\omega)}{h} \right| \leq M; \quad |\xi'_t(\omega)| \leq M$$

y además para $\omega \in \bar{A}$

$$\left| \frac{\xi_{t+h}(\omega) - \xi_t(\omega)}{h} - \xi'_t(\omega) \right| \leq \varepsilon$$

entonces

$$\left| \frac{M\xi_{t+h}(\omega)-M\xi_t(\omega)}{h} - M\xi'_t(\omega) \right|$$

$$\leq M\left|\frac{\xi_{t+h}(\omega)-\xi_t(\omega)}{h} - \xi'_t(\omega)\right|$$

$$= P(A)M\left\{\left|\frac{\xi_{t+h}(\omega)-\xi_t(\omega)}{h} - \xi'_t(\omega)\right| \Big\| A\right\}$$

$$+ P(\bar{A})M\left\{\left|\frac{\xi_{t+h}(\omega)-\xi_t(\omega)}{h} - \xi'_t(\omega)\right| \Big\| \bar{A}\right\}$$

$$\leq 2MP(A) + \varepsilon$$

ya que es posible escoger cualquiera $\varepsilon > 0$ y la probabilidad $P(A)$ es tan pequeña como se quiera para un h suficientemente pequeño, entonces

$$\frac{d}{dt} M\xi_t(\omega) = \lim_{h\to 0} \frac{M\xi_{t+h}(\omega)-M\xi_t(\omega)}{h} = M\xi'_t(\omega)$$

que era lo que quería demostrar.

Demostración del Teorema 2. Sea

$$S_n = \frac{1}{n}\sum_{K=1}^{n} \xi_{t+Kh}(\omega), \quad h = \frac{b-a}{n}$$

Ya que S_n converge a $S = \int_a^b \xi_t(\omega)dt$ entonces para todo $\varepsilon > 0$ se puede escoger un N tal que para todo $n \geqslant N$

$$P\{|S_n - S| > \varepsilon\} < \varepsilon .$$

Si denotamos $A = \{|S_n - S| > \varepsilon\}$ y

$$S_n^* = \frac{1}{n}\sum_{k=1}^{n} M\xi_{t+kh}(\omega) = MS_n$$

Entonces

$$|S_n^* - MS| = |M(S_n - S)| \leqslant M|S_n - S|$$

$$= P(A)M\{|S_n - S|/A\} + P(\bar{A})M\{|S_n - S|/\bar{A}\}$$

$$\leqslant 2KP(A) + \varepsilon \leqslant (2K+1)\varepsilon .$$

En consecuencia S_n^* tiende a MS de donde se desprende la igualdad

$$\int_a^b M\xi_t(\omega)dt = \lim_n S_n^* = MS.$$

Al Teorema 2 puede ser sin ninguna nueva dificultad generalizado para integrales dobles y triples.

Daremos una aplicación de este teorema para un problema geométrico de la teoría de las probabilidades.

Sea $G = G(x)$ una región cuadrable en un plano y supongamos que cada suceso elemental $\omega \in \Omega$ se pone en correspondencia una determinada región cuadrable $G(\omega)$ del plano. Por S_n denotaremos el área de la región G, y por $P(x,y)$ denotaremos la probabilidad de que el punto (x,y) pertenezca a la región G. Entonces $MS_G = \iint P(x,y)\,dx\,dy$. Para la demostración es suficiente anotar que

$$S_G = \iint \delta_G(x,y)\,dx\,dy$$

$$P(x,y) = M\delta_G(x,y).$$

En donde $\delta_G(x,y)$ es la función característica de la región G $[\delta_G(x,y) = 1$ en G y $\delta_G(x,y) = 0$ fuera de $G]$[1].

*

(1) Comp. A.Kolmogórov und N. Leontowitsch, Zur Berechnung der mittlerer Brownschen Fläche, Physik. Zeitschr. d. Sonjetunion, V.4 (1933).

CAPITULO V

PROBABILIDAD CONDICIONAL
Y
ESPERANZA MATEMATICA

§1. PROBABILIDAD CONDICIONAL.

En el capítulo I, §6 definimos la probabilidad condicional $P(B|U)$ del suceso B con respecto a las pruebas U. En este caso se suponía que U tenía solo un número finito de posibles resultados diferentes. Podemos sin embargo definir $P(B|U)$ también para el caso cuando las pruebas U tiene un conjunto infinito de posibles resultados o sea para el desarrollo de conjunto en un número infinito de subconjuntos disyuntos. En particular este desarrollo se obtiene si se examina cualquier función $\xi = \xi(\omega)$ de omega (ω) y determinan en calidad de elementos de desarrollo $U = U_\xi$ los conjuntos $\{\omega : \xi(\omega) = \text{const}\}$.

La probabilidad condicional $P(B|U_\xi)$ se denotará también por $P(B|\xi)$. Cualquier desarrollo U del conjunto Ω puede definirse como el desarrollo U_ξ al cual se le induce la función $\xi = \xi(\omega)$ si acaba ω se pone en correspondencia en calidad de $\xi(\omega)$ entonces el conjunto de desarrollos U que contiene a ω.

Dos funciones ξ y η de ω definen una y solo un mismo desarrollo $(U_\xi = U_\eta)$ del conjunto Ω si existe una correspondencia biunívoca $y = f(x)$ entre sus valores tal que $\eta(\omega) = f(\xi(\omega))$. El lector puede facilmente mostrar que las magnitudes aleatorias bajo las definiciones $P(B|\xi)$ y $P(B|\eta)$ en este caso coinciden consecuencia en lo fundamental, ellas se definen con un mismo desarrollo ($U_\xi = U_\eta$).

Para la definición $P(B|\xi)$, podemos tomar la siguiente igualdad

$$P(B \mid \xi \in A) = M[P(B \mid \xi) \mid \xi \in A]. \qquad (1)$$

Se puede mostrar facilmente que en el caso de conjuntos finitos X de posibles valores de ξ la igualdad (1) se cumple para cualquier conjunto A (además $P(B|\xi)$ se define conforme al parágrafo 6 del Cap.I). En el caso general para cuando la variable aleatoria $P(B|\xi)$ aún no está definida) demostraremos que existe una y solo una variable aleatoria $P(B|\xi)$, y la existencia de otra supone la equivalencia de esta con la primera variable aleatoria. La variable aleatoria $P(B|\xi)$ se define como una función (de Borel) de ξ que cumple para cada A de F_ξ y $P_\xi(A) > 0$ la ecuación (1).

La función $P(B|\xi)$ de ξ definida de ésta manera la llamaremos Probabilidad condicional de B respecto a ξ (o para un ξ dado). Los valores de $P(B|\xi)$ cuando $\xi = x$ lo denotaremos $P(B|\xi = x)$.

DEMOSTRACION DE LA EXISTENCIA Y UNICIDAD DE $P(B|\xi)$.

Multiplicando (1) por $P\{\xi \in A\} = P_\xi(A)$ obtenemos a la izquierda

$$P\{\xi \in A\}P(B|\xi \in A) = P(B \cap \{\xi \in A\}) = P(B \cap \xi^{-1}(A))$$

y a la derecha

$$P\{\xi \in A\}M[P(B|\xi)/\xi \in A] = \int_{\{\omega:\xi \in A\}} P(B|\xi)P(d\omega) = \int_A P(B|\xi = x)P_\xi(dx)$$

En consecuencia

$$P(B\cap\xi^{-1}(A)) = \int_A P(B|\xi = x)P_\xi(dx) \qquad (2)$$

Lo contrario de (2) se deduce la fórmula (1).

En el caso de que $P_\xi(A) = 0$ para el cual (1) no tiene sentido la igualdad (2) es trivial. Así de la exigencia (2) es equivalente a (1).

Según la prioridad 9 (Cap. IV, §1) las magnitudes aleatorias $\eta = \eta(\omega)$ unívocamente se definen por los valores de las integrales.

Ya que $(B|\xi = x)$ es una variable aleatoria definida en el campo probabilístico (X,F_ξ,P_ξ) entonces, de aquí se deduce que la fórmula (2) define esta variable.

Nos queda por demostrar la existencia de $P(B|\xi)$. Para esta aplicaremos el siguiente teorema de Radon-Nikodyn(1).

TEOREMA. *Sea $\mathfrak{F}$ en álgebra de Borel de conjuntos en Y, P una función de conjuntos no negativos finita-aditiva y definida en $(Y,\mathfrak{F})$, $\tilde{P}$ una segunda función de conjuntos también definida en $(Y,\mathfrak{F})$ y finita-aditiva y además de $P(A) = 0$ se deduce que $\tilde{P}(A) = 0$. Entonces existe una función medible (con respecto a $\mathfrak{F}$) $\phi = \phi(y)$ (en la terminología teórico-probabilístico una variable aleatoria) que cumple que para cada con junto A de $\mathfrak{F}$ la igualdad*

$$\tilde{P}(A) = \int_A \phi(y)P(dy).$$

---

(1) O.Nikodym, Sur une généralisation des integrales de M.J Radon, Fund. Math. V.15 (1930), 168 (théoreme III). Ver también Kolmogórov y Fomín.

Para la utilización de nuestro teorema basta con demostrar que

1. $\tilde{P}(A) = P(B \cap \xi^{-1}(A))$ es finita-aditiva en (X, F_ξ);

2. De $P(A) \neq 0$ se deduce que $P_\xi(A) > 0$.

La afirmación **2.** se desprende por el siguiente hecho

$$0 \leqslant P(B \cap \xi^{-1}(A)) \leqslant P(\xi^{-1}(A)) = P_\xi(A).$$

Para la demostración de la afirmación **1.** pongamos $A = \sum_n A_n$. Entonces

$$\xi^{-1}(A) = \sum_n \xi^{-1}(A_n)$$

y

$$B \cap \xi^{-1}(A) = \sum_n B \cap \xi^{-1}(A_n).$$

Ya que P es finita aditiva entonces

$$P(B \cap \xi^{-1}(A)) = \sum_n P(B \cap \xi^{-1}(A_n))$$

que era lo que se quería demostrar.

De la igualdad (1) se desprende en particular (suponiendo $A = X$). La importante fórmula

$$P(B) = M[P(B|\xi)]. \tag{3}$$

Ahora pasaremos a la demostración de dos propiedades fundamentales de la probabilidad fundamental.

TEOREMA 1. *Casi siempre* $0 \leqslant P(B|\xi) \leqslant 1.$ (4)

TEOREMA 2. *Si los sucesos* $B, B_1, \ldots$ *pertenece a* $\mathcal{F}$

$$B = \sum_n B_n$$

Entonces, casi siempre

$$P(B|\xi) = \sum_n P(B_n \mid \xi) \tag{5}$$

Estas dos propiedades de $P(B|\xi)$ corresponden a dos propiedades características de la función probabilística $P(B)$; siempre $0 \leqslant P(B) \leqslant 1$ y $P(B)$ es finita aditiva, estas permiten trasladar a las probabilidades condicionales $P(\cdot|\xi)$ muchas propiedades esenciales de las probabilidades absolutas $P(\cdot)$.

Sin embargo es bueno recordar $P(B|\xi)$ para un conjunto fijo B es una magnitud que se define solo como una exactitud de equivalencia.

Demostración Teorema 1. Supondremos en contraposición a la afirmación a demostrar que en el conjunto M con $P_\xi(M) > 0$ se cumple la desigualdad $P(B|\xi) \geqslant 1+\varepsilon$, entonces por la fórmula (1):

$$P(B|\xi \in M) = M[P(B|\xi) \mid \xi \in M] \geqslant 1+\varepsilon,$$

lo cual es claramente imposible. Igualmente se demuestra que casi siempre $P(B|\xi) \geqslant 0$.

Demostración Teorema 2. De la convergencia de serie

$$\sum_n M| P(B_n|\xi) | = \sum_n M[P(B_n|\xi)] = \sum_n P(B_n) = P(B)$$

Por la probabilidad (5) de las esperanzas matemáticas (cap. IV parágrafo 2) se deduce que la serie $\sum_n P(B_n|\xi)$ converge casi siempre. Ya que la serie

$$\sum_n M[P(B_n|\xi)/\xi \in A] = \sum_n P(B_n|\xi \in A) = P(B|\xi \in A)$$

Converge para cualquier A tal que $P_\xi(A) > 0$ entonces según la probabilidad 5 de las esperanzas matemáticas se deduce que para todo conjunto A del tipo mostrado tiene lugar la relación

$$M[\sum_n P(B_n|\xi) \in A] = \sum_n M[P(B_n|\xi)/\xi \in A] = P(B|\xi \in A)$$

$$= M[P(B|\xi)/\xi \in A].$$

De la cual directamente se desprende la igualdad (5).

