

YURI ROZANOV

PROCESOS ALEATORIOS

curso resumido



EDITORIAL MIR · MOSCU



Ю. А. РОЗАНОВ

СЛУЧАЙНЫЕ ПРОЦЕССЫ

ИЗДАТЕЛЬСТВО «НАУКА»
МОСКВА

YU. A. ROZANOV

PROCESOS
ALEATORIOS

curso resumido

EDITORIAL MIR MOSCU

CDU 519.2=60

Traducción del ruso

Impreso en la URSS
1973

На испанском языке

© Traducción al español. Editorial Mir. 1973

P 0223—367
041 (01)—73

INDICE

Del autor	7
---------------------	---

CAPITULO I

INTRODUCCIÓN A LOS CONCEPTOS FUNDAMENTALES DE LA TEORÍA DE PROBABILIDADES

§ 1. Espacio de sucesos elementales y probabilidad	9
1. Prueba con resultados equiprobables (9). 2. Espacio de los sucesos elementales (15). 3. Propiedades fundamentales de la probabilidad. Aditividad y continuidad (19). 4. Concordancia entre el modelo y la prueba (23).	
§ 2. Independencia y probabilidad condicional	25
1. Concepto de independencia (25). 2. Probabilidad condicional (29).	
§ 3. Magnitudes aleatorias y distribución de las probabilidades. Independencia	33
1. Distribuciones discreta y continua (33). 2. Distribución conjunta de las probabilidades (35). 3. Transformaciones de las magnitudes aleatorias (39). 4. Distribuciones condicionales de probabilidades (42). 5. Magnitudes aleatorias multidimensionales (44).	
§ 4. Esperanzas matemáticas de las magnitudes aleatorias	45
1. Esperanza matemática, definición y algunas fórmulas (45). 2. Momentos, dispersión y desigualdad de Chébishev (51). 3. Esperanza matemática condicional (53). 4. Distancia media cuadrática y coeficiente de correlación (57). 5. Algunos teoremas sobre la convergencia (61).	
§ 5. Series infinitas de pruebas independientes y ley de los grandes números	
1. Ley de los grandes números (67). 2. Probabilidad y frecuencia (70).	

CAPITULO II

ALGUNAS DISTRIBUCIONES DE PROBABILIDADES

§ 1. Elección aleatoria y variación aleatoria	72
1. Algunas fórmulas combinatorias (72). 2. Distintas distribuciones de las partículas independientes en el espacio físico (76).	
§ 2. Distribución de Poisson. Corriente uniforme de sucesos y tiempo de espera	84
1. Distribución de Poisson en las partículas (84). 2. Tiempo de espera del suceso aleatorio (88).	
§ 3. Pruebas de Bernoulli. Movimiento browniano y distribuciones de probabilidades ligadas a éste	95
1. Pruebas de Bernoulli y distribución binomial. Aproximaciones de Poisson y normal (95). 2. Proceso del movimiento browniano. Distribución de probabilidades del máximo y momento de su consecución (100).	
§ 4. Distribución normal de probabilidades y distribuciones ligadas a ella	108
1. Distribución normal multidimensional (108). 2. Valuación de	

los parámetros de la distribución normal. La distribución χ^2 y la distribución de Student (115).

- § 5. **Distribución de probabilidades y funciones características** 121
 1. Funciones características y sus propiedades fundamentales (121).
 2. Convergencia de las distribuciones de probabilidades (127).

CAPITULO III

ALGUNOS MODELOS DE PROCESOS ALEATORIOS

- § 1. **Algunas definiciones y ejemplos** 135
 1. Definición general del proceso aleatorio (135). 2. Los procesos aleatorios de Márkov (136).
- § 2. **Cadenas de Márkov. Clasificación de los estados. Distribuciones estacionarias** 141
 1. Probabilidades de paso (141). 2. Estados reversibles e irreversibles (145). 3. Tiempo medio de estancia en un estado. Clasificación de los estados (149). 4. Teorema ergódico (convergencia hacia la distribución estacionaria) (153).
- § 3. **Cadenas de Márkov (tiempo continuo)** 161
 1. Ecuaciones diferenciales para las probabilidades de paso (161).
 2. Coeficiente de ergodicidad y convergencia hacia la distribución estacionaria (167).
- § 4. **Procesos que se ramifican** 170
 1. Ecuación diferencial de las funciones generatrices (170). 2. Efectos de degeneración y explosión (176).
- § 5. **Algunos procesos de servicios en masa y fluctuaciones aleatorias** 177
 1. Procesos de restablecimiento (177). 2. Sucesión de sumas de las magnitudes independientes. Distribución del máximo (183). 3. Procesos aleatorios en los sistemas con una línea de servicio (191).
- § 6. **Procesos aleatorios en los sistemas lineales** 199
 1. Algunas observaciones auxiliares (199). 2. Integral estocástica (202). 3. Convergencia hacia el proceso estacionario (207). 4. Procesos de efecto fraccionario (209).
- § 7. **Procesos aleatorios estacionarios** 214
 1. Representación espectral de los procesos estacionarios y transformación de Fourier (214). 2. Transformaciones lineales. Ejemplos (224).
- § 8. **Procesos de difusión** 232
 1. Procesos aleatorios representados por la integral estocástica Ito (232). 2. Ecuaciones diferenciales de Kolmogórov (245).

CAPITULO IV

ALGUNAS TAREAS DE PRONOSTICACIÓN, FILTRADO Y REGULACIÓN DE LOS PROCESOS ALEATORIOS

- § 1. **Tarea general sobre la aproximación óptima. Ejemplos** 252
- § 2. **Pronosticación y filtrado de los procesos aleatorios estacionarios** 259
 1. Tarea de pronosticación lineal (259). 2. Filtrado lineal (valuación del valor medio) (263).
- § 3. **Esperanzas matemáticas condicionales y algunas tareas de pronosticación y de filtrado** 271
 1. Una vez más sobre las esperanzas matemáticas condicionales (271). 2. Papel de las probabilidades a posteriori en algunas tareas de pronosticación y filtrado (276).
- Índice temático** 283

DEL AUTOR

Este libro ha sido escrito paralelamente al Curso de lecciones sobre la teoría de los procesos aleatorios que expuse en el Instituto físico-técnico de Moscú en los años 1969-1970. Tengo la esperanza de que será útil, en uno u otro grado, para todo el que desee dominar los resultados y métodos fundamentales de esta teoría.

Quisiera expresar aquí mi reconocimiento a Yu. V. Prójorov que leyó el manuscrito del libro e hizo una serie de observaciones críticas útiles, que me obligaron a examinar otra vez, más detalladamente, todo el contenido e introducir muchos cambios. Mi reconocimiento a O. V. Viskov que tomó en sí el trabajo de redacción del libro.

Yu. A. Rózanov

CAPITULO I

INTRODUCCIÓN A LOS CONCEPTOS FUNDAMENTALES DE LA TEORÍA DE PROBABILIDADES

§ 1. ESPACIO DE SUCEOS ELEMENTALES Y PROBABILIDAD

1. Prueba con resultados equiprobables. Empezamos nuestro curso con la descripción de algunas pruebas elementales, en las cuales el resultado no se puede de antemano predeterminar con certidumbre, sino sólo se puede hablar sobre la probabilidad de tal o cual resultado.

Lanzamiento de la moneda. A esta prueba se recurre, corrientemente, al realizar un sorteo, cuando la moneda simétrica se lanza con un fuerte giro y como resultado cae de forma aleatoria «cara» o «cruz» (fig. 1).

Lanzamiento del dado de juego. Este ensayo es una parte componente del juego popular infantil conocido por el nombre «Quien va despacio, va más lejos», «Circo», etc., en el cual, después del lanzamiento correspondiente del cubo (dado de juego) regular con las caras numeradas, en dependencia del número de puntos que han caído, el jugador desplaza su figura (según las reglas de juego) en el número correspondiente de pasos (casillas). Durante el lanzamiento del dado de juego cae una u otra cara, con tal o cual número de puntos $a = 1, 2, \dots, 6$



Fig. 1.

Algo más compleja es la prueba que se realiza en el conocido juego de azar cuando cada uno de los jugadores lanza por turno dos dados de juego, y gana el que tiene la suma mayor de puntos. Al lanzar dos dados de juego cae de una forma aleatoria una de las combinaciones posibles (a, b) , donde a significa el número de puntos que han caído en el primer dado, y b el número de puntos en el segundo dado.

Juego a la ruleta. Imagínese un disco pesado, dividido en N sectores regulares, que se encuentra en posición horizontal y puede girar fácilmente alrededor de su eje; a lo largo de la circunferencia por el borde del disco se tiene un hundimiento (canal) uniforme en el cual se encuentra una bolita pequeña que se desplaza libremente (fig. 2). A cada paso por separado se le comunica al disco una rotación fuerte, con la cual rueda la bolita por el canal. Después de pararse el disco, se para también la bolita, cayendo en uno de los sectores (designados en el disco con los números desde 1 hasta N). En dependencia del número aleatorio que cayó $v = 1, \dots, N$ y de la «puesta hecha» el jugador recibe el premio correspondiente.

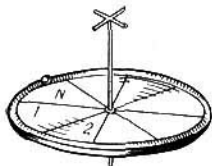


Fig. 2.

Con relación a cada una de las pruebas descritas anteriormente (lanzamiento de la moneda o del dado de juego, el «lanzamiento» de la bolita en el juego a la ruleta), se puede decir lo siguiente: primero, el resultado de la prueba es aleatorio; segundo, se tiene un número finito de resultados

diferentes que se excluyen mutuamente uno al otro; tercero, todos los resultados son equiprobables.

Cualquier prueba de tal naturaleza la denominaremos *prueba con un número finito de resultados equiprobables*, llamando a estos resultados *sucesos elementales* y designándolos en lo sucesivo por ω .

Designemos por N el número total de sucesos elementales. De acuerdo a que todos estos sucesos son equiprobables, consideraremos que la probabilidad de cada uno de ellos es igual a $1/N$.

Considerando tal o cual suceso A , que se realiza (o no se realiza) en dependencia del resultado aleatorio ω , designemos por $N(A)$ el número total de resultados elementales ω , que conducen a la realización de este suceso. Está claro, que el suceso A será tanto más probable¹⁾ cuanto mayor sea el número $N(A)$. Consideramos que la probabilidad del suceso A , representándola por $P(A)$, es proporcional a la magnitud $N(A)$; más exactamente definimos a esta probabilidad así:

$$P(A) = \frac{N(A)}{N}. \quad (1.1)$$

Por ejemplo, la probabilidad de que caiga «cara» al lanzar la moneda es igual a $1/2$; la probabilidad del suceso de que «al lanzar el dado de juego cae un número par de puntos» también es igual a $1/2$; al lanzar dos dados de juego la probabilidad del suceso A de que

¹⁾ La expresión del tipo «el suceso A es más probable que el suceso B », que se comprende intuitivamente debe ser interpretada, a nuestro parecer, en el sentido de que, cuando la prueba se repita reiteradamente el suceso A se verificará más frecuentemente, que el suceso B .

«la suma de los puntos que cayeron sea mayor de 10» es igual a $1/12$, ya que el número de todos los resultados elementales posibles ω , es decir, de combinaciones distintas de puntos (a, b) , donde $a, b = 1, \dots, 6$, es $N = 36$, y el número de resultados, para los cuales se realiza el suceso A (es decir, cuando $a + b > 10$), es $N(A) = 3$, ya que al suceso A sólo llevan los resultados $(5, 6)$, $(6, 5)$, $(6, 6)$.

Volvamos a la prueba con el disco giratorio, sobre el cual se encuentra la bolita (véase «Juego a la ruleta»). Después de pararse el disco la bolita ocupa una posición determinada, que se puede describir por la coordenada angular ω , $0 \leq \omega \leq 2\pi$, si consideramos la bolita como un punto sobre la circunferencia.

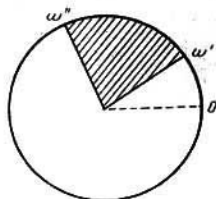


Fig. 3.

El punto aleatorio ω cae en cualquiera de los N sectores iguales, en que está dividido el disco, con una misma probabilidad $1/N$. Considerando que la probabilidad de que caiga en cualquiera de los sectores $A = [\omega', \omega'']$ (fig. 3) es una misma para todos los sectores con la misma distancia angular $\omega'' - \omega'$, determinemos esta probabilidad $P(A)$ así:

$$P(A) = \frac{\omega'' - \omega'}{2\pi}.$$

Luego, desarrollando la circunferencia (de longitud L) en el segmento $[0, L]$ de la recta real $-\infty < x < \infty$, sobre el punto $\xi = \frac{\omega}{2\pi} L$ diremos que, en dependencia del caso, cae en tal o cual intervalo $[x', x'']$ de este segmento, y precisamente, cae en cualquier intervalo de longitud $x'' - x'$ con una misma probabilidad igual a $\frac{x'' - x'}{L}$.

El valor ξ que es la coordenada del punto considerado depende del caso, hablando de otra forma, es una *magnitud aleatoria*. La probabilidad de que ella tome un valor en el intervalo $[x', x'']$ (designado más adelante por $P\{x' \leq \xi \leq x''\}$) se puede determinar por la fórmula siguiente:

$$P\{x' \leq \xi \leq x''\} = \int_{x'}^{x''} p_{\xi}(x) dx, \quad (1.2)$$

donde la función subintegral $p_{\xi}(x)$, llamada en adelante *densidad de probabilidad* para la magnitud casual ξ que consideramos, es

$$p_{\xi}(x) = \begin{cases} \frac{1}{L} & \text{para } 0 \leq x \leq L, \\ 0 & \text{para } x < 0, x > L. \end{cases} \quad (1.3)$$

En el ejemplo expuesto la magnitud ξ está distribuida uniformemente en el segmento $[0, L]$.

Sea $y = \varphi(x)$ una función cualquiera en el segmento $[0, L]$, que tiene una derivada positiva $\varphi'(x)$ continua a trozos; ella transforma biunívocamente el segmento $[0, L]$ en otro determinado segmento $[a, b]$ (que puede incluso ser infinito) (fig. 4). Examinemos la nueva

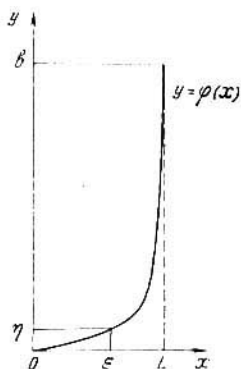


Fig. 4. Se ve fácilmente que para la función $y = \varphi(x)$ de la forma indicada el punto aleatorio $\eta = \varphi(\xi)$ cae con gran probabilidad en la mitad inferior del segmento $[a, b]$, $a = 0$.

magnitud aleatoria $\eta = \varphi(\xi)$. ¿Qué probabilidad existe de que el valor de η descansa en el intervalo $\{y' \leq \eta \leq y''\}$?

Designemos por $x = \psi(y)$ la función en el segmento $[a, b]$, que es inversa de la función inicial $y = \varphi(x)$, $0 \leq x \leq L$. Evidentemente, los sucesos $\{y' \leq \eta \leq y''\}$ y $\{\psi(y') \leq \xi \leq \psi(y'')\}$ coinciden, y la probabilidad del suceso $\{\psi(y') \leq \xi \leq \psi(y'')\}$, de acuerdo a la fórmula

(1.2) es igual a $\int_{\psi(y')}^{\psi(y'')} p_{\xi}(x) dx$. Utilizando

el cambio de variable $x = \psi(y)$, obtenemos que la probabilidad buscada del suceso $\{y' \leq \eta \leq y''\}$ es

$$\begin{aligned} P\{y' \leq \eta \leq y''\} &= \int_{\psi(y')}^{\psi(y'')} p_{\xi}(x) dx = \\ &= \int_{y'}^{y''} p_{\xi}[\psi(y)] \psi'(y) dy = \int_{y'}^{y''} p_{\eta}(y) dy, \end{aligned}$$

donde la función $p_{\eta}(y)$ tiene la forma

$$p_{\eta}(y) = \begin{cases} p_{\xi}[\psi(y)] \frac{1}{\varphi'[\psi(y)]} & \text{para } a \leq y \leq b, \\ 0 & \text{para } y < a, y > b, \end{cases} \quad (1.4)$$

y en nuestro caso (cuando $p_{\xi}(x) = \frac{1}{L}$, $0 \leq x \leq L$)

$$p_{\eta}(y) = \frac{1}{L} \cdot \frac{1}{\varphi'[\psi(y)]}, \quad a \leq y \leq b.$$

Vemos que la magnitud aleatoria η puede tener una densidad de probabilidad muy compleja $p_{\eta}(y)$, $-\infty < y < \infty$, con ayuda de la cual la probabilidad de caer en tal o cual intervalo $[y', y'']$ se determina como

$$P\{y' \leq \eta \leq y''\} = \int_{y'}^{y''} p_{\eta}(y) dy \quad (p_{\eta}(y) \geq 0, \int_{-\infty}^{\infty} p_{\eta}(y) dy = 1)$$

Supongamos ahora que tenemos dos magnitudes aleatorias ξ_1 y ξ_2 , análogas a la magnitud aleatoria ξ examinada anteriormente, que no están ligadas entre sí de ningún modo. Digamos, el punto ξ_1 se lanza «al azar» sobre el segmento $[0, L_1]$ del eje $-\infty < x_1 < \infty$, y el punto ξ_2 , sobre el segmento $[0, L_2]$ del eje $-\infty < x_2 < \infty$ del sistema de coordenadas cartesianas en el plano.

Considerando a (ξ_1, ξ_2) como un punto aleatorio en el plano con las coordenadas ξ_1 y ξ_2 , análogamente a la fórmula (1.2) determinamos la probabilidad de caída de este punto en una u otra zona A (con frontera lisa a trozos) del modo siguiente:

$$P(A) = \iint_A p_{\xi_1 \xi_2}(x_1, x_2) dx_1 dx_2 \quad (1.5)$$

donde la función subintegral $p_{\xi_1 \xi_2}(x_1, x_2)$ es

$$p_{\xi_1 \xi_2}(x_1, x_2) = \begin{cases} \frac{1}{L_1 L_2} & \text{para } 0 \leq x_1 \leq L_1, 0 \leq x_2 \leq L_2, \\ 0 & \text{para los restantes } x_1, x_2. \end{cases} \quad (1.6)$$

Para la zona A del rectángulo $\Omega = \{0 \leq x_1 \leq L_1, 0 \leq x_2 \leq L_2\}$ donde, solamente, cae el punto (ξ_1, ξ_2) la probabilidad $P(A)$ de caída en A determinada por la fórmula (1.5) es numéricamente igual a la relación entre el área de la zona A y el área de todo el rectángulo Ω .

Ejemplo. Supongamos que lanzamos al azar sobre el segmento de longitud L dos puntos independientemente uno del otro. ¿Qué probabilidad hay de que la distancia entre ellos no sea mayor que l ?

Designemos por ξ_1, ξ_2 las coordenadas de estos puntos en el segmento $[0, L]$. Luego, tomemos ξ_1 en el segmento $(0, L)$ del eje x_1 , ξ_2 en el segmento $(0, L)$ del eje x_2 y consideremos que el punto (ξ_1, ξ_2) se lanza al azar en el cuadrado $\Omega = \{0 \leq x_1 \leq L, 0 \leq x_2 \leq L\}$. Evidentemente, la probabilidad buscada de que $\Delta = |\xi_1 - \xi_2| \leq l$, coincide con la probabilidad $P(A)$ de caída del punto (ξ_1, ξ_2) en la zona A , limitada por las rectas $x_2 = l + x_1$ y $x_2 = -l + x_1$ (véase la fig. 5, donde esta zona ha quedado sin rayado). De tal modo, de acuerdo con la fórmula (1.5) la probabilidad $P(A)$ puede ser expresada así:

$$P(A) = \iint_A \frac{1}{L^2} dx_1 dx_2 = \frac{2Ll - l^2}{L^2},$$

donde $2Ll - l^2$ es el área de la zona A .

Ejemplo. (Tarea de Buffon sobre la aguja). Supongamos que en el plano rayado con líneas paralelas, que están una de otra a la distancia L , se lanza al azar una «aguja», o sea un segmento de longitud l , $l \leq L$. ¿Cuál será la probabilidad de que el segmento lanzado corte a una de las líneas existentes?

Designemos por ξ_1 el ángulo de inclinación del segmento «caído» respecto a la dirección de las líneas, por ξ_2 la distancia de su extremo inferior hasta la línea más próxima de arriba (fig. 6, a). Las condiciones de la prueba son tales que la magnitud aleatoria ξ_1 está distribuida uniformemente, en el segmento $[0, \pi]$, lo mismo que la magnitud aleatoria ξ_2 en el segmento $[0, L]$. Intuitivamente está claro que las magnitudes ξ_1 y ξ_2 no están ligadas entre sí; consideramos que el punto (ξ_1, ξ_2) cae en tal o cual zona A en el plano con la misma probabilidad, como si este punto (ξ_1, ξ_2) se lanzase al azar sobre el rectángulo $\Omega = \{0 \leq$

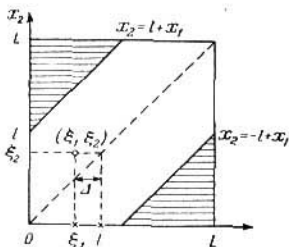


Fig. 5.

$\leq x_1 \leq \pi, 0 \leq x_2 \leq L\}$. Evidentemente, el segmento de longitud l corta una de las líneas cuando, y sólo cuando

$$\xi_2 \leq l \operatorname{sen} \xi_1,$$

es decir, cuando el punto correspondiente (ξ_1, ξ_2) caiga en la zona A limitada por arriba con la curva $x_2 = l \operatorname{sen} x_1$ (fig. 6, b). De acuerdo

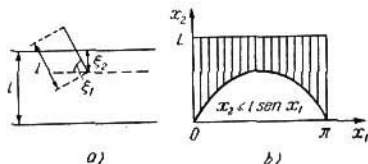


Fig. 6.

a nuestra fórmula propuesta (1.5), la probabilidad de caída del punto (ξ_1, ξ_2) en esta zona será

$$P(A) = \iint_A \frac{1}{\pi L} dx_1 dx_2 = \frac{2l}{\pi L}.$$

Tomemos un par de magnitudes aleatorias arbitrarias ξ_1 y ξ_2 para las cuales la probabilidad de caída del punto (ξ_1, ξ_2) en tal o cual zona A está determinada por la igualdad (1.5), donde $p_{\xi_1 \xi_2}(x_1, x_2)$, $-\infty < x_1, x_2 < \infty$, es una función no negativa, que satisface a la condición $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p_{\xi_1 \xi_2}(x_1, x_2) dx_1 dx_2 = 1$. A esta

función la llamaremos *densidad de probabilidad* de las magnitudes aleatorias ξ_1, ξ_2 .

Supongamos, que nos interesan determinadas magnitudes $\eta_1 = \varphi_1(\xi_1, \xi_2)$ y $\eta_2 = \varphi_2(\xi_1, \xi_2)$, que son funciones de las magnitudes aleatorias iniciales ξ_1, ξ_2 . Sea

$$y_1 = \varphi_1(x_1, x_2), \quad y_2 = \varphi_2(x_1, x_2)$$

la transformación biunívoca del plano con un jacobiano distinto de cero

$$|I| = \begin{vmatrix} \frac{\partial \varphi_1}{\partial x_1} & \frac{\partial \varphi_1}{\partial x_2} \\ \frac{\partial \varphi_2}{\partial x_1} & \frac{\partial \varphi_2}{\partial x_2} \end{vmatrix}$$

y

$$x_1 = \psi_1(y_1, y_2), \quad x_2 = \psi_2(y_1, y_2)$$

la transformación inversa.

En la transformación examinada, la zona con el límite liso a trozos se transforma en una zona del mismo tipo. Para cualquier zona B , la caída del punto (η_1, η_2) en ella significa la caída del punto correspondiente (ξ_1, ξ_2) en la zona A , que es la imagen recíproca de B en la representación $y_1 = \varphi_1(x_1, x_2)$, $y_2 = \varphi_2(x_1, x_2)$, es decir, los sucesos $\{(\eta_1, \eta_2) \in B\}$ y $\{(\xi_1, \xi_2) \in A\}$ coinciden. Por consiguiente, la probabilidad $\mathbf{P}(B)$ de caída del punto (η_1, η_2) en la zona B , de acuerdo con la fórmula (1.5) será

$$\mathbf{P}(B) = \int_A p_{\xi_1, \xi_2}(x_1, x_2) dx_1 dx_2,$$

de donde después del cambio de variables $x_1 = \psi_1(y_1, y_2)$, $x_2 = \psi_2(y_1, y_2)$ obtenemos la fórmula del mismo tipo que (1.5):

$$\mathbf{P}(B) = \int_B p_{\eta_1, \eta_2}(y_1, y_2) dy_1 dy_2,$$

donde

$$p_{\eta_1, \eta_2}(y_1, y_2) = p_{\xi_1, \xi_2}[\psi_1(y_1, y_2), \psi_2(y_1, y_2)] |I|^{-1} \quad (1.7)$$

es, según la definición dada anteriormente, la *densidad de probabilidad* de las magnitudes aleatorias η_1, η_2 :

$$p_{\eta_1, \eta_2}(y_1, y_2) \geq 0, \quad \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p_{\eta_1, \eta_2}(y_1, y_2) dy_1 dy_2 = 1.$$

2. Espacio de los sucesos elementales. Al examinar un par de magnitudes aleatorias ξ_1, ξ_2 con una densidad de probabilidad $p_{\xi_1, \xi_2}(x_1, x_2)$ consideramos que nos ocupamos con una prueba donde se observa

un punto aleatorio (ξ_1, ξ_2) . Los «resultados elementales» de tal prueba consistente cada uno de ellos en que las magnitudes aleatorias ξ_1, ξ_2 que toman determinados valores x_1, x_2 pueden ser descritos por los puntos correspondientes (x_1, x_2) en el plano, y todos los sucesos posibles del tipo $\{(\xi_1, \xi_2) \in A\}$, o sea, «el punto aleatorio (ξ_1, ξ_2) cae en la zona A », en su esencia se describen por los conjuntos correspondientes A . Sin detenernos en los detalles físicos de las condiciones de la prueba, se puede describir el «mecanismo aleatorio» que actúa sobre ella, determinando las probabilidades $P(A)$ de todos los sucesos posibles A (véase la fórmula (1.5).)

En este ejemplo claramente se manifiestan todos los principales rasgos del modelo teórico-probabilístico que es en un amplio sentido una prueba con resultados aleatorios. Al describir tal modelo se indican todos los resultados elementales posibles ω , que satisfacen a la siguiente exigencia: al final de la prueba se obtiene uno y sólo uno de estos resultados (ellos también se llaman *sucesos elementales* y todo el conjunto Ω se llama *espacio de los sucesos elementales*). Además de esto, se describe de un modo correspondiente el mecanismo de aleatoriedad en la prueba dada, es decir, se dan precisamente las probabilidades $P(A)$.

$$0 \leq P(A) \leq 1, \quad (1.8.)$$

de tales o cuales sucesos A ; con ello la probabilidad del suceso cierto se considera igual a 1, la probabilidad del suceso imposible es igual a 0 (más adelante se hablará con más detalle de qué modo se dan las probabilidades).

Cualquier suceso A tomado por separado sólo nos interesa desde el punto de vista si ocurre como resultado de la prueba dada o no ocurre; desde este punto de vista podemos no hacer diferencia entre los sucesos A_1 y A_2 , para los cuales la aparición de A_1 también significa la realización de A_2 , y viceversa, la aparición de A_2 también significa la realización de A_1 .

Si examinamos el suceso A^1 ligado con la prueba Ω (como se hizo en ejemplos descritos anteriormente) se puede separar el conjunto de todos aquellos resultados elementales $\omega \in \Omega$, que conducen a la aparición del suceso A ; designemos este conjunto por el mismo símbolo A , como en el caso del suceso examinado. Es evidente que la aparición del suceso $\{\omega \in A\}$, que consiste en que el resultado ω de nuestra prueba que entra en el conjunto separado $A \subseteq \Omega$ significa exactamente la aparición del propio suceso A .

En efecto, si ocurre el suceso A , entonces tiene lugar algún resultado ω que conduce a la aparición de este suceso, y todos estos resultados componen el conjunto $A \subseteq \Omega$, de modo que $\omega \in A$; por otro

¹⁾ Es decir, aquel suceso que aparece (o no aparece) a base del resultado aleatorio $\omega \in \Omega$.

lado, si ocurre el suceso $\{\omega \in A\}$, es decir, si tiene lugar cualquier resultado elemental $\omega \in A$, entonces según la propia definición del conjunto $A \subseteq \Omega$, esto significa que el resultado dado ω conduce a la aparición del suceso A .

Así pues, los sucesos A y $\{\omega \in A\}$ coinciden de hecho, y con ello, cualquier suceso ligado con la prueba examinada Ω puede ser descrito por el conjunto correspondiente A de resultados elementales $\omega \in \Omega$, cada uno de los cuales conduce a la aparición de este suceso. Se puede identificar convencionalmente el suceso A con el correspondiente conjunto $A \subseteq \Omega$. (Señalemos una vez más, que A , como un conjunto determinado de resultados elementales ω de la prueba Ω examinada, significa el suceso que ocurre cuando, y sólo cuando se realiza cualquier resultado elemental ω que pertenece al conjunto indicado $A \subseteq \Omega$.) El suceso cierto A que se realiza para cualquier resultado elemental $\omega \in \Omega$, se puede identificar con todo el espacio de los sucesos elementales Ω ; es cómodo introducir también el suceso imposible, identificándolo con el conjunto vacío \emptyset .

Tal tratamiento desde el punto de vista de la teoría de conjuntos permite determinar, claramente, las distintas operaciones sobre los sucesos. Se llama precisamente *unión* o *suma de los sucesos* A_1 y A_2 al suceso $A_1 \cup A_2$, que significa que ocurre aunque sea uno de los sucesos A_1 o A_2 (el suceso $A_1 \cup A_2$, como conjunto de resultados elementales $\omega \in \Omega$, es la unión de los conjuntos correspondientes A_1 y A_2); análogamente se determina la unión $\bigcup_k A_k$ de los sucesos

A_1, A_2, \dots . Se llama *intersección* o *producto de los sucesos* A_1 y A_2 el suceso $A_1 \cap A_2$ (designado también así: $A_1 \cdot A_2$), que consiste en que se realizan ambos sucesos A_1 y A_2 (como un conjunto en el espacio de los sucesos elementales Ω , el suceso $A_1 \cap A_2$ es la intersección de los conjuntos correspondientes $A_1, A_2 \subseteq \Omega$); análogamente se determina la intersección $\bigcap_k A_k$ de muchos sucesos A_1, A_2, \dots .

Se llama *complementario* del suceso A el suceso \bar{A} , que significa que el propio suceso A no se realiza (\bar{A} es el complemento del conjunto A en el espacio de los sucesos elementales Ω). Se llama *diferencia de los sucesos* A_1 y A_2 el suceso $A_1 \setminus A_2$, que significa que el suceso A_1 se realiza y el suceso A_2 no se realiza ($A_1 \setminus A_2$ es la diferencia de los conjuntos $A_1, A_2 \subseteq \Omega$). Luego, llamamos *incompatibles* o *mutuamente excluyentes* los sucesos A_1 y A_2 si ellos se excluyen mutuamente el uno al otro: cuando aparece A_1 no se realiza el suceso A_2 y viceversa, cuando aparece A_2 no se realiza el suceso A_1 (A_1 y A_2 como conjuntos en el espacio Ω no se cortan).

En el lenguaje de la teoría de conjuntos se pueden expresar otras relaciones entre los sucesos: por ejemplo, en vez de decir «el suceso A_1 lleva consigo al suceso A_2 », se puede decir que «el suceso A_1 pertenece al suceso A_2 », ya que uno y otro significan que al aparecer

cualquier resultado elemental $\omega \in \Omega$, que conduce al suceso A_1 ($\omega \in A_1$), también aparece el suceso A_2 (es decir, $\omega \in A_2$), y en la terminología de la teoría de conjuntos esto significa, que el conjunto A_1 pertenece al conjunto correspondiente A_2 : $A_1 \subseteq A_2$.

Algunas de las relaciones entre los distintos sucesos, indicadas anteriormente, están representadas gráficamente en la fig. 7, donde los sucesos elementales ω son los puntos del cuadrado Ω .

Como ejemplo, volvemos de nuevo a la prueba del lanzamiento de dos dados de juego. El suceso A , «cae una suma par de puntos»,

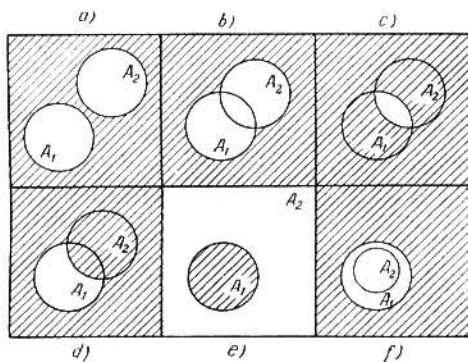


Fig. 7. a) A_1 y A_2 son sucesos incompatibles (que no se cortan. b) La figura sin rayado representa la unión $A_1 \cup A_2$. c) La figura sin rayado representa la intersección $A_1 \cap A_2$. d) La figura sin rayado representa la diferencia $A_1 \setminus A_2$. e) El suceso A_2 es complementario del suceso A_1 . f) El suceso A_1 contiene al suceso A_2 .

es la unión de los sucesos que no se cortan A_1 , «en cada dado cae un número par de puntos» y los sucesos A_2 , «en cada dado cae un número impar de puntos»; siendo $A_1 = A \setminus A_2$ y $A_2 = A \setminus A_1$. El suceso \bar{A} «cae una suma impar de puntos», es complementario del suceso A . El suceso \bar{A}_1 «cae, aunque sea, en un dado un número impar de puntos» es complementario de A_1 . El suceso \bar{A}_2 — «cae, aunque sea, en un dado un número par de puntos» es el suceso complementario de A_2 ; siendo $\bar{A}_1 \setminus \bar{A} = \bar{A}_1 \cap A = A_2$ y $\bar{A}_2 \setminus \bar{A} = \bar{A}_2 \cap A = A_1$.

No es difícil comprender que existe la siguiente ley general en los enlaces entre los distintos sucesos. Precisamente, si $A_1 \subseteq A_2$, entonces $\bar{A}_2 \subseteq \bar{A}_1$; si $A = A_1 \cup A_2$, entonces $\bar{A} = \bar{A}_1 \cap \bar{A}_2$; si $A = A_1 \cap A_2$, entonces $\bar{A} = \bar{A}_1 \cup \bar{A}_2$. En general, si es justa una relación determinada, expresada por el signo de igualdad =, inclu-

sión \subseteq (o \supseteq), unión \cup e intersección \cap , entonces será también justa la relación, que se obtiene de la inicial sustituyendo los signos de inclusión (\subseteq y \supseteq) por sus inversos (\supseteq y \subseteq), la unión \cup por la intersección \cap y viceversa, la intersección \cap por la unión \cup , sustituyendo al mismo tiempo los sucesos correspondientes por sus complementarios. Por ejemplo, son equivalentes las siguientes relaciones

$$\begin{aligned}\bar{A} &= \bar{A}_1 \cup (A_1 \cap A_2), & A &= \overline{\bar{A}_1 \cup (A_1 \cap A_2)}, \\ A &= A_1 \cap (\bar{A}_1 \cap A_2), & A &= A_1 \cap (\bar{A}_1 \cup \bar{A}_2)\end{aligned}$$

(observamos que aquí A coincide con la diferencia de los sucesos A_1, A_2 es decir, $A = A_1 \setminus A_2$).

3. Propiedades fundamentales de la probabilidad. Aditividad y continuidad. Examinemos cualesquiera sucesos que no se cortan A_1 y A_2 , que aparecen con las probabilidades $\mathbf{P}(A_1)$ y $\mathbf{P}(A_2)$. ¿Cuál es la probabilidad del suceso $A_1 \cup A_2$, que representa su unión?

Si los sucesos examinados ligados a la prueba tienen un número finito de resultados equiprobables, entonces de acuerdo con la fórmula general (1.1) la propiedad buscada será

$$\mathbf{P}(A_1 \cup A_2) = \mathbf{P}(A_1) + \mathbf{P}(A_2), \quad (1.9)$$

ya que el suceso $A_1 \cup A_2$ es la unión de todos los resultados elementales, que entran en A_1 o en A_2 , de modo que su número es $N(A_1 \cup A_2) = N(A_1) + N(A_2)$ y

$$\frac{N(A_1 \cup A_2)}{N} = \frac{N(A_1)}{N} + \frac{N(A_2)}{N}.$$

A esa misma igualdad (1.9) llegamos, considerando a los sucesos del tipo $\{(\xi_1, \xi_2) \in A\}$, ligados con la observación de ciertas magnitudes aleatorias ξ_1 y ξ_2 , cuando cada suceso de tal género se describe por el correspondiente conjunto A sobre el plano, y los sucesos que no se cortan A_1 y A_2 son en este sentido conjuntos que no se cortan. Precisamente, según la fórmula (1.5)

$$\begin{aligned}\mathbf{P}(A_1 \cup A_2) &= \int_{A_1 \cup A_2} \int p_{\xi_1 \xi_2}(x_1, x_2) dx_1 dx_2 = \\ &= \int_{A_1} \int p_{\xi_1 \xi_2}(x_1, x_2) dx_1 dx_2 + \int_{A_2} \int p_{\xi_1 \xi_2}(x_1, x_2) dx_1 dx_2 = \mathbf{P}(A_1) + \mathbf{P}(A_2).\end{aligned}$$

En el modelo teórico-probabilístico la relación (1.9) para los sucesos que no se cortan A_1 y A_2 , se admite en calidad de axioma; o sea, la llamada *ley de adición de las probabilidades*.

La propiedad de la aditividad expresada en ella es completamente análoga a las propiedades de longitud, superficie, volumen, etc. Debido a esta propiedad, por ejemplo, el suceso complementario \bar{A}

de cualquier suceso A con probabilidad $\mathbf{P}(A)$ tiene la probabilidad

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

(la probabilidad del suceso cierto $A \cup \bar{A}$ es igual a 1);

$$\mathbf{P}(A_1) \leq \mathbf{P}(A_2) \quad \text{para} \quad A_1 \subseteq A_2,$$

$$\mathbf{P}(A_1 \cup A_2) = \mathbf{P}(A_1) + \mathbf{P}(A_2) - \mathbf{P}(A_1 \cap A_2),$$

para cualesquiera sucesos A_1 y A_2 , etc.

Está claro, que la ley de adición de las probabilidades se extiende a cualquier número finito de sucesos que no se cortan A_1, A_2, \dots, A_n ; precisamente, aplicando sucesivamente la fórmula (1.9) los pares de sucesos que no se cortan A_1 y

$(\bigcup_{h=2}^n A_h), A_2$ y $(\bigcup_{h=3}^n A_h), \dots, A_{n-1}$ y A_n obtenemos como resultado

$$\mathbf{P}\left(\bigcup_{h=1}^n A_h\right) = \sum_{h=1}^n \mathbf{P}(A_h). \quad (1.10)$$

Imaginémonos luego que se tiene una sucesión infinita de sucesos, cada uno de los cuales arrastra consigo a todos los anteriores:

$$A_1 \supseteq A_2 \supseteq \dots \supseteq A_{n-1} \supseteq A_n \supseteq \dots$$

¿Qué probabilidad hay de que se realicen todos estos sucesos A_1, A_2, \dots ?

Según las condiciones, la realización de todos los sucesos A_1, A_2, \dots, A_n es equivalente a la aparición del suceso A_n , de modo que $\bigcap_{h=1}^n A_h = A_n$ y $\mathbf{P}\left(\bigcap_{h=1}^n A_h\right) = \mathbf{P}(A_n)$. Ya que $\mathbf{P}(A_{n+1}) \leq \mathbf{P}(A_n)$ cuando $A_{n+1} \subseteq A_n$, para la sucesión monótona decreciente de probabilidades $\mathbf{P}(A_n)$, $n = 1, 2, \dots$, existe el límite $\lim_{n \rightarrow \infty} \mathbf{P}(A_n)$. En el esquema general de la teoría de probabilidades se supone que este límite coincide con la probabilidad de realización de todos estos sucesos A_1, A_2, \dots , es decir, con la probabilidad del suceso $A = \bigcap_{h=1}^{\infty} A_h$:

$$\mathbf{P}(A) = \lim_{n \rightarrow \infty} \mathbf{P}(A_n); \quad (1.11)$$

ésta es la propiedad llamada *continuidad de la probabilidad*.

Demostremos que la aditividad y la continuidad expresadas respectivamente por las relaciones (1.10) y (1.11) en su conjunto tienen el mismo significado que la *aditividad numerable*: para cualquier número (finito o numerable) de sucesos que no se cortan A_1, A_2, \dots

$$\mathbf{P}\left(\bigcup_h A_h\right) = \sum_h \mathbf{P}(A_h). \quad (1.12)$$

En efecto, para cualesquiera de los sucesos que no se cortan $A_1, A_2, \dots, A_n, \dots$, los sucesos $B_1 = A_1, B_2 = A_1 \cup A_2, \dots, B_n = \bigcup_{k=1}^n A_k$ forman una sucesión monótona creciente en el sentido de que $B_1 \subseteq B_2 \subseteq \dots \subseteq B_n \subseteq \dots$, y los sucesos complementarios $\bar{B}_1, \bar{B}_2, \dots, \bar{B}_n, \dots$ son tales que $\bar{B}_1 \supseteq \bar{B}_2 \supseteq \dots \supseteq \bar{B}_n \supseteq \dots$. Con ello, el suceso $A = \bigcup_k A_k$ coincide con el suceso $\bigcup_k B_k$, y el suceso complementario \bar{A} coincide con el suceso $\bigcap_k \bar{B}_k$. Según la propiedad de continuidad (1.11) $\mathbf{P}(\bar{A}) = \lim_{n \rightarrow \infty} \mathbf{P}(\bar{B}_n)$. Ahora bien, $\mathbf{P}(\bar{A}) = 1 - \mathbf{P}(A)$, $\mathbf{P}(\bar{B}_n) = 1 - \mathbf{P}(B_n)$, y según la propiedad de la aditividad (1.10) $\mathbf{P}(B_n) = \sum_{k=1}^n \mathbf{P}(A_k)$, $n = 1, 2, \dots$. Por consiguiente,

$$\begin{aligned} \mathbf{P}(A) &= 1 - \mathbf{P}(\bar{A}) = \lim_{n \rightarrow \infty} [1 - \mathbf{P}(\bar{B}_n)] = \\ &= \lim_{n \rightarrow \infty} \mathbf{P}(B_n) = \lim_{n \rightarrow \infty} \sum_{k=1}^n \mathbf{P}(A_k) = \sum_{k=1}^{\infty} \mathbf{P}(A_k), \end{aligned}$$

es decir, tiene lugar la igualdad (1.12).

Por otro lado, si $A_1 \supseteq A_2 \supseteq \dots$ es cualquier sucesión de sucesos monótona decreciente (la aparición de A_n arrastra consigo la aparición de A_1, \dots, A_{n-1}), entonces los sucesos complementarios $\bar{A}_1, \bar{A}_2, \dots$ forman una sucesión monótona creciente $\bar{A}_1 \subseteq \bar{A}_2 \subseteq \dots$ y su unión $\bar{A} = \bigcup_k \bar{A}_k$ puede ser representada como la unión de los sucesos que no se cortan $B_1 = \bar{A}_1, B_2 = \bar{A}_2 \setminus \bar{A}_1, \dots, B_n = \bar{A}_n \setminus \bar{A}_{n-1} \dots : \bar{A} = \bigcup_k B_k$. De acuerdo con la fórmula (1.12)

$$\mathbf{P}(\bar{A}) = \sum_{k=1}^{\infty} \mathbf{P}(B_k) = \sum_{k=1}^{\infty} [\mathbf{P}(\bar{A}_k) - \mathbf{P}(\bar{A}_{k-1})] = \lim_{n \rightarrow \infty} \mathbf{P}(\bar{A}_n),$$

y por consiguiente, tiene lugar la relación (1.11):

$$\mathbf{P}(A) = \lim_{n \rightarrow \infty} \mathbf{P}(A_n).$$

Ejemplo (juego hasta la primera pérdida). Imaginémosnos que se juega a «cara» o «cruz», cuando a cada paso el jugador gana o pierde en dependencia de si acertó o no acertó el resultado del lanzamiento de la moneda simétrica (tenemos dos resultados equiprobables «cara» o «cruz»). Supongamos que el juego se prolonga solamente hasta la primera pérdida. Cada uno de los resultados durante lanzamiento repetido n veces se puede describir por la sucesión del tipo $CCX \dots \dots XCX$, donde C ó X significan la caída de «cara» o «cruz» durante

el paso correspondiente; consideramos que todos los resultados posibles son equiprobables. Su número total es $N = 2^n$, de modo que la probabilidad de cada uno de los resultados, durante el lanzamiento de la moneda repetido n veces, será igual a 2^{-n} .

