V. K. ZAJÁROV
B. A. SEVASTIÁNOV
V. P. CHISTIAKOV

Teoría de las probabilidades

EDITORIAL MIR

В.К. ЗАХАРОВ
Б.А. СЕВАСТЬЯНОВ
В.П. ЧИСТЯКОВ

Теория вероятностей

ИЗДАТЕЛЬСТВО «НАУКА»
МОСКВА

V. K. ZAJÁROV
B. A. SEVASTIÁNOV
V. P. CHISTIAKOV

Teoría de las probabilidades

EDITORIAL MIR
MOSCÚ

Traducido del ruso por A. I. Samojvalov

A nuestros lectores:

Mir edita libros soviéticos traducidos al español, inglés, francés, árabe y otros idiomas extranjeros. Entre ellos figuran las mejores obras de las distintas ramas de la ciencia y la técnica: manuales para los centros de enseñanza superior y escuelas tecnológicas; literatura sobre ciencias naturales y médicas. También se incluyen monografías, libros de divulgación científica y ciencia ficción.

Dirijan sus opiniones a la Editorial Mir, 1 Rizhski per., 2, 129820, Moscú, I-110, GSP, URSS.

Impreso en la URSS

На испанском языке

Índice

Prefacio

Al escribir un curso breve de la teoría de las probabilidades para quienes se dedican a ciencias aplicadas reviste gran importancia la labor de selección. El compendio no debe contener conceptos difíciles ni teoremas que se utilizan poco o no pueden ser empleados en la práctica por los especialistas respectivos. Por otro lado, no se permite tampoco abreviar demasiado el manual, convirtiéndolo en una guía que no sea capaz de dar una idea clara acerca del contenido y los métodos de la teoría de las probabilidades.

Los autores procuraban, sobre todo, formular exactamente los modelos probabilísticos principales, mostrar las condiciones de su aplicación y los métodos de cálculo de las probabilidades y las esperanzas matemáticas. Estas cuestiones se exponen bastante detalladamente, citando los ejemplos correspondientes.

En muchos compendios de la teoría de las probabilidades el concepto de la variable aleatoria no se utiliza prácticamente y se examinan sólo los problemas, para cuya resolución no se necesita saber más que la ley de distribución. Por eso el contenido de tal curso se hace, en gran medida más pobre y la teoría de las probabilidades se reduce a problemas sueltos del análisis matemático. En el presente manual se da primeramente la definición matemática de la variable aleatoria para modelos matemáticos más simples y luego, una definición general. Se presta mucha atención a variables aleatorias que se representan en forma de las funciones de magnitudes más elementales, en particular, en forma de las sumas de indicadores.

El estudio del curso de la teoría de las probabilidades debe acompañarse, obligatoriamente, con la resolución de problemas respectivos. Al fin de cada párrafo hay problemas que los estudiantes deben resolver por sí mismos; al fin del libro se dan las respuestas.

El material tratado en el manual es suficiente, siempre que se aprenda debidamente, para resolver muchos problemas que se encuentran en la práctica. El estudio de la materia en cuestión puede llegar a ser mucho más eficaz, si la teoría de las probabilidades se utiliza sistemáticamente, y en un alto nivel profesional, en ciclos de conferencias especiales dictadas por los profesores de las cátedras formativas y si se realiza un trabajo científico conjunto por los estudiantes y profesores de las cátedras matemáticas y formativas.

§ 1. Modelos matemáticos de los fenómenos aleatorios

1.1. Modelos matemáticos. Es bien conocido el papel extraordinario que desempeñan las matemáticas en el estudio de las leyes del mundo real. El esquema de la fig. 1 muestra el lugar que ocupan las matemáticas en la investigación de los fenómenos reales. En este caso es muy importante subrayar que las matemáticas no tienen nada que ver con los mismos fenómenos reales, sino que sólo con sus modelos matemáticos. La ligazón de las matemáticas con los fenómenos del mundo que nos rodea se realiza en dos direcciones. Primeramente, abstrayéndonos de muchos hechos secundarios, construimos un modelo matemático que refleja las regularidades principales del fenómeno que se estudia. En este modelo se utilizan conceptos matemáticos y se formulan axiomas, a los cuales satisfacen estos conceptos. Luego, en el marco del modelo matemático construido, de los axiomas se deducen varios corolarios enunciados en forma de teoremas y lemas. Y, por último, nuevos hechos matemáticos obtenidos en el modelo vienen interpretados en las nociones primarias de los fenómenos reales. Esto permite comprobar la utilidad del modelo matemático y emplear en la práctica los cálculos matemáticos efectuados en el modelo. Las conclusiones prácticas serán suficientemente fiables, si el modelo construido representa los aspectos esenciales del fenómeno en estudio. En calidad de ejemplo de un modelo matemático que con éxito presta sus servicios se puede citar la mecánica construida sobre el sistema de axiomas de Newton.

1.2. Fenómenos aleatorios. El ejemplo del modelo que hemos mencionado arriba se refiere a fenómenos regulares, es decir, a tales fenómenos cuyo resultado se determina unívocamente por ciertas condiciones. Sin embargo, sabemos que para un amplio círculo de fenómenos se observan resultados no unívocos durante la repetición de una prueba, aunque se conserven condiciones principales de su realización. La no univocidad del resultado es originada por la influencia de un gran número de causas, cada una de las cuales no puede cambiarlo considerablemente. Los sucesos vinculados con tales fenómenos se llaman aleatorios. Son, por

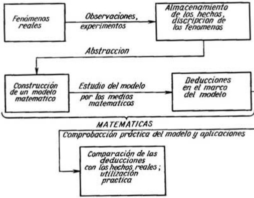

Fig. 1.

ejemplo, sucesos aleatorios la salida del "escudo" al arrojar una moneda, los resultados de mediciones, la duración de una conversación telefónica, etc. Vamos a estudiar sucesos aleatorios en masa, o sea, sucesos que aparecen como resultado de la realización de condiciones que se pueden reproducir (aunque sea en principio) muchas veces.

De la experiencia cotidiana sabemos que unos sucesos aleatorios se producen bastante frecuentemente, otros con menor frecuencia o muy raramente. Estas características de sucesos son demasiado indefinidas. Una característica experimental más objetiva de un suceso aleatorio (designémoslo, por ejemplo, con A) es la frecuencia $h_n(A)$, igual a la relación entre el número de pruebas n_A en las cuales el suceso A ha llegado y el número total de pruebas n, o sea, $h_n(A) = n_A/n$. Se ha establecido experimentalmente que para muchos sucesos la frecuencia, al aumentar n, deviene casi constante. Esta propiedad se denomina *estabilidad estadística de frecuencias* de un suceso aleatorio. Por regla general, los sucesos aleatorios masivos poseen la propiedad de estabilidad de las frecuencias. Ahora bien, con cada suceso A se puede enlazar cierto número $\mathbf{P}(A)$ al que se aproxima la frecuencia y considerar este número como la probabilidad del suceso A. Tal descripción de la probabilidad es bastante indefinida. Para que esta descripción tenga el sentido exacto, debemos construir un modelo matemático del fenómeno aleatorio. Con este fin es necesario, ante todo, dar una descripción matemática de la prueba, para cuyos resultados deseamos hallar las probabilidades.

La parte de las matemáticas, en la cual se estudian los modelos matemáticos de fenómenos aleatorios, ha recibido el nombre de teoría de las probabilidades.

1.3. Espacio de los sucesos elementales. Comencemos con examinar unos ejemplos simples.

EJEMPLO 1. *Arrojamiento de un dado una sola vez.* Como resultado de esta prueba pueden producirse diferentes sucesos: "sale el dos", "sale el seis", "el número de puntos salidos es par", etc. Vamos a distinguir sucesos elementales (no descomponibles) y sucesos compuestos (o simplemente sucesos). Por ejemplo, decir que el número de puntos salidos es par es lo mismo que decir que como resultado de la prueba sale el dos, el cuatro o el seis. Designemos por ω_k el suceso consistente en la salida de k puntos. En esta prueba los sucesos $\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6$ son sucesos elementales. Los sucesos compuestos, o simplemente sucesos, pueden ser descritos como subconjuntos del conjunto de los sucesos elementales: $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}$. Así, el suceso A = {sale un número par de puntos} se expresa por sucesos elementales del modo siguiente: $A = \{\omega_2, \omega_4, \omega_6\}$.

EJEMPLO 2. *El arrojar tres veces de una moneda.* Al tirar cada vez la moneda vamos a escribir el resultado de la prueba, designando la salida del escudo con el símbolo "1" y la de la cifra con el símbolo "0". Así, la notación 010 designará el resultado del experimento en el cual al lanzar la moneda la primera y tercera vez habrá salido la cifra y al lanzarla la segunda vez, habrá salido el escudo. El conjunto de sucesos elementales Ω se compone de 8 sucesos elementales:

$$\Omega = \{000, \ 001, \ 010, \ 011, \ 100, \ 101, \ 110, \ 111\}. \qquad (1)$$

Para denotar los sucesos elementales pueden utilizarse cualesquiera símbolos cómodos. Por ejemplo, en vez de 000, 001, ... , 111 pueden emplearse los números 0, 1, ... 7 cuya notación binaria corresponde a (1); o bien designar los sucesos elementales más formalmente: $\omega_1, \omega_2, \dots, \omega_8$. El suceso A = {en el primer arrojamiento ha salido el escudo} es, en esta prueba, compuesto: $A = \{100, 101, 110, 111\}$. Todo subconjunto del conjunto Ω se puede interpretar como cierto suceso de un experimento real. Por ejemplo, el suceso $B = \{110, 101, 011\}$ consiste en que han salido exactamente dos escudos.

EJEMPLO 3. *Tiro contra un blanco plano.* Supongamos que contra un blanco plano se produce un disparo. En esta prueba los puntos del blanco son sucesos elementales. Introduzcamos en el plano del blanco el sistema de coordenadas rectangulates uOv. Entonces el conjunto de los sucesos elementales $\omega = (u, v)$ se puede escribir en la forma

$$\Omega = \{\omega\} = \{(u, v): \quad -\infty < u < +\infty, \quad -\infty < v < +\infty\}.$$

El suceso A = {la bala ha alcanzado el círculo de radio unitario} es el subconjunto de Ω:

$$A = \{(u, v): \quad u^2 + v^2 \leqslant 1\}.$$

En todos los ejemplos examinados hemos introducido los sucesos elementales que no son más que los casos posibles de la prueba o la observación. Todos los sucesos relacionados con este experimento pueden ser descritos con ayuda de *los sucesos elementales* ω. La colección de todos los sucesos elementales Ω la llamaremos *espacio de sucesos elementales.*

Así, pues, en el caso general
llamaremos espacio de sucesos elementales a un conjunto arbitrario $\Omega = \{\omega\}$ *y los elementos* ω *de este conjunto los denominaremos sucesos elementales.*

Puesto que existe una gran variedad de fenómenos aleatorios, no es posible dar una definición general más concreta del espacio de sucesos elementales. Para la descripción de cada experimento real el conjunto Ω se elige del modo más conveniente.

1.4. Álgebra de sucesos. En los ejemplos anteriormente citados, en calidad de sucesos compuestos, o simplemente sucesos, hemos examinado los subconjuntos del conjunto Ω. Claro está que si Ω es finito o numerable,

$$\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\} \quad \text{o bien} \quad \Omega = \{\omega_1, \omega_2, \ldots, \omega_n, \ldots\},$$

entonces se llama suceso aleatorio, o simplemente suceso, a todo subconjunto del conjunto Ω. Siempre que se trate de un Ω arbitrario llamaremos

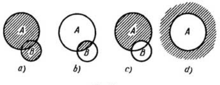

Fig. 2.

sucesos sólo los subconjuntos que forman parte de cierta clase $\mathscr{A}$ de los subconjuntos de la cual quedará determinada después de introducir las operaciones sobre sucesos coincidentes con las operaciones sobre conjuntos.

Se denomina *suma* $A + B$ (o $A \cup B$) de dos sucesos A y B al suceso que se compone de todos los sucesos elementales pertenecientes, por lo menos, a uno de los sucesos A o B. Se puede decir que en una prueba real el suceso correspondiente a $A + B$ consiste en que se ha producido, como mínimo, uno de los sucesos A o B.

Supongamos que en el ejemplo 3 los impactos del círculo mayor o menor (fig. 2) son los sucesos A y B, respectivamente. Entonces la región rayada de la fig. 2,*a* es el suceso $A + B$.

Se llama *producto* AB (o $A \cap B$) el suceso que está constituido por los sucesos elementales pertenecientes tanto a A como a B. El suceso AB se produce si y sólo si se producen A y B (fig. 2,*b*). Se denomina *diferencia* $A \setminus B$ el suceso formado por los elementos del conjunto A que no pertenecen a B (fig. 2,*c*). El suceso $A \setminus B$ consiste en que A se ha producido y B no se ha producido.

El conjunto Ω se dice *cierto*; el conjunto vacío $\varnothing$ se dice suceso *imposible*. El suceso $\bar{A} = \Omega \setminus A$ se llama opuesto al suceso A (fig. 2,*d*). El suceso $\bar{A}$ significa que A no se ha producido.

Los sucesos A y B son *incompatibles* si $AB = \varnothing$. El hecho de que A es el subconjunto de B notaremos así: $A \subset B$ (o $B \supset A$). Esto quiere decir que de la llegada del suceso A se deriva la llegada de B. En el ejemplo 3, si $A \subset B$, entonces el impacto en la región A que se contiene en B significa asimismo el impacto en B. Cuando $A \subset B$, diremos que el suceso A lleva tras de sí el suceso B o que el suceso B se deduce del suceso A. Si $A \subset B$ y $B \subset A$, diremos que los sucesos A y B son *equivalentes* y escribiremos $A = B$. La pertenencia de un elemento a un conjunto se designa con el símbolo $\in$. Por ejemplo, $\omega \in \Omega$.

Los conceptos de producto y suma de los sucesos se extienden a sucesiones infinitas de sucesos.

El suceso

$$A_1 + A_2 + \dots + A_n + \dots = \bigcup_{n=1}^{\infty} A_n$$

se compone de los sucesos elementales pertenecientes, por lo menos, a uno de los sucesos A_n, $n = 1, 2, \dots$. El suceso $\bigcap\limits_{n=1}^{\infty} A_n = A_1A_2 \dots$

... A_n ... está compuesto por los sucesos elementales pertenecientes a cada suceso A_n, $n = 1, 2, \ldots$

Para sucesos arbitrarios es fácil comprobar directamente de la definición que

$$AA = A, \quad A + A = A, \quad A\bar{A} = \varnothing.$$

Frecuentemente resultan útiles las igualdades siguientes:

$$\overline{A + B} = \bar{A}\bar{B}, \quad (A + B)C = AC + BC.$$

Vamos a demostrar, por ejemplo, la segunda igualdad. Es necesario cerciorarse de que los conjuntos que se encuentran en ambos miembros de la igualdad están constituidos por los mismos elementos. Sea ω arbitrario $\in (A + B)C$. Entonces $\omega \in A + B$ y $\omega \in C$. De $\omega \in A + B$ se deduce que ω pertenece, por lo menos, a un sumando. Sea, por ejemplo, $\omega \in A$. De $\omega \in A$ y $\omega \in C$ resulta, según la definición de producto de los sucesos, que $\omega \in AC$ y, por consiguiente, $\omega \in AC + BC$. Ahora bien, todo elemento del conjunto $(A + B)C$ es el elemento del conjunto $AC + BC$, o sea, $(A + B)C \subset AC + BC$. Al suponer que $\omega \in AC + BC$, mostremos, utilizando los razonamientos análogos a los citados más arriba, que cada elemento de $AC + BC$ es el elemento de $(A + B)C$. De aquí se deduce la igualdad que se demuestra, puesto que los conjuntos de sus miembros izquierdo y derecho están formados por los mismos elementos. Antes de proceder a la demostración de las igualdades es útil, considerando A, B y C conjuntos sobre un plano, hacer los dibujos de los conjuntos que se hallan en los miembros izquierdo y derecho de las igualdades en cuestión.

Ahora vamos a definir ciertas clases de los subconjuntos Ω. Ya hemos señalado que la probabilidad se considera como función de un suceso. Para cualesquiera subconjuntos no siempre se logra definir la probabilidad. En los casos en que la clase de subconjuntos se debe restringir vamos a suponer que como resultado de cualesquiera operaciones introducidas anteriormente vuelve a obtenerse un conjunto de la clase dada.

Sea Ω un espacio arbitrario de sucesos elementales y $\mathscr{A}$, cierta clase de los subconjuntos Ω.

La clase de los subconjuntos $\mathscr{A}$ se llama *álgebra de sucesos* si $\Omega \in \mathscr{A}$ y si $AB \in \mathscr{A}$, $A + B \in \mathscr{A}$, $A \setminus B \in \mathscr{A}$ cualesquiera que sean $A \in \mathscr{A}$ y $B \in \mathscr{A}$.

1.5. Probabilidad. Pasemos a definir la probabilidad. Notemos primeramente que no existe una definición general de la probabilidad que permita de una sola vez encontrar su valor numérico. En calidad de la definición general se enuncia una serie de axiomas a los cuales debe satisfacer la probabilidad que viene determinada para cada prueba concreta o para cada suceso aleatorio.

La función numérica P definida sobre el álgebra de sucesos $\mathscr{A}$ se llama probabilidad si se cumplen los axiomas siguientes:

A1. AXIOMA DE NO NEGATIVIDAD.

Para todo $A \in \mathscr{A}\ \mathbf{P}(A) \geqslant 0$.

A2. AXIOMA DE NORMALIZABILIDAD.

$$P(\Omega) = 1.$$

A3. AXIOMA DE ADITIVIDAD. *Si A y B son incompatibles, entonces*

$$\mathbf{P}(A + B) = \mathbf{P}(A) + \mathbf{P}(B).$$

Señalemos que los axiomas A1—A3 son absolutamente necesarios si queremos que la probabilidad $\mathbf{P}(A)$ del suceso A sea un número alrededor del cual oscilen las frecuencias $h_n(A)$ para un gran número de pruebas n, ya que las mismas frecuencias $h_n(A)$ satisfacen A1—A3. En efecto, supongamos que cierta prueba ha sido repetida n veces. Designemos con n_A el número de pruebas en las cuales se ha producido el suceso real A. Es evidente que

$$h_n(A) = \frac{n_A}{n} \geqslant 0, \quad h_n(\Omega) = \frac{n_\Omega}{n} = \frac{n}{n} = 1.$$

Si los sucesos reales A y B son incompatibles, ellos se han producido en diferentes experimentos y, por lo tanto, $n_{A+B} = n_A + n_B$. De aquí

$$h_n(A + B) = \frac{n_{A+B}}{n} = \frac{n_A}{n} + \frac{n_B}{n} = h_n(A) + h_n(B)$$

lo que corresponde a A3.

Para resolver los problemas relacionados con sucesiones infinitas de sucesos se requiere añadir a los axiomas citados el axioma siguiente.

A4. AXIOMA AMPLIADO DE ADITIVIDAD. *Si en la sucesión* $A_1, A_2, \ldots, A_n, \ldots$ *los sucesos son incompatibles dos a dos (o sea,* $A_iA_j = \varnothing$

para $i \neq j$*) y* $A = \bigcup_{n=1}^{\infty} A_n \in \mathscr{A}$, *entonces*

$$\mathbf{P}(A) = \sum_{n=1}^{\infty} \mathbf{P}(A_n).$$

El triplete $(\Omega, \mathscr{A}, \mathbf{P})$ en el cual $\mathbf{P}$ satisface A1—A4 y el conjunto $\mathscr{A}$ es no sólo el álgebra de sucesos, sino contiene también sumas y productos numerables de sucesos se llama *espacio probabilístico*.

El sistema de axiomas A1—A4 del espacio probabilístico ofrece el modelo matemático más general de los fenómenos aleatorios.

Citemos ahora algunos importantes casos particulares de espacios probabilísticos. A continuación nuevos conceptos, que se introducen, de la teoría de las probabilidades nos permitirán ampliar la colección de casos particulares y procedimientos de su construcción.

1.6. Espacio probabilístico finito. Sea $\Omega = \{\omega\}$ un conjunto finito (por ejemplo, $\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\}$, donde N es un número natural); $\{p(\omega)$: $\omega \in \Omega\}$ es la colección de números que satisfagan las condiciones

$$p(\omega) \geqslant 0, \quad \omega \in \Omega; \qquad \sum_{\omega \in \Omega} p(\omega) = 1. \tag{2}$$

Designemos con $\mathscr{A}$ el conjunto de todos los subconjuntos de Ω. Llamaremos probabilidad del suceso $A \in \mathscr{A}$ al número $\mathbf{P}(A)$ definido por la fórmula

$$\mathbf{P}(A) = \sum_{\omega \in A} p(\omega) = \sum_{l=1}^{k} p(\omega_{i_l}), \tag{3}$$

donde el suceso $A = \{\omega_{i_1}, \omega_{i_2}, \ldots, \omega_{i_k}\}$. Si $A = \varnothing$, entonces, por definición suponemos que $\mathbf{P}(A) = 0$. Los números $\{p(\omega)\}$ son las probabilidades de sucesos elementales; vamos a llamarlos simplemente *probabilidades elementales*. Ahora bien,
la probabilidad del suceso A es igual a la suma de las probabilidades elementales $p(\omega)$ en las cuales ω forman parte de A.

De la definición $\mathbf{P}(A)$ resulta directamente que se cumplen los axiomas A1, A2: $\mathbf{P}(A) \geqslant 0$, $\mathbf{P}(\Omega) = 1$. Si $AB = \varnothing$, entonces

$$\mathbf{P}(A + B) = \sum_{\omega \in A + B} p(\omega) = \sum_{\omega \in A} p(\omega) + \sum_{\omega \in B} p(\omega) = \mathbf{P}(A) + \mathbf{P}(B)$$

y, por lo tanto, el axioma A3 también se cumple.

El espacio probabilístico finito, definido por nosotros, lo llamaremos, a veces, *esquema finito*. En un esquema finito la probabilidad se determina unívocamente por las probabilidades elementales. En muchos casos el esquema finito sirve de buen modelo matemático de los fenómenos aleatorios.

A continuación distintos casos particulares de un modelo matemático general de fenómenos aleatorios los llamaremos, a menudo, esquemas indicando sus peculiaridades características (esquema finito, esquema de pruebas independientes (§ 3), etc).

El análisis de las aplicaciones del esquema finito a la descripción de los fenómenos reales lo comenzaremos con un caso particular de este esquema, caso en el cual las probabilidades $p(\omega)$, $\omega \in \Omega$ son iguales. Si convenimos en designar con $|M|$ el número de elementos del conjunto M, entonces de la fórmula (3), suponiendo $p(\omega) = 1/N$, obtenemos

$$\mathbf{P}(A) = \frac{|A|}{|\Omega|} = \frac{k}{N}, \tag{4}$$

donde $\Omega = \{\omega_1, \ldots, \omega_N\}$, $A = \{\omega_{i_1}, \ldots, \omega_{i_k}\}$. La definición (4) se denomina *definición clásica de la probabilidad*. Así, pues, según (4), *la probabilidad de un suceso aleatorio A es igual a la relación entre el número de los sucesos elementales, cuando el suceso A se produce, y el número total de los sucesos elementales.*

La definición clásica de la probabilidad sirve de buen modelo matemático de los fenómenos aleatorios, para los cuales los casos de la prueba son, en cualquier sentido, simétricos y por eso es natural suponer que son equiposibles. En general, esta suposición está justificada en los problemas referentes a los juegos de azar, las loterías, etc. Esto se explica por el hecho de que al fabricar dados, naipes y organizar las loterías se cuida de observar la equiprobabilidad de diferentes resultados. Estos requisitos deben reunirlos asimismo la organización de un control parcial y las investigaciones estadísticas parciales.