Finalmente anotaremos dos casos particulares. Sea $\xi(\omega) \equiv C$-const, entonces casi siempre $P(A|C) = P(A)$. Si ponemos $\xi(\omega) = \omega$ entonces obtenemos inmediatamente que $P(A|\omega)$ casi siempre es igual a 1 en A e igual a cero en $\bar{A}$ en consecuencia $P(A|\omega)$ es una función característica del conjunto A.

*

§2. EXPLICACION DE LA PARADOJA DE BOREL.

En calidad del conjunto fundamental Ω escogemos el conjunto de todos los puntos de una superficie esférica. Por F tomaremos el conjunto de todos los conjuntos de Borel de superficie esférica Finalmente sea $P(A)$ una medida proporcional del conjunto A. Escogemos ahora dos puntos diametralmente opuestos en calidad de polos. Entonces cada meridiano circular unívocamente se define por su correspondiente longitud geográfica $\psi (0 \leqslant \psi < \pi)$. Ya que ψ sea mide solo de 0 hasta π, o sea examinaremos meridianos circulares completos (y no semimeridianos) entonces la amplitud θ deberá medirse de $-\pi$ a π (y no de $-\frac{\pi}{2}$ a $\frac{\pi}{2}$). Borel propuso el siguiente problema: definir la probabilidad condicional para la amplitud θ, $-\pi \leqslant \theta < \pi$ para una longitud dada ψ es fácil calcular que

$$P(\theta_1 \leqslant \theta < \theta_2|\psi) = \frac{1}{4}\int_{\theta_1}^{\theta_2} |\cos\theta|\, d\theta$$

lo que quiere decir que la distribución probabilística para θ dado en ψ no es uniforme.

Si ahora suponemos que la distribución condicional de la probabilidad para θ según la hipótesis que ω está en el meridiano circular, debe ser uniforme entonces obtenemos una contradicción.

Este hecho muestra que el concepto de probabilidad condicional respecto a una hipótesis aislada cuya probabilidad es igual a cero es inadmisible: solo entonces nosotros obtenemos en el meridiano circular una distribución probabilística para θ, si nosotros examinamos este meridiano circular en calidad de elemento del desarrollo de toda la superficie específica en los meridianos circulares con los polos dados.

*

§3. PROBABILIDAD CONDICIONAL RESPECTO A UNA VARIABLE ALEATORIA

Si $\xi = \xi(\omega)$ es una variable aleatoria entonces la probabilidad condicional $P(B|\xi)$ se puede definir también de una forma elemental. En el caso de que $P(B) = 0$ pondremos $P(B|\xi) = 0$ sea ahora $P(B) > 0$. Entonces a la fórmula (2) del §1 se le puede dar la siguiente fórmula

$$P(B)P(\xi \in A|B) = \int_A P(B|\xi = x)\; P_\xi(dx) \tag{1}$$

ó

$$P(B)P_{\xi}(A|B) = \int_{A} P(B|\xi = x)P_{\xi}(dx).$$

De aquí se deduce directamente que

$$P(B)P_{\xi}(a|B) = \int_{-\infty}^{a} P(B|\xi = x)F_{\xi}(dx) \tag{2}$$

De acuerdo con un teorema de Lebesgue(1) de (2) se deduce que

$$P(B|\xi = x) = P(B)\lim_{h\to 0} \frac{F_{\xi}(x+h\ B)-F_{\xi}(x\ B)}{F_{\xi}(x+h)-F_{\xi}(h)} \tag{3}$$

con exactitud en el conjunto H de puntos x tales que $P_{\xi}(H) = 0$. Si examinamos ahora la fórmula (3) como definición de $P(B|\xi = x)$ y en caso de no existencia del límite en la parte derecha de (3) colocar $P(B|\xi = x) = 0$, entonces la nueva variable cumple todas las exigencias del parágrafo §1. Si además, existen las densidades probabilísticas $f_{\xi}(x)$, $f_{\xi}(x|B)$ y si $f_{\xi}(x) > 0$ entonces la fórmula (3) se convierte en lo siguiente:

$$P(B|\xi = x) = P(B) \cdot \frac{f_{\xi}(x|B)}{f_{\xi}(x)} \tag{4}$$

De la fórmula (3) igualmente se deduce que es existencia del límite (3) y de la densidad probabilística $f_{\xi}(x)$ tienen como consecuencia la existencia de $f_{\xi}(x|B)$ y además

$$P(B)f_{\xi}(x|B) \leqslant f_{\xi}(x). \tag{5}$$

Si $P(B) > 0$ entonces de (4) se deduce

$$f_{\xi}(x|B) = \frac{P(B|\xi = x)f_{\xi}(x)}{P(B)} \tag{6}$$

---

(1) Lebesgue "Integración..."

En el caso $f_\xi(x) = 0$ de acuerdo con (5), $f_\xi(x|B) = 0$ y en consecuencia (6) también es cierto y si además la distribución ξ es absolutamente continua, entonces

$$\begin{aligned} P(B) &= M[P(B|\xi)] = \int_\Omega P(B|\xi)P(d\omega) \\ &= \int_{-\infty}^{\infty} P(B|\xi = x)F_\xi(dx) = \int_{-\infty}^{\infty} P(B|\xi = x)f_\xi(x)\,dx. \end{aligned} \tag{7}$$

De (6) y (7) se deduce que

$$f_\xi(x|B) = \frac{P(B|\xi = x)f_\xi(x)}{\int_{-\infty}^{\infty} P(B|\xi = x)f_\xi(x)\,dx} \tag{8}$$

Esta igualdad nos da el así llamado teorema de Bayes para las distribuciones absolutamente continuas.

Las suposiciones para las cuales este teorema es cierto son las siguientes: $P(B|\xi = x)$ es medible en el sentido de Borel y definida según la fórmula (3), la distribución ξ es absolutamente continua (en el punto si existe la densidad probabilística $f_\xi(x)$).

*

§4. ESPERANZA MATEMATICA CONDICIONAL.

Sea $\xi = \xi(\omega)$ una función cualquiera de ω, y $\eta = \eta(\omega)$ una variable aleatoria. La variable aleatoria $M(\eta|\xi)$, representadas como función de ξ que cumple para todo conjunto A de $\mathfrak{F}_\xi$ con $P_\xi(A) > 0$ la condición

$$M(\eta|\xi \in A) = M[M(\eta|\xi)/\xi \in A]. \tag{1}$$

Se llama (en caso de que ella exista) esperanza matemática condicional de la variable η dado un valor conocido ξ.

Multiplicando (1) por $P_\xi(A)$ obtenemos

$$\int_{\{\xi\in A\}} \eta(\omega)P(d\omega) = \int_{\{\xi\in A\}} M(\eta|\xi)P(d\omega). \qquad (2)$$

Lo contrario de (2) se deduce la fórmula (1). En caso de $P_\xi(A) = 0$ para el cual (1) no tiene sentido, (2) es trivial de la misma forma que para la probabilidad condicional (ver §1 se demuestra que $M(\eta|\xi)$ se define por (2) unívocamente (con exactitud de equivalencia).

Los valores de $M(\eta|\xi)$ para $\xi(\omega) = x$ los denotaremos por $M(\eta|\xi = x)$. Anotaremos también $M(\eta|\xi)$ igual que $P(B|\xi)$ depende solo del desarrollo $\mathfrak{U}_\xi$ y puede ser denotado por $M(\eta|\mathfrak{U}_\xi)$.

En la definición de $M(\eta|\xi)$ se supone la existencia $M\eta$ (si ponemos $A = X$ entonces $M(\eta|\xi \in A) = M\eta$). Denotaremos que la existencia de $M\eta$ es suficiente para la existencia de $M(\eta|\xi)$ para ello es suficiente demostrar que según el teorema de Rador-Nikodyn (ver parágrafo §1):

la función de conjunto $Q(A) = \int_{\{\xi\in A\}} \eta(\omega)P(d\omega)$ es finita-adtiviva en $\mathfrak{F}_\xi$ y absolutamente continua con respecto $P_\xi(A)$. El primer hecho se demuestra exactamente como en el caso de las probabilidades condicionales (ver §1). La segunda condición-continuidad absoluta-consiste en que de $Q(A) \neq 0$ debe deducirse la desigualdad $P_\xi(A) > 0$. Si suponemos que $P_\xi(A) = P(\xi \in A) = 0$, entonces es claro que

$$Q(A) = \int_{\{\xi\in A\}} \eta(\omega)P(d\omega) = 0 .$$

De esta forma nuestra segunda condición también se cumple.

Si en la igualdad (1) poner $A = X$, entonces obtenemos la fórmula

$$M\eta = M[M(\eta|\xi)]. \tag{3}$$

Se demuestra así que casi siempre

$$M[a\eta + b\zeta|\xi] = aM(\eta|\xi) + bM(\zeta|\xi) \tag{4}$$

donde a y b son dos constantes, $M|\eta| < \infty$, $M|\zeta| < \infty$. (La demostración de este hecho se propone al lector).

Si ξ y ζ son dos funciones de los sucesos elementales ω, entonces el par (ξ,ζ), puede también examinarse como una función de ω, y tiene lugar la siguiente importante igualdad:

$$M[M(\eta|\xi,\zeta)/\xi] = M(\eta,\xi) \tag{5}$$

En efecto $M(\eta,\xi)$ se define a través de la relación

$$M(\eta|\xi \in A) = M[M(\eta|\xi)/\xi \in A].$$

En consecuencia, es necesario demostrar que la variable $M[M(\eta|\xi,\zeta)/\xi]$ cumple la igualdad

$$M(\eta|\xi \in A) = M\{M[M(\eta|\xi,\zeta)/\xi]/\xi \in A\}. \tag{6}$$

De la definición $M(\eta|\xi,\zeta)$ se deduce que

$$M(\eta|\xi \in A) = M[M(\eta|\xi,\zeta)/\xi \in A]. \tag{7}$$

De la definición $M|M(\eta|\xi,\zeta)/\xi|$ se deduce que

$$M[M(\eta|\xi,\zeta)/\xi \in A] = M\{M[M(\eta|\xi,\zeta)/\xi]/\xi \in A\}. \tag{8}$$

La igualdad (6) es una consecuencia de las igualdades (7) y (8) y con esto se obtiene lo que queríamos demostrar.

Si colocamos $\eta(\omega)$ igual a 1 en B e igual a cero fuera de B entonces

$$M(\eta|\xi) = P(B|\xi),$$

$$M(\eta|\xi,\zeta) = P(B|\xi,\zeta).$$

En este caso de la fórmula (5) obtenemos la fórmula

$$M[(P(B|\xi,\zeta)|\xi] = P(B|\xi). \tag{9}$$

La esperanza matemática condicional se puede definir también directamente a través de su correspondiente probabilidad condicional.

Para esto es necesario examinar la siguiente suma

$$S_\lambda(\xi) = \sum_{k=-\infty}^{\infty} k\lambda P[k\lambda \leqslant \eta < (k+1)\lambda|\xi]\left(=\sum_{k=-\infty}^{\infty} R_k\right) \tag{10}$$

Si $M(\eta)$ existe entonces la serie (10) converge casi siempre en efecto según la fórmula (3), §1.

$$M|R_k| = k\lambda P\{k\lambda \leqslant \eta < (k+1)\lambda\}$$

y la convergencia de la serie

$$\sum_{k=-\infty}^{\infty} |k\lambda| P\{k\lambda \leqslant \eta < (k+1)\lambda\} = \sum_{k=-\infty}^{\infty} M|R_k|$$

Es una condición necesaria para la existencia de $M\eta$ (ver Cap.IV, §1), de esta convergencia se deduce que la serie (10) converge casi siempre (Cap.IV §2).