Para cualquier estrategia del jugador (digamos, el siempre predice «cara») el juego hasta la primera pérdida (hasta que caiga por primera vez «cruz») tiene un número infinito de resultados: la pérdida puede ser ya en el primer paso, puede ocurrir en el segundo paso y así sucesivamente. Cada resultado elemental se determina por el número n de pasos hasta la primera pérdida: $n = 1, 2, \dots$ siendo evidentemente igual a 2^{-n} la probabilidad de que el número indicado sea precisamente n . De tal forma como modelo teórico-probabilístico del juego descrito se puede tomar el espacio de los sucesos elementales, que coincide con el conjunto de todos los números naturales $n = 1, 2, \dots$, adjudicándole al resultado elemental n la probabilidad 2^{-n} . Como en cualquier otra prueba, la probabilidad del suceso cierto (que une a todos los resultados elementales) deberá ser igual a 1, lo que obtenemos precisamente de acuerdo con la fórmula general (1.12):

$$\sum_{n=1}^{\infty} \frac{1}{2^n} = \frac{1}{2} \cdot \frac{1}{1 - \frac{1}{2}} = 1.$$

Para tener un cuadro más completo se podría admitir también el resultado $n = \infty$, que significa que en la sucesión infinita de lanzamientos no cae ni una vez «cruz»; en el cuadro de nuestro esquema teórico-probabilístico tal resultado tiene una probabilidad igual a cero.

Hallemos la probabilidad de que el número de lanzamientos hasta el primer acierto con «cruz» sea par. Evidentemente, de acuerdo con la fórmula (1.12) esta probabilidad es

$$\sum_{k=1}^{\infty} \frac{1}{2^{2k}} = \frac{1}{4} \cdot \frac{1}{1 - \frac{1}{4}} = \frac{1}{3},$$

(y no $1/2$ como se podría pensar equivocadamente, partiendo de la simetría de «cara» y «cruz»; a propósito, la probabilidad de que caiga «cruz» en el primer lanzamiento ya es igual a $1/2$, pero ella puede aparecer por primera vez también en el lanzamiento tercero, quinto, etc).

Así pues, el modelo teórico-probabilístico representa en sí el espacio Ω de los resultados elementales ω descritos de una u otra forma, en el cual están determinadas las probabilidades $\mathbf{P}(A)$ de los sucesos examinados $A \subseteq \Omega$, que satisfacen la condición de la aditividad numerable (1.12).

El problema fundamental, que de una u otra forma surge en la mayoría de las tareas teórico-probabilísticas, consiste en lo siguiente:

se dan las probabilidades $\mathbf{P}(A)$, hablando convencionalmente, de algunos sucesos simples A , cuyo conjunto designamos por γ ; se exige hallar las probabilidades $\mathbf{P}(B)$ de otros sucesos B , ligados de algún modo con los sucesos $A \in \gamma$ (más exactamente, obtenidos de los sucesos iniciales $A \in \gamma$ por medio de operaciones sucesivas de unión, intersección, paso a los sucesos complementarios, etc.).

Naturalmente, en cada caso concreto esta cuestión se resuelve distintamente, pero la pregunta general es: ¿cuándo pueden ser determinadas, en principio, las probabilidades $\mathbf{P}(B)$ a partir de las probabilidades dadas $\mathbf{P}(A)$, $A \in \gamma$. Esta cuestión importante para fundamentar la teoría de las probabilidades se resuelve positivamente, si el sistema inicial de sucesos γ posee la propiedad de que junto con los sucesos A_1, A_2 contiene también su intersección $A_1 \cap A_2$, y para cualesquiera sucesos $A, A_1 \in \gamma$ tales que A_1 pertenezca a A , la diferencia $A \setminus A_1$ se puede «desarrollar», en una suma finita de los sucesos, que no se cortan A_2, \dots, A_n de γ (tal sistema γ algunas veces se llama *desarrollable*). En el caso del sistema desarrollable γ tiene lugar la fórmula siguiente¹⁾:

$$\mathbf{P}(B) = \inf_{A \subset B} \sum_h \mathbf{P}(A_h), \quad (1.13)$$

donde el límite inferior se toma para todos los sucesos A_h de γ , $k = 1, 2, \dots$, que dan en suma el suceso $A = \bigcup_h A_h$ que contiene a B .

De ejemplo del sistema desarrollable de sucesos, en la prueba sobre la observación de una determinada magnitud aleatoria ξ , puede servir el conjunto de sucesos del tipo $A = \{x' < \xi \leq x''\}$, o sea, «el punto ξ cae en el segmento abierto por la izquierda $(x', x'']$ »; en el ensayo sobre la observación de un par de magnitudes aleatorias ξ_1, ξ_2 , se considera sistema desarrollable el conjunto de todos los sucesos del tipo $A = \{x'_1 < \xi_1 \leq x''_1, x'_2 < \xi_2 \leq x''_2\}$, o sea, «el punto con coordenadas (ξ_1, ξ_2) cae en el rectángulo $(x'_1, x''_1) \times (x'_2, x''_2)$ con los lados que forman el ángulo inferior izquierdo excluidos».

En todo lo sucesivo, al examinar tal o cual modelo teórico probabilístico concreto supondremos que las probabilidades de los sucesos que nos interesan, en principio, están determinadas y la tarea consiste en hallarlas en una forma explícita.

4. Concordancia entre el modelo y la prueba. En las aplicaciones, al utilizar tal o cual modelo teórico probabilístico surge la pregunta, de cómo corresponde este modelo a la situación real de las cosas, a la prueba real.

Ejemplo (probabilidad del nacimiento de un niño). Se puede pensar que el nacimiento de un niño o una niña en cada uno de los casos es

¹⁾ Véase, por ejemplo, el libro de I. I. Guijman, A. V. Skorjod «Introducción a la teoría de los procesos aleatorios» Moscú «Nauka» 1965 (pág. 121 y en adelante).

un suceso equiprobable (es decir, el nacimiento de un niño, desde el punto de vista de la probabilidad, es análogo a la prueba del lanzamiento de la moneda: con la probabilidad $1/2$ nace un niño y con la misma probabilidad nace una niña).

Examinaremos cada nacimiento como una prueba, durante la cual nace un niño con la probabilidad p y nace una niña con la probabilidad $1 - p$, con ello consideraremos que el resultado de una prueba de ningún modo influye en el resultado de la otra (en otras palabras, examinaremos los nacimientos como pruebas independientes).

Durante estas n «pruebas» el número m de niños nacidos es aleatorio. La probabilidad de que la frecuencia m/n del nacimiento de un niño se desvíe de $p = 1/2$ en un valor $\delta > 0$, es igual a (véase más adelante el § 3, cap. II)

$$P \left\{ \frac{m}{n} - p > \delta \right\} \approx 1 - \Phi(2\delta\sqrt{n}),$$

donde $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-u^2/2} du$, además, en la igualdad aproximada

indicada el error no supera a $1/\sqrt{n}$. Elegimos δ de tal modo, que $1 - \Phi(2\delta\sqrt{n}) \leq \varepsilon$, donde ε es un número tan pequeño que se debe despreciar la posibilidad de aparición del suceso que tenga una probabilidad no mayor que ε . (Por ejemplo, para $\varepsilon = 0,002$, se puede tomar $x = 2\delta\sqrt{n} \geq 3$ de las tablas de la función $\Phi(x)$; véase la pág. 101.)

En casi todos los países ya hace mucho tiempo que se registra el nacimiento de cada niño de modo que ya se tiene una gran cantidad de datos estadísticos. Hablando de Suiza, desde el año 1871 hasta el 1900 nacieron 1 359 671 niños y 1 285 086 niñas¹⁾. De acuerdo con estos datos, para $n = 2\,644\,757$ se tuvo $m = 1\,359\,671$ y $m/n = 0,5141$. ¿Está esto de acuerdo con el modelo elegido, en el cual $p = 1/2$, o en realidad la probabilidad del nacimiento de un niño es mayor que $1/2$?

Para $p = 1/2$ tenemos

$$\frac{m}{n} - \frac{1}{2} = 0,0141 > \delta = \frac{3}{2\sqrt{2\,644\,757}},$$

y de este modo, dentro del marco del modelo aceptado por nosotros (con probabilidad del nacimiento de un niño $p = 1/2$) «observamos» el suceso prácticamente imposible $\left\{ \frac{m}{n} - \frac{1}{2} > \delta \right\}$, más exactamente, la probabilidad de este proceso no supera a 0,002. Al mismo tiempo,

¹⁾ Véase el libro de B. L. Van der Warden «Matematicheskaya Statistika» IL-Moscú 1960 (pág. 43).

si la probabilidad del nacimiento de un niño es en realidad mayor que $1/2$ (digamos, $p = 0,51$), entonces en este suceso $\left\{ \frac{m}{n} - \frac{1}{2} > \delta \right\}$ no hay nada extraño, como podemos calcular, él tiene una probabilidad próxima a 1. Basándose en estos datos, se debe **d e s e c h a r** la hipótesis de que $p = 1/2$ y considerar $p > \frac{1}{2}$ (en cierto sentido la mejor valuación de la probabilidad del nacimiento de un niño, según los datos estadísticos expuestos, es la **f r e c u e n c i a** $\frac{m}{n} = 0,5141$).

En nuestro curso estudiaremos los distintos modelos teórico-probabilísticos, sin examinar la cuestión en qué grado corresponden a tal o cual fenómeno real (la formulación matemática exacta y la solución de semejantes cuestiones son objeto de la ciencia especial llamada estadística matemática).

§ 2. INDEPENDENCIA
Y PROBABILIDAD
CONDICIONAL

1. Concepto de independencia. Examinemos dos pruebas independientes Ω_1 y Ω_2 cada una de las cuales tiene un número finito de resultados equiprobables. La independencia aquí se comprende en el sentido de que ninguno de los resultados elementales ω_1 de la primera prueba influye de ningún modo en el resultado ω_2 de la segunda prueba (y viceversa). De ejemplo puede servir (véase el comienzo del § 1) los dos lanzamientos de la moneda (ω_1 es el primer resultado, ω_2 es el segundo resultado de los lanzamientos); el lanzamiento Ω_1 de la moneda y el lanzamiento Ω_2 del dado de juego, etc.

Examinemos el suceso A_1 ligado con la primera prueba Ω_1 y el suceso A_2 ligado con la segunda prueba Ω_2 . ¿Cuál será la probabilidad del suceso $A = A_1 \cdot A_2$ ligado con la prueba, que es un conjunto de pruebas independientes Ω_1, Ω_2 ?

El resultado elemental de las pruebas examinadas se describe por el par (ω_1, ω_2) , donde ω_1 es el resultado elemental en la primera prueba, ω_2 es el resultado en la segunda prueba. La intuición indica que para los resultados independientes equiprobables $\omega_1 \in \Omega_1$ y $\omega_2 \in \Omega_2$, los resultados elementales $\omega = (\omega_1, \omega_2)$ también deben ser equiprobables. Si el número total de resultados en la primera prueba es N_1 y en la segunda prueba N_2 , entonces el número total de pares posibles $\omega = (\omega_1, \omega_2)$ será evidentemente igual al producto $N_1 \cdot N_2$. De tal modo, el número de resultados elementales $\omega = (\omega_1, \omega_2)$ es $N = N_1 \cdot N_2$. El suceso $A = A_1 \cdot A_2$ significa la realización de ambos sucesos A_1 y A_2 que será, cuando, y sólo cuando tengan lugar los

resultados elementales $\omega_1 \in A_1$ y $\omega_2 \in A_2$. Evidentemente, en total se tienen $N(A) = N_1(A_1) N_2(A_2)$ resultados elementales $\omega = (\omega_1, \omega_2)$ que conducen a la aparición del suceso $A = A_1 \cdot A_2$ (aquí $N_1(A_1)$ es el número de sucesos elementales ω_1 que conducen al suceso A_1 en la primera prueba, $N_2(A_2)$ es el número de resultados elementales ω_2 que conducen al suceso A_2 en la segunda prueba. Según la fórmula general (1.1), la probabilidad del suceso $A = A_1 \cdot A_2$ se debe determinar como

$$P(A) = \frac{N(A)}{N} = \frac{N_1(A_1) \cdot N_2(A_2)}{N_1 \cdot N_2},$$

que se expresa directamente a través de las probabilidades $P(A_1)$, $P(A_2)$ de los sucesos separados A_1, A_2 como

$$P(A) = P(A_1) P(A_2). \quad (2.1)$$

Así pues, al examinar los sucesos del tipo $A = A_1 \cdot A_2$, donde los sucesos separados A_1 y A_2 están ligados con las pruebas independientes, para la probabilidad $P(A)$ tendremos la fórmula (2.1).

Cierto es, que hasta ahora, sólo hemos examinado las pruebas independientes con un número finito de resultados equiprobables.

Examinemos ahora las pruebas independientes Ω_1 y Ω_2 en las que se observa la magnitud aleatoria ξ_1 en Ω_1 y la magnitud aleatoria ξ_2 en Ω_2 (las magnitudes ξ_1 y ξ_2 no están ligadas entre sí de ningún modo).

Sea $p_1(x)$ la densidad de la distribución de probabilidades de la magnitud aleatoria ξ_1 de modo que la probabilidad de cualquier suceso del tipo $A_1 = \{x'_1 \leq \xi_1 \leq x''_1\}$, o sea, «el punto aleatorio ξ_1 cae en el intervalo $[x'_1, x''_1]$ », se da por la fórmula (1.2):

$$P(A_1) = \int_{A_1} p_1(x) dx.$$

Representémonos mentalmente, que el segmento $[a, b]$ de los valores posibles de la magnitud aleatoria ξ_1 (este segmento puede ser también infinito) está dividido en N_1 partes de tal forma que los aciertos en cualquiera de los intervalos de división $\Delta_k = (x_k, x_{k+1})$, $k = 1, 2, \dots, N_1$ son sucesos equiprobables:

$$P\{x_k < \xi \leq x_{k+1}\} = \int_{\Delta_k} p_1(x) dx = \frac{1}{N_1}.$$

Eligiendo a N_1 lo suficientemente grande, se puede determinar con cualquier exactitud la probabilidad del suceso A_1 (probabilidad de caída del punto aleatorio ξ_1 en el intervalo $[x'_1, x''_1]$), por la fórmula siguiente:

$$P(A_1) \approx \sum_{\Delta_k} \int p_1(x) dx \frac{N_1(A_1)}{N_1},$$

donde la suma se realiza en aquellos k , para los cuales $\Delta_k \subseteq [x'_1, x''_1]$ y $N_1(A_1)$ es el número de intervalos de la división que caen en el segmento considerado $[x'_1, x''_1]$. Esto significa de hecho, que podemos determinar con cualquier grado de exactitud, las probabilidades de diferentes sucesos ligados con la prueba Ω_1 , examinándola convencionalmente como prueba con un número finito de resultados equiprobables, cada uno de los cuales representa en sí el acierto del punto aleatorio ξ_1 en el intervalo correspondiente de la división Δ_k , $k = 1, \dots, N_1$.

Lo mismo se puede decir sobre la prueba Ω_2 donde se observa la magnitud aleatoria ξ_2 con la densidad de distribución de probabilidades $p_2(x)$, y se examina el suceso $A_2 = \{x'_2 \leq \xi_2 \leq x''_2\}$. Utilizando la fórmula (2.1) para los sucesos con un número finito de resultados equiprobables, para valores suficientemente grandes de N_1 y N_2 (que significan el número de intervalos de la división en cada prueba por separado), tendremos con cualquier grado de exactitud: para el suceso $A = A_1 \cdot A_2$

$$\mathbf{P}(A) \approx \frac{N_1(A_1)}{N_1} \frac{N_2(A_2)}{N_2} \approx \mathbf{P}(A_1) \cdot \mathbf{P}(A_2).$$

Esto indica que en realidad debe tener lugar la igualdad exacta

$$\mathbf{P}(A) = \mathbf{P}(A_1) \cdot \mathbf{P}(A_2),$$

es decir, de nuevo llegamos a la fórmula (2.1), obtenida anteriormente, sólo para las pruebas independientes con un número finito de resultados equiprobables.

Examinemos, por último, los sucesos arbitrarios A_1 y A_2 ligados con algunas pruebas independientes Ω_1 y Ω_2 . Igual que anteriormente, al hablar sobre la independencia, aquí tenemos en cuenta las condiciones físicas concretas, para las cuales ninguno de los resultados de la primera prueba influye de ningún modo sobre el resultado de la segunda (y viceversa).

Está claro que la probabilidad del suceso $A = A_1 \cdot A_2$ en las pruebas independientes dadas Ω_1, Ω_2 deberá ser la misma que en las pruebas independientes cualesquiera Ω'_1, Ω'_2 , cuando nosotros examinamos cualquier suceso $A' = A'_1 \cdot A'_2$, donde los sucesos $A'_1 \subseteq \Omega'_1$ y $A'_2 \subseteq \Omega'_2$ tienen la misma probabilidad que los sucesos tomados al principio $A_1 \subseteq \Omega_1$ y $A_2 \subseteq \Omega_2$. Ahora bien, para cualesquiera probabilidades $p_1 = \mathbf{P}(A_1)$ y $p_2 = \mathbf{P}(A_2)$ se pueden realizar las pruebas independientes Ω'_1 y Ω'_2 del tipo descrito anteriormente (en los que se observan determinadas magnitudes aleatorias ξ_1 y ξ_2 con densidad de distribución $p_1(x)$ y $p_2(x)$), eligiendo los sucesos correspondientes A'_1 y A'_2 con las mismas probabilidades $p_1 = \mathbf{P}(A'_1)$ y $p_2 = \mathbf{P}(A'_2)$. Anteriormente ya se mostró que debe cumplirse la relación $\mathbf{P}(A'_1 \cdot A'_2) = p_1 \cdot p_2$ y por consiguiente $\mathbf{P}(A_1 \cdot A_2) = p_1 \cdot p_2$.

De tal modo, la fórmula (2.1) debe tener lugar para los sucesos arbitrarios A_1, A_2 , ligados con las pruebas independientes Ω_1, Ω_2 .

En la teoría de probabilidades, como en cualquier otra disciplina matemática, se opera con conceptos que se formulan dentro del cuadro del modelo teórico correspondiente. El modelo teórico probabilístico general se describe como un determinado espacio Ω de resultados elementales ω , donde para determinar la clase de conjuntos $A \subseteq \Omega$, llamados sucesos, las probabilidades indicadas $\mathbf{P}(A)$ se someten a la regla de adición (1.12). En tal descripción no se dice una palabra sobre las determinadas «condiciones físicas concretas» cuyo análisis permitiese juzgar sobre la independencia de tales o cuales sucesos $A_1, A_2 \subseteq \Omega$.

Dentro del cuadro del modelo teórico-probabilístico general se considera señal de independencia de los sucesos A_1, A_2 al cumplimiento de la propia relación (2.1). Se llaman independientes precisamente, los sucesos A_1 y A_2 si

$$\mathbf{P}(A_1 \cdot A_2) = \mathbf{P}(A_1) \cdot \mathbf{P}(A_2). \quad (2.2)$$

Ejemplo. Supongamos que de la baraja de naipes se saca al azar una carta. Tenemos una prueba concreta, que cabe en el esquema de un número N finito de resultados equiprobables (N es el número de naipes en la baraja, digamos $N = 52$). Examinemos los sucesos A_1 , o sea, «se saca un naipe del palo de bastos» y A_2 «se saca un caballo». A la primera prueba le favorecen $N(A_1) = 13$ resultados elementales, ya que en el montón se tienen 13 naipes del palo de bastos, y por eso $\mathbf{P}(A_1) = 13/52 = 1/4$. Al segundo suceso le favorecen $N(A_2) = 4$ resultados elementales, ya que se tienen 4 caballos y por eso $\mathbf{P}(A_2) = 4/52 = 1/13$. Por último, al suceso $A_1 \cdot A_2$, o sea, «se saca el caballo de bastos», le favorecen, exactamente, un resultado elemental, y $\mathbf{P}(A_1 \cdot A_2) = 1/52$. Se ve que tiene lugar la relación (2.2) y, por consiguiente, los sucesos A_1 y A_2 son independientes.

Este ejemplo sencillo ya demuestra que el concepto de independencia, dado anteriormente, permite con exactitud matemática utilizarlo allí, donde, a pesar de toda la sencillez «de las condiciones físicas concretas» de la prueba examinada, es difícil abordar de otra forma la cuestión sobre la independencia de tales o cuales sucesos A_1 y A_2 .

A propósito, si en el ejemplo examinado anteriormente cambiamos las condiciones de la prueba, añadiendo a los 52 naipes de la baraja corriente M naipes en blanco («carte blanche») entonces los sucesos A_1 , «que sean bastos» y A_2 , «que sea caballo» serán de forma evidente dependientes; para un gran M (digamos, $M = 1000$) casi con seguridad el naipe sacado será «carte blanche», y la probabilidad de sacar los bastos es muy pequeña; para la condición de que el naipe sacado sea el caballo (suceso A_2), el suceso A_1 (que el naipe sacado sean bastos)

se hace completamente real, se debe considerar que la probabilidad del suceso A_1 en las condiciones A_2 será igual a $1/4$.

En general, los sucesos A_1, A_2, \dots se llaman *independientes entre sí*, (abreviadamente: *independientes*) si la probabilidad de la intersección $A_{i_1} \dots A_{i_n}$ para cualesquiera valores distintos i_1, \dots, \dots, i_n es

$$P(A_{i_1} \dots A_{i_n}) = P(A_{i_1}) \dots P(A_{i_n}). \quad (2.3)$$

Supongamos que se examinan determinadas pruebas Ω_1, Ω_2 , con los resultados posibles $\omega_1, \omega_2, \dots$. Se puede determinar la «prueba compleja» Ω como el espacio de los resultados elementales $\omega = (\omega_1, \omega_2, \dots)$, donde ω_1 es el resultado elemental en Ω_1 , ω_2 es el resultado elemental en Ω_2 , etc. El mecanismo de aleatoriedad de tal prueba se describe corrientemente con las probabilidades $P(A_{i_1} \dots A_{i_n})$ de todas las intersecciones posibles de los sucesos $A_{i_1} \subseteq \Omega_{i_1}, A_{i_2} \subseteq \Omega_{i_2}, \dots$. Si para cualesquiera sucesos A_{i_1}, \dots, A_{i_n} (ligados con diferentes pruebas $\Omega_{i_1}, \dots, \Omega_{i_n}$), tiene lugar la relación (2.3), entonces los sucesos $\Omega_1, \Omega_2, \dots$ se llaman *independientes*.

2. Probabilidad condicional. Junto al concepto de independencia, uno de los más importantes conceptos en la teoría de probabilidades es la llamada *probabilidad condicional*, que da la característica general del enlace de distintos sucesos.

Sean los sucesos A_1, A_2 ligados con la prueba Ω , que tienen un número finito N de resultados equiprobables ω . Supongamos que nos es conocida la aparición del suceso A_2 (pero no sabemos el resultado elemental ω de la prueba dada). ¿Cuál es la probabilidad del suceso A_1 en estas nuevas condiciones, cuando es conocida la aparición del suceso A_2 ?

Está claro, que en las nuevas condiciones pueden haber, en efecto, sólo $N(A_2)$ resultados elementales distintos (cada uno de los cuales puede conducir a la aparición del suceso A_2). Si designamos por $N(A_1 \cdot A_2)$ el número de aquellos resultados elementales del número de los resultados $\omega \in A_2$ que conducen al suceso A_1 , entonces sería natural determinar la probabilidad del suceso A_1 , (en condiciones de aparición del suceso A_2) como la relación $\frac{N(A_1 \cdot A_2)}{N(A_2)}$ (compárese con la fórmula general (1.1)). Pero observemos que el número $N(A_1 \cdot A_2)$ coincide con el número de todos los resultados elementales $\omega \in \Omega$ que conducen a la aparición de ambos sucesos A_1, A_2 (es decir, al suceso $A_1 \cap A_2$). Tomando en consideración, que

$$P(A_1 \cdot A_2) = \frac{N(A_1 \cdot A_2)}{N}, \quad P(A_2) = \frac{N(A_2)}{N},$$

podemos expresar la probabilidad del suceso A_1 , determinada anteriormente, en condiciones de la aparición de A_2 (designada en adelante

por $\mathbf{P}(A_1 | A_2)$ por la fórmula

$$\mathbf{P}(A_1 | A_2) = \frac{\mathbf{P}(A_1 \cdot A_2)}{\mathbf{P}(A_2)}. \quad (2.4)$$

Esta fórmula también tiene sentido en el caso general (para $\mathbf{P}(A_2) > 0$); ella determina la probabilidad condicional $\mathbf{P}(A_1 | A_2)$ del suceso A_1 en condiciones de la aparición del suceso A_2 .

En vista de la importancia de este concepto lo aclaramos, aún más, en un modelo. Imaginémonos que la prueba examinada consiste en la observación de un punto aleatorio ξ en un determinado espacio fásico X , actuando precisamente, el «mecanismo de aleatoriedad» del siguiente modo: la probabilidad de caída del punto aleatorio en la zona $A \subseteq X$ (suceso A) es proporcional al volumen de esta zona, que se determina como la integral por A de una función determinada positiva integrable $p(x)$:

$$\mathbf{P}(A) = \frac{1}{V} \int_A p(x) dx,$$

donde V es el volumen de todo el espacio fásico:

$$V = \int_X p(x) dx.$$

Supongamos que al observador le es conocido, que el punto aleatorio ξ cae en una determinada zona A_2 (suceso A_2). ¿Cuál será en estas nuevas condiciones la aleatoriedad de caída en una u otra zona A_1 ?

Está claro, que bajo la condición de caída en la zona A_2 , el punto ξ podrá caer en la zona A_1 , sólo en el caso, cuando se tiene una intersección no vacía $A_1 \cap A_2$, siendo la caída en A_1 en las nuevas condiciones, equivalente a la caída en la intersección indicada $A_1 \cap A_2$, que se realiza con la probabilidad proporcional al volumen correspondiente

$\int_{A_1 \cap A_2} p(x) dx$. Por consiguiente, en condiciones de la caída en la zona A_2 , el punto aleatorio ξ cae en la zona A_1 con una probabilidad, igual a $\frac{1}{V(A_2)} \int_{A_1 \cap A_2} p(x) dx$, donde $V(A_2)$ es el volumen de la zona A_2 , es decir $V(A_2) = \int_{A_2} p(x) dx$. Vemos que esta probabilidad $\mathbf{P}(A_1 | A_2)$ coincide con la relación $\frac{\mathbf{P}(A_1 \cdot A_2)}{\mathbf{P}(A_2)}$.

Partiendo de la probabilidad condicional $\mathbf{P}(A_1 | A_2)$, se puede introducir el siguiente concepto de independenciam: el suceso A_1 no depende del suceso A_2 , [si la probabilidad de aparición de A_1 no se cambia al aparecer el suceso A_2 , más exactamente, si la probabilidad

condicional $\mathbf{P}(A_1 | A_2)$ coincide con la probabilidad inicial $\mathbf{P}(A_1)$ del suceso examinado A_1 , es decir:

$$\mathbf{P}(A_1 | A_2) = \mathbf{P}(A_1).$$

Pero como vemos, en la fórmula (2.4) esta igualdad es equivalente a la igualdad (2.2) y de nuevo llegamos al concepto anterior de independencia de los sucesos A_1 y A_2 .

En las aplicaciones, cuando se trata de hallar la probabilidad $\mathbf{P}(A)$ de tal o cual suceso A , ligado con una determinada prueba compleja Ω , es cómodo suponer condicionalmente, la realización, de uno u otro modo, del suceso elegido B , si la introducción de esta condición simplifica la prueba y permite determinar la probabilidad condicional $\mathbf{P}(A | B)$.

Supongamos que tenemos, precisamente, un tal sistema llamado *sistema completo* de sucesos B_1, B_2, \dots de modo que como resultado de la prueba examinada ocurre, u n o y s ó l o u n o, de los sucesos B_1, B_2, \dots (en otras palabras, estos sucesos no se cortan y su unión es el suceso cierto). Después de determinar las probabilidades condicionales $\mathbf{P}(A | B_k)$ del suceso A que nos interesa, en diferentes condiciones $B_k, k = 1, 2, \dots$, se puede calcular la probabilidad $\mathbf{P}(A)$ del suceso A por la siguiente *fórmula de la probabilidad total*:

$$\mathbf{P}(A) = \sum_k \mathbf{P}(A | B_k) \mathbf{P}(B_k). \quad (2.5)$$

Esta fórmula se puede deducir fácilmente, representando el suceso A como la unión de los sucesos que se cortan $A_k = A \cdot B_k, k = 1, 2, \dots$:

$$1: A \cdot \bigcup_k B_k = \bigcup_k (A \cdot B_k),$$

y utilizando la regla de adición de probabilidades (véase (1.12)), según la cual

$$\mathbf{P}(A) = \sum_k \mathbf{P}(A \cdot B_k).$$

Expresando los sumandos $\mathbf{P}(A \cdot B_k)$ a través de la probabilidad $\mathbf{P}(A \cdot B_k) = \mathbf{P}(A | B_k) \cdot \mathbf{P}(B_k)$, obtenemos la igualdad (2.5)

Para ilustrar las posibilidades que abre la utilización de la fórmula de la probabilidad total, detengámonos aunque sea en un ejemplo.

Ejemplo (tarea sobre el arruinamiento del jugador). Examinemos el juego llamado de «cara» o «cruz» cuando el jugador elige la «cara» o la «cruz». Si cae la cara de la moneda designada por el jugador, entonces gana recibiendo, digamos, 1 rublo; en caso contrario pierde lo mismo. Supongamos que el capital inicial del jugador consta de x rublos y el jugador se plantea el objetivo de aumentarlo hasta una determinada suma de $a > x$ rublos. El juego se termina cuando el jugador o bien reúne la suma a determinada anteriormente, o bien

se arruina, perdiendo todo el capital que tenía. ¿Cuál será la probabilidad de que, al fin o al cabo, el jugador se arruine, ya que no reunió la suma a de rublos deseada?

Es claro, que esta probabilidad depende del capital inicial x y de la suma final a . Designando por $p(x)$ la probabilidad de que, teniendo el jugador x rublos, a pesar de todo se arruina. Entonces la probabilidad de arruinamiento en condiciones de ganancia en el primer paso en nuestras designaciones será $p(x+1)$, ya que después de la ganancia el capital se hace igual a $x+1$. Análogamente, la probabilidad del arruinamiento en condiciones de pérdida en el primer paso es igual a $p(x-1)$, ya que después de la pérdida el capital del jugador quedará igual a $x-1$. Designemos por B_1 el suceso, consistente en que el jugador ganó en el primer paso, por B_2 el suceso consistente en que perdió y sea A el suceso que significa el arruinamiento del jugador. Las probabilidades condicionales del arruinamiento se expresan en las designaciones admitidas por nosotros así:

$$\mathbf{P}(A | B_1) = p(x+1), \quad \mathbf{P}(A | B_2) = p(x-1).$$

Los sucesos B_1 y B_2 forman un sistema completo, ya que en el primer paso el jugador o bien gana o pierde, siendo $\mathbf{P}(B_1) = \mathbf{P}(B_2) = 1/2$. La fórmula de la probabilidad total da la siguiente relación de las probabilidades buscadas $p(x)$:

$$p(x) = \frac{1}{2} [p(x+1) + p(x-1)]$$

para todos los $x = 1, \dots, a-1$ (evidentemente se debe establecer que $p(0) = 1$ y $p(a) = 0$).

La solución de la ecuación

$$f(x+1) = 2f(x) - f(x-1), \quad x = 1, 2, \dots,$$

con relación a la función $f(x)$, la cual para $1 \leq x \leq a-1$ satisface a la probabilidad $p(x)$, se determina sucesivamente a través de $y_0 = f(0)$, $y_1 = f(1)$ y para todos los $x = 2, 3, \dots$, y, por consiguiente, puede existir sólo una solución con las condiciones iniciales dadas $y_0 = f(0)$, $y_1 = f(1)$. Como se puede comprobar fácilmente, esta solución tiene la forma

$$f(x) = y_0 - (y_1 - y_0)x,$$

y para $y_0 = 1$, $y_1 = p(1)$ obtenemos

$$p(x) = 1 + (1 - p(1))x,$$

de donde, teniendo en cuenta que $p(a) = 1 - (1 - p(1))a = 0$, hallamos $p(1) = 1 - \frac{1}{a}$ y finalmente obtenemos

$$p(x) = 1 - \frac{x}{a}, \quad x = 0, 1, \dots, a.$$

§ 3. MAGNITUDES ALEATORIAS
Y DISTRIBUCION
DE LAS PROBABILIDADES.
INDEPENDENCIA

1. Distribuciones discreta y continua. En el cuadro de tal o cual modelo teórico-probabilístico, que se describe con el espacio correspondiente de sucesos elementales Ω y con unas probabilidades determinadas $\mathbf{P}(A)$ de los sucesos $A \subseteq \Omega$, cuando se examina tal o cual magnitud aleatoria ξ , suponemos que su dependencia del suceso (más exactamente, del resultado elemental $\omega \in \Omega$) es tal, que están determinadas las probabilidades de todos los sucesos posibles del tipo $\{x' \leq \xi \leq x''\}$. Aquí tenemos presente la magnitud aleatoria real, o sea la función real $\xi = \xi(\omega)$ del resultado elemental $\omega \in \Omega$.

Inmediatamente se debe decir, que en las tareas teórico-probabilísticas, como regla, la dependencia explícita $\xi = \xi(\omega)$ de $\omega \in \Omega$ no juega un papel importante. En efecto, la característica importante de dependencia entre la magnitud ξ que nos interesa y el caso dado, nos la dan las probabilidades

$$\mathbf{P} \{x' \leq \xi \leq x''\} \quad (3.1)$$

de todos los sucesos posibles $A = \{x' \leq \xi \leq x''\}$; en su conjunto, ellos muestran cómo está distribuida la probabilidad de caída del punto aleatorio ξ en tal o cual intervalo $[x', x'']$, en otras palabras, se da la *distribución de probabilidades de la magnitud aleatoria* ξ .

Al examinar una magnitud aleatoria dada ξ , se puede tomar la recta real E^1 , en calidad de espacio de los sucesos elementales, fijando cada suceso elemental $\{\xi = x\}$ por el punto correspondiente $x \in E^1$, y determinando formalmente la dependencia entre ξ y el suceso elemental $x \in E^1$ por la fórmula $\xi(x) \equiv x$; el «mecanismo de aleatoriedad» de tal prueba se describe por la distribución de las probabilidades (3.1).

Señalemos una vez más, que la dependencia explícita del suceso elemental (digamos, del tipo $\xi(x) \equiv x$) no juega un papel importante desde el punto de vista del comportamiento probable de la magnitud aleatoria ξ ; por ejemplo, a las mismas leyes de probabilidad se someten la magnitud aleatoria $\xi(x) \equiv x$, $-\infty < x < \infty$, con una densidad de distribución uniforme ($p_{\xi}(x) = \frac{1}{L}$ para $0 \leq x \leq L$, $p_{\xi}(x) = 0$ para $x < 0$, $x > L$), y el punto aleatorio sobre la circunferencia de longitud L , que determina la posición de la bolita en el juego a la ruleta (véase el § 1).

Supongamos que se tiene un número finito o numerable de valores de x , cada uno de los cuales la magnitud aleatoria ξ puede tomarlo con la correspondiente probabilidad

$$P_{\xi}(x) = \mathbf{P} \{\xi = x\} \quad \left(\sum_x P_{\xi}(x) = 1 \right),$$

es decir, la magnitud aleatoria ξ toma uno de los valores indicados x con la probabilidad 1 (de hecho estos valores de x son los únicos posibles para ξ). Tal magnitud aleatoria se llama *discreta*; también se llama discreta su distribución de probabilidades.

Para cualesquiera x', x'' ($x' \leq x''$)

$$\mathbf{P}\{x' \leq \xi \leq x''\} = \sum_{x'}^{x''} P_{\xi}(x), \quad (3.2)$$

donde la suma se extiende a los límites señalados, al número finito o numerable de valores posibles de x para los cuales $P_{\xi}(x) > 0$.

La distribución de probabilidades de la magnitud aleatoria ξ se llama *continua*, si se tiene la densidad de probabilidad $p_{\xi}(x)$; es decir para cualesquiera x', x'' ($x' \leq x''$) (véase (1.2))

$$\mathbf{P}\{x' \leq \xi \leq x''\} = \int_{x'}^{x''} p_{\xi}(x) dx \quad (3.3)$$

(la densidad de la probabilidad $p_{\xi}(x)$, o de otro modo: *la densidad de distribución de probabilidades* es tal función no negativa e integrable

para la cual $\int_{-\infty}^{+\infty} p_{\xi}(x) dx = 1$).

La magnitud ξ , distribuida continuamente, toma cada uno de los valores fijados de x , sólo con una probabilidad igual a cero:

$$\mathbf{P}\{\xi = x\} = \lim_{\substack{x' \rightarrow x-0 \\ x'' \rightarrow x+0}} \int_{x'}^{x''} p_{\xi}(x) dx = 0,$$

y para cualquier punto de continuidad de la función $p_{\xi}(x)$ tenemos:

$$\mathbf{P}\{x' \leq \xi \leq x''\} \sim p_{\xi}(x) \cdot \Delta x$$

para $\Delta x = x'' - x' \rightarrow 0$, $x' \leq x \leq x''$.

Naturalmente, la magnitud aleatoria ξ puede no relacionarse ni al tipo discreto, ni al continuo (véase el ejemplo de la pág. 42). Se puede dar la distribución general de las probabilidades de la magnitud aleatoria ξ con ayuda de la llamada *función de distribución*

$$F_{\xi}(x) = \mathbf{P}\{\xi \leq x\}, \quad -\infty < x < \infty \quad (3.4)$$

y, precisamente, para cualesquiera x', x'' ($x' \leq x''$)

$$\mathbf{P}\{x' \leq \xi \leq x''\} = F_{\xi}(x'') - F_{\xi}(x' - 0),$$

donde $F_{\xi}(x - 0)$ significa el límite $\lim_{h \rightarrow 0} F_{\xi}(x - h)$ para $h > 0$.

La función $F_{\xi}(x)$ que satisface a la igualdad (3.4) es no negativa, monótona creciente (no decreciente) y continua por la derecha

$(F_{\xi}(x) = F_{\xi}(x + 0))$, siendo

$$\lim_{x \rightarrow -\infty} F_{\xi}(x) = 0, \quad \lim_{x \rightarrow \infty} F_{\xi}(x) = 1.$$

Las propiedades enumeradas se pueden deducir, fácilmente de la aditividad y continuidad de las probabilidades. Precisamente, el suceso $\{\xi \leq x\}$ es la intersección de los sucesos monótonos decrecientes $\{\xi \leq x_n\}$, donde x_n , $n = 1, 2, \dots$, es cualquier sucesión monótona decreciente, que converge en x , y según la propiedad de continuidad de la probabilidad $\lim_{n \rightarrow \infty} P\{\xi \leq x_n\} = P\{\xi \leq x\}$; según esta misma propiedad, para los sucesos $\{\xi \leq x_n\}$, $x_n \rightarrow -\infty$, que dan en su intersección el suceso imposible (vacío) $\lim_{n \rightarrow \infty} P\{\xi \leq x_n\} = 0$; para los sucesos monótonos crecientes $\{\xi \leq x_n\}$, $x_n \rightarrow +\infty$, que dan en su suma el suceso cierto $\{\xi < +\infty\}$, $\lim_{n \rightarrow \infty} P\{\xi \leq x_n\} = 1$. Luego, el suceso $\{\xi \leq x''\}$ es la unión de los sucesos que no se cortan $\{\xi \leq x\}$ y $\{x < \xi \leq x''\}$, de modo que

$$P\{x < \xi \leq x''\} = P\{\xi \leq x''\} - P\{\xi \leq x\}$$

y

$$P\{x' \leq \xi \leq x''\} = \lim_{x \rightarrow x'-0} P\{x < \xi \leq x''\} = F_{\xi}(x'') - F_{\xi}(x'-0).$$

Señalemos que para la magnitud discreta ξ

$$F_{\xi}(x) = \sum_{y \leq x} P_{\xi}(y),$$

y para la magnitud aleatoria distribuida continuamente

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

donde $p_{\xi}(x)$ es la densidad de probabilidad correspondiente, que coincide en sus puntos de continuidad con la derivada de la función de distribución:

$$p_{\xi}(x) = F'_{\xi}(x). \quad (3.5)$$

2. Distribución conjunta de las probabilidades. Examinemos dos magnitudes aleatorias ξ_1 y ξ_2 dependientes del resultado elemental $\omega \in \Omega$ de una misma prueba.

En calidad de modelo formal se puede tomar la prueba, consistente en la observación de las magnitudes aleatorias ξ_1 , ξ_2 , y más exactamente, en calidad de espacio de los sucesos elementales se puede tomar el plano real E^2 , fijando cada suceso elemental $\{\xi_1 = x_1, \xi_2 = x_2\}$ con el punto correspondiente $(x_1, x_2) \in E^2$ y determinando, formalmente, la dependencia entre las magnitudes ξ_1 , ξ_2 y el resultado ele-

mental $(x_1, x_2) \in E^2$ por la fórmula $\xi_1(x_1, x_2) \equiv x_1$, $\xi_2(x_1, x_2) \equiv x_2$; el «mecanismo de aleatoriedad» de tal prueba se da por la *distribución conjunta de las probabilidades*, es decir, las probabilidades

$$P \{x'_1 \leq \xi_1 \leq x''_1; x'_2 \leq \xi_2 \leq x''_2\}$$

de todos los sucesos posibles $A = \{x'_1 \leq \xi_1 \leq x''_1; x'_2 \leq \xi_2 \leq x''_2\}$.

Señalemos que ξ_1, ξ_2 se pueden considerar como coordenadas de una magnitud vectorial aleatoria $\xi = (\xi_1, \xi_2)$ con una distribución de probabilidades, dadas por las probabilidades de caída del punto ξ en las distintas zonas rectangulares del tipo $A = [x'_1, x''_1] \times [x'_2, x''_2]$:

$$P \{\xi \in A\} = P \{x'_1 \leq \xi_1 \leq x''_1; x'_2 \leq \xi_2 \leq x''_2\}.$$

Para las magnitudes discretas (ξ_1, ξ_2) la distribución de probabilidades se dan por las probabilidades

$P_{\xi_1, \xi_2}(x_1, x_2) = P \{\xi_1 = x_1, \xi_2 = x_2\}$, $-\infty < x_1, x_2 < \infty$,
y precisamente,

$$P \{x'_1 \leq \xi_1 \leq x''_1; x'_2 \leq \xi_2 \leq x''_2\} = \sum_{x'_1}^{x''_1} \sum_{x'_2}^{x''_2} P_{\xi_1, \xi_2}(x_1, x_2), \quad (3.6)$$

donde en los límites indicados la suma se realiza por el número finito o numerable de valores (x_1, x_2) para los cuales $P_{\xi_1, \xi_2}(x_1, x_2) > 0$.

Las distribuciones de las probabilidades de cada una de estas magnitudes ξ_1, ξ_2 pueden ser determinadas por separado según las fórmulas

$$y \left. \begin{aligned} P_{\xi_1}(x_1) &= \sum_{-\infty < x_2 < \infty} P_{\xi_1, \xi_2}(x_1, x_2) \\ P_{\xi_2}(x_2) &= \sum_{-\infty < x_1 < \infty} P_{\xi_1, \xi_2}(x_1, x_2), \end{aligned} \right\} \quad (3.7)$$

que se obtienen de (3.6) para $x'_1 = -\infty, x''_1 = \infty$ y $x'_2 = -\infty, x''_2 = \infty$.