Al emplear la fórmula (4) son frecuentemente útiles diversas fórmulas combinatorias. Citemos las más difundidas de ellas.

De un conjunto finito $\{a_1, a_2, \ldots, a_n\}$ compuesto por n elementos diferentes pueden formarse distintas colecciones constituidas por m $(m < n)$ elementos.

Las colecciones ordenadas se llaman *variaciones* y las no ordenadas, *combinaciones*. Por ejemplo, del conjunto $\{1, 2, 3\}$, escogiendo dos a dos elementos $(n = 3, m = 2)$ se pueden formar 6 variaciones ((1, 2), (2, 1), (1, 3) (3, 1), (2, 3), (3, 2)) y 3 combinaciones ((1, 2), (1, 3), (2, 3)). La combinación (1, 2) se puede escribir en la forma de (2, 1), ya que al formar las combinaciones por definición el orden de los elementos no se tiene en cuenta. Las variaciones de n elementos n a n se denominan *permutaciones*. Diversas permutaciones contienen los mismos elementos dispuestos en un orden diferente.

El número de variaciones, que pueden formarse al elegir por diferentes procedimientos m elementos de n, se designa con A_n^m y el número de combinaciones se designa con el símbolo C_n^m o bien $\binom{n}{m}$. Los números A_n^m y C_n^m pueden ser hallados según las fórmulas

$$A_n^m = n^{[m]}, \quad C_n^m = \frac{n^{[m]}}{m!},$$

donde $m! = 1 \cdot 2 \cdot \ldots \cdot m$ y *la potencia generalizada* $n^{[m]}$ se determina por la fórmula $n^{[m]} = n(n-1) \ldots (n-m+1)$. Frecuentemente son útiles las fórmulas siguientes:

$$C_n^m = \frac{n!}{m!(n-m)!}, \quad C_n^m = C_n^{n-m}.$$

EJEMPLO 4. De una urna que contiene M bolas blancas y $N—M$ bolas negras se sacan a la ventura n bolas. Hallemos la probabilidad de que entre las n bolas escogidas sean exactamente m bolas blancas.

La palabra "a la ventura" ("a la suerte" o "al azar") se encuentra en las descripciones de pruebas con bastante frecuencia. En el problema dado se supone que las bolas estaban bien mezcladas, que todas son de un mismo radio e igualmente lisas y no se distinguen sino que por el color. En este caso es racional suponer la equiprobabilidad de los sucesos elementales y aprovecharse de la definición clásica de la probabilidad.

Es natural tomar por sucesos elementales cualesquiera subconjuntos de n a n elementos sacados del conjunto de N bolas. Del curso escolar de las matemáticas se sabe que el número de tales subconjuntos es igual a C_N^n. Ahora bien, en la fórmula (4) hace falta poner $|\Omega| = C_N^n$. Cada colección de bolas que forma parte del suceso que nos interesa (desígnémoslo con A_m) consta de dos partes: 1) m bolas blancas y 2) $n—m$ bolas negras. Todas estas colecciones se pueden obtener del modo siguiente. Primeramente saquemos las partes de colecciones formadas por las bolas blancas; el número de tales partes es C_M^m; luego compongamos las partes de colecciones de bolas negras; el número de tales partes es C_{N-M}^{n-m}. La unión de cualquier parte de la colección de bolas blancas con cualquier parte de la colección de bolas negras ofrece una colección completa de bolas pertenecientes a A_m. Por consiguiente, $|A_m| = C_M^m C_{N-M}^{n-m}$ y por la fórmula (4)

$$\mathbf{P}(A_m) = P_n(m, N, M) = \frac{C_M^m C_{N-M}^{n-m}}{C_N^n}. \tag{5}$$

Aquí y a continuación se supone que $C_n^m = 0$ cuando $m > n$. La colección de los números $P_n(0, N, M)$, $P_n(1, N, M)$, ... se llama *distribución hipergeométrica.*

En las aplicaciones al control parcial el papel de las bolas lo desempeñan N artículos del lote que se comprueba. El número M de artículos defectuosos (bolas blancas) se desconoce. Puede resultar que es imposible comprobar todos los artículos sin excepción: son demasiado muchos o la comprobación lleva a la destrucción del artículo (por ejemplo, durante la verificación se requiere determinar el plazo de servicio de una bombilla). Entonces de todo el lote de artículos se escoge para la comprobación una pequeña parte formada de n artículos. Si entre los artículos elegidos resultaron m artículos defectuosos, entonces se supone que $\frac{M}{N} \approx \frac{m}{n}$. (Tal elección de un valor aproximado será fundamentada en el § 12, ejemplo 5.) A veces se necesita estimar un parámetro desconocido N. Sea N el número desconocido de peces en cierto depósito de agua. Se puede pescar M peces, marcarlos y volver a meterlos en el depósito. Según el número m de peces marcados en la pesca repetida de n peces se puede, basándose en la igualdad aproximada $\frac{M}{N} \approx \frac{m}{n}$, sacar las conclusiones sobre la magnitud N: $N \approx M \cdot \frac{n}{m}$.

Con la mayor frecuencia en la teoría de las probabilidades los modelos matemáticos que tienen aplicaciones en los más diferentes campos se formulan en los términos de esquemas de urna, en los términos de colocaciones de mostacillas o partículas en los cajones. En general, tal forma de descripción no origina dificultades en su utilización práctica. La enunciación de un modelo matemático en los términos que se emplean en cierto campo estrecho de aplicaciones será más embarazosa para otros consumidores que las enunciaciones tradicionales que contienen bolas y urnas, bien conocidas por todos. Para los especialistas del control parcial "las bolas blancas" son, por lo visto, más atractivas que "los peces marcados".

A continuación nos atendremos, por regla, a los términos tradicionales para la teoría de las probabilidades.

EJEMPLO 5. Hagamos una descripción más detallada de la prueba expuesta en el ejemplo precedente. Supongamos que N bolas contenidas en la urna llevan los números de orden $1, 2, \dots, N$. De la urna, a la ventura se sacan, una tras otra, n bolas sin volver a meter las bolas extraídas. En dicho experimento los sucesos elementales ω son las cadenas $\omega = (x_1, x_2, \dots, x_n)$ formadas por los números $1, 2, \dots, N$, con ello entre $x_1, \dots, x_n$ no hay números iguales. El número de elementos $|\Omega|$ del conjunto Ω se puede hallar directamente o haciéndose uso de la fórmula para el número de variaciones de N elementos dispuestos n a n:

$$|\Omega| = A_N^n = N(N-1)\dots(N-n+1).$$

La expresión dada en el segundo miembro de esta igualdad se llama *potencia generalizada del número N*. Emplearemos la siguiente designación:

$$m^{[k]} = m(m-1)\dots(m-k+1). \tag{6}$$

Si la organización de la prueba descrita es conveniente, se puede considerar equiprobables los sucesos elementales y utilizar la definición clásica de la probabilidad. El esquema descrito de la selección se denomina *esquema de selección aleatoria sin devolución*.

Respecto al ejemplo en cuestión los sucesos elementales del ejemplo precedente son más "grandes". En el ejemplo precedente hemos examinado las colecciones no ordenadas de los números de orden y, por lo tanto, de cada tal colección se puede obtener, teniendo en cuenta la disposición, el número igual de $n!$ colecciones que son sucesos elementales en el esquema que se analiza. Así, pues, a cada suceso elemental del ejemplo 4 le corresponde $n!$ sucesos elementales del ejemplo 5. Esto lleva a la coincidencia de las probabilidades de los sucesos que se han determinado en ambos ejemplos.

EJEMPLO 6. De N fichas iguales que llevan los números de orden 0, 1, 2, ..., $N-1$ se sacan, una tras otra, las n fichas. Después de cada extracción la ficha sacada se devuelve. Al igual que en el ejemplo 1, siempre que la organización de la prueba sea acertada, es natural considerar equiprobables todos los casos posibles y utilizar la definición clásica.

Si los números de orden de las fichas extraídas se anotan, como resultado obtenemos una cadena ω de n números: $\omega = (x_1, x_2, \dots, x_n)$, donde x_i pueden tomar un valor cualquiera de $\{0, 1, \dots, N-1\}$ independientemente uno del otro. El conjunto de todas las cadenas ω es en este problema el espacio de sucesos elementales Ω. Supondremos que todos los sucesos elementales son equiprobables. El esquema descrito de la selección se llama *esquema de selección aleatoria con devolución*.

Hallemos el número de sucesos elementales $|\Omega|$ del conjunto Ω. Vamos a partir el conjunto de cadenas $\omega = (x_1, \dots, x_n)$ entre N conjuntos disjuntos en cuanto al valor de x_1:

$$\{\omega: x_1 = 0\}, \quad \{\omega: x_1 = 1\}, \dots, \{\omega: x_1 = N-1\}.$$

A su vez, cada uno de estos conjuntos se puede partir entre N grupos respecto a x_2 y, por consiguiente, respecto a los valores del par $\{x_1 x_2\}$ se forman N^2 grupos; respecto a los valores del triplete $\{x_1, x_2, x_3\}$ obtenemos

N^3 particiones, etc. Como resultado de n particiones obtenemos N^n grupos, cada uno de los cuales contiene un elemento. Ahora bien, $|\Omega| = N^n$, $p(\omega) = N^{-n}$ y la probabilidad se determina por la fórmula (4). Calculemos la probabilidad del suceso A_m = {entre n fichas sacadas exactamente m fichas llevan los números de orden que no superan $M - 1$}. Entre n lugares de la cadena $\omega = (x_1, \ldots, x_n)$ se puede por C_n^m procedimientos escoger los lugares en los cuales colocaremos los números de orden de las fichas que no exceden $M - 1$. Estos lugares pueden ser llenados por M^m métodos. Los $n - m$ lugares restantes los llenaremos, por $(N - M)^{n-m}$ métodos, de fichas cuyos números de orden son mayores que $M - 1$. Así, pues, $|A_m| = C_n^m M^m (N - M)^{n-m}$ y según la fórmula (4) encontramos

$$P(A_m) = \frac{C_n^m M^m (N - M)^{n-m}}{N^n} = C_n^m \left(\frac{M}{N}\right)^m \left(1 - \frac{M}{N}\right)^{n-m},$$

$$m = 0, 1, 2, \ldots, n.$$

Los números obtenidos como resultado de una bastante buena realización de la selección aleatoria con devolución para $N = 10$ se llaman ***números aleatorios uniformemente distribuidos*** (o simplemente números aleatorios). Las tablas de tales números (véanse [1], [11], [14]) pueden utilizarse para la simulación de los fenómenos aleatorios. Supongamos, por ejemplo, que es necesario simular los resultados de n pruebas consistentes en el arrojamiento de una moneda simétrica. Convengamos en considerar que a la salida del escudo le corresponde una cifra par (0, 2, 4, 6, 8). Al sustituir las cifras pares por la letra "E" (escudo) y las impares por "C" (cifra) obtenemos la sucesión de dos símbolos: E y C. Es evidente que son equiprobables todas las sucesiones posibles obtenidas por el método indicado. Designemos con $n_E(n)$ el número de "escudos" en la sucesión cuya longitud es igual a n. A título de ejemplo en la tabla de números aleatorios se han encontrado los valores de $n_E = n_E(n)$ y se han calculado las frecuencias $h_E(n) = n_E/n$. Los resultados se dan en la siguiente tabla:

n	5	10	20	30	40	50
n_E	3	7	13	15	21	25
$h_E(n)$	0,60	0,70	0,65	0,50	0,53	0,50

n	60	70	80	90	100
n_E	32	34	37	43	47
$h_E(n)$	0,53	0,49	0,46	0,48	0,47

El gráfico de la función $h_E(n)$ se muestra en la fig. 3.

EJEMPLO 7. Supongamos que un dado ha sido lanzado una vez. En este caso es natural poner $\Omega = \{1, 2, 3, 4, 5, 6\}$. Si el dado es simétrico, entonces las probabilidades de los sucesos elementales se consideran iguales: $p(1) = p(2) = \dots = p(6) = 1/6$. Así que obtenemos la definición clásica de la probabilidad. Si el dado no es simétrico, tenemos que utilizar el esquema finito con diferentes $p(i)$, $i = 1, \dots, 6$ que pueden ser estimadas experimentalmente. La construcción de las estimaciones de parámetros desconocidos es una de las tareas de la estadística matemática (véase el § 12).

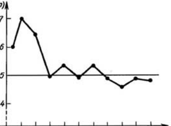

Fig. 3.

1.7. Espacio probabilístico numerable. Sea $\Omega = \{\omega\}$ un conjunto numerable ($\Omega = \{\omega_1, \omega_2, \dots, \omega_n, \dots\}$); sea $\mathscr{A}$ el conjunto de todos los subconjuntos Ω y sea $\{p(\omega)\}$ una colección de los números que satisfacen las condiciones

$$p(\omega) \geqslant 0, \ \omega \in \Omega; \qquad \sum_{n=1}^{\infty} p(\omega_n) = 1.$$

Llamaremos probabilidad del suceso $A = \{\omega_{i_1}, \omega_{i_2}, \ldots, \omega_{i_l}, \ldots\}$ el número $\mathbf{P}(A)$ que se determina por las fórmulas

$$\mathbf{P}(A) = \sum_{\omega \in A} p(\omega) = \sum_{l=1}^{\infty} p(\omega_{i_l}), \quad A \neq \varnothing, \tag{7}$$

y $\mathbf{P}(\varnothing) = 0$. Ahora bien,
la probabilidad del suceso A es igual a la suma de la serie formada por las probabilidades elementales $p(\omega)$, en las cuales ω forman parte de A.

Los axiomas A1—A4 se comprueban al igual que en el esquema finito. El orden de numeración de los sucesos elementales no influye en la definición, ya que $p(\omega) \geqslant 0$ y la suma de la serie que entra en (7) no varía al cambiar el orden de sumación. *El espacio probabilístico numerable* construido lo llamaremos, a veces, *esquema numerable*.

Nótese que la probabilidad $\mathbf{P}(A)$, al igual que en un esquema finito, se determina unívocamente por las probabilidades de los sucesos elementales $p(\omega)$. Es esquema finito es un caso particular del esquema numerable con $p(\omega_k) = 0, k \geqslant N + 1$. Los esquemas numerable y finito se llaman *esquema discreto o espacio probabilístico discreto*.

EJEMPLO 8. Una moneda va arrojada repetidamente hasta que salga dos veces seguidas un mismo lado. Supongamos que necesitamos determinar la probabilidad de los sucesos: B = {la prueba se finalizará con un número par de lanzamientos}, A = {la prueba durará 5 lanzamientos, como máximo}. Pongamos $\Omega = \{n : n \geqslant 2\}$, donde los números naturales n vamos a interpretarlos como duración de la prueba. Procuremos introducir naturalmente las probabilidades de los sucesos elementales $p(n)$. Al arrojar la moneda n veces son posibles 2^n diferentes casos de la prueba. Supondremos que estos casos son equiprobables. Entre 2^n casos hay dos que corresponden al hecho de que la moneda por primera vez salga dos veces seguidas con un mismo lado en la n-ésima prueba. Así, pues, es natural suponer que $p(n) = 2/2^n = 2^{-n+1}$ $(n \geqslant 2)$.

Los razonamientos expuestos sobre la probabilidad $p(n)$ no son una demostración matemática; se puede estimarlos sólo como ciertas consideraciones sugestivas. No es difícil comprobar que los números escogidos satisfacen las condiciones

$$p(n) = 2^{-n+1} > 0, \quad n \geqslant 2; \qquad \sum_{n=2}^{\infty} p(n) = 2 \sum_{n=2}^{\infty} \frac{1}{2^n} = 1.$$

Ahora podemos determinar la probabilidad con ayuda de la fórmula (7). Los sucesos A y B examinados en este ejemplo pueden ser representados en forma de

$$A = \{2, 3, 4, 5\}, \quad B = \{2, 4, 6, 8, \ldots\}.$$

Por consiguiente,

$$\mathbf{P}(A) = p(2) + p(3) + p(4) + p(5) = \frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3} + \frac{1}{2^4} = \frac{15}{16},$$

$$\mathbf{P}(B) = \sum_{k=1}^{\infty} p(2k) = 2 \sum_{k=1}^{\infty} \frac{1}{2^{2k}} = \frac{2}{3}.$$

1.8. Espacio probabilístico continuo. Pongamos $\Omega = \{u = (u_1, \ldots, u_n)\colon u \in G\}$, donde G es la región cuadrable del espacio euclídeo n-dimensional. Designemos con $\mathscr{A}$ el sistema de subconjuntos cuadrables de la región G. Del curso "Análisis" sabemos que la suma, el producto y la diferencia de las figuras cuadrables son figuras cuadrables. Ahora bien, $\mathscr{A}$ es el álgebra de sucesos. Sea $\pi(u_1, \ldots, u_n) \geqslant 0$ una función integrable sobre la región G y la integral de esta función sobre la región G es igual a 1.

Lamaremos probabilidad del suceso A el número $\mathbf{P}(A)$ determinado por la fórmula

$$\mathbf{P}(A) = \int_A \ldots \int \pi(u_1 \ldots, u_n) du_1 \ldots du_n, \tag{8}$$

donde la integral del segundo miembro (8) es una integral de Riemann múltiple de n. Utilizando las propiedades de las integrales no es difícil comprobar que la función $\mathbf{P}(A)$ satisface los axiomas A1—A3.

Nótese que en vez de la región finita G se puede examinar el espacio n-dimensional y la integral en (8) entender como impropia. La función $\mathbf{P}(A)$ determinada sobre el álgebra de $\mathscr{A}$ se puede prolongar sobre un sistema de conjuntos más amplio que contenga sumas y productos numerables de los sucesos.

El espacio probabilístico construido lo llamaremos, a veces, *modelo probabilístico continuo* o *simplemente esquema continuo.*

Examinemos un caso particular del esquema continuo general, poniendo $\pi(u_1, \dots, u_n) = 1/m(G)$ si $u \in G$ y $\pi(u_1, \dots, u_n) = 0$, si $u \notin G$. Aquí $m(G)$ es el área o el volumen de la región G. Con tal elección de la función $\pi(u)$ la fórmula (8) se escribirá del modo siguiente:

$$\mathbf{P}(A) = m(A)/m(G). \tag{9}$$

Tal definición de la probabilidad se llama *geométrica*. La definición geométrica de la probabilidad se puede considerar como generalización de la definición clásica para experimentos con un número infinito de casos. A continuación, a veces haremos uso de la expresión el "punto está distribuido uniformemente sobre el conjunto G". Esto quiere decir que la probabilidad se debe calcular según la fórmula (9).

A título de la ilustración cómo se usa la definición geométrica examinemos el siguiente ejemplo.

EJEMPLO 9. En el intervalo de tiempo $[0, T]$ deben llegar a un dispositivo de servicio dos pedidos. Si la diferencia entre los instantes de llegada de los pedidos es menor que t, el segundo pedido se pierde. Se necesita hallar la probabilidad de la pérdida del pedido.

Designemos con t_1 y t_2 los instantes de llegada de los pedidos. Pongamos $\Omega = \{(t_1, t_2): 0 \leqslant t_1 \leqslant T, 0 \leqslant t_2 \leqslant T\}$; $A = \{$el pedido se perderá$\}$. Entonces $A = \{(t_1, t_2): |t_1 - t_2| < t\}$ (fig. 4). Si se hace uso de la definición geométrica, se necesita calcular el área $m(\Omega) = T^2$ y $m(A) = T^2 - (T - t)^2$. Según la fórmula (9) hallamos $\mathbf{P}(A) = 1 - (1 - t/T)^2$. El empleo de la definición geométrica de las probabilidades entraña ciertas suposiciones de la ley de llegada de los pedidos. El hecho de que el modelo escogido de un fenómeno aleatorio corresponda a la realidad puede ser estimado basándose en los experimentos.

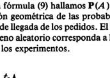

Fig. 4.

Ahora obtendremos ciertos corolarios simples de los axiomas A1—A4. De los axiomas A2, A3 y de la igualdad $A + \bar{A} = \Omega$ resulta que

$$\mathbf{P}(\bar{A}) = 1 - \mathbf{P}(A). \tag{10}$$

Suponiendo aquí $A = \Omega$, obtenemos $\mathbf{P}(\varnothing) = 0$. Para cualesquiera sucesos A y B tiene lugar la fórmula

$$\mathbf{P}(A + B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(AB). \tag{11}$$

En efecto, representemos los sucesos $A + B$ y B en la forma $A + B = A + B\bar{A}$ y $B = B\bar{A} + BA$. Los sucesos contenidos en los segundos miembros de estas igualdades son incompatibles y, por lo tanto,

$$\mathbf{P}(A + B) = \mathbf{P}(A) + \mathbf{P}(B\bar{A}), \quad \mathbf{P}(B) = \mathbf{P}(B\bar{A}) + \mathbf{P}(AB).$$

De aquí se deduce fácilmente (11). Haciendo uso de la fórmula (11), obtenemos la fórmula para $\mathbf{P}(A_1 + A_2 + A_3)$. Pongamos $A = A_1 + A_2$, $B = A_3$. Entonces de la fórmula (11) y de la fórmula $AB = A_1A_3 + A_2A_3$ resulta que

$$\mathbf{P}(A_1 + A_2 + A_3) = \mathbf{P}(A_1 + A_2) + \mathbf{P}(A_3) - \mathbf{P}(A_1A_3 + A_2A_3).$$

De aquí aplicando a las probabilidades $\mathbf{P}(A_1 + A_2)$ y $\mathbf{P}(A_1A_3 + A_2A_3)$ la fórmula (11) y haciendo uso de la igualdad $(A_1A_3)(A_2A_3) = A_1A_2A_3$, obtenemos

$$\mathbf{P}(A_1 + A_2 + A_3) = \mathbf{P}(A_1) + \mathbf{P}(A_2) + \mathbf{P}(A_3) - {}$$
$$- \mathbf{P}(A_1A_2) - \mathbf{P}(A_1A_3) - \mathbf{P}(A_2A_3) + \mathbf{P}(A_1A_2A_3).$$

La fórmula general para la probabilidad de la suma de n sucesos tiene el siguiente aspecto:

$$\mathbf{P}(A_1 + A_2 + \ldots + A_n) = \sum_{k=1}^{n} \mathbf{P}(A_k) - \sum_{1 \leqslant k < l \leqslant n} \mathbf{P}(A_kA_l) + {}$$
$$+ \sum_{1 \leqslant k < l < m \leqslant n} \mathbf{P}(A_kA_lA_m) - \ldots + (-1)^{n+1}\mathbf{P}(A_1 \ldots A_n) =$$
$$= \sum_{r=1}^{\infty} (-1)^{r+1} \sum_{1 \leqslant l_1 < \ldots < l_r \leqslant n} \mathbf{P}(A_{l_1} \ldots A_{l_r}). \tag{12}$$

Al demostrar por inducción la fórmula (12) el paso de $n - 1$ a n se lleva a cabo de un modo análogo al paso de 2 a 3, citado anteriormente.