Seguidamente demostraremos como la teoría de la integral de Lebesgue que de la convergencia de (10) para un λ se deduce la convergencia para cualquier λ y que en el caso de la con vergencia de la serie (10) $S_\lambda(\xi)$ tiende a un determinado límite cuando $\lambda \to 0$[1]. Podemos entonces en calidad de definición escribir

$$M(\eta|\xi) = \lim_{\lambda\to o} S_\lambda(\xi) \tag{11}$$

Para demostrar que la relación definida en (11) de la esperanza condicional $M(\eta|\xi)$ cumple la condición establecida con anterioridad es necesario solo comprobar que la variable $M(\eta|\xi)$ definida en (11) cumple la igualdad (1). Esta demostración se realiza de la siguiente manera

$$M[M(\eta|\xi)/\xi \in A] = \lim_{\lambda\to o} M[S_\lambda(\xi)/\xi \in A]$$
$$= \lim_{\lambda\to o} \sum_{k=-\infty}^{\infty} k\lambda P[k\lambda \leqslant \eta < (k+1)\lambda|\xi \in A] = M[\eta|\xi \in A]$$

El cambio de signos de la esperanza matemática y el limit es permisible en estos cálculos ya que $S_\lambda(\xi)$ converge uniformemente a $M(\eta|\xi)$ cuando $\lambda \to o$ (una consecuencia sencilla de las esperanzas matemáticas §2). El cambio de los signos de las experiencias matemáticas y la sumatoria también es justificada ya que

$$\sum_{k=-\infty}^{\infty} M\{|k\lambda|P(k\lambda \leqslant \eta < (k+1)\lambda/\xi)/\xi \in A\}$$
$$= \sum_{k=-\infty}^{\infty} |k\lambda|P\{k\lambda \leqslant \eta < (k+1)\lambda/\xi \in A\}.$$

Es convergente (aplicación directa de la prop.5 de las esperanzas matemáticas).

---

(1) Aquí examinamos solo la sucesión contable de valores λ; en este caso todas las probabilidades $P\{k\lambda \leqslant \eta < (k+1)\lambda/\xi\}$ están definidas casi siempre para todas estos valores.

En lugar de (11) podemos escribir

$$M(\eta|\xi) = \int_{\Omega} \eta(\omega) P(d\omega/\xi) \tag{12}$$

No debemos sin embargo olvidar que (12) es una integral en el sentido de parágrafo 1 en el Cap.IV ya que (12) es solo una notación simbólica.

Si ξ es una variable aleatoria, entonces la función de ξ y y

$$F\eta(y|\xi) = P(\eta < y|\xi).$$

Nosotros la llamaremos función de distribución condicional de η para ξ conocido. La función $F_\eta(y|\xi)$ definida casi siempre para cada y. Si $y_1 \leqslant y_2$ entonces casi siempre

$$F_\eta(y_1|\xi) \leqslant F_\eta(y_2|\xi).$$

De (11) y (10) se deduce que casi siempre

$$M(\eta|\xi) = \lim_{\lambda\to o} \sum_{k=-\infty}^{\infty} k\lambda[F_\eta((k+1)\lambda/\xi) - F_\eta(k\lambda/\xi)]. \tag{13}$$

Este hecho lo podemos simbólicamente expresar por la fórmula

$$M(\eta|\xi) = \int_{-\infty}^{\infty} yF_\eta(dy|\xi). \tag{14}$$

Con la ayuda de la nueva definición de la esperanza matemática (10) y (11) facilmente se demuestra que para la función real $f(\xi)$ con $M|f(\xi)\eta| < \infty$ se cumple (casi siempre) la igualdad

$$M[f(\xi)\eta|\xi] = f(\xi)M[\eta|\xi].$$

*

CAPITULO VI

INDEPENDENCIA. LEY DE LOS GRANDES NUMEROS

§1. INDEPENDENCIA.

DEFINICION 1. *Dos funciones* $\xi = \xi(\omega)$ *y* $\eta = \eta(\omega)$ *se llaman independientes si para todo par de conjuntos* A *de* $\mathfrak{F}_\xi$ *y* B *de* $\mathfrak{F}_\eta$ *se verifica la siguiente igualdad:*

$$P(\xi \in A, \eta \in B) = P(\xi \in A)P(\eta \in B)\,. \tag{1}$$

Si los conjuntos X y Y estan formados solo de un número finito de elementos

$$X = x_1+\ldots+x_n,$$
$$Y = y_1+\ldots+y_n,$$

entonces nuestra definición de independencia de ξ y η coincide, de acuerdo con el §5 del primer capítulo, con la definición de independencia del desarrollo

$$\Omega = \sum_{k=1}^{n} \{\omega : \xi(\omega) = x_k\},$$
$$\Omega = \sum_{k=1}^{m}\{\omega: \eta(\omega) = y_k\}.$$

Para la independencia de ξ y η es necesario y suficiente el cumplimiento de la siguiente condición: Para cualquier conjunto A escogido de $\mathfrak{F}_\xi$ casi siempre se cumple la igualdad

$$P(\xi \in A|\eta) = P(\xi \in A). \tag{2}$$

En efecto, en el caso cuando $P_\eta(B) = 0$, ambas igualdades (1) y (2) se cumplen. En consecuencia es suficiente demostrar su equivalencia en el caso $P_\eta(B) > 0$. En este caso (1) es equivalente a la relación

$$P(\xi \in A|\eta \in B) = P(\xi \in A) \qquad (3)$$

y lo que indica la relación

$$M[P(\xi \in A|\eta)/\eta \in B] = P(\xi \in A). \qquad (4)$$

Por otra parte, es claro que de (2) se deduce la igualdad (4). Lo contrario, ya que $(\xi \in A|\eta)$ unívocamente se define de (4) con precisión hasta de conjuntos con probabilidad cero, entonces de (4) casi siempre se deduce la igualdad (2).

DEFINICION 2. *Sea $\sum$ el conjunto de funciones $\xi_\nu(\omega)$, donde $\nu \in N$. Estas funciones se llaman independientes en conjunto si se cumplen las siguientes condiciones: sean $\sum'$ y $\sum''$ dos subconjuntos de $\sum$ con intercepción vacía, A' (A'' respectivamente) un conjunto de $\mathcal{F}$ definido con la relación entre $\xi_\nu(\omega)$ de $\sum'$ ($\sum''$ respectivamente); entonces*

$$P(A' \cap A'') = P(A')P(A'').$$

El conjunto de todas las $\xi_\nu(\omega)$ de $\sum'$ (de $\sum''$ respectivamente) se puede examinar como las coordenadas de alguna función ξ' (de ξ'' respectivamente). La definición 2 exige solo la independencia de ξ'' y ξ' en el sentido de la definición 1, para cualquier escogencia de los conjuntos con intercepción vacía $\sum'$ y $\sum''$.

Si $\xi_1, \xi_2, \ldots, \xi_n$ son independientes, entonces siempre

$$P(\xi_1 \in A_1, \xi_2 \in A_2, \ldots, \xi_n \in \xi_n) = P(\xi_1 \in A_1)P(\xi_2 \in A_2)\ldots P(\xi_n \in A_n), \qquad (5)$$

si cada conjunto A_k pertenece a su respectiva $\mathcal{F}_{\xi_k}$ (la demostración mediante inducción). Esta igualdad sin embargo en el caso general de ninguna manera es suficiente para la independencia de $\xi_1, \xi_2, \ldots, \xi_n$.

La igualdad (5) sin dificultades se generaliza al caso de producto contable.

De la independencia de las variables aleatorias ξ_{ν_n} en cada grupo finito $(\xi_{\nu_1}, \xi_{\nu_2}, \ldots, \xi_{\nu_n})$ en general no se deduce la independencia de todos los ξ_ν.

Por último, es fácil observar que la independencia de las funciones ξ_ν es una propiedad de los correspondientes desarrollos $\mathcal{U}_{\xi_\nu}$. Si además ξ'_ν es una función unívocamente definida por los correspondientes ξ_ν, entonces de la independencia de ξ_ν se deduce la independencia de ξ'_ν.

*

§2. INDEPENDENCIA DE LAS VARIABLES ALEATORIAS.

Si $\xi_1, \xi_2, \ldots, \xi_n$ son variables aleatorias independientes entonces de la igualdad (2) del anterior parágrafo se deduce en particular la fórmula

$$F_{\xi_1, \xi_2, \ldots, \xi_n}(x_1, \ldots, x_n) = F_{\xi_1}(x_1)\ F_{\xi_2}(x_2) \ldots F_{\xi_n}(x_n) \qquad (1)$$

TEOREMA 1. *Si el álgebra de conjuntos $\mathcal{F}_{\xi_1, \xi_2, \ldots, \xi_n}$ esta formado solo de conjuntos de Borel del espacio $\mathbf{R}^n$, entonces la condición (1) es también suficiente para la independencia de las variables $\xi_1, \xi_2, \ldots, \xi_n$.*

Demostración. Sea $\xi' = (\xi_{i_1},\xi_{i_2},\ldots,\xi_{i_k})$ y $\xi'' = (\xi_{j_1},\xi_{j_2},\ldots,\xi_{j_m})$ dos subsistemas disyuntos de las variables $\xi_1,\xi_2,\ldots,\xi_n$. En base a la fórmula (1) se sigue demostrar que para cualquier par de conjuntos de Borel A' y A'' respectivamente de $\mathbf{R}^k$ y $\mathbf{R}^m$ se cumple la igualdad

$$P(\xi' \in A',\ \xi'' \in A'') = P(\xi' \in A')P(\xi'' \in A'') \qquad (2)$$

Para los conjuntos de la forma

$$A' = \{(x_{i_1},\ldots,x_{i_n}) : x_{i_1} < a_1,\ldots,x_{i_k} < a_k\}$$

$$A'' = \{(x_{j_1},\ldots,x_{j_m}) : x_{j_1} < b_1,\ldots,x_{j_m} < b_m\}$$

esto se deduce inmediatamente de (1). Luego se demuestra que esta propiedad se conserva para la suma y la diferencia de conjuntos de la forma señalada, de donde la igualdad (2) se deduce para todos los conjuntos de Borel.

TEOREMA 2. *Sea $\xi = \{\xi_\nu\}$ un conjunto cualquiera (en general infinito) de variables aleatorias. Si el álgebra de conjuntos $\mathfrak{F}_\xi$ coincide con la σ-álgebra $\sigma(\mathfrak{F}^N)$ (N es el conjunto de todos los ν)*(1) *entonces el conjunto de igualdades*

$$F_{\xi_{\nu_1},\ldots,\xi_{\nu_n}}(x_1,\ldots,x_n) = F_{\xi_{\nu_1}}(x_1),\ldots,F_{\xi_{\nu_n}}(x_n) \qquad (3)$$

es necesario y suficiente para la independencia de las variables ξ_ν, $\nu \in N$.

Demostración. La necesidad de esta condición se deduce inmediatamente de la fórmula (1). Demostraremos ahora que ella también es suficiente.

Sean N' y N'' dos subconjuntos disyuntos del conjunto N de

---

(1) Ver §4 Cap.III.

todos los subíndices ν, A' (respectivamente A'') es un conjunto de $\sigma(\mathcal{F}^N)$, definido por la relación entre ξ_ν con los subíndices ν de N' (N'' respectivamente). Sigue demostrar que entonces se cumple la igualdad

$$P(A' \cap A'') = P(A')P(A''). \tag{4}$$

En efecto si A' y A'' son conjuntos cilíndricos, entonces estaremos tratando con relaciones entre un conjunto finito de variables ξ_ν y la igualdad (4) representa en este caso una consecuencia sencilla de las anteriores resultados (fórmula (2)). Lo mismo de la relación (4) se conserva para la suma y diferencia de conjuntos A' (A'' respectivamente), entonces (4) queda demostrado para todos los conjuntos de $\sigma(\mathcal{F}^N)$.

Supongàmos ahora que para cada ν de algún conjunto N esta dada a priori la función de distribución $F_\nu = F_\nu(x)$.