Se dice que las magnitudes aleatorias ξ_1, ξ_2 tienen una *distribución conjunta continua de probabilidades*, si existe la *densidad de probabilidad*, o de otro modo, la *densidad de la distribución conjunta* $p_{\xi_1, \xi_2}(x_1, x_2)$ es una función no negativa integrable, de un par de variables $(x_1, x_2) \in E^2$,

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p_{\xi_1, \xi_2}(x_1, x_2) dx_1 dx_2 = 1,$$

de modo que

$$P \{x'_1 \leq \xi_1 \leq x''_1; x'_2 \leq \xi_2 \leq x''_2\} = \int_{x'_1}^{x''_1} \int_{x'_2}^{x''_2} p_{\xi_1, \xi_2}(x_1, x_2) dx_1 dx_2. \quad (3.8)$$

La fórmula (3.8) se puede volver a escribir en la forma

$$P \{(\xi_1, \xi_2) \in A\} = \iint_A p_{\xi_1, \xi_2}(x_1, x_2) dx_1 dx_2. \quad (3.9)$$

Mostremos, que ésta se puede extender desde los rectángulos $A = [x'_1, x''_1] \times [x'_2, x''_2]$ a las zonas de estructura más compleja, por ejemplo, a cualesquiera zonas A con frontera lisa a trozos.

Observemos primeramente, que la probabilidad de caer en el rectángulo cerrado $[x'_1, x''_1] \times [x'_2, x''_2]$ y en el rectángulo semiabierto $(x'_1, x''_1] \times (x'_2, x''_2]$ es una misma, ya que la probabilidad de que el punto aleatorio (ξ_1, ξ_2) del plano E^2 caiga en la frontera del rectángulo es igual a cero. Esto se puede deducir, fácilmente, de la fórmula (3.8) y de la propiedad de continuidad de la probabilidad.

Pero tratándose del rectángulo semiabierto del tipo $(x'_1, x''_1] \times (x'_2, x''_2]$, utilizando la ley de adición de probabilidades, podemos extender la fórmula (3.8) a los polígonos A , que admiten la división en rectángulos que no se cortan A_h del tipo indicado:

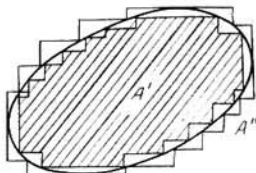


Fig. 8.

$$P \{(\xi_1, \xi_2) \in A\} = \sum_h P \{(\xi_1, \xi_2) \in A_h\} = \sum_h \iint_{A_h} p_{\xi_1, \xi_2}(x_1, x_2) dx_1 dx_2 = \iint_A p_{\xi_1, \xi_2}(x_1, x_2) dx_1 dx_2.$$

Pongamos atención a la circunstancia de que los rectángulos semiabiertos A_h quedan sin cortarse, incluso al tomar contacto uno con otro (lo que no se puede decir de los rectángulos cerrados).

A continuación, examinando ahora la zona A con frontera lisa a trozos, podemos inscribir en ella el polígono A' ($A' \subseteq A$) (fig. 8) del tipo examinado anteriormente, de modo que para cualquier valor dado de $\varepsilon > 0$

$$P \{\xi \in A'\} = \iint_{A'} p_{\xi_1, \xi_2}(x_1, x_2) dx_1 dx_2 \geq \underline{\underline{P}} \geq \iint_A p_{\xi_1, \xi_2}(x_1, x_2) dx_1 dx_2 - \varepsilon.$$

Ya que para $A' \subseteq A$ el suceso $\{\xi \in A'\}$ pertenece al suceso $\{\xi \in A\}$, entonces $P \{\xi \in A\} \geq P \{\xi \in A'\}$ y tanto más

$$P \{\xi \in A\} \geq \iint_A p_{\xi_1, \xi_2}(x_1, x_2) dx_1 dx_2 - \varepsilon.$$

De modo completamente análogo, utilizando el correspondiente polígono A'' , circunscrito junto a la zona A ($A \subseteq A''$), se puede obtener la desigualdad

$$P\{\xi \in A\} \leq \int_A \int p_{\xi_1, \xi_2}(x_1, x_2) dx_1 dx_2 + \varepsilon.$$

Debido a la arbitrariedad de ε , de aquí sacamos la conclusión de que en realidad

$$P\{\xi \in A\} = \int_A \int p_{\xi_1, \xi_2}(x_1, x_2) dx_1 dx_2,$$

lo que se exigía demostrar.

Señalemos, que si las magnitudes aleatorias ξ_1, ξ_2 tienen una distribución conjunta de probabilidades con la densidad $p_{\xi_1, \xi_2}(x_1, x_2)$, entonces, cada una de ellas por separado también tiene densidad de distribución de probabilidades; precisamente, utilizando la fórmula (3.8), para $x_2' = -\infty, x_2'' = \infty$ y $x_1' = -\infty, x_1'' = \infty$, no es difícil comprobar que

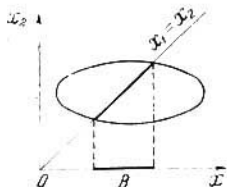


Fig. 9.

$$\left. \begin{aligned} p_{\xi_1}(x_1) &= \int_{-\infty}^{+\infty} p_{\xi_1, \xi_2}(x_1, x_2) dx_2 \\ \text{y} \\ p_{\xi_2}(x_2) &= \int_{-\infty}^{+\infty} p_{\xi_1, \xi_2}(x_1, x_2) dx_1. \end{aligned} \right\} (3.10)$$

La afirmación inversa, hablando en general es falsa. Para las magnitudes aleatorias ξ_1 y ξ_2 que tienen por separado las densidades de distribución $p_{\xi_1}(x_1)$ y $p_{\xi_2}(x_2)$, la densidad de distribución conjunta puede no existir. Por ejemplo, así será, precisamente, en el caso cuando $\xi_1 = \xi_2$ y $P\{(\xi_1, \xi_2) \in A\} = 0$ para cualquier zona A sobre el plano (x_1, x_2) , que no se corta con la recta $x_1 = x_2$; para la zona arbitraria A

$$P\{(\xi_1, \xi_2) \in A\} = \int_B p_{\xi_1}(x) dx,$$

donde B significa un conjunto sobre el eje x_1 , que se obtiene como imagen de la intersección A con la recta $x_1 = x_2$ en su representación $(x_1, x_2) \rightarrow x_1$ (fig. 9).

Las magnitudes aleatorias ξ_1 y ξ_2 se llaman *independientes*, si para cualesquiera intervalos $[x_1', x_1'']$ y $[x_2', x_2'']$ los sucesos $\{x_1' \leq \xi_1 \leq x_1''\}$ y $\{x_2' \leq \xi_2 \leq x_2''\}$ son independientes, es decir,

$$P\{x_1' \leq \xi_1 \leq x_1''; x_2' \leq \xi_2 \leq x_2''\} = P\{x_1' \leq \xi_1 \leq x_1''\} \times P\{x_2' \leq \xi_2 \leq x_2''\}.$$

Como se ve fácilmente, la independencia de las magnitudes discretas ξ_1 y ξ_2 significa, que la distribución conjunta de probabilidades se determina por la fórmula

$$P_{\xi_1, \xi_2}(x_1, x_2) = P_{\xi_1}(x_1) \cdot P_{\xi_2}(x_2). \quad (3.11)$$

En caso de la distribución continua de las magnitudes independientes ξ_1 y ξ_2

$$\begin{aligned} P\{x'_1 \leq \xi_1 \leq x''_1; x'_2 \leq \xi_2 \leq x''_2\} &= \int_{x'_1}^{x''_1} p_{\xi_1}(x_1) \times \\ &\times dx_1 \cdot \int_{x'_2}^{x''_2} p_{\xi_2}(x_2) dx_2 = \int_{x'_1}^{x''_1} \int_{x'_2}^{x''_2} [p_{\xi_1}(x_1) p_{\xi_2}(x_2)] dx_1 dx_2 \end{aligned}$$

para todos los intervalos $[x'_1, x''_1]$ y $[x'_2, x''_2]$. Vemos que se tiene una densidad de distribución conjunta, y precisamente

$$p_{\xi_1, \xi_2}(x_1, x_2) = p_{\xi_1}(x_1) p_{\xi_2}(x_2). \quad (3.12)$$

También se ve fácilmente, que si la densidad de distribución conjunta $p_{\xi_1, \xi_2}(x_1, x_2)$ se expresa por la fórmula (3.12), entonces, las magnitudes aleatorias ξ_1 y ξ_2 son independientes.

3. Transformación de las magnitudes aleatorias. Examinemos la cuestión de cómo varía la distribución de probabilidades de las magnitudes aleatorias durante tal o cual transformación.

Anteriormente, en el § 1 se mostró que para la magnitud aleatoria ξ con densidad de probabilidad $p_{\xi}(x)$, durante la transformación mutuamente unívoca $y = \varphi(x)$, donde $\varphi'(x) \neq 0$, la densidad de probabilidad $p_{\eta}(y)$ de la magnitud $\eta = \varphi(\xi)$ se determina por la fórmula (1.4). El método de sustitución de la variable, utilizado al deducir esta fórmula, también se puede utilizar en el caso cuando en determinados puntos $\varphi'(x) = 0$ y la transformación $y = \varphi(x)$ no es mutuamente unívoca.

Ejemplo. Examinemos la transformación $y = x^2$. Evidentemente, para la magnitud aleatoria $\eta = \xi^2$ tenemos (fig. 10)

$$\begin{aligned} P\{y' \leq \eta \leq y''\} &= P\{-x'' \leq \xi \leq -x'\} + P\{x' \leq \xi \leq x''\} = \\ &= \int_{-x''}^{-x'} p_{\xi}(x) dx + \int_{x'}^{x''} p_{\xi}(x) dx = \int_{x'}^{x''} [p_{\xi}(-x) + p_{\xi}(x)] dx = \\ &= \int_{\sqrt{y'}}^{\sqrt{y''}} [p_{\xi}(-\sqrt{y}) + p_{\xi}(\sqrt{y})] \frac{1}{2\sqrt{y}} dy, \end{aligned}$$

y de tal modo, la densidad de probabilidad de la magnitud aleatoria η es

$$p_{\eta}(y) = \begin{cases} [p_{\xi_1}(-\sqrt{y}) + p_{\xi_1}(\sqrt{y})] \frac{1}{2\sqrt{y}}, & y > 0, \\ 0, & y < 0. \end{cases}$$

En el § 1 fue mostrado que para las magnitudes aleatorias (ξ_1, ξ_2) con densidad de probabilidad $p_{\xi_1, \xi_2}(x_1, x_2)$ durante la transformación mutuamente unívoca del plano $y_1 = \varphi_1(x_1, x_2)$, $y_2 = \varphi_2(x_1, x_2)$ con un jacobiano no degenerado

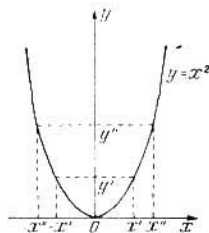


Fig. 10.

$$|I| = \begin{vmatrix} \frac{\partial \varphi_1}{\partial x_1} & \frac{\partial \varphi_1}{\partial x_2} \\ \frac{\partial \varphi_2}{\partial x_1} & \frac{\partial \varphi_2}{\partial x_2} \end{vmatrix}$$

la densidad de probabilidad $p_{\eta_1, \eta_2}(y_1, y_2)$ de las magnitudes aleatorias $\eta_1 = \varphi_1(\xi_1, \xi_2)$, $\eta_2 = \varphi_2(\xi_1, \xi_2)$ se determina por la fórmula (1.7). Esta fórmula también tiene sentido para la transformación, en la cual el jacobiano $|I|$ se convierte en cero en determinadas líneas separadas, que dividen el plano E^2 en zonas que no se cortan, dentro de las cuales el jacobiano $|I|$ no se hace cero.

La fórmula (1.7) también se puede utilizar cuando se trata de la transformación del tipo $\eta = \varphi(\xi_1, \xi_2)$. Si se puede introducir, precisamente, la transformación auxiliar $\eta_1 = \eta$, $\eta_2 = \varphi_2(\xi_1, \xi_2)$ que conduce a la densidad $p_{\eta_1, \eta_2}(y_1, y_2)$ del tipo (1.7), entonces la densidad de probabilidad $p_{\eta}(y)$ de la magnitud η será (véase (3.10))

$$p_{\eta_1}(y_1) = \int_{-\infty}^{\infty} p_{\xi_1, \xi_2}[\psi_1(y_1, y_2), \psi_2(y_1, y_2)] |I|^{-1} dy_2. \quad (3.13)$$

Ejemplo. Sean ξ_1 y ξ_2 magnitudes aleatorias independientes con densidades de distribución de probabilidades $p_1(x)$ y $p_2(x)$. Hallemos la distribución de probabilidades de su relación ξ_1/ξ_2 .

Observemos primeramente, que ya que $\xi_2 = 0$ solamente con una probabilidad igual a cero, la magnitud $\eta_1 = \xi_1/\xi_2$ toma valores finitos con una probabilidad 1. Examinemos la transformación

$$y_1 = \frac{x_1}{x_2}, \quad y_2 = x_2 \quad \left(\eta_1 = \frac{\xi_1}{\xi_2}, \quad \eta_2 = \xi_2 \right)$$

con el jacobiano $|I|^{-1} = |y_2|$. La densidad de probabilidad $p_{\eta_1, \eta_2}(y_1, y_2)$ de las magnitudes η_1, η_2 en cada una de las zonas $y_2 < 0$

y $y_2 > 0$ se puede determinar por la fórmula (1.7)

$$p_{\eta_1, \eta_2}(y_1, y_2) = \begin{cases} p_1(y_1 \cdot y_2) p_2(y_2) y_2 & \text{para } y_2 > 0 \\ -p_1(y_1 \cdot y_2) p_2(y_2) y_2 & \text{para } y_2 < 0. \end{cases}$$

Teniendo en cuenta que la densidad de probabilidad de la magnitud η_1 tomada por separado es

$$p_{\eta_1}(y_1) = \int_{-\infty}^{\infty} p_{\eta_1, \eta_2}(y_1, y_2) dy_2, \quad -\infty < y_1 < \infty,$$

en el caso considerado tenemos

$$p_{\eta_1}(y) = \int_0^{\infty} p_1(yx) p_2(x) x dx - \int_{-\infty}^0 p_1(yx) p_2(x) x dx. \quad (3.14)$$

Ejemplo. Sean ξ_1 y ξ_2 magnitudes aleatorias independientes con densidades de probabilidad $p_1(x)$ y $p_2(x)$. ¿Cuál será la distribución de probabilidades de su suma $\xi_1 + \xi_2$?

La transformación $\varphi_1(x_1, x_2) = x_1 + x_2$ se puede complementar hasta la transformación no degenerada del plano, suponiendo que $\varphi_2(x_1, x_2) = x_2$. Con ello $|I| = 1$ y

$$\psi_2(y_1, y_2) = y_2, \quad \psi_1(y_1, y_2) = y_1 - y_2.$$

Para las magnitudes independientes ξ_1, ξ_2 la densidad conjunta de probabilidad es

$p_{\xi_1, \xi_2}(x_1, x_2) = p_1(x_1) p_2(x_2)$ y para las magnitudes $\eta_1 = \xi_1 + \xi_2, \eta_2 = \xi_2$ la densidad de probabilidad será $p_{\eta_1, \eta_2}(y_1, y_2) = p_1(y_1 - y_2) \cdot p_2(y_2)$. Por consiguiente, la densidad buscada $p_{\eta_1}(y)$ es

$$p_{\eta_1}(y) = \int_{-\infty}^{\infty} p_1(y-x) p_2(x) dx. \quad (3.15)$$

Esta expresión de la densidad de distribución de probabilidades de la magnitud aleatoria $\xi_1 + \xi_2$ se llama *envoltura* o *composición* de las densidades $p_1(x)$ y $p_2(x)$.

Sean, en caso particular, ξ_1, ξ_2 magnitudes aleatorias independientes con distribución uniforme en el segmento $[0, 1]$. La densidad de cada una de ellas es

$$p(x) = \begin{cases} 1 & \text{para } 0 \leq x \leq 1, \\ 0 & \text{para } x < 0, x > 1, \end{cases}$$

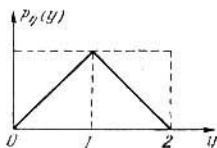


Fig. 11.

y para la suma $\eta = \xi_1 + \xi_2$ obtenemos de la fórmula (3.15) el tal llamado *triángulo de distribución* con densidad (fig. 11)

$$p_{\eta}(y) = \begin{cases} \int_0^y dy = y & \text{para } 0 \leq y \leq 1, \\ \int_{y-1}^1 dx = 2-y & \text{para } 1 \leq y \leq 2 \\ 0 & \text{para } y < 0, y > 2. \end{cases}$$

4. Distribuciones condicionales de probabilidades. El enlace entre distintas magnitudes ξ y η (o de la magnitud ξ con un determinado suceso B), se puede caracterizar por la tal llamada *distribución condicional de probabilidades*. En aquellas condiciones cuando aparece, precisamente, el suceso B , el comportamiento probabilístico de la magnitud aleatoria ξ se describe por las probabilidades condicionales $\mathbf{P}\{x' \leq \xi \leq x'' \mid B\}$, hablando en general, se diferencian de las probabilidades iniciales $\mathbf{P}\{x' \leq \xi \leq x''\}$.

Ejemplo. Imaginémonos que sólo se puede observar una de las dos pruebas Ω_1 o Ω_2 , en cada una de las cuales se examina una determinada magnitud aleatoria (ξ_1 en Ω_1 y ξ_2 en Ω_2). Con objeto de elegir una de las pruebas Ω_1 o Ω_2 , el observador echa a la suerte, complementariamente, de tal modo que se elige a Ω_1 con la probabilidad dada q_1 y se elige a Ω_2 con la probabilidad q_2 ($q_1 + q_2 = 1$). En total, en lugar de las magnitudes ξ_1 y ξ_2 aparece la magnitud aleatoria de la forma $\xi = \eta\xi_1 + (1 - \eta)\xi_2$, donde η no depende de ξ_1 , ξ_2 y toma con las probabilidades q_1 , q_2 los valores 1 ó 0, respectivamente. Se ve fácilmente, que para la condición $\eta = 1$, esta magnitud ξ coincide con ξ_1 y para la condición $\eta = 0$, con la magnitud ξ_2 . De acuerdo con esto, la distribución de probabilidades de la magnitud aleatoria ξ es tal que

$$\mathbf{P}\{x' \leq \xi \leq x'' \mid \eta = 1\} = \mathbf{P}\{x' \leq \xi_1 \leq x''\},$$

$$\mathbf{P}\{x' \leq \xi \leq x'' \mid \eta = 0\} = \mathbf{P}\{x' \leq \xi_2 \leq x''\},$$

y según la fórmula de la probabilidad total

$$\mathbf{P}\{x' \leq \xi \leq x''\} = q_1 \cdot \mathbf{P}\{x' \leq \xi_1 \leq x''\} + q_2 \cdot \mathbf{P}\{x' \leq \xi_2 \leq x''\}.$$

Si decimos que ξ_1 es una magnitud discreta que toma distintos valores de x con las correspondientes probabilidades $P_{\xi_1}(x)$, y ξ_2 es una magnitud distribuida continuamente con una densidad de probabilidad $p_{\xi_2}(x)$, entonces

$$\mathbf{P}\{x' \leq \xi \leq x''\} = q_1 \cdot \sum_{x'}^{x''} P_{\xi_1}(x) + q_2 \cdot \int_{x'}^{x''} p_{\xi_2}(x) dx.$$

Si ξ y η son magnitudes aleatorias discretas, entonces para el suceso B del tipo $\{\eta = y\}$ la magnitud ξ ligada con η ya no tendrá la distribución de probabilidades anterior $P_{\xi}(x)$, sino la tal llamada distribución condicional $P_{\xi}(x|y)$ que de acuerdo con la fórmula (2.4) se determina así:

$$P_{\xi}(x|y) = \frac{P_{\xi, \eta}(x, y)}{P_{\eta}(y)} = \frac{P_{\xi, \eta}(x, y)}{\sum_x P_{\xi, \eta}(x, y)}, \quad -\infty < x < \infty, \quad (3.16)$$

donde x recorre todos los valores posibles de la magnitud discreta ξ , y $P_{\xi, \eta}(x, y)$ es la distribución conjunta de probabilidades de ξ y η . Análogamente, si ξ y η tienen una densidad de distribución conjunta $p_{\xi, \eta}(x, y)$, entonces consideraremos que para el valor fijado $\eta = y$, la magnitud ξ tiene una densidad de distribución condicional $p_{\xi}(x|y)$, que determinamos por la fórmula

$$p_{\xi}(x|y) = \frac{p_{\xi, \eta}(x, y)}{p_{\eta}(y)} = \frac{p_{\xi, \eta}(x, y)}{\int_{-\infty}^{\infty} p_{\xi, \eta}(x, y) dx}, \quad -\infty < x < \infty. \quad (3.17)$$

De las fórmulas (3.16) y (3.17) se ve, directamente, que la magnitud aleatoria ξ no depende de η , cuando, y sólo cuando su distribución de probabilidades condicional (para la condición $\eta = y$) no depende de y y coincide con la distribución inicial (incondicional).

Señalemos las siguientes igualdades (compárese con la fórmula de la probabilidad total):

$$P_{\xi}(x) = \sum_y P_{\xi}(x|y) P_{\eta}(y) \quad (3.18)$$

para las magnitudes discretas ξ y η ,

$$p_{\xi}(x) = \int_{-\infty}^{\infty} p_{\xi}(x|y) p_{\eta}(y) dy \quad (3.19)$$

para las magnitudes ξ y η con densidades de distribución de probabilidades $p_{\xi}(x)$ y $p_{\eta}(y)$.

Para ilustrar, cuán útil puede ser la consideración de tal o cual distribución condicional, nos dirigimos de nuevo a la suma de magnitudes aleatorias independientes ξ_1, ξ_2 con densidad de probabilidad $p_1(x), p_2(x)$. Hallemos la distribución de la magnitud $\eta = \xi_1 + \xi_2$.

Señalemos, primeramente, que la distribución condicional de probabilidades de la magnitud η en condiciones, cuando $\xi_2 = x$, coincide con la distribución de la magnitud $\xi_1 + x$, que tiene una densidad $p_1(y - x)$, $-\infty < y < \infty$ (recordemos, que ξ_1 no depende de ξ_2 , de modo que la densidad condicional de distribución de la magnitud ξ_1 coincide con la densidad incondicional $p_1(y)$). Según

la fórmula general (3.19) obtenemos la expresión ya conocida (3.15):

$$p_{\eta}(y) = \int_{-\infty}^{\infty} p_1(y-x) p_2(x) dx.$$

5. Magnitudes aleatorias multidimensionales. El conjunto de magnitudes aleatorias ξ_1, \dots, ξ_n (que toman los valores numéricos x_1, \dots, x_n) suele ser cómodo examinarlo como una magnitud aleatoria vectorial $\xi = (\xi_1, \dots, \xi_n)$ con las coordenadas ξ_1, \dots, ξ_n , que en dependencia del caso toma tales o cuales valores $x = (x_1, \dots, x_n)$ en el espacio vectorial E^n de n dimensiones.

Para las magnitudes discretas, la probabilidad de caer el punto ξ en tal o cual zona $A \subseteq E^n$ se determina así:

$$P\{\xi \in A\} = \sum_{x \in A} P_{\xi}(x), \quad (3.20)$$

donde las probabilidades

$$P_{\xi}(x) = P\{\xi = x\} = P\{\xi_1 = x_1, \dots, \xi_n = x_n\} = P_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) \quad (3.21)$$

dan la *distribución conjunta de probabilidades* de las magnitudes ξ_1, \dots, ξ_n . Para las magnitudes distribuidas continuamente ξ_1, \dots, ξ_n con la densidad de probabilidad

$$p_{\xi}(x) = p_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) \\ (p_{\xi}(x) \geq 0, \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} p_{\xi}(x) dx_1, \dots, dx_n = 1)$$

la probabilidad del suceso $\{\xi \in A\}$ se determina así:

$$P\{\xi \in A\} = \int_A \dots \int p_{\xi}(x) dx. \quad (3.22)$$

La diferencia con la fórmula (3.9) expuesta anteriormente, aquí sólo consiste en que en lugar de la variable $x = (x_1, x_2)$ en el plano E^2 tenemos $x = (x_1, \dots, x_n) \in E^n$. Todo lo dicho anteriormente, sobre las magnitudes unidimensionales con cambios análogos se extiende a las magnitudes multidimensionales. Particularmente, de una forma evidente se extienden al caso de n variables (x_1, \dots, x_n) las fórmulas (3.7), (3.10) y también las fórmulas de transformación (1.4), (1.7) y (3.13).

La independencia de las magnitudes vectoriales ξ_1 y ξ_2 significa las mismas propiedades (expresadas por las relaciones (3.11); (3.12)), que para las magnitudes corrientes. Sin ningún cambio se determinan también las distribuciones condicionales de probabilidades de la

magnitud vectorial ξ en relación a la magnitud también vectorial η (véase (3.16), (3.17)).

El único concepto, esencialmente nuevo, para el conjunto de las magnitudes aleatorias ξ_1, ξ_2, \dots es su *independencia mutua* (abreviadamente: *independencia*). Precisamente, las magnitudes aleatorias (unidimensionales) ξ_1, ξ_2, \dots se llaman mutuamente independientes, si para cualesquiera $x'_1, x''_1; x'_2, x''_2; \dots$ son mutuamente independientes los sucesos

$$\{x'_1 \leq \xi_1 \leq x''_1\}, \{x'_2 \leq \xi_2 \leq x''_2\}, \dots$$

Para las magnitudes discretas ξ_1, \dots, ξ_n la independencia significa que

$$P_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) = P_{\xi_1}(x_1) \dots P_{\xi_n}(x_n), \quad (3.23)$$

y para las magnitudes distribuidas continuamente con las densidades de probabilidad $p_{\xi_1}(x_1), \dots, p_{\xi_n}(x_n)$, significa que su distribución conjunta de probabilidades tiene la densidad

$$p_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) = p_{\xi_1}(x_1) \dots p_{\xi_n}(x_n). \quad (3.24)$$

§ 4. ESPERANZAS
MATEMATICAS
DE LAS MAGNITUDES
ALEATORIAS

1. Esperanza matemática, definición y algunas fórmulas. Sea ξ una magnitud aleatoria discreta, que toma cada valor posible de x con la correspondiente probabilidad $P_\xi(x)$. Se dice que la magnitud aleatoria ξ tiene una *esperanza matemática* finita $M\xi$:

$$M\xi = \sum_{-\infty}^{\infty} x P_\xi(x) \quad (4.1)$$

(llamada también *valor medio* de esta magnitud aleatoria), si la serie en (4.1) converge absolutamente ($\sum_{-\infty}^{\infty} |x| P_\xi(x) < \infty$).

Ejemplo. El valor medio de la magnitud aleatoria ξ , que toma uno de los N valores posibles $x = x_1, \dots, x_N$ con una misma probabilidad (igual a $\frac{1}{N}$), es

$$M\xi = \frac{1}{N} \sum_{h=1}^N x_h.$$

Examinemos la magnitud aleatoria $\eta = \varphi(\xi)$ que tiene esperanza matemática, donde $y = \varphi(x)$ es una función determinada de la variable x . Para el valor medio $M\varphi(\xi)$ tiene lugar la fórmula

$$M\varphi(\xi) = \sum_{-\infty}^{\infty} \varphi(x) P_{\xi}(x), \quad (4.2)$$

donde la sumación se efectúa por todos los valores de x , para los cuales las probabilidades $P_{\xi}(x) > 0$.

En efecto, la magnitud discreta $\eta = \varphi(\xi)$ puede tomar, solamente, los valores $y = \varphi(x)$, donde x recorre los valores posibles de la magnitud discreta ξ , siendo $P_{\eta}(y) = \sum_{x:\varphi(x)=y} P_{\xi}(x)$, se ve fácilmente

que si la serie $\sum_{-\infty}^{\infty} y P_{\eta}(y)$ converge absolutamente, entonces

$$M\eta = \sum_{-\infty}^{\infty} y P_{\eta}(y) = \sum_y y \sum_{x:\varphi(x)=y} P_{\xi}(x) = \sum_{-\infty}^{\infty} \varphi(x) P_{\xi}(x).$$

Exactamente igual, para el valor medio de la magnitud aleatoria $\eta = \varphi(\xi_1, \xi_2)$, que es una determinada función de las magnitudes ξ_1, ξ_2 con la distribución conjunta de probabilidades $P_{\xi_1, \xi_2}(x_1, x_2)$, tenemos

$$M\varphi(\xi_1, \xi_2) = \sum_{-\infty}^{\infty} \sum_{-\infty}^{\infty} \varphi(x_1, x_2) P_{\xi_1, \xi_2}(x_1, x_2) \quad (4.3)$$

(formalmente (4.3) coincide con (4.2), si suponemos que

$$\xi = (\xi_1, \xi_2) \text{ y } x = (x_1, x_2)).$$

Según la fórmula (4.2), la esperanza matemática de la magnitud $|\xi|$ es $M|\xi| = \sum_{-\infty}^{\infty} |x| P_{\xi}(x)$, de modo que la condición de la

convergencia absoluta de la serie $\sum_{-\infty}^{\infty} x P_{\xi}(x)$ puede ser expresada de la siguiente forma: $M|\xi| < \infty$ (señalemos que la expresión $M|\xi| = \sum_{-\infty}^{\infty} |x| P_{\xi}(x)$ tiene sentido para cualesquiera magnitudes, solamente, puede ocurrir que $M|\xi| = \infty$).

La esperanza matemática posee las siguientes propiedades.

El valor medio de una magnitud constante es igual a sí misma y en particular.

$$M1 = 1.$$

Si existe la esperanza matemática $M\xi$, entonces para cualquier factor constante k , tendremos para la magnitud $k \cdot \xi$ que

$$M(k \cdot \xi) = k \cdot M\xi.$$

Para cualesquiera magnitudes ξ_1 y ξ_2 (que tienen las esperanzas matemáticas $M\xi_1$ y $M\xi_2$)

$$M(\xi_1 + \xi_2) = M\xi_1 + M\xi_2; \quad (4.4)$$

si $\xi_1 \leq \xi_2$, entonces

$$M\xi_1 \leq M\xi_2; \quad (4.5)$$

si las magnitudes aleatorias ξ_1 y ξ_2 son independientes, entonces

$$M(\xi_1 \cdot \xi_2) = M\xi_1 \cdot M\xi_2. \quad (4.6)$$

Para cerciorarse de la justeza de las relaciones indicadas utilizemos las fórmulas (4.2), (4.3). Para $\varphi(x) = k \cdot x$ tenemos

$$M\varphi(\xi) = \sum_{-\infty}^{\infty} k \cdot x P_{\xi}(x) - k \cdot \sum_{-\infty}^{\infty} x P_{\xi}(x) = k M\xi.$$

Tomando $\varphi(x_1, x_2) = x_1 + x_2$, obtenemos

$$M\varphi(\xi_1, \xi_2) = \sum_{-\infty}^{\infty} x_1 \sum_{-\infty}^{\infty} P_{\xi_1, \xi_2}(x_1, x_2) + \sum_{-\infty}^{\infty} x_2 \sum_{-\infty}^{\infty} P_{\xi_1, \xi_2}(x_1, x_2),$$

de donde (véase también (3.7)) se ve, que la igualdad (4.4) en efecto tiene lugar. Para cualquier función no negativa $\varphi(x_1, x_2)$ en todos los valores de x_1, x_2 para los cuales las probabilidades $P_{\xi_1, \xi_2}(x_1, x_2) > 0$,

$$M\varphi(\xi_1, \xi_2) = \sum_{-\infty}^{\infty} \varphi(x_1, x_2) P_{\xi_1, \xi_2}(x_1, x_2) \geq 0,$$

y, en particular, $M(\xi_2 - \xi_1) \geq 0$ para $\xi_2 - \xi_1 \geq 0$, de donde se deduce que en este caso $M\xi_1 \leq M\xi_2$. Para las magnitudes independientes ξ_1 y ξ_2 la distribución conjunta de probabilidades se determina por la fórmula (3.11) y para $\varphi(x_1, x_2) = x_1 \cdot x_2$ tenemos

$$\begin{aligned} M\varphi(\xi_1, \xi_2) &= \sum_{-\infty}^{\infty} \sum_{-\infty}^{\infty} (x_1 \cdot x_2) P_{\xi_1}(x_1) P_{\xi_2}(x_2) = \\ &= \sum_{-\infty}^{\infty} x_1 P_{\xi_1}(x_1) \cdot \sum_{-\infty}^{\infty} x_2 P_{\xi_2}(x_2) = M\xi_1 \cdot M\xi_2. \end{aligned}$$

La fórmula (4.4) se extiende por inducción a cualquier número de magnitudes aleatorias ξ_1, \dots, ξ_n que tienen los valores medios:

$$M(\xi_1 + \dots + \xi_n) = M\xi_1 + \dots + M\xi_n;$$

también se puede decir lo mismo sobre la fórmula (4.6); para cualesquiera magnitudes mutuamente independientes ξ_1, \dots, ξ_n

$$M(\xi_1 \cdot \dots \cdot \xi_n) = M\xi_1 \cdot \dots \cdot M\xi_n.$$

Volvamos a la magnitud aleatoria arbitraria ξ (no obligatoriamente discreta). La definición de la esperanza matemática $M\xi$ está basada en que cualquier magnitud ξ se puede aproximar con magnitudes discretas, tan exactamente, como se desee. Por ejemplo, si dividimos la recta real $-\infty < x < \infty$ con los puntos $x_{k,n}$ ($-\infty < k < \infty$, $\sup_k |x_{k,n} - x_{k-1,n}| = \varepsilon_n$) y determinamos las magnitudes discretas aleatorias ξ_n así:

$$\xi_n = x_{k,n} \text{ para } x_{k-1,n} < \xi \leq x_{k,n},$$

entonces, evidentemente $|\xi_n - \xi| \leq \varepsilon_n$ y para $\varepsilon_n \rightarrow 0$ la sucesión de magnitudes discretas ξ_n , $n = 1, 2, \dots$, da la aproximación tan exacta como se desee para la magnitud inicial ξ . Si las magnitudes ξ_n tienen las esperanzas matemáticas $M\xi_n$ (es decir, $M|\xi_n| < \infty$), entonces, según la desigualdad general (4.5)

$$\begin{aligned} |M\xi_n - M\xi_m| &\leq M|\xi_n - \xi_m| \leq \varepsilon_n + \varepsilon_m \rightarrow 0 \\ (|\xi_n - \xi_m| &\leq |\xi_n - \xi| + |\xi_m - \xi| \leq \varepsilon_n + \varepsilon_m) \end{aligned}$$

y, por consiguiente, existe el límite $\lim_{n \rightarrow \infty} M\xi_n$. Este límite se llama *esperanza matemática* (o *valor medio*) de la magnitud aleatoria ξ :

$$M\xi = \lim_{n \rightarrow \infty} \sum_{-\infty}^{\infty} x_{k,n} P(x_{k-1,n} < \xi \leq x_{k,n}). \quad (4.7)$$

(Se ve, fácilmente, que para cualquier sucesión de magnitudes ξ_n , $n = 1, 2, \dots$, del tipo indicado, o en otras palabras, para cualquier división de la recta real en intervalos $(x_{k-1,n}, x_{k,n}]$, para $\sup_k |x_{k,n} - x_{k-1,n}| \rightarrow 0$, el valor límite $M\xi$ en (4.7) será el mismo).

Evidentemente, las relaciones (4.4)–(4.6) se conservan durante el paso límite utilizado por nosotros desde las magnitudes discretas hasta las magnitudes aleatorias arbitrarias.

Demostremos que *para la magnitud aleatoria ξ con la densidad de probabilidad $p_\xi(x)$ la esperanza matemática será*

$$M\xi = \int_{-\infty}^{\infty} x p_\xi(x) dx \quad (4.8)$$

(donde la integral converge absolutamente: $\int_{-\infty}^{\infty} |x| p_\xi(x) dx < \infty$).

En efecto, suponiendo $\varphi(x) = x$ y $\varphi_n(x) = x_{k,n}$ para $x_{k-1,n} < x \leq x_{k,n}$ tenemos

$$M\xi_n = \sum_{-\infty}^{\infty} \int_{x_{k-1,n}}^{x_{k,n}} x_{k,n} p_\xi(x) dx = \int_{-\infty}^{\infty} \varphi_n(x) p_\xi(x) dx,$$

donde

$$\left| \int_{-\infty}^{\infty} \varphi_n(x) p_{\xi}(x) dx - \int_{-\infty}^{\infty} \varphi(x) p_{\xi}(x) dx \right| \leq \int_{-\infty}^{\infty} |\varphi_n(x) - \varphi(x)| \times \\ \times p_{\xi}(x) dx \leq \varepsilon_n \int_{-\infty}^{\infty} p_{\xi}(x) dx = \varepsilon_n \rightarrow 0 \text{ para } n \rightarrow \infty$$

y, por consiguiente, $M_{\xi_n} \rightarrow \int_{-\infty}^{\infty} \varphi(x) p_{\xi}(x) dx$, es decir $M_{\xi} =$

$$= \int_{-\infty}^{\infty} x p_{\xi}(x) dx. \text{ (Se ve fácilmente, que la condición } M|\xi_n| =$$

$$= \int_{-\infty}^{\infty} |\varphi_n(x)| p_{\xi}(x) dx < \infty \text{ presupuesta en la definición de la espe-}$$

ranza matemática M_{ξ} es equivalente a que $\int_{-\infty}^{\infty} |\varphi(x)| p_{\xi}(x) dx < \infty,$

$\varphi(x) = x$.)

Utilizando un método análogo de aproximación en la función $y = \varphi(x)$ con funciones constantes a trozos $\varphi_n(x)$, para la magnitud aleatoria $\eta = \varphi(\xi)$ en condiciones que $\int_{-\infty}^{\infty} |\varphi(x)| p_{\xi}(x) dx < \infty$ se puede deducir que

$$M_{\varphi(\xi)} = \int_{-\infty}^{\infty} \varphi(x) p_{\xi}(x) dx. \quad (4.9)$$

La misma fórmula tiene lugar cuando $\xi = (\xi_1, \xi_2)$ para la magnitud aleatoria $\eta = \varphi(\xi_1, \xi_2)$: o más exacto,

$$M_{\varphi(\xi_1, \xi_2)} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \varphi(x_1, x_2) p_{\xi_1, \xi_2}(x_1, x_2) dx_1 dx_2, \quad (4.10)$$

donde $p_{\xi_1, \xi_2}(x_1, x_2)$ es la densidad de probabilidad de las magnitudes ξ_1, ξ_2 , con ello, la condición de existencia de la esperanza matemática finita ($M|\varphi(\xi_1, \xi_2)| < \infty$) es equivalente a que

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |\varphi(x_1, x_2)| p_{\xi_1, \xi_2}(x_1, x_2) dx_1 dx_2 < \infty.$$

Ejemplo. Sea ξ la magnitud aleatoria distribuida uniformemente en el segmento $[a, b]$, es decir, que tiene una densidad de probabilidad

de la forma

$$p_{\xi}(x) = \begin{cases} \frac{1}{b-a} & \text{para } a \leq x \leq b, \\ 0 & \text{para } x < a, x > b. \end{cases}$$

Entonces su valor medio es

$$M\xi = \int_a^b x p_{\xi}(x) dx = \frac{a+b}{2}.$$

Examinemos la magnitud aleatoria ξ de «tipo combinado» (véase el ejemplo en la pág. 42): ξ toma un valor determinado de x con la probabilidad correspondiente $P_{\xi}(x)$ ($\sum_{-\infty}^{\infty} P_{\xi}(x) = q_1$), y en general cae en tal o cual intervalo $[x', x'']$ con la probabilidad

$$P\{x' \leq \xi \leq x''\} = \sum_{x'}^{x''} P_{\xi}(x) + \int_{x'}^{x''} p_{\xi}(x) dx,$$

donde $p_{\xi}(x)$ es una función no negativa, $\int_{-\infty}^{\infty} p_{\xi}(x) dx = q_2$, $q_1 +$

$+ q_2 = 1$, de modo que las probabilidades $P_{\xi}(x) = q_1^{-1} \cdot P_{\xi}(x)$, $-\infty < x < \infty$, y la densidad $p_{\xi_2}(x) = q_2^{-1} \cdot p_{\xi}(x)$, $-\infty < x < \infty$, dan, correspondientemente, las distribuciones de probabilidades de la magnitud discreta, digamos ξ_1 y la magnitud ξ_2 distribuida continuamente. Para tal magnitud ξ , la esperanza matemática $M\xi$ puede ser expresada por la fórmula (compárese con (4.1), (4.8))

$$M\xi = q_1 M\xi_1 + q_2 M\xi_2 = \sum_{-\infty}^{\infty} x P_{\xi}(x) + \int_{-\infty}^{\infty} x p_{\xi}(x) dx. \quad (4.11)$$

El concepto de esperanza matemática $M\xi$ también se extiende a las magnitudes aleatorias ξ , que toman valores complejos: $\xi = \xi_1 + i\xi_2$, donde ξ_1 y ξ_2 son magnitudes aleatorias reales. Precisamente, se supone que

$$M\xi = M\xi_1 + iM\xi_2.$$

Se ve, fácilmente, que todas las propiedades citadas de las esperanzas matemáticas también tienen lugar para las magnitudes complejas de la forma $\xi = \xi_1 + i\xi_2$ (y en general, para las funciones de valores complejos $\varphi(\xi_1, \xi_2)$; véanse en particular, las fórmulas (4.3) y (4.10)).

Como conclusión de este apartado observemos, que las fórmulas, anteriormente indicadas, (4.2), (4.3) y (4.9), (4.10), de forma evidente

se extienden a las magnitudes aleatorias multidimensionales $\xi = (\xi_1, \dots, \xi_n)$.

2. Momentos, dispersión y desigualdad de Chebishev. Se llama *momento* (de orden k) de la magnitud aleatoria ξ al valor medio $M\xi^k$. De acuerdo a las fórmulas generales (4.2) y (4.9)

$$M\xi^k = \sum_{-\infty}^{\infty} x^k P_{\xi}(x)$$

para las magnitudes discretas y

$$M\xi^k = \int_{-\infty}^{\infty} x^k p_{\xi}(x) dx$$

para las magnitudes distribuidas continuamente.

Se llama *segundo momento* $M\xi^2$ al valor de la *media cuadrática* de la magnitud aleatoria ξ .

Es justa la siguiente *desigualdad (Cauchy—Bunjakovsky)*:

$$M|\xi_1 \cdot \xi_2| \leq \sqrt{M\xi_1^2} \cdot \sqrt{M\xi_2^2} \quad (4.12)$$

para cualesquiera magnitudes ξ_1, ξ_2 (qué tienen un segundo momento finito).

En realidad, ya que $|\xi_1 \cdot \xi_2| \leq 1/2(\xi_1^2 + \xi_2^2)$, de la limitación de los segundos momentos $M\xi_1^2$ y $M\xi_2^2$ se deduce, que $M|\xi_1 \cdot \xi_2| < \infty$. Luego, la forma cuadrática determinada positivamente

$$M(x_1|\xi_1| + x_2|\xi_2|)^2 = x_1^2 \cdot M\xi_1^2 + 2x_1x_2 \cdot M|\xi_1\xi_2| + x_2^2 \cdot M\xi_2^2$$

de las variables x_1, x_2 tiene un determinante no negativo, igual a $M\xi_1^2 \cdot M\xi_2^2 - (M|\xi_1 \cdot \xi_2|)^2$, lo que da la desigualdad (4.12).

Suponiendo que $\xi_1 = \xi, \xi_2 = 1$, obtenemos de la fórmula (4.12) que

$$M|\xi| \leq \sqrt{M\xi^2}. \quad (4.13)$$

De (4.12) también se deduce la siguiente desigualdad importante

$$\sqrt{M|\xi_1 - \xi_2|^2} \leq \sqrt{M\xi_1^2} + \sqrt{M\xi_2^2}. \quad (4.14)$$

En efecto,

$$M|\xi_1 + \xi_2|^2 = M\xi_1^2 + 2M\xi_1\xi_2 + M\xi_2^2 \leq M\xi_1^2 + 2\sqrt{M\xi_1^2} \sqrt{M\xi_2^2} + M\xi_2^2 = (\sqrt{M\xi_1^2} + \sqrt{M\xi_2^2})^2.$$

Luego, tiene lugar la siguiente *desigualdad de Chebishev*:

$$P\{|\xi| > \varepsilon\} \leq \frac{M|\xi|^2}{\varepsilon^2}, \quad (4.15)$$

para cualquiera $\varepsilon > 0$.