Directamente de (11) obtenemos para cualesquiera A y B

$$\mathbf{P}(A + B) \leqslant \mathbf{P}(A) + \mathbf{P}(B). \tag{13}$$

Si $A \subset B$, entonces

$$\mathbf{P}(A) \leqslant \mathbf{P}(B), \tag{14}$$

ya que $B = A + B\bar{A}$, $\mathbf{P}(B) = \mathbf{P}(A) + \mathbf{P}(B\bar{A})$ y $\mathbf{P}(B\bar{A}) \geqslant 0$. Si los sucesos $A_1, A_2, \ldots, A_n$ son incompatibles dos a dos, o sea, $A_i A_j = \varnothing$ para todos valores de $i \neq j$, entonces

$$\mathbf{P}(A_1 + A_2 + \ldots + A_n) = \mathbf{P}(A_1) + \mathbf{P}(A_2) + \ldots + \mathbf{P}(A_n). \tag{15}$$

Esta fórmula se demuestra por inducción con ayuda del axioma de aditividad A3; además, (15) se puede obtener como caso particular de (12). Por inducción, de (13) se deduce que

$$\mathbf{P}(A_1 + \ldots + A_n) \leqslant \mathbf{P}(A_1) + \ldots + \mathbf{P}(A_n).$$

Examinemos algunos ejemplos que ilustran la utilización de los corolarios obtenidos de los axiomas.

EJEMPLO 10. Estimemos la probabilidad de que se obtenga cualquier premio en el "Sportlotó", poseyendo una ficha. Un participante de la lotería de 49 números de orden marca 6 números. Después de que el participante ha entregado su ficha, se realiza la selección de $n = 6$ números de orden. Sea $A_m = \{m$ números de orden marcados por el participante llegan a ser seleccionados$\}$. Si la cantidad de los números de orden marcados por el participante y seleccionados en la lotería es más de 2, éste obtiene un premio. Así, pues, $A = \{$la obtención de un premio cualquiera$\} = A_3 + A_4 + A_5 + A_6$ y, por consiguiente, $\bar{A} = A_0 + A_1 + A_2$. Según la fórmula (5) con $N = 49$, $M = 6$, $n = 6$ encontramos

$$\mathbf{P}(A_0) = 0{,}435965, \quad \mathbf{P}(A_1) = 0{,}413019, \quad \mathbf{P}(A_2) = 0{,}132378,$$
$$\mathbf{P}(A_3) = 0{,}017650, \quad \mathbf{P}(A_4) = 0{,}000969, \quad \mathbf{P}(A_5) = 0{,}000018,$$
$$\mathbf{P}(A_6) = 0{,}00000007151.$$

De aquí, utilizando (15), obtenemos

$$\mathbf{P}(\bar{A}) = \mathbf{P}(A_0) + \mathbf{P}(A_1) + \mathbf{P}(A_2) = 0{,}981362$$

y

$$\mathbf{P}(A) = 1 - \mathbf{P}(\bar{A}) = 0{,}018638.$$

EJEMPLO 11. Hallemos la probabilidad de que entre n números aleatorios (véase el ejemplo 6) la cifra "1" se encuentre, por lo menos, una vez.

El conjunto de sucesos elementales Ω es constituido por las sucesiones que se pueden componer de todas las 10 cifras; $\Omega = \{(x_1, x_2, \ldots, x_n)\}$, donde $x_k = 0, 1, \ldots, 9$, $k = 1, \ldots, n$. En el ejemplo 6 hemos mostrado que $|\Omega| = 10^n$. Pongamos A = {entre n números aleatorios se ha encontrado al menos una cifra "1"}. Entonces $\bar{A}$ = {entre n números aleatorios la cifra "1" no se ha encontrado ninguna vez}. Así, pues, en $\bar{A}$ entran las cadenas formadas solamente por 9 cifras, puesto que "1" no puede ser utilizada al componer las cadenas. Razonando al igual que en el ejemplo 6, al calcular $|\Omega|$ obtenemos $|\bar{A}| = 9^n$. Por consiguiente, por la fórmula (4) tenemos $\mathbf{P}(\bar{A}) = 9^n/10^n$. De aquí, haciendo uso de la fórmula (10), hallamos

$$\mathbf{P}(A) = 1 - \mathbf{P}(\bar{A}) = 1 - (9/10)^n.$$

EJEMPLO 12. Se arrojan dos dados. Hallemos la probabilidad del suceso A = { al menos en uno de los dados salen más de dos puntos}.

El conjunto de todos los casos posibles se puede escribir en forma de

$$\Omega = \{(i, j): i, j = 1, 2, \dots, 6\},$$

donde el suceso elemental (i, j) corresponde a la salida de i puntos en un dado y j puntos en el segundo. No es difícil comprobar que $|\Omega| = 6 \cdot 6 = 36$. Designemos con B_1 el suceso consistente en que en el primer dado no salen más de dos puntos y con B_2, el suceso análogo para el segundo dado. Entonces

$$B_1 = \{(i, j): \quad i = 1, 2; \quad j = 1, 2, \dots, 6\},$$
$$B_2 = \{(i, j): \quad i = 1, \dots, 6; \quad j = 1, 2\}$$

y, por lo tanto, $|B_1| = |B_2| = 12$. Puesto que $B_1B_2 = \{(i, j): i, j = 1, 2\}$, entonces $|B_1B_2| = 4$. Ahora bien,

$$\mathbf{P}(B_1) = \mathbf{P}(B_2) = \frac{12}{36} = \frac{1}{3}, \quad \mathbf{P}(B_1B_2) = \frac{1}{9}.$$

De aquí por la fórmula (11) hallamos

$$\mathbf{P}(A) = \mathbf{P}(B_1 + B_2) = \mathbf{P}(B_1) + \mathbf{P}(B_2) - \mathbf{P}(B_1B_2) = \frac{1}{3} + \frac{1}{3} - \frac{1}{9} = \frac{5}{9}.$$

EJEMPLO 13. Examinemos cierto sistema complejo formado por n bloques. Supongamos que en cierto modelo probabilístico están determinados los sucesos A_k = {el k-ésimo bloque no se averiará durante el período en cuestión} y sus probabilidades $\mathbf{P}(A_k)$, $k = 1, 2, \dots, n$. Por ejemplo, las probabilidades $\mathbf{P}(A_k)$ han sido halladas aproximadamente según los resultados de pruebas de diferentes bloques. Evaluemos la probabilidad del suceso A = {el sistema no se averiará durante el período en cuestión} $= A_1A_2 \dots A_n$. Puesto que

$$\bar{A} = \bigcup_{k=1}^{n} \bar{A}_k, \text{ entonces}$$

$$1 - \mathbf{P}(A) = \mathbf{P}(\bar{A}) = \mathbf{P}\left(\bigcup_{k=1}^{n} \bar{A}_k\right) \leqslant \sum_{k=1}^{n} \mathbf{P}(\bar{A}_k).$$

Además, para todo valor de k

$$A_1 \dots A_n \subset A_k, \quad \mathbf{P}(A) \leqslant \mathbf{P}(A_k).$$

Finalmente obtenemos

$$1 - \sum_{k=1}^{n} (1 - \mathbf{P}(A_k)) \leqslant \mathbf{P}(A) \leqslant \min_k \mathbf{P}(A_k).$$

Problemas

1. Mostrar que $A \setminus (B \setminus C) \neq A \setminus B + C$. Hallar una expresión más simple para $A \setminus (B \setminus C)$.

2. Sean A, B y C tres sucesos arbitrarios. Hallar las expresiones para los sucesos consistentes en que entre A, B y C

1) se ha producido solamente A,

2) se han producido A y B, mientras que C no se ha producido,

3) se han producido todos los tres sucesos,

4) se ha producido al menos uno de los sucesos,

5) no se ha producido ninguno de los sucesos.

3. Entre 100 artículos del lote en cuestión hay 5 defectuosos. Hallar la probabilidad de que entre 10 artículos escogidos a la suerte no se contenga más que un artículo defectuoso.

4. Hallar la probabilidad de que los cumpleaños de 12 personas estén distribuidos entre diferentes meses del año.

5. Un participante de la lotería "Sportlotó" ha llenado dos fichas de modo que todos los números de orden tachados en ambas fichas sean diferentes. Hallar la probabilidad de que el participante no haya adivinado ningún número.

6. Las tres estaciones de radio tienen el permiso de trabajar en cualquiera de tres frecuencias asignadas. Hallar las probabilidades de los sucesos: A = {todas las estaciones de radio están funcionando en diferentes frecuencias}; B = {las dos estaciones de radio están funcionando en iguales frecuencias y la tercera, en otra frecuencia}. Suponer que todas las selecciones posibles de las frecuencias son equiprobables.

7. Según el esquema de selección aleatoria con devolución, de un conjunto de números enteros $(1, 2, \dots, N)$ se escogen dos números ξ y η. Designemos con p_N la probabilidad del suceso $\xi^2 + \eta^2 \leqslant N^2$. Hallar $\lim_{N \to \infty} p_N$.

8. El punto A está distribuido uniformemente en el cuadrado del lado 1. Hallar las probabilidades de los sucesos:

a) la distancia entre el punto A y el lado fijado del cuadrado no supera x;

b) la distancia entre el punto A y el lado más próximo del cuadrado no supera x.

9. Al entarimado formado por triángulos perfectos del lado a se lanza a la suerte una moneda de radio r $\left(r < a\dfrac{\sqrt{3}}{6}\right)$. Hallar la probabilidad de que la moneda no toque el límite de ninguno de los triángulos.

10. Una barra de largo l se ha rompido en dos puntos escogidos a la ventura. ¿A qué es igual la probabilidad de que de los segmentos obtenidos se pueda formar un triángulo?

11. Cada uno de n sobres lleva puesta una carta a n destinatarios. En cada sobre se ha escrito a la suerte uno de n destinatarios. Hallar la probabilidad p_n de que al menos una carta llegue al destinatario respectivo, calcular $\lim\limits_{n\to\infty} p_n$.

Indicación: utilizar la fórmula (12).

§ 2. Probabilidades condicionales. Independencia de sucesos

2.1. Probabilidades condicionales. Antes de dar una definición formal, vamos a exponer ciertas consideraciones que puedan ayudar a comprender la naturalidad de la misma.

Supongamos que se realizan n pruebas. Designemos con n_A el número de experimentos, en los cuales se ha producido el suceso A. Supongamos asimismo que entre n_A pruebas junto con A se ha producido n_{AB} veces el suceso B. Respecto a los experimentos con el caso A la frecuencia del suceso B es la relación $h_n(B \mid A) = n_{AB}/n_A$. Si en esta relación dividimos el numerador y el denominador por n, obtenemos la expresión que liga $h_n(B \mid A)$ con las frecuencias incondicionales $h_n(A)$ y $h_n(AB)$:

$$h_n(B \mid A) = \frac{n_{AB}/n}{n_A/n} = \frac{h_n(AB)}{h_n(A)}. \qquad (1)$$

Cuando los valores de n son grandes las frecuencias $h_n(A)$ y $h_n(AB)$ son próximas a las probabilidades $\mathbf{P}(A)$ y $\mathbf{P}(AB)$. La igualdad (1) se puede poner por base de la definición de la probabilidad condicional. Sea $\mathbf{P}(A) > 0$.

El número $\mathbf{P}(B \mid A)$ *definido por la igualdad*

$$\mathbf{P}(B \mid A) = \frac{\mathbf{P}(AB)}{\mathbf{P}(A)} \qquad (2)$$

se llama probabilidad condicional del suceso B a la condición A.

La probabilidad condicional $\mathbf{P}(B \mid A)$ vamos a designarla, a veces, con $\mathbf{P}_A(B)$.

EJEMPLO 1. En un almacén están reunidos N artículos de dos talleres, con ello el primer taller ha suministrado N_1 artículos entre los cuales M_1 artículos son de primera calidad y el segundo taller ha suministrado N_2 artículos, entre ellos M_2 artículos de primera calidad. Un artículo recibido del almacén fue escogido a la suerte. Hallemos la probabilidad del suceso $B = \{$fue recibido un artículo de primera calidad$\}$. Según la definición clásica de la probabilidad tenemos, evidentemente, $\mathbf{P}(B) = (M_1 + M_2)/N$. Supongamos que hemos logrado determinar qué taller había producido el artículo recibido; por ejemplo, hemos revelado que el artículo había sido fabricado por el primer taller. Designemos este suceso con A. Disponiendo de esta información adicional, es natural buscar la probabilidad condicional $\mathbf{P}(B \mid A)$. Según la fórmula (2) hallamos

$$\mathbf{P}(B \mid A) = \frac{\mathbf{P}(AB)}{\mathbf{P}(A)} = \frac{M_1/N}{N_1/N} = \frac{M_1}{N_1},$$

o sea, en este caso la probabilidad condicional es una porción de artículos de primera calidad entre los fabricados por el primer taller.

Sea fijado cierto suceso A con $\mathbf{P}(A) > 0$. La función $\mathbf{P}_A(B) = \mathbf{P}(B \mid A) = \mathbf{P}(AB)/\mathbf{P}(A)$ determinada para todos los valores de $B \in \mathscr{A}$, satisface los axiomas A1—A4. Ahora bien, para $\mathbf{P}_A(B)$ son válidos todos los corolarios (10)—(15) obtenidos en el § 1 directamente de los axiomas.

La igual (2) puede ser escrita en forma

$$\mathbf{P}(AB) = \mathbf{P}(A)\mathbf{P}(B \mid A). \tag{3}$$

De aquí, por inducción, es fácil obtener una fórmula más general

$$(\mathbf{P}A_1A_2 \dots A_n) = \\ = \mathbf{P}(A_1)\mathbf{P}(A_2 \mid A_1) \dots \mathbf{P}(A_n \mid A_1A_2 \dots A_{n-1}). \tag{4}$$

En efecto, para $n = 2$ la fórmula (4) coincide con (3). Supongamos que la fórmula (4) queda demostrada para $n - 1$ factores. Entonces para n factores la fórmula (4) se deduce de la suposición de la inducción y de la igualdad (3), en la cual es necesario poner $A = A_1A_2 \dots A_{n-1}$, $B = A_n$.

Las fórmulas (3) y (4) son, evidentemente, inútiles para calcular las probabilidades de los productos de sucesos, ya que los segundos miembros de estas fórmulas contienen las probabilidades condicionales, para cuya determinación hace falta saber las probabilidades de los productos. La utilidad de las fórmulas (3) y (4) se revela al construir modelos matemáticos de diversas series de experimentos que serán examinados en el § 3. Aquí nos limitamos a examinar diferentes ejemplos.

En las definiciones clásica y geométrica de la probabilidad hemos partido de "la equiposibilidad" de los casos de una prueba. Ahora vamos a citar los ejemplos de los espacios probabilísticos, cuya construcción está basada sobre el conocimiento de los valores de probabilidades condicionales.

En las aplicaciones resulta frecuentemente que las probabilidades condicionales se asignan de un modo natural por las condiciones de la prueba o se determinan aproximadamente al realizar experimentos adicionales. Por ejemplo, al llevar a cabo una serie de n pruebas un suceso real A se ha producido n_A veces y n_{AB} veces el suceso A se ha producido junto con un suceso real B, entonces en calidad de un valor aproximado de la probabilidad condicional $\mathbf{P}(B \mid A)$ se puede tomar la frecuencia $h_n(B \mid A) = n_{AB}/n_A$. En las series de experimentos consistentes en pasar bolas de una urna a la otra o en extraer las bolas de una urna es natural asignar las probabilidades condicionales de que se obtenga la muestra de tipo determinado de la urna en cuestión, si de las pruebas precedentes se conoce la composición de la misma. A partir de los valores atribuidos de las probabilidades condicionales, pueden ser calculadas las probabilidades de los sucesos elementales.

EJEMPLO 2. De la primera urna que contiene 5 bolas blancas y 3 negras se pasa, a la suerte, una bola de color desconocido a la segunda urna que contiene 2 bolas blancas y 3 negras. Acto seguido, de la segunda urna se saca una bola. Vamos a construir un modelo matemático de la prueba descrita. Designemos con A_1 y A_2 los sucesos consistentes en que de la primera y la segunda urna han sido extraídas las bolas blancas. Tomando en consideración la equiposibilidad de los casos, es natural estimar que $\mathbf{P}(A_1) = 5/8$. Si de la primera urna ha pasado a la segunda una bola blanca, entonces en la segunda urna habrán 3 bolas blancas y 3 negras. Por consiguiente, a esta condición es necesario poner $\mathbf{P}(A_2/A_1) = 3/6 = 1/2$. Análogamente, determinamos la probabilidad condicional del suceso A_2 al pasar una bola negra: $\mathbf{P}(A_2/\bar{A}_1) = 4/6 = 2/3$. De los razonamientos expuestos (que no son una demostración matemática) se puede llegar a la conclusión de que en el modelo matemático que deseamos construir deben cumplirse las igualdades

$$\mathbf{P}(A_1) = 5/8, \quad \mathbf{P}(A_2 \mid A_1) = 1{,}2 \quad \mathbf{P}(A_2 \mid \bar{A}_1) = 2/3. \quad (5)$$

Así, pues, se necesita hallar un espacio probabilístico conveniente $(\Omega, \mathscr{A}, \mathbf{P})$, cuya álgebra $\mathscr{A}$ contiene los sucesos A_1, A_2 y la probabilidad $\mathbf{P}$ se determina de modo que se cumplan las condiciones (5).

Pongamos $\Omega = \{A_1A_2, \bar{A}_1A_2, A_1\bar{A}_2, \bar{A}_1\bar{A}_2\}$, $\mathscr{A}$ es el conjunto de todos los subconjuntos de Ω y

$$\begin{aligned}
\mathbf{P}(A_1A_2) &= \mathbf{P}(A_1)\mathbf{P}(A_2 \mid A_1) = 5/8 \cdot 1/2,\\
\mathbf{P}(\bar{A}_1A_2) &= \mathbf{P}(\bar{A}_1)\mathbf{P}(A_2 \mid \bar{A}_1) = 3/8 \cdot 2/3,\\
\mathbf{P}(A_1\bar{A}_2) &= \mathbf{P}(A_1)\mathbf{P}(\bar{A}_2 \mid A_1) = 5/8 \cdot 1/2,\\
\mathbf{P}(\bar{A}_1\bar{A}_2) &= \mathbf{P}(\bar{A}_1)\mathbf{P}(\bar{A}_2 \mid \bar{A}_1) = 3/8 \cdot 1/3.
\end{aligned} \tag{6}$$

Las probabilidades (6) determinan unívocamente la probabilidad de todo subconjunto de Ω.

La construcción de los espacios probabilísticos en casos más generales, con la utilización de las probabilidades condicionales y la independencia de sucesos, será examinada en el § 3.

2.2. Fórmula de la probabilidad completa. Fórmula de Bayes.

Diremos que los sucesos $B_1, B_2, \dots, B_n$ forman una *partición* si

$$\Omega = B_1 + B_2 + \dots + B_n \tag{7}$$

y $B_iB_j = \varnothing$ para todos valores de $i \neq j$.

Para toda partición (7) tiene lugar la siguiente fórmula (*fórmula de la probabilidad completa*):

$$\mathbf{P}(A) = \sum_{k=1}^{n} \mathbf{P}(B_k)\mathbf{P}(A \mid B_k). \tag{8}$$

Nótese que para demostrar esta fórmula A se puede representar en forma de la siguiente suma de los sucesos incompatibles dos a dos:

$$A = AB_1 + AB_2 + \dots + AB_n.$$

De aquí, haciendo uso del axioma A3 y de la fórmula (3), obtenemos la fórmula (8):

$$\mathbf{P}(A) = \sum_{k=1}^{n} \mathbf{P}(AB_k) = \sum_{k=1}^{n} \mathbf{P}(B_k)\mathbf{P}(A \mid B_k).$$

Sustituyendo en la igualdad

$$\mathbf{P}(B_k \mid A) = \frac{\mathbf{P}(AB_k)}{\mathbf{P}(A)} = \frac{\mathbf{P}(B_k)\mathbf{P}(A \mid B_k)}{\mathbf{P}(A)}$$

la probabilidad $\mathbf{P}(A)$ por la fórmula (8), obtenemos *la fórmula de Bayes*

$$\mathbf{P}(B_k \mid A) = \frac{\mathbf{P}(B_k)\mathbf{P}(A \mid B_k)}{\sum_{i=1}^{n} \mathbf{P}(B_i)\mathbf{P}(A \mid B_i)}. \tag{9}$$

Las fórmulas de Bayes se interpretan con frecuencia como fórmulas que permiten por las probabilidades *a priori* (conocidas previamente, antes de realizarse un experimento) $\mathbf{P}(B_k)$ y por los resultados del experimento (llegada del suceso A) hallar las probabilidades *a posteriori* (calculadas después del experimento) $\mathbf{P}(B_k \mid A)$. Un ejemplo interesante que indica las aplicaciones posibles de las fórmulas de Bayes se contiene en el problema 10.

La fórmula de la probabilidad completa, al igual que las (3) y (4), puede ser utilizada para construir un espacio probabilístico conveniente según las probabilidades condicionales prefijadas. A continuación se examinan dos ejemplos, en los cuales es natural considerar asignadas las probabilidades condicionales.

EJEMPLO 3. En una fábrica que produce pernos la primera máquina hace el 25%, la segunda el 35% y la tercera el 40% de todos los artículos. El desecho constituye el 5%, el 4% y el 2%, respectivamente. Hallar

a) la probabilidad de que un perno seleccionado a la suerte sea defectuoso,

b) la probabilidad de que si un perno escogido a la ventura es defectuoso, esté fabricado por la primera, la segunda o la tercera máquina.

Designemos con A el suceso consistente en que un perno escogido a la suerte es defectuoso y con B_1, B_2, B_3 sucesos consistentes en que este perno ha sido producido por la primera, la segunda y la tercera máquina, respectivamente. Es evidente que la fórmula (8) es aplicable. Ahora bien, utilizando los datos del problema, obtenemos

$$\mathbf{P}(A) = \mathbf{P}(B_1)\mathbf{P}(A \mid B_1) + \mathbf{P}(B_2)\mathbf{P}(A \mid B_2) + \\ + \mathbf{P}(B_3)\mathbf{P}(A \mid B_3) = 0{,}25\cdot 0{,}05 + 0{,}35\cdot 0{,}04 + 0{,}40\cdot 0{,}02 = 0{,}0345.$$

A estos mismos sucesos pueden ser aplicadas las fórmulas de Bayes (9), cuando $n = 3$ para $k = 1, 2, 3$:

$$\mathbf{P}(B_1 \mid A) = \frac{0{,}25\cdot 0{,}05}{0{,}0345} = \frac{125}{345},$$

$$\mathbf{P}(B_2 \mid A) = \frac{0{,}35 \cdot 0{,}04}{0{,}0345} = \frac{140}{345},$$

$$\mathbf{P}(B_3 \mid A) = \frac{0{,}40 \cdot 0{,}02}{0{,}0345} = \frac{80}{345}.$$

EJEMPLO 4. De una urna que contenía M bolas blancas y $N—M$ negras se ha perdido una bola de color desconocido. Hallemos la probabilidad de que una bola extraída de la urna después de la pérdida sea blanca.