TEOREMA 3. *Existe un campo probabilístico $(\Omega,\mathcal{F},P)$ y definidas en el las variables aleatorias ξ_ν, $\nu \in N$ las cuales son independientes en conjunto y cada una de las variables ξ_ν tiene en calidad de función de distribución a priori la función dada $F_\nu(x)$.*

Demostración. Como conjunto fundamental Ω tomamos el espacio $\mathbf{R}^N$ (conjunto de todos los sistemas $\omega = \{x_\nu\}$ de números reales x_ν, $\nu \in N$) y en calidad de $\mathcal{F}$ tomamos el σ-álgebra $\sigma(\mathcal{F}^N)$ (1). Pongamos además $\xi_\nu(\omega) = x_\nu$ y definamos la función de distribución $F_{\nu_1,\ldots,\nu_n}$ con la igualdad:

$$F_{\nu_1,\ldots,\nu_n}(x_1,\ldots,x_n) = F_{\nu_1}(x_1)\ldots F_{\nu_n}(x_n).$$

---

(1) Ver §4 Cap.III .

Entonces, de acuerdo al teorema fundamental de (§4, Cap. III); en $(\Omega,\mathcal{F})$ unívocamente queda definida la función probabilística P, además $P\{\xi_\nu < x\} = F_\nu(x)$, para todo $\nu \in N$.

Anotemos además que, como se veía arriba, de la independencia de todo grupo finito de variables ξ_ν (igualdad (3)) se concluye la independencia de todas las ξ_ν en $\sigma(\mathcal{F}^N)$. En campos voluntarios probabilísticos esta propiedad puede perderse.

En la conclusión de este parágrafo daremos algunos indicios de independencia para dos variables aleatorias.

Si dos variables aleatorias ξ y η son independientes y si $M\xi$ y $M\eta$ son finitas, entonces casi siempre

$$\begin{aligned} M(\eta|\xi) &= M\eta , \\ M(\xi|\eta) &= M\xi . \end{aligned} \tag{5}$$

Estas fórmulas representan conclusiones inmediatas de la segunda definición de las esperanzas matemáticas condicionales (fórmulas (10) y (11) §4 del Cap.V). En consecuencia, en el caso de la independencia de ξ y η las magnitudes

$$\delta^2 = \frac{M[\eta-M(\eta|\xi)]^2}{D\eta}, \qquad g^2 = \frac{M[\xi-M(\xi|\eta)]^2}{D\xi}$$

son iguales a cero (se supone que $\xi > 0$ y $\eta > 0$). El número δ^2 se llama relación de correlación de η con respecto a ξ, g^2-relación de correlación de ξ con respecto a η. (Pearson).

De (5) se concluye la igualdad

$$M\xi\eta = M\xi \cdot M\eta \tag{6}$$

para cuya demostración es necesario utilizar las fórmulas (15) de §4 del Cap.V.

$$M\xi\eta = M[M(\xi\eta)|\xi] = M[\xi M(\eta|\xi)] = M[\xi\cdot M\eta] = M\xi\cdot M\eta$$

En consecuencia en caso de independencia de ξ y η la magnitud

$$\rho = \frac{M\xi\eta - M\xi M\eta}{\sqrt{D\xi\cdot D\eta}}$$

también es igual a cero. Como es sabido ρ es el coeficiente de correlación entre ξ y η.

Si las dos variables aleatorias ξ y η cumplen la igualdad (6) entonces se llaman nocorrelacionadas. Para la suma

$$S = \xi_1 + \xi_2 +\ldots+ \xi_n$$

de variables nocorrelacionados en parejas $\xi_1,\ldots,\xi_n$ es fácil calcular que

$$DS = D\xi_1 + D\xi_2 +\ldots+ D\xi_n \tag{7}$$

En particular, la igualdad (7) la verifica para las variables independientes $\xi_1,\xi_2,\ldots,\xi_n$.

*

§3. LEY DE LOS GRANDES NUMEROS.

DEFINICION. *Las variables aleatorias* η_n *de la sucesión*

$$\eta_1,\eta_2,\ldots,\eta_n$$

se llama estable si existe una sucesión de números

$$d_1,d_2,\ldots,d_n,$$

tal que para todo cualquier número positivo ε

$$P\{|\eta_n - d_n| \geqslant \varepsilon\} \to 0, \quad n \to \infty .$$

Si existen todos los $M\eta_n$ y si se puede escribir

$$d_n = M\eta_n$$

entonces la estabilidad es normal.

Si todos los η_n son uniformemente acotadas entonces de

$$\{|\eta_n - d_n| \geqslant \varepsilon\} \to 0, \quad n \to \infty \tag{1}$$

se concluye la relación

$$|M\eta_n - d_n| \to 0, \qquad n \to \infty$$

y en consecuencia

$$\{|\eta_n - M\eta_n| \geqslant \varepsilon\} \to 0, \quad n \to \infty \tag{2}$$

De esta forma la estabilidad de la sucesión acotada es necesariamente normal.

Sea

$$\sigma^2 = D\eta_n (= M(\eta_n - M\eta_n)^2).$$

Por la desigualdad de Chebishev

$$P\{|\eta_n - M\eta_n| \geqslant \varepsilon\} \leqslant \frac{\sigma_n^2}{\varepsilon^2}.$$

En consecuencia la condición de Markov:

$$\sigma_n^2 \to 0, \quad n \to \infty \tag{3}$$

es suficiente para la estabilidad normal. Si $\eta_n - M\eta_n$ es uniformemente acotada,

$$|\eta_n - M\eta_n| \leqslant C ,$$

entonces según la desigualdad (9) de §3 del cuarto capítulo.

$$P\{|\eta_n - M\eta_n| \geqslant \varepsilon\} \geqslant \frac{\sigma^2 - \varepsilon^2}{C^2}$$

En este caso la condición de Markov (3) es también necesaria para la estabilidad normal de η_n.

Si
$$\eta_n = \frac{\xi_1 + \xi_2 + \ldots + \xi_n}{n}$$

y las variables ξ_n son nocorrelacionados en parejas, entonces

$$\mathcal{D}\eta_n = \frac{1}{n^2}\left[\mathcal{D}\xi_1 + \mathcal{D}\xi_2 + \ldots + \mathcal{D}\xi_n\right].$$

En consecuencia en este caso para la estabilidad normal de la media aritmética de η_n, o sea para todo $\varepsilon > 0$

$$\lim_n P\{\left|\frac{\xi_1+\ldots+\xi_n}{n} - \frac{M\xi_1+\ldots+M\xi_n}{n}\right| \geqslant \varepsilon\} = 0 \,,$$

es suficiente el cumplimiento de la siguiente condición

$$\lim_n \frac{1}{n^2}\sum_{i=1}^{n} \mathcal{D}\xi_i = 0 \qquad (4)$$

(teorema de Chebishev). En particular la condición (4) se cumple si todas las variables ξ_n son uniformemente acotadas.

1. Se pueden generalizar este teorema al caso de correlación débil de las variables ξ_n. Si suponemos que el coeficiente de correlación ρ_{mn}(1) entre ξ_m y ξ_n cumple la desigualdad

$$\rho_{mn} \leqslant C(|m-n|)$$

y $C(k) \geqslant 0$, entonces la estabilidad normal de las medias arit

(1) Es claro que siempre $\rho_{nn} = 1$.

méticas, o sea para todo $\varepsilon > 0$

$$\lim_n P\{|\frac{\xi_1+..+\xi_n}{n} \quad \frac{M\xi_1+..+M\xi_n}{n}| \geqslant \varepsilon\} = 0$$

es suficiente el complimiento de la condición[1].

$$\lim_n \frac{C_n}{n^2} \sum_{i=1}^{n} D\xi_i = 0 \,, \tag{5}$$

donde $C_n = \sum_{k=0}^{n-1} C(k)$.

2. En caso de sumandos independientes ξ_n se puede dar las condiciones necesaria y suficiente para la estabilidad de la media aritmética de η_n.

Para cada ξ_n existe la constante m_n (mediana de ξ_n) que cumple las siguientes condiciones

$$P(\xi_n \leqslant m_n) \leqslant \frac{1}{2} \,,$$
$$P(\xi_n \geqslant m_n) \leqslant \frac{1}{2} \,.$$

Pongamos

$$\xi_{nk} = \begin{cases} \xi \,, & \text{si} \quad |\xi_k - m_k| \leqslant n \,, \\ 0 \,, & \text{si} \quad |\xi_k - m_k| > n \,, \end{cases}$$

$$\eta_n^* = \frac{\xi_{n1}+\ldots+\xi_{nn}}{n}$$

TEOREMA. [1] pag.105. *Sea $\xi_1, \xi_2, \ldots,$ una sucesión de variables aleatorias independientes conjuntamente. Entonces la con*

(viene pag. 103)
(1) Comparar con A. Khintchine, Sur la loi forte de grandes nombres, C.R. de l'Acad.Sci., Paris, V.186 (1928), 285.

dición

$$\sum_{k=1}^{n} P\{|\xi_k - m_k| > n\} = \sum_{k=1}^{n} P\{\xi_{nk} \neq \xi_k\} \to 0, \quad n \to \infty \tag{6}$$

$$\frac{1}{n^2} \sum_{k=1}^{n} D\xi_{nk} \to 0 , \qquad n \to \infty \tag{7}$$

es necesario y suficiente para la estabilidad de las variables η_n, $n = 1,2,\ldots$

Además si las constantes d_n, $n = 1,2,\ldots$ *se pueden tomar iguales a* $M\eta_n$, *así que en caso de*

$$M\eta_n^* - M\eta_n \to 0, \quad n \to \infty ,$$

(y solo en este caso) para la estabilidad normal.

Demostración. La suficiencia de las condiciones (6) y (7) se establece sencillamente. En efecto ya que

$$P(\eta_n \neq \eta_n^*) \leq \sum_{k=1}^{n} P(\xi_{nk} \neq \xi_k) \to 0, \quad n \to \infty ,$$

y de acuerdo a la desigualdad de Chebishev,

$$P\{|\eta_n^* - M\eta_n| \geq \varepsilon\} \leq \frac{1}{n^2 \varepsilon^2} \sum_{k=1}^{n} D\xi_{nk} \to 0, \quad n \to \infty ,$$

entonces

$$P\{|\eta_n - M\eta_n^*| \geq \varepsilon\} \to 0, \quad n \to \infty$$

Para la demostración de la necesidad es necesario una serie de supuestos auxiliares.

---

(1) Comparar con A. Kolmogórov, Uber die Summer durch der Zufall bestimnter unabhangiger Grossen, Math. Ann., V.99 (1928), pp.309-319.

LEMA 1. *Sean* $A_1, A_2, \ldots, A_n$ *sucesos independientes* $P(A_i) > 0$, $i = 1,2,\ldots,n$ *y para alguna* $u \geqslant 0$, $P[\bigcup_{i=1} A_i] \geqslant u$. *Si además el suceso* U *es tal que para cada* $i = 1,2,\ldots,n$

$$P(U \mid A_i) \geqslant u\,,$$

entonces

$$P(U) \geqslant \frac{1}{9}\,u^2 \quad . \tag{8}$$

Demostración. Si existe un i tal que $P(A_i) \geqslant \frac{1}{3}\,u$, entonces

$$P(U) \geqslant P(U \mid A_i)P(A_i) \geqslant \frac{1}{3}\,u^2.$$

Supongamos ahora que para todo $i = 1,2,\ldots,n$

$$P(A_i) < \frac{1}{3}u\,.$$

Entonces se encuentra un k tal que

$$\frac{1}{3}\,u \leqslant P(A_1 \cup \ldots \cup A_k) \leqslant \frac{2}{3}\,u,$$

y por lo tanto para todo $i \leqslant k$

$$P(A_2 \cup \ldots \cup A_{i-1} | A_i) = P(A_1 \cup \ldots \cup A_{i-1}) \leqslant P(A_1 \cup \ldots \cup A_k) \leqslant \frac{2}{3}\,u,$$

$$P(\,U \cap \overline{(A_1 \cup \ldots \cup A_{i-1})} \mid A_i) \geqslant \frac{1}{3}\,u,$$

$$P(\,U \cap \overline{(A_1 \cup \ldots \cup A_{i-1})} \cap A_i) \geqslant \frac{1}{3}\,u P(A_i).$$

De donde

$$P(U) \geqslant \sum_{i=1}^{k} P(\,U \cap \overline{(A_1 \cup \ldots \cup A_{i-1})} \cap A_i) \geqslant \frac{1}{3}\,u \sum_{i=1}^{k} P(A_i)$$

$$\geqslant \frac{1}{3}\,u P(A_1 \cup \ldots \cup A_k) \geqslant \frac{1}{9}\,u^2.$$