Esta se puede deducir fácilmente, de la desigualdad general (4.5). En efecto, si suponemos que

$$\xi_1 = \begin{cases} 0 & \text{para } |\xi| \leq \varepsilon, \\ \varepsilon & \text{para } |\xi| > \varepsilon, \end{cases}$$

entonces, evidentemente, $\xi_1 \leq |\xi|$ y

$$M|\xi|^2 \geq M\xi_1^2 = \varepsilon^2 \cdot P\{|\xi| > \varepsilon\},$$

que es equivalente a la relación (4.15).

La desigualdad de Chebishev muestra que si el valor de la media cuadrática $M\xi^2$ es pequeño en relación a ε^2 , digamos $M\xi^2/\varepsilon^2 \leq \delta$, y se puede despreciar, prácticamente, la posibilidad de realización del suceso $\{|\xi| > \varepsilon\}$ de pequeña probabilidad δ , entonces también será pequeña la propia magnitud aleatoria ξ , a saber: $|\xi| \leq \varepsilon$; en particular, si $M|\xi|^2 = 0$, entonces $\xi = 0$ con la probabilidad 1 (en general, si $\xi \geq 0$ y $M\xi = M|\sqrt{\xi}|^2 = 0$, entonces $\xi = 0$ con la probabilidad 1).

Examinemos la diferencia $\xi - a$, donde $a = M\xi$ es la esperanza matemática de la magnitud ξ ; se llama *dispersión* de la magnitud aleatoria ξ , y se designa por $D\xi$ al valor medio cuadrático de la diferencia $\xi - a$, es decir

$$D\xi = M(\xi - a)^2.$$

Según la desigualdad de Chebishev,

$$P\{|\xi - a| > \varepsilon\} \leq \frac{D\xi}{\varepsilon^2},$$

de modo que para una dispersión relativamente pequeña $D\xi$, la magnitud aleatoria ξ toma con una gran probabilidad un valor cercano al a (en este sentido también se diferencia poco de la constante a):

$$P\{|\xi - a| \leq \varepsilon\} \geq 1 - \frac{D\xi}{\varepsilon^2},$$

en particular, si $D\xi = 0$, entonces $\xi = a$ con la probabilidad 1 (evidentemente, la dispersión de la magnitud constante $\xi = a$ será siempre igual a 0).

Señalemos las siguientes propiedades de la dispersión. Tiene lugar la fórmula

$$D\xi = M\xi^2 - a^2 \quad (a = M\xi).$$

(En realidad, $M(\xi - a)^2 = M\xi^2 - 2a \cdot M\xi + a^2 = M\xi^2 - a^2$).

Para cualquier factor constante k , tenemos para la magnitud aleatoria $k \cdot \xi$ que

$$D(k \cdot \xi) = k^2 \cdot D\xi.$$

Sean ξ_1 y ξ_2 magnitudes aleatorias independientes. Entonces

$$D(\xi_1 + \xi_2) = D\xi_1 + D\xi_2. \quad (4.16)$$

En efecto, si suponemos que $a_1 = M\xi_1$ y $a_2 = M\xi_2$, entonces, según la relación (4.6) tendremos

$$M(\xi_1 - a_1)(\xi_2 - a_2) = M(\xi_1 - a_1) \cdot M(\xi_2 - a_2) = 0$$

y

$$\begin{aligned} D(\xi_1 + \xi_2) &= M[(\xi_1 - a_1) + (\xi_2 - a_2)]^2 = \\ &= M(\xi_1 - a_1)^2 + 2M(\xi_1 - a_1)(\xi_2 - a_2) + \\ &+ M(\xi_2 - a_2)^2 = M(\xi_1 - a_1)^2 + M(\xi_2 - a_2)^2 = D\xi_1 + D\xi_2. \end{aligned}$$

Evidentemente, la fórmula (4.16) se extiende a cualquier número de magnitudes aleatorias mutuamente independientes ξ_1, \dots, ξ_n :

$$D(\xi_1 + \dots + \xi_n) = D\xi_1 + \dots + D\xi_n.$$

3. Esperanza matemática condicional. Si es conocido, que aparece el suceso determinado B , el comportamiento probable de la magnitud aleatoria ξ (ligada de algún modo al suceso B) se describe por las probabilidades condicionales $P\{x' \leq \xi \leq x'' | B\}$. El valor medio de la magnitud ξ en relación a tal distribución condicional de probabilidades se llama *esperanza matemática condicional* (o *valor medio condicional*) y se designa así: $M(\xi | B)$.

Por ejemplo, para la magnitud aleatoria $\xi = \eta\xi_1 + (1 - \eta)\xi_2$, donde η no depende de ξ_1, ξ_2 y toma los valores correspondientes 1 ó 0 con las probabilidades q_1, q_2 ($q_1 + q_2 = 1$) (véase el ejemplo de la pág. 42), su *esperanza matemática condicional*, al realizarse el suceso $\{\eta = 1\}$ (en este caso ξ coincide con ξ_1), es $M(\xi | \eta = 1) = M\xi_1$ y, al ocurrir el suceso $\{\eta = 0\}$ (en este caso ξ coincide con ξ_2), es $M(\xi | \eta = 0) = M\xi_2$.

Para la magnitud discreta ξ con la distribución condicional $P_\xi(x | B) = P\{\xi = x | B\}$ la *esperanza matemática condicional*, según la fórmula (4.1), es

$$M(\xi | B) = \sum_{-\infty}^{\infty} x P_\xi(x | B).$$

En particular, si se trata del suceso B del tipo $\{\eta = y\}$, entonces, para las magnitudes discretas ξ y η el valor medio correspondiente es

$$M(\xi | y) = \sum_{-\infty}^{\infty} x P_\xi(x | y),$$

donde $P_\xi(x | y)$ se da por la fórmula (3.16). El valor medio condicional de la magnitud ξ para las magnitudes distribuidas continuamente,

con la condición $\{\eta = y\}$, será (compárese con (4.8))

$$M(\xi | y) = \int_{-\infty}^{\infty} x p_{\xi}(x | y) dx,$$

donde la densidad condicional de probabilidad $p_{\xi}(x | y)$ se da por la fórmula (3.17). Análogamente se determina el valor medio condicional $M(\xi | y_1, \dots, y_n)$ en relación a determinadas magnitudes η_1, \dots, η_n (con la condición de que $\eta_1 = y_1, \dots, \eta_n = y_n$). La distribución condicional de probabilidades correspondiente se da para las magnitudes discretas $\xi, \eta_1, \dots, \eta_n$ por las probabilidades

$$P_{\xi}(x | y_1, \dots, y_n) = \frac{P_{\xi, \eta_1, \dots, \eta_n}(x, y_1, \dots, y_n)}{P_{\eta_1, \dots, \eta_n}(y_1, \dots, y_n)},$$

y para las magnitudes distribuidas continuamente por la densidad de probabilidad

$$p_{\xi}(x | y_1, \dots, y_n) = \frac{p_{\xi, \eta_1, \dots, \eta_n}(x, y_1, \dots, y_n)}{p_{\eta_1, \dots, \eta_n}(y_1, \dots, y_n)}.$$

En adelante para la unificación de las designaciones en el caso multidimensional, en lugar de (y_1, \dots, y_n) simplemente escribiremos y . También con este fin, en lugar del suceso B examinaremos la magnitud η , que será igual a 1 cuando aparezca B e igual a 0 en caso contrario; entonces $M(\xi | B) = M(\xi | y)$, $y = 1$.

Evidentemente, para un valor fijado de y la esperanza matemática condicional $M(\xi | y)$ posee todas las propiedades de las esperanzas matemáticas establecidas anteriormente. En particular,

$$M(k \cdot \xi | y) = k \cdot M(\xi | y)$$

para cualquier constante k ,

$$M(\xi_1 + \xi_2 | y) = M(\xi_1 | y) + M(\xi_2 | y),$$

etc. (véanse (4.2)–(4.11)). Señalemos especialmente las siguientes propiedades:

a) si ξ no depende de la magnitud η , entonces

$$M(\xi | y) = M\xi; \quad (4.17)$$

b) si ξ_1 no depende del par de magnitudes (ξ_2, η) , entonces

$$M(\xi_1 \cdot \xi_2 | y) = M(\xi_1) \cdot M(\xi_2 | y). \quad (4.18)$$

La igualdad (4.17) es evidente. La igualdad (4.18) se obtiene, exactamente, igual que (4.6), ya que incluso en condiciones $\{\eta = y\}$ las magnitudes ξ_1 y ξ_2 quedan independientes, su distribución con-

condicional conjunta es tal que

$$P_{\xi_1, \xi_2}(x_1, x_2 | y) = \frac{P_{\xi_1, \xi_2, \eta}(x_1, x_2, y)}{P_\eta(y)} = \frac{P_{\xi_1}(x_1) \cdot P_{\xi_2, \eta}(x_2, y)}{P_\eta(y)} = P_{\xi_1}(x_1) \cdot P_{\xi_2}(x_2 | y)$$

(para las magnitudes discretas) y

$$p_{\xi_1, \xi_2}(x_1, x_2 | y) = \frac{p_{\xi_1, \xi_2, \eta}(x_1, x_2, y)}{p_\eta(y)} = \frac{p_{\xi_1}(x_1) \cdot p_{\xi_2, \eta}(x_2, y)}{p_\eta(y)} = p_{\xi_1}(x_1) p_{\xi_2}(x_2 | y)$$

(para las magnitudes distribuidas continuamente con una densidad conjunta $p_{\xi_1, \xi_2, \eta}(x_1, x_2, y)$). Naturalmente, la igualdad (4.18) es también justa para la magnitud vectorial $\eta = (\eta_1, \dots, \eta_n)$, cuando $y = (y_1, \dots, y_n)$.

Luego señalemos la fórmula de la *esperanza matemática total*, análoga a (2.5):

$$M\xi = \sum_k M(\xi | B_k) \cdot P(B_k),$$

donde B_1, B_2, \dots , es el sistema total de sucesos que no se cortan. Del mismo tipo son las siguientes fórmulas:

$$M\xi = \sum_{-\infty}^{\infty} M(\xi | y) P_\eta(y)$$

para la magnitud discreta η con la distribución $P_\eta(y)$;

$$M\xi = \int_{-\infty}^{\infty} M(\xi | y) p_\eta(y) dy$$

para la magnitud η que tiene la densidad de probabilidad $p_\eta(y)$ (estas igualdades se deducen, directamente, de la propia definición de la esperanza matemática condicional).

Si consideramos el valor medio condicional $M(\xi | y)$ como función de la variable y , y en lugar de y ponemos el valor de la magnitud aleatoria η , entonces $M(\xi | \eta)$ se puede considerar como una magnitud aleatoria (que es una función determinada de η). Con tal interpretación la fórmula de la esperanza matemática total toma la siguiente forma:

$$M[M(\xi | \eta)] = M\xi. \quad (4.19)$$

Observemos también la tal llamada fórmula de la *esperanza matemática repetida*¹⁾ que generaliza la igualdad (4.19) en el caso, cuando

¹⁾ Véase más detalladamente sobre la cuestión por ejemplo, el libro de I. I. Guijman y A. V. Skorjod, citado anteriormente.

la distribución inicial de probabilidades por sí misma es condicional:

$$\mathbf{M} \{ \mathbf{M} [\xi | \eta] | \zeta \} = \mathbf{M} \{ \xi | \zeta \},$$

donde la magnitud aleatoria (multidimensional) ξ es una función de η .

Señalemos la igualdad importante siguiente:

$$\mathbf{M} (\varphi (\eta) \cdot \xi | \eta) = \varphi (\eta) \cdot \mathbf{M} (\xi | \eta), \quad (4.20)$$

donde $\varphi (\eta)$ significa una magnitud aleatoria, que es una función de η .

En efecto, para cada y fijada, el valor de $\varphi (y)$ es una constante, de modo que

$$\mathbf{M} [\varphi (y) \cdot \xi | y] = \varphi (y) \cdot \mathbf{M} (\xi | y).$$

Para la magnitud aleatoria ξ que es el indicador del suceso A ,

$$\xi = \begin{cases} 1 & \text{para la aparición de } A, \\ 0 & \text{en caso contrario,} \end{cases}$$

es evidente que $\mathbf{M} \xi = \mathbf{P} (A)$. La esperanza matemática condicional de tal magnitud ξ para el valor fijado de η , no es otra cosa que la probabilidad condicional del suceso A :

$$\mathbf{M} (\xi | \eta) = \mathbf{P} (A | \eta);$$

en este caso la relación (4.19) da la siguiente fórmula de la probabilidad total:

$$\mathbf{P} (A) = \mathbf{M} [\mathbf{P} (A | \eta)]. \quad (4.21)$$

Hasta ahora no hemos tocado la cuestión sobre si existe la esperanza matemática condicional finita $\mathbf{M} (\xi | \eta)$ para tal o cual valor de η . La existencia de $\mathbf{M} (\xi | \eta)$ significa que $\mathbf{M} (| \xi | | \eta) < \infty$ (véase el punto 1 en que se define la esperanza matemática). Si existe el valor medio incondicional finito $\mathbf{M} | \xi |$ ($\mathbf{M} | \xi | < \infty$), entonces la fórmula (4.20) de la esperanza matemática total nos da la siguiente relación:

$$\mathbf{M} [\mathbf{M} (| \xi | | \eta)] = \mathbf{M} | \xi | < \infty,$$

de donde se deduce que $\mathbf{M} (| \xi | | \eta) < \infty$ con la probabilidad 1.

En efecto, la probabilidad de que la magnitud aleatoria η tome un valor, para el cual $\xi = \mathbf{M} (| \xi | | \eta)$ no sea mayor que N , según la desigualdad de Chebishev es

$$\mathbf{P} \{ \xi \leq N \} = \mathbf{P} \{ \sqrt{\xi} \leq \sqrt{N} \} > 1 - \frac{\mathbf{M} \xi}{N} = 1 - \frac{\mathbf{M} | \xi |}{N},$$

de aquí inmediatamente se deduce, que

$$\mathbf{P} \{ \mathbf{M} (| \xi | | \eta) < \infty \} = \lim_{N \rightarrow \infty} \mathbf{P} \{ \mathbf{M} (| \xi | | \eta) < N \} = 1.$$

Así pues, si $\mathbf{M} |\xi| < \infty$, entonces existe la esperanza matemática condicional $\mathbf{M}(\xi | \eta)$ para cualquier valor de la magnitud aleatoria η con la probabilidad 1.

4. Distancia media cuadrática y coeficiente de correlación. Designemos por H al conjunto de todas las magnitudes aleatorias ξ , para las cuales $\mathbf{M}\xi^2 < \infty$. Como se deduce de la desigualdad (4.14), para las magnitudes arbitrarias $\xi_1, \xi_2 \in H$, en H entra cualquier combinación lineal $\xi = c_1\xi_1 + c_2\xi_2$ y, en particular, la diferencia $\xi_1 - \xi_2$.

Introduzcamos la distancia media cuadrática entre las magnitudes $\xi_1, \xi_2 \in H$, determinándola como

$$\|\xi_1 - \xi_2\| = \sqrt{\mathbf{M}|\xi_1 - \xi_2|^2}.$$

Esta es análoga a la distancia media cuadrática admitida en el análisis funcional, entre las funciones $\varphi_1(x)$ y $\varphi_2(x)$ de la variable x , $a \leq x \leq b$, que se determina así:

$$\|\varphi_1 - \varphi_2\| = \left(\int_a^b |\varphi_1(x) - \varphi_2(x)|^2 f(x) dx \right)^{1/2},$$

donde $f(x) \geq 0$ es una determinada función llamada peso; aún más,

para la condición $\int_a^b f(x) dx = 1$ tenemos (véase la fórmula (4.9))

$$\|\varphi_1 - \varphi_2\| = (\mathbf{M}|\varphi_1(\xi) - \varphi_2(\xi)|^2)^{1/2},$$

donde ξ , $a \leq \xi \leq b$, es una magnitud aleatoria con la densidad de probabilidad $f(x)$, $a \leq x \leq b$.

La distancia media cuadrática $\|\xi_1 - \xi_2\|$ muestra en grado conocido la gran diferencia entre sí de las magnitudes correspondientes ξ_1, ξ_2 ; en particular, se puede afirmar a base de la desigualdad de Chebishev, que los valores ξ_1, ξ_2 con una probabilidad no menor que $1 - \frac{1}{\varepsilon^2} \|\xi_1 - \xi_2\|^2$ se diferenciarán en no más de ε :

$$\mathbf{P}\{|\xi_1 - \xi_2| \leq \varepsilon\} \geq 1 - \frac{1}{\varepsilon^2} \|\xi_1 - \xi_2\|^2.$$

En este sentido, como ya se indicó, la dispersión $\mathbf{D}\xi = \|\xi - a\|^2$ muestra, en los distintos resultados aleatorios ω , cuán grande es la desviación de la magnitud $\xi(\omega)$ desde su valor medio $a = \mathbf{M}\xi$; a propósito, para cualquier constante c_0

$$\begin{aligned} \mathbf{M}(\xi - c_0)^2 &= \mathbf{M}[(\xi - a) + (a - c_0)]^2 = \\ &= \mathbf{M}(\xi - a)^2 + 2(a - c_0) \cdot \mathbf{M}(\xi - a) + (a - c_0)^2 = \\ &= \mathbf{M}(\xi - a)^2 + (a - c_0)^2 \geq \mathbf{M}(\xi - a)^2 = \mathbf{D}\xi, \end{aligned}$$

de modo que el valor medio $a = \mathbf{M}\xi$ es aquella constante, de la cual se desvía lo menos posible la magnitud aleatoria ξ en el sentido de

distancia media cuadrática:

$$\min_{c_0} \|\xi - c_0\| = \|\xi - a\|. \quad (4.22)$$

Examinemos las magnitudes aleatorias ξ_1 y ξ_2 . Planteemos la siguiente tarea: hallar una combinación del tipo $\hat{c}_1 + \hat{c}_2\xi_2$ (donde \hat{c}_1 y \hat{c}_2 son determinadas constantes), que nos dé la mejor aproximación de la magnitud aleatoria ξ_1 en tal sentido que

$$\mathbf{M}(\xi_1 - \hat{c}_1 - \hat{c}_2\xi_2)^2 = \min_{c_1, c_2} \mathbf{M}(\xi_1 - c_1 - c_2\xi_2)^2 \quad (4.23)$$

(el mínimo se toma por todas las constantes c_1 y c_2).

Supongamos que

$$r = \frac{\mathbf{M}(\xi_1 - a_1)(\xi_2 - a_2)}{\sigma_1\sigma_2}, \quad (4.24)$$

donde

$$a_1 = \mathbf{M}\xi_1, \quad \sigma_1^2 = \mathbf{D}\xi_1 \quad \text{y} \quad a_2 = \mathbf{M}\xi_2, \quad \sigma_2^2 = \mathbf{D}\xi_2.$$

Pasemos para comodidad a las magnitudes aleatorias normalizadas

$$\eta_1 = \frac{\xi_1 - a_1}{\sigma_1} \quad \text{y} \quad \eta_2 = \frac{\xi_2 - a_2}{\sigma_2}.$$

Para cualesquiera constantes c_1 y c_2 tenemos

$$\begin{aligned} \mathbf{M}(\eta_1 - c_1 - c_2\eta_2)^2 &= \mathbf{M}[(\eta_1 - r\eta_2) - c_1 + (r - c_2)\eta_2]^2 = \\ &= (1 - r^2) + c_1^2 + (r - c_2)^2. \end{aligned}$$

Se ve que el mínimo de la expresión $\mathbf{M}(\eta_1 - c_1 - c_2\eta_2)^2$ se alcanza, cuando $c_1 = 0$ y $c_2 = r$:

$$\min_{c_1, c_2} \mathbf{M}(\eta_1 - c_1 - c_2\eta_2)^2 = \|\eta_1 - r\eta_2\|^2 = 1 - r^2. \quad (4.25)$$

Expresando la diferencia $\eta_1 - r\eta_2$ a través de las magnitudes iniciales ξ_1 y ξ_2 obtenemos

$$\eta_1 - r\eta_2 = \frac{1}{\sigma_1} \left[\xi_1 - a_1 - r \frac{\sigma_1}{\sigma_2} (\xi_2 - a_2) \right],$$

y evidentemente la combinación lineal buscada $\hat{c}_1 + \hat{c}_2\xi_2$ en (4.23) es

$$\hat{c}_1 + \hat{c}_2\xi_2 = a_1 + r \frac{\sigma_1}{\sigma_2} (\xi_2 - a_2). \quad (4.26)$$

Aquí a_1 y a_2 son las esperanzas matemáticas de las magnitudes aleatorias ξ_1 y ξ_2 , σ_1^2 y σ_2^2 son sus dispersiones, y la constante r determinada por la igualdad (4.24) es el tal llamado *coeficiente de correlación* de estas magnitudes aleatorias, que es la característica de enlace más elemental de entre ξ_1 y ξ_2 .

Como se ve de la fórmula (4.25) el coeficiente de correlación r siempre se halla entre los límites $-1 \leq r \leq 1$, con ello, si $r = -1$ ó $r = 1$, entonces la magnitud aleatoria ξ_1 es, sencillamente, una combinación lineal de la forma $\xi_1 = \hat{c}_1 + \hat{c}_2 \xi_2$.

En efecto, si $r = -1$ ó $r = 1$, entonces, según la fórmula (4.25) el valor de la media cuadrática de la magnitud $\xi_1 - \hat{c}_1 - \hat{c}_2 \xi_2$ es

$$M(\xi_1 - \hat{c}_1 - \hat{c}_2 \xi_2)^2 = \sigma_1^2 (1 - r^2) = 0$$

y, por consiguiente, $\xi_1 - \hat{c}_1 - \hat{c}_2 \xi_2 = 0$ con la probabilidad 1.

Si las magnitudes aleatorias ξ_1 y ξ_2 son independientes, su coeficiente de correlación es igual a cero, ya que

$$M(\xi_1 - a_1)(\xi_2 - a_2) = M(\xi_1 - a_1) \cdot M(\xi_2 - a_2) = 0.$$

Debemos señalar, que la condición $r = 0$, hablando en general, no lleva consigo la independencia de las magnitudes aleatorias ξ_1 y ξ_2 .

Sea, por ejemplo, ξ_1 una magnitud distribuida simétricamente, con una densidad $p_1(x)$ tal que $p(-x) = p(x)$, $-\infty < x < \infty$ y sea $\xi_2 = |\xi_1|$. Entonces, a pesar de que la magnitud ξ_2 también es una función de ξ_1 , el coeficiente de correlación de las magnitudes ξ_1 y ξ_2 será igual a cero, ya que

$$a_1 = M\xi_1 = \int_{-\infty}^{\infty} xp_1(x) dx = 0$$

y como se ve fácilmente,

$$M(\xi_1 - a_1)(\xi_2 - a_2) = M(\xi_1 \cdot \xi_2) = \int_{-\infty}^{\infty} x|x|p_1(x) dx = 0.$$

Examinemos la tarea general sobre la aproximación óptima de la magnitud aleatoria ξ_0 con ayuda de combinaciones lineales $\sum_{k=1}^n c_k \xi_k$ de otras magnitudes aleatorias ξ_1, \dots, ξ_n . La tarea consiste en hallar, según las características dadas

$$a_k = M\xi_k, \quad \sigma_k^2 = M(\xi_k - a_k)^2, \\ r_{kj} = \frac{M(\xi_k - a_k)(\xi_j - a_j)}{\sigma_k \sigma_j} \quad (k, j = 0, 1, \dots, n),$$

tales coeficientes $\hat{c}_0, \hat{c}_1, \dots, \hat{c}_n$ para los cuales

$$\|\xi_0 - (\hat{c}_0 + \sum_{k=1}^n \hat{c}_k \xi_k)\| = \min_{c_0, \dots, c_n} \|\xi_0 - (c_0 + \sum_{k=1}^n c_k \xi_k)\|. \quad (4.27)$$

De la relación general (4.22) obtenida anteriormente se ve, que si han sido hallados los coeficientes correspondientes $\hat{c}_1, \dots, \hat{c}_n$,

entonces, la constante \hat{c}_0 es necesario elegirla del siguiente modo:

$$\hat{c}_0 = \mathbf{M} \left(\xi_0 - \sum_{k=1}^n \hat{c}_k \xi_k \right) = a_0 - \sum_{k=1}^n \hat{c}_k a_k.$$

Sin limitar la generalidad, se puede considerar $a_k = 0$, $k = 0, 1, \dots, n$, y entonces también $\hat{c}_0 = 0$.

Así pues, examinemos las magnitudes aleatorias $\xi_0, \xi_1, \dots, \xi_n$ con valores medios nulos y con una matriz de correlación $\{R_{kj}\}$ cuyas componentes son

$$R_{kj} = \mathbf{M} \xi_k \xi_j \quad (R_{kk} = \sigma_k^2; R_{kj} = \sigma_k r_{kj} \sigma_j) \\ (k, j = 0, 1, \dots, n). \quad (4.28)$$

La matriz de correlación $\{R_{kj}\}$ está determinada positivamente: para cualesquiera valores reales de c_0, c_1, \dots, c_n

$$\sum_{k,j=0}^n c_k c_j R_{kj} = \mathbf{M} \left(\sum_{k=0}^n c_k \xi_k \right)^2 \geq 0, \quad (4.29)$$

y con su ayuda, en el espacio lineal H de todas las magnitudes aleatorias de la forma

$$\xi = \sum_{k=0}^n c_k \xi_k \quad (4.30)$$

se puede introducir el producto escalar, determinándolo para cualesquiera $\xi' = \sum_{k=0}^n c'_k \xi_k$ y $\xi'' = \sum_{k=0}^n c''_k \xi_k$ como

$$(\xi', \xi'') = \mathbf{M} \xi' \xi'' = \sum_{k,j=0}^n c'_k c''_j R_{kj}. \quad (4.31)$$

La distancia media cuadrática, introducida anteriormente, es la distancia en el espacio de Euclides H con el producto escalar (4.31), a saber:

$$\|\xi' - \xi''\| = \sqrt{(\xi' - \xi'', \xi' - \xi'')}.$$

Como es conocido, en el espacio de Euclides H se tiene realmente a magnitud

$$\hat{\xi}_0 = \sum_{k=1}^n \hat{c}_k \xi_k,$$

que satisface la condición (4.27) planteada anteriormente,

$$\|\xi_0 - \hat{\xi}_0\| = \min_{c_1, \dots, c_n} \left\| \xi_0 - \sum_{k=1}^n c_k \xi_k \right\|;$$

que representa en sí, geoméricamente, una proyección del elemento $\xi_0 \in H$ sobre el subespacio de todos los elementos de la forma

$$\sum_{k=1}^n c_k \xi_k.$$

Esta magnitud ξ_0 puede ser hallada de las condiciones de ortogonalidad de la diferencia $\xi_0 - \xi_0$ respecto a todas las magnitudes ξ_k , $k = 1, \dots, n$, lo que da el siguiente sistema de ecuaciones lineales para hallar los coeficientes \hat{c}_k , $k = 1, \dots, n$:

$$\left(\xi_0 - \sum_{k=1}^n \hat{c}_k \xi_k, \xi_j \right) = 0, \quad j = 1, \dots, n,$$

o más detalladamente,

$$\sum_{k=1}^n R_{kj} \hat{c}_k = R_{0j}, \quad j = 1, \dots, n. \quad (4.32)$$

(Ciertamente, desde el mismo comienzo, es natural examinar el caso de las magnitudes linealmente independientes ξ_1, \dots, ξ_n para las cuales el determinante del sistema (4.32) es distinto de cero.)

5. Algunos teoremas sobre la convergencia. Examinemos la sucesión de magnitudes aleatorias ξ_n , $n = 1, 2, \dots$. Se dice que la magnitud aleatoria ξ es el límite en la media cuadrática de la sucesión ξ_n :

$$\text{l.i.m}_{n \rightarrow \infty} \xi_n = \xi,$$

$$\text{si } \|\xi_n - \xi\| = \sqrt{M|\xi_n - \xi|^2} \rightarrow 0 \text{ para } n \rightarrow \infty.$$

Para la sucesión ξ_n , $n = 1, 2, \dots$ convergente en su media cuadrática, según la desigualdad de Chebishev también tenemos la convergencia por la probabilidad

$$P\{|\xi_n - \xi| > \varepsilon\} \leq \frac{1}{\varepsilon^2} \|\xi_n - \xi\|^2 \rightarrow 0 \text{ para } n \rightarrow \infty,$$

es decir, por pequeños que fuesen $\varepsilon > 0$ y $\delta > 0$, para los valores de n suficientemente grandes, la probabilidad de cada uno de los sucesos separados $\{|\xi_n - \xi| > \varepsilon\}$ será menor que δ . De tal modo, si despreciamos el suceso poco probable $A_n = \{|\xi_n - \xi| > \varepsilon\}$, al examinar la magnitud aleatoria ξ_n , entonces tendremos que $\xi_n \approx \xi$ con una exactitud hasta de ε .

Recordemos que examinamos las magnitudes aleatorias $\xi_n = \xi_n(\omega)$, $n = 1, 2, \dots$, dependientes del resultado elemental $\omega \in \Omega$. Para tal o cual resultado elemental ω tenemos la sucesión correspondiente de números $\xi_n(\omega)$, $n = 1, 2, \dots$, llamada *sucesión seleccionada*.

Fijemos un determinado ε . Como ya fue señalado, para la sucesión ξ_n , $n = 1, 2, \dots$, convergente en la media cuadrática, cada suceso por separado $A_n \subseteq \Omega$, para el cual $|\xi_n(\omega) - \xi(\omega)| > \varepsilon$, es poco probable para los grandes valores de n ($\mathbf{P}(A_n) \rightarrow 0$ para $n \rightarrow \infty$). Sin embargo, puede ocurrir que en la sucesión A_n , $n = 1, 2, \dots$, se realicen un número infinito de sucesos, es decir, en la sucesión seleccionada $\xi_n(\omega)$, $n = 1, 2, \dots$, habrá un número infinito de miembros (designémoslos por $\xi_{n_k}(\omega)$) tales que $|\xi_{n_k}(\omega) - \xi(\omega)| > \varepsilon$ y, por consiguiente, tal sucesión $\xi_n(\omega)$, $n = 1, 2, \dots$, en general no converge hacia el valor $\xi(\omega)$ de la magnitud límite ξ . Más adelante, se da un ejemplo sencillo de una sucesión de magnitudes aleatorias ξ_n , $n = 1, 2, \dots$, que converge en la media cuadrática, para las cuales la sucesión seleccionada $\xi_n(\omega)$, $n = 1, 2, \dots$, no converge para ninguno de los resultados elementales ω .

Ejemplo. Supongamos que el espacio de los resultados elementales ω es el segmento $0 \leq \omega \leq 1$ con una distribución uniforme de probabilidades y la sucesión ξ_n , $n = 1, 2, \dots$, está compuesta de magnitudes aleatorias $\xi_{mk} = \xi_{mk}(\omega)$ de la forma

$$\xi_{mk}(\omega) = \begin{cases} 1, & \text{si } \frac{k-1}{m} < \omega \leq \frac{k}{m} \\ 0, & \text{para los restantes } \omega, \end{cases}$$

donde $k = 1, \dots, m$, $m = 1, 2, \dots$. La sucesión de magnitudes aleatorias ξ_n , $n = 1, 2, \dots$, converge hacia cero en la media cuadrática, ya que $M\xi_{mk}^2 = \frac{1}{m} \rightarrow 0$ para $m \rightarrow \infty$. Al mismo tiempo, para cualquier resultado elemental ω se tiene un número infinito de valores $\xi_{mk}(\omega)$ con los números m, k : $\frac{k-1}{m} < \omega \leq \frac{k}{m}$ tales que $\xi_{mk} = 1$. Se ve que para ninguno de los resultados elementales ω , la sucesión seleccionada $\xi_n(\omega)$, $n = 1, 2, \dots$, no converge hacia cero.

Se dice, que la sucesión de magnitudes aleatorias ξ_n , $n = 1, 2, \dots$, converge hacia la magnitud aleatoria ξ con una probabilidad 1, si para la sucesión seleccionada $\xi_n(\omega)$, $n = 1, 2, \dots$,

$$\lim_{n \rightarrow \infty} \xi_n(\omega) = \xi(\omega) \quad (4.33)$$

y para cualquier resultado elemental ω , a excepción, puede ser, de algunos resultados ω que tienen en su conjunto la probabilidad 0, es decir, si la relación límite (4.33) es un suceso de probabilidad 1.

La convergencia $\xi_n(\omega) \rightarrow \xi(\omega)$ significa, que $|\xi_n(\omega) - \xi(\omega)| \leq \varepsilon$ con cualquier $\varepsilon > 0$ para todos los n suficientemente grandes, en otras palabras, $|\xi_n(\omega) - \xi(\omega)| > \varepsilon$ solamente para un número finito de miembros de la sucesión $\xi_n(\omega)$, $n = 1, 2, \dots$. De tal modo, si $\xi_n(\omega) \rightarrow \xi(\omega)$ con la probabilidad 1, entonces para cualquier

$\varepsilon > 0$ se realiza con la probabilidad 1, sólo un número finito de sucesos $A_n^\varepsilon = \{|\xi_n(\omega) - \xi(\omega)| > \varepsilon\}$, $n = 1, 2, \dots$

Mostremos que ésta es la condición suficiente para que $\xi_n(\omega) \rightarrow \xi(\omega)$ cuando $n \rightarrow \infty$ con la probabilidad 1. Supongamos que $\varepsilon = \frac{1}{k}$ y designemos por B_k el suceso que significa que en la sucesión A_n^ε , $n = 1, 2, \dots$, se realiza sólo un número finito de sucesos $A_n^\varepsilon = \{|\xi_n(\omega) - \xi(\omega)| > \frac{1}{k}\}$, $n = 1, 2, \dots$. Evidentemente, B_k , $k = 1, 2, \dots$, es una sucesión monótona decreciente de sucesos $B_1 \supseteq B_2 \supseteq \dots$, y para el suceso $B = \bigcap_k B_k$, se deduce de la propiedad de continuidad, que $\mathbf{P}(B) = \lim_{k \rightarrow \infty} \mathbf{P}(B_k)$; para $\mathbf{P}(B_k) = 1$ tenemos $\mathbf{P}(B) = 1$; está claro que $\xi_n(\omega) \rightarrow \xi(\omega)$ cuando $n \rightarrow \infty$ para cualquier resultado elemental $\omega \in B$.

Lema (Borel—Cantelli). Sea A_n , $n = 1, 2, \dots$, una sucesión de sucesos arbitrarios A_n de probabilidad $\mathbf{P}(A_n) = p_n$. Si

$$\sum_{n=1}^{\infty} p_n < \infty, \quad (4.34)$$

entonces ocurre, con la probabilidad 1 sólo un número finito de tales sucesos.

Demostración: Supongamos que B significa que ocurre un número finito de sucesos, entre los A_n , $n = 1, 2, \dots$, y B_n significa que ocurre aunque sea sólo un suceso A_k , $k \geq n$ ($B_n = \bigcup_{k \geq n} A_k$). Está claro, que el suceso B aparece cuando, y sólo cuando ocurren todos los sucesos B_n , $n = 1, 2, \dots$ ($B = \bigcap_n B_n$). Ya que $B_1 \supseteq B_2 \supseteq \dots$, entonces $\mathbf{P}(B) = \lim_{n \rightarrow \infty} \mathbf{P}(B_n)$. La probabilidad de la unión de cualesquiera sucesos siempre no supera la suma de sus probabilidades de modo que

$$\mathbf{P}(B_n) \leq \sum_{k=n}^{\infty} \mathbf{P}(A_k) = \sum_{k=n}^{\infty} p_k,$$

donde según la condición (4.34) $\sum_{k=n}^{\infty} p_k \rightarrow 0$ para $n \rightarrow \infty$. Por consiguiente,

$$\mathbf{P}(B) = \lim_{n \rightarrow \infty} \mathbf{P}(B_n) = 0$$

y el suceso \bar{B} complementario del B , que significa que en la sucesión A_1, A_2, \dots ocurre sólo un número finito de sucesos, tiene una probabilidad $\mathbf{P}(\bar{B}) = 1$.

Teorema 1. Si la sucesión de magnitudes aleatorias ξ_n , $n = 1, 2, \dots$, converge en su media cuadrática hacia la magnitud aleatoria ξ .

«bastante rápidamente», a saber, si

$$\sum_{n=1}^{\infty} \|\xi_n - \xi\|^2 < \infty,$$

entonces ella converge también con la probabilidad 1.

Demostración. Para cualquier $\varepsilon > 0$ en el suceso $A_n = \{|\xi_n(\omega) - \xi(\omega)| > \varepsilon\}$, tenemos según la desigualdad de Chebichev que $P(A_n) \leq \|\xi_n - \xi\|^2 / \varepsilon^2$ y, por consiguiente,

$$\sum_{n=1}^{\infty} P(A_n) \leq \frac{1}{\varepsilon^2} \sum_{n=1}^{\infty} \|\xi_n - \xi\|^2 < \infty.$$

Según el lema de Borel—Cantelli ocurre, con la probabilidad 1, solamente un número finito de sucesos A_n , $n = 1, 2, \dots$, y esto significa, como indicamos anteriormente, que $\lim_{n \rightarrow \infty} \xi_n(\omega) = \xi(\omega)$ con la probabilidad 1.

La distancia media cuadrática $\|\xi_1 - \xi_2\| = \sqrt{M|\xi_1 - \xi_2|^2}$, como corresponde a una verdadera distancia entre los puntos del espacio H , (véase el punto 4 anterior), satisface la *desigualdad del triángulo*: para tres puntos cualesquiera $\xi, \eta, \zeta \in H$, la distancia entre ξ, η no supera a la suma de las distancias entre ξ, ζ y η, ζ :

$$\|\xi - \eta\| \leq \|\xi - \zeta\| + \|\eta - \zeta\|. \quad (4.35)$$

Esto se deduce directamente de la desigualdad general (4.14) aplicada a la distancia $\xi_1 = \xi - \zeta$ y $\xi_2 = \eta - \zeta$.

Se cumple la siguiente proposición.

Teorema 2. *La sucesión de magnitudes aleatorias ξ_n , $n = 1, 2, \dots$, de H converge en su media cuadrática hacia una magnitud aleatoria $\xi \in H$ cuando, y sólo cuando*

$$\lim_{n, m \rightarrow \infty} \|\xi_n - \xi_m\| = 0. \quad (4.36)$$

La necesidad de la condición (4.36) se deduce de la desigualdad del triángulo: si la sucesión ξ_n , $n = 1, 2, \dots$, converge en la media cuadrática hacia la magnitud ξ , entonces

$$\|\xi_n - \xi_m\| \leq \|\xi_n - \xi\| + \|\xi_m - \xi\| \rightarrow 0 \text{ para } n, m \rightarrow \infty.$$

La demostración de que para la condición (4.36) existe la magnitud $\xi \in H$ tal que $\xi = \lim_{n \rightarrow \infty} \xi_n$, exige la aplicación de un aparato matemático especial (teoría de la medida), y por eso omitimos la demostración¹⁾.

¹⁾ Véase, por ejemplo, el libro de I. I. Guijman, A. V. Skorojod (pág. 108). Señalemos que el teorema 2, nosotros lo utilizamos, sólo al determinar las llamadas integrales estocásticas (véanse las págs. 204, 225, 234).

Examinemos la sucesión ξ_n , $n = 1, 2, \dots$, que converge en la media cuadrática hacia la magnitud aleatoria ξ . Las magnitudes $\xi_n \in H$ tienen unas esperanzas matemáticas finitas $M\xi_n$, ya que según la desigualdad (4.13) $M|\xi_n| \leq \sqrt{M\xi_n^2} < \infty$. Mostremos que

$$\lim_{n \rightarrow \infty} M\xi_n = M\xi. \quad (4.37)$$

En efecto, como ya se indicó, la magnitud $\xi \in H$ tiene una esperanza matemática finita $M\xi$ y

$$|M\xi_n - M\xi| \leq M|\xi_n - \xi| \leq \sqrt{M|\xi_n - \xi|^2} \rightarrow 0$$

para $n \rightarrow \infty$.

Sea η_n , $n = 1, 2, \dots$, otra sucesión convergente en la media cuadrática (hacia la magnitud aleatoria η).

Teorema 3. *Tiene lugar la siguiente relación límite*

$$\lim_{n \rightarrow \infty} M\xi_n \eta_n = M\xi \eta. \quad (4.38)$$

Demostración. Para cualesquiera magnitudes $\xi, \eta \in H$ existe la esperanza matemática finita $M\xi\eta$, ya que según la igualdad de Cauchy—Bunjakovsky $M|\xi\eta| \leq \sqrt{M|\xi|^2} \cdot \sqrt{M|\eta|^2}$. Luego, los valores de las medias cuadráticas $\|\xi_n\| = \sqrt{M\xi_n^2}$ y $\|\eta_n\| = \sqrt{M\eta_n^2}$ están limitadas por una determinada constante C :

$$\begin{aligned} \|\xi_n\| &\leq \|\xi\| + \|\xi_n - \xi\| \leq C \\ \|\eta_n\| &\leq \|\eta\| + \|\eta_n - \eta\| \leq C, \end{aligned}$$

por cuanto $\|\xi_n - \xi\| \rightarrow 0$, $\|\eta_n - \eta\| \rightarrow 0$ para $n \rightarrow \infty$. Utilizando de nuevo la desigualdad de Cauchy—Bunjakovsky obtenemos

$$\begin{aligned} |M\xi_n \eta_n - M\xi \eta| &\leq |M\xi_n \eta_n - M\xi_n \eta| + |M\xi_n \eta - M\xi \eta| \leq \\ &\leq M(|\xi_n| \cdot |\eta_n - \eta|) + M(|\eta| \cdot |\xi_n - \xi|) \leq \\ &\leq \|\xi_n\| \cdot \|\eta_n - \eta\| + \|\eta\| \cdot \|\xi_n - \xi\| \leq \\ &\leq C(\|\xi_n - \xi\| + \|\eta_n - \eta\|) \rightarrow \end{aligned}$$

para $n \rightarrow \infty$ lo que se exigía demostrar.

Exponemos sin demostración¹⁾ dos teoremas más sobre la convergencia de las esperanzas matemáticas.

Sea $\xi_1 \leq \xi_2 \leq \dots$ una sucesión monótona creciente de magnitudes aleatorias no negativas. Para cada resultado elemental ω

¹⁾ Véase el libro citado anteriormente de I. I. Guijman, A. V. Skorojod (pág. 102). El teorema 4 siguiente (más exacto, su corolario (4.39), sólo se utiliza para la demostración en las págs. 191 y 234 y el teorema 5 en las págs. 151 y 166.

la sucesión seleccionada monótona creciente $\xi_n(\omega)$, $n = 1, 2, \dots$, tiene un límite finito o infinito $\xi(\omega) \geq 0$. Supongamos que las magnitudes ξ_n tienen una esperanza matemática finita $M\xi_n$, $n = 1, 2, \dots$. Para esta sucesión monótona creciente existe un límite finito o infinito $\lim_{n \rightarrow \infty} M\xi_n$.

Teorema 4. Si la sucesión $M\xi_n$, $n = 1, 2, \dots$, es limitada, entonces la magnitud aleatoria límite $\xi = \lim_{n \rightarrow \infty} \xi_n$ es finita con una probabilidad 1 y

$$\lim_{n \rightarrow \infty} M\xi_n = M\xi.$$

Como consecuencia de este teorema obtenemos el siguiente resultado.