Sea B_k el suceso consistente en que se han perdido k bolas blancas $(k = 0, 1)$; sea A el suceso consistente en que la bola extraída entre las bolas quedadas es blanca. Pongamos

$$\mathbf{P}(B_0) = \frac{N - M}{N}, \qquad \mathbf{P}(B_1) = \frac{M}{N},$$

$$\mathbf{P}(A \mid B_0) = \frac{M}{N - 1}, \qquad \mathbf{P}(A \mid B_1) = \frac{M - 1}{N - 1}.$$

Según la fórmula de la probabilidad completa

$$\mathbf{P}(A) = \frac{N - M}{N}\frac{M}{N - 1} + \frac{M}{N}\frac{M - 1}{N - 1} = \frac{M}{N}.$$

Notemos que la probabilidad de que de la urna salga una bola blanca antes de ocurrir la pérdida es asimismo igual a M/N.

2.3. Independencia de sucesos. En un modelo matemático el concepto de independencia es cómodo introducirlo con ayuda del concepto de probabilidad condicional. Diremos que el suceso B no depende del suceso A si

$$\mathbf{P}(B \mid A) = \mathbf{P}(B), \quad \mathbf{P}(A) > 0. \tag{10}$$

Si $\mathbf{P}(B) > 0$, de las igualdades (10) y (2) resulta que $\mathbf{P}(A \mid B) = \mathbf{P}(A)$. Por lo tanto, si $\mathbf{P}(A) > 0$ y $\mathbf{P}(B) > 0$, el concepto de independencia de dos sucesos es simétrico, o sea, si B no depende de A, tampoco A depende de B. En comparación con (10) una definición más cómoda de la independencia es la siguiente.

Los sucesos A y B son independientes, si

$$\mathbf{P}(AB) = \mathbf{P}(A)\mathbf{P}(B). \tag{11}$$

Esta definición es equivalente a (10), si $\mathbf{P}(A) > 0$.

Demos la generalización de la definición (11) para varios sucesos.

Llamaremos los sucesos $A_1, A_2, \dots, A_n$, recíprocamente independientes (o independientes en conjunto, o simplemente independientes) si para todas las combinaciones de los índices $1 \leqslant i_1 < i_2 < \dots < i_k \leqslant n$ $(k = 2, \dots, n)$ tenemos

$$\mathbf{P}(A_{i_1}A_{i_2}\dots A_{i_k}) = \mathbf{P}(A_{i_1})\mathbf{P}(A_{i_2})\dots\mathbf{P}(A_{i_k}). \qquad (12)$$

Si (12) se cumple solamente para $k = 2$, los sucesos se llaman *independientes dos a dos.* Notemos que de la independencia dos a dos no se deduce la independencia recíproca (véase el problema 4).

Ahora examinemos el nexo entre la independencia teórico-probabilística introducida por las igualdades (11) y (12) y la independencia causal de los sucesos reales. Supongamos que al realizar n observaciones los sucesos A, B y AB se han producido n_A, n_B y n_{AB} veces, respectivamente. De la propiedad de estabilidad de las frecuencias se deducen las igualdades aproximadas:

$$\frac{n_A}{n} \approx \mathbf{P}(A), \quad \frac{n_B}{n} \approx \mathbf{P}(B), \quad \frac{n_{AB}}{n} \approx \mathbf{P}(AB),$$

$$\frac{n_{AB}}{n_A} \approx \mathbf{P}(A \mid B) = \frac{\mathbf{P}(AB)}{\mathbf{P}(B)}.$$

Por consiguiente, para los sucesos A y B, que son independientes en el sentido teórico-probabilístico, de la igualdad $\mathbf{P}(A \mid B) = \mathbf{P}(A)$ conviene esperar el cumplimiento de la igualdad aproximada $n_{AB}/n_B \approx n_A/n$ ó, lo que es equivalente,

$$\frac{n_{AB}}{n} \approx \frac{n_A}{n}\frac{n_B}{n}.$$

Esta propiedad está determinada para los sucesos A y B de independencia causal por una práctica de larga duración.

Al construir un modelo matemático se utiliza con frecuencia el siguiente principio: *los sucesos de independencia causal son independientes en el sentido teórico-probabilístico.* Notemos que este principio no puede ser teorema, puesto que no está formulado en los términos del modelo matemático. Las igualdades (11) y (12) no suelen usarse para establecer la independencia, sino para determinar los productos de los sucesos al construir un modelo matemático de las pruebas cuyos casos son los sucesos de independencia causal. Tales construcciones se realizan en el § 3.

2.4. Aplicación de la fórmula de la probabilidad completa.

Vamos a examinar un ejemplo que da cierta idea sobre la utilización de la fórmula de la probabilidad completa al investigar procesos aleatorios con tiempo ininterrumpido.

EJEMPLO 5. La probabilidad de que a una línea telefónica llegue una llamada en el período $(t, t+h)$ es igual a $\alpha h + o(h)$, $h \to 0$; la probabilidad de que ninguna llamada llegue en el período $(t, t+h)$ es igual a $1 - \alpha h + o(h)$. Si la línea está ocupada, la llamada se pierde. Si en el instante t continúa la conversación, entonces en el período $(t, t+h)$ ésta terminará con la probabilidad $\beta h + o(h)$, $h \to 0$. Las llamadas llegan independientemente una de la otra. Hallemos la probabilidad $P_0(t)$ de que en el instante t la línea esté libre y la probabilidad $P_1(t)$ de que la línea quede ocupada.

Supongamos, a la vez, que el espacio probabilístico conveniente existe y que para los sucesos que nos interesan

$$B_t^{(0)} = \{\text{en el instante } t \text{ la línea está libre}\},$$
$$B_t^{(1)} = \{\text{en el instante } t \text{ la línea queda ocupada}\}$$

a cada t las probabilidades $P_0(t) = \mathbf{P}(B_t^{(0)})$, $P_1(t) = \mathbf{P}(B_t^{(1)})$ están determinadas y son continuas en t. Puesto que

$$B_t^{(0)} + B_t^{(1)} = \Omega, \quad B_t^{(0)}B_t^{(1)} = \varnothing,$$

entonces por la fórmula de la probabilidad completa (11)

$$\mathbf{P}(B_{t+h}^{(0)}) = \mathbf{P}(B_t^{(0)})\mathbf{P}(B_{t+h}^{(0)} \mid B_t^{(0)}) + \mathbf{P}(B_t^{(1)})\,\mathbf{P}(B_{t+h}^{(0)} \mid B_t^{(1)}). \tag{13}$$

La línea, libre en el instante t, quedará libre en el instante $t + h$, si en el período $(t, t+h)$ no lleguen las llamadas. Puesto que otros sucesos, para los cuales la línea quedará libre en el instante $t + h$, tienen la probabilidad $o(h)$, entonces

$$\mathbf{P}(B_{t+h}^{(0)} \mid B_t^{(0)}) = 1 - \alpha h + o(h).$$

La línea, ocupada en el instante t, estará libre para el instante $t + h$, si la conversación termina y en el periodo $(t, t+h)$ no llegarán nuevas llamadas. La probabilidad de este suceso es igual a

$$(\beta h + o(h))(1 - \alpha h + o(h)) = \beta h + o(h).$$

Esta probabilidad da el aporte principal en $\mathbf{P}(B_{t+h}^{(0)} \mid B_t^{(1)})$. La suma de los demás sumandos es igual a $o(h)$. Así, pues,

$$\mathbf{P}(B_{t+h}^{(0)} \mid B_t^{(1)}) = \beta h + o(h).$$

Sustituyendo las probabilidades condicionales encontradas en (13) y utilizando las designaciones $P_0(t)$ y $P_1(t)$, obtenemos

$$P_0(t + h) = (1 - \alpha h)P_0(t) + \beta h P_1(t) + o(h).$$

De aquí

$$\frac{P_0(t + h) - P_0(t)}{h} = -\alpha P_0(t) + \beta P_1(t) + o(1).$$

Pasando al límite cuando $h \to 0$, obtenemos

$$\frac{dP_0(t)}{dt} = -\alpha P_0(t) + \beta P_1(t). \tag{14}$$

Análogamente hallamos la ecuación para $P_1(t)$:

$$\frac{dP_1(t)}{dt} = \alpha P_0(t) - \beta P_1(t). \tag{15}$$

Suponiendo $P_0(0) = 1$ y $P_1(0) = 0$, encontramos la solución del sistema (14)—(15):

$$P_0(t) = \frac{\beta}{\alpha + \beta} + \frac{\alpha}{\alpha + \beta} e^{-(\alpha + \beta)t},$$

$$P_1(t) = \frac{\beta}{\alpha - \beta} - \frac{\alpha}{\alpha + \beta} e^{-(\alpha + \beta)t}. \tag{16}$$

Cuando $t \to \infty$, la probabilidad $P_1(t)$ tiende hacia la magnitud $\alpha/(\alpha + \beta)$. Es natural que con el crecimiento de la intensidad de las llamadas α que llegan, la probabilidad de que la línea quede ocupada aumenta.

Problemas

1. Se han lanzado dos dados. ¿Cuál es la probabilidad de que salgan dos "3", si se sabe que la suma de los puntos obtenidos se divide por tres?

2. En el conjunto de números $\{1, 2, \ldots, N\}$ se escogen tres números según el esquema de selección aleatoria sin devolución. Hallar la probabilidad de que el tercer número se encuentre en el intervalo formado por los dos primeros, si se conoce que el primer número es menor que el segundo.

3. Demostrar que los sucesos A y $\bar{B}$ son independientes, si son independientes los sucesos A y B.

4. Un punto aleatorio (ξ_1, ξ_2) tiene la distribución uniforme en el cuadrado $\{(x_1, x_2): 0 \leqslant x_1 \leqslant 1, 0 \leqslant x_2 \leqslant 1\}$. Sea

$$C_1 = \{\xi_1 \leqslant 1/2\},\ C_2 = \{\xi_2 \leqslant 1/2\} \quad C_3 = \{(\xi_1 - 1/2)\cdot(\xi_2 - 1/2) < 0\}.$$

Mostrar que los sucesos C_1, C_2, C_3 son independientes dos a dos. ¿Son los sucesos C_1, C_2, C_3 independientes recíprocamente? ¿Son dependientes los sucesos C_1C_2 y C_3?

5. Los sucesos A_1, A_2, A_3, A_4 son · recíprocamente independientes; $\mathbf{P}(A_k) = p_k$, $k = 1, 2, 3, 4$. Hallar las probabilidades de los sucesos: 1) $A_1\bar{A}_3A_4$; 2) $A_1 + A_2$; 3) $(A_1 + A_2)(A_3 + A_4)$.

6. Un circuito eléctrico se compone de elementos A_k, $k = 1, 2, 3$, según el esquema representado en la fig. 5. Si se avería un elemento cualquiera, en el lugar de su conexión el circuito se rompe. La probabilidad de que durante el período dado falle el elemento A_k es igual a p_k, $k = 1, 2, 3$. Se supone que los elementos fallan o no fallan independientemente uno del otro. Hallar la probabilidad de que en el período que se examina por el circuito pase la corriente.

7. Un sistema simplificado para controlar los artículos consiste en dos comprobaciones independientes. Como resultado de la k-ésima comprobación ($k = 1, 2$) un artículo que satisface la norma se desecha con la probabilidad β_k y un artículo defectuoso se recibe con la probabilidad α_k. Un artículo se recibe, si pasa ambas comprobaciones. Hallar la probabilidad de los sucesos:

1) un artículo defectuoso quedará recibido,

2) un artículo que satisface la norma quedará desechado.

8. La primera urna contiene 2 bolas blancas y 4 bolas negras y la segunda, 3 blancas y 1 negra. Dos bolas pasan de la primera urna a la segunda. Hallar la probabilidad de que la bola extraída de la segunda urna, después de pasar a ella dos bolas de la primera, sea blanca.

9. Los artículos se suministran para la comprobación descrita en el problema 7. Suponiendo que cada artículo satisface la norma con la probabilidad p, hallar las probabilidades siguientes:

1) la probabilidad de que un artículo enviado para la comprobación no quede desechado;

2) la probabilidad de que un artículo no desechado satisface la norma.

10. Supongamos que el 5% de todos los hombres y el 0,25% de todas las mujeres son daltonianos. Una persona escogida a la ventura resulta ser daltoniana.

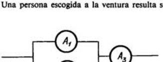

Fig. 5.

¿Cuál es la probabilidad de que esta persona sea un hombre? (Se considera que la cantidad de los hombres y las mujeres es igual.)

11. Por un canal de comunicación puede ser transmitida una de las tres sucesiones de letras: AAAA, BBBB, CCCC; es sabido que las probabilidades de cada una de las sucesiones son iguales a 0,3; 0,4; 0,3, respectivamente. Debido a los ruidos una letra se recibe correctamente con probabilidad 0,6. Las probabilidades de que una letra transmitida se tome por dos otras son iguales a 0,2 y 0,2. Se supone que las letras se distorsionan independientemente una de la otra. Hallar la probabilidad de que se haya transmitido la sucesión AAAA si en el dispositivo de recepción se recibe ABCA.

12. La probabilidad de que una molécula que ha chocado en el instante $t = 0$ con otra molécula y que no tenía otros choques hasta el instante t experimente una colisión en el intervalo de tiempo $(t, t + h)$ es igual a $\lambda h + o(h)$, $h \to 0$. Hallar la probabilidad de que el tiempo del recorrido libre sea mayor que t.

13. Variemos las condiciones de trabajo de la línea telefónica descrita en el ejemplo 5. Supondremos que si la línea esté ocupada las llamadas no se pierden, sino se ponen en la cola. Designemos con $P_k(t)$ la probabilidad de que en el instante t una llamada se reciba y $k - 1$ llamadas formen la cola ($k \geqslant 1$); $P_0(t)$ es la probabilidad de que la línea esté libre. Plantear para $P_k(t)$ las ecuaciones diferenciales. Hallar $\lim\limits_{t \to \infty} P_k(t) = \pi_k$, si $\theta = \alpha/\beta < 1$.

§ 3. Sucesión de pruebas

3.1. Definición de la sucesión de pruebas independientes. Se llama generalmente prueba (observación, experimento) el conjunto de ciertas condiciones cuando pueden producirse algunos sucesos aleatorios. Como resultado de cada prueba aparece uno de algunos sucesos incompatibles dos a dos que denominaremos casos. Por ejemplo, la prueba consistente en el control de un artículo acabado puede terminar en uno de los dos casos: un artículo será útil o defectuoso.

Describiremos un modelo teórico-probabilístico de la sucesión de pruebas independientes. Supongamos que en cada prueba puede producirse uno de r casos, $1, 2, \ldots, r$, y los sucesos relacionados con diferentes pruebas son de independencia causal. Entonces en el modelo matemático que vamos a construir los sucesos respectivos deben ser independientes en el sentido teórico-probabilístico (véase el § 12, (2)). El resultado de n pruebas se puede escribir en forma de la cadena $x_1, x_2, \ldots, x_n$, donde x_t es el caso de la t-ésima prueba. Por el conjunto Ω se puede tomar el conjunto de todas las cadenas posibles $\omega = (x_1, x_2, \ldots, x_n)$. Ahora bien,

$$\Omega = \{\omega\}, \quad \omega = (x_1, x_2, \ldots, x_n),$$
$$x_t \in \{1, 2, \ldots, r\}, \quad t = 1, 2, \ldots, n. \tag{1}$$

El suceso $A_i(t)$ = {en la t-ésima prueba se ha producido el caso i} ahora puede ser expresado por sucesos elementales como subconjunto Ω:

$$A_i(t) = \{\omega = (x_1, x_2, \dots, x_n): x_t = i\},$$
$$t = 1, \dots, n, \quad i = 1, \dots, r.$$

Por otro lado, el suceso elemental $\omega = (x_1, \dots, x_n)$ se representa como producto de los sucesos $A_i(t)$:

$$\omega = (x_1, \dots, x_n) = A_{x_1}(1)A_{x_2}(2) \dots A_{x_n}(n). \tag{2}$$

A partir de las fórmulas (2) y (12) del § 2, las probabilidades elementales $p(\omega)$ se determinan mediante la igualdad:

$$p(\omega) = p_{x_1}p_{x_2} \cdots p_{x_n}, \tag{3}$$

donde las probabilidades de los casos de diversas pruebas $p_k = \mathbf{P}(A_k(t))$, $t = 1, \dots, n$, $k = 1, \dots, r$, satisfacen las probabilidades

$$p_k \geqslant 0, \quad k = 1, \dots, r; \qquad \sum_{k=1}^{r} p_k = 1. \tag{4}$$

Se llama sucesión de pruebas independientes el esquema probabilístico finito, en el cual las probabilidades de sucesos elementales se determinan por la fórmula (3) *como productos de probabilidades de los casos de diferentes pruebas.*

La sucesión de pruebas independientes la llamaremos, a veces, *esquema de pruebas independientes* o bien *esquema polinomial.*

Con frecuencia se tienen que examinar las sucesiones de pruebas con dos casos: un artículo es útil o defectuoso; con un billete de lotería se recibe un premio o no; en el periodo de tiempo que se examina un aparato ha funcionado normalmente o ha fallado, etc.

Un esquema particular de pruebas independientes, en el cual cada prueba no puede terminar, sino que en uno de dos casos se denomina esquema de Bernoulli.

Estos casos suelen llamarse "favorables" y "desfavorables" y sus probabilidades se designan con p y $q = 1 - p$ $(0 \leqslant p \leqslant 1)$, respectivamente. Para brevedad, los casos favorables y desfavorables los designaremos con Y y H o bien 1 y 0, respectivamente. En el esquema de Bernoulli con n pruebas tenemos

$$\Omega = \{\omega\}, \quad \omega = (x_1, \dots, x_n), x_t \in \{0; 1\},$$
$$t = 1, \dots, n. \tag{5}$$

Es evidente que el número de casos favorables (o el número "1") en la cadena $\omega = (x_1, x_2, \dots, x_n)$ es igual a la suma $x_1 + x_2 + \dots + x_n$. Las probabilidades elementales definidas por la fórmula (3) tienen para el esquema de Bernoulli el siguiente aspecto:

$$p(\omega) = p^{x_1 + x_2 + \dots + x_n} q^{n - (x_1 + \dots + x_n)},$$
$$\omega = (x_1, \dots, x_n). \tag{6}$$

A menudo para el esquema de Bernoulli representa interés el suceso.

B_m = {en n pruebas se han producido exactamente m casos favorables}.

TEOREMA 1. *La probabilidad $\mathbf{P}(B_m)$ de que en n pruebas del esquema de Bernoulli se produzcan exactamente m casos favorables se determina por medio de la fórmula*

$$\mathbf{P}(B_m) = C_n^m p^m q^{n-m}, \quad m = 0, 1, \dots, n, \tag{7}$$

donde p es la probabilidad del caso favorable en una prueba.

DEMOSTRACIÓN. El suceso B_m se determina como subconjunto Ω:

$$B_m = \{\omega = (x_1, \dots, x_n): x_1 + x_2 + \dots + x_n = m\}.$$

Por lo tanto, $\mathbf{P}(B_m) = \sum_{\omega \in B_m} p(\omega)$. Puesto que según la fórmula (6) las probabilidades elementales $p(\omega) = p^m q^{n-m}$ con cualquier valor de $\omega \in B_m$, entonces para calcular la probabilidad queda determinar el número $|B_m|$ de los sucesos elementales que entran en B_m. El número $|B_m|$ coincide, evidentemente, con el número de los procedimientos de escoger m lugares para "1" en la cadena ω, ya que los lugares quedados se llenan unívocamente de "0". De aquí $|B_m| = C_n^m$ y $\mathbf{P}(B_m) = C_n^m p^m q^{n-m}$.

EJEMPLO 1. Un sistema compuesto por n bloques funciona irreprochablemente, si durante el período que se examina no fallan más que dos bloques. Hallar la probabilidad de que el sistema quede en buen estado suponiendo que los fallos de los bloques son sucesos independientes y la probabilidad de que falle cada bloque es igual a p.

En calidad de modelo hacemos uso del esquema de Bernoulli con n pruebas. Cada prueba consiste en el funcionamiento de uno de los bloques en el período en cuestión. Para comodidad de la utilización de la fórmula (7) llamaremos "caso favorable" la avería de un bloque. Sea A = {el sistema funciona irreprochablemente}. Entonces $A = B_0 + B_1 + B_2$, donde B_m = {la avería de m bloques}. De aquí, valiéndose de la fórmula (7), obtenemos

$$\mathbf{P}(A) = \mathbf{P}(B_0) + \mathbf{P}(B_1) + \mathbf{P}(B_2) =$$
$$= q^n + npq^{n-1} + \frac{n(n-1)}{2}p^2q^{n-2}.$$

EJEMPLO 2. Un dado se arroja 10 veces. Hallar la probabilidad $\mathbf{P}(B_3)$ de que salgan exactamente los tres "6". En la fórmula (7) hace falta poner $n = 10$, $m = 3$, $p = 1/6$, $q = 5/6$. Entonces

$$\mathbf{P}(B_3) = C^3_{10}\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^7 = \frac{10\cdot 9\cdot 8\cdot 5^7}{1\cdot 2\cdot 3\cdot 6^{10}} = 0{,}155045\ldots$$

Obtenemos la generalización de la fórmula (7) para un esquema polinomial.

TEOREMA 2. *La probabilidad $P_n(m_1, \ldots, m_r)$ de que en las pruebas de un esquema polinomial el caso "1" se produzca m_1 veces, el caso "2", m_2 veces, ..., el caso "r" m_r veces se determina por la igualdad*

$$P_n(m_1, \ldots, m_r) = \frac{n!}{m_1!m_2!\ldots m_r!}p_1^{m_1}p_2^{m_2}\ldots p_r^{m_r}, \tag{8}$$

donde p_k es la probabilidad del caso k en una prueba ($k = 1, \ldots, r$); m_1, ..., m_r son los números enteros no negativos que satisfacen la igualdad $m_1 + m_2 + \ldots + m_r = n$.

DEMOSTRACIÓN. Sea B = {entre n pruebas el caso "1" se produce m_1 veces, ..., el caso "r" se produce m_r veces}. Para todo suceso elemental $\omega \in B$ tenemos $p(\omega) = p_1^{m_1}p_2^{m_2}\ldots, p_r^{m_r}$. El número $|B|$ de los sucesos elementales que forman parte de B no es difícil calcularlo. El caso "1" en m_1 lugares de la cadena $\omega = (x_1, \ldots, x_r)$ se puede disponer por $C_n^{m_1}$ procedimientos; el caso "2" se puede disponer en los $n - m_1$ lugares quedados por $C_{n-m_1}^{m_2}$ procedimientos, etc. Así, pues,

$$|B| = C_n^{m_1}C_{n-m_1}^{m_2}\ldots C_{n-m_1-\ldots-m_{r-1}}^{m_r} = \frac{n!}{m_1!m_2!\ldots m_r!}.$$

Por consiguiente,

$$\mathbf{P}(B) = \sum_{\omega\in B} p(\omega) = |B|p_1^{m_1}\ldots p_r^{m_r} = \frac{n!}{m_1!\ldots m_r!}p_1^{m_1}\ldots p_r^{m_r}.$$

El teorema queda demostrado.

La fórmula (7) se obtiene de la fórmula (8) si se pone $r = 2$, $p_1 = p$, $p_2 = 1 - p = q$.

Vamos a examinar algunos casos particulares importantes del esquema de pruebas independientes.