LEMA 2. *Sean* $\xi_1, \xi_2, \ldots, \xi_n$ *variables aleatorias independientes, acotadas* $|\xi_i| \leqslant C$, $i = 1,\ldots,n$ *con medias cero. Enton-*

ces para todo $\alpha > 0$ *y todo entero* m

$$P\{\max_{k\leq n} |\xi_1+..+\xi_k| \geq m(\alpha D + C)\} < \frac{1}{\alpha^{2m}} \tag{9}$$

donde

$$D^2 = \sum_{i=1}^{n} D\xi_i .$$

Demostración. Denotamos

$$R = \alpha D + C, \qquad S_o = 0, \qquad S_k = \xi_1+\ldots+\xi_k ,$$

$$B_{ik}=\{\omega : |S_j| < iR, \qquad j < k, \qquad |S_k| \geq iR\},$$

$$B_i = \sum_{k=1}^{n} B_{ik} .$$

notando que en el conjunto B_{ik}

$$|S_k| \leq iR + C ,$$

obtenemos

$$\begin{aligned} P(B_{i+1} \mid B_{ik}) &= \{\max_{1\leq k\leq n} |S_k| \geq (i+1)R \mid B_{ik}\} \\ &\leq P\{\max_{k+1\leq p\leq n} |\sum_{j=k+1}^{p} \xi_j| \geq \alpha D \mid B_{ik}\} \\ &= P\{\max_{k+1\leq p\leq n} |\sum_{j=k+1}^{p} \xi_j| \geq \alpha D\}. \end{aligned}$$

De la desigualdad (3) del teorema 2 demostrada más adelante en §5 se deduce que

$$P\{\max_{k+1\leq p\leq n} |\sum_{j=k+1}^{p} \xi_j| \geq \alpha D\} \leq \frac{\sum_{j=k+1}^{p} D\xi_j}{\alpha^2 D^2} \leq \frac{D^2}{\alpha D^2} = \frac{1}{\alpha^2}$$

Por eso $P(B_{i+1}|B_{ik}) \leq \frac{1}{\alpha^2}$ para cualquier $k = 1,2,\ldots,n$, lo que indica que

$$P(B_{i+1} \mid B_i) \leqslant \frac{1}{\alpha^2}$$

y

$$P\{\max_{1<k<n} |S_k| \geqslant mR\} = P(B_m) = P(B_m|B_{m-1})P(B_{m-1}|B_{m-2})\ldots P(B_1|B_o)$$

$$\leqslant \frac{1}{\alpha^{2m}}$$

LEMA 2. *Sean* $\xi_1, \xi_2, \ldots, \xi_n$ *variables aleatorias independientes y acotadas además* $|\xi_i - M\xi_i| \leqslant C$, $i = 1,2,\ldots,n$. *Entonces*

$$P\{|\xi_1 + \ldots + \xi_n| \geqslant a\} \geqslant \frac{1}{1600}\left[1 - \frac{4a^2 + C^2}{\sum_{i=1}^{n} D\xi_i}\right] \qquad (10)$$

Demostración. Denotemos por $S = \xi_1 + \ldots + \xi_n$, $D^2 = \sum_{i=1}^{n} D\xi_i$. Si $C > D$ y $2a > D$, entonces la parte derecha en (10) es negativa y la desigualdad es evidente.

Sea ahora simultáneamente $C \leqslant D$, $2a \leqslant D$. Entonces es suficiente mostrar que

$$P\left(|S| \geqslant \frac{D}{2}\right) \geqslant \frac{1}{1600},$$

ya que es evidente

$$P(|S| > a) \geqslant P(|S| > \frac{D}{2}) \geqslant \frac{1}{1600} \geqslant \frac{1}{1600}\left[1 - \frac{4a^2 + C^2}{D^2}\right].$$

Denotemos por $A = \{|S| \geqslant \frac{D}{2}\}$. Si $|MS| \geqslant 2D$, entonces

$$D^2 \geqslant P(\bar{A})M\{(S-MS)^2|\bar{A}\} \geqslant (2D - \frac{D}{2})^2 P(\bar{A}) = \frac{9}{4} D^2 P(\bar{A})$$

y por lo tanto

$$P(A) \geqslant \frac{1}{2}.$$

Supongamos ahora que $|MS| < 2D$. Denotando ,

$$A_m = \{\omega : 3m \leqslant |S-MS| < 3(m+1)D, \quad |S| \geqslant \frac{D}{2}\}$$

y aplicando el lema 2 encontramos

$$P(A_m) \leqslant P\{|S-MS| \geqslant 3mD\} = P\{|S-MS| \geqslant m(2D+D)\}$$

$$\leqslant P\{|S-MS| \geqslant m(2D+C)\} \leqslant \frac{1}{2^{2m}}.$$

De donde

$$M(S^2|A_m)P(A_m) = M[(S-MS)^2 + 2(S-MS)MS + (MS)^2|A_m)P(A_m)$$

$$< 25\frac{(m+1)}{2^{2m}}D^2.$$

En el conjunto $A' = \sum_{m=0}^{5} A_m$

$$|S| \leqslant |S-MS| + |MS| \leqslant 20 \cdot D$$

Por eso

$$M(S^2|A')P(A') \leqslant 400\,D^2P(A).$$

Es claro que

$$M(S^2|\bar{A})P(\bar{A}) \leqslant \frac{1}{4}D^2.$$

En consecuencia

$$D^2 \leqslant MS^2 = M[S^2|\bar{A}]P(\bar{A}) + M[S^2/A']P(A') + \sum_{m=6}^{\infty} M[S^2|A_m]P(A_m)$$

$$\leqslant \frac{1}{4}D^2 + 400\,D^2P(A) + \sum_{m=6}^{\infty}\frac{(m+1)^2}{2^{2m}}\,25D^2$$

y por lo tanto

$$P(A) \geqslant \frac{1}{1600}$$

Demostración del teorema (la condición necesaria)

Sea la sucesión d_n, $n = 1,2,\ldots$ tal que para todo $\varepsilon > 0$

$$P(|\eta_n - d_n| \geq \varepsilon) \to 0, \quad n \to \infty .$$

Mostraremos entonces que

$$\sum_{k=1}^{n} P(|\xi_k - m_k| \geq n\varepsilon) \to 0, \quad n \to \infty . \qquad (11)$$

Denotemos para cada n

$$U = \{|\eta_n - d_n| \geq \frac{\varepsilon}{2}\},$$

$$A_k = \{|\xi_k - m_k| \geq n\varepsilon\} ,$$

$$B_k = \{\left|\frac{\sum_{i \neq k} \xi_i + m_k}{n} - d_n\right| \geq \frac{\varepsilon}{2}\} = \{|(\eta_n - d_n) + \frac{m_k - \xi_k}{n}| \geq \frac{\varepsilon}{2}\}.$$

Ya que m_k es la mediana de ξ_k, entonces

$$P(U \mid B_k) \geq \frac{1}{2} .$$

Para un n suficientemente grande

$$P(U) \leq \frac{1}{4} ,$$

por eso

$$\frac{1}{4} \geq P(U) \geq P(U \mid B_k)P(K_k) \geq \frac{1}{2} P(B_k),$$

o sea

$$P(B_k) \leq \frac{1}{2} .$$

Si el suceso A_k ocurre y B_k no, entonces ocurre el suceso U y esto indica que

$$P(B_k \mid A_k) + P(U \mid A_k) \geq 1.$$

Pero

$$P(B_k \mid A_k) = P(B_k) \leq \frac{1}{2} .$$

En consecuencia

$$P(U|A_k) \geq 1 - P(B_k|A_k) \geq \frac{1}{2} \geq \frac{1}{2} P(\bigcup_{k=1}^{n} A_k)$$

Aplicamos el lema 1, tomando

$$u = \frac{1}{2} P(\bigcup_{k=1}^{n} A_k).$$

Entonces

$$P(U) \geq \frac{1}{36} \left[P(\bigcup_{k=1}^{n} A_k)\right]^2 \tag{12}$$

Los sucesos $A_1,\ldots,A_n$ son independientes, por eso

$$P(\bigcup_{k=1}^{n} A_k) = 1 - \prod_{k=1}^{n} (1 - P(A_k)) \tag{13}$$

Ya que según las condiciones $P(U) \to 0$, $n \to \infty$, entonces, de (12) y de (13) obtenemos la relación buscada (11).

Pongamos ahora

$$\bar{\xi}_{nk} = \begin{cases} \xi_k, & \text{si } |\xi_k - m_k| \leq n\,, \\ m_k, & \text{si } |\xi_k - m_k| > n\,, \end{cases}$$

$$\bar{\xi}_{nk}(\varepsilon) = \begin{cases} \xi_k, & \text{si } |\xi_k - m_k| \leq \varepsilon n \\ m_k, & \text{si } |\xi_k - m_k| > \varepsilon n\,. \end{cases}$$

De (11) se deduce que si $\{|\eta_n - d_n| \geq \varepsilon\} \to 0$, $n \to \infty$, entonces

$$P\{|\frac{1}{n} \sum_{k=1}^{m} \bar{\xi}_{nk} - d_n| > \varepsilon\} \to 0\,,$$

$$P\{|\frac{1}{n} \sum_{k=1}^{n} \bar{\xi}_{nk}(\varepsilon) - d_n| \geq \varepsilon\} \to 0\,.$$

Denotemos por $\bar{\xi}_{nk}(\varepsilon) = \xi_{nk}(\varepsilon) - M\xi_{nk}(\varepsilon)$. Entonces $|\bar{\xi}_{nk}(\varepsilon)| \leqslant 2n\varepsilon$ y por el lema 3

$$\{|\frac{1}{n}\sum_{k=1}^{n} \xi_{nk}(\varepsilon) - d_n| \geqslant \varepsilon\} = P\{|\sum_{k=1}^{n} \xi_{nk}(\varepsilon) - nd_n| \geqslant \varepsilon n\}$$

$$= \{|\sum_{k=1}^{n} (\xi_{nk}(\varepsilon) - d_n)| \geqslant \varepsilon n\} \geqslant \frac{1}{1600}\left[1 - \frac{8\varepsilon^2 n^2}{\sum_{k=1}^{n} D\,\bar{\xi}_{nk}(\varepsilon)}\right]$$

de donde

$$\lim_{n} \sup \frac{1}{n^2}\sum_{k=1}^{n} D\bar{\xi}_{nk}(\varepsilon) \leqslant 8\varepsilon^2. \qquad (14)$$

Para $\varepsilon \leqslant 1$

$$|D\bar{\xi}_{nk}(\varepsilon) - D\bar{\xi}_{nk}| \leqslant 8n^2 P\{|\xi_k - m_k| > n\varepsilon\} .$$

Entonces de (11), (13) y (14) se deduce que

$$\lim \sup_{n} \frac{1}{n^2}\sum_{k=1}^{n} D\bar{\xi}_{nk} \leqslant 8\varepsilon^2 ,$$

lo cual indica, en virtud a la arbitrariedad de $\varepsilon \in [0,1]$ que

$$\lim_{n} \frac{1}{n^2}\sum_{k=1}^{n} D\xi_{nk} = \lim_{n} \frac{1}{n^2}\sum_{k=1}^{n} D\bar{\xi}_{nk} = 0.$$

3. La siguiente generalización del teorema de Chebishev se obtiene, si se supone que η_n de una u otra forma depende de los resultados de algunas n pruebas

$$\mathfrak{U}_1, \mathfrak{U}_2, \mathfrak{U}_3, \ldots, \mathfrak{U}_n$$

tal que después de cada resultado cualquiera de todas las n pruebas η_n toma un valor determinado. La idea general de todos los teoremas conocidas con el nombre de Ley de los grandes "nú meros" consiste en lo siguiente si la dependencia de la variables η_n de cada prueba por separado U_k, $k = 1,2,\ldots,n$ es muy pequeña para un n grande, entonces las variables η_n son estables. Si examinamos

$$\beta_{nk}^2 = M[M(\eta_n \mid U_1,U_2,\ldots,U_k) - M(\eta_n \mid U_1,U_2,\ldots,U_{k-1})]^2$$

como medida de dependencia de las variables η_n de las pruebas U_k, entonces, la idea general, arriba relacionada, de la ley de los grandes números puede ser caracterizada en las siguientes raciocinios(1).