Supongamos que la sucesión de magnitudes aleatorias no negativas ξ_k , $k = 1, 2, \dots$, sea tal que la serie de sus esperanzas matemáticas $\sum_{k=1}^{\infty} M\xi_k$ es convergente. Entonces con la probabilidad 1

$$\sum_{k=1}^{\infty} \xi_k < \infty$$

y

$$M\left(\sum_{k=1}^{\infty} \xi_k\right) = \sum_{k=1}^{\infty} M\xi_k. \quad (4.39)$$

En efecto, las sumas parciales $\sum_{k=1}^n \xi_k$, $n = 1, 2, \dots$, forman una sucesión monótona creciente, para la cual las esperanzas matemáticas

$M\left(\sum_{k=1}^n \xi_k\right) = \sum_{k=1}^n M\xi_k$, $n = 1, 2, \dots$, están limitadas: $M\left(\sum_{k=1}^n \xi_k\right) \leq \sum_{k=1}^{\infty} M\xi_k < \infty$, y según el teorema 4 existe, con la probabilidad 1,

el límite finito $\lim_{n \rightarrow \infty} \sum_{k=1}^n \xi_k = \sum_{k=1}^{\infty} \xi_k$ siendo

$$M\left[\lim_{n \rightarrow \infty} \sum_{k=1}^n \xi_k\right] = \lim_{n \rightarrow \infty} M\left[\sum_{k=1}^n \xi_k\right] = \lim_{n \rightarrow \infty} \sum_{k=1}^n M\xi_k.$$

Teorema 5. Supongamos que la sucesión de magnitudes aleatorias ξ_n , $n = 1, 2, \dots$, converge, con la probabilidad 1, hacia una mag-

nitid ξ , siendo $|\xi_n| \leq \eta$, donde la magnitud aleatoria η tiene una esperanza matemática finita ($M|\eta| < \infty$). Entonces las magnitudes ξ_n y ξ también tienen esperanzas matemáticas, siendo

$$\lim_{n \rightarrow \infty} M\xi_n = M\xi.$$

§ 5. SERIES INFINITAS
DE PRUEBAS INDEPENDIENTES
Y LEY DE LOS
GRANDES NUMEROS

1. Ley de los grandes números. Examinemos la sucesión de magnitudes aleatorias independientes

$$\xi_1, \xi_2, \dots, \quad (5.1)$$

que tienen una misma distribución de probabilidades (para claridad nos podemos imaginar que se ejecuta una serie de pruebas iguales independientes $\Omega_1, \Omega_2, \dots$, y en cada prueba Ω_n se observa la magnitud correspondiente ξ_n , $n = 1, 2, \dots$).

Supongamos que existe la esperanza matemática y la dispersión:

$$a = M\xi_n, \quad \sigma^2 = D\xi_n, \quad n = 1, 2, \dots,$$

y examinemos el valor medio «empírico»

$$\frac{1}{n} \sum_{h=1}^n \xi_h = \frac{S_n}{n} \quad (5.2)$$

(que el «observador» puede calcular después de realizarse las pruebas $\Omega_1, \dots, \Omega_n$). Para la suma $S_n = \sum_{h=1}^n \xi_h$ tenemos

$$M \frac{S_n}{n} = \frac{1}{n} \sum_{h=1}^n M\xi_h = a,$$

$$D \frac{S_n}{n} = M \left(\frac{S_n}{n} - a \right)^2 = \frac{1}{n^2} = \sum_{h=1}^n D\xi_h = \frac{\sigma^2}{n}.$$

Vemos que

$$\left\| \frac{1}{n} \sum_{h=1}^n \xi_h - a \right\| = \frac{\sigma}{\sqrt{n}} \rightarrow 0 \text{ para } n \rightarrow \infty, \quad (5.3)$$

es decir, la media empírica $\frac{1}{n} \sum_{h=1}^n \xi_h$ converge en la media cuadrática hacia la esperanza matemática a :

$$\text{l.í.m.}_{n \rightarrow \infty} \frac{1}{n} \sum_{h=1}^n \xi_h = a.$$

De tal modo, para cualquier $\varepsilon > 0$ la probabilidad de que la media empírica $\frac{1}{n} \sum_{h=1}^n \xi_h$, para los valores n suficientemente grandes, se diferencie de la esperanza matemática a en no más de ε , es casi igual a 1, más exacto,

$$P \left\{ \left| \frac{1}{n} \sum_{h=1}^n \xi_h - a \right| \leq \varepsilon \right\} \geq 1 - \delta$$

siendo $\delta > 0$ tan pequeño como se quiera, para los valores n suficientemente grandes; con ello la desigualdad de Chebishev da la siguiente valuación de la probabilidad indicada:

$$P \left\{ \left| \frac{1}{n} \sum_{h=1}^n \xi_h - a \right| \leq \varepsilon \right\} \geq 1 - \frac{\sigma^2}{n\varepsilon^2}. \quad (5.4)$$

En la vida real siempre despreciamos la posibilidad de que aparezcan los sucesos, cuya probabilidad es suficientemente pequeña. Si despreciamos el suceso poco probable $\left| \frac{1}{n} \sum_{h=1}^n \xi_h - a \right| > \varepsilon$ (la probabilidad no es mayor de un δ suficientemente pequeño), entonces se puede considerar, que para cualquier n fijado suficientemente grande digamos $n > \frac{\sigma^2}{\delta\varepsilon^2}$

$$a - \varepsilon \leq \frac{1}{n} \sum_{h=1}^n \xi_h \leq a + \varepsilon,$$

de modo que, con una exactitud hasta de ε , la media empírica $\frac{1}{n} \sum_{h=1}^n \xi_h$ coincide con el valor medio de a :

$$\frac{1}{n} \sum_{h=1}^n \xi_h \approx a. \quad (5.5)$$

Esta ley interesante (expresada por las relaciones (5.3) y (5.4) lleva el nombre de *ley de los grandes números*.

Propiamente hablando, la ley indicada de los grandes números la hemos deducido, no sólo para las magnitudes aleatorias independientes, sino también para las magnitudes aleatorias no correlativas ξ_k , $k = 1, 2, \dots$ ($M\xi_k = a$, $D\xi_k = \sigma^2$). Para tales magnitudes también se puede obtener el siguiente resultado:

Teorema (ley reforzada de los grandes números). *Con la probabilidad 1*

$$\frac{1}{n} \sum_{k=1}^n \xi_k \rightarrow a \text{ para } n \rightarrow \infty. \quad (5.6)$$

D e m o s t r a c i ó n¹⁾. Sin limitar la generalidad se puede considerar $a = 0$. Según la desigualdad de Chebishev para cualquier $\varepsilon > 0$

$$\sum_{m=1}^{\infty} \mathbf{P} \left\{ \left| \frac{S_{m^2}}{m^2} \right| > \varepsilon \right\} \leq \frac{m^2 \sigma^2}{m^4 \varepsilon^2} = \frac{\sigma^2}{\varepsilon^2} \cdot \frac{1}{m^2},$$

de modo que

$$\sum_{m=1}^{\infty} \mathbf{P} \left\{ \left| \frac{S_{m^2}}{m^2} \right| > \varepsilon \right\} \leq \frac{\sigma^2}{\varepsilon^2} \sum_{m=1}^{\infty} \frac{1}{m^2} < \infty;$$

según el lema de Borel—Cantelli, con la probabilidad 1 puede ocurrir un solo número finito de sucesos $A_m = \left\{ \left| \frac{S_{m^2}}{m^2} \right| > \varepsilon \right\}$, $m = 1, 2, \dots$ y de tal modo $\frac{S_{m^2}}{m^2} \rightarrow 0$ con la probabilidad 1 (para $m \rightarrow \infty$). Después, sea

$$\eta_m = \max_{m^2+1 \leq k \leq (m+1)^2} |\xi_{m^2} + \dots + \xi_k|.$$

Evidentemente,

$$\begin{aligned} \mathbf{P} \left\{ \left| \frac{\eta_m}{m^2} \right| > \varepsilon \right\} &\leq \sum_{k=m^2+1}^{(m+1)^2} \mathbf{P} \left\{ \left| \frac{\xi_{m^2} + \dots + \xi_k}{m^2} \right| > \varepsilon \right\} \leq \\ &\leq \sum_{k=m^2+1}^{(m+1)^2} \frac{(k-m^2) \sigma^2}{m^4 \varepsilon^2} \leq 2m \cdot \frac{2m \sigma^2}{m^4 \varepsilon^2} = \frac{4\sigma^2}{\varepsilon^2} \cdot \frac{1}{m^2} \end{aligned}$$

y

$$\sum_{m=1}^{\infty} \mathbf{P} \left\{ \left| \frac{\eta_m}{m^2} \right| > \varepsilon \right\} \leq \frac{4\sigma^2}{\varepsilon^2} \sum_{m=1}^{\infty} \frac{1}{m^2} < \infty;$$

¹⁾ La demostración expuesta aquí ha sido comunicada al autor por Yu. V. Prójorov.

por consiguiente, con la probabilidad 1, puede ocurrir, sólo un número finito de sucesos $B_m = \left\{ \left| \frac{\eta_m}{m^2} \right| > \varepsilon \right\}$, $m = 1, 2, \dots$, y de tal modo, $\frac{\eta_m}{m^2} \rightarrow 0$ con la probabilidad 1 (para $m \rightarrow \infty$). Como resumen, para cualquier n , $m^2 \leq n \leq (m+1)^2$, tenemos

$$\left| \frac{S_n}{n} \right| \leq \left| \frac{S_{m^2}}{m^2} \right| + \frac{\eta_m}{m^2} \rightarrow 0$$

con la probabilidad 1 para $n \rightarrow \infty$, lo que se exigía demostrar.

También tiene lugar el siguiente resultado fundamental¹⁾.

Ley reforzada de los grandes números. Para la sucesión de magnitudes aleatorias independientes, igualmente distribuidas, ξ_1, ξ_2, \dots , que tienen una esperanza matemática finita a , con la probabilidad 1

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \xi_k = a. \quad (5.7)$$

2. Probabilidad y frecuencia. Dirijámonos a un determinado suceso A , ligado con una prueba determinada. Supongamos que se tiene la posibilidad de reproducir repetidas veces esta prueba, de modo que las pruebas Ω_1, Ω_2 son independientes.

Introduzcamos las magnitudes aleatorias ξ_1, ξ_2, \dots donde cada magnitud ξ_k está solamente ligada con su prueba correspondiente Ω_k definida del siguiente modo: $\xi_k = 1$, si en la k -ésima prueba ocurre el suceso A , y $\xi_k = 0$ en caso contrario. Ahora tenemos

$$M\xi_k = p, \quad D\xi_k = p(1-p),$$

donde $p = P(A)$ es la probabilidad del suceso A en cada prueba por separado.

La media empírica

$$\frac{1}{n} \sum_{k=1}^n \xi_k = \frac{n}{A}$$

representa la frecuencia del suceso A en las pruebas examinadas, que se define como la relación entre el número $n(A)$ de aquellas pruebas, en las cuales ocurrió este suceso y el número total n de pruebas. Según la ley de los grandes números

$$\frac{n(A)_n}{n} \rightarrow P(A) \text{ para } n \rightarrow \infty \quad (5.8)$$

¹⁾ Véase, por ejemplo, el libro de V. Feller «Introducción a la teoría de probabilidades y sus aplicaciones», tomo I, II, Moscú 1964 (pág. 266). Aquí se tiene en cuenta que las magnitudes ξ_1, ξ_2, \dots , no tienen limitada obligatoriamente la dispersión σ^2 .

de manera que en una gran serie de pruebas independientes, la frecuencia $\frac{n(A)}{n}$ del suceso A coincide, prácticamente, con la probabilidad de este suceso:

$$\frac{n(A)}{n} \approx P(A). \quad (5.9)$$

Esta ley (que recibió su afirmación en los múltiples experimentos de la práctica) nos da una base para valorar experimentalmente, en las aplicaciones la probabilidad conocida $P(A)$ de tal o cual suceso A , como $P(A) \approx \frac{n(A)}{n}$ (véase el ejemplo en la pág. 23).

Más adelante se expone la tabla 1 (tomada del libro, citado anteriormente, de W. Feller, véase la pág. 33) en donde se dan los resultados de 100 series con 100 pruebas, en cada una de las cuales se lanzó una moneda simétrica. De esta tabla se ve que la frecuencia $\frac{n(A)}{n}$ de la caída con «cruz» (suceso A) en cada serie coincide con gran exactitud con la probabilidad $P(A) = 1/2$.

Tabla 1

Número de pruebas	Número de cruces										Número total de cruces
0—1000	54	46	53	55	46	54	41	48	51	53	501
2000	48	46	40	53	49	49	48	54	53	45	485
3000	43	52	58	51	51	50	52	50	53	49	509
4000	58	60	54	55	50	48	47	57	52	55	536
5000	48	51	51	49	44	52	50	46	53	41	485
6000	49	50	45	52	52	48	47	47	47	51	488
7000	45	47	41	51	49	59	50	55	53	50	500
8000	53	52	46	52	44	51	48	51	46	54	497
9000	45	47	46	52	47	48	59	57	45	48	494
10 000	47	41	51	48	59	51	52	55	39	41	484

CAPITULO II

ALGUNAS DISTRIBUCIONES DE PROBABILIDADES

§ 1. ELECCION ALEATORIA Y VARIACION ALEATORIA

1. Algunas fórmulas combinatorias. Al calcular las probabilidades aportan una gran ayuda las fórmulas combinatorias. Expongamos las más importantes de ellas.

Combinaciones de elementos elegidos de diferentes grupos. Supongamos que se tiene r grupos diferentes, compuestos por determinados elementos. El primer grupo contiene n_1 elementos a_1, \dots, a_{n_1} ; el segundo contiene n_2 elementos b_1, b_2, \dots, b_{n_2} ; ... el último grupo r contiene n_r elementos c_1, c_2, \dots, c_{n_r} . Se componen todas las combinaciones posibles de r elementos pertenecientes a los diferentes grupos de tal modo que en una combinación separada entra sólo un elemento de cada grupo. Ellas tienen la forma

$$(a, b, \dots, c).$$

Las combinaciones (a, b, \dots, c) y $(\tilde{a}, \tilde{b}, \dots, \tilde{c})$ se consideran distintas, si se tiene aunque un solo par de elementos distintos entre sí a y \tilde{a} , b y \tilde{b} , ... , c y \tilde{c} . El número de tales combinaciones es

$$N = n_1 \cdot n_2 \cdot \dots \cdot n_r \quad (1.1)$$

(para $r = 2$ la fórmula (1.1) es evidente, y para un número arbitrario r se establece fácilmente por inducción).

Elección con retorno. Supongamos que se tiene n objetos diferentes a_1, a_2, \dots, a_n . Elijamos sucesivamente r objetos de este conjunto de tal modo que cada objeto elegido se fija y se retorna atrás. Resultado de tal elección es la combinación de la forma

$$(a_{i_1}, a_{i_2}, \dots, a_{i_r}).$$

Las combinaciones $(a_{i_1}, a_{i_2}, \dots, a_{i_r})$ y $(a_{j_1}, a_{j_2}, \dots, a_{j_r})$ se consideran distintas, si en cualquier paso fueron elegidos distintos objetos, es decir $a_{i_k} \neq a_{j_k}$, aunque sea para un solo k . Nos podemos imaginar que se tienen r grupos iguales de n elementos, y en el paso k se elige

el elemento del k -ésimo grupo. En este caso particular, la fórmula (1.1) da la siguiente expresión para el número de combinaciones diferentes N :

$$N = n^r. \quad (1.2)$$

Número de variaciones. Elección sin retorno. Supongamos que r objetos diferentes se colocan por determinadas celdas. Con ello las celdas son distinguibles por el observador y su número es igual a n . Supongamos que los objetos se colocan de tal modo que en cada celda cae no más de un objeto. Numeremos todos los objetos que se tienen y las celdas. Entonces cada variación la podemos describir por la combinación (i_1, i_2, \dots, i_r) , donde i_1 es el número de la celda en donde cae el 1º objeto, i_2 es el número de la celda donde cae el 2º objeto, etc., i_r es el número de la celda donde cae el último objeto r . En la combinación (i_1, i_2, \dots, i_r) el elemento i_1 puede ser cualquiera de los n elementos (celdas), i_2 es uno de los $n - 1$ elementos restantes, etc., i_r puede ser uno de los $n - r + 1$ elementos posibles. De acuerdo con la fórmula general (1.1) en total existen

$$N = n(n - 1) \dots (n - r + 1) \quad (1.3)$$

distintas variaciones de r elementos por n celdas.

Supongamos, que se tiene el conjunto de n elementos diferentes a_1, a_2, \dots, a_n . De este conjunto se eligen r elementos. Se puede considerar que ellos se eligen a la vez y se colocan en un orden determinado, o se puede considerar que los elementos se eligen sucesivamente uno después de otro, con ello el elemento elegido no se retorna atrás. Como resultado de tal elección se forma la combinación $(a_{i_1}, a_{i_2}, \dots, a_{i_r})$, donde a_{i_1} es el primero de los elementos elegidos, a_{i_2} es el segundo de los elementos elegidos, etc., a_{i_r} es el último de los elementos elegidos; a_{i_1} puede ser cualquiera de los n elementos que se tienen, a_{i_2} , cualquiera de los $n - 1$ elementos restantes, etc., a_{i_r} , cualquiera de los $n - r + 1$ elementos posibles. En total existen

$$N = n(n - 1) \dots (n - r + 1)$$

distintas elecciones $(a_{i_1}, a_{i_2}, \dots, a_{i_r})$ de r objetos, elegidos, sin retorno, del conjunto de volumen n .

Para $r = n$, el número de variaciones coincide con el número de permutaciones de n elementos:

$$N = n(n - 1) \dots 1 = n! \quad (1.4)$$

Número de combinaciones. Supongamos que se colocan r objetos no distinguibles entre sí por n celdas (en cada celda cae no más de un objeto). Entonces el número de variaciones distinguibles coincide con el número de grupos distintos por r celdas, que se puede formar del conjunto general de volumen n (en un grupo se agrupan todas las celdas ocupadas), y es igual al número de combinaciones C_n^r

de n elementos por r :

$$N = C_n^r = \frac{n!}{r!(n-r)!} \quad (1.5)$$

Más general es el caso, cuando el conjunto de n elementos se divide en un número fijo k de grupos, siendo también fijo el número de elementos en cada grupo, digamos, en el grupo con el número i se tienen, justamente, n_i elementos. En lo restante, los elementos se agrupan lo más arbitrariamente. El número de combinaciones diferentes de n elementos por n_1, n_2, \dots, n_k elementos ($n_1 + n_2 + \dots + n_k = n$) es

$$N = \frac{n!}{n_1! n_2! \dots n_k!} \quad (1.6)$$

La fórmula (1.5) puede ser obtenida de la siguiente forma. Supongamos temporalmente, que todos los objetos examinados son diferentes; entonces el número de variaciones diferentes (véase (1.3)) será igual a $n(n-1) \dots (n-r+1)$. Para cualquier grupo i_1, \dots, i_r de celdas ocupadas, entre aquel número figuran también las variaciones que se obtienen permutando r objetos por las celdas indicadas; el total de tales permutaciones será $r!$. Por consiguiente, el número de combinaciones C_n^r de n celdas por r^1 , que nos interesa, es $r!$ veces menor que el número indicado de variaciones $n(n-1) \dots (n-r+1)$, a saber:

$$C_n^r = \frac{n(n-1) \dots (n-r+1)}{r!} = \frac{n!}{r!(n-r)!}$$

Para la demostración de la fórmula (1.6) la volvemos a escribir en la forma:

$$N = C_n^{n_1} C_{n-n_1}^{n_2} \dots C_{n-n_1-n_2-\dots-n_{k-2}}^{n_{k-1}}$$

La formación de grupos de n_1, n_2, \dots, n_k elementos se puede realizar consecutivamente. Primeramente, formemos el grupo de n_1 elementos (esto se puede hacer de $N_1 = C_n^{n_1}$ modos diferentes); luego de los $n - n_1$ elementos restantes formamos el grupo de n_2 elementos (esto se puede hacer de $N_2 = C_{n-n_1}^{n_2}$ modos diferentes), etc.; los elementos restantes al final $n - n_1 - \dots - n_{k-2} = n_{k-1} + n_k$ se pueden dividir en dos grupos de n_{k-1} y n_k elementos, de $N_{k-1} = C_{n-n_1-\dots-n_{k-2}}^{n_{k-1}}$ modos diferentes. Combinando a cada paso los modos de formación de grupos, según la fórmula general (1.1), en total obtenemos el número indicado $N = N_1 N_2 \dots N_{k-1}$ de todas las combinaciones posibles.

¹⁾ El término C_n^r se suele leer así: «el número de combinaciones de n elementos (celdas) tomados de r en r , (N , del T .)»

En todas las igualdades expuestas anteriormente, se encuentra la expresión $n! = 1 \cdot 2 \cdot \dots \cdot (n-1) n$, para la cual se tiene la siguiente fórmula de Stirling¹⁾:

$$n! \sim \sqrt{2\pi n} n^n e^{-n}, \quad (1.7)$$

donde $e = 2,718\dots$ es la base de los logaritmos naturales (aquí y en adelante la relación $\alpha_n \sim \beta_n$ entre las magnitudes variables α_n y β_n significa que $\frac{\alpha_n}{\beta_n} \rightarrow 1$ para $n \rightarrow \infty$). Más exacta que (1.7) es la relación

$$e^{\frac{1}{12n+1}} \leq \frac{n!}{\sqrt{2\pi n} n^n e^{-n}} \leq e^{\frac{1}{12n}},$$

justa para todos los $n = 1, 2, \dots$.

Ejemplo. Un grupo de pasajeros de r personas se sientan en el tren eléctrico suburbano, compuesto de $n \geq r$ vagones. Supongamos que cada pasajero elige su vagón casualmente y se encuentra con una misma probabilidad en cualquiera de los vagones. ¿Cuál será la probabilidad de que todos ellos entren en un mismo vagón? ¿Qué probabilidad de que entren en diferentes vagones?

Cada pasajero puede elegir uno de los n vagones, de modo que todas las posibles combinaciones (i_1, i_2, \dots, i_r) , donde i_k es el número del vagón, elegido por el k -ésimo pasajero, alcanzan el valor $N = n^r$.

El suceso A , o sea «todos los r pasajeros entran en un vagón» ocurre en cualquiera de los $N(A) = n$ resultados, en los cuales $i_1 = i_2 = \dots = i_r$, y su probabilidad es

$$P(A) = \frac{N(A)}{N} = \frac{1}{n^{r-1}}.$$

El suceso B , o sea «en ninguno de los vagones entra más de uno de los pasajeros examinados», aparece cuando, y sólo cuando todos los elementos i_k , $k = 1, \dots, r$, son distintos entre sí. Se puede considerar que se realiza la elección (s i n r e t o r n o) de r elementos del conjunto de volumen n : el primer pasajero elige cualquiera de los n vagones, el segundo, cualquiera de los $(n-1)$ vagones restantes, etc. Por consiguiente, el número de combinaciones (i_1, i_2, \dots, i_r) que conducen a la aparición del suceso B , es igual a $n(n-1) \dots \dots (n-r+1)$, y la probabilidad de este suceso es

$$P(B) = \frac{n(n-1) \dots (n-r+1)}{n^r}.$$

Ejemplo. Imaginémonos que se tienen n llaves, una de las cuales entra en la cerradura. Las llaves se prueban sucesivamente. ¿Cuál será la probabilidad de que la llave útil caiga en el paso k ?

¹⁾ Véase, por ejemplo, el libro de V. Feller citado anteriormente.

Consideramos que las llaves están dispuestas en orden aleatorio (i_1, i_2, \dots, i_n) siendo todas estas disposiciones equiprobables. Supongamos que la llave útil tiene el número i . Entonces la probabilidad que nos interesa será igual a la probabilidad de que, en la disposición aleatoria de los números $1, \dots, n$, esté situado el número i en el lugar k . El número de todas las variaciones posibles (i_1, \dots, i_n) que se tienen es igual a $n!$, y el número de variaciones con el elemento fijado $i_k = i$ es igual a $(n-1)!$, de modo que la probabilidad buscada es

$$\frac{(n-1)!}{n!} = \frac{1}{n}.$$

Ejemplo. Examinemos el juego de «préférence»¹⁾, cuando los 32 naipes mayores de la baraja de naipes están distribuidos de un modo aleatorio entre los tres jugadores, que recibieron 10 naipes, y «la toma» de dos cartas, que se le concede a uno de los jugadores que luego anuncia la categoría del juego. Supongamos que junto con «la toma» el jugador tiene 5 naipes mayores de un mismo palo (digamos copas), excluyendo «la dama». Al anunciar la categoría de juego hay que tener en cuenta la posibilidad de que alguno de los compañeros tenga los tres naipes restantes del palo de copas, incluyendo «la dama» (más abreviado, que tenga la «tercera dama»). ¿Cuál es la probabilidad de este suceso?

En total existen $C_{20}^{10} = \frac{20!}{10! 10!}$ distribuciones equiprobables de 20 naipes en dos grupos de 10 naipes, que se entregan a cada dos compañeros jugadores. Si fijamos la combinación de «la tercera dama» en cualquier jugador determinado, entonces el número de distribuciones de naipes, compatibles con ésta, entre dos jugadores es igual al número de combinaciones $C_{17}^7 = \frac{17!}{7! 10!}$ de los 17 naipes restantes en grupos de 7 (o lo que es igual, de 17 tomados de 10 en 10). Por lo tanto, «la tercera dama» la tendrá un determinado jugador con la probabilidad

$$\frac{C_{17}^7}{C_{20}^{10}} = \frac{8 \cdot 9 \cdot 10}{18 \cdot 19 \cdot 20} = \frac{2}{19},$$

y cada uno de los dos jugadores, con la probabilidad doble, igual a $\frac{4}{19}$.

2. Distintas distribuciones de las partículas independientes en el espacio fásico. Examinemos el siguiente modelo sencillo. Imaginémonos el proceso del cocido de la masa para los bollitos. En la masa se echan las uvas pasas, luego todo se mezcla minuciosamente y como resultado de lo cual las uvas pasas que son «partículas» se distribuyen de un modo aleatorio dentro del «espacio fásico» ocupado por la masa.

¹⁾ El juego de «préférence» con naipes franceses es muy parecido al juego español del «tresillo». (N. del T.)

número, evidentemente, es igual al número de combinaciones $C_{n+r-1}^{n-1} = \frac{(n+r-1)!}{(n-1)!r!}$ de $n-1+r$ elementos en grupos de $n-1$ y, por consiguiente, cada resultado por separado (r_1, \dots, r_n) tendrá la probabilidad

$$\frac{(n-1)!r!}{(n+r-1)!}.$$

¿Cuál de las dos soluciones dadas antes es verdadera? Si consideramos que el «mecanismo aleatorio», descrito, anteriormente, (del entremezclado repetido), actúa de tal modo que cada uva pasa, por separado, cae en cualquiera de las n celdas con la misma probabilidad, siendo prácticamente, independiente del comportamiento de las otras uvas pasas, entonces la probabilidad de la distribución (r_1, \dots, r_n) será

$$P(r_1, \dots, r_n) = \frac{r!}{r_1! r_2! \dots r_n!} n^{-r}.$$

Esto se ve especialmente claro en el caso, cuando $n = r = 2$. Cada una de las dos uvas pasas (independiente una de otra) cae en dos cualesquiera celdas con la probabilidad $1/2$ y

$$\begin{aligned} P(2, 0) &= P(0, 2) = 1/4, \\ P(1, 1) &= 1/2, \end{aligned}$$

de modo que las posibles distribuciones $(2, 0)$, $(0, 2)$, $(1, 1)$ no son de ningún modo equiprobables.

En el caso de distribución de las uvas pasas en la masa, el modelo de distribuciones equiprobables (i_1, \dots, i_r) descrito antes, evidentemente, no es satisfactorio, si el volumen de cada celda por separado es menor que el volumen de todas las partículas, o sea de las uvas pasas: (por ejemplo, en una celda no caben todas las partículas). Examinemos un nuevo modelo, en el cual las celdas elegidas son tan pequeñas (comparables con las medidas de las mismas uvas pasas) que en cada celda no puede caer más de una partícula.

Para la limitación indicada, cada distribución de las partículas puede ser descrita por la combinación (i_1, \dots, i_r) indicando las celdas ocupadas i_1, \dots, i_r (se considera naturalmente, que $r \leq n$) y el número total de distribuciones distintas es igual al número de combinaciones C_n^r con n elementos en grupos de r . Si consideramos que todas las distribuciones posibles son equiprobables (así está planteada la cuestión en nuestro modelo con las uvas pasas), entonces la probabilidad de cada distribución por separado (i_1, \dots, i_r) es

$$\frac{(n-r)!r!}{n!}.$$

La cuestión sobre la distribución de las partículas independientes en el espacio físico (uvas pasas en la masa) puede ser planteada del

siguiente modo. Se elige una zona determinada del espacio fásico de volumen v . ¿Cuál será la probabilidad de que caiga en esta zona tal o cual número de partículas?

Dividamos imaginariamente todo el espacio fásico en celdas tan pequeñas que en cada una de ellas puede caer no más de una partícula. Consideremos igual que anteriormente, que todas las posibles distribuciones de partículas (i_1, i_2, \dots, i_r) , donde i_r es el número de la celda ocupada por la partícula r , son equiprobables. Si consideramos, convencionalmente, las celdas ocupadas y libres como bolas blancas y negras, entonces nos podemos imaginar al espacio fásico como una determinada caja, que contiene r bolas blancas y $n - r$ negras, que se han remezclado repetida y minuciosamente (r es el número de todas las partículas en el espacio fásico, n es el número de todas las celdas, $n \geq r$).

Sea m el número de celdas de la zona designada en el espacio fásico v y ξ es el número de partículas que caen, según el caso, en esta zona (número de celdas ocupadas entre las m indicadas). La probabilidad $P_{\xi}(k)$ de que el número ξ sea igual a k , $0 \leq k \leq m$, es la misma que para el número ξ de bolas blancas, que suelen encontrarse entre las m bolas sacadas al azar durante su elección aleatoria de la caja (urna) con r bolas blancas y $n - r$ bolas negras.

La expresión «sacarla al azar» significa más exactamente, que m bolas cualesquiera se eligen con la misma probabilidad (si se numeran todas las bolas existentes, se eligen con la misma probabilidad cualesquiera bolas j_1, \dots, j_m). El número N de resultados distintos (cada uno de los cuales representa en sí a la elección de tales o cuales bolas j_1, \dots, j_m) es igual al número de combinaciones $C_n^m = \frac{n!}{m!(n-m)!}$. A nosotros nos interesa el suceso A , que consiste en que entre las m bolas sacadas se encuentren justamente k bolas blancas.

Para hallar la probabilidad de este suceso, es necesario calcular el número $N(A)$ de resultados, en los cuales aparece el suceso A , $(P(A) = \frac{N(A)}{N})$. Esto se puede hacer del siguiente modo. Separemos mentalmente los dos conjuntos de bolas: r blancas y $n - r$ negras. Las bolas blancas sacadas serán iguales a k , cuando del primer conjunto se elijan k bolas y del segundo, $m - k$ bolas. Designemos a las bolas blancas por a_1, \dots, a_r y las negras, por b_1, \dots, b_{n-r} ; cada elección de bolas blancas $a = (a_{i_1}, \dots, a_{i_k})$ y de negras $b = (b_{j_1}, \dots, b_{j_{m-k}})$ puede considerarse como un par (a, b) y el número $N(A)$ coincide con el número de pares diferentes que se pueden formar con distintos elementos a y b . El número total de elementos a , o las distintas elecciones en grupos de k , con las r bolas blancas existentes, es igual al número de combinaciones $C_r^k = \frac{r!}{k!(r-k)!}$ y, análogamente, el número total de elementos b es C_{n-r}^{m-k} ; por consiguiente, el

número $N(A)$ de pares diferentes (a, b) es igual al producto $C_r^h \cdot C_{n-r}^{m-h}$. Como resultado obtenemos para la probabilidad buscada $P(A)$ la siguiente expresión:

$$P(A) = \frac{C_r^h C_{n-r}^{m-h}}{C_n^m}.$$

De hecho, hemos encontrado la distribución de las probabilidades de la magnitud aleatoria ξ , que es igual al número de bolas blancas elegidas al azar de la partida de volumen m (al número de partículas en la zona designada separada del espacio fásico):

$$P_\xi(k) = \frac{C_r^k C_{n-r}^{m-k}}{C_n^m}, \quad k = 0, \dots, m \quad (1.8)$$

(los parámetros n, r , y m están fijados). Esta es la llamada *distribución hipergeométrica de las probabilidades*.

Para subrayar la importancia de esta distribución examinemos solamente un ejemplo.

Ejemplo (control por elección de la producción). Imaginémosnos que se tiene una gran partida de piezas, que puede ser admitida o rechazada. Está claro, que no siempre es posible la comprobación de cada una de las piezas (por ejemplo, puede ocurrir que la comprobación haga a la pieza inservible para utilización futura). En estos casos se puede emplear el método de control por elección de la producción. Digamos, el controlador elige, casualmente, de toda la partida m piezas, las comprueba y si el número de las rechazadas entre ellas supera a un número crítico m^* entonces se rechaza toda la partida. ¿Qué ocurre durante tal control, qué partidas y con qué probabilidad son rechazadas? Si en toda la partida hay n piezas (siendo rechazadas en ella justamente r), entonces la probabilidad de que entre las m piezas elegidas al azar resulten ser rechazadas justamente k , se da evidentemente, por la fórmula (1.8) en la que ξ significa el número aleatorio de piezas rechazadas en la partida elegida para el control.

En la tarea inicial sobre la distribución de las partículas en el espacio fásico, se consideran parámetros al número r de partículas existentes y el volumen v de la zona considerada del espacio fásico (del volumen total V). Es incluso más cómodo considerar dados no los números r y V , sino el número medio λ de partículas en la unidad de volumen ($\lambda = r/V$), lo que da la posibilidad, como veremos más adelante, de hallar la distribución de las partículas también en el espacio fásico infinito (más exacto, para $V \rightarrow \infty, r \rightarrow \infty$). Desde este punto de vista la fórmula (1.8) sufre el inconveniente de que en ella figuran «los parámetros sobrantes» n y m . Más adelante se mostrará de hecho (véase también el siguiente § 2) que en las condiciones, cuando la zona considerada del espacio fásico es suficientemente

grande, en relación con la medida de cada una de las partículas, el número ξ de partículas en tal zona tiene, aproximadamente, la *distribución de probabilidades de Poisson*:

$$P_{\xi}(k) \approx \frac{(\lambda v)^k}{k!} e^{-\lambda v},$$

más exacto, para $n, m, r \rightarrow \infty$ y con la constante

$$a = \lambda v = r \frac{v}{V} = r \frac{m}{n}$$

(a es el número medio de partículas en la zona de volumen v), se tiene la siguiente expresión límite para las probabilidades determinadas por la fórmulas (1.8) $P_{\xi}(k)$ (véase más adelante la fórmula (1.13)):

$$\lim P_{\xi}(k) = \frac{a^k}{k!} e^{-a}, \quad k = 0, 1, \dots \quad (1.9)$$

Volvamos de nuevo a la caja (o como dicen corrientemente a «la urna») con r bolas blancas y $n - r$ negras. En «el esquema de la urna» utilizado por nosotros se eligen m bolas sin retorno. Examinemos un esquema análogo, pero con aquel cambio, que durante la extracción sucesiva de m bolas, la bola sacada cada vez se retorna de nuevo a la urna (elección con retorno). ¿Cuál será la probabilidad de que entre las m bolas sacadas haya justamente k blancas?

Convengamos en que cada bola de la urna tenga su número. Entonces una elección del volumen m puede ser descrita por la sucesión (i_1, \dots, i_m) , donde i_p significa el número de la bola sacada en el paso p , y durante la elección con retorno para i_p se tienen n valores posibles ($i_p = 1, 2, \dots, n$). En total habrá n^m elecciones diferentes (i_1, \dots, i_m) y todas ellas serán equiprobables.

Examinemos las elecciones en las que hay justamente k bolas blancas. Calculemos cuántas elecciones hay (i_1, \dots, i_m) con bolas blancas en los lugares fijados m_1, \dots, m_k (es decir, aquellas elecciones en las cuales las bolas i_{m_1}, \dots, i_{m_k} son blancas y las restantes bolas $m - k$, negras). Para cada una de las bolas blancas i_{m_1}, \dots, i_{m_k} se tiene r valores posibles (en la urna hay r bolas blancas), y para cada una de las bolas negras hay $n - r$ valores posibles. Por consiguiente, habrá $r^k \cdot (n - r)^{m-k}$ elecciones distintas (i_1, \dots, i_m) con k bolas blancas en los lugares fijados m_1, \dots, m_k . Pero entre los valores posibles $1, \dots, m$ podrán existir valores distintos m_1, \dots, m_k elegidos de C_m^k maneras, de modo que el número total de elecciones (i_1, \dots, i_m) con m bolas blancas será igual a $C_m^k r^k (n - r)^{m-k}$. Por consiguiente, la probabilidad de que entre las m bolas extraídas haya k blancas será

$$C_m^k \frac{r^k (n - r)^{m-k}}{n^m} = C_m^k p^k (1 - p)^{m-k},$$

donde $p = r/n$ coincide con la probabilidad de extraer una bola blanca en cada paso por separado.

Así pues, la magnitud aleatoria ξ que es igual al número de bolas blancas en la elección (con retorno) del volumen m , tiene la distribución de probabilidades

$$P_{\xi}(k) = C_m^k p^k (1-p)^{m-k}, \quad k = 0, 1, \dots, m, \quad (1.10)$$

donde p es la probabilidad de sacar una bola blanca en cada paso por separado. Esta es la tal llamada *distribución binomial*¹⁾ o como se dice frecuentemente, *la distribución de Bernoulli*.

Está claro, que para un gran número de bolas n y con un volumen pequeño de elección m , la elección sin retorno y la elección con retorno deben dar, prácticamente, un mismo resultado: la probabilidad de que hayan sido extraídas k bolas blancas será aproximadamente, la misma en ambos casos. Esto se reafirma también con la relación límite

$$\frac{C_{n_1}^k \cdot C_{n_2-n_1}^{m-k}}{C_n^m} = \frac{m!}{k!(m-k)!} \left[\frac{n_1(n_1-1)\dots(n_1-k+1)}{n(n-1)\dots(n-k+1)} \right] \times \\ \times \left[\frac{n_2(n_2-1)\dots(n_2-m+k+1)}{(n-k)(n-k-1)\dots(n-m+1)} \right] \sim \frac{m!}{k!(m-k)!} p^k (1-p)^{m-k},$$

donde $n_1 = r$, $n_2 = n - r$ (compárense las fórmulas (1.8), (1.10)), cuando la parte $p = n_1/n$ de bolas blancas queda constante en la urna, al aumentar ilimitadamente, el número n de bolas (recordemos, que p es la probabilidad de que la bola extraída de la urna al azar sea blanca).

El valor medio $M\xi$, durante la distribución binomial del número ξ de bolas blancas en la elección de volumen m , será

$$M\xi = \sum_{k=1}^m k \frac{m!}{k!(m-k)!} p^k (1-p)^{m-k} = mp \sum_{j=0}^{m-1} \times \\ \times \frac{(m-1)!}{j!(m-1-j)!} p^j (1-p)^{m-1-j} = mp [p + (1-p)]^{m-1} = mp. \quad (1.11)$$

¿Cuál será la distribución de las bolas blancas en la elección de volumen m , cuando la parte $p = r/n$ de bolas blancas disminuye en

1) $C_m^k = \frac{m!}{k!(m-k)!}$, $k=0, \dots, m$ son los coeficientes binomiales en el

desarrollo de $(p+q)^m = \sum_{k=0}^m C_m^k p^k q^{m-k}$ ($q=1-p$).

la urna ($p \rightarrow 0$), pero, en cambio se aumenta en volumen de la elección m ($m \rightarrow \infty$) de tal modo, que el valor medio de bolas $a = mp$ quede constante? Tenemos

$$P_{\xi}(0) = (1-p)^m = \left(1 - \frac{a}{m}\right)^m \rightarrow e^{-a},$$

para cada valor fijado $k = 1, 2, \dots$

$$\frac{P_{\xi}(k)}{P_{\xi}(k-1)} = \frac{C_m^k p^k (1-p)^{m-k}}{C_m^{k-1} p^{k-1} (1-p)^{m-k+1}} = \frac{a - (k-1)p}{k(1-p)} \rightarrow \frac{a}{k},$$

y por eso

$$\begin{aligned} P_{\xi}(1) &\rightarrow \frac{a}{1!} e^{-a}, \\ P_{\xi}(2) &\rightarrow \frac{a^2}{2!} e^{-a}, \\ &\dots \dots \dots \\ P_{\xi}(k) &\rightarrow \frac{a^k}{k!} e^{-a}, \\ &\dots \dots \dots \end{aligned} \quad (1.12)$$

De tal modo, la magnitud aleatoria ξ tiene en el límite la distribución de probabilidades

$$P_{\xi}(k) = \frac{a^k}{k!} e^{-a}, \quad k = 0, 1, \dots \quad (1.13)$$

Esta es la tal llamada *distribución de Poisson*.

A pesar de que en el límite se trata de hecho de una elección infinita ($m \rightarrow \infty$) y la cantidad ξ de «bolas blancas extraídas» puede ser tan grande como se quiera, su valor medio es igual a a :

$$M\xi = \sum_{k=1}^{\infty} k \frac{a^k}{k!} e^{-a} = a \sum_{k=0}^{\infty} \frac{a^k}{k!} e^{-a} = a. \quad (1.14)$$

Señalemos también que

$$\begin{aligned} D\xi &= M(\xi - a)^2 = M\xi^2 - a^2 = \sum_{k=1}^{\infty} k^2 \frac{a^k}{k!} e^{-a} - a^2 = \\ &= a \sum_{k=1}^{\infty} \left[(k-1) \frac{a^{k-1} e^{-a}}{(k-1)!} + \frac{a^{k-1}}{(k-1)!} e^{-a} \right] - a^2 = a^2 + a - a^2 = a. \end{aligned} \quad (1.15)$$

§ 2. DISTRIBUCION
DE POISSON.
CORRIENTE UNIFORME
DE SUCESOS
Y TIEMPO DE ESPERA

1. Distribución de Poisson en las partículas. Volvamos de nuevo al problema de la distribución de las partículas. Imaginémosnos que en una zona determinada del espacio V de Euclides se lanzan al azar r partículas, independientemente unas de otras, más exacto, si designamos por ξ_k la situación de la partícula k en el espacio ($k = 1, \dots, \dots, r$), entonces las magnitudes vectoriales aleatorias ξ_1, \dots, ξ_r son independientes y cada una de ellas tiene la distribución uniforme de probabilidades con la densidad

$$p(x) = \begin{cases} \frac{1}{|V|} & \text{para } x \in V \\ 0 & \text{para las restantes } x, \end{cases} \quad (2.1)$$

donde $|V|$ significa el volumen de la zona V . Designemos por λ al número medio de partículas en la unidad de volumen

$$\lambda = \frac{r}{|V|}.$$

Designemos por $\xi(v)$ el número de partículas, que caen en la zona correspondiente $v \subseteq V$.

Mostremos que en un «espacio fásico» V que se ensancha, ilimitadamente, y a la vez aumenta el número r de partículas (de modo, que el número medio de partículas por unidad de volumen queda, constantemente, igual a λ), la distribución límite de la magnitud aleatoria $\xi(v)$ será la de Poisson

$$P\{\xi(v) = k\} = \frac{(\lambda|v|)^k}{k!} e^{-\lambda|v|}, \quad k = 0, 1, \dots, \quad (2.2)$$

donde $|v|$ significa el volumen de la zona v y, aún más, para las zonas que no se cortan v_1, \dots, v_n

$$P\{\xi(v_1) = k_1, \dots, \xi(v_n) = k_n\} = \frac{(\lambda|v_1|)^{k_1} \dots (\lambda|v_n|)^{k_n}}{k_1! \dots k_n!} e^{-\lambda(|v_1| + \dots + |v_n|)}. \quad (2.3)$$

La última relación significa junto con la fórmula (2.2), que para las zonas que no se cortan v_1, \dots, v_n , las magnitudes aleatorias $\xi(v_1), \dots, \xi(v_n)$ son *independientes*. Esto es intuitivamente claro: durante la existencia de un número infinito de partículas independien-

tes, la caída de tal o cual número finito de partículas en la zona v_1 , no influye en el número de partículas que caen en las zonas v_2, \dots, v_n que no se cortan con v_1 .