Números aleatorios. Examinemos una sucesión de pruebas independientes con $r = 10$ casos que designemos con 1, 2, ... , 9, 0. Las probabilidades de estos casos en diferentes pruebas se designan con $p_1, p_2, \dots, p_{10}$, respectivamente.

Si se pone $p_1 = p_2 = \dots = p_{10} = 1/10$, *entonces el esquema probabilístico obtenido coincide con el introducido en el § 1 para definir los números aleatorios a base de la definición clásica de la probabilidad.*

EJEMPLO 3. Hallar la probabilidad $P_{20}(10, 2, 3, 5)$ de que entre 20 números aleatorios hay exactamente 10 cifras pares, dos "tripletes" y tres "septenas". Para calcular la probabilidad indicada es cómodo reducir el esquema finito que determina los números aleatorios a la sucesión de pruebas independientes en cada una de las cuales es posible la producción de uno de los cuatro casos posibles: 1) la cifra par, 2) el triplete, 3) la septena y 4) todo lo demás. Las probabilidades de estos casos son iguales a $p_1 = 5/10, p_2 = p_3 = 1/10$ y $p_4 = 3/10$, respectivamente. Según la fórmula (8) obtenemos

$$P_{20}(10, 2, 3, 5) = \frac{20!}{10!2!3!5!}\left(\frac{1}{2}\right)^{10}\left(\frac{1}{10}\right)^{5}\left(\frac{3}{10}\right)^{5}.$$

En una sucesión de pruebas independientes n pruebas se interpretan, a veces, como disposición de n partículas en $r = N$ células; la probabilidad $p_i(i = 1, \dots, N)$ es la probabilidad de que una partícula se encuentre en la i-ésima célula. Si $p_1 = p_2 = \dots = p_N = 1/N$ el esquema de disposición de las partículas en las células se llama *equiprobable* (a veces se trata de la disposición de "mostacillas en los cajones" en vez de la de "partículas en las células").

Entre los problemas de disposición figuran también diferentes problemas "de cumpleaños".

EJEMPLO 4. En un grupo hay 25 estudiantes. Calculemos la probabilidad de que existan al menos dos estudiantes, cuyos cumpleaños coincidan. Vamos a suponer que los cumpleaños son independientes y el hecho de que un cumpleaños tenga lugar en un día determinado es igual a 1/365. Sea que el suceso $A = \{$hay, por lo menos, dos estudiantes, cuyos cumpleaños coinciidan$\}$. Entonces $\bar{A} = \{$todos los cumpleaños tienen lugar en diferentes días del año$\}$. En el problema dado los sucesos elementales ω son las cadenas $\omega = (x_1, x_2, \dots, x_{25})$, donde x_1 es el día del año en que se cumple el aniversario del nacimiento del primer estudiante; x_2 es el día en que se cumple el aniversario del nacimiento del segundo estudiante, etc. Entonces, por definición del esquema de pruebas independientes, para todo valor de $p(\omega) = (1/365)^{25}$ y $\mathbf{P}(\bar{A}) = |\bar{A}|/365^{25}$, donde $|\bar{A}|$ es el número de los sucesos elementales que forman parte de $\bar{A}$. Puesto que para ω que entran en $\bar{A}$ entre $x_1, \ldots, x_{25}$ no hay tales que coincidan, el valor de x_1 se puede elegir por 365 métodos; el valor de x_2, por 364 métodos (un día cualquiera a excepción del elegido para el cumpleaños del primer estudiante: el valor de x_3 se elige por 363 procedimientos, etc. Como resultado encontramos $|\bar{A}| = 365 \cdot 364 \cdot \ldots \cdot 341$. De aquí que $\mathbf{P}(\bar{A}) = ((365 \cdot 364 \cdot \ldots \ldots \cdot 341)/365^{25} = 0{,}43130 \ldots$ y, por lo tanto,

$$\mathbf{P}(A) = 1 - \mathbf{P}(\bar{A}) = 0{,}56869 \ldots$$

3.2. Definición general de la sucesión de pruebas. Aduzcamos la descripción general de un modelo teórico probabilístico de la sucesión de pruebas (no obligatoriamente independientes) en los términos de variación del estado de cierta "partícula" (o de un sistema físico). Supongamos que una partícula puede encontrarse en uno de r estados. Examinemos n estados sucesivos de la partícula. Tomemos por conjunto $\Omega = \{\omega\}$ el conjunto de todas las trayectorias posibles de la partícula $\omega = (x_1, x_2, \ldots, x_n)$, donde x_t es el estado de la partícula en el instante t, $t = 1, 2, \ldots, n$. Por lo tanto, tenemos

$$\Omega = \{\omega\}, \ \omega = (x_1, \ldots, x_n), \quad x_t \in \{1, 2, \ldots, r\}, \quad t = 1, \ldots, n. \tag{9}$$

Si se conocen las probabilidades condicionales $P_{t+1}(x_{t+1} \mid x_1, \ldots, x_t)$ de que en el instante $t + 1$ la partícula se encuentre en el estado x_{t+1} a condición de que en los instantes $1, 2, \ldots, t$ sus estados sean $x_1, x_2, \ldots, x_t$, entonces de acuerdo con la fórmula (6) del § 2 la probabilidad del suceso elemental $\omega = (x_1, \ldots, x_n)$ se determina unívocamente:

$$p(\omega) = p_1(x_1) p_2(x_2 \mid x_1) p_3(x_3 \mid x_1, x_2) \ldots \ldots p_n(x_n \mid x_1, \ldots, x_{n-1}). \tag{10}$$

Aquí $p_1(x_1)$ es la probabilidad de que en el instante $t = 1$ la partícula se encuentre en el estado x_1.

El esquema probabilístico finito, en el cual el conjunto Ω se determina por la igualdad (9) *y las probabilidades elementales según la fórmula* (10) *se llama sucesión de n pruebas con r casos.*

Para un esquema de pruebas sucesivas la descripción del conjunto Ω en forma de un conjunto de trayectorias $\omega = (x_1, \ldots, x_n)$ de cierta partícula es cómoda, puesto que se distingue por su claridad. Esta descripción figurada se puede aplicar a los esquemas más diferentes que no tienen nada que ver con las "partículas" y sus "estados". Por ejemplo, en los términos de probabilidades condicionales se define de un modo natural el esquema de selección aleatoria sin devolución de n elementos entre N elementos (véase el § 1, ejemplo 5). En este caso los sucesos elementales $\omega = (x_1, x_2, \dots, x_n)$ se determinan como variaciones de N elementos n a n y las probabilidades condicionales $p_t(x_t \mid x_1, \dots, x_{t-1})$ se determinan del modo siguiente: si salen t bolas, entonces en la prueba siguiente cualquiera de las $N - t$ bolas quedadas se extrae con la probabilidad $1/(N - t)$. Así, pues, en este caso $p_1(x_1) = \frac{1}{N}$, $p_t(x_t \mid x_1, \dots, x_{t-1}) = \frac{1}{N - t}$ y la definición de las probabilidades elementales

$$p(\omega) = p_1(x_1)p_2(x_2 \mid x_1)p_3(x_3 \mid x_1, x_2) \dots p_n(x_n \mid x_1, \dots, x_{n-1}) =$$
$$= \frac{1}{N}\frac{1}{N-1} \dots \frac{1}{N-n+1} = \frac{1}{N^{[n]}}$$

coincide con la dada en el § 1 (ejemplo 5).

EJEMPLO 5. Entre n llaves es sólo una que abre la puerta. Según el esquema de selección aleatoria sin devolución las llaves se extraen hasta que aparezca la llave necesaria. Hallar la probabilidad de los sucesos A_k = {la llave requerida aparece por primera vez en la k-ésima prueba}.

Hacemos uso del esquema de selección aleatoria sin devolución, definido en los términos de las probabilidades condicionales. Como $A_k \subset \bar{A}_i$ cuando $i \le k$, entonces el suceso A_k se puede representar en la forma $A_k = \bar{A}_1\bar{A}_2 \dots \bar{A}_{k-1}A_k$. Entonces

$$\mathbf{P}(A_k) = \mathbf{P}(\bar{A}_1 \dots \bar{A}_{k-1}A_k) = \mathbf{P}(\bar{A}_1)\mathbf{P}(\bar{A}_2 \mid \bar{A}_1) \times$$
$$\times \mathbf{P}(\bar{A}_3 \mid \bar{A}_1\bar{A}_2) \dots \mathbf{P}(\bar{A}_{k-1} \mid \bar{A}_1 \dots \bar{A}_{k-2}) \times$$
$$\times \mathbf{P}(A_k \mid \bar{A}_1\bar{A}_2 \dots \bar{A}_{k-1}).$$

No es difícil hallar las probabilidades condicionales que entran en esta fórmula, ya que según los datos del problema las probabilidades de que aparezca cualquiera de las bolas quedadas para la prueba dada son iguales. Ahora bien,

$$\mathbf{P}(\bar{A}_1) = \frac{n-1}{n}, \quad \mathbf{P}(\bar{A}_2 \mid \bar{A}_1) = \frac{n-2}{n-1}, \dots$$
$$\dots, \mathbf{P}(\bar{A}_{k-1} \mid \bar{A}_1 \dots \bar{A}_{k-2}) = \frac{n-k+1}{n-k+2},$$
$$\mathbf{P}(A_k \mid \bar{A}_1 \dots \bar{A}_{k-1}) = \frac{1}{n-k+1}$$

y, por consiguiente,

$$\mathbf{P}(A_k) = \frac{n-1}{n} \cdot \frac{n-2}{n-1} \cdot \frac{n-3}{n-2} \cdot \ldots \cdot \frac{n-k+1}{n-k+2} \cdot \frac{1}{n-k+1} = \frac{1}{n}.$$

El esquema general de pruebas contiene, en calidad de un caso particular, el de pruebas independientes, en el cual

$$p_t(x_t \mid x_1, x_2, \ldots, x_{t-1}) = p_t(x_t), \quad t = 1, \ldots, n. \tag{11}$$

Este esquema de pruebas independientes es natural llamarlo *no homogéneo* a distinción del esquema de pruebas independientes introducido en el punto 3.1, ya que las probabilidades de los casos en el punto 3.1 no dependen del número de orden de las pruebas t.

En conclusión cabe indicar un otro importante caso particular del esquema general de la sucesión de pruebas. Si las probabilidades condicionales $p_t(x_t \mid x_1, \ldots, x_{t-1})$ no dependen sino que del resultado de la última prueba, o sea,

$$p_t(x_t \mid x_1, \ldots, x_{t-1}) = p_t(x_t \mid x_{t-1}),$$

entonces la sucesión de pruebas se denomina *cadena de Márkov*.

Problemas

1. Dos jugadores extraen por turno las bolas (sin devolución) de una urna que contiene 2 bolas blancas y 4 negras. Gana el primero en sacar una bola blanca. Hallar las probabilidades de que gane uno u otro participante.

2. En la transmisión de una comunicación la probabilidad de que se distorsione un signo es igual a 1/100. Suponiendo que la distorsión de los signos es independiente, hallar la probabilidad de que una comunicación compuesta por 5 signos: a) no se someta a una distorsión; b) contenga exactamente una distorsión; c) contenga al menos dos distorsiones.

3. En una lotería la probabilidad de que un billete sea premiado es igual a 1/5. Suponiendo que los premios que caen en los diferentes billetes son independientes, determinar el número de los billetes que se necesitan comprar para que la probabilidad $Q(n)$ de obtener al menos un premio sea no menor que 0,9. Hallar $Q(n)$ para este n.

4. Cada una de 10 pruebas independientes consiste en arrojar tres dados. Hallar la probabilidad de que en cuatro pruebas aparezcan cada vez exactamente los dos "6".

§ 4. Teoremas del límite en el esquema de Bernoulli

4.1. Teorema de Poisson. En las aplicaciones se necesita con frecuencia calcular las probabilidades de diferentes sucesos relacionados con el número de casos favorables μ_n en n pruebas de Bernoulli, cuando los valores de n son grandes. En este caso llega a ser difícil realizar los cálculos con ayuda de la fórmula (7) del § 3. Las dificultades aparecen también, si se deben sumar las probabilidades de los sucesos $\{\mu_n = m\}$. Se reduce a la sumación el cálculo de las probabilidades de los sucesos $\{a < \mu_n < b\}$. En efecto,

$$\mathbf{P}\{a < \mu_n < b\} = \sum_{a < m < b} C_n^m p^m q^{n-m}. \tag{1}$$

Las dificultades durante los cálculos surgen también cuando los valores de p o q son pequeños.

A veces, cuando los valores de n son grandes, se logra sustituir la fórmula (7) del § 3 por cualquier fórmula asintótica aproximada. Citemos tres teoremas límites que contienen las fórmulas asintóticas para las probabilidades $\mathbf{P}\{\mu_n = m\}$ y $\mathbf{P}\{a < \mu_n < b\}$, cuando $n \to \infty$.

TEOREMA 1 (teorema de Poisson). *Si $n \to \infty$ y $p \to 0$ de modo que $np \to \lambda$, $0 < \lambda < \infty$, entonces*

$$\mathbf{P}\{\mu_n = m\} = P_n(m) = C_n^m p^m q^{n-m} \to p_m(\lambda) = \frac{\lambda^m}{m!} e^{-\lambda}$$

para todo valor constante de m, $m = 0, 1, 2, \ldots$

DEMOSTRACIÓN. Suponiendo $np = \lambda_n$ representemos la probabilidad $P_n(m)$ en forma

$$\mathbf{P}\{\mu_n = m\} = \frac{n(n-1)\ldots(n-m+1)}{m!}\left(\frac{\lambda_n}{n}\right)^m \times$$

$$\times \left(1 - \frac{\lambda_n}{n}\right)^{n-m} = \frac{\lambda_n^m}{m!}\left(1 - \frac{\lambda_n}{n}\right)^n \left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\ldots$$

$$\ldots\left(1 - \frac{m-1}{n}\right)\left(1 - \frac{\lambda_n}{n}\right)^{-m}$$

De aquí, cuando $n \to \infty$, obtenemos la afirmación del teorema.

Ahora bien, siempre que los valores de n sean grandes y los de p, pequeños, podemos hacer uso de la fórmula aproximada

$$\mathbf{P}\{\mu_n = m\} \approx \frac{\lambda_n^m}{m!} e^{-\lambda_n}, \quad \lambda_n = np. \tag{2}$$

Para comparar los valores exactos y aproximados citemos la tabla siguiente:

m	0	1	2	3	4	5
$p_m(np)$	0,1353	0,2707	0,2707	0,1805	0,0902	0,0361
$P_{10}(m)$	0,1074	0,2684	0,3020	0,2013	0,0880	0,0264
$P_{100}(m)$	0,1326	0,2707	0,2734	0,1823	0,0902	0,0353

Cuando $n = 10$ las probabilidades $P_n(m)$ están calculadas para $p = 1/5$ y cuando $n = 100$, para $p = 1/50$.

Si el valor de q es pequeño, la aproximación de Poisson se puede utilizar para el número de casos desfavorables.

EJEMPLO 1. En la tabla de números aleatorios las cifras están agrupadas dos a dos. Hallar el valor aproximado de la probabilidad de que entre 100 pares el par 00 se encuentre dos veces, como máximo.

Hacemos uso del esquema de Bernoulli con $n = 100$ pruebas. Si la aparición del par 00 se llama "caso favorable", hace falta poner $p = 0{,}01$, $q = 0{,}99$. El valor exacto de la probabilidad del suceso $\{\mu_{100} \leqslant 2\}$, donde μ_{100} es el número de los pares 00, se encuentra según la fórmula (7) del § 2.

$$\mathbf{P}\{\mu_{100} \leqslant 2\} = \mathbf{P}\{\mu_{100} = 0\} + \mathbf{P}\{\mu_{100} = 1\} + \mathbf{P}\{\mu_{100} = 2\} = 0{,}99^{100} + 100 \cdot 0{,}01 \cdot 0{,}99^{99} + \frac{100 \cdot 99}{2} \cdot 0{,}01^2 \cdot 0{,}99^{98}.$$

En este caso se pueden utilizar el teorema 1, la fórmula (2) con $\lambda_n = 0{,}01 \cdot 100 = 1$ y la tabla 3. Obtenemos

$$\mathbf{P}\{\mu_{100} \leqslant 2\} \approx 0{,}3679 + 0{,}3679 + 0{,}1839 = 0{,}9197.$$

4.2. Teoremas de Moivre—Laplace.

TEOREMA 2 (teorema local de Moivre—Laplace). *Sea que* $p(0 < p < 1)$ *constante,* $x_m = (m - np)/\sqrt{npq}$, $C > 0$ *una constante cualquiera. Si* $|x_m| \leqslant C$, *entonces para* $n \to \infty$

$$\mathbf{P}\{\mu_n = m\} = C_n^m p^m q^{n-m} = \frac{1}{\sqrt{2\pi}} e^{-\frac{x_m^2}{2}} \Delta x_m (1 + \varepsilon_{n,m}), \tag{3}$$

donde $\Delta x_m = 1/\sqrt{npq}$, $|\varepsilon_{n,m}| < C_1/\sqrt{n}$, $C_1 > 0$ *es cierta constante.*

Así, pues, en calidad de valores aproximados de las probabilidades $P_n(m) = P\{\mu_n = m\}$ se pueden tomar los valores

$$\varphi\left(\frac{m - np}{\sqrt{npq}}\right) \frac{1}{\sqrt{npq}},$$

donde

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

En la fig. 6 se muestran los gráficos de las funciones

$$y = \varphi\left(\frac{u - np}{\sqrt{npq}}\right) \frac{1}{\sqrt{npq}}$$

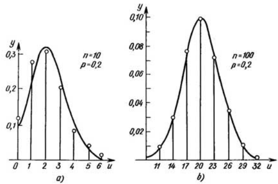

Fig. 6.

y en los puntos $u = m, m = 0, 1, 2, \ldots$ se ponen los segmentos verticales, cuyas longitudes son iguales a los valores de $P_n(m)$:

a) Los valores de $P_{10}(m)$ para $m = 0; 1; 2; 3; 4; 5; 6$ son iguales a 0,1074; 0,2684; 0,3020; 0,2013; 0,0881; 0,0264; 0,0055; los valores aproximados correspondientes son iguales a 0,0905; 0,2308; 0,3153; 0,2308; 0,0905; 0,0191; 0,0021.

b) Los valores de $P_{100}(m)$ para $m = 11; 14; 17; 20; 23; 26; 29; 32$ son iguales a 0,0069; 0,0335; 0,0789; 0,0993; 0,0720; 0,0316; 0,0088; 0,0016; los valores aproximados correspondientes valen 0,0079; 0,0324; 0,0742; 0,0997; 0,0742; 0,0324; 0,0079; 0,0011.

La demostración del teorema local se da en el suplemento.

Cabe señalar que en las hipótesis del teorema 2 a partir del hecho de que $n \to \infty$ se deduce la tendencia de m hacia la infinidad.

TEOREMA 3. (teorema integral de Moivre—Laplace). *Si en cada prueba la probabilidad de un caso favorable p, $0 < p < 1$, es constante, entonces cuando $n \to \infty$*

$$\mathbf{P}\left\{a \leqslant \frac{\mu_n - np}{\sqrt{npq}} \leqslant b\right\} - \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}}\, dx \to 0 \tag{4}$$

es uniforme sobre a, b, $-\infty \leqslant a \leqslant b \leqslant +\infty$.

DEMOSTRACIÓN. Vamos a demostrar este teorema para los valores fijos de a y b, $-\infty < a \leqslant b < +\infty$*. Es evidente que la probabilidad del suceso $\left\{a \leqslant \dfrac{\mu_n - np}{\sqrt{npq}} \leqslant b\right\}$ se puede representar en forma

$$\mathbf{P}\left\{a \leqslant \frac{\mu_n - np}{\sqrt{npq}} \leqslant b\right\} = P_n(a, b) = \sum_{a \leqslant x_m \leqslant b} P_n(m), \tag{5}$$

donde $x_m = \dfrac{m - np}{\sqrt{npq}}$ y la sumación se realiza con los mismos valores de m para los cuales $x_m \in [a, b]$. Aplicando a los sumandos (5) el teorema local, obtenemos

$$P_n(a, b) = S_n + T_n, \tag{6}$$

---

* La demostración de la convergencia uniforme sobre a y b se da, por ejemplo, en [4].

donde

$$S_n = \sum_{x_m \in [a, b]} \frac{1}{\sqrt{2\pi}} e^{-\frac{x_m^2}{2}} \Delta x_m,$$

$$T_n = \sum_{x_m \in [a, b]} \frac{1}{\sqrt{2\pi}} e^{-\frac{x_m^2}{2}} \Delta x_m \varepsilon_{n, m}.$$

Nótese que S_n se distingue no más que por dos sumandos de una suma integral convenientemente elegida que corresponda a la integral

$\frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx$. Por lo tanto,

$$\lim_{n \to \infty} S_n = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx. \tag{7}$$

Para T_n tenemos

$$|T_n| = \sum_{x_m \in [a, b]} \frac{1}{\sqrt{2\pi}} e^{-\frac{x_m^2}{2}} \Delta x_m \varepsilon_{n, m} \leqslant$$

$$\leqslant \sum_{x_m \in [a, b]} \Delta x_m \varepsilon_{n, m} = O\left(\frac{1}{\sqrt{n}}\right).$$

Por lo tanto, cuando $n \to \infty$

$$T_n \to 0. \tag{8}$$

La afirmación del teorema para los valores constantes de a, b, $-\infty < a \leqslant b < +\infty$, se deduce de las fórmulas (5)—(8).