Sea

$$\zeta_{nk} = M[\eta_n \mid U_1,\ldots,U_k] - M[\eta_n \mid U_1,\ldots,U_{n-1}].$$

Entonces

$$\eta_n - M\eta_n = \zeta_{n1} + \zeta_{n2} + \ldots + \zeta_{nn},$$

$$M\zeta_{nk} = MM[\eta_n \mid U_1,\ldots,U_k] - MM[\eta_n \mid U_1,\ldots,U_{k-1}]$$

$$= M\eta_n - M\eta_n = 0$$

$$D\zeta_{nk} = M\zeta_{nk}^2 = \beta_{nk}^2 .$$

Es fácil calcular que las variables aleatorias ζ_{nk}, $k = 1,2,\ldots,n$ son no correlacionadas. En efecto si $i < k$ entonces(2)

$$M[\zeta_{ni}\zeta_{nk} \mid U_1,\ldots,U_{n-1}] = \zeta_{ni}M[\zeta_{nk} \mid U_1,\ldots,U_{k-1}] =$$

---

(1) Comp. A. Kolmogórov , Sur la loi des grandes nombres, Rend. Acad. Lincei, Nº 9 (1929), 470-474.

(2) Aplicación de la fórmula (15) §4, Cap.V

$$= \zeta_{ni} M\{M(\eta_n \mid u_1,\ldots,u_k) - M(\eta_n \mid u_1,\ldots,u_{k-1}) \mid u_1,\ldots,u_{k-1}\}$$

$$\zeta_{ni}[M(\eta_n \mid u_1,\ldots,u_{k-1}) - M(\eta_n \mid u_1,\ldots,u_{k-1})] = 0$$

y en consecuencia

$$M\zeta_{ni}\zeta_{nk} = 0, \qquad i < k$$

y así

$$D\eta_n = \sum_{i=1}^{n} D\zeta_{ni} = \sum_{i=1}^{n} \beta_{ni}^2$$

De esta manera la condición

$$\sum_{i=1}^{n} \beta_{ni}^2 \to 0, \qquad n \to \infty$$

es suficiente para la estabilidad normal de la variable η_n.

*

§4. OBSERVACIONES AL CONCEPTO DE ESPERANZA MATEMATICA

Definimos la esperanza matemática de la variable ξ como

$$M\xi = \int_{\Omega} \xi(\omega) Pd(\omega) = \int_{-\infty}^{\infty} xF\xi(dx).$$

En este caso la integral de la parte derecha se entiende como

$$M\xi = \int_{-\infty}^{\infty} xF\xi(d\omega) = \lim_{\substack{a\to-\infty \\ b\to+\infty}} \int_{x}^{b} xF\xi(dx) \tag{1}$$

La idea sugiere al mismo tiempo considerar la expresión

$$\tilde{M}\xi = \lim_{a\to\infty} \int_{-a}^{a} xF\xi(dx) \tag{2}$$

en calidad de esperanza matemática generalizada. Sin embargo, en este caso se pierden algunas propiedades sencillas de las esperanzas matemáticas. Por ejemplo, la fórmula

$$\tilde{M}[\xi+\eta] = \tilde{M}\xi + \tilde{M}\eta$$

no siempre es cierta. Nosotros vemos que si nos limitamos a ciertas suposiciones la definición (2) es completamente natural y útil.

En efecto, podemos discutir el problema como sigue: Sea

$$\xi_1,\ \xi_2,\ldots,\xi_n,\ldots$$

una sucesión de variables aleatorias independientes que tienen la misma función de distribución $F\xi(x) = F\xi_n(x)$, $n = 1,2,\ldots$ que ξ. Pongamos

$$\eta_n = \frac{\xi_1+\xi_2+\ldots+\xi_n}{n}$$

Preguntamos ahora si existe una constante $M°\xi$ tal que para cada $\varepsilon > 0$

$$\lim_{n} P\{|\eta_n - M°\xi| > \varepsilon\} = 0\,. \tag{3}$$

Respuesta. Si existe esta constante $M°\xi$ entonces, ella se expresa $M°\xi = \tilde{M}\xi$. La condición necesaria y suficiente para la existencia de $M°\xi$ estan contenidas en la existencia del límite (2) y el cumplimiento de la relación

$$P(|\xi| > n) = o\left(\frac{1}{n}\right)\,. \tag{4}$$

Este resultado se desprende directamente del siguiente teorema.

TEOREMA. *Sea* $\xi_1, \xi_2, \ldots$ *una sucesión de variables aleatorias independientes e igualmente distribuídas con función de distribución* $F(x)$. *Para la estabilidad de las medias aritméticas*

$$\eta_n = \frac{\xi_1 + \xi_2 + \ldots + \xi_n}{n}, \qquad n = 1, 2, \ldots$$

es necesario y suficiente que se cumpla la condición (4). *En este caso de estabilidad podemos colocar*

$$d_n = \int_{-n}^{n} xF(dx) \; .$$

Demostración. Denotemos

$$\xi_{nk} = \begin{cases} \xi_k, & \text{si} \quad |\xi_k - m| < n \\ 0, & \text{si} \quad |\xi_k - m| < n \; , \end{cases}$$

donde m es la mediana de la variable aleatoria ξ_1. De la condición (4) se desprende que

$$\sum_{k=1}^{n} P(\xi_{nk} \neq \xi_k) = nP(|\xi_1 - m| > n) \to 0, \qquad n \to \infty$$

y

$$\overline{\lim_{n}} \, \frac{1}{n} \int_{\{x: |x-m| \leq n\}} x^2 F(dx) = \overline{\lim_{n}} \, \frac{1}{n} \int_{|x| \leq n} x^2 F(dx). \qquad (5)$$

Pero

$$\frac{1}{n} \int_{|x| \leq n} x^2 F(dx) = \frac{1}{n} \sum_{k=1}^{n} \int_{k-1 \leq |x| < k} x^2 F(dx) \leq$$

$$\leq \frac{1}{n}\sum_{k=1}^{n} K^2 P(K-1 < |\xi_1| \leq K)$$

$$\leq \frac{1}{n}\sum_{k=1}^{n}\Big[\sum_{\ell=1}^{k} \ell P(K-1 < |\xi_1| < K)\Big]$$

$$\leq \frac{2}{n}\sum_{\ell=1}^{n} \ell P(\ell-1 < |\xi_1| \leq \ell) \leq \frac{2}{n}\sum_{\ell=1}^{n} \ell P(|\xi_1| > \ell - 1) \to 0.$$

ya que

$$\ell P\{|\xi_1| > \ell-1\} \to 0, \qquad \ell \to \infty .$$

Por eso

$$\frac{1}{n^2}\sum_{k=1}^{n} D\xi_{nk} \leq \frac{1}{n^2}\sum_{k=1}^{n} M\xi_{nk}$$

$$= \frac{1}{n^2}\int_{|x-m|\leq n} x^2 F(dx) \to 0, \qquad n \to \infty.$$

De esta forma las condiciones (6) y (7) del teorema del pará-grafo (3) se cumple para todo $\varepsilon > 0$,

$$P(|\eta_n - M\eta_n^*| \geq \varepsilon) \to 0, \qquad n \to \infty$$

donde

$$M\eta_n^* = M\,\frac{\xi_{n1}+\ldots+\xi_{nn}}{n} = \int_{|x-m|\leq n} xF(dx).$$

Pero por la condición (4)

$$\int_{|x-m|\leq n} xF(dx) - \int_{|x|\leq n} xF(dx) \to 0, \qquad n \to \infty,$$

por eso

$$P(|\eta_n - d_n| \geq \varepsilon) \to 0, \qquad n \to \infty ,$$

donde

$$d_n = \int_{-n}^{n} xF(dx).$$

Lo contrario, si la sucesión η_n, $n = 1,2,\ldots$ es estable entonces en virtud a la relación (11) del parágrafo (3)

$$nP\{|\xi_1 - m| \geqslant n\} \to 0, \qquad n \to \infty,$$

de donde se desprende la condición (4).

Si existe la esperanza matemática en el sentido inicial (fórmula (1)) entonces la condición (4) siempre se cumple[1] y $M\xi = M^{o}\xi$.

De esta manera $M^{o}(\xi)$ es la generalización formal de la esperanza matemática $M(\xi)$. Para ella se conservan las propiedades del 1 al 7 (Cap. IV, §2), sin embargo, en el caso general de la existencia $M^{o}(\xi)$ no se deduce la existencia $M^{o}|\xi|$.

Para demostrar que el nuevo concepto de la esperanza matemática es en realidad más general que el anterior es suficiente el siguiente ejemplo. Tomemos las densidades de probabilidad $f_\xi(x)$ igual

$$f_\xi(x) = \frac{c}{(|x|+2)^2 \ln(|x|+2)},$$

además la constante c está determinada por la condición

$$\int_{-\infty}^{\infty} f_\xi(x)\,dx = 1.$$

---

(1) Comparar con A. Kolmogórov , Bemerkungen zu meiner Arbeit Über die Summen zufälleger Grössen, Math. Ann., V. 102 (1930), p. 484-488, Teorema XII.

Es fácil calcular que en este caso se cumple la condición (4). La fórmula (2) nos da el valor

$$M^{\circ}\xi = 0.$$

Sin embargo, la integral

$$\int_{-\infty}^{\infty} |x| F_{\xi}(dx) = \int_{-\infty}^{\infty} |x| f_{\xi}(x)\, dx$$

diverge.

Si Ω es el segmento $[0,1]$ y P la medida Lebesgue, entonces $M^{\circ}\xi$ no es otra cosa que la A-integral(1)

$$(A) \int_{0}^{1} \xi(\omega) P(d\omega).$$

*

§5. LEY FUERTE DE LOS GRANDES NUMEROS, CONVERGENCIA DE SERIES.

DEFINICION 1. *La sucesión de variables aleatorias* η_n

$$\eta_1, \eta_2, \ldots, \eta_n, \ldots$$

es fuertemente estable si existe una sucesión

$$d_1,\ d_2, \ldots, d_n, \ldots$$

tal que las variables aleatorias $\eta_n - d_n$ *casi siempre tiende a cero cuando* $n \to \infty$:

$$P\{\lim_{n} (\eta_n - d_n) = 0\} = 1.$$

---

(1) Ver N.K. Bari, Series trigonométricas. Edic. Fismat, 1961.

De la estabilidad fuerte se deduce, claramente, la estabilidad ordinaria. Si se puede escoger

$$d_n = M\eta_n \ ,$$

entonces decimos que la estabilidad fuerte es normal.

TEOREMA 1. *Sean $\xi_1, \xi_2, \ldots$ una sucesión de variables aleatorias independientes. Para la estabilidad fuerte normal de las medias aritméticas*

$$\eta_n = \frac{\xi_1 + \ldots + \xi_n}{n} \ ; \quad n = 1,2,\ldots, \tag{1}$$

es suficiente que se cumpla la condición[1]

$$\sum_{n=1}^{\infty} \frac{D\xi_n}{n^2} < \infty \tag{2}$$

Esta condición es la mejor, en el sentido que para cualquier serie de constantes no negativas $b_1, b_2, \ldots$ tales que $\sum_{n=1}^{\infty} \frac{b_n}{n^2} = \infty$ se puede construir una serie de variables aleatorias independientes $\xi_1, \ldots, \xi_n$ tal que $D\xi_n = b_n$ y ser respectivas medias aritméticas η_n, $n = 1,2,\ldots$ no son fuertemente estables. Para la demostración es necesario la desigualdad (3) del siguiente teorema.

TEOREMA 2. *Sea $\xi_1, \xi_2, \ldots, \xi_n$ variables aleatorias independientes con media igual a cero. Entonces para cualquier $a > 0$*

$$P\{\max_{1 \leq k \leq n} |\xi_1 + \ldots + \xi| \geq a\} \leq \frac{\sum_{k=1}^{n} D\xi_k}{2} \tag{3}$$

y si además las variables aleatorias $\xi_1, \ldots, \xi_n$ son acotadas

---

(1) A. Kolmogórov, Sur la loi forte des grandes nombres C.R. Acad. Sci., Paris, V. 191 (1930), 910-911.