Pasemos a la deducción de la fórmula (2.3). Cuando están fijados V y r , cada una de las partículas existentes r cae en una de las zonas $v_i \subseteq V$ con las correspondientes probabilidades $p_i = \frac{|v_i|}{|V|}$, $i = 0, 1, \dots, n$, donde por v_0 habíamos designado al complemento $v_1 \cup v_2 \cup \dots \cup v_n$ hasta toda la zona V . Dividamos nuestras partículas ξ_1, \dots, ξ_r en $n+1$ grupos de k_1, \dots, k_n, r_0 partículas en el grupo correspondiente $k_1 + \dots + k_n = k$, $r_0 = r - k$; el número de variantes distintas en tal división es igual a $\frac{r!}{k_1! \dots k_n! r_0!}$ (véase (1.6)). El suceso $\{\xi(v_1) = k_1, \dots, \xi(v_n) = k_n\}$ aparece entonces y sólo entonces, cuando un grupo determinado de k_1 partículas cae en la zona v_1 , un grupo determinado de k_2 partículas cae en la zona v_2 , etc. (las restantes $r_0 = r - k$ partículas caen en la zona v_0). Una partícula tomada por separado, independientemente de las otras partículas, cae en la zona correspondiente v_i con la probabilidad indicada anteriormente p_i , $i = 0, 1, \dots, n$, y por consiguiente, la probabilidad de que un grupo determinado de k_1 partículas caiga en v_1 , otro grupo de k_2 partículas caiga en v_2 , etc. será igual al producto $p_1^{k_1} \cdot p_2^{k_2} \cdot \dots \cdot p_n^{k_n} p_0^{r_0}$. Las distintas variantes de división en los grupos indicados dan $\frac{r!}{k_1! \dots k_n! r_0!}$ resultados incompatibles, cada uno de los cuales tiene la probabilidad indicada anteriormente $p_1^{k_1} \cdot \dots \cdot p_n^{k_n} p_0^{r_0}$ y por consiguiente, el suceso $\{\xi(v_1) = k_1, \dots, \xi(v_n) = k_n\}$, que coincide con la unión de esos resultados, tiene la probabilidad

$$\frac{r!}{k_1! \dots k_n! r_0!} p_1^{k_1} \dots p_n^{k_n} p_0^{r_0}. \quad (2.4)$$

Utilicemos la fórmula de Sterling según la cual

$$r! \sim \sqrt{2\pi r} r^r e^{-r}, \quad (r-k)! \sim \sqrt{2\pi(r-k)} (r-k)^{r-k} e^{-(r-k)}.$$

Para los valores fijados k_1, \dots, k_n y $r \rightarrow \infty$ tenemos

$$\sqrt{\frac{r}{r-k}} \rightarrow 1, \quad \left(\frac{r}{r-k}\right)^{r-k} = \left(1 + \frac{k}{r-k}\right)^{r-k} \rightarrow e^k,$$

$$p_i^{k_i} r^{k_i} = \left(\frac{|v_i|}{|V|} r\right)^{k_i} = (\lambda |v_i|)^{k_i}$$

$$i = 1, \dots, n,$$

$$|v_0| = |V| - (|v_1| + \dots + |v_n|)$$

y

$$\begin{aligned} p_0^{r_0} &= \left(1 - \frac{|v_1| + \dots + |v_n|}{|V|}\right)^{r-k} = \\ &= \left(1 - \frac{\lambda(|v_1| + \dots + |v_n|)}{r}\right)^{r-k} \rightarrow e^{-\lambda(|v_1| + \dots + |v_n|)}. \end{aligned}$$

En total para $r \rightarrow \infty$ y el valor constante $\lambda = r/V$ tenemos

$$\frac{r!}{k_1! \dots k_n! k_0!} p_1^{k_1} \dots p_n^{k_n} p_0^{k_0} \rightarrow \frac{(\lambda |v_1|)^{k_1} \dots (\lambda |v_n|)^{k_n}}{k_1! \dots k_n!} \times e^{-\lambda(|v_1| + \dots + |v_n|)},$$

es decir, en el límite obtenemos la fórmula (2.3).

La distribución de Poisson (2.3) se determina por un solo parámetro λ , $0 < \lambda < \infty$, que es el número medio de partículas en la unidad de volumen: $\lambda = M\xi(v)/|v|$ (véase (1.14)). Esta distribución posee las siguientes propiedades específicas:

a) para las zonas que no se cortan v_1, \dots, v_n , las magnitudes aleatorias $\xi(v_1), \dots, \xi(v_n)$ son mutuamente independientes;

b) la probabilidad de caer en la zona v tal o cual número de partículas no depende de la disposición de la zona v en el espacio fásico (sólo depende de su volumen $|v|$);

c) la probabilidad de caer una partícula en la zona de volumen pequeño $|v|$, con exactitud hasta de un infinitamente pequeño de orden superior, es proporcional a $|v|$:

$$P\{\xi(v) = 1\} = \lambda |v| e^{-\lambda|v|} = \lambda |v| + o(|v|),$$

y la probabilidad de caer más de una partícula en la zona de volumen $|v|$ es una cantidad infinitamente pequeña de orden superior para $|v| \rightarrow 0$:

$$P\{\xi(0) > 1\} = \sum_{k=2}^{\infty} \frac{(\lambda |v|)^k}{k!} e^{-\lambda|v|} = o(|v|).$$

Mostremos, que para cualquier distribución de las partículas, que satisfaga a las condiciones a), b) y c) indicadas, anteriormente, es obligatoria la distribución de Poisson.

En esencia, sólo hay que mostrar que tiene lugar la fórmula (2.2) (de donde para la condición a) se deduce también la fórmula más general (2.3)). Utilicemos el mismo modo de división de la zona v en pequeñas celdas Δ_k , $k = 1, \dots, n$ (de volumen $|\Delta_k| = \frac{|v|}{n}$), como en el punto 2 del párrafo anterior (compárese con (1.9) y demás). Debido a las condiciones a) y b), la probabilidad del suceso

$$\begin{aligned} \xi(\Delta_{i_p}) = 1, \quad p = 1, \dots, k, \quad \xi(\Delta_{i_p}) = 0, \\ p = k+1, \dots, n \end{aligned}$$

es una misma para todas las combinaciones $(i_1, \dots, i_k), (i_{k+1}, \dots, i_n)$ y es igual al producto

$$(P\{\xi(\Delta_i) = 1\})^k (P\{\xi(\Delta_i) = 0\})^{n-k};$$

y debido a la condición c), para $n \rightarrow \infty$ la probabilidad de que caiga más de una partícula en cualquier celda no supera a la suma

$$\sum_{k=1}^n \mathbf{P} \{ \xi(\Delta_k) > 1 \} = n \cdot o \{ |\Delta_k| \} \rightarrow 0,$$

y por consiguiente, suponiendo que

$$p = \mathbf{P} \{ \xi(\Delta_1) = 1 \} = \lambda \frac{|v|}{n} + o \left(\frac{1}{n} \right),$$

para $n \rightarrow \infty$ obtenemos (véase (1.13))

$$\begin{aligned} \mathbf{P} \{ \xi(v) = k \} &\sim 1104 \quad 168 \quad 2 \\ &\sim \sum_{(i_1, \dots, i_k)} \mathbf{P} \{ \xi(\Delta_{i_p}) = 1, p=1, \dots, k, \xi(\Delta_{i_p}) = 0, p=k+1, \dots, n \} = \\ &= C_n^k (\mathbf{P} \{ \xi(\Delta_1) = 1 \})^k (\mathbf{P} \{ \xi(\Delta_1) = 0 \})^{n-k} \sim \\ &\sim C_n^k p^k (1-p)^{n-k} \sim \frac{a^k}{k!} e^{-a}, \end{aligned}$$

donde $a = np = \lambda |v|$.

Señalemos además la siguiente característica de la distribución de Poisson: *bajo la condición de que caigan justamente r partículas en una determinada zona V , su distribución dentro de esta zona será tal, como si las hubiesen tirado al azar en V , independientemente una de otra; más exactamente, las magnitudes vectoriales ξ_1, \dots, ξ_r , que describen la posición de estas partículas, serán independientes y estarán distribuidas uniformemente en la zona V (con una densidad de probabilidad de la forma (2.1)).*

En efecto, para cualesquiera zonas que no se cortan v_1, \dots, v_n la probabilidad convencional $\mathbf{P} \{ \xi(v_1) = k_1, \dots, \xi(v_n) = k_n \mid \xi(V) = r \}$ es

$$\begin{aligned} &\frac{\mathbf{P} \{ \xi(v_1) = k_1, \dots, \xi(v_n) = k_n, \xi(v_0) = r_0 \}}{\mathbf{P} \{ \xi(V) = r \}} = \\ &= \frac{\frac{(\lambda |v_1|)^{k_1}}{k_1!} \dots \frac{(\lambda |v_n|)^{k_n}}{k_n!} e^{-\lambda(|v_1| + \dots + |v_n| + |v_0|)}}{(\lambda |V|)^r e^{-\lambda|V|}} = \\ &= \frac{r!}{k_1! \dots k_n! r_0!} p_1^{k_1} \dots p_n^{k_n} p_0^{r_0}, \end{aligned}$$

lo que demuestra nuestra afirmación (compárese con (2.4)).

Ejemplo (proceso de la desintegración radiactiva). Como es conocido, el radio (Ra) con el transcurso del tiempo se transforma en radón (Rn). El núcleo del átomo de Ra que se desintegra emite la llamada partícula α (núcleo del átomo de helio He). La desintegración de un átomo por separado de Ra se realiza independientemente del estado de otros átomos de la substancia y la emisión de rayos α se representa

en sí a un flujo de una gran cantidad de partículas independientes. Es natural esperar, que la distribución de las partículas α en el tiempo será la distribución de Poisson; la probabilidad de que el número $\xi(\Delta)$ de partículas α emitidas sea igual a k durante el intervalo de tiempo Δ , deberá ser

$$P\{\xi(\Delta) = k\} = \frac{\lambda |\Delta|^k}{|k!|} e^{-\lambda|\Delta|}, \quad k = 0, 1, \dots, \quad (2.5)$$

donde $|\Delta|$ es la amplitud del intervalo de tiempo examinado, y λ es el número medio de partículas α emitidas en la unidad de tiempo: $\lambda = \frac{M\xi(\Delta)}{\Delta}$ (el proceso de desintegración radioactiva será examinado más detalladamente en el siguiente punto 2). Este resultado se confirma bien con los datos experimentales¹⁾.

Naturalmente, la distribución de Poisson se origina no sólo al examinar el flujo uniforme de partículas independientes. Digamos, se puede tratar de una corriente homogénea de exigencias que se reciben en un determinado sistema de servicio (por ejemplo, al surtidor de gasolina llegan los autos, se reciben peticiones en el buró de información, en la estación telefónica se registra un flujo de fallos en los abonados, etc.). En general, la cuestión puede tratar sobre una corriente homogénea de sucesos independientes, registrados en el tiempo: los momentos de aparición de estos sucesos se puede interpretar como «partículas» distribuidas aleatoriamente sobre la recta real (eje del tiempo). Si en el intervalo de tiempo Δ se cumplen, para el número de sucesos $\xi(\Delta)$, las condiciones descritas anteriormente a), b) y c):

a) para los intervalos de tiempo que no se cortan $\Delta_1, \dots, \Delta_n$, las magnitudes aleatorias $\xi(\Delta_1), \dots, \xi(\Delta_n)$ serán independientes;

b) la probabilidad de tal o cual suceso en el intervalo Δ no depende del origen de recuento del tiempo (no depende de la posición de Δ en el eje de tiempo);

c) la probabilidad de aparición del suceso en un pequeño intervalo de tiempo Δ es proporcional a la amplitud $|\Delta|$ de este intervalo, y la probabilidad de aparición de más de un suceso, tiene un orden infinitesimal superior en comparación con $|\Delta|$, entonces a tal corriente de sucesos se la llama *de Poisson* y la distribución de probabilidades $\xi(\Delta)$ se describe por la fórmula (2.5).

2. Tiempo de espera del suceso aleatorio. Examinemos los lanzamientos sucesivos de la moneda hasta la primera caída «cruz», considerando que la moneda se lanza una vez en la unidad de tiempo. Está claro que la duración de la prueba (tiempo de espera de la «cruz») es aleatorio casual, con ello, debido a la independencia de las pruebas por separado (lanzamiento de la moneda) en cualquier momento

¹⁾ Véase, por ejemplo, el libro citado anteriormente de W. Feller.

$t < 0$, la posición del observador no será nada mejor (en sentido del tiempo de espera de «cruz») que en el momento inicial $t = 0$: la probabilidad de esperar después del momento t un tiempo más s , es la misma que esperar este mismo tiempo s a partir del momento inicial; más exactamente, si τ es el tiempo de espera, entonces para la condición $\tau > t$ deberá ser

$$\mathbf{P}(\tau > t + s \mid \tau > t) = \mathbf{P}(\tau > s). \quad (2.6)$$

En efecto, para cualquier $t = 0, 1, \dots$, el suceso $\{\tau > t\}$ significa que en t pruebas independientes no caerá ni una vez «cruz» y

$$\mathbf{P}\{\tau > t\} = (1 - p)^t, \quad t = 0, 1, \dots, \quad (2.7)$$

donde $p = 1/2$ es la probabilidad de que caiga «cruz» en cada prueba por separado, por eso

$$\mathbf{P}\{\tau > t + s \mid \tau > t\} = \frac{\mathbf{P}\{\tau > t + s\}}{\mathbf{P}\{\tau > t\}} = (1 - p)^s = \mathbf{P}\{\tau > s\}.$$

La relación (2.6), considerada por sí misma, es característica para el tiempo de espera τ de tal suceso A , cuya espera hasta el momento t de ningún modo aproxima, prácticamente, su aparición. En esta situación, en el ejemplo más elemental que fue expuesto anteriormente, la distribución de la magnitud aleatoria τ será tal que

$$\mathbf{P}\{\tau > t\} = e^{-\lambda t}, \quad t \geq 0, \quad (2.8)$$

donde λ es un determinado parámetro ($\lambda \geq 0$); para el valor discreto de $t = 0, 1, \dots$, la fórmula (2.8) puede ser escrita en la forma (2.7) si ponemos $p = 1 - e^{-\lambda}$.

Para el valor discreto de t esta distribución de probabilidades se llama distribución geométrica:

$$P_{\tau}(k) = e^{-\lambda(k-1)} - e^{-\lambda k} = p(1 - p)^{k-1}, \quad k = 1, 2, \dots \quad (2.9)$$

Para el valor continuo de t , $0 \leq t < \infty$, la distribución (2.8) tiene una densidad de probabilidad

$$p_{\tau}(t) = \begin{cases} \lambda e^{-\lambda t} & \text{para } t \geq 0, \\ 0 & \text{para } t < 0 \end{cases} \quad (2.10)$$

(ya que la función de distribución τ es $F_{\tau}(t) = 1 - \mathbf{P}\{\tau > t\}$ y $p_{\tau}(t) = F'_{\tau}(t)$); tal distribución se llama *exponencial*. El parámetro indicado λ tiene un sentido probabilístico sencillo, es decir, $1/\lambda$ es el tiempo medio de espera

$$\frac{1}{\lambda} = \mathbf{M}\tau = \int_0^{\infty} t p_{\tau}(t) dt. \quad (2.11)$$

Deduzcamos la fórmula (2.8) de la relación (2.6) suponiendo que $\tau = 0$, sólo cuando la probabilidad es igual a cero, y que en caso del tiempo continuo, la magnitud τ tiene para $t \geq 0$ una densidad de probabilidad continua. Determinemos la función $f(t)$ por la igualdad $f(t) = P\{\tau > t\}$; según la fórmula de la probabilidad total

$$P\{\tau > t + s\} = P\{\tau > s\} \cdot P\{\tau > t\},$$

y, por consiguiente, la función $f(t)$ será tal, que para cualesquiera $s, t \geq 0$

$$f(s + t) = f(s) f(t),$$

o bien

$$\log f(s + t) = \log f(s) + \log f(t),$$

siendo $f(0) = P\{\tau > 0\} = 1$. Para el valor discreto de $t = 0, 1, \dots$ obtenemos de aquí que $\log f(t), \log f(0) = 0$ es una función lineal:

$$\log f(t) = t \log f(1)$$

y

$$f(t) = (1 - p)^t = e^{-\lambda t}, \quad \text{donde } p = 1 - f(1) = P\{\tau = 1\}.$$

Para el tiempo continuo t (en condiciones de que existe la densidad $p_\tau(t) = -f'(t + 0), t \geq 0$) la diferenciación por s nos da

$$\frac{f'(t+s)}{f(t+s)} = \frac{f'(s)}{f(s)},$$

de donde para $s = 0$ obtenemos

$$\frac{f'(t)}{f(t)} = -\lambda,$$

donde $-\lambda = f'(0) \leq 0$, ya que $f(0) = 1$ es el valor máximo de la función $f(t) = P\{\tau > t\}, t \geq 0$. Por consiguiente, $f(t) = e^{-\lambda t}$, lo que se exigía demostrar.

Ejemplo (modelo de la desintegración radiactiva). Anteriormente ya recordamos que la distribución de las partículas α emitidas por el radio es la distribución de Poisson. Supongamos, que la transformación del radio Ra en radón Rn (en la que se emiten partículas α) se realiza de tal modo que en el intervalo de tiempo (t_0, t_1) un átomo de Ra tomado por separado se transforma en un átomo de Rn con una determinada probabilidad $p = p(t)$, dependiente de la amplitud $t = t_1 - t_0$ del intervalo examinado.

Si hasta el momento t_1 no ocurrió la transición Ra \rightarrow Rn, entonces en el nuevo momento inicial t_1 trataremos del mismo átomo de Ra y la transición Ra \rightarrow Rn en el tiempo sucesivo $s = t_2 - t_1$, de acuerdo a nuestra suposición, deberá realizarse con la correspondiente probabilidad $p(s)$. Evidentemente, el tiempo τ de espera de la desintegración Ra \rightarrow Rn del átomo dado de Ra (empezando desde el momento

t_0) satisface a la relación (2.6), ya que la probabilidad condicional $\mathbf{P}\{\tau > t + s \mid \tau > t\}$ es la probabilidad de que el átomo Ra conservándose hasta el momento t_1 , no se desintegra tampoco en el siguiente intervalo de tiempo (t_1, t_2) , y por suposición, esta probabilidad (así como la probabilidad $\mathbf{P}\{\tau > s\}$) es igual a $1 - p(s)$, $s = t_2 - t_1$. Por consiguiente, el tiempo de espera τ tiene la distribución exponencial de probabilidades (2.8).

Aclaremos el sentido físico de la constante correspondiente λ . Ya sabemos, que $1/\lambda$ es el tiempo medio de espera de la desintegración $\text{Ra} \rightarrow \text{Rn}$ para un átomo por separado (véase (2.11)), es natural considerar, que la constante λ es una misma para todos los átomos de Ra.

Designemos por n_0 la cantidad de radio (digamos, el número de átomos de Ra) en el momento inicial t_0 . Cada átomo de Ra por separado se desintegra ($\text{Ra} \rightarrow \text{Rn}$) en el tiempo t sucesivo con la probabilidad

$$p(t) = 1 - e^{-\lambda t},$$

y si designamos por $\xi(t)$ al número de átomos Ra desintegrados en este tiempo (que coincide con el número de partículas α emitidas) entonces el valor medio $\mathbf{M}\xi(t)$ será

$$\mathbf{M}\xi(t) = n_0 p(t) = n_0 (1 - e^{-\lambda t}).$$

Por consiguiente, la cantidad restante de radio después del tiempo t será en su media

$$n(t) = \mathbf{M}(n_0 - \xi(t)) = n_0 - \mathbf{M}\xi(t) = n_0 e^{-\lambda t}.$$

La dependencia exponencial de t puede ser caracterizada no sólo por el exponente λ , sino también por la llamada constante T de semi-desintegración, definida como el tiempo durante el cual se desintegra exactamente, la mitad de la substancia inicial; para hallar T tenemos la ecuación

$$n(T) = \frac{n_0}{2},$$

que da el valor $T = \frac{\log 2}{\lambda}$ (para el radio ha sido hallado experimentalmente $T = 1590$ años).

Examinemos ahora un modelo general de la corriente de sucesos de Poisson, en el cual se supone que, independientemente de cuándo y cuántos sucesos se realizaron hasta el momento corriente t_0 , en el intervalo de tiempo (t_0, t_1) aparecen exactamente k sucesos con la correspondiente probabilidad $\frac{(\lambda t)^k}{k!} e^{-\lambda t}$, $k = 0, 1, \dots$ (donde $t = t_1 - t_0$).

Vemos que el tiempo aleatorio de espera (después del momento t_0) del suceso de turno tiene una distribución exponencial de probabilidades

con el parámetro λ :

$$P \{ \tau_1 - t_0 > t \} = P \{ \xi(t_0, t_1) = 0 \} = e^{-\lambda t}, \quad (2.12)$$

donde τ_1 significa el momento de aparición del suceso (el primero después del momento t_0), y $\xi(t_0, t_1)$ es el número de sucesos en el intervalo de tiempo (t_0, t_1) ; el parámetro λ de esta distribución exponencial es el mismo que en la corriente inicial de Poisson (λ es el número medio de sucesos ocurridos en la unidad de tiempo).

Por consiguiente, en la corriente de sucesos de Poisson, incluso después de una larga espera del suceso de turno, éste se realiza en el siguiente intervalo de tiempo (t_1, t_2) con la misma probabilidad $1 - e^{-\lambda s}$, $s = t_2 - t_1$, como si no hubiese habido nunca ninguna espera y t_1 , fuese el momento inicial de observación ($t_1 = t_0$):

$$P \{ \tau_1 - t_0 > t + s \mid \tau_1 - t_0 > t \} = \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}} = e^{-\lambda s} = P \{ \tau_1 - t_0 > s \}.$$

Supongamos que τ_1, \dots, τ_n son los momentos aleatorios, en que se realizan, siguiendo uno a otro, los sucesos de la corriente de Poisson examinada. Tomemos un momento arbitrario t_1 , $\tau_1 \leq t_1 < \tau_2$. Después del momento t_1 , el suceso de turno aparece, independiente de τ_1 , durante el tiempo t con la probabilidad $1 - e^{-\lambda t}$, y por eso, la probabilidad condicional de que $\{ \tau_2 - t_1 \leq t \}$ para cualquier τ_1 , $\tau_1 \leq t_1$ es

$$P \{ \tau_2 - t_1 \leq t \mid \tau_1, \tau_1 \leq t_1 \} = 1 - e^{-\lambda t}.$$

De aquí, para $\tau_1 = t_1$ obtenemos

$$P \{ \tau_2 - t_1 \leq t \mid \tau_1 = t_1 \} = P \{ \tau_2 - \tau_1 \leq t \mid \tau_1 = t_1 \} = 1 - e^{-\lambda t}.$$

Se ve que la densidad condicional de probabilidad $p(t \mid t_1)$, $t \geq 0$ de la magnitud $\tau_2 - \tau_1$ a condición de que $\tau_1 = t_1$ para todos los $t_1 \geq t_0$ es una misma e igual a $\lambda e^{-\lambda t}$. Por consiguiente, (según la fórmula (3.1)) la densidad «incondicional» es

$$p(t) = \int_{t_0}^{\infty} p(t \mid t_1) p_{\tau_1}(t_1) dt_1 = \lambda e^{-\lambda t},$$

es decir, la amplitud del intervalo de tiempo $\tau_2 - \tau_1$ tiene también la distribución exponencial con el mismo parámetro λ :

$$P \{ \tau_2 - \tau_1 > t \} = e^{-\lambda t}, \quad t \geq 0. \quad (2.13)$$

Consideraciones análogas, referentes a los momentos τ_{n-1} y τ_n , permiten concluir que independientemente de los momentos de aparición de los sucesos anteriores $\tau_1, \tau_2, \dots, \tau_{n-1}$ (lo que significa

también que, independientemente de los intervalos $\tau_2 - \tau_1, \dots$
 $\dots, \tau_{n-1} - \tau_{n-2}, \tau_1 - t_0$

$$P \{ \tau_n - \tau_{n-1} > t \mid \tau_1, \tau_2, \dots, \tau_{n-1} \} = P \{ \tau_n - \tau_{n-1} > t \} = e^{-\lambda t}. \quad (2.14)$$

Como resultado llegamos a la siguiente conclusión: los intervalos $\tau_1 - t_0, \tau_2 - \tau_1, \dots$, que separan los sucesos de la corriente de Poisson que aparecen sucesivamente, son magnitudes aleatorias independientes, que tienen la misma distribución exponencial con el parámetro λ (λ es el número medio de sucesos realizados en la unidad de tiempo).

Sean ξ_1, ξ_2, \dots magnitudes aleatorias independientes que tienen la misma distribución exponencial con parámetro λ ; supongamos que

$$S_n = \xi_1 + \dots + \xi_n. \quad (2.15)$$

Si $\xi_1 = \tau_1 - t_0, \xi_2 = \tau_2 - \tau_1, \dots$ son los intervalos en los sucesos de la corriente de Poisson, entonces la magnitud aleatoria S_n es igual al tiempo de espera del suceso n según el cálculo: $S_n = \tau_n - t_0$. Evidentemente, el suceso $\{S_n \leq t\}$ significa que en el tiempo t por lo menos se realizó n sucesos y, por consiguiente, la función de la distribución de probabilidades de la magnitud no negativa S_n es

$$P \{ S_n \leq t \} = 1 - \sum_{k=0}^{n-1} \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \quad t \geq 0.$$

Diferenciando esta expresión por t , hallamos la densidad de probabilidad correspondiente, que tiene la forma

$$p(t) = \lambda \frac{(\lambda t)^{n-1}}{(n-1)!} e^{-\lambda t}, \quad t \geq 0 \quad (2.16)$$

($p(t) = 0$ para $t < 0$).

La distribución de probabilidades obtenida entra en la familia de las tal llamadas «distribuciones gamma» (en adelante nos pondremos en conocimiento con algunas de ellas). La densidad general de la «distribución gamma» se da por la fórmula

$$p(t) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\beta t} & \text{para } t \geq 0, \\ 0 & \text{para } t < 0, \end{cases} \quad (2.17)$$

donde α, β son parámetros positivos, y

$$\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1} e^{-t} dt \quad (2.18)$$

$$(\Gamma(\alpha + 1) = \alpha \Gamma(\alpha), \quad \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}, \quad \Gamma(n + 1) = n!)$$

es la llamada función gamma de Euler. La distribución (2.16) se obtiene para $\alpha = n$, $\beta = \lambda$; para $n = 1$ la cuestión se refiere a la distribución exponencial.

Como conclusión de este punto nos detendremos en un hecho, que puede parecer paradójico a primera vista. Precisamente, examinemos, como anteriormente, la corriente de sucesos de Poisson: independientemente, de cuándo y de cuántos sucesos se realizaron (o se realizarán) fuera de un determinado intervalo de tiempo (s , t), en este intervalo ocurren k sucesos con la correspondiente probabilidad

$$\frac{(\lambda(t-s))^k}{k!} e^{-\lambda(t-s)}, \quad k = 0, 1, \dots$$

Supongamos que los sucesos se registran en el eje de tiempo $-\infty < t < \infty$ y el observador, que llegó en determinado momento t_0 , se interesa por los intervalos entre el suceso anterior (el momento de su aparición lo designamos por τ_{-1}) y el suceso siguiente cuyo momento de aparición lo designamos por τ_1 , como anteriormente.

Nosotros establecimos que los intervalos (digamos $\tau_2 - \tau_1$, $\tau_3 - \tau_2$, etc.) entre los sucesos consecutivos, lo mismo que la magnitud $\tau_1 - t_0$, tienen una distribución exponencial con el parámetro λ . Se podría pensar que la diferencia $(\tau_1 - \tau_{-1})$ tiene la misma distribución, aunque, surge en seguida la incompreensión debido a la desigualdad $(\tau_1 - \tau_{-1}) > (\tau_1 - t_0)$, de la cual se deduce, que las magnitudes indicadas no pueden tener una misma distribución de probabilidades. Se puede abordar la solución del problema sobre la distribución de la magnitud $(\tau_1 - \tau_{-1})$, teniendo en cuenta la circunstancia de que la dirección del transcurso del tiempo no influye en nada sobre las leyes probabilísticas de la corriente de sucesos de Poisson, ellas serán también las mismas para la corriente de sucesos «invertida hacia atrás» (formalmente, se obtiene con el cambio de la variable t por $-t$, $-\infty < t < \infty$). Por consiguiente, para cualquier momento t_0 fijado, de modo que $\tau_{-1} < t_0 < \tau_1$ no sólo, la magnitud $(\tau_1 - t_0)$, que es el tiempo de espera del suceso de turno, tendrá la distribución exponencial de probabilidades, sino que también deberá poseer esta propiedad la magnitud $(t_0 - \tau_{-1})$, que es el tiempo de espera del suceso de turno en la «corriente invertida». Además de esto, como ya indicamos, la magnitud $(\tau_1 - t_0)$ no depende de cuándo se realizó, precisamente, el suceso anterior al momento t_0 y de tal modo, para la condición $\tau_{-1} < t_0 < \tau_1$ las magnitudes $(\tau_1 - t_0)$ y $(t_0 - \tau_{-1})$ serán independientes. Por consiguiente, la distribución de la diferencia $(\tau_1 - \tau_{-1})$, que es el intervalo entre los sucesos vecinos (a condición de que $\tau_{-1} < t_0 < \tau_1$, para un determinado momento fijo t_0), será la misma, que en la suma $(\tau_1 - t_0) + (t_0 - \tau_{-1})$ de dos magnitudes independientes, que tienen la misma distribución exponencial (la densidad de probabilidad correspondiente se da por la fórmula (2.16) para $n = 2$).

§ 3. PRUEBAS DE BERNOULLI.
MOVIMIENTO BROWNIANO
Y DISTRIBUCIONES
DE PROBABILIDADES LIGADAS
A ESTE

1. Pruebas de Bernoulli y distribución binomial. Aproximaciones de Poisson y normal. Las pruebas iguales e independientes entre sí, en cada una de las cuales se examina un determinado suceso A , que aparece con una misma probabilidad positiva $p = \mathbf{P}(A)$, se denominan *pruebas de Bernoulli*. El propio suceso A se llama, convencionalmente, «éxito» y el suceso complementario \bar{A} que aparece en cada una de las pruebas examinadas con la probabilidad $q = 1 - p$, convencionalmente se llama «fallo».

Designemos al «éxito» en la prueba por separado con la unidad (1), y al «fallo» con el cero (0). Entonces cada resultado elemental $\omega = (\omega_1, \dots, \omega_n)$ de n pruebas da la sucesión correspondiente de ceros y unidades ($\omega_i = 1$ con la probabilidad p y $\omega_i = 0$ con la probabilidad $1 - p$). La probabilidad $\mathbf{P}(\omega)$ del resultado elemental ω , para el cual aparece justamente k veces el «éxito» y $n - k$ veces aparece el «fallo», debido a la independencia de las pruebas por separado, es

$$\mathbf{P}(\omega) = p^k q^{n-k}.$$

Examinemos la magnitud aleatoria ξ , igual al número total de «éxitos» en n pruebas de Bernoulli: $\xi(\omega) = k$, si aparece justamente k veces el «éxito» en el resultado elemental ω .

El número de resultados distintos ω , que conducen a un mismo número k de «éxitos», es igual al número de combinaciones de n tomados en grupos de k , que alcanza a $C_n^k = \frac{n!}{k!(n-k)!}$. Todos estos resultados ω tienen una misma probabilidad $\mathbf{P}(\omega) = p^k q^{n-k}$, de modo que el suceso $\{\xi = k\}$ tiene la probabilidad $C_n^k p^k q^{n-k}$. Por lo tanto, la distribución de las probabilidades de la magnitud aleatoria ξ se da por la fórmula

$$P_\xi(k) = C_n^k p^k q^{n-k}, \quad k = 0, \dots, n. \quad (3.1)$$

Esta es la *distribución binomial* (fig. 12), que ya apareció antes (véase (1.10)) al examinar el esquema de la urna, en el cual el «éxito» era la extracción de la bola blanca, y el «fallo» la extracción de la bola negra. Ella se da por dos parámetros: por la probabilidad de un «éxito» por separado p y por el número n de pruebas.

Es útil indicar, que la magnitud aleatoria ξ es la suma de n magnitudes *independientes* ξ_1, \dots, ξ_n que se determinan del siguiente

modo: $\xi_k = 1$ si aparece el «éxito» en la prueba k -ésima y $\xi_k = 0$, si en esta prueba aparece el «fallo»:

$$\xi = \xi_1 + \dots + \xi_n. \quad (3.2)$$

Si observamos que

$$M\xi_k = p, \\ D\xi_k = M\xi_k^2 - (M\xi_k)^2 = p - p^2 = p(1-p) = pq,$$

entonces utilizando (3.2), obtenemos para la esperanza matemática y la dispersión de la magnitud aleatoria ξ la siguiente expresión:

$$M\xi = np, \quad D\xi = npq. \quad (3.3)$$

Para un gran número n de pruebas y una probabilidad p relativamente pequeña, cuando cada uno de los éxitos es un suceso relativa-

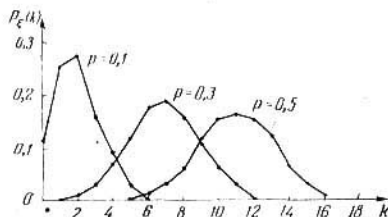


Fig. 12. Los puntos en los gráficos significan las probabilidades binomial $P_{\xi}(k) = C_n^k p^k (1-p)^{n-k}$, $k = 0, 1, \dots, n$, para $n = 20$, que responden a parámetro $p = 0, 1, 0,3, 0,5$, respectivamente.

mente raro, pero el número medio de «éxitos» np es suficientemente grande, se puede considerar que

$$P_{\xi}(k) \approx \frac{a^k}{k!} e^{-a}, \quad k = 0, 1, \dots, \quad (3.4)$$

donde $a = np$ es el número medio de éxitos (véase (1.12)). Esta es la tal llamada *aproximación de Poisson para la distribución binomial* (fig. 13).

En las aplicaciones, las pruebas de Bernoulli sirven, frecuentemente, para la determinación experimental de la probabilidad

$$p = P(A)$$

de tal o cual suceso A (del «éxito» en una prueba por separado) para cuya valuación empírica se toma la frecuencia observada del «éxito»

$$p \approx \frac{n(A)}{n} \left(= \frac{1}{n} \sum_{k=1}^n \xi_k \right). \quad (3.5)$$

Debido a la ley de los grandes números

$$\rho = \lim_{n \rightarrow \infty} \frac{n(A)}{n},$$

con ello, la desviación media cuadrática de la valuación $\frac{n(A)}{n}$ de la probabilidad desconocida $\mathbf{P}(A)$, tomada por el observador, desde su

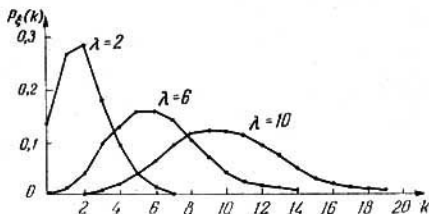


Fig. 13. Los puntos en los gráficos significan las probabilidades de Poisson $P_{\xi}(k) = \frac{\lambda^k}{k!} e^{-\lambda}$, $k = 0, 1, \dots$, que responden al parámetro $\lambda = 2, 6, 10$, respectivamente ($\lambda = np$ para $n = 20$ y $p = 0,1, 0,3, 0,5$; compárese con la fig. 12).

verdadero valor p , llega hasta

$$\left\| \frac{n(A)}{n} - p \right\| = \frac{\sqrt{DS_n}}{n} = \sqrt{\frac{pq}{n}},$$

donde $S_n = \sum_{k=1}^n \xi_k$ sólo es la nueva designación para el número de «éxitos» en las n pruebas independientes. De tal modo en la representación

$$\frac{n(A)}{n} - p = \sqrt{\frac{pq}{n}} \cdot S_n^*$$

la magnitud aleatoria $S_n^* = \frac{S_n - MS_n}{\sqrt{DS_n}}$ tiene el valor medio 0 y la dispersión 1: se ve que la desviación probable de la frecuencia $\frac{n(A)}{n}$ del valor desconocido de p es del orden $\frac{1}{\sqrt{n}}$. Más exacto es el resultado siguiente.

Teorema (de Moivre—Laplace). Para $n \rightarrow \infty$ la magnitud aleatoria S_n^* tiene una distribución de probabilidades límite, a saber

$$\lim_{n \rightarrow \infty} \mathbf{P}\{x' \leq S_n^* \leq x''\} = \frac{1}{\sqrt{2\pi}} \int_{x'}^{x''} e^{-x^2/2} dx \quad (3.6)$$

para cualesquiera valores fijados x', x'' , ($x' \leq x''$).

Esta es la llamada *distribución de probabilidades normal* (o de Gauss) con la densidad

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad -\infty < x < \infty. \quad (3.7)$$

Pasemos a la demostración de la relación límite (3.6). La magnitud normalizada S_n^* , según el caso, toma uno de los valores x de la forma

$$x = \frac{k-np}{\sqrt{npq}}, \quad k=0, 1, \dots, n,$$

que dividen los segmentos $\left[\frac{-np}{\sqrt{npq}}, \frac{nq}{\sqrt{npq}} \right]$ en intervalos iguales de longitud

$$\Delta x = \frac{1}{\sqrt{npq}},$$

siendo

$$P\{S_n^* = x\} = P_{S_n}(k) = \frac{n!}{k!(n-k)!} p^k q^{n-k}, \quad k=0, 1, \dots, n.$$

Evidentemente, para $n \rightarrow \infty$

$$k = np + \sqrt{npq}x \rightarrow \infty, \quad n-k = nq - \sqrt{npq}x \rightarrow \infty$$

uniformemente según x , $x' \leq x \leq x''$. Utilizando la fórmula de Sterling obtenemos, que

$$\begin{aligned} P_{S_n}(k) &\sim \frac{\sqrt{2\pi n} n^n e^{-n}}{\sqrt{2\pi k} k^k e^{-k} \sqrt{2\pi (n-k)} (n-k)^{n-k} e^{-(n-k)}} = \\ &= \frac{1}{\sqrt{2\pi}} \sqrt{\frac{n}{k(n-k)}} \left(\frac{np}{k}\right)^k \left(\frac{nq}{n-k}\right)^{n-k} \end{aligned}$$

Luego,

$$\frac{k}{np} = 1 + \sqrt{\frac{q}{np}} x, \quad \frac{n-k}{nq} = 1 - \sqrt{\frac{p}{nq}} x$$

y utilizando el desarrollo

$$\ln(1 + \alpha_n) \sim \alpha_n - \frac{1}{2} \alpha_n^2$$

(para $\alpha_n \rightarrow 0$) obtenemos que

$$\begin{aligned} \ln\left(\frac{k}{np}\right)^{-k} &= -k \ln\left(1 + \sqrt{\frac{q}{np}} x\right) \sim -(np + \sqrt{npq}x) \times \\ &\quad \times \left[\sqrt{\frac{q}{np}} x - \frac{1}{2} \frac{q}{np} x^2\right]; \end{aligned}$$

$$\begin{aligned} \ln\left(\frac{n-k}{nq}\right)^{-(n-k)} &= -(n-k) \ln\left(1 - \sqrt{\frac{p}{nq}} x\right) \sim \\ &\sim -(nq - \sqrt{npq}x) \left[-\sqrt{\frac{p}{nq}} x - \frac{1}{2} \frac{p}{nq} x^2\right] \end{aligned}$$

Sumando estas expresiones llegamos a la siguiente relación:

$$\lim_{n \rightarrow \infty} \ln \left[\left(\frac{np}{k} \right)^k \left(\frac{nq}{n-k} \right)^{n-k} \right] = -\frac{1}{2} x^2,$$

y, por consiguiente,

$$\lim_{n \rightarrow \infty} \left(\frac{np}{k} \right)^k \left(\frac{nq}{n-k} \right)^{n-k} = e^{-x^2/2}.$$

Luego,

$$\sqrt{\frac{n}{k(n-k)}} \sim \sqrt{\frac{n}{npnq}} = \frac{1}{\sqrt{npq}} = \Delta x,$$

y como resultado obtenemos, que

$$P\{S_n^* = x\} \sim \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \cdot \Delta x, \quad \Delta x = \frac{1}{\sqrt{npq}}$$

siendo uniforme por todos los $x = \frac{k-np}{\sqrt{npq}}$, $x' \leq x \leq x''$. Por consiguiente,

$$P\{S_n^* = x\} = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \cdot \Delta x + o(\Delta x)$$

uniforme por todos los $x = \frac{k-np}{\sqrt{npq}}$, $x' \leq x \leq x''$ y para $n \rightarrow \infty$

$$\begin{aligned} P\{x' \leq S_n^* \leq x''\} &= \sum_{x'}^{x''} P\{S_n^* = x\} = \frac{1}{\sqrt{2\pi}} \times \\ &\times \sum_{x'}^{x''} [e^{-x^2/2} \Delta x + o(\Delta x)] = \frac{1}{\sqrt{2\pi}} \times \\ &\times \sum_{x'}^{x''} e^{-x^2/2} \Delta x + o(1) \sim \frac{1}{\sqrt{2\pi}} \int_{x'}^{x''} e^{-x^2/2} dx, \end{aligned}$$

lo que se exigía demostrar.

De tal modo, para la distribución de probabilidades de la magnitud aleatoria S_n^* tenemos la siguiente aproximación (normal):

$$P\{x' \leq S_n^* \leq x''\} \approx \Phi(x'') - \Phi(x'),$$

donde $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-u^2/2} du$ es la función de distribución de probabilidades normal (con la densidad (3.7)) (fig. 14). Una aproximación algo más exacta de la fórmula¹⁾ es

$$P\{x' \leq S_n^* \leq x''\} \approx \Phi\left(x'' + \frac{\Delta x}{2}\right) - \Phi\left(x' - \frac{\Delta x}{2}\right),$$

¹⁾ Véase, por ejemplo, el libro citado anteriormente de V. Feller.

donde $\Delta x = \frac{1}{\sqrt{npq}}$. Para la diferencia de las funciones de distribuciones $F(x) = P\{S_n^* \leq x\}$ y $\Phi(x)$ se tiene la siguiente valuación (véase más adelante (4.19)):

$$\sup_x |F(x) - \Phi(x)| \leq C \frac{1}{\sqrt{n}}, \quad (3.8)$$

donde $C \leq \frac{p^2 + q^2}{\sqrt{pq}}$.

Para comodidad de los lectores más adelante se expone la tabla

2 de la función de distribución normal $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-u^2/2} du$.

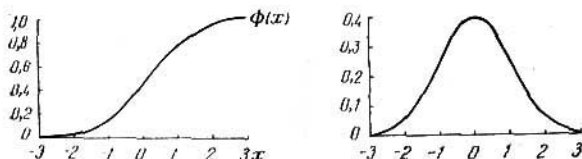


Fig. 14. Vista general de la función de distribución normal de probabilidades $\Phi(x)$ con los parámetros $(0,1)$ y de la densidad de probabilidad respectiva.

2. Proceso del movimiento browniano. Distribución de probabilidades del máximo y momento de su consecución. Imaginémos a una partícula suspendida en un líquido homogéneo. Ella sufre unos choques caóticos con las moléculas del líquido en resultado de lo cual se encuentra en un movimiento continuamente desordenado.

De modelo análogo discreto de tal proceso puede servir el siguiente modelo de la fluctuación aleatoria. La partícula cambia su posición sólo en momentos discretos de tiempo, múltiplos de Δt . El cambio de posición ocurre de tal forma que encontrándose en el punto x , la partícula independientemente del comportamiento anterior pase con las mismas probabilidades a uno de los puntos vecinos $x + \Delta x$ ó $x - \Delta x$, siendo una misma la desviación Δx para todos los puntos x (se trata sólo de una de las coordenadas de la partícula, en el caso de la fluctuación aleatoria unidimensional). En el límite, cuando de un modo determinado $\Delta t \rightarrow 0$, $\Delta x \rightarrow 0$ se obtiene la fluctuación aleatoria c o n t i n u a, para el proceso físico del movimiento browniano²⁾.