La fórmula aproximada

$$\mathbf{P}\left\{a \leqslant \frac{\mu_n - np}{\sqrt{npq}} \leqslant b\right\} \approx \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx = \Phi(b) - \Phi(a), \tag{9}$$

donde

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt, \tag{10}$$

ofrece una buena aproximación cuando n es suficientemente grande, mientras que los valores de p y q no son muy próximos a cero. Con frecuencia la aproximación normal se utiliza cuando $npq > 20$. El valor numérico de la integral (10) se puede hallar haciendo uso de las tablas para la función

$$\Phi_0(x) = \frac{1}{\sqrt{2\pi}} \int_{0}^{x} e^{-\frac{t^2}{2}} dt. \tag{11}$$

Cuando los valores de npq son pequeños, la fórmula aproximada (9) debe ser sustituida por la siguiente (véase [13], cap. 7, § 2):

$$\mathbf{P}\left\{a \leqslant \frac{\mu_n - np}{\sqrt{npq}} \leqslant b\right\} \approx \Phi\left(b + \frac{1}{2\sqrt{npq}}\right) - \Phi\left(a - \frac{1}{2\sqrt{npq}}\right). \tag{12}$$

A continuación citemos las tablas de los valores exactos y aproximados de las probabilidades de los sucesos $\{m_1 \leqslant \mu_n \leqslant m_2\}$. Pongamos

$$P(m_1, m_2) = \mathbf{P}\{m_1 \leqslant \mu_n \leqslant m_2\} = \mathbf{P}\left\{x_1 \leqslant \frac{\mu_n - np}{\sqrt{npq}} \leqslant x_2\right\},$$

$$P_0(m_1, m_2) = \Phi(x_2) - \Phi(x_1),$$

$$P_1(m_1, m_2) = \Phi\left(x_2 + \frac{h}{2}\right) - \Phi\left(x_1 - \frac{h}{2}\right),$$

donde $h = 1/\sqrt{npq}$, $x_1 = (m_1 - np)h$, $x_2 = (m_2 - np)h$.

a) $n = 10$

p	$m_1; m_2$	0; 2	3; 5	6; 8	9; 10
	$P(m_1, m_2)$	0,6778	0,3158	0,0064	0,0000
0,2	$P_0(m_1, m_2)$	0,4429	0,2059	0,0007	0,0000
	$P_1(m_1, m_2)$	0,6316	0,3455	0,0029	0,0000
	$P(m_1, m_2)$	0,0547	0,5684	0,3662	0,0107
0,5	$P_0(m_1, m_2)$	0,0277	0,3980	0,2356	0,0057
	$P_1(m_1, m_2)$	0,0558	0,5696	0,3651	0,0133

b) $n = 100, p = 0,2$

$m_1; m_2$	0; 9	10; 15	16; 20	21; 25	26; 30	31; 40
$P(m_1, m_2)$	0,0023	0,1262	0,4310	0,3531	0,0814	0,0061
$P_0(m_1, m_2)$	0,0030	0,0994	0,3413	0,2957	0,0606	0,0030
$P_1(m_1, m_2)$	0,0041	0,1250	0,4223	0,3480	0,0795	0,0043

c) $n = 100, p = 0{,}5$

$m_1; m_2$	30; 35	36; 40	41; 45	46; 50
$P(m_1; m_2)$	0,0017	0,0267	0,1557	0,3557
$P_0(m_1, m_2)$	0,0013	0,0202	0,1227	0,2881
$P_1(m_1, m_2)$	0,0018	0,0269	0,1563	0,3558

La fórmula (9) permite también estimar la proximidad de la frecuencia y la probabilidad. Sea que p la probabilidad de un caso favorable en el esquema de Bernoulli y μ_n el número total de los casos favorables. Se llama frecuencia del caso favorable la relación μ_n/n. Vamos a evaluar la probabilidad del suceso $\{|\mu_n/n - p| < \Delta\}$. Si n es suficientemente grande, se puede hacer uso de la fórmula (9). Entonces

$$\mathbf{P}\left\{\left|\frac{\mu_n}{n} - p\right| < \Delta\right\} = \mathbf{P}\left\{-\Delta\sqrt{\frac{n}{pq}} < \frac{\mu_n - np}{\sqrt{npq}} < \Delta\sqrt{\frac{n}{pq}}\right\} \approx$$

$$\approx \frac{1}{\sqrt{2\pi}} \int_{-\Delta\sqrt{n/pq}}^{\Delta\sqrt{n/pq}} e^{-\frac{x^2}{2}}\, dx = 2\Phi_0\left(\Delta\sqrt{\frac{n}{pq}}\right), \tag{13}$$

puesto que la función $\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ es par. El valor de $\Phi_0\left(\Delta\sqrt{\frac{n}{pq}}\right)$ se halla con auyda de la tabla 1 colocada en el fin del libro.

EJEMPLO 2. Dos monedas se arrojan $n = 1000$ veces. Sea que μ_n el número de salidas de la combinación "escudo — escudo". Encontremos el valor aproximado de la probabilidad de que el número de las combinaciones "escudo — escudo" se halle entre 236 y 264, es decir, estimemos la probabilidad $\mathbf{P}\{236 < \mu_n < 264\}$.

Utilicemos el esquema de Bernoulli como modelo matemático. Llamaremos ."caso favorable" la aparición de la combinación "escudo — escudo". Entonces $p = 1/4$, $q = 3/4$. Apliquemos el teorema integral de Moivre—Laplace. Empleando el valor de $np = 250$, $\sqrt{npq} \approx 13{,}7$, escribamos el suceso $\{236 < \mu_n < 264\}$ en la forma usada en el teorema 3:

$$\{236 < \mu_n < 264\} = \left\{\frac{236 - 250}{13{,}7} < \frac{\mu_n - np}{\sqrt{npq}} < \frac{264 - 250}{13{,}7}\right\} =$$

$$= \left\{\left|\frac{\mu_n - np}{\sqrt{npq}}\right| < 1{,}02\right\}.$$

Según la fórmula (9) encontramos

$$\mathbf{P}\{236 < \mu_n < 264\} = \mathbf{P}\left\{-1{,}02 < \frac{\mu_n - np}{\sqrt{npq}} < 1{,}02\right\} \approx$$

$$\approx \frac{1}{\sqrt{2\pi}} \int_{-1{,}02}^{1{,}02} e^{-\frac{x^2}{2}}\, dx = 2\Phi_0(1{,}02) = 2 \cdot 0{,}3461 = 0{,}6922.$$

El valor de $\Phi_0(1{,}02)$ se ha determinado con ayuda de la tabla 1.

Con frecuencia aparece el problema inverso: ¿cuántas pruebas se necesita llevar a cabo para que la frecuencia μ_n/n se distinga de la probabilidad p no más que en Δ, con la probabilidad $1 - 2\alpha$ (α tiene un pequeño valor)? Es natural que en tales problemas p se considere incógnita. Entonces para escoger el valor mínimo de n, para el cual la probabilidad de desviación sea igual a $1 - 2\alpha$, hace falta de acuerdo con (13) resolver la ecuación

$$2\Phi_0\left(\Delta\sqrt{\frac{n}{pq}}\right) = 1 - 2\alpha.$$

La solución dependerá de p desconocida. Se puede liberarse de esta dependencia si se exige que

$$\mathbf{P}\left\{\left|\frac{\mu_n}{n} - p\right| < \Delta\right\} \geqslant 1 - 2\alpha.$$

Entonces, utilizando la desigualdad $pq \leqslant 1/4$, de (13) obtenemos

$$\mathbf{P}\left\{\left|\frac{\mu_n}{n} - p\right| < \Delta\right\} \approx 2\Phi_0\left(\Delta\sqrt{\frac{n}{pq}}\right) \geqslant 2\Phi_0(2\Delta\sqrt{n}) = 1 - 2\alpha,$$

y para determinar n tenemos la ecuación $\Phi_0(2\Delta\sqrt{n}) = \dfrac{1 - 2\alpha}{2}$. Según la tabla pueden ser encontrados los valores de u_α para los cuales $\Phi_0(u_\alpha) = \dfrac{1 - 2\alpha}{2}$. Entonces $2\Delta\sqrt{n} = u_\alpha$ y $n \geqslant u_\alpha^2/4\Delta^2$. Con frecuencia se utilizan los valores de 2α iguales a 0,05 y 0,01. Para estos valores los valores respectivos de u_α valen 1,960 y 2,576.

Problemas

1. Hallar una expresión aproximada de la probabilidad de que el número de salidas de "1", si un dado perfecto se lanza 12 000 veces, quede encerrado entre 1900 y 2150.
2. Suponiendo que la probabilidad del nacimiento de un niño vale 0,515, hallar la probabilidad de que entre 10 000 neonatos la cantidad de niños no sea mayor que la de niñas.
3. Un libro de 500 páginas contiene 50 errores. Utilizando el esquema de Bernoulli en calidad de modelo matemático de la distribución de los errores, estimar la probabilidad de que en una página determinada haya tres errores, como mínimo. Calcular los valores exactos y la aproximación de Poisson. Comparar los resultados.
4. La probabilidad de que se fabriquen taladros de fragilidad demasiada (defecto) vale 0,02. Los taladros se embalan en las cajas cada una de las cuales contiene 100 piezas. Al hacer uso de la ley de Poisson, determinar la probabilidad de que
   a) en una caja no haya taladros defectuosos;
   b) la cantidad de taladros defectuosos no exceda de dos.
5. ¿Cuántas cifras aleatorias se necesita tomar para que entre ellas la cifra "6" aparezca al menos una vez con una probabilidad no menor que 0,9?
6. ¿Cuántas veces se debe lanzar una moneda para que la frecuencia de la salida del escudo se distinga de 1/2 no más que en Δ con una probabilidad no menor que $1-2\alpha$ ($2\alpha = 0,05$, $\Delta = 0,1$)?

§ 5. Variables aleatorias

5.1. Variables aleatorias en el esquema finito. Examinemos un espacio probabilístico finito $(\Omega, \mathscr{A}, \mathbf{P})$.

Se llama variable aleatoria la función numérica $\xi = \xi(\omega)$ *de un suceso elemental* $\omega \in \Omega$.

Designaremos las variables aleatorias con las letras griegas $\xi, \eta, \zeta, \mu, \nu$, ... , a veces las variables aleatorias se anotan con las letras latinas mayúsculas X, Y, Z, ...

En un modelo de un espacio probabilístico finito cada suceso elemental $\omega \in \Omega$ se confronta con un caso aleatorio de cierto fenómeno aleatorio, mientras que una variable aleatoria $\xi = \xi(\omega)$ asigna a estos casos aleatorios los valores numéricos. Por lo tanto, en función del caso la variable aleatoria toma diversos valores numéricos.

EJEMPLO 1. Al arrojar un dado son posibles seis casos elementales. Designemos con ω_i un suceso elemental consistente en que salen i puntos. Entonces $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}$ y la variable aleatoria $\xi(\omega_i) = i$ es igual al número de los puntos que salen en el dado lanzado.

EJEMPLO 2. En un esquema de pruebas independientes de Bernoulli (véase el § 3) el conjunto Ω se compone de sucesos elementales $\omega = (x_1, x_2, \ldots, x_n)$, donde $x_i = 1$, si en la i-ésima prueba se produce un caso favorable, y $x_i = 0$, si se produce un caso desfavorable. La variable aleatoria $\mu = \mu(\omega) = x_1 + x_2 + \ldots + x_n$ es igual al número de los casos favorables al realizar n pruebas en el esquema de Bernoulli.

Toda constante C puede considerarse como un caso particular de la variable aleatoria $\xi = \xi(\omega) \equiv C$. Tales magnitudes aleatorias se llaman *degeneradas*. Las magnitudes aleatorias más simples, distintas de las degeneradas, son los indicadores. Con cada suceso $A \in \mathscr{A}$ se puede vincular una variable aleatoria

$$\chi_A = \chi_A(\omega) = \begin{cases} 1, & \text{si} \quad \omega \in A, \\ 0, & \text{si} \quad \omega \notin A, \end{cases}$$

que se denomina *indicador* del suceso A.

EJEMPLO 3. En n pruebas del esquema de Bernoulli (véase el ejemplo 2) determinemos los sucesos A_i = {en la i-ésima prueba se ha producido el "caso favorable"}, $i = 1, 2, \ldots, n$. La variable aleatoria μ_n, igual al número de los casos favorables, puede ser representada en forma de suma de los indicadores:

$$\mu_n = \chi_{A_1} + \chi_{A_2} + \ldots + \chi_{A_n}. \tag{1}$$

En efecto, $\mu_n(x_1, \ldots, x_n) = x_1 + x_2 + \ldots + x_n$ y, por otro lado, $\chi_{A_i}(x_1, \ldots, x_n) = x_i$. La representación (1) se utiliza frecuentemente al investigar la variable μ_n.

Los indicadores satisfacen las siguientes propiedades que se comprueban fácilmente:

$$\chi_\varnothing \equiv 0, \quad \chi_\Omega \equiv 1, \quad \chi_{AB} = \chi_A \chi_B, \quad \chi_{\bar{A}} = 1 - \chi_A.$$

Si los sucesos A y B son incompatibles, entonces

$$\chi_{A+B} = \chi_A + \chi_B.$$

En efecto, $\chi_{A+B}(\omega) = 1$ es equivalente a la condición $\omega \in A + B$. Esto quiere decir que o bien $\omega \in A$ y $\omega \in B$, o bien $\omega \in A$ y $\omega \in B$, o sea, $\chi_A = 1$ y $\chi_B = 0$ o bien $\chi_A = 0$ y $\chi_B = 1$.

Por inducción es fácil determinar que

$$\chi_{A_1 A_2 \ldots A_n} = \chi_{A_1} \chi_{A_2} \cdots \chi_{A_n},$$

para cualesquiera sucesos $A_1, \ldots, A_n$, y

$$\chi_{A_1 + \ldots + A_n} = \chi_{A_1} + \ldots + \chi_{A_n},$$

para los sucesos $A_1, A_2, \ldots, A_n$ incompatibles dos a dos.

Sean $x_1 < x_2 < \ldots < x_n$ todos los valores posibles de una variable aleatoria ξ. Anotemos por $A_i = \{\omega \colon \xi(\omega) = x_i\}$ el suceso consistente en todos los sucesos elementales ω, para los cuales $\xi(\omega) = x_i$. Como x_i son todos diferentes, los sucesos A_i y A_j, para $i \neq j$, son incompatibles. Puesto que $x_1, x_2, \ldots, x_k$ comprenden todos los valores posibles de ξ, entonces $\sum_{i=1}^{k} A_i = \Omega$. Así, pues, con la variable aleatoria ξ puede ser vinculada la partición α_ξ que está compuesta por los conjuntos

$$A_i = \{\omega \colon \xi(\omega) = x_i\},$$
$$i = 1, \ldots, k, \tag{2}$$

sobre los cuales $\xi(\omega)$ es constante. La partición α_ξ la llamaremos *engendrada por la variable aleatoria* ξ.

Con ayuda de la partición (2) la variable aleatoria ξ puede representarse en forma de una combinación lineal de los indicadores

$$\xi = \xi(\omega) = \sum_{i=1}^{k} x_i \chi_{A_i}(\omega), \tag{3}$$

ya que el primer miembro y el segundo miembro de (3) toman un mismo valor de x_i cuando $\omega \in A_i$.

EJEMPLO 4. En los ejemplos 2 y 3 pongamos $n = 3$. Entonces $\Omega = \{000, 001, 010, 011, 100, 101, 110, 111\}$. La variable aleatoria μ_3 (el número de los casos favorables en tres pruebas) determina la partición siguiente α_{μ_3}:

$$A_0 = \{\omega \colon \mu_3 = 0\} = \{000\},$$
$$A_1 = \{\omega \colon \mu_3 = 1\} = \{001, 010, 100\},$$
$$A_2 = \{\omega \colon \mu_3 = 2\} = \{011, 101, 110\},$$
$$A_3 = \{\omega \colon \mu_3 = 3\} = \{111\}.$$

La representación (2) para μ_3 tiene la forma

$$\mu_3 = 0 \cdot \chi_{A_3} + \chi_{A_1} + 2\chi_{A_2} + 3\chi_{A_3}.$$

Para todo valor de $\omega \in \Omega$ exactamente una de las variables $\chi_{A_0}, \chi_{A_1}, \chi_{A_2}, \chi_{A_3}$ se distingue de 0.

Llamaremos *ley de distribución* $\mathbf{P}_\xi(B)$ de la variable aleatoria ξ la probabilidad $\mathbf{P}_\xi(B) = \mathbf{P}\{\xi \in B\}$ determinada para cada conjunto numérico B. La ley de distribución de la variable aleatoria ξ se determina por sus valores $x_1, x_2, \ldots, x_k$ y por las probabilidades $\mathbf{P}\{\xi = x_i\}$ de estos valores. Designemos $\mathbf{P}\{\xi = x_i\} = p_i$. Entonces la ley de distribución puede ser determinada con ayuda de la tabla

x_1	x_2	...	x_k
p_1	p_2	...	p_k

(4)

cuya línea superior se compone de diversos números x_i y los números de la línea inferior satisfacen las condiciones $p_i \geqslant 0$, $\sum_{i=1}^{k} p_i = 1$. Estas condiciones se derivan de que $p_i = \mathbf{P}(A_i)$ son las probabilidades de la partición $\sum_{i=1}^{k} A_i = \Omega$, de donde a base del axioma de aditividad, resulta que $\sum_{i=1}^{k} \mathbf{P}(A_i) = \mathbf{P}(\Omega) = 1$. Con ayuda de la tabla (4) se puede determinar la probabilidad

$$\mathbf{P}\{\xi \in B\} = \sum_{i:\, x_i \in B} p_i, \tag{5}$$

para todo conjunto numérico B.

En la teoría de las probabilidades se examinan frecuentemente las variables aleatorias ξ con la ley de distribución (5) sin indicar el espacio probabilístico $(\Omega, \mathscr{A}, \mathbf{P})$ ni la función $\xi(\omega)$ que define la variable aleatoria. En este caso se supone que existe un espacio probabilístico $(\Omega, \mathscr{A}, \mathbf{P})$ cualquiera, sobre el cual puede definirse la función $\xi = \xi(\omega)$ de modo que la tabla (3) asigne su ley de distribución. La selección del espacio probabilístico se determina cada vez por la esencia del problema o la simplicidad del esquema que se obtiene. El espacio probabilístico elemental relacionado con la ley de distribución (5) será el espacio, en el cual el conjunto de los sucesos elementales no es más que el conjunto de los valores de la variable aleatoria $\Omega = \{x_1, x_2, \dots, x_k\}$ con las probabilidades elementales $p(x_i) = p_i$. Entonces la variable aleatoria se define por la función $\xi(x_i) = x_i$.

La ley de distribución del indicador χ_A del suceso A está definida por la tabla

$$\begin{array}{cc} \hline 0 & 1 \\ \hline 1 - \mathbf{P}(A) & \mathbf{P}(A) \\ \hline \end{array} \tag{6}$$

A cada variable aleatoria le corresponde una ley de distribución. Diferentes variables aleatorias pueden tener una misma ley de distribución. Volvamos a examinar el ejemplo 1 que trata del lanzamiento de un dado. Designemos los sucesos $A = \{\omega_1, \omega_3, \omega_5\}$, $B = \{\omega_2, \omega_4, \omega_6\}$, o sea, el suceso A = {el número de los puntos salidos es impar} y el suceso B = {el número de los puntos salidos es par}. Con estos sucesos están vinculadas diferentes variables aleatorias χ_A y χ_B. Sin embargo, en virtud de la igualdad $\mathbf{P}(A) = \mathbf{P}(B) = 1/2$ estas diversas variables aleatorias tienen la misma ley de distribución (6).

Por brevedad, la ley de distribución de ξ se llama a veces simplemente *ley* o *distribución*. Con frecuencia se denomina ley de distribución la tabla (4) que la define.

Citemos algunas leyes de distribución.

1. *Ley binomial* para el número de los vasos favorables μ_n en n pruebas independientes según el esquema de Bernoulli

$$\mathbf{P}\{\mu_n = m\} = C_n^m p^m (1 - p)^{n - m}, \quad m = 0, 1, \dots, n$$

(véase (7) del § 3).

2. *Distribución hipergeométrica*, o sea, la distribución del número de las bolas blancas ξ al seleccionar, sin devolver, el volumen n de la urna que contiene M bolas blancas y $N - M$ bolas negras (véase el punto 1.6, ejemplo 4)

$$\mathbf{P}\{\xi = m\} = \frac{C_M^m C_{N-M}^{n-m}}{C_N^n}, \quad m = 0, 1, \dots, \min(n, M).$$

3. *Distribución uniforme* sobre $\{1, 2, \ldots, N\}$:

$$\mathbf{P}\{\xi = m\} = \frac{1}{N}, \quad m = 1, 2, \ldots, N.$$

Con ayuda de la función numérica $y = g(x)$ y la variable aleatoria $\xi = \xi(\omega)$ se puede obtener una nueva variable aleatoria $\eta = \eta(\omega) = g(\xi(\omega))$. Si la función $g(x)$ es tal que a distintos valores de ξ les corresponden distintos valores de η (o sea, $g(x_i) \neq g(x_j)$ para $x_i \neq x_j$), entonces la partición (2) engendrada por ξ coincide con la engendrada por η de modo que

$$A_i = \{\omega: \xi(\omega) = x_i\} = \{\omega: g(\xi(\omega)) = g(x_i)\}.$$

En este caso la ley de distribución de $\eta = g(\xi)$ se asigna por la tabla

$g(x_1)$	$g(x_2)$	$\ldots$	$g(x_k)$
p_1	p_2	$\ldots$	p_k

(7)

y la misma variable aleatoria η es representable en forma de la suma

$$\eta = g(\xi) = \sum_{i=1}^{k} g(x_i)\chi_{A_i}. \tag{8}$$

No obstante, si para ciertos valores de $i \neq j$ $g(x_i) = g(x_j)$, la ley de distribución de $\eta = g(\xi)$ será otra, aunque es fácil obtenerla de la tabla (7). Designemos con $y_1 < y_2 < \ldots < y_m$ todos los valores diferentes de la variable $\eta = g(\xi)$. Entonces la ley de distribución de η está definida por la tabla

y_1	y_2	$\ldots$	y_m
q_1	q_2	$\ldots$	q_m

(9)

donde en la línea superior están escritos todos los valores diferentes de $g(x_i)$ de la línea superior de (7) y en la línea inferior q_j es la suma de los valores de p_i, para los cuales $g(x_i) = y_j$. Nótese que la igualdad (8) es válida también en el caso en que $g(x_i)$ coinciden para ciertos valores de x_i.

EJEMPLO 5. Sea que ξ una variable aleatoria con la ley de distribución $\mathbf{P}\{\xi = k\} = \frac{1}{10}$, $k = 0, 1, 2, \dots, 9$ y la función $g(x) = (x - 5)^2$. Entonces la variable aleatoria $\eta = (\xi - 5)^2$ tendrá la siguiente ley de distribución: $\mathbf{P}\{\eta = 0\} = \mathbf{P}\{\eta = 25\} = 1/10$, $\mathbf{P}\{\eta = 1\} = \mathbf{P}\{\eta = 4\} = \mathbf{P}\{\eta = 9\} = \mathbf{P}\{\eta = 16\} = 1/5$. Calculemos, por ejemplo, $\mathbf{P}\{\eta = 4\}$. Tenemos

$$\mathbf{P}\{\eta = 4\} = \mathbf{P}\{(\xi - 5)^2 = 4\} = \mathbf{P}\{\xi = 3\} + \mathbf{P}(\xi = 7) = 1/10 + 1/10 = 1/5.$$

5.2. Variables aleatorias en el esquema numerable. Supongamos que en un espacio probabilístico $(\Omega, \mathscr{A}, \mathbf{P})$ el conjunto

$$\Omega = \{\omega\} = \{\omega_1, \omega_2, \dots, \omega_n, \dots\}$$

es numerable, $\mathscr{A}$ se compone de todos los subconjuntos Ω y la probabilidad $\mathbf{P}$ viene dada por las probabilidades elementales $p(\omega)$. En este caso la variable aleatoria ξ asimismo se define como la función numérica $\xi = \xi(\omega)$ de un suceso elemental. El conjunto de diferentes valores de x_i de la variable aleatoria ξ puede ser finito o numerable. Análogamente a (2) con la variable aleatoria ξ se puede vincular la partición α_ξ compuesta por los conjuntos

$$A_i = \{\omega: \xi(\omega) = x_i\}, \quad i = 1, 2, \dots,$$

sin embargo, el número de tales conjuntos puede ser numerable. Entonces la partición α_ξ engendrada por ξ se llama *numerable*. La representación de ξ en forma de la suma análoga a (3) ofrece la serie

$$\xi = \xi(\omega) = \sum_{i=1}^{\infty} x_i \chi_{A_i}, \tag{10}$$

en el cual el segundo miembro no tiene para cada ω más que un término no nulo. En este caso la ley de distribución $\mathbf{P}\{\xi \in B\}$ está definida por una tabla análoga a (4) y consistente en dos líneas infinitas que determinan las probabilidades $\mathbf{P}\{\xi = x_i\} = p_i$ $\sum\limits_{i=1}^{\infty} p_i = 1$.