$|\xi_i| \leq c$, $i = 1,\ldots,n$, *entonces tiene lugar la siguiente desigualdad*

$$P\{\max_{1\leq k\leq m} \{\xi_1+\ldots+\xi_k\} \geq a\} \geq 1 - \frac{(c+a)^2}{\sum_{k=1}^{n} D\xi_k} \tag{4}$$

Demostración del Teorema 2. Pongamos $S_o = 0$, $S_k = \xi_1+\ldots+\xi_k$,

$$A = \{\max_{1\leq k\leq n} |S_k| \geq a\}, \quad A_k = \{\max_{j\leq k\leq 1} |S_j| < a, \quad |A_k| \geq a\},$$

donde $k = 1,2,\ldots,n$. Entonces $\sum_{k=1}^{n} A_k = A$ y denotando I_B es indicador del conjunto B, encontramos

$$\begin{aligned}
\sum_{k=1}^{n} D\xi_k = MS_n^2 \geq M[S_n^2\, I_A] &= \sum_{k=1}^{n} M[S_n^2\, I_{Ak}] \\
&= \sum_{k=1}^{n} M\{[S_k + (S_n - S_k)]^2 I_{Ak}\} = \sum_{k=1}^{n} MI_{Ak}\, S_k^2 \\
&+ \sum_{k=1}^{n} M\{[S_k+(S_n - S_k)]^2 I_{Ak}\} = \sum_{k=1}^{n} MI_{Ak}\, S_k^2 \\
&+ \sum_{k=1}^{n} MI_{Ak}(S_n - S_k)^2 + 2\sum_{k=1}^{n} M[I_{Ak}S_k(S_n - S_k)] \\
&\geq a^2 \sum P(A_k) = a^2 P(A),
\end{aligned} \tag{5}$$

donde nosotros utilizamos que:

$$M[I_{Ak}S_k(S_n - S_k)] = M[I_{Ak}S_k\, M(S_n - S_k \mid \xi_1,\ldots,\xi_k)] = 0.$$

De (5) se desprende la desigualdad pedida (3). Para establecer (4), anotamos por una parte que:

$$M[S_n^2 I_A] = MS_n^2 - MS_n^2 I_{\bar{A}}$$
$$\geq \sum_{k=1}^{n} D\xi_k - a^2 P(\bar{A}) = \sum_{k=1}^{n} D\xi_k - a^2 + a^2 P(A) \tag{6}$$

Por otra parte, utilizando que en los conjuntos $A_k |S_{k-1}| \leq a$ o sea $|S_k| \leq a + c$. Entonces

$$M[S_n^2 I_A] = \sum_{k=1}^{n} M[S_k^2 I_{A_k}] + \sum_{k=1}^{n} M[I_{A_k}(S_A - S_k)^2]$$
$$\leq (a+c)^2 \sum_{k=1}^{n} P(A_k) + \sum_{k=1}^{n} P(A_k) \sum_{j=k+1}^{n} D\xi_j \leq [(a+c)^2 + \sum_{j=1}^{n} D\xi_j] P(A). \tag{7}$$

De (6) y (7) obtenemos la desigualdad buscada

$$P(A) \geq \frac{\sum_{k=1}^{n} D\xi_k - a}{(a+c)^2 + \sum_{k=1}^{n} D\xi_k - a^2} = 1 - \frac{(a+c)^2}{(a+c)^2 + \sum_{k=1}^{n} D\xi_k - a^2} \geq 1 - \frac{(a+c)^2}{\sum_{k=1}^{n} D\xi_k}$$

Demostración del Teorema 1. Sin restricciones podemos considerar $M\xi_n = 0$, $n = 1, \ldots$. Entonces para la demostración de que da media aritmética $\eta_n = \frac{S_n}{n}$ converge con probabilidad 1 hacia cero es suficiente demostrar que para cualquier $\varepsilon > 0$ la probabilidad

$$P\{\lim_n \sup |\frac{S_n}{n}| \geq \varepsilon\} = 0.$$

En virtud de la desigualdad (3)

$$P_m \equiv P\{\max_{2^m \leq n < 2^{m+1}} |\frac{S_n}{n}| \geq \varepsilon\}$$
$$\leq P\{\max_{1 \leq n < 2^{m+1}} |S_n| \geq 2^m \varepsilon\} \leq (\frac{1}{2^m \varepsilon})^2 \sum_{n=1}^{2^{m+1}} D\xi_n.$$

Por eso

$$P \equiv P\{\lim\sup |\frac{S_n}{n}| \geq \varepsilon\} \leq \sum_{m=0}^{\infty} P_m$$

$$\leq \frac{1}{\varepsilon^2} \sum_{m=0}^{\infty} \left(\frac{1}{2^m}\right)^2 \sum_{n=1}^{2^{m+1}} D\xi_n \leq \frac{1}{\varepsilon^2} \sum_{i=0}^{\infty} \sum_{m=i}^{\infty} \left(\frac{1}{2^m}\right)^2 \sum_{n=2^i}^{2^{i+1}} D\xi_n$$

$$\leq \frac{1}{\varepsilon^2} \sum_{i=0}^{\infty} \left(\frac{1}{2^{i-1}}\right)^2 \sum_{n=2^i}^{2^{i+1}} D\xi_n \leq \frac{8}{\varepsilon^2} \sum_{n=1}^{\infty} \frac{D\xi_n}{n^2} .$$

Anotemos ahora que la probabilidad P no cambia si reemplazamos por ceros cualquier número finito de primeros términos de la sucesión $\xi_1,\xi_2,\ldots,\xi_n$. O sea para cualquier N

$$P \leq \frac{8}{\varepsilon^2} \sum_{n=N}^{\infty} \frac{D\xi_n}{n^2}$$

lo cual en virtud de la convergencia de la serie $\sum_{n=1}^{n} \frac{D\xi_n}{n}$ se demuestra que $P = 0$.

Para la demostración de la última afirmación del teorema construimos una sucesión de variables aleatorias independientes $\xi_1,\xi_2,\ldots$, de la siguiente manera:

Si $\frac{b_n}{n^2} < 1$ entonces tomamos $\xi_n = n$, $\xi_n = -n$, $\xi_n = 0$ con probabilidades

$$\frac{b_n}{n^2}, \quad \frac{b_n}{2n^2}, \quad 1 - \frac{b_n}{n^2}$$

correspondiente.

Si $\frac{b_n}{n^2} \geq 1$, entonces ponemos $\xi_n = \sqrt{b_n}$, $\xi_n = -\sqrt{b_n}$ con probabilidad igual a $\frac{1}{2}$.

Entonces $M\xi_n = 0$, $M\xi_n^2 = b_n$. Si suponemos ahora que en el conjunto con probabilidad positiva $\frac{S_n}{n} \to 0$, $n \to \infty$ entonces de la igualdad

$$\frac{\xi_n}{n} = \frac{S_n}{n} - \left(\frac{n-1}{n}\right)\frac{S_{n-1}}{n-1}$$

se concluirá que con probabilidad positiva $\frac{\xi_n}{n} \to 0$, $n \to \infty$. Pero en virtud de la divergencia de la serie $\sum_{n=1}^{\infty} \frac{b_n}{n^2}$, la serie

$$\sum_{n=1}^{\infty} P\{|\frac{\xi_n}{n}| \geqslant 1\} = \infty \quad .$$

De donde por el teorema de Borel-Cantelly[1] se deduce que $\frac{\xi_n}{n}$ no puede con probabilidad positiva converger a cero.

Para el caso cuando las variables aleatorias independientes $\xi_1, \xi_2, \ldots$ tienen una misma función de distribución $F = F(x)$ se puede dar la condición necesaria y suficiente para la estabilidad fuerte de las medidas aritméticas.

TEOREMA 3. *Si $\xi_1, \xi_2, \ldots$ es una sucesión de variables aleatorias independientes igualmente distribuidas, entonces la condición $M|\xi_1| < \infty$ es necesaria y suficiente para la estabilidad fuerte normal de las medias aritméticas $\eta_n = \frac{S_n}{n}$, $n = 1, 2, \ldots$*

Demostración. Sea $M|\xi_1| < \infty$; mostraremos entonces que con probabilidad uno

$$\frac{S_n}{n} \to M\xi_1, \qquad n \to \infty.$$

Paralelamente a las variables ξ_n analizaremos las variables

$$\xi_n^* = \begin{cases} \xi_n & \text{si} \quad |\xi_n| \leqslant n \\ 0 & \text{si} \quad |\xi_n| > n \end{cases}$$

---

(1) Aquí y en lo sucesivo se utiliza la variante del lema Borel-Cantelly: si $A_1, A_2, \ldots$ es una sucesión de sucesos tales que sus indicadores $I_{A_1}, I_{A_2}, \ldots$ son independientes, entonces $\sum_{n=1}^{\infty} I_{A_n} < \infty$ con probabilidad uno si y solo si $\sum_{n=1}^{\infty} P(A_n) < \infty$.

Entonces

$$\sum_{n=1}^{\infty} P\{\xi_n \neq \xi_n^*\} = \sum_{n=1}^{\infty} P\{|\xi_n| > n\} = \sum_{n=1}^{\infty} P\{|\xi_1| > n\}$$

$$= \sum_{n=1}^{\infty} \sum_{k \geqslant n} P\{k < |\xi_1| \leqslant k+1\} = \sum_{n=1}^{\infty} nP\{n < |\xi_1| \leqslant n+1\}$$

$$\leqslant \sum_{k=0}^{\infty} \int_{\{k<|x|\leqslant k+1\}} |x| F(dx) \leqslant \int_{-\infty}^{\infty} |x| F(dx) = M|\xi_1| < \infty.$$

Por eso según el teorema de Borel-Cantelly con probabilidad 1 ocurre solo un número finito de sucesos. $A_n = \{\xi_n \neq \xi_n^*\}$, $n = 1,2,\ldots$ De donde se deduce que es suficiente demostrar que con probabilidad 1 que

$$\frac{\xi_1^* + \ldots + \xi_n^*}{n} \to M\xi_1 .$$

Ya que $M\xi_n^* \to M\xi_1$. Entonces de aquí no es difícil deducir

$$\frac{M\xi_1^* + \ldots + M\xi_n^*}{n} \to M\xi_1 ,$$

esto indica que es necesario establecer que con probabilidad 1

$$\frac{\tilde{\xi}_1 + \ldots + \tilde{\xi}_n}{n} \to 0,$$

donde:

$$\tilde{\xi}_n = \xi_n^* - M\xi_n^* .$$

Para la demostración de esto es suficiente solo comprobar que para la sucesión $\tilde{\xi}_1, \tilde{\xi}_2, \ldots$ se cumple la condición $\sum_{n=1}^{\infty} \frac{M\tilde{\xi}_n^2}{n^2} < \infty$ del Teorema 1.

En efecto

$$\sum_{n=1}^{\infty} \frac{M\tilde{\xi}_n^2}{n^2} \leqslant \sum_{n=1}^{\infty} \frac{M(\xi_n^*)^2}{n^2} = \sum_{n=1}^{\infty} \frac{1}{n^2} \int x^2 F(dx) =$$

$$= \sum_{n=1}^{\infty} \frac{1}{n^2} \{ \sum_{k=1}^{\infty} \int_{\{k-1<|x|<k\}} x^2 F(dx) \}$$

$$\leq \sum_{k=1}^{\infty} \frac{1}{n^2} \sum_{k=1}^{n} \int_{\{k-1<|x|\leq k\}} |x| F(dx)$$

$$= \sum_{k=1}^{\infty} \int_{\{k-1<|x|\leq k\}} |x| F(dx) \sum_{n\geq k} \frac{1}{n^2}$$

$$\leq 2 \sum_{k=1}^{\infty} \int_{\{k-1\leq |x|<k\}} |x| F(dx) = 2M|\xi_1| < \infty .$$

Así la condición (2) del teorema 1 se cumple y en consecuencia con probabilidad $\frac{\xi_1+..+\xi_n}{n} \to 0,\ n \to \infty.$

Supongamos ahora que con probabilidad uno $\frac{S_n}{n} \to M\xi_1,\ n \to \infty$.

Entonces de la igualdad

$$\frac{\xi_n}{n} = \frac{S_n}{n} - (\frac{n-1}{n}) \frac{S_{n-1}}{n-1} ,$$

se deduce que con probabilidad uno $\frac{\xi_n}{n} \to 0,\ n \to \infty$.