²⁾ Véase, por ejemplo, el libro de A. Ya. Jinchin «Leyes asintóticas de la teoría de probabilidades» ONTI. 1936 (cap. 3).

Tabla 2

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0,0	0,500 000	1,5	0,933 193	3,0	0,998 650
0,1	0,539 828	1,6	0,945 201	3,1	0,999 032
0,2	0,579 260	1,7	0,955 435	3,2	0,999 313
0,3	0,617 911	1,8	0,964 070	3,3	0,999 517
0,4	0,655 422	1,9	0,971 283	3,4	0,999 663
0,5	0,691 462	2,0	0,977 250	3,5	0,999 767
0,6	0,725 747	2,1	0,982 136	3,6	0,999 841
0,7	0,758 036	2,2	0,986 097	3,7	0,999 892
0,8	0,788 145	2,3	0,989 276	3,8	0,999 928
0,9	0,815 940	2,4	0,991 802	3,9	0,999 952
1,0	0,841 345	2,5	0,993 790	4,0	0,999 968
1,1	0,864 334	2,6	0,995 339	4,1	0,999 979
1,2	0,884 930	2,7	0,996 533	4,2	0,999 987
1,3	0,903 200	2,8	0,997 445	4,3	0,999 991
1,4	0,919 243	2,9	0,998 134	4,4	0,999 995
				4,5	0,999 997

Designemos por $\xi(t)$ la posición de la partícula browniana en el momento de tiempo t . Supongamos que en el momento inicial de tiempo $t = 0$ la partícula se encuentra en el punto $x = 0$.

Durante la fluctuación discreta en el tiempo t ella realiza $n = \frac{t}{\Delta t}$ pasos, entre los cuales se realiza un número aleatorio determinado de pasos en dirección positiva. Si designamos por S_n al número de pasos en dirección positiva, entonces el desplazamiento total en dirección positiva será $S_n \Delta x$ y en dirección negativa, $(n - S_n) \Delta x$. De tal modo, el desplazamiento total $\xi(t)$ en el tiempo $t = n \Delta t$ está ligado con el número S_n por la siguiente igualdad:

$$\xi(t) = |S_n \Delta x - (n - S_n) \Delta x| = (2S_n - n) \Delta x.$$

Si consideramos que $\xi(0) = 0$, entonces

$$\xi(s+t) = |\xi(s) - \xi(0)| + |\xi(t+s) - \xi(s)|$$

para cualesquiera $s, t \geq 0$. Evidentemente, en el modelo de la fluctuación aleatoria descrito, las magnitudes $\xi(s) - \xi(0)$ y $\xi(t+s) - \xi(s)$ son independientes, siendo la distribución de probabilidades del incremento $\xi(t+s) - \xi(s)$ exactamente igual que la del incremento $\xi(t) - \xi(0)$. Por eso, para la dispersión $D\xi(t+s)$

con cualesquiera $s, t \geq 0$, tiene lugar la igualdad

$$\begin{aligned} D\xi(t+s) &= D|\xi(s) - \xi(0)| + D|\xi(t+s) - \xi(s)| = \\ &= D\xi(s) + D\xi(t). \end{aligned}$$

Se ve que la dispersión $D\xi(t)$ como función de t , varía linealmente con el crecimiento de t y de tal modo

$$D\xi(t) = \sigma^2 \cdot t, \quad 0 \leq t \leq \infty,$$

donde σ^2 es una determinada constante llamada *coeficiente de difusión*. Por otra parte, se puede calcular, fácilmente, que la dispersión del desplazamiento durante el tiempo t (o de otra forma, durante n pasos, $n = \frac{t}{\Delta t}$) es $D\xi(t) = (\Delta x)^2 \frac{t}{\Delta t}$. Como resultado obtenemos la siguiente relación entre Δx y Δt :

$$\frac{(\Delta x)^2}{\Delta t} = \sigma^2.$$

Los pasos realizados por la partícula no dependen uno de otro, y los podemos considerar como pruebas de Bernoulli, con la probabilidad del «éxito» $p = 1/2$, que la tomamos como un paso en dirección positiva. Entonces S_n , que es el número de pasos en dirección positiva, será igual al número de «éxitos» en las n pruebas de Bernoulli. Con esto, la posición de la partícula en el momento t estará ligada a la magnitud normalizada $S_n^* = \frac{1}{\sqrt{n}}(2S_n - n)$ del siguiente modo:

$$\xi(t) = S_n^* \sqrt{n} \Delta x = S_n^* \sqrt{t} \frac{\Delta x}{\sqrt{\Delta t}} = S_n^* \sigma \sqrt{t}.$$

Utilizando el teorema de Moivre—Laplace (véase la relación (3.6)), obtenemos que la distribución de probabilidades de la magnitud aleatoria $\xi(t)$ en el límite se da por la fórmula

$$P \left\{ x' \leq \frac{\xi(t)}{\sigma \sqrt{t}} \leq x'' \right\} = \lim_{\Delta t \rightarrow 0} P \{ x' \leq S_n^* \leq x'' \} = \frac{1}{\sqrt{2\pi}} \int_{x'}^{x''} e^{-x^2/2} dx. \quad (3.9)$$

La fórmula (3.9) da no sólo la distribución de probabilidades de la magnitud $\xi(t)$, que es la posición de la partícula browniana en el momento t , sino también la de cualquier desplazamiento de la partícula que fluctúa aleatoriamente en el tiempo t , ya que debido a la homogeneidad del proceso examinado, para cualquier valor de $\xi(s)$, el incremento $\xi(t+s) - \xi(s)$ tiene la misma distribución de probabilidades que el incremento $\xi(t) - \xi(0) = \xi(t)$, a saber:

$$P \left\{ x' \leq \frac{\xi(t+s) - \xi(s)}{\sigma \sqrt{t}} \leq x'' \right\} = \frac{1}{\sqrt{2\pi}} \int_{x'}^{x''} e^{-x^2/2} dx.$$

En la fig. 15¹⁾ están representadas algunas trayectorias experimentales del movimiento browniano.

El modelado del movimiento browniano se puede realizar con ayuda de la llamada «tabla de Galton». Este dispositivo representa en sí una tabla lisa con clavitos dispuestos simétricamente; se lanza desde arriba una bolita («partícula browniana») que en su movimiento hacia abajo tropieza con los clavitos («moléculas») y se desplaza cada

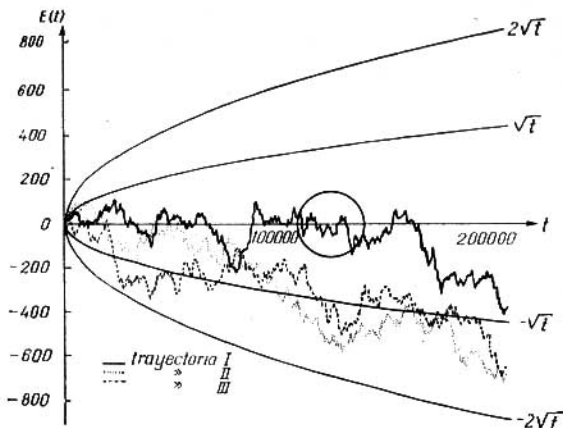


Fig. 15. Trayectorias experimentales del movimiento browniano con coeficiente de difusión $\sigma = 1$.

vez con la misma probabilidad en una distancia Δ a la izquierda o a la derecha; después de n choques (n es igual al número de clavitos según la vertical) la bolita ocupa una determinada posición ξ sobre la recta horizontal OX , cayendo en el intervalo $[x', x'']$ con la probabilidad

$$P\{x' \leq \xi \leq x''\} \approx \frac{1}{\sqrt{2\pi}} \int_{x'}^{x''} e^{-x^2/2} dx$$

(consideramos que en la escala elegida $\Delta \sqrt{n} = 1$); de hecho la bolita cae en una de las celdas que dividen al eje OX en pequeños intervalos $[x_h, x_{h+1}]$ de igual longitud Δx .

Si tomamos un número N de bolitas, suficientemente grande, entonces de acuerdo a la ley de los grandes números, el número N_h

¹⁾ Véase H. O. A. Wold. (Editor), Bibliography on time series and stochastic processes, Oliver and Boyd, Edinburgh and London, 1965, (págs. 10-11).

de aquellas que caen en la celda k $[x_k, x_{k+1}]$, deberá ser tal que $\frac{N_k}{N} \approx \approx \mathbf{P} \{x_k \leq \xi \leq x_{k+1}\} \approx \frac{1}{\sqrt{2\pi}} e^{-x_k^2/2} \Delta x$.

En el modelo discreto, para cualquier valor fijado $\xi(s) = a$, el movimiento de la partícula browniana a partir del momento s (cuando ella se encuentra en el punto a) no depende de su comportamiento anterior a este momento. Suponiendo que esta propiedad también se conserva en el proceso continuo límite del movimiento browniano, hallamos la distribución de probabilidades de la magnitud aleatoria τ_a , o sea, el momento en que por primera vez la partícula browniana alcanza el punto $x = a$ (considerando, como anteriormente, que la partícula se encuentra en el momento inicial $t = 0$ en el punto $x = 0$).

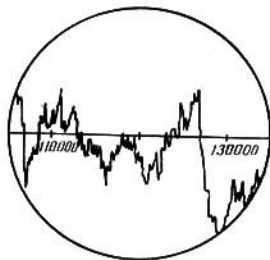


Fig. 15a. Parte de la fig. 15 aumentada 12 veces.

Está claro, que la partícula browniana durante el movimiento en dirección positiva, obedece a la misma ley, que durante el movimiento en dirección negativa (en el modelo discreto la partícula se mueve a cada paso a la derecha o a la izquierda con la misma probabilidad).

Por eso, las magnitudes τ_a y τ_{-a} (momentos de alcanzar a los puntos a y $-a$), al salir del punto inicial $x = 0$, tienen la misma distribución de probabilidades. Consideramos que $a > 0$ y hallaremos la probabilidad $\mathbf{P} \{\tau_a \leq t\}$.

La partícula puede encontrarse en el momento t más a la derecha del punto a , sólo con la condición de que, en determinado momento $\tau_a \leq t$, se encuentre en este punto (ya que durante el movimiento browniano continuo, la partícula no puede «saltar» por encima de a). Esto significa formalmente que el suceso $\{\xi(t) \geq a\}$ está contenido en el suceso $\{\tau_a \leq t\}$ y, por consiguiente,

$$\mathbf{P} \{\xi(t) \geq a | \tau_a \leq t\} = \frac{\mathbf{P} \{\xi(t) \geq a\}}{\mathbf{P} \{\tau_a \leq t\}}.$$

Evidentemente, la probabilidad condicional $\mathbf{P} \{\xi(t) \geq a | \tau_a \leq t\}$, a condición de que la partícula se encuentre en a en el momento τ_a ($\tau_a \leq t$), coincide con la probabilidad de que ella se encuentre más a la derecha de a en el momento t después de salir de a . Pero, de las consideraciones de simetría expuestas anteriormente, se deduce que la probabilidad de encontrarse en el momento t , más a la derecha del punto de partida a , es la misma que la probabilidad de encontrarse en este momento, más a la izquierda de a , e igual a $1/2$. De tal forma,

$P\{\xi(t) \geq a \mid \tau_a \leq t\} = 1/2$ y considerando, para sencillez, al coeficiente de difusión σ^2 igual a 1, de la igualdad obtenida antes y de la fórmula (3.9) obtenemos la función de distribución de la magnitud τ_a :

$$F_{\tau_a}(t) = P\{\tau_a \leq t\} = 2P\{\xi(t) \geq a\} = \sqrt{\frac{2}{\pi}} \int_{at^{-1/2}}^{\infty} e^{-x^2/2} dx, \quad t > 0.$$

Diferenciando por t la función de distribución, hallamos la densidad de probabilidad correspondiente:

$$p_{\tau_a}(t) = \frac{a}{\sqrt{2\pi}} t^{-3/2} e^{-a^2/2t} \quad 0 \leq t < \infty \quad (3.10)$$

($p_{\tau_a}(t) = 0$ para $t < 0$).

Es interesante observar que para cualquier punto a la magnitud τ_a es finita con la probabilidad 1:

$$P\{\tau_a < \infty\} = \lim_{t \rightarrow \infty} P\{\tau_a \leq t\} = \sqrt{\frac{2}{\pi}} \int_0^{\infty} e^{-x^2/2} dx = 1,$$

es decir, la partícula browniana cae más tarde o más pronto en cualquier punto a (en un determinado momento aleatorio de tiempo $\tau_a < \infty$).

Conociendo la distribución de la magnitud τ_x , o sea, el momento de alcanzar el punto x , se puede hallar también, inmediatamente, la distribución de probabilidades de la magnitud de *máximo desplazamiento* de la partícula browniana en el tiempo fijado t . Evidentemente,

$$\begin{aligned} P\{\max_{0 \leq s \leq t} \xi(s) \geq x\} &= P\{\tau_x \leq t\} = \sqrt{\frac{2}{\pi}} \int_{xt^{-1/2}}^{\infty} e^{-u^2/2} du = \\ &= \sqrt{\frac{2}{\pi t}} \int_x^{\infty} e^{-u^2/2t} du \end{aligned}$$

y, por consiguiente, la función de distribución de la magnitud $\xi = \max_{0 \leq s \leq t} \xi(s)$ será

$$F_{\xi}(x) = 1 - P\{\max_{0 \leq s \leq t} \xi(s) \geq x\} = \sqrt{\frac{2}{\pi t}} \int_0^x e^{-u^2/2t} du$$

(recordemos que $\sqrt{\frac{2}{\pi t}} \int_0^{\infty} e^{-u^2/2t} dt = \frac{2}{\sqrt{2\pi}} \int_0^{\infty} e^{-u^2/2} du = 1$). Se ve, que la magnitud $\xi = \max_{0 \leq s \leq t} \xi(s)$ tiene la densidad de probabilidad

$$p_{\xi}(x) = \sqrt{\frac{2}{\pi t}} e^{-x^2/2t}, \quad 0 \leq x < \infty \quad (3.11)$$

($p_{\bar{\xi}}(x) = 0$ para $x < 0$, ya que $\xi \geq \xi(0) = 0$). Esta es la llamada *ley normal doble* de distribución de probabilidades (como se ve fácilmente, $\mathbf{P}\{\xi \geq x\} = 2\mathbf{P}\{\xi(t) \geq x\}$).

Evidentemente, la magnitud $\min_{0 \leq s \leq t} \xi(s)$ tiene una distribución de probabilidades análoga, y precisamente, su densidad de probabilidad será

$$p(x) = \sqrt{\frac{2}{\pi t}} e^{-x^2/2t}, \quad -\infty < x \leq 0$$

$p(x) = 0$ para $x > 0$). Es interesante indicar que

$$\mathbf{P}\{\max_{0 \leq s \leq t} \xi(s) > 0\} = \mathbf{P}\{\min_{0 \leq s \leq t} \xi(s) < 0\} = 1$$

y por lo tanto, al salir del punto $x = 0$, la partícula browniana, en cualquier tiempo t , tan pequeño como se quiera, estará tanto más a la derecha del punto inicial $x = 0$ como más a la izquierda de este punto. Examinando la trayectoria de la partícula browniana en el tiempo (es decir, el gráfico de la función $\xi = \xi(t)$, $t \geq 0$), de aquí se puede deducir que esta trayectoria continua en un intervalo de tiempo $(0, t)$ tan pequeño como se quiera, corta infinitas veces al nivel $x = 0$, tomando un número infinito de veces valores tanto positivos como negativos (en otras palabras, la partícula browniana retorna un número infinito de veces al punto inicial $x = 0$).

Suponiendo que es continua, la trayectoria de la partícula browniana $\xi(u)$, $0 \leq u \leq t$, alcanza su máximo absoluto en un punto determinado τ , $0 \leq \tau \leq t$ (tendremos en consideración el primer punto del máximo, si son varios). Hallemos la distribución de la magnitud aleatoria τ .

Supongamos que se tiene la densidad de la distribución conjunta de probabilidades de las magnitudes aleatorias τ y $\xi = \xi(\tau)$ ($\xi = \max_{0 \leq u \leq t} \xi(u)$). Mostremos que entonces, la densidad tiene la forma

$$p_{\tau, \xi}(s, x) = \frac{1}{\pi \sqrt{s(t-s)}} \frac{x}{s} e^{-x^2/2s}, \quad (3.12)$$

$$0 < s < t, \quad 0 \leq x < \infty.$$

Para ello, examinemos al principio la distribución conjunta de probabilidades de las magnitudes aleatorias τ_a y ξ , donde τ_a , como anteriormente, significa el momento en que la partícula browniana alcanza por primera vez el punto $a > 0$.

Después de caer la partícula browniana en el punto a su comportamiento ulterior se somete a las mismas leyes, como si este punto fuese el inicial desde el propio comienzo. Por eso la magnitud $\xi = \max_{0 \leq u \leq t} \xi(u)$, que para la condición $\tau_a = s$, $0 < s \leq t$, coincide con la magnitud $\max_{s \leq u \leq t} \xi(u)$, tiene, para la condición indicada,

la misma distribución de probabilidades que la magnitud $a + \max_{0 \leq u \leq t-s} \xi(u)$, y de acuerdo con la fórmula establecida anteriormente (3.11), tiene la densidad convencional de distribución

$$p_{\xi}(x|s) = \sqrt{\frac{2}{\pi(t-s)}} e^{-(x-a)^2/2(t-s)}, \\ a \leq x < \infty.$$

De aquí se deduce, que la densidad $p_{\tau_a, \xi}(s, x)$ de la distribución conjunta de probabilidades de las magnitudes τ_a, ξ para $0 < s < t$, $a \leq x < \infty$ tiene la forma (véase (3.17), cap. I):

$$p_{\tau_a, \xi}(s, x) = p_{\tau_a}(s) p_{\xi}(x|s) = \frac{1}{\pi \sqrt{s(t-s)}} \frac{a}{s} e^{-a^2/2s} e^{-(x-a)^2/2(t-s)}.$$

Por otra parte, para la condición $\xi = \max_{0 \leq u \leq t} \xi(u) = a$ el punto τ del máximo coincide con el momento τ_a , y por consiguiente, para la condición indicada, las magnitudes τ, τ_a tienen la misma distribución de probabilidades. De aquí se desprende que la densidad de probabilidad de las magnitudes (τ, ξ) en el punto $\tau = s, \xi = a$, coincide con la densidad de probabilidad de las magnitudes (τ_a, ξ) en el mismo punto (s, a) ya que

$$p_{\tau, \xi}(s, a) = p_{\tau}(s|a) p_{\xi}(a) = p_{\tau_a}(s|a) p_{\xi}(a) = p_{\tau_a, \xi}(s, a),$$

donde $p_{\tau}(s|x)$ y $p_{\tau_a}(s|x)$ representan a las densidades convencionales de las distribuciones de probabilidades de las magnitudes τ y τ_a , para la condición $\xi = x$. De la fórmula hallada anteriormente para la densidad $p_{\tau_a, \xi}(x, s)$, cuando $x = a$, obtenemos

$$p_{\tau, \xi}(s, a) = \frac{1}{\pi \sqrt{s(t-s)}} \frac{a}{s} e^{-a^2/2s}, \quad 0 < s < t, \quad 0 < a < \infty,$$

lo que nos da la fórmula indicada anteriormente (3.12). La densidad de una magnitud τ tomada por separado, o sea del punto máximo de la trayectoria browniana $\xi(s)$ en el intervalo de tiempo $0 \leq s \leq t$ será

$$p_{\tau}(s) = \int_0^{\infty} p_{\tau, \xi}(s, x) dx = \frac{1}{\pi \sqrt{s(t-s)}} \int_0^{\infty} \frac{x}{s} e^{-x^2/2s} dx = \\ = \frac{1}{\pi \sqrt{s(t-s)}}, \quad 0 < s < t.$$

La función de distribución de probabilidades correspondiente tiene la forma

$$F_{\tau}(s) = \int_0^s \frac{du}{\pi \sqrt{u(t-u)}} = \frac{2}{\pi} \arcsen \sqrt{\frac{s}{t}}, \quad 0 \leq s \leq t.$$

Esta ley de distribución de probabilidades lleva el nombre de *ley del arco seno*. La misma distribución de probabilidades tiene también, evidentemente, el punto del mínimo de la trayectoria $\xi(s)$, $0 \leq s \leq t$.

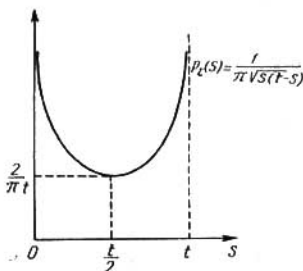


Fig. 16.

Se ve (fig. 16) que el comportamiento más probable de la partícula browniana es aquel, para el cual el punto extremal de su trayectoria está situado en la proximidad de los extremos del intervalo examinado $[0, t]$.

§ 4. DISTRIBUCION NORMAL DE PROBABILIDADES Y DISTRIBUCIONES LIGADAS A ELLA

1. Distribución normal multidimensional. Examinemos la *distribución normal* de probabilidades con la densidad

$$p_0(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx, \quad -\infty < x < \infty. \quad (4.1)$$

La densidad de probabilidad $p_0(x)$ es una función par, y por eso

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-x^2/2} dx = 0$$

(es decir, el valor medio es igual a cero). Para calcular la dispersión

$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-x^2/2} dx$, recurrimos a la "integral de probabilidad"

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} dx = 1.$$

Pongamos $x\sqrt{u}$ en lugar de x , para todos los valores $u > 0$ tenemos

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-ux^2/2} dx \equiv \frac{1}{\sqrt{u}}.$$

Diferenciando por el parámetro u en el punto $u = 1$ obtenemos que la dispersión es

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-x^2/2} dx = 1.$$

Sea ξ_0 la magnitud aleatoria, que tiene la distribución normal de probabilidades (4.1); como acabamos de mostrar

$$M\xi_0 = 0, \quad D\xi_0 = 1.$$

Examinemos la magnitud aleatoria $\xi = \sigma\xi_0 + a$, $\sigma > 0$. Está claro que

$$M\xi = a, \quad D\xi = \sigma^2, \quad (4.2)$$

y la densidad de la distribución de probabilidades de la magnitud ξ será

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-a)^2/2\sigma^2}, \quad -\infty < x < \infty. \quad (4.3)$$

La distribución de probabilidades con tal densidad se llama *distribución normal* (o de Gauss); se determina con dos parámetros (a , σ), donde el parámetro a es el valor medio:

$$a = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} x e^{-(x-a)^2/2\sigma^2} dx, \quad (4.4)$$

y el parámetro σ , así llamada *desviación standard* (normalizada), es la raíz positiva de la dispersión

$$\sigma^2 = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (x-a)^2 e^{-(x-a)^2/2\sigma^2} dx. \quad (4.5)$$

Para los valores de los parámetros $a = 0$ y $\sigma = 1$, la distribución (4.3) corrientemente se llama distribución de Gauss standard.

Examinemos n magnitudes aleatorias independientes $\xi_{01}, \dots, \xi_{0n}$, que tienen una misma distribución de Gauss (standard). Su densidad conjunta de probabilidades es

$$p_0(x_1, \dots, x_n) = \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2} \sum_{k=1}^n x_k^2}, \quad -\infty < x_1, \dots, x_n < \infty. \quad (4.6)$$

Sea

$$\xi_i = \sum_{h=1}^n \sigma_{ih} \xi_{0h} + a_i, \quad i = 1, \dots, n, \quad (4.7)$$

una transformación lineal no degenerada de las magnitudes ξ_{0h} , $h = 1, \dots, n$. Evidentemente,

$$M\xi_i = a_i, \quad M(\xi_i - a_i)(\xi_j - a_j) = \sum_{h=1}^n \sigma_{ih} \sigma_{jh}, \quad i, j = 1, \dots, n, \quad (4.8)$$

ya que las magnitudes independientes ξ_{0h} , $h = 1, \dots, n$, con los valores medios nulos son no correlacionados:

$$M\xi_{0h} = 0, \quad M(\xi_{0h} \xi_{0l}) = M\xi_{0h} \cdot M\xi_{0l} = 0 \text{ para } h \neq l, \quad h, l = 1, \dots, n.$$

Recordemos que la matriz $R = \{R_{ij}\}$ con los elementos

$$R_{ij} = M(\xi_i - a_i)(\xi_j - a_j), \quad i, j = 1, \dots, n,$$

se llama *matriz de correlación* de las magnitudes aleatorias ξ_1, \dots

\dots, ξ_n ($r_{ij} = \frac{R_{ij}}{\sqrt{R_{ii}R_{jj}}}$ es el coeficiente de correlación entre las

magnitudes ξ_i, ξ_j). Las fórmulas (4.8) muestran que la matriz de correlación R en el caso examinado tiene la forma

$$R = \sigma \sigma^*, \quad (4.9)$$

donde $\sigma = \{\sigma_{ij}\}$ es la matriz de la transformación lineal (4.7) y σ^* es la matriz conjugada con σ (con los elementos $\sigma_{ij}^* = \sigma_{ji}$, $i, j = 1, \dots, n$). Los determinantes $|R|$ y $|\sigma|$ de las matrices indicadas están ligados por la igualdad $|R| = |\sigma|^2$; siendo $|\sigma|$ el jacobiano de la transformación (4.7) y, por consiguiente, la densidad conjunta de distribución de las magnitudes aleatorias ξ_1, \dots, ξ_n puede ser descrita por la fórmula

$$p(x_1, \dots, x_n) = \frac{1}{(2\pi)^{n/2} |R|} \exp \left\{ -\frac{1}{2} \sum_{i,j=1}^n b_{ij} (x_i - a_i)(x_j - a_j) \right\}, \quad (4.10)$$

donde $\sum_{i,j=1}^n b_{ij} (x_i - a_i)(x_j - a_j)$ es aquella forma cuadrática en la

que se transforma la suma de los cuadrados $\sum_{h=1}^n x_h^2$ durante la transformación lineal inversa a (4.7). Se ve fácilmente, que la matriz $\{b_{ij}\}$ de esta forma cuadrática es la inversa de la matriz de correlación $\{R_{ij}\}$, ya que la suma de los cuadrados $\sum_{h=1}^n x_h^2$ se transforma

en la forma cuadrática

$$\sum_{h=1}^n \left[\sum_{j=1}^n c_{hj} (x_j - a_j) \right]^2 = \sum_{i,j=1}^n \left(\sum_{h=1}^n c_{hi} c_{hj} \right) (x_i - a_i) (x_j - a_j),$$

donde $\{c_{hj}\}$ es la matriz inversa a $\{\sigma_{ij}\}$, de modo que

$$\{b_{ij}\} = \{c_{ih}\}^* \{c_{hj}\} = \{ \{\sigma_{ih}\} \{\sigma_{hj}\}^* \}^{-1} = \{R_{ij}\}^{-1}.$$

Como ya se indicó anteriormente, la matriz de correlación $\{R_{ij}\}$ está determinada positivamente: para cualesquiera valores reales c_1, \dots, \dots, c_n

$$\sum_{i,j=1}^n R_{ij} c_i c_j = \mathbf{M} \left[\sum_{j=1}^n c_j (\xi_j - a_j) \right]^2 \geq 0;$$

la matriz inversa $\{b_{ij}\} = R^{-1}$ posee esta misma propiedad. Para cualquier matriz no degenerada, determinada positivamente, $\{b_{ij}\}$,

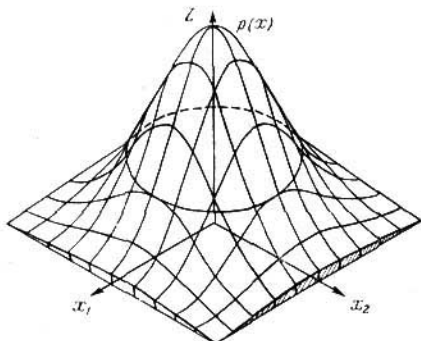


Fig. 17. Vista general de la distribución normal bidimensional (de la densidad de probabilidades $p(x)$); en la figura están indicadas las líneas de intersección de la superficie $z = p(x)$ con los distintos planos.

la fórmula (4.10), donde $|R|^{-1}$ significa el determinante de la matriz $\{b_{ij}\}$ y a_1, \dots, a_n son constantes arbitrarias, da una determinada densidad de probabilidad (esta forma tiene, precisamente, la densidad de probabilidad de las magnitudes ξ_1, \dots, ξ_n , ligadas con las magnitudes de Gauss standards $\xi_{01}, \dots, \xi_{0n}$ a través de la transformación lineal (4.7) con matriz $\{\sigma_{hj}\}$ que es la raíz cuadrada de la matriz $R = \{b_{ij}\}^{-1}$).

La distribución de probabilidades con la densidad (4.10), igualmente que las magnitudes aleatorias ξ_1, \dots, ξ_n con tal distribución de probabilidades, se llama *distribución normal* (o *de Gauss*) (fig. 17).

Evidentemente, durante cualquier transformación lineal no degenerada, las magnitudes de Gauss ξ_1, \dots, ξ_n se transforman en las magnitudes de Gauss η_1, \dots, η_n : si

$$\eta_k = \sum_{j=1}^n c_{kj} \xi_j, \quad k=1, \dots, n,$$

entonces la densidad de la probabilidad de la magnitud η_1, \dots, η_n se da por la fórmula del tipo (4.10), a saber:

$$p_{\eta_1, \dots, \eta_n}(y_1, \dots, y_n) = \frac{1}{(2\pi)^{n/2} |\tilde{R}|^{1/2}} \times \\ \times \exp \left\{ -\frac{1}{2} \sum_{i,j=1}^n \tilde{b}_{ij} (y_i - \tilde{a}_i) (y_j - \tilde{a}_j) \right\}, \quad (4.11)$$

donde

$$\tilde{a}_i = \sum_{j=1}^n c_{ij} a_j, \quad i=1, \dots, n, \\ \{\tilde{b}_{ij}\} = \{c_{ki}\}^{-1} \{b_{kl}\} \{c_{lj}\}^{-1}$$

y $|\tilde{R}|$ es el determinante de la matriz $\tilde{R} = \{\tilde{b}_{ij}\}^{-1}$ (véase la fórmula (1.7) para la transformación de la densidad de probabilidades, cap. I).

Examinemos las magnitudes aleatorias de Gauss $\xi_0, \xi_1, \dots, \xi_n$ y hallemos la distribución condicional de probabilidades de la magnitud ξ_0 cuando están fijados $\xi_1, \xi_2, \dots, \xi_n$.

Consideremos la magnitud $\hat{\xi}_0 = a_0 + \sum_{k=1}^n \hat{c}_k (\xi_k - a_k)$, donde $a_k = M\xi_k$, $k=0, 1, \dots, n$, y los coeficientes $\hat{c}_1, \dots, \hat{c}_n$ satisfacen al sistema de ecuaciones lineales

$$\sum_{k=1}^n R_{kj} \hat{c}_k = R_{0j}, \quad j=1, \dots, n \quad (4.12)$$

($\hat{\xi}_0$ es la proyección de la magnitud ξ_0 en el espacio de todas las magnitudes del tipo $\sum_{k=1}^n c_k \xi_k$, véase (4.32) cap. I).

La transformación lineal no degenerada

$$\eta_0 = \hat{\xi}_0 - \xi_0, \quad \eta_k = \xi_k - a_k, \quad k=1, \dots, n,$$

nos da las magnitudes de Gauss $\eta_0, \eta_1, \dots, \eta_n$ tales que $M\eta_0\eta_k = 0$ para todos los $k=1, \dots, n$; su matriz de correlación tiene la forma

$$\tilde{R} = \begin{pmatrix} \sigma^2 & 0 \\ 0 & R \end{pmatrix},$$

donde $\sigma^2 = M\eta_0^2$ y $R = \{R_{ij}\}$ es la matriz de correlación de las magnitudes η_1, \dots, η_n . Se puede comprobar fácilmente que la matriz inversa a \hat{R} será

$$\hat{b} = \begin{pmatrix} \sigma^{-2} & 0 \\ 0 & R^{-1} \end{pmatrix}$$

y, por consiguiente, la densidad de probabilidad de las magnitudes de Gauss $\eta_0, \eta_1, \dots, \eta_n$ tiene la forma

$$p_{\eta_0, \eta_1, \dots, \eta_n}(y, y_1, \dots, y_n) = \frac{1}{\sqrt{2\pi}\sigma} e^{-y^2/2\sigma^2} \times \\ \times \frac{1}{(2\pi)^{(n-1)/2} |R|^{1/2}} \exp \left\{ -\frac{1}{2} \sum_{i,j=1}^n b_{ij} y_i y_j \right\}. \quad (4.13)$$

Como se observa con facilidad ésta descompone en el producto de la densidad de probabilidad

$$p_{\eta_0}(y) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} p_{\eta_0, \eta_1, \dots, \eta_n}(y, y_1, \dots, y_n) dy_1 \dots dy_n = \\ = \frac{1}{\sqrt{2\pi}\sigma} e^{-y^2/2\sigma^2} \quad (4.14)$$

de la magnitud η_0 y la densidad de probabilidad

$$p_{\eta_1, \dots, \eta_n}(y_1, \dots, y_n) = \int_{-\infty}^{\infty} p_{\eta_0, \eta_1, \dots, \eta_n}(y, y_1, \dots, y_n) dy = \\ = \frac{1}{(2\pi)^{(n-1)/2} |R|^{1/2}} \exp \left\{ -\frac{1}{2} \sum_{i,j=1}^n b_{ij} y_i y_j \right\} \quad (4.15)$$

de las magnitudes η_1, \dots, η_n . Esto significa que la magnitud η_0 no depende de η_1, \dots, η_n . Por esto la distribución condicional de probabilidades de esta magnitud (para cualesquiera valores fijados η_1, \dots, η_n) tiene una misma densidad normal $p_{\eta_0}(y)$ como la indicada anteriormente.

A nosotros nos interesa la magnitud aleatoria $\xi_0 = \eta_0 + \hat{\xi}_0$, donde $\hat{\xi}_0 = a_0 + \sum_{k=1}^n \hat{c}_k \eta_k$ y $\eta_k = \xi_k - a_k$, $k = 1, \dots, n$. Evidentemente, para los valores fijados $\xi_1 = x_1, \dots, \xi_n = x_n$, la magnitud ξ_0 se diferencia de η_0 por el sumando constante $\hat{x}_0 = a_0 + \sum_{k=1}^n \hat{c}_k (x_k - a_k)$ y, por consiguiente, tiene la densidad condicional de probabilidad

$$p_{\xi_0}(x | x_1, \dots, x_n) = p_{\eta_0}(x - \hat{x}_0) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x - \hat{x}_0)^2/2\sigma^2}, \quad -\infty < x < \infty. \quad (4.16)$$

Esta es la densidad normal de probabilidad que responde al valor medio

$$\begin{aligned}\hat{x}_0 &= \mathbf{M}(\xi_0 | x_1, \dots, x_n) = \int_{-\infty}^{\infty} x p_{\xi_0}(x | x_1, \dots, x_n) dx = \\ &= a_0 + \sum_{h=1}^n \hat{c}_h (x_h - a_h) \text{ y a la dispersión} \\ \sigma^2 &= \mathbf{M}[\xi_0 - a_0 - \sum_{h=1}^n \hat{c}_h (\xi_h - a_h)]^2.\end{aligned}$$

Así pues, hemos hallado la distribución condicional de la magnitud ξ_0 para los valores fijados ξ_1, \dots, ξ_n . De paso hemos establecido efectivamente los siguientes hechos: *si las magnitudes de Gauss no están correlacionadas, entonces son independientes; si la magnitud multidimensional $(\xi_0, \xi_1, \dots, \xi_n)$ tiene una distribución normal, entonces la magnitud multidimensional (ξ_1, \dots, ξ_n) de un número menor de dimensiones, tendrá una distribución de probabilidades del mismo tipo, y en particular, las componentes ξ_1, \dots, ξ_n tomadas por separado tienen una distribución normal de probabilidades del tipo (4.3) compárese con (4.13)—(4.15).*

Nos podemos imaginar fácilmente, que en las aplicaciones frecuentemente tratamos magnitudes aleatorias compuestas de un gran número de componentes independientes. Resulta que para unas condiciones amplias tales magnitudes tienen una distribución normal de probabilidades.

Este hecho fundamental se expresa matemáticamente en la forma de tal o cual (en dependencia de las condiciones) *teorema límite central* (que es un caso muy particular del teorema de Moivre—Laplace). De ejemplo, puede servir el teorema límite central para la suma $S_n = \sum_{k=1}^n \xi_k$ de sumandos independientes ξ_1, ξ_2, \dots , que satisfacen a la condición de Liapunov:

$$\frac{1}{B_n^3} \sum_{k=1}^n \mathbf{M}|\xi_k - a_k|^3 \rightarrow 0 \text{ para } n \rightarrow \infty \quad (4.17)$$

($a_k = \mathbf{M}\xi_k$, $k = 1, 2, \dots$; $B_n = \sqrt{DS_n}$); precisamente, con la condición (4.17) para la distribución normal de probabilidades de la suma normalizada

$$S_n^* = \frac{S_n - \mathbf{M}S_n}{\sqrt{DS_n}}$$

tiene lugar la relación límite siguiente¹⁾:

$$\lim_{n \rightarrow \infty} \mathbf{P} \{x' \leq S_n^* \leq x''\} = \frac{1}{\sqrt{2\pi}} \int_{x'}^{x''} e^{-x^2/2} dx. \quad (4.18)$$

Se puede precisar que si las magnitudes aleatorias ξ_1, ξ_2, \dots tienen la misma distribución de probabilidades, entonces para la función de distribución $F_n(x) = \mathbf{P} \{S_n^* \leq x\}$ de la magnitud S_n^* es justa la siguiente valuación:

$$\sup_{-\infty < x < \infty} |F_n(x) - \Phi(x)| \leq C_0 \frac{\mu}{\sigma^3} \frac{1}{\sqrt{n}}, \quad (4.19)$$

donde $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-u^2/2} du$ es la función de distribución normal y C_0 es una constante²⁾

$$(\mu = \mathbf{M} |\xi_k - a|^3, \quad a = \mathbf{M} \xi_k, \quad \sigma = \sqrt{D \xi_k} \quad k = 1, 2, \dots).$$

2. Valuación de los parámetros de la distribución normal. La distribución χ^2 y la distribución de Student. La densidad normal de probabilidad se determina por dos parámetros: por el valor medio de a y la desviación standard (varianza o dispersión) σ :

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-a)^2/2\sigma^2}, \quad -\infty < x < \infty$$

(véase (4.3) — (4.5)).

Supongamos que tratamos la magnitud aleatoria distribuida según la ley normal, cuyos parámetros (a, σ) son desconocidos. Se pueden obtener las valuaciones estadísticas de estos parámetros. Precisamente, si ξ_1, \dots, ξ_n son magnitudes de Gauss independientes con una media a y la dispersión σ , entonces en calidad de valuación para la esperanza matemática desconocida a (según los valores de elección de que disponemos ξ_1, \dots, ξ_n) se puede tomar la «media empírica»

$$\hat{a} = \frac{1}{n} \sum_{k=1}^n \xi_k \quad (4.20)$$

y para la dispersión σ^2 , la magnitud

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{k=1}^n (\xi_k - \hat{a})^2. \quad (4.21)$$

¹⁾ Véase más adelante el § 5.

²⁾ Está calculado que $C_0 < 1$.

El valor medio empírico $\hat{a} = \frac{1}{n} \sum_{k=1}^n \xi_k$, siendo una combinación

lineal de magnitudes de Gauss independientes ξ_1, \dots, ξ_n , tiene la distribución normal de probabilidades con los parámetros a , $\frac{\sigma}{\sqrt{n}}$ ($a = M\hat{a}$, $\frac{\sigma^2}{n} = D\hat{a}$).

Para hallar la distribución de probabilidades de «la dispersión empírica» $\hat{\sigma}^2 = \frac{1}{n} \sum_{k=1}^n (\xi_k - \hat{a})^2$ nos dirigimos a la identidad

$$n\hat{\sigma}^2 = \sum_{j=1}^n \xi_j^2 - n\hat{a}^2 = \sum_{j=1}^n \xi_j^2 - \eta^2,$$

donde

$$\eta = \frac{\xi_1}{\sqrt{n}} + \dots + \frac{\xi_n}{\sqrt{n}}.$$

Utilizando cualquier transformación ortogonal de las variables ξ_1, \dots, ξ_n

$$\eta_k = \sum_{j=1}^n c_{kj} \xi_j, \quad k = 1, \dots, n, \quad (4.22)$$

en la cual $\eta_n = \frac{\xi_1}{\sqrt{n}} + \dots + \frac{\xi_n}{\sqrt{n}}$ obtenemos que

$$n\hat{\sigma}^2 = \sum_{j=1}^n \xi_j^2 - \eta^2 = \sum_{k=1}^n \eta_k^2 - \eta_n^2 = \sum_{k=1}^{n-1} \eta_k^2,$$

ya que en la transformación ortogonal la suma de los cuadrados $\sum_{j=1}^n \xi_j^2$ se transforma en suma de los cuadrados $\sum_{k=1}^n \eta_k^2$.

Evidentemente, «la dispersión empírica»

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{k=1}^n \left(\xi_k - \frac{1}{n} \sum_{j=1}^n \xi_j \right)^2$$

no se cambia al pasar de las magnitudes ξ_1, \dots, ξ_n a las $\xi_1 - a, \dots, \xi_n - a$, de modo que sin limitar la generalidad, se puede considerar el valor medio a igual a cero. Entonces, la densidad conjunta de probabilidad de las magnitudes de Gauss independientes ξ_1, \dots, ξ_n será

$$p(x_1, \dots, x_n) = \frac{1}{(2\pi)^{n/2} \sigma^n} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{j=1}^n x_j^2 \right\},$$

y la densidad conjunta de probabilidad de las magnitudes η_1, \dots, η_n , que se obtiene con la transformación ortogonal (4.22), es

$$p(y_1, \dots, y_n) = \frac{1}{(2\pi)^{n/2} \sigma^n} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{k=1}^n y_k^2 \right\}$$

(véase la fórmula de transformación de la densidad (4.11)). Se ve, que η_1, \dots, η_n son magnitudes de Gauss independientes entre sí y,

por consiguiente, $n\hat{\sigma}^2 = \sum_{k=1}^{n-1} \eta_k^2$ es la suma de los cuadrados de las $n-1$ magnitudes de Gauss independientes entre sí η_k con los parámetros $(0, \sigma)$. También vemos que las *valuaciones examinadas*

$$\hat{a} = \frac{1}{n} \eta_n \quad y \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{k=1}^{n-1} \eta_k^2 \quad (4.23)$$

son *magnitudes independientes*.

La distribución de probabilidades de la magnitud aleatoria

$$\chi^2 = \sum_{k=1}^n \xi_{0k}^2, \quad (4.24)$$

donde $\xi_{01}, \dots, \xi_{0n}$ son las magnitudes de Gauss independientes *standards*, lleva el nombre de *ji-cuadrado distribución*; por n se designa el número de grados de libertad. Esta distribución tiene la densidad de la forma ¹⁾

$$p(x) = \frac{1}{2^{n/2} \Gamma\left(\frac{n}{2}\right)} x^{n/2-1} e^{-x/2}, \quad 0 \leq x < \infty, \quad (4.25)$$

y es un caso particular de la distribución Γ (véase (2.17)) (fig. 18).

Señalemos que de la fórmula (4.25) se puede obtener, fácilmente, la expresión para la densidad de probabilidad de la magnitud aleatoria

$\chi = \sqrt{\sum_{k=1}^n \xi_{0k}^2}$, que es igual al módulo del vector $(\xi_{01}, \dots, \xi_{0n})$:

$$p_\chi(x) = 2xp(x^2) = \frac{1}{2^{(n-1)/2} \Gamma\left(\frac{n}{2}\right)} x^{n/2} e^{-x^2/2}, \quad 0 \leq x < \infty; \quad (4.26)$$

para $n=2$ esta es la tal llamada *distribución de Rayleigh*, para $n=3$ es la *distribución de Maxwell*.

Así pues, para apreciar los valores \hat{a} , $\hat{\sigma}$ de los parámetros a , σ de la distribución normal (véase (4.23)) hemos establecido que \hat{a}

¹⁾ La fórmula (4.25) será deducida en el siguiente § 5, véase (5.21).

tiene una distribución normal de probabilidades con los parámetros $(a, \frac{\sigma}{\sqrt{n}})$ y $n \frac{\hat{\sigma}^2}{\sigma^2}$ tiene la χ^2 -distribución con $n-1$ grados de libertad, siendo \hat{a} y $\hat{\sigma}$ magnitudes aleatorias independientes.