Citemos dos distribuciones que se encuentran con frecuencia en diferentes modelos.

1. *La distribución de Poisson* se determina por las probabilidades

$$\mathbf{P}\{\xi = m\} = \frac{a^m}{m!} e^{-a}, \quad m = 0, 1, \ldots,$$

que dependen del parámetro $a > 0$. Es fácil de ver que en este caso se cumple la condicion $\sum_{m=0}^{\infty} \frac{a^m}{m!} e^{-a} = 1$.

2. *La distribución geométrica* depende del parámetro $0 < p < 1$ y se determina por las probabilidades

$$\mathbf{P}\{\xi = m\} = p^m q, \quad m = 0, 1, 2, \ldots, \quad q = 1 - p.$$

En este caso también está cumplida la condición $\sum_{m=0}^{\infty} p^m q = 1$.

5.3. Variables aleatorias en el esquema general. Función de distribución. Si se trata de un espacio probabilístico arbitrario $(\Omega, \mathscr{A}, \mathbf{P})$, se llama *variable aleatoria* ξ tal función $\xi = \xi(\omega)$ de un suceso elemental ω para la cual, cualquiera que sea el valor de x, el conjunto $\{\omega: \xi(\omega) \leqslant x\} \in \mathscr{A}$, o sea, la desigualdad $\{\xi \leqslant x\}$ es un suceso. La probabilidad de este suceso $\mathbf{P}\{\xi \leqslant x\}$ se denomina *función de distribución*. La anotaremos por $F_\xi(x)$, y, a veces, simplemente $F(x)$.

EJEMPLO 6. Supongamos que en un espacio probabilístico $(\Omega, \mathscr{A}, \mathbf{P})$ Ω se compone de los puntos (x, y) de un cuadrado unitario del plano $0 \leqslant x \leqslant 1$, $0 \leqslant y \leqslant 1$, $\mathscr{A}$ son los conjuntos cuadrables de este cuadrado y la probabilidad $\mathbf{P}(A) = |A|$ es el área del conjunto A. Determinemos la variable aleatoria $\xi = \chi_A$, donde χ_A es el indicador del suceso A (o sea, $\chi_A(\omega) = 1$, si $\omega \in A$ y $\chi_A(\omega) = 0$, si $\omega \notin A$). La variable aleatoria ξ tiene solamente dos valores posibles $\xi = 0$ y $\xi = 1$. Su función de distribución se define por las igualdades

$$F_\xi(x) = \begin{cases} 0, & \text{para } x < 0, \\ \mathbf{P}(\bar{A}) = 1 - |A|, & \text{para } 0 \leqslant x < 1, \\ 1, & \text{para } x \geqslant 1; \end{cases}$$

el gráfico se muestra en la fig. 7. Si el conjunto B no es cuadrable, entonces la función de un suceso elemental $g(\omega)$ definida por las igualdades

$$g(\omega) = \begin{cases} 0, & \text{si } \omega \in B, \\ 1, & \text{si } \omega \notin B \end{cases}$$

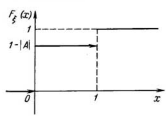

Fig. 7.

no será una variable aleatoria, puesto que el conjunto $\{\omega: g(\omega) \leqslant 1/2\}$ coincide con el conjunto B, el cual no es un suceso y para el cual no está determinada la probabilidad.

Propiedades de la función de distribución. La función de distribución $F_\xi(x)$ asigna la ley de distribución

$$\mathbf{P}_\xi(B) = \mathbf{P}\{\xi \in B\} \tag{11}$$

para una colección bastante rica de conjuntos numéricos B. Sea que $x_1 < x_2$. Entonces el suceso $\{\xi \leqslant x_2\}$ puede ser representado en la suma de los sucesos incompatibles $\{\xi \leqslant x_1\} + \{x_1 < \xi \leqslant x_2\}$; aplicando el teorema de adición A3, tenemos $\mathbf{P}(\xi \leqslant x_2\} = \mathbf{P}\{\xi \leqslant x_1\} + \mathbf{P}\{x_1 < \xi \leqslant x_2\}$, de donde obtenemos

$$\mathbf{P}\{x_1 < \xi \leqslant x_2\} = F_\xi(x_2) - F_\xi(x_1). \tag{12}$$

Ahora bien, con ayuda de la fórmula (12) la función de distribución $F_\xi(x)$ determina la probabilidad $\mathbf{P}\{\xi \in B\}$, para $B = (x_1, x_2)$, o sea para el semiintervalo abierto a la izquierda y cerrado a la derecha. Con auyda de la función de distribución $F_\xi(x)$ se puede determinar la probabilidad de que ξ se encuentre en los conjuntos B representables en forma de la suma de intervalos disjuntos (finitos o infinitos). Sea $-\infty \leqslant x_1 < y_1 \leqslant x_2 < y_2 \leqslant \ldots \leqslant x_N < y_N \leqslant +\infty$ y $B = \bigcup_{n=1}^{N} (x_n, y_n)$. Entonces de acuerdo con el axioma de adición A3 y la fórmula (12) tenemos

$$\mathbf{P}_\xi(B) = \mathbf{P}\{\xi \in B\} = \sum_{n=1}^{N} [F_\xi(y_n) - F_\xi(x_n)]. \tag{13}$$

La fórmula (13) puede ser extendida también a la suma del número numerable de los intervalos $(x_k, y_k\rangle$. Para esto se necesita aplicar el axioma ampliado de adición A4.

Si $x_1 = -\infty$ o $y_N = +\infty$, entonces en (13) aparecen las expresiones $F_\xi(-\infty)$ y $F_\xi(+\infty)$. Vamos a determinarlas como límites $F_\xi(+\infty) = \lim\limits_{x\to+\infty} F_\xi(x)$, $F_\xi(-\infty) = \lim\limits_{x\to-\infty} F_\xi(x)$ y demostrar que $F_\xi(+\infty) = 1$ y $F(-\infty) = 0$. Introduzcamos los sucesos $A_n = \{-n < \xi \leqslant n\}$, $n = 1, 2, \ldots$, $A_0 = \varnothing$ y $B_n = A_n \setminus A_{n-1} = \{n - 1 < \xi \leqslant n\} + \{-n < \xi \leqslant -(n-1)\}$. No es difícil ver que los sucesos B_n son incompatibles dos a dos y $\Omega = \sum\limits_{n=1}^{\infty} B_n$, por eso conforme al axioma ampliado de adición A4

$$1 = \mathbf{P}(\Omega) = \sum_{n=1}^{\infty} \mathbf{P}(B_n) = \lim_{N\to\infty} \sum_{n=1}^{N} \mathbf{P}(B_n) =$$

$$= \lim_{N\to\infty} \mathbf{P}(A_N) = \lim_{N\to\infty} F_\xi(N) - \lim_{N\to\infty} F(-N) =$$

$$= F_\xi(+\infty) - F_\xi(-\infty), \qquad (14)$$

puesto que $A_N = \sum\limits_{n=1}^{N} B_n$ y $\mathbf{P}(A_N) = F_\xi(N) - F_\xi(-N)$. Todos los valores de $F_\xi(x)$ están entre 0 y 1, por eso $0 \leqslant F_\xi(\mp\infty) \leqslant 1$ y de (14) se deduce $F_\xi(+\infty) = 1$, $F_\xi(-\infty) = 0$.

Designemos $F_\xi(x - 0)$ el límite $\lim\limits_{n\to\infty} F_\xi\left(x - \dfrac{1}{n}\right)$, es cual se llama *límite en el punto x a la izquierda*. Mostremos que $F_\xi(x - 0) = \mathbf{P}\{\xi < x\}$. En efecto, el suceso $\{\xi < x\}$ es representable en la forma numerable de los sucesos incompatibles dos a dos $\sum\limits_{n=1}^{\infty} \left\{x - \dfrac{1}{n-1} < \xi \leqslant x - \dfrac{1}{n}\right\}$ (cuando $n = 1$ suponemos que $\left\{-\dfrac{1}{n-1} < \xi \leqslant x - \dfrac{1}{n}\right\} = \{\xi \leqslant x - 1\}$). Según el axioma A4

$$\mathbf{P}\{\xi < x\} = \sum_{n=1}^{\infty} \mathbf{P}\left\{x - \frac{1}{n-1} < \xi \leqslant x - \frac{1}{n}\right\} =$$

$$= \lim_{N\to\infty} \sum_{n=1}^{N} \mathbf{P}\left(x - \frac{1}{n-1} < \xi \leqslant x - \frac{1}{n}\right\} =$$

$$= \lim_{N\to\infty} F_\xi\left(x - \frac{1}{N}\right) = F_\xi(x - 0).$$

Ahora podemos demostrar que

$$\mathbf{P}\{\xi = x\} = F_\xi(x) - F_\xi(x - 0). \tag{15}$$

En efecto, (15) se deduce fácilmente de $\{\xi \leqslant x\} = \{\xi < x\} + \{\xi = x\}$ y del axioma A3. Con ayuda de las fórmulas (12) y (15) no es difícil obtener ahora las probabilidades de llegar a parar en los intervalos abierto y cerrado. Como para $x_1 < x_2$ tienen lugar las relaciones $\{x_1 \leqslant \xi \leqslant x_2\} = \{\xi = x_1\} + \{x_1 < \xi \leqslant x_2\}$ y $\{x_1 < \xi \leqslant x_2\} = \{x_1 < \xi < x_2\} + \{\xi = x_2\}$, entonces de acuerdo con el axioma A3 tenemos

$$\mathbf{P}\{x_1 \leqslant \xi \leqslant x_2\} = F_\xi(x_2) - F_\xi(x_1 - 0),$$

$$\mathbf{P}\{x_1 < \xi < x_2\} = F_\xi(x_2 - 0) - F_\xi(x_1).$$

Puede ser mostrado que el límite en el punto x *a la derecha* $F_\xi(x + 0) = \lim\limits_{n \to \infty} F_\xi\left(x + \dfrac{1}{n}\right)$ de la función de distribución es igual a $F_\xi(x)$ (esta propiedad significa que la función de distribución es *continua a la derecha*). Para esto es necesario aplicar el teorema A4 a la suma

$$\{x < \xi \leqslant x + 1\} = \sum_{n=1}^{\infty} \left\{x + \frac{1}{n+1} < \xi \leqslant x + \frac{1}{n}\right\}.$$

Por lo tanto, la función de distribución $F(x)$ posee las propiedades siguientes:

1) $0 \leqslant F(x) \leqslant 1$, para todos los valores de x;
2) $F(x_1) \leqslant F(x_2)$, si $x_1 \leqslant x_2$;
3) $F(-\infty) = 0$, $F(+\infty) = 1$;
4) $F(x + 0) = F(x)$.

Se puede mostrar que una función $F(x)$ cualquiera que posea las propiedades citadas puede ser función de distribución de cierta variable aleatoria.

Variables aleatorias discretas y continuas. Sobre un espacio probabilístico general $(\Omega, \mathscr{A}, \mathbf{P})$ mediante la fórmula (9) se puede prefijar una variable aleatoria ξ, si por A_i, $i = 1, 2, \ldots$, se entienden los sucesos que forman una partición numerable o finita, o sea, los sucesos incompatibles dos a dos, para los cuales $\sum\limits_{i=1}^{\infty} A_i = \Omega$.

En este caso la ley de distribución $\mathbf{P}_\xi(B) = \mathbf{P}\{\xi \in B\}$ se asigna por las probabilidades $\mathbf{P}\{\xi = x_i\}$ de modo que

$$\mathbf{P}_\xi(B) = \sum_{x_i \in B} \mathbf{P}\{\xi = x_i\}. \tag{16}$$

Tales variables aleatorias y sus leyes de distribución (16) se llaman *discretas*. Todas las variables aleatorias de los puntos 5.1 y 5.2 determinadas sobre un espacio probabilístico finito o numerable son discretas.

Otra clase importante de variables aleatorias está constituida por así llamadas variables aleatorias *continuas* ξ, para las cuales $\mathbf{P}\{\xi = x\} = 0$ con cualesquiera valores de x. En este caso la ley de distribución $P_\xi(B)$ no se puede prefijar mediante la fórmula (16). Examinaremos tales variables aleatorias continuas ξ, cuya ley de distribución se puede prefijar con ayuda de la fórmula

$$\mathbf{P}\{\xi \in B\} = \int_B p_\xi(x)dx, \tag{17}$$

donde $p_\xi(x) \geqslant 0$ es la función integrable. La función $p_\xi(x)$ se denomina *densidad* de distribución de la variable aleatoria ξ. De (17) resulta (cuando $B = (-\infty, \infty)$) que

$$\int_{-\infty}^{\infty} p_\xi(x)dx = 1. \tag{18}$$

En los casos prácticamente importantes las funciones $p_\xi(x)$ son continuas o continuas a trozos. Si $B = (-\infty, x)$, entonces de (16) obtenemos la relación entre la densidad $p_\xi(x)$ y la función de distribución

$$F_\xi(x) = \int_{-\infty}^{x} p_\xi(u)du. \tag{19}$$

De esta fórmula se deduce que en el punto x, en el cual es continua la probabilidad $p_\xi(x)$

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

y para $\Delta x > 0$ y $\Delta x \to 0$

$$\mathbf{P}\{x < \xi < x + \Delta x\} = p_\xi(x)\Delta x + o(\Delta x).$$

Toda función $p(x)$ continua a trozos no negativa que posea la propiedad (17) puede ser densidad de cierta distribución de las propiedades.

Citemos algunos de las leyes de distribución de variables aleatorias continuas que se encuentran con frecuencia.

1. *La distribución uniforme* sobre $[a, b]$ está prefijada por la densidad

$$p(x) = \begin{cases} \dfrac{1}{b-a}, & \text{si} \quad 0 \leqslant x \leqslant b, \\ 0, & \text{si} \quad x < a \text{ o bien } x > b. \end{cases}$$

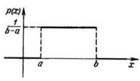

Fig. 8.

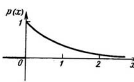

Fig. 9.

El gráfico de la densidad $y = p(x)$ de distribución uniforme se muestra en la fig. 8.

2. *La distribución exponencial* está prefijada por la densidad que depende del parámetro $\lambda > 0$

$$p(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{si} \quad x \geqslant 0, \\ 0, & \text{si} \quad x < 0. \end{cases}$$

El gráfico $y = p(x)$ está representado en la fig. 9.

3. *La distribución normal* o *gausiana* está prefijada por la densidad que depende de los parámetros $-\infty < a < \infty$ y $\sigma > 0$:

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}}$$

El gráfico $y = p(x)$ aparece en la fig. 10.

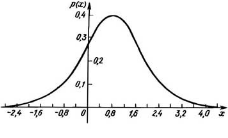

Fig. 10.

Al hacer uso de las fórmulas (17), (18) ó (19) hay que tener presente que la densidad $p(x)$ se prefija frecuentemente por diferentes fórmulas analíticas en distintos intervalos.

EJEMPLO 7. Calculemos la función de distribución de una variable aleatoria que está distribuida uniformemente en el segmento $[a, b]$. Para esto se emplean la fórmula (19) y la densidad antes prefijada $p(x)$ de la distribución uniforme. Cuando $x < 0$ tenemos $F(x) = \int_{-\infty}^{x} p(u)du = 0$. Si $a \leqslant x \leqslant b$, entonces

$$F(x) = \int_{-\infty}^{x} p(u)du = \int_{-\infty}^{a} p(u)du + \int_{a}^{x} p(u)du =$$
$$= 0 + \frac{1}{b-a}\int_{a}^{x} du = \frac{x-a}{b-a}.$$

Si $x > b$, entonces

$$F(x) = \int_{-\infty}^{a} p(u)du + \int_{a}^{b} p(u)du + \int_{b}^{x} p(u)du =$$
$$= 0 + \frac{1}{b-a}\int_{a}^{b} du + 0 = 1.$$

Así, pues,

$$F(x) = \begin{cases} 0, & \text{si } x < a, \\ \dfrac{x-a}{b-a}, & \text{si } a \leqslant x \leqslant b, \\ 1, & \text{si } x > b. \end{cases}$$

El gráfico de la función de distribución $y = F(x)$ se muestra en la fig. 11.

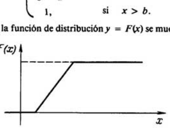

Fig. 11.

5.4. Funciones de variables aleatorias. Sea ξ una variable aleatoria continua con la función de distribución $F_\xi(x)$ y la densidad $p_\xi(x)$. Si $g(x)$ es la función numérica del argumento real x, se puede formar la variable aleatoria $\eta = g(\xi)$, que es la función de ξ. Se pregunta ¿cómo por la ley de distribución de ξ se halla la ley de distribución de η? Citemos algunos ejemplos que muestran cómo hace falta proceder en el caso general. Examinaremos las variables aleatorias continuas, para las cuales $\mathbf{P}\{\xi = x\} = 0$, por eso la función de distribución $F_\xi(x)$ será continua en todos los puntos y $\mathbf{P}\{\xi \leqslant x\} = \mathbf{P}\{\xi < x\} = F_\xi(x)$.

EJEMPLO 8. Examinemos la función lineal $g(x) = ax + b$. Pongamos $\eta = a\xi + b$. Si $a > 0$, entonces, según la definición de la función de distribución,

$$F_\eta(x) = \mathbf{P}\{\eta < x\} = \mathbf{P}\{a\xi + b < x\} = \mathbf{P}\left\{\xi < \frac{x-b}{a}\right\} = F_\xi\left(\frac{x-b}{a}\right),$$

de donde $p_\eta(x) = F'_\eta(x) = F'_\xi\left(\frac{x-b}{a}\right)\frac{1}{a} = \frac{1}{a}p_\xi\left(\frac{x-b}{a}\right)$. Si $a < 0$, tenemos

$$F_\eta(x) = \mathbf{P}\{\eta < x\} = \mathbf{P}\{a\xi + b < x\} = \mathbf{P}\left\{\xi > \frac{x-b}{a}\right\} = 1 - F_\xi\left(\frac{x-b}{a}\right)$$

y

$$p_\eta(x) = F'_\eta(x) = -\frac{1}{a}F'_\xi\left(\frac{x-b}{a}\right) = -\frac{1}{a}p_\xi\left(\frac{x-b}{a}\right).$$

Al unir estos dos casos tenemos

$$p_\eta(x) = \frac{1}{|a|}p_\xi\left(\frac{x-b}{a}\right).$$

EJEMPLO 9. Examinemos una función monótona creciente $g(x) = x^3$. Pongamos $\eta = \xi^3$. Por definición de la función de distribución tenemos

$$F_\eta(x) = \mathbf{P}\{\eta < x\} = \mathbf{P}\{\xi^3 < x\} = \mathbf{P}\{\xi < x^{1/3}\} = F_\xi\{x^{1/3}\}.$$

Derivando esta igualdad obtenemos la expresión para la densidad $p_\eta(x)$:

$$p_\eta(x) = F_\eta'(x) = F_\xi'(x^{1/3}) \cdot \frac{1}{3}x^{-2/3} = \frac{1}{3}x^{-2/3}p_\xi(x^{1/3}).$$

EJEMPLO 10. Examinemos una función monótona decreciente $g(x) = e^{-x}$. Hallemos $F_\eta(x)$ y $p_\eta(x)$ de la variable aleatoria $\eta = e^{-\xi}$. Para $x > 0$ tenemos

$$F_\eta(x) = \mathbf{P}\{\eta < x\} = \mathbf{P}\{e^{-\xi} < x\} = \mathbf{P}\{\xi > -\ln x\} =$$
$$= 1 - \mathbf{P}\{\xi \leqslant -\ln x\} = 1 - F_\xi(-\ln x),$$

de donde

$$p_\eta(x) = F_\eta'(x) = \frac{1}{x}F_\xi'(-\ln x) = \frac{1}{x}p_\xi(-\ln x).$$

Cuando $x < 0$, $F_\eta(x) = \mathbf{P}\{\eta < x\} = 0$, ya que $\eta = e^{-\xi} \geqslant 0$ para todos los valores de ξ. Por eso sobre el semieje negativo $p_\eta(x) = F_\eta'(x) = 0$.

EJEMPLO 11. Examinemos una función no monótona $g(x) = x^2$. Hallemos $F_\eta(x)$ y $p_\eta(x)$ de la variable aleatoria $\eta = \xi^2$. Para $x > 0$ tenemos

$$F_\eta(x) = \mathbf{P}\{\eta < x\} = \mathbf{P}\{\xi^2 < x\} = \mathbf{P}\{-\sqrt{x} < \xi < \sqrt{x}\} =$$
$$= \mathbf{P}\{\xi < \sqrt{x}\} - \mathbf{P}\{\xi \leqslant -\sqrt{x}\} = F_\xi(\sqrt{x}) - F_\xi(-\sqrt{x}),$$

de donde

$$p_\eta(x) = F_\eta'(x) = F_\xi'(\sqrt{x})\frac{1}{2\sqrt{x}} + \frac{1}{2\sqrt{x}}F_\xi'(-\sqrt{x}) =$$
$$= \frac{1}{2\sqrt{x}}(p_\xi\sqrt{x}) + p_\xi(-\sqrt{x})).$$

Cuando $x \leqslant 0$ tenemos $F_\eta(x) = 0$ y $p_\eta(x) = 0$.

Problemas

1. Se arroja un dado. Designemos A = {el número de los puntos salidos es par}, B = {el número de los puntos salidos se divide por 3}. Hallar: a) la ley de distribución de χ_A; b) la ley de distribución de χ_B; c) la partición engendrada por la variable aleatoria $\xi = \chi_A + \chi_B$ y su ley de distribución.

2. Sobre un tablero de ajedrez vacío de coloca a la suerte un alfil. Las probabilidades de que el alfil se ponga en cada escaque se considera igual. Hallar la ley de distribución de ξ, o sea, del número de las casillas de movilidad.

3. Entre las 28 fichas de un dominó se escoge a la suerte y de un modo equiprobable una. Hallar la ley de distribución de la suma de los puntos ξ marcados en ambos cuadrados de la ficha.

4. Una variable aleatoria ξ tiene la distribución de Poisson con parámetro a. Hallar la probabilidad de los sucesos: $A = \{\xi \text{ es par}\}$, $B = \{\xi \text{ es impar}\}$. *Indicación*. Hacer uso de las series

$$e^{a} = \sum_{n=0}^{\infty} \frac{a^n}{n!}, \quad e^{-a} = \sum_{n=0}^{\infty} \frac{(-1)^n a^n}{n!}.$$

5. La densidad de distribución de ξ está definida por la fórmula

$$p_{\xi}(x) = \begin{cases} Cx^{-3/2}, & \text{para} \quad x \geqslant 1, \\ 0, & \text{para,} \quad x < 1. \end{cases}$$

Hallar C y la función de distribución $F_{\xi}(x)$.

6. Se efectúa el tiro contra un blanco circular de radio r. Hallar la función de distribución $F_{\xi}(x)$ y la densidad $p_{\xi}(x)$ de la variable aleatoria ξ, igual a la distancia comprendida entre el centro del blanco y el punto de impacto, si este último está distribuido uniformemente en el blanco.

7. Un blanco de radio r está dividido mediante diez circunferencias concéntricas de radios $r_k = \frac{k}{10} r$, $k = 1, 2, \ldots, 10$, entre 10 partes que llevan, a partir del centro, los números de orden 10, 9, ... , 3, 2, 1. Al hacer el impacto en la zona respectiva, el número de orden ν de la zona indica el número de puntos obtenidos por el tirador. En las condiciones del problema 6 hallar la ley de distribución de ν.