Por eso para cada $\varepsilon > 0$

$\{ |\frac{\xi_n}{n}| > \varepsilon$ para un número infinito de $n\} = 0,$

lo cual es equivalente según el teorema de Borel-Cantelly a la desigualdad

$$\sum_{n=1}^{\infty} P\{ |\frac{\xi_n}{n}| > \varepsilon \} < \infty .$$

Pero en el caso que las variables tengan una misma distribución esta condición es equivalente, como es fácil de verificar, a la condición $M|\xi_1| < \infty$.

El teorema 3 queda demostrada.

Supongamos ahora $\xi_1, \xi_2, \ldots$ que es una sucesión de variables aleatorias independientes. Entonces la probabilidad de que la serie $\sum_{n=1}^{\infty}$ converja es igual a uno o a cero (ver el Teorema suplementario).

TEOREMA 4. *Para la convergencia con probabilidad uno de la serie $\sum_{n=1}^{\infty} \xi_n$ es suficiente la convergencia simultanea de las siguientes series*

$$\sum_{n=1}^{\infty} M\xi_n, \quad \sum_{n=1}^{\infty} D\xi_n .$$

Si además las variables aleatorias $\xi_1,\xi_2,\ldots$ son uniformemente acotadas $|\xi_n| \leqslant c$, $n = 1,2,\ldots$, entonces esta condición es además necesaria.

Demostración. De la convergencia de la serie $\sum_{n=1}^{\infty} D\xi_n$ y la desigualdad (3) se deduce que con probabilidad uno la serie $\sum_{n=1}^{\infty} M\xi_n$ se desprende además la afirmación exigida sobre la convergencia con probabilidad uno de la serie $\sum_{n=1}^{\infty} \xi_n$.

Supongamos, ahora, que las variables $\xi_1,\xi_2,\ldots$ son uniformemente acotadas y la serie $\sum_{n=1}^{\infty} \xi_n$ converge con probabilidad uno. Construyamos la sucesión de variables aleatorias independientes $\tilde{\xi}_1,\tilde{\xi}_2,\ldots$ las cuales tienen igual distribución que ξ_1, $\xi_2,\ldots$. Entonces con probabilidad uno converge también cada una de las series $\sum_{n=1}^{\infty} \tilde{\xi}_n$, $\sum_{n=1}^{\infty} (\xi_n-\tilde{\xi}_n)$.

Ya que $M(\xi_n-\tilde{\xi}_n) = 0$, entonces de la convergencia con probabilidad uno de la serie $\sum_{n=1}^{\infty} (\xi_n-\tilde{\xi}_n)$ y la desigualdad (4) se deduce que $\sum_{n=1}^{\infty} D(\xi_n-\tilde{\xi}_n) < \infty$. Lo cual indica que

$$\sum_{n=1}^{\infty} D\xi_n = \frac{1}{2} \sum_{n=1}^{\infty} D(\xi_n-\tilde{\xi}_n) < \infty .$$

En virtud de la primera parte del teorema con probabilidad uno converge la serie $\sum_{n=1}^{\infty}(\xi_n-\tilde{\xi}_n)$ que junto con la convergencia de la serie $\sum_{n=1}^{\infty}\xi_n$ permiten la convergencia de la serie $\sum_{n=1}^{\infty}M\xi_n$.

El teorema 4 queda demostrado.

Supongamos que

$$\xi_n^c = \begin{cases} \xi_n, & \text{si } |\xi_n| \leq c, \\ 0, & \text{si } |\xi_n| > c. \end{cases}$$

TEOREMA 5. *Para la convergencia de la serie*

$$\sum_{n=1}^{\infty} \xi_n$$

de variables aleatorias $\xi_1, \xi_2, \ldots$ *es necesario y suficiente la convergencia simultánea para un* $c > 0$ *de cada una de estas tres series*(1)

$$\sum_{n=1}^{\infty} P\{|\xi_n| > c\}, \qquad \sum_{n=1}^{\infty} M\xi_n^c, \qquad \sum_{n=1}^{\infty} D\xi_n^c.$$

Demostración. Suficiencia. Ya que $\sum_{n=1}^{\infty}P\{|\xi_n| > c\} < \infty$, entonces según el lema de Borel-Cantelly casi siempre $\xi_n = \xi_n^c$ para toda n, excluída posiblemente solo un número finito. Entonces por el teorema 4 converge la serie $\sum_{n=1}^{\infty}\xi_n$ con probabilidad uno, o sea converge y $\sum_{n=1}^{\infty}\xi_n$.

Necesidad. Si la serie $\sum_{n=1}^{\infty}\xi_n$ converge con probabilidad uno, entonces $\xi_n \to 0$ casi siempre, $n \to \infty$. Por eso para toda $c > 0$ puede ocurrir un gran número finito de sucesos $\{|\xi_n| > c\}$, o sea

(1) Comparar con A. Khintchine und A. Kolmogórov , Über Konvergenz vonRoihen, Manual Mat., Vol. 32 (1925), 668-677.

según el lema Borel-Cantelly $\sum_{n=1}^{\infty} P\{|\xi_n| > c\} < \infty$. Además de la convergencia de la serie $\sum_{n=1}^{\infty} \xi$ se deduce que la serie $\sum_{n=1}^{\infty} \xi_n^c$ también converge. Entonces la convergencia de las series $\sum_{n=1}^{\infty} M\xi_n^c$ y $\sum_{n=1}^{\infty} D\xi_n^c$ se deducen del teorema 4.

*

Ley de ceros y unos en la teoría de las probabilidades.

Vimos ya, varios casos en los cuales son conocidas las probabilidades extremales con la necesidad de ser iguales a cero o la unidad. Por ejemplo, la probabilidad de la convergencia de la serie de variables aleatorias independientes puede tomar solo estos dos valores(1). Demostraremos ahora, un teorea general que abarca muchos de estos casos.

TEOREMA (Ley de ceros y unos). Sea ξ_1, ξ_2, *variables aleatorias cualquiera, y* $f(x) = f(x), x_2, \ldots$ *la función de Borel*(2) *de las variables* $x = (x_1, x_2, \ldots$ *tal que la probabilidad condicional*

$$P\{f(\xi_1, \xi_2, \ldots) \equiv 0 \mid \xi_1, \xi_2, \ldots \xi_n\} ,$$

de la relación

---

(1) Ver Cap. VI §5. También tiene lugar para la probabilidad $P\{\eta_n - d_n \to 0,\ n \to \infty\}$ en la lev fuerte de los grandes números, por lo menos cuando las variables ξ_n son independientes en conjunto.

(2) Por función de Bera se entiende la función que puede ser representada a partir de polinomios, por pasos límites sucesivos integrados.

$$f(\xi_1,\xi_2,\ldots) = 0,$$

para los n primeras variables conocidas $\xi_1,\ldots,\xi_n$ *es igual a la probabilidad absoluta*

$$P\{f(\xi_1,\xi_2,\ldots) = 0\} \qquad (1)$$

para todo n. En estas condiciones la probabilidad (1) es igual a 1 o a 0.

En particular las unidades de probabilidad se cumplen si las variables aleatorias $\xi_1,\xi_2,\ldots$ *son independientes y los valores de la función* f x *son invariables su cambiamos solo un número finito de variables* x_n.

Demostración. Denotemos por $\xi = (\xi_1,\xi_2,\ldots)$ y sea

$$A = \{\omega : f(\xi) = 0\}$$

Junto con este suceso examinamos el álgebra $\mathfrak{N}$ de todos los sucesos, que pueden ser definidos por cualquier relación entre un número finito de variables ξ_n. Si el suceso B pertenece a $\mathfrak{N}$ entonces por las condiciones del teorema

$$P(A \mid B) = P(A). \qquad (2)$$

En el caso $P(A) = 0$ nuestro teorema esta demostrado.

Supongamos que $P(A) > 0$. Entonces de (2) se deduce la fórmula

$$P(B \mid A) = \frac{P(A|B)P(B)}{P(A)} = P(B) \qquad (3)$$

y así $P(B)$ y $P(B \mid A)$ son dos funciones de conjuntos finita-aditivas que coinciden en $\mathfrak{N}$, en consecuencia ellos deben ser iguales entre si en cada conjunto de Borel de la dilatación $\sigma(\mathfrak{N})$ del álgebra $\mathfrak{N}$. Por eso, en particular

$$P(A) = P(A \mid A) = 1 .$$

El teorema queda demostrado.

Algunos otros casos en los cuales se puede afirmar sobre probabilidades conocidas que ellas toman solo valores de 0 o 1 fueron descubiertos por P. Levi[1]

*

BIBLIOGRAFIA

[1] Bernstein, S.N., *Experimento de axiomatización de los fundamentos de la teoría de las probabilidades*, 1917.

[2] Bernstein, S.N., *Teoría de las probabilidades*, 2a. ed. Moscú, 1939.

[1] Borel, E., *Les probabilites denombrables et leurs applications arithmetiques*, Rend. Circ. mat., Palermo, T.27, 1909.

[2] Borel, E., *Principes et formules classiques*, Fasc. 1 du tome I dur Traité des probabilites par E. Borel et divers auteurs, Paris, cauthier-Villars, 1925.

[3] Borel, E., *Applications a l'arithmetique et a la theoria des fonctions*, Fasc. 1 du tome II du traite des probabilites par E. Borel et ldives auteurs, Paris, cauthier-Villors, 1920.

(1) Ver sobre esto a P. Lévy, Sur un théoreme de A. Khi tchine, Bull. des Sci. Math., V. 55 (1931), 145-160, Theorema 2.

[1] Cantelli, F.P., *Una teoría astralla del calcolo delle probabilita*, Giorn.Inst. Ital. Altuari, T.3 1932.

[2] Cantelli, F.P., *Sulla legge dei grandi numeri*, Mem. Acad. Linoei, T.11, 1916.

[3] Cantelli, F.P., *Sulla probabilita come limite della frenquenza*, Rend. Accad. Lincci, T.26, 1917.

[1] Copeland, H., *The theory of probability from the point of view of admissible numbers*, Ann. Math. Statist. T 3 1932. pp. 143-156

[1] Dörge, K., *Zu der von Mises gegebenen Begründung der Wahrschein lich Keitsrechnung*, Math.Z. T.32, 1930,pp.232-258.

[1] Frechet, M., *Sur la convergence en probabilite*, Metron, T.8, 1930.

[2] Frechet, M., *Recherches theoriques modernes, sur theorie des probabilites*, Fasc.3 du Tome I du traite des probabilites par E. Borel et divers auteurs, Paris, Gathier-Villars.

[1] Kolmogoroff, A., *A uber die analytischen methoden in der Wahrsheinlichkeitsrechnung*, Math. An.. T. 104, 1931.

[2] Kolmogoroff, A., *Teoria general de la medida y teoría de las probabilidades*, 1929.

[1] Levy, P., *Calcul des probabilites*, Paris, Gauthier-Villars, 1925.

[1] Lomnicki, A., *Nouveaux fondements du Calcul des probabilites*, Fundam. Math. T 4, 1923.

[1] Von Mises, R., *Wahrsheinlichkeitsrechning*, Leipzig, Wein, Fr. Deuticke, 1931.

[2] Von Mises, R., *Grundlagen der Wahrsheinlichkeitsrechnung*, Math., Z., T 5, 1919.

[3] Von Mises, R., *Wahrsheinlichtkeitsrchnung*, Statistik und wahrheit wein, julius Springer, 1928.

[1] Reichenbach, H., *Axiomatik der Wahrscheinlichkeitsrechnung* Math. Z., T 34, 1932, pp.3-89

[1] Slutsky, E., *Uber stochastishe asymptoten und grenzwerte,* Metron, T.5, 1925, pp. 3-89.

[2] Slutsky, E., *Acerca de las bases lógicas de la teoría de probabilidades,* Biestnik Statistiki, Tomo 12, 1922 pp. 13-21.

[1] Steinhauss, M., *Les probabilites denombrables et leur rapport a la theorie de la mesure,* Fundam. Math. T 4, 1923, pp. 286-310.

[1] Tornier, *Wahrscheinlichkeitsrechnung und zehlentheorie,* J. reine angew, Math., T 160, 1929.

[2] Tornier, *Grundlagen der Wahrscheinlichkeitsrechnung,* Acta Math. T 60, 1933, pp. 239-380.

*