Si el parámetro σ es conocido, entonces para cualquier α tan pequeño como se quiera se puede indicar Δ de modo tal que el valor desco-

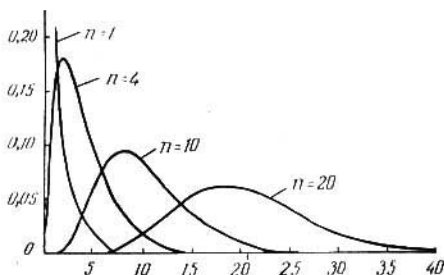


Fig. 18. Vista general de las densidades de probabilidad χ^2 de la distribución con n grados de libertad ($n = 1, 4, 10, 20$).

nocido del parámetro a se hallará dentro de los límites garantizados

$$\hat{a} - \frac{\sigma}{\sqrt{n}} \Delta \leq a \leq \hat{a} + \frac{\sigma}{\sqrt{n}} \Delta \quad (4.27)$$

con una probabilidad no menor de $1 - \alpha$, donde Δ se determina según α de la relación

$$P \left\{ \left| \frac{\hat{a} - a}{\sigma} \sqrt{n} \right| \leq \Delta \right\} = 2\Phi(\Delta) - 1 = 1 - \alpha$$

(Δ puede ser hallado en la tabla de la función normal de distribución $\Phi(x)$).

Si son desconocidos los dos parámetros (a, σ), entonces los límites garantizados para el valor medio de a , pueden ser establecidos sobre la base de la distribución que satisface a la relación

$$\tau = \sqrt{n-1} (\hat{a} - a) / \hat{\sigma} = \eta_n \left/ \sqrt{\frac{1}{n-1} \sum_{k=1}^{n-1} \eta_k^2} \right.,$$

donde η_1, \dots, η_n son las magnitudes independientes de Gauss con los parámetros $(0, \sigma)$.

Está claro que la magnitud $\tau = \eta_n \left/ \sqrt{\frac{1}{n-1} \sum_{k=1}^{n-1} \eta_k^2} \right.$ no depende

de σ , y su distribución de probabilidades coincide con la distribución de la relación $\xi_{0n} / \sqrt{\frac{1}{n-1} \sum_{j=1}^{n-1} \xi_{0j}^2}$ para las magnitudes de Gauss "standards" independientes $\xi_{01} = \frac{\eta_1}{\sigma}, \dots, \xi_{0n} = \frac{\eta_n}{\sigma}$. Esta distribución que se denomina *distribución de Student* (o *distribución t*)

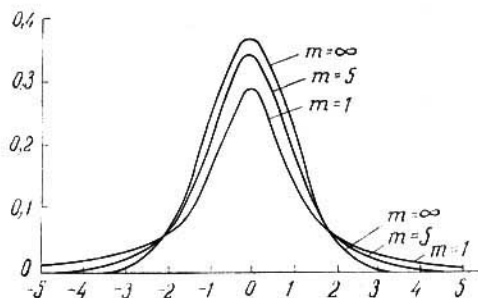


Fig. 19. Vista general de las densidades de probabilidad de la distribución *t* de Studente con los parámetros $m = 1, 5$; a $m = \infty$ le corresponde la densidad de la distribución de probabilidades normal standard.

(fig. 19), se determina por el parámetro $m = n - 1$ y, como se indicará más adelante, tiene una densidad de la forma

$$p(x) = \frac{1}{\sqrt{\pi m}} \frac{\Gamma\left(\frac{m+1}{2}\right)}{\Gamma\left(\frac{m}{2}\right)} \left(1 + \frac{x^2}{m}\right)^{-(m+1)/2}, \quad 0 \leq x < \infty. \quad (4.28)$$

Igualmente que durante la determinación del «intervalo confidencial» (4.27), se puede hallar el valor correspondiente Δ para el cual

$$P \left\{ \left| \sqrt{n-1} \frac{\hat{a}-a}{\hat{\sigma}} \right| \leq \Delta \right\} = P \{ |\tau| \leq \Delta \} = 1 - \alpha,$$

y, por consiguiente, con la probabilidad $1 - \alpha$ se puede garantizar que

$$\hat{a} - \frac{\hat{\sigma}}{\sqrt{n-1}} \Delta \leq a \leq \hat{a} + \frac{\hat{\sigma}}{\sqrt{n-1}} \Delta. \quad (4.29)$$

En conclusión deduzcamos la fórmula (4.28) para la densidad de la distribución *t* (la densidad de la probabilidad de la relación $\eta =$

= ξ_1/ξ_2 de las magnitudes de Gauss independientes ξ_1 con los parámetros (0, 1) y de la magnitud $\xi_2 = \sqrt{\frac{1}{m} \chi^2}$, donde la magnitud χ^2 tiene la χ^2 -distribución con m grados de libertad).

De la fórmula (4.26) hallamos fácilmente que la densidad $p_2(x)$ de la magnitud ξ_2 es

$$p_2(x) = \frac{m^{m/2}}{2^{m/2-1} \Gamma\left(\frac{m}{2}\right)} x^{m-1} e^{-mx^2/2}, \quad x > 0.$$

Para la densidad de probabilidad de la relación $\eta = \xi_1/\xi_2$, teniendo en cuenta que $p_2(x) = 0$ para $x < 0$, tendremos la siguiente expresión:

$$p_\eta(y) = \int_0^\infty p_1(yx) p_2(x) x dx$$

(véase (3.14), cap. I). Poniendo aquí la expresión hallada anteriormente, para $p_2(x)$ y

$$p_1(yx) = \frac{1}{\sqrt{2\pi}} e^{-y^2 x^2/2}, \quad -\infty < x < \infty,$$

obtenemos que

$$p_\eta(y) = \int_0^\infty \frac{m^{m/2}}{\sqrt{\pi} 2^{m/2} \Gamma\left(\frac{m}{2}\right)} x^{m-1} e^{-\frac{m}{2} x^2 \left(1 + \frac{y^2}{m}\right)} dx.$$

Haciendo en la integral el cambio de variable $u = \frac{m}{2} x^2 \times \left(1 + \frac{y^2}{m}\right)$, llegamos a la expresión

$$p_\eta(y) = \frac{1}{\sqrt{\pi m} \Gamma\left(\frac{m}{2}\right)} \left(1 + \frac{y^2}{m}\right)^{-(m+1)/2} \int_0^\infty u^{(m-1)/2} e^{-u} du,$$

en el cual

$$\int_0^\infty u^{(m-1)/2} e^{-u} du = \Gamma\left(\frac{m+1}{2}\right)$$

es el valor de la función Γ en el punto $(m+1)/2$. Hemos obtenido la fórmula (4.28);

§ 5. DISTRIBUCION
DE PROBABILIDADES
Y FUNCIONES
CARACTERISTICAS

1. **Funciones características y sus propiedades fundamentales.** La función

$$\varphi(t) = Me^{it\xi}, \quad -\infty < t < \infty, \quad (5.1)$$

cuyo valor en el punto t coincide con la esperanza matemática $Me^{it\xi}$, donde ξ es la magnitud aleatoria real, se denomina *función característica* de esta magnitud ξ (o de la correspondiente distribución de probabilidades con la función de distribución $F(x) = P\{\xi \leq x\}$).

Supongamos que la distribución de probabilidades está concentrada en los puntos enteros $k = 0, \pm 1, \pm 2, \dots$ (es decir, $\sum_{k=-\infty}^{\infty} P(k) = 1$, donde $P(k) = P\{\xi = k\}$, $k = 0, \pm 1, \pm 2, \dots$), entonces la función característica

$$\varphi(t) = \sum_{k=-\infty}^{\infty} e^{itk} P(k) \quad (5.2)$$

es periódica (con período 2π) y la fórmula (5.2) da su descomposición en serie de Fourier con los coeficientes

$$P(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-ikt} \varphi(t) dt, \quad k = 0, \pm 1, \pm 2, \dots \quad (5.3)$$

Supongamos que la distribución de probabilidades tiene la densidad $p(x)$, entonces la función característica es

$$\varphi(t) = \int_{-\infty}^{\infty} e^{itx} p(x) dx \quad (5.4)$$

y representa en sí la integral de Fourier de la función $p(x)$, $-\infty < x < \infty$, y para la función absolutamente integrable $\varphi(t) = \left(\int_{-\infty}^{\infty} |\varphi(t)| dt < \infty \right)$ la densidad $p(x)$ puede ser obtenida con ayuda de la transformación inversa de Fourier:

$$p(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ixt} \varphi(t) dt, \quad -\infty < x < \infty. \quad (5.5)$$

Exponemos la fórmula de inversión general¹⁾, que liga la función característica $\varphi(t)$ y la función de distribución correspondiente $F(x)$:

$$F(x'') - F(x') = \lim_{\sigma \rightarrow 0} \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-ix''t} - e^{-ix't}}{-it} e^{-\sigma^2 t^2 / 2} \varphi(t) dt \quad (5.6)$$

para cualesquiera puntos x' y x'' en los cuales la función de distribución $F(x)$ es continua (es decir, $\mathbf{P}\{\xi = x'\} = \mathbf{P}\{\xi = x''\} = 0$).

Señalemos aquí que se tiene no más de n valores diferentes de x , para los cuales $\mathbf{P}\{\xi = x\} \geq \frac{1}{n}$, y esto significa, que se tiene un número finito y numerable de valores de x , para los cuales $\mathbf{P}\{\xi = x\} > 0$; es evidente que para cualquier x' , x'' se hallarán sucesiones de puntos $x_n \rightarrow x' - 0$, $x_n'' \rightarrow x'' + 0$ que satisfacen a la condición $\mathbf{P}\{\xi = x_n'\} = \mathbf{P}\{\xi = x_n''\} = 0$, y debido a la continuidad de la probabilidad,

$$\lim_{n \rightarrow \infty} \mathbf{P}\{x_n' \leq \xi \leq x_n''\} = \mathbf{P}\{x' \leq \xi \leq x''\} = F(x'') - F(x' - 0), \quad (5.7)$$

lo que da, junto con la fórmula de inversión, el siguiente resultado: la distribución de probabilidades se determina unívocamente por la función característica.

Examinemos el desarrollo

$$e^{it\xi} = \sum_{k=0}^n i^k \frac{\xi^k}{k!} t^k + \theta \frac{\xi^{n+1}}{(n+1)!} t^{n+1},$$

donde θ es una magnitud limitada, $|\theta| \leq 1$. Si se tienen los momentos finitos $\mu_k = \mathbf{M}|\xi|^k$, $k = 1, \dots, n+1$, entonces, evidentemente,

$$\mathbf{M}e^{it\xi} = \sum_{k=0}^n i^k \frac{\mathbf{M}\xi^k}{k!} t^k + \frac{\mathbf{M}(\theta\xi^{n+1})}{(n+1)!} t^{n+1},$$

y de tal modo, la función característica $\varphi(t)$ admite el desarrollo

$$\varphi(t) = \sum_{k=0}^n i^k \frac{\mu_k}{k!} t^k + \frac{R_n}{(n+1)!} t^{n+1}, \quad -\infty < t < \infty, \quad (5.8)$$

donde $|R_n| \leq \mathbf{M}|\xi|^{n+1}$ y $\mu_0 = \varphi(0) = 1$,

$$\mu_k = i^{-k} \varphi^{(k)}(0), \quad k = 1, \dots, n.$$

La fórmula (5.8) permite calcular los momentos $\mu_k = \mathbf{M}\xi^k$ según las derivadas de la función característica de la magnitud aleatoria ξ .

¹⁾ Semejantes fórmulas de inversión serán examinadas más adelante, véase la pág. 130.

Examinemos algunos ejemplos.

Ejemplo (distribución uniforme). Para densidad de probabilidad

$$\rho(x) = \begin{cases} \frac{1}{b-a} & \text{para } a \leq x \leq b \\ 0 & \text{para } x < a, x > b \end{cases}$$

la función característica es

$$\varphi(t) = \frac{1}{b-a} \int_a^b e^{itx} dx = \frac{e^{ibt} - e^{iat}}{(b-a)it}, \\ -\infty < t < \infty.$$

Ejemplo (distribución de Poisson). Hallemos la función característica de la función de distribución de Poisson

$$P(k) = \frac{a^k}{k!} e^{-a}, \quad k = 0, 1, \dots$$

Tenemos

$$\varphi(t) = \sum_{k=0}^{\infty} e^{itk} P(k) = e^{-a} \sum_{k=0}^{\infty} \frac{(ae^{it})^k}{k!} = \\ = e^{-a} e^{ae^{it}} = e^{a(e^{it}-1)}, \quad -\infty < t < \infty. \quad (5.9)$$

Ejemplo (distribución normal). Hallemos la función característica de la distribución normal con la densidad

$$\rho(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad -\infty < x < \infty.$$

Para el parámetro real z tenemos

$$\int_{-\infty}^{\infty} e^{zx} \rho(x) dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{zx-x^2/2} dx = e^{z^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(x+z)^2/2} dx = \\ = e^{z^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} dx = e^{z^2/2},$$

de modo que la integral $\int_{-\infty}^{\infty} e^{zx} \rho(x) dx = e^{z^2/2}$ es una función analítica de z que se prolonga unívocamente por todo el plano complejo de la variable z . Colocando el valor it en lugar de z obtenemos, que la función característica buscada es

$$\varphi(t) = \int_{-\infty}^{\infty} e^{itx} \rho(x) dx = e^{-t^2/2}, \quad -\infty < t < \infty.$$

Señalemos que en la transformación lineal $\xi = \sigma\xi_0 + a$ la función característica $\varphi(t)$ de la magnitud ξ se obtiene de la función característica $\varphi_0(t)$ de la magnitud ξ_0 según la fórmula

$$\varphi(t) = M e^{it(\sigma\xi_0 + a)} = e^{ita} \varphi_0(\sigma t), \quad (5.10)$$

de modo que la función característica de la magnitud de Gauss ξ con el valor medio a y la dispersión σ^2 ($\xi = \sigma\xi_0 + a$, donde ξ_0 es la magnitud de Gauss con los parámetros (0,1)) es

$$\varphi(t) = e^{ita - \sigma^2 t^2 / 2}, \quad -\infty < t < \infty. \quad (5.11)$$

Ejemplo (distribución simétrica exponencial y distribución de Cauchy). Examinemos la distribución de probabilidades con la densidad

$$p(x) = \frac{a}{2} e^{-a|x|}, \quad -\infty < x < \infty \quad (5.12)$$

(a es un determinado parámetro positivo). La función característica es

$$\begin{aligned} \varphi(t) &= \frac{a}{2} \int_0^{\infty} (e^{itx} + e^{-itx}) e^{-ax} dx = \frac{a}{2} \int_0^{\infty} e^{(a+it)x} dx + \\ &+ \frac{a}{2} \int_0^{\infty} e^{(a-it)x} dx = \frac{a}{2} \left(\frac{1}{a+it} - \frac{1}{a-it} \right) = \frac{a^2}{a^2 + t^2}, \quad -\infty < t < \infty. \end{aligned} \quad (5.13)$$

Indiquemos, que para $a = 1$ la función característica (5.13) se diferencia, de la densidad de probabilidad sólo por un factor positivo,

$$p(x) = \frac{1}{\pi} (1 + x^2)^{-1}, \quad -\infty < x < \infty, \quad (5.14)$$

que se obtiene de la fórmula general (4.28) para la distribución de Student cuando $m = 1$. Esta densidad nos da la distribución de probabilidades de la relación de dos magnitudes de Gauss independientes con parámetros (0,1) y también la distribución de probabilidades de la magnitud $\operatorname{tg} \xi$, donde ξ es la magnitud aleatoria, distribuida uniformemente en el segmento $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ y otras; la distribución con tal densidad de probabilidad se llama *distribución de Cauchy*.

Comparando a (5.12)—(5.14) y utilizando la fórmula de inversión (5.5), obtenemos para la función característica $\varphi(t)$ de la distribución de Cauchy, la expresión

$$\begin{aligned} \varphi(t) &= \frac{1}{\pi} \int_{-\infty}^{\infty} e^{itx} (1 + x^2)^{-1} dx = 2 \left[\frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itu} (1 + u^2)^{-1} du \right] = \\ &= e^{-|t|}, \quad -\infty < t < \infty, \end{aligned} \quad (5.15)$$

Indiquemos una propiedad importante de las funciones características: *la suma de las magnitudes aleatorias independientes* $\xi = \xi_1 + \dots + \xi_n$ *tiene la función característica de la forma*

$$\varphi(t) = \varphi_1(t) \dots \varphi_n(t), \quad (5.16)$$

igual al producto de las funciones características $\varphi_1(t), \dots, \varphi_n(t)$ de los sumandos aislados ξ_1, \dots, ξ_n , que es un corolario de la propiedad general de las esperanzas matemáticas

$$M e^{it\xi} = M(e^{it\xi_1} \dots e^{it\xi_n}) = (M e^{it\xi_1} \dots M e^{it\xi_n})$$

(véase la fórmula (4.6), cap. I).

Ejemplo (distribución de triángulo). Hallemos la función característica de la distribución de probabilidades con densidad

$$p(x) = \begin{cases} 1-|x| & \text{para } |x| \leq 1, \\ 0 & \text{para } |x| > 1, \end{cases}$$

que coincide con la densidad de probabilidad de la suma $\xi = \xi_1 + \xi_2$ de las magnitudes aleatorias independientes ξ_1 y ξ_2 distribuidas uniformemente en los segmentos $[-1, 0]$ y $[0, 1]$ y que tienen las funciones características

$$\varphi_1(t) = \frac{1-e^{-it}}{it} \quad \text{y} \quad \varphi_2(t) = \frac{e^{it}-1}{it}.$$

De acuerdo a la fórmula (5.16) la función característica buscada es

$$\varphi(t) = \varphi_1(t)\varphi_2(t) = \frac{|1-e^{it}|^2}{t^2} = \frac{2(1-\cos t)}{t^2}, \quad -\infty < t < \infty. \quad (5.17)$$

Ejemplo (distribución binomial). Hallemos la función característica de la distribución binomial o sea la distribución de la suma $\xi = \xi_1 + \dots + \xi_n$ de magnitudes independientes igualmente distribuidas de la forma

$$\xi_k = \begin{cases} 1 & \text{con probabilidad } p, \\ 0 & \text{con probabilidad } 1-p. \end{cases}$$

La función característica de cada sumando aislado es igual a $(1-p) + e^{it}p$, y por consiguiente, la función característica buscada es

$$\varphi(t) = [1 + p(e^{it} - 1)]^n, \quad -\infty < t < \infty. \quad (5.18)$$

Ejemplo (χ^2 -distribución). Hallemos la función característica de la magnitud aleatoria

$$\chi^2 = \xi_1^2 + \dots + \xi_n^2.$$

o sea, de la suma de los cuadrados de magnitudes de Gauss independientes con parámetros (0, 1) que tienen la densidad de probabilidad la forma

$$p_0(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad -\infty < x < \infty.$$

La densidad de probabilidad de cada una de las magnitudes ξ_k tiene la forma (véase el ejemplo de la pág. 39)

$$p(y) = \frac{1}{\sqrt{2\pi y}} e^{-y/2}, \quad 0 \leq y < \infty$$

($p(y) = 0$, para $y < 0$) y la función característica de cada sumando aislado ξ_k será

$$\begin{aligned} \int_0^{\infty} e^{it y} p(y) dy &= \frac{1}{\sqrt{2\pi}} \int_0^{\infty} y^{-1/2} e^{-(1-2it)y/2} dy = \\ &= (1-2it)^{-1/2} \frac{1}{\sqrt{\pi}} \int_0^{\infty} x^{-1/2} e^{-x} dx = \\ &= (1-2it)^{-1/2} \frac{\Gamma(1/2)}{\sqrt{\pi}} = (1-2it)^{-1/2}, \end{aligned} \quad (5.19)$$

puesto que el valor de la función Γ en el punto $1/2$ es igual a $\sqrt{\pi}$. La función característica de la χ^2 -distribución con n grados de libertad será, por consiguiente,

$$\varphi(t) = (1-2it)^{-n/2}, \quad -\infty < t < \infty. \quad (5.20)$$

Para $n > 1$ la función $|\varphi(t)|$, $-\infty < t < \infty$, es integrable y la densidad de la χ^2 -distribución se podría hallar por la fórmula de inversión (5.5). Pero es más sencillo volver a las expresiones de (5.19). Se observa fácilmente, que si en lugar de $y^{-1/2}$ en la expresión inicial de la densidad $p(y)$, hubiese $y^{n/2-1}$, entonces, para el correspondiente factor normalizado D tendríamos

$$D \int_0^{\infty} e^{it y} y^{n/2-1} e^{-y/2} dy = (1-2it)^{-n/2}.$$

Ya que la función característica $\varphi(t) = (1-2it)^{-n/2}$ determina unívocamente la χ^2 -distribución correspondiente, entonces, se debe concluir que esta distribución tiene la densidad de probabilidad

$$p_n(x) = \frac{1}{2^{n/2} \Gamma\left(\frac{n}{2}\right)} x^{n/2-1} e^{-x/2}, \quad 0 \leq x < \infty, \quad (5.21)$$

donde el factor normalizador $D = \frac{1}{2^{n/2} \Gamma\left(\frac{n}{2}\right)}$ se determina de la condición

$$\int_0^{\infty} p_n(x) dx = D \int_0^{\infty} x^{n/2-1} e^{-x/2} dx = D \cdot 2^{n/2} \int_0^{\infty} y^{n/2-1} e^{-y} dy = \\ = D 2^{n/2} \Gamma\left(\frac{n}{2}\right) = 1.$$

Ejemplo (estabilidad de la distribución de Cauchy). Sea ξ_1, ξ_2, \dots , una sucesión de magnitudes aleatorias independientes que tienen la distribución de probabilidades de Cauchy. La función característica de cada una de ellas es (véase (5.15))

$$\varphi(t) = e^{-|t|}, \quad -\infty < t < \infty.$$

Examinemos la media aritmética

$$\xi = \frac{1}{n} \sum_{k=1}^n \xi_k.$$

Se ve fácilmente que la función característica $\varphi_{\xi}(t)$ de la magnitud ξ es

$$\varphi_{\xi}(t) = \varphi\left(\frac{t}{n}\right)^n = e^{-|t|},$$

y de tal modo, para cualquier n , la media aritmética $\frac{1}{n} \sum_{k=1}^n \xi_k$ tiene

la misma distribución de probabilidades, que cada una de las magnitudes ξ_k , $k = 1, 2, \dots$ (Señalamos que las magnitudes indicadas no tienen esperanza matemática; compárese con la ley de los grandes números.)

2. Convergencia de las distribuciones de probabilidades. Se dice que la distribución de probabilidades, de las magnitudes aleatorias ξ_n , $n = 1, 2, \dots$ converge débilmente hacia la distribución de probabilidades con la función de distribución $F(x)$, si

$$P\{x' \leq \xi_n \leq x''\} \rightarrow F(x'') - F(x') \quad \text{para } n \rightarrow \infty \quad (5.22)$$

para cualesquiera puntos x', x'' en los cuales $F(x)$ es continua.

Puede servir de ejemplo la convergencia débil de la distribución de sumas normalizadas $S_n^* = \frac{S_n - MS_n}{\sqrt{DS_n}}$, $n = 1, 2, \dots$, en el esquema de pruebas de Bernoulli, hacia la ley normal con función de

distribución $F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-u^2/2} du$ (véase (3.6)).

Evidentemente, para cualquier punto de continuidad de $F(x)$ y para $x' = x'' = x$ de la relación (5.22) se deduce que

$$P\{\xi_n = x\} \rightarrow 0 \quad \text{para } n \rightarrow \infty,$$

y, de tal modo, la convergencia débil también significa que

$$P\{x' < \xi_n \leq x''\} = F_{\xi_n}(x'') - F_{\xi_n}(x') \rightarrow F(x'') - F(x'). \quad (5.23)$$

(para cualesquiera puntos de continuidad de $F(x)$).

Consideremos especialmente las magnitudes de números enteros, cuyos valores posibles son los enteros $k = 0, \pm 1, \dots$. En este caso, evidentemente, la distribución límite de probabilidades deberá estar concentrada en los puntos de números enteros (es decir, $F(x) = \sum_{k \leq x} P(k)$, $\sum_k P(k) = 1$) y la relación (5.22) es equivalente a que

$$P\{\xi_n = k\} \rightarrow P(k) \quad (5.24)$$

para cada k aislado ($P(k)$, $k = 0, \pm 1, \dots$, es la distribución límite de probabilidades).

En efecto, en cualquier intervalo finito $[x', x'']$ cae, solamente, un número finito de puntos enteros, y de la relación (5.22) se deduce que

$$P\{x' \leq \xi_n \leq x''\} = \sum_{x' \leq k \leq x''} \{P_{\xi_n} = k\} \rightarrow \sum_{x' \leq k \leq x''} P(k) = F(x'') - F(x'),$$

por otra parte, $k - \varepsilon$, $k + \varepsilon$ para cualquier entero k y $0 < \varepsilon < 1$ son puntos de continuidad de la función $F(x) = \sum_{x \leq k} P(k)$, siendo

$P\{k - \varepsilon \leq \xi_n \leq k + \varepsilon\} = P\{\xi_n = k\}$ y, por consiguiente, para una convergencia débil se deberá cumplir la relación límite (5.24).

Puede servir de ejemplo la convergencia de las distribuciones de magnitudes ξ_n o sea, el número de éxitos en las n pruebas de Bernoulli, hacia la distribución de probabilidades de Poisson, cuando para $n \rightarrow \infty$ el número medio de éxitos $a = np$ queda constante:

$$P\{\xi_n = k\} \rightarrow \frac{a^k}{k!} e^{-a}, \quad k = 0, 1, \dots \quad (\text{véase (3.4)}).$$

Establezcamos algunos teoremas sobre la convergencia débil de la distribución de probabilidades.

Para mayor claridad introduzcamos la magnitud aleatoria ξ , cuya distribución de probabilidades figurará como distribución límite. Naturalmente, no se trata de ninguna convergencia de las propias magnitudes aleatorias ξ_n , $n = 1, 2, \dots$, sino que se examinará sólo la convergencia de su distribución de probabilidades.

Teorema 1. Si la sucesión de las distribuciones de probabilidades de las magnitudes ξ_n , $n = 1, 2, \dots$, converge débilmente hacia la distribución de la magnitud ξ , entonces para cualquier función continua

$$\text{limitada } u = u(x), \quad -\infty < x < \infty, \\ \lim_{n \rightarrow \infty} \mathbf{M}u(\xi_n) = \mathbf{M}u(\xi). \quad (5.25)$$

Demostración. En condiciones de la convergencia débil, las distribuciones examinadas de probabilidades son *compactas* en el sentido, de que para cualquier $\varepsilon > 0$ se encuentra un tal valor finito de a que

$$\mathbf{P}\{|\xi_n| \leq a\} > 1 - \varepsilon \quad (5.26)$$

a la vez para todos los $n = 1, 2, \dots$

En efecto, tomando un intervalo lo suficientemente grande $[-a_0, a_0]$ para el cual $\mathbf{P}\{|\xi| \leq a_0\} \geq 1 - \frac{\varepsilon}{2}$ y $\mathbf{P}\{|\xi_n| \leq a_0\} \rightarrow \mathbf{P}\{|\xi| \leq a_0\}$, obtenemos que para todos los valores suficientemente grandes de n ($n \geq n_0$) y $a \geq a_0$ se cumplirá la condición (5.26); y para cada número finito de los valores restantes $n = 1, \dots, n_0$, con la correspondiente elección de a , también se cumplirá la condición (5.26).

Tomemos una función continua limitada arbitraria $u = u(x)$, $|u(x)| \leq K$. En el intervalo finito $[-a, a]$ se la puede aproximar con la exactitud que se quiera, con las funciones «escalonadas» del tipo

$$u_\varepsilon(x) = \begin{cases} u(x_k) & \text{para } x_{k-1} < x \leq x_k \quad (k = 1, \dots, N), \\ 0 & \text{para } x < -a, \quad x > a, \end{cases}$$

donde $a = x_0 \leq x_1 \leq \dots \leq x_N = a$ son los puntos de continuidad de la función $F(x)$ tales que $|u(x) - u(x_k)| \leq \varepsilon$ para $x_{k-1} < x \leq x_k$. De la condición de compacidad (5.26) se deduce que

$$|\mathbf{M}u(\xi_n) - \mathbf{M}u_\varepsilon(\xi_n)| \leq M |u(\xi_n) - u_\varepsilon(\xi_n)| \leq \\ \leq K \mathbf{P}\{|\xi_n| > a\} + \varepsilon \leq (K + 1) \varepsilon$$

para todos los $n = 1, 2, \dots$, y análogamente,

$$|\mathbf{M}u(\xi) - \mathbf{M}u_\varepsilon(\xi)| \leq (K + 1) \varepsilon.$$

Para cualquier valor fijado N , de la relación (5.23) obtenemos que

$$\mathbf{M}u_\varepsilon(\xi_n) = \sum_{h=1}^N u(x_h) \mathbf{P}\{x_{h-1} < \xi_n \leq x_h\} \rightarrow \\ \rightarrow \sum_{h=1}^N u(x_h) \mathbf{P}\{x_{h-1} < \xi \leq x_h\} = \mathbf{M}u_\varepsilon(\xi),$$

en conclusión obtenemos, que cualquiera que sea el valor tomado de antemano $\varepsilon > 0$,

$$\lim_{n \rightarrow \infty} |\mathbf{M}u(\xi_n) - \mathbf{M}u(\xi)| \leq 2(K + 1) \varepsilon,$$

de donde se deduce la relación (5.25).

Teorema 2. Supongamos que la sucesión de magnitudes aleatorias ξ_n , $n = 1, 2, \dots$, converge en su media cuadrática hacia la mag-

itud aleatoria ξ ; entonces la distribución de probabilidades de estas magnitudes converge débilmente hacia la distribución límite de la magnitud ξ .

D e m o s t r a c i ó n. Para cualesquiera ε y δ tan pequeños como se quiera y para los valores de n suficientemente grandes

$$\mathbf{P}\{|\xi_n - \xi| > \varepsilon\} \leq \frac{M|\xi_n - \xi|^2}{\varepsilon^2} \leq \delta.$$

En estas condiciones, evidentemente,

$$\mathbf{P}\{x' + \varepsilon \leq \xi \leq x'' - \varepsilon\} - \delta \leq \mathbf{P}\{x' \leq \xi_n \leq x''\} \leq \mathbf{P}\{x' - \varepsilon \leq \xi \leq x'' + \varepsilon\} + \delta,$$

de donde, para los puntos de continuidad x' , x'' de la función de distribución $F_\xi(x)$ cuando $\varepsilon \rightarrow 0$, obtenemos

$$\mathbf{P}\{x' \leq \xi \leq x''\} - \delta \leq \mathbf{P}\{x' \leq \xi_n \leq x''\} \leq \mathbf{P}\{x' \leq \xi \leq x''\} + \delta;$$

está claro que $\mathbf{P}\{x' \leq \xi_n \leq x''\} \rightarrow \mathbf{P}\{x' \leq \xi \leq x''\}$ para $n \rightarrow \infty$.

Utilizando el teorema 2 deducimos la fórmula de inversión (5.6).

Examinemos, precisamente, la sucesión de magnitudes aleatorias $\xi_n = \xi + \Delta_n$, donde Δ_n son magnitudes aleatorias independientes de ξ que convergen en su media cuadrática hacia cero y que tienen las funciones características absolutamente integrables

$\delta_n(t)$, $\int_{-\infty}^{\infty} |\delta_n(t)| dt < \infty$ (por ejemplo, éstas pueden ser magnitudes de Gauss con los parámetros $(0, \sigma_n)$, donde $\sigma_n \rightarrow 0$; en este caso $\delta_n(t) = e^{-\sigma_n^2 t^2/2}$, véase (5.11)). Las funciones características

$\varphi_n(t) = \varphi(t) \cdot \delta_n(t)$ de las magnitudes $\xi_n = \xi + \Delta_n$, donde $\varphi(t)$ es la función característica de la magnitud que nos interesa ξ , son también integrables absolutamente, porque $|\varphi(t)| \leq 1$ y $|\varphi_n(t)| \leq |\delta_n(t)|$. Por consiguiente, las magnitudes ξ_n tienen densidades de probabilidad $\rho_n(x)$ que están ligadas con las funciones características $\varphi_n(t)$ a través de la transformación de Fourier (véase (5.5)):

$$\rho_n(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ixt} \varphi_n(t) dt.$$

Integrando por x , $x' \leq x \leq x''$ obtenemos que

$$\begin{aligned} \mathbf{P}\{x' \leq \xi_n \leq x''\} &= \int_{x'}^{x''} \rho_n(x) dx = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left[\int_{x'}^{x''} e^{-itx} dx \right] \varphi_n(t) dt = \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-ix''t} - e^{-ix't}}{it} \varphi(t) \delta_n(t) dt, \end{aligned}$$

y según el teorema 2

$$\begin{aligned} P\{x' \leq \xi \leq x''\} &= \lim_{n \rightarrow \infty} P\{x' \leq \xi_n \leq x''\} = \\ &= \lim_{n \rightarrow \infty} \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-ix''t} - e^{-ix't}}{-it} \varphi(t) \delta_n(t) dt \quad (5.27) \end{aligned}$$

para cualesquiera puntos de continuidad de la función de distribución $F(x)$ de la magnitud ξ .

La relación límite (5.27) representa en sí una fórmula de inversión, cuyo caso particular es la fórmula (5.6).

Consideremos la conocida *igualdad de Parseval* para la transformación de Fourier:

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \varphi_1(t) \varphi_2(t) dt = \int_{-\infty}^{\infty} \rho_1(x-y) \rho_2(y) dy, \quad (5.28)$$

donde $\rho_1(x)$, $\rho_2(x)$ son ciertas funciones absolutamente integrables, $\varphi_1(t)$, $\varphi_2(t)$ son sus transformaciones de Fourier (determinadas por la fórmula general (5.4)), con ello se supone que el producto $\varphi(t) = \varphi_1(t) \varphi_2(t)$, también es una función absolutamente integrable.

Detengámonos brevemente en la deducción de la igualdad (5.28). Como se puede comprobar fácilmente, $\varphi_1(t) \varphi_2(t)$ es la convolución de la transformación de Fourier

$$\begin{aligned} \rho(x) &= \int_{-\infty}^{\infty} \rho_1(x-y) \rho_2(y) dy: \\ \int_{-\infty}^{\infty} e^{itx} \left[\int_{-\infty}^{\infty} \rho_1(x-y) \rho_2(y) dy \right] dx &= \int_{-\infty}^{\infty} e^{ity} \left[\int_{-\infty}^{\infty} e^{it(x-y)} \times \right. \\ &\quad \left. \times \rho_1(x-y) dx \right] \rho_2(y) dy = \varphi_1(t) \int_{-\infty}^{\infty} e^{ity} \rho_2(y) dy = \varphi_1(t) \varphi_2(t), \end{aligned}$$

de modo que en condiciones de absoluta integrabilidad de la función $\varphi(t) = \varphi_1(t) \varphi_2(t)$, la igualdad (5.28) no es otra cosa que la fórmula (5.5) para la transformación inversa de Fourier.

Para la función real $\rho_2(x)$ su transformación de Fourier $\varphi_2(t)$ será tal que $\varphi_2(-t) = \overline{\varphi_2(t)}$ y la igualdad de Parseval se podrá obtener en la siguiente forma:

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} e^{itx} \varphi_1(t) \overline{\varphi_2(t)} dt = \int_{-\infty}^{\infty} \rho_1(x+y) \rho_2(y) dy,$$

de donde para $x=0$ y $p_1 = p_2 = p$, $\varphi_1 = \varphi_2 = \varphi$ obtenemos la llamada *identidad de Plancherel*:

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} |\varphi|^2 dt = \int_{-\infty}^{\infty} |p(x)|^2 dx \quad (5.29)$$

(se supone que

$$\int_{-\infty}^{\infty} |\varphi(t)|^2 dt < \infty).$$

De acuerdo con la identidad de Plancherel la distancia media cuadrática

$$\|p_1(x) - p_2(x)\| = \sqrt{\int_{-\infty}^{\infty} |p_1(x) - p_2(x)|^2 dx}$$

entre las funciones $p_1(x)$ y $p_2(x)$ se diferencia, sólo en un factor constante de la distancia del mismo tipo para sus transformaciones de Fourier $\varphi_1(t)$ y $\varphi_2(t)$; suponiendo, precisamente, en la fórmula (5.29) $p(x) = p_1(x) - p_2(x)$ y $\varphi(t) = \varphi_1(t) - \varphi_2(t)$ obtenemos

$$\int_{-\infty}^{\infty} |p_1(x) - p_2(x)|^2 dx = \frac{1}{2\pi} \int_{-\infty}^{\infty} |\varphi_1(t) - \varphi_2(t)|^2 dt. \quad (5.30)$$

Utilizando este hecho demostraremos la siguiente proposición.

Teorema 3. *La sucesión de distribuciones de probabilidades con las funciones características $\varphi_n(t)$, $n = 1, 2, \dots$, converge débilmente hacia la distribución con la función característica $\varphi(t)$, entonces y sólo entonces, cuando*

$$\varphi_n(t) \rightarrow \varphi(t) \quad \text{para } n \rightarrow \infty \quad (5.31)$$

uniformemente según t en cada intervalo finito.

Demostración. Supongamos que se trata de las distribuciones de probabilidades de las magnitudes aleatorias ξ_n , $n = 1, 2, \dots$, que convergen débilmente hacia la distribución de la magnitud aleatoria ξ . Refiriéndonos a la demostración del teorema 1, observamos fácilmente, que para las funciones continuas $u(x) = e^{itx}$ (donde el parámetro t cambia en un determinado intervalo finito $|t| \leq T$)

$$\varphi_n(t) = M e^{it\xi_n} \rightarrow M e^{it\xi} = \varphi(t)$$

uniformemente según t , ya que

$$|e^{itx} - e^{itx_k}| \leq C |x - x_k|, \quad k = 1, \dots, N,$$

donde la constante C depende sólo de T . Por otra parte, si se ha cumplido la condición (5.31), entonces utilizando el mismo método que fue aplicado a la deducción de la fórmula de inversión (5.27), para las magnitudes aleatorias $\eta_n = \xi_n + \Delta$ y $\eta = \xi + \Delta$ con las funciones características $\psi_n(t) = \varphi_n(t) \delta(t)$ y $\psi(t) = \varphi(t) \delta(t)$ (donde $\delta(t) = e^{-\sigma^2 t^2/2}$ es la función característica de la magnitud de Gauss Δ , independiente de ξ_n y ξ) tendremos

$$\int_{-\infty}^{\infty} |\psi_n(t) - \psi(t)|^2 dt = \int_{-\infty}^{\infty} |\varphi_n(t) - \varphi(t)|^2 \delta(t)^2 dt \rightarrow 0 \text{ para } n \rightarrow \infty,$$

ya que para cualquier $\varepsilon > 0$ se encuentra un tal T que

$$\int_{|t|>T} |\varphi_n(t) - \varphi(t)|^2 \delta(t)^2 dt \leq 2 \int_{|t|>T} \delta(t)^2 dt \leq \varepsilon,$$

y para $\varphi_n(t) \rightarrow \varphi(t)$ uniformemente según t , $|t| \leq T$,

$$\int_{-T}^T |\varphi_n(t) - \varphi(t)|^2 \delta(t)^2 dt \rightarrow 0.$$

Por consiguiente, para las densidades $p_n(x)$ y $p(x)$ de distribuciones de probabilidades de las magnitudes η_n y η

$$\int_{-\infty}^{\infty} |p_n(x) - p(x)|^2 dx = \frac{1}{2\pi} \int_{-\infty}^{\infty} |\psi_n(t) - \psi(t)|^2 dt \rightarrow 0,$$

de donde se deduce la convergencia débil de estas distribuciones: para cualquier intervalo finito $[x', x'']$

$$\mathbf{P}\{x' \leq \eta_n \leq x''\} = \int_{x'}^{x''} p_n(x) dx \rightarrow \int_{x'}^{x''} p(x) dx = \mathbf{P}\{x' \leq \eta \leq x''\}.$$

Se ve fácilmente, que también tiene lugar la convergencia débil para las distribuciones de probabilidades de las magnitudes ξ_n , $n = 1, 2, \dots$ (véase el teorema 2), ya que $\eta_n = \xi_n + \Delta$ y $\eta = \xi + \Delta$ se diferencian en su media cuadrática, de las magnitudes iniciales ξ_n y ξ todo lo pequeño que se desee $\sigma = \sqrt{M\Delta^2}$. Esto es lo que se exigía demostrar.

Demostremos en conclusión el teorema límite central en la siguiente forma. Sea

$$S_n^* = \sum_{k=1}^n \xi_{k,n}, \quad n = 1, 2, \dots,$$

una sucesión de sumas de magnitudes independientes ξ_{kn} , $k=1, \dots, n$ tales que $M\xi_{kn}=0$, $\sigma_{kn}^2=M\xi_{kn}^2 \rightarrow 0$ para $n \rightarrow \infty$ uniformemente por k) $\sum_{k=1}^n \sigma_{kn}^2=1$. Entonces, para la condición de Liapunov

$$\sum_{k=1}^n M|\xi_{kn}|^3 \rightarrow 0 \text{ para } n \rightarrow \infty, \quad (5.32)$$

las distribuciones de probabilidades de las magnitudes S_n^* , $n=1, 2, \dots$, convergen débilmente hacia la distribución normal standard, a saber:

$$\lim_{n \rightarrow \infty} P\{x' \leq S_n^* \leq x''\} = \frac{1}{\sqrt{2\pi}} \int_{x'}^{x''} e^{-x^2/2} dx$$

(compárese con (4.17), donde figuran las magnitudes $\xi_{kn} = (\xi_k - a_k)/B_n$, $k=1, 2, \dots$).

Demostremos esto apoyándonos en el teorema 3. Demostremos precisamente, que las funciones características $\varphi_n(t)$ de las magnitudes S_n^* convergen hacia la función característica $\varphi(t) = e^{-t^2/2}$ de la distribución normal standard.

Para las funciones características $\varphi_{kn}(t)$ de los sumandos independientes ξ_{kn} , tenemos según la fórmula (5.8)

$$\varphi_{kn}(t) = 1 - \frac{\sigma_{kn}^2}{2} t^2 + \frac{R_{kn}}{6} t^3,$$

donde $|R_{kn}| \leq M|\xi_{kn}|^3$. Ya que $\sigma_{kn}^2 \rightarrow 0$, $R_{kn} \rightarrow 0$ para $n \rightarrow \infty$ uniformemente por k , entonces $\varphi_{kn}(t) \rightarrow 1$ uniformemente por k y t en cada intervalo finito; en particular, $|\varphi_{kn}(t) - 1| \leq \frac{1}{2}$ para valores de n suficientemente grandes y tiene sentido hablar sobre el logaritmo de las funciones, $\varphi_{kn}(t)$. Para $\varphi_n(t) = \prod_{k=1}^n \varphi_{kn}(t)$ cuando $n \rightarrow \infty$, tenemos

$$\begin{aligned} \ln \varphi_n(t) &= \sum_{k=1}^n \ln \varphi_{kn}(t) = \sum_{k=1}^n \ln \left(1 - \frac{\sigma_{kn}^2}{2} t^2 + \frac{R_{kn}}{6} t^3 \right) \sim \\ &\sim \sum_{k=1}^n \left(-\frac{\sigma_{kn}^2}{2} t^2 + \frac{R_{kn}}{6} t^3 \right) = -\frac{t^2}{2} + \left(\sum_{k=1}^n R_{kn} \right) \frac{t^3}{6} \sim -\frac{t^2}{2} \end{aligned}$$

uniformemente por t en cada intervalo finito, ya que para la condición (5.32)

$$\left| \sum_{k=1}^n R_{kn} \right| \leq \sum_{k=1}^n M|\xi_{kn}|^3 \rightarrow 0.$$

Por consiguiente,

$$\varphi_n(t) \rightarrow e^{-t^2/2}, \text{ para } n \rightarrow \infty$$

uniformemente en cada intervalo finito, que es lo que se exigía demostrar.