8. Una variable aleatoria ξ tiene la distribución uniforme en el segmento $[1, 2]$. Hallar la probabilidad $\mathbf{P}\{2 < \xi^2 < 5\}$.

9. Una variable aleatoria ξ tiene la distribución uniforme en el segmento $[-1, 3]$. Hallar la probabilidad $\mathbf{P}\{\xi^2 \leqslant 2\}$.

10. Una variable aleatoria ξ tiene la distribución exponencial con densidad $p_{\xi}(x) = e^{-x}, x \geqslant 0$. Hallar la función de distribución y la densidad de la variable aleatoria $\eta = e^{-\xi}$.

11. Una variable aleatoria ξ está distribuida uniformemente en el segmento $[-1, 2]$. Hallar la densidad $p_{\eta}(x)$ de distribución $\eta = \xi^2$.

§ 6. Distribuciones compatibles de variables aleatorias

6.1. Leyes multidimensionales de distribución. En modelos probabilísticos se necesita con frecuencia examinar simultáneamente varias variables aleatorias. Por ejemplo, al efectuar el tiro contra un blanco plano un punto de impacto aleatorio tiene dos coordenadas ξ y η, las cuales son variables aleatorias; durante las investigaciones antropométricas como parámetros principales del cuerpo humano se consideran el peso, la estatura y el volumen del pecho, los cuales al escoger a la suerte una persona de cualquier conjunto asimismo son aleatorias. Al fabricar una gran cantidad de cualesquiera piezas en una máquina herramienta, diferentes dimensiones de las mismas también ofrecen un ejemplo de variables aleatorias.

En estos casos sobre el espacio probabilístico $(\Omega, \mathscr{A}, \mathbf{P})$ de un modelo matemático están asignadas varias variables aleatorias $\xi_1, \xi_2, \dots, \xi_r$ que, a veces, se examinan con comodidad como coordenadas de un punto aleatorio o de un vector aleatorio $\xi = (\xi_1, \dots, \xi_r)$ en un espacio r-dimensional R^r. Se llama ley de distribución compatible de estas variables aleatorias las probabilidades de que el punto ξ se encuentre en el conjunto r-dimensional B:

$$P_\xi(B) = \mathbf{P}\{\xi \in B\}, \tag{1}$$

probabilidad que se examina en dependencia del conjunto B. La ley de distribución (1) se denomina también *multidimensional* o, lo que es más exactamente, *r-dimensional*. Analicemos dos métodos de prefijar la ley (1). Supongamos que hay una colección finita o numerable de vectores $x(i) = (x_1(i), x_2(i), \dots, x_r(i))$ y se dan las probabilidades $\mathbf{P}\{\xi = x(i)\}$ que satisfacen la condición $\sum\limits_{i=1}^{\infty} \mathbf{P}\{\xi = x(i)\} = 1$. La ley de distribución (1) prefijada por la fórmula

$$P_\xi(B) = \sum_{x(i) \in B} \mathbf{P}\{\xi = x(i)\} \tag{2}$$

se llama *ley de distribución discreta*.

EJEMPLO 1. En el esquema de n pruebas independientes con r casos, examinado en el § 3, hemos introducido la distribución polinomial. Se puede considerarlo como distribución r-dimensional de las variables aleatorias $\mu_1, \mu_2, \dots, \mu_r$, donde μ_i es el número de los i-ésimos casos en n pruebas. Entonces

$$\mathbf{P}\{\mu_1 = m_1, \dots, \mu_r = m_r\} = \frac{n!}{m_1! \dots m_r!} p_1^{m_1} \dots p_r^{m_r},$$

si $m_1 + \dots + m_r = n$, y

$$\mathbf{P}\{\mu = m\} = 0$$

en los demás casos (aquí $\mu = (\mu_1, \dots, \mu_r)$ $m = (m_1, \dots, m_r)$). Otra clase de las leyes de distribución viene prefijada por la *densidad r-dimensional* $p_\xi(x) = p_{\xi_1 \dots \xi_r}(x_1, \dots, x_r) \geqslant 0$ que satisface la condición $\int\limits_{R^r} \dots \int p_{\xi_1 \dots \xi_r}(x_1, \dots, x_r)dx_1 \dots dx_r = 1$:

$$\mathbf{P}_\xi(B) = \int\limits_B \dots \int p_\xi(x_1, \dots, x_r)dx_1 \dots dx_r. \qquad (3)$$

Si $B = \{(u_1, \dots, u_r): x_i \leqslant u_i \leqslant x_i + \Delta x_i, i = 1, \dots, r\}$, $(x_1, \dots, x_r)$ es el punto de continuidad de la densidad, entonces para $\Delta x_i \to 0$, $i = 1, \dots, r$, aplicando a la integral del segundo miembro (3) el teorema del valor medio, obtenemos

$$\mathbf{P}_\xi(B) = p_{\xi_1 \dots \xi_r}(x_1, \dots, x_r)\Delta x_1 \dots \Delta x_r + o(\Delta x_1 \dots \Delta x_r).$$

EJEMPLO 2. Supongamos que durante el tiro contra un blanco plano cuadrado $|x| \leqslant a$, $|y| \leqslant a$ el punto de impacto (ξ, η) tiene la densidad bidimensional siguiente:

$$p_{\xi_\eta}(x, y) = \begin{cases} \dfrac{1}{4a^2}, & \text{si } |x| \leqslant a \text{ y } |y| \leqslant a, \\ 0 & \text{en los demás casos.} \end{cases}$$

Entonces la probabilidad de dar en el círculo $x^2 + y^2 \leqslant a^2$ se calcula, según la fórmula (3), del modo siguiente:

$$\mathbf{P}\{\xi^2 + \eta^2 \leqslant a^2\} = \iint\limits_{x^2 + y^2 \leqslant a^2} p_{\xi_\eta}(x, y)dx\,dy = \frac{\pi a^2}{4a^2} = \frac{\pi}{4}.$$

Suponiendo que en las fórmulas (2) ó (3) el conjunto B es igual a $\{y = (y_1, \dots, y_r): y_i \leqslant x_i, i = 1, \dots, r\}$, donde $x_1, \dots, x_r$ son los números dados, obtenemos la probabilidad del suceso $\{\xi_1 \leqslant x_1, \dots, \xi_r \leqslant x_r\}$ que depende de $x_1, x_2, \dots, x_r$; esta probabilidad se llama *función de distribución r-dimensional* de las variables aleatorias $\xi_1, \dots, \xi_r$ y se designa

$$F_\xi(x) = F_{\xi_1 \dots \xi_r}(x_1, \dots, x_r) = \mathbf{P}\{\xi_1 \leqslant x_1, \dots, \xi_r \leqslant x_r\}. \qquad (4)$$

Se puede demostrar que la función de distribución r-dimensional (4) prefija unívocamente la ley de distribución $\mathbf{P}_\xi(B)$. Por ejemplo, si hay una densidad r-dimensional $p_{\xi_1 \dots \xi_r}(x_1, \dots, x_r)$, entonces conforme a (3) y (4) la función de distribución $F_{\xi_1 \ldots \xi_r}(x_1, \ldots, x_r)$ está prefijada por la integral:

$$F_{\xi_1 \ldots \xi_r}(x_1, \ldots, x_r) =$$
$$= \int_{-\infty}^{x_1} du_1 \int_{-\infty}^{x_2} du_2 \ldots \int_{-\infty}^{x_r} p_{\xi_1 \ldots \xi_r}(u_1, \ldots, u_r) du_r, \quad (5)$$

de donde, derivando respecto a $x_1, x_2, \ldots, x_r$, obtenemos la igualdad

$$\frac{\partial^r F_{\xi_1 \ldots \xi_r}(x_1, \ldots, x_r)}{\partial x_1 \ldots \partial x_r} = p_{\xi_1 \ldots \xi_r}(x_1, \ldots, x_r) \quad (6)$$

que es válida en todos los puntos de continuidad de la densidad $p_{\xi_1 \ldots \xi_r}$. Ahora bien, conociendo la función de distribución $F_{\xi_1 \ldots \xi_r}$, según la fórmula (6) puede ser hallada la densidad $p_{\xi_1 \ldots \xi_r}$ y conforme a esta densidad por la fórmula (3) se define la ley de distribución $\mathbf{P}_\xi(B)$.

Con ayuda de la ley de distribución r-dimensional $\mathbf{P}_{\xi_1 \ldots \xi_r}$ se puede definir toda ley de distribución m-dimensional $\mathbf{P}_{\xi_1 \ldots \xi_m}$, $m < r$. Si la ley de distribución está prefijada por la función de distribución (4), entonces, suponiendo $x_{m+1} \to +\infty, \ldots, x_r \to +\infty$, obtenemos

$$F_{\xi_1 \ldots \xi_m}(x_1, \ldots, x_m) = \mathbf{P}\{\xi_1 \leqslant x_1, \ldots, \xi_m \leqslant x_m\} =$$
$$= \lim_{\substack{x_{m+1} \to +\infty \\ \cdots\cdots\cdots \\ x_r \to +\infty}} F_{\xi_1 \ldots \xi_r}(x_1, \ldots, x_r). \quad (7)$$

Designando con $F_{\xi_1 \ldots \xi_r}(x_1, \ldots, x_m, +\infty, \ldots, +\infty)$ el límite a la derecha en (7), se puede escribir (7) del modo siguiente:

$$F_{\xi_1 \ldots \xi_m}(x_1, \ldots, x_m) = F_{\xi_1 \ldots \xi_r}(x_1, \ldots, x_m, +\infty, \ldots, +\infty). \quad (8)$$

Suponiendo $x_{m+1} \to +\infty, \ldots, x_r \to +\infty$ en la fórmula (5) y teniendo en cuenta la fórmula (8), podemos obtener la densidad de distribución

$$p_{\xi_1 \ldots \xi_m}(x_1, \ldots, x_m) =$$
$$= \int_{-\infty}^{\infty} dx_{m+1} \ldots \int_{-\infty}^{\infty} p_{\xi_1 \ldots \xi_r}(x_1, \ldots, x_m, x_{m+1}, \ldots, x_r) dx_r. \quad (9)$$

En particular, con ayuda de las fórmulas del tipo (8) y (9), de una distribución r-dimensional se puede obtener distribuciones unidimensionales de diferentes componentes:

$$F_{\xi_1}(x_1) = F_{\xi_1 \dots \xi_r}(x_1, +\infty, \dots, +\infty),$$
$$F_{\xi_2}(x_2) = F_{\xi_1 \dots \xi_r}(+\infty, x_2, +\infty, \dots, +\infty),$$
....
$$F_{\xi_r}(x_r) = F_{\xi_1 \dots \xi_r}(+\infty, \dots, +\infty, x_r),$$

y

$$p_{\xi_1}(x) = \int_{R^{r-1}} p_{\xi_1 \dots \xi_r}(x, x_2, \dots, x_r)dx_2 \dots dx_r,$$

$$p_{\xi_2}(x) = \int_{R^{r-1}} p_{\xi_1 \dots \xi_r}(x_1, x, x_3, \dots, x_r)dx_1 dx_3 \dots dx_r,$$
....
$$p_{\xi_r}(x) = \int_{R^{r-1}} p_{\xi_1 \dots \xi_r}(x_1, \dots, x_{r-1}x)dx_1 \dots dx_{r-1}.$$

6.2. Independencia de variables aleatorias. Las variables aleatorias ξ_1, $\xi_2, \dots, \xi_r$ se llaman *independientes* si para conjuntos numéricos cualesquiera $B_1, B_2, \dots, B_r$, para los cuales están determinadas las probabilidades de los sucesos $\{\xi_i \in B_i\}$, tiene ligar la igualdad

$$\mathbf{P}\{\xi_1 \in B_1, \xi_2 \in B_2, \dots, \xi_r \in B_r\} = \\ = \mathbf{P}\{\xi_1 \in B_1\}\mathbf{P}\{\xi_2 \in B_2\} \dots \mathbf{P}\{\xi_r \in B_r\}. \quad (10)$$

Si la igualdad (10) se altera para cualesquiera conjuntos B_i, las variables aleatorias $\xi_1, \dots, \xi_r$ se llaman *dependientes*.

En particular, si $B_i = \{y_i : y_i \leqslant x_i\}$, entonces para las variables aleatorias independientes

$$F_{\xi_1 \dots \xi_r}(x_1, x_2, \dots, x_r) = F_{\xi_1}(x_1)F_{\xi_2}(x_2) \dots F_{\xi_r}(x_r), \quad (11)$$

o sea, la función de distribución r-dimensional es igual al producto de las funciones de distribución unidimensionales. Se puede mostrar que la condición (11) puede ser tomada por definición de la independencia en el caso general. Si existe una densidad r-dimensional $p_{\xi_1 \dots \xi_r}(x_1, \dots, x_r)$, entonces con ayuda de la derivación de la igualdad (11) respecto a $x_1, \dots, x_r$ obtenemos que para las variables aleatorias independientes

$$p_{\xi_1 \dots \xi_r}(x_1, \dots, x_r) = p_{\xi_1}(x_1)p_{\xi_2}(x_2) \dots p_{\xi_r}(x_r). \quad (12)$$

En caso de una distribución discreta, de la definición de la independencia (10) se deduce la igualdad

$$\mathbf{P}\{\xi_1 = x_1, \xi_2 = x_2, \dots, \xi_r = x_r\} = \\ = \mathbf{P}\{\xi_1 = x_1\}\mathbf{P}\{\xi_2 = x_2\} \dots \mathbf{P}\{\xi_r = x_r\}. \quad (13)$$

Las condiciones (12) y (13) son equivalentes a la definición de la independencia (10) en el caso de una distribución continua y discreta, respectivamente. En efecto, si se cumple la igualdad (12), entonces

$$\mathbf{P}\{\xi_1 \in B_1, \dots, \xi_r \in B_r\} = \\ = \int_{B_1} dx_1 \dots \int_{B_r} p_{\xi_1 \dots \xi_r}(x_1, \dots, x_r)dx_1 \dots dx_r = \\ = \int_{B_1} p_{\xi_1}(x_1)dx_1 \cdot \int_{B_2} p_{\xi_2}(x_2)dx_2 \dots \int_{B_r} p_{\xi_r}(x_r)dx_r = \\ = P\{\xi_1 \in B_1\} \dots P\{\xi_r \in B_r\}.$$

Análogamente se demuestra la equivalencia de las condiciones de independencia (10) y (13) para distribuciones discretas.

En el punto 6.1 hemos mostrado que la distribución multidimensional determina las distribuciones unidimensionales. En el caso general, con ayuda de las distribuciones unidimensionales $\xi_1, \dots, \xi_r$ no se puede reconstruir la distribución multidimensional. Tal reconstrucción se puede hacer solamente en el caso en que $\xi_1, \dots, \xi_r$ son independientes. Al construir los modelos de fenómenos aleatorios esta circunstancia se utiliza cuando es sabido que los fenómenos aleatorios relacionados con las variables aleatorias son causalmente independientes. Entonces, como hemos mostrado en el § 3, al construir el espacio probabilístico general en el cual están definidos todos los valores de ξ_i estas variables aleatorias es natural considerarlas independientes en el sentido teórico probabilístico.

Las variables aleatorias independientes poseen la propiedad común siguiente.

TEOREMA 1. *Si las variables aleatorias $\eta_1, \dots, \eta_r$, son las funciones $\eta_i = g_i(\xi_i)$ de las variables aleatorias $\xi_1, \dots, \xi_r$, entonces ellas son independientes.*

DEMOSTRACIÓN. Examinemos los conjuntos numéricos B_i, para los cuales $\{\eta_i \in B_i\}$ son sucesos. Designemos con B_i' tal conjunto numérico que $y_i = g_i(x_i) \in B_i'$ en aquellos puntos x_i, y solamente en ellos, para los cuales $x_i \in B_i'$. Entonces

$$\{\eta_i \in B_i\} = \{g(\xi_i) \in B_i\} = \{\xi_i \in B_i'\}.$$

Como $\xi_1, \ldots, \xi_r$ son independientes, entonces, según (10), los sucesos $\{\xi_i \in B_i'\}$ son independientes; por consiguiente, los sucesos $\{\eta_i \in B_i\}$ y esto significa que las variables aleatorias $\eta_1, \ldots, \eta_r$ son independientes. El teorema queda demostrado.

6.3. Convolución de distribuciones. Si se da una distribución multidimensional discreta o continua $\mathbf{P}_{\xi_1 \ldots \xi_r}(B)$, entonces con ayuda de las fórmulas (2) ó (3) se puede obtener la distribución de la variable aleatoria $\eta = g(\xi_1, \ldots, \xi_r)$, donde $g(x_1, \ldots, x_r)$ es la función numérica de $x_1, \ldots, x_r$. Examinemos un caso particular importante cuando las variables aleatorias ξ_1, ξ_2 son independientes y $g(x_1, x_2) = x_1 + x_2$. Si hay densidades $p_{\xi_1}(x_1)$ y $p_{\xi_2}(x_2)$, entonces en virtud de la independencia $p_{\xi_1\xi_2}(x_1, x_2) = p_{\xi_1}(x_1)p_{\xi_2}(x_2)$. Calculemos con ayuda de la fórmula (3) la función de distribución de la suma $\xi_1 + \xi_2$, poniendo

$$F_{\xi_1 + \xi_2}(z) = P\{\xi_1 + \xi_2 \leqslant z\} =$$

$$= \iint\limits_{x_1 + x_2 \leqslant z} p_{\xi_1}(x_1)p_{\xi_2}(x_2)dx_1dx_2 =$$

$$= \int\limits_{-\infty}^{\infty} p_{\xi_1}(x_1)dx_1 \int\limits_{-\infty}^{z - x_1} p_{\xi_2}(x_2)dx_2.$$

Efectuando en la integral interior la sustitución $x_2 = u - x_1$ y cambiando el orden de integración, obtenemos

$$F_{\xi_1 + \xi_2}(z) = \int\limits_{-\infty}^{z} du \int\limits_{-\infty}^{\infty} p_{\xi_1}(x_1)p_{\xi_2}(u - x_1)dx_1.$$

Derivando esta igualdad respecto a z, llegamos a la fórmula *de convolución o de composición para las densidades*:

$$p_{\xi_1 + \xi_2}(z) = \int\limits_{-\infty}^{\infty} p_{\xi_1}(x_1)p_{\xi_2}(z - x_1)dx_1. \tag{14}$$

Conociendo las densidades de dos variables aleatorias ξ_1 y ξ_2, con ayuda de la fórmula (14) se puede encontrar la densidad de su suma $\xi_1 + \xi_2$. Al utilizar la fórmula (14) es necesario tener presente que las densidades se definen con frecuencia por diferentes fórmulas analíticas en diferentes partes.

EJEMPLO 3. Al ir al instituto un estudiante toma el metro y un autobús. En el metro tiene que esperar un tren no más de dos minutos, la espera del autobús dura diez minutos, como máximo. Considerando los tiempos de espera ξ y η del tren y del autobús como variables aleatorias independientes distribuidas uniformemente en los intervalos [0, 2] y [0, 10], respectivamente, hallar la densidad de distribución de la espera total $\xi + \eta$. Las densidades $p_\xi(x)$ y $p_\eta(x)$ están definidas por las fórmulas

$$p_\xi(x) = \begin{cases} 1/2, & 0 \leqslant x \leqslant 2, \\ 0 & \text{en los demás casos;} \end{cases}$$

$$p_\eta(x) = \begin{cases} 1/10, & 0 \leqslant x \leqslant 10, \\ 0 & \text{en los demás casos.} \end{cases}$$

Por la fórmula (14) tenemos

$$p_{\xi+\eta}(x) = \int_{-\infty}^{\infty} p_\xi(u) p_\eta(x-u)\,du = \frac{1}{2}\int_0^2 p_\eta(x-u)\,du,$$

puesto que $p_\xi(u) = 0$ cuando $u < 0$ y $u > 2$. Cuando $x < 0$, $p_\eta(x - u) = 0$, por eso $p_{\xi+\eta}(x) = 0$. Cuando $0 \leqslant x \leqslant 2$,

$$p_{\xi+\eta}(x) = \frac{1}{2}\int_0^x p_\eta(x-u)\,du = \frac{1}{2}\cdot\frac{1}{10}\int_0^x du = \frac{x}{20}.$$

Cuando $2 \leqslant x \leqslant 10$,

$$p_{\xi+\eta}(x) = \frac{1}{2}\int_0^x p_\eta(x-u)\,du = \frac{2}{20} = \frac{1}{10}.$$

Cuando $10 \leqslant x \leqslant 12$,

$$p_{\xi+\eta}(x) = \frac{1}{2}\int_{x-10}^{2} p_\eta(x-u)\,du = \frac{12-x}{20}.$$

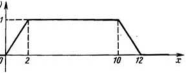

Fig. 12.

Y, por fin, para $x > 12$ obtenemos $p_{\xi+\eta}(x) = 0$. El gráfico de la densidad $p_{\xi+\eta}(x)$ se muestra en la fig. 12.

Para las variables aleatorias independientes ξ_1 y ξ_2 con distribución discreta existe la fórmula de convolución o composición, análoga a (14):

$$\mathbf{P}\{\xi + \eta = x\} = \sum_{x_i} \mathbf{P}\{\xi = x_i\}\mathbf{P}\{\eta = x - x_i\}, \tag{15}$$

donde x_i son los puntos en los cuales $\mathbf{P}\{\xi = x_i\} > 0$. La demostración de (15) se obtiene fácilmente de la fórmula (2) y de la condición de independencia (13).

EJEMPLO 4. Supongamos que las variables aleatorias ξ_1 y ξ_2 son independientes y están distribuidas uniformemente con parámetros (a_1, σ_1) y (a_2, σ_2), respectivamente. Hallemos la ley de distribución de $\xi = \xi_1 + \xi_2$. Sustituyendo en la fórmula (14) las densidades de distribución de ξ_1 y ξ_2, obtenemos

$$p_\xi(x) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(u-a_1)^2}{2\sigma_1^2}} \cdot \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(x-u-a_2)^2}{2\sigma_2^2}} du.$$

De aquí, puesto que

$$\frac{(u-a_1)^2}{\sigma_1^2} + \frac{(x-u-a_2)^2}{\sigma_2^2} =$$

$$= \frac{\sigma_1^2+\sigma_2^2}{\sigma_1^2\sigma_1^2}\left(u - \frac{a_1\sigma_2^2 + (x-a_2)\sigma_1^2}{\sigma_1\sigma_2\sqrt{\sigma_1^2+\sigma_2^2}}\right) + \frac{(x-a_1-a_2)^2}{\sigma_1^2+\sigma_2^2},$$

resulta que

$$p_\xi(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}} \cdot \frac{\sigma}{\sqrt{2\pi}\sigma_1\sigma_2} \int_{-\infty}^{\infty} e^{-\frac{\sigma^2(u-C)^2}{2\sigma_1^2\sigma_2^2}} du,$$

donde $a = a_1 + a_2$, $\sigma^2 = \sigma_1^2 + \sigma_2^2$ y

$$C = \frac{a_1\sigma_2^2 + (x-a_2)\sigma_1^2}{\sigma_1\sigma_2\sqrt{\sigma_1^2+\sigma_2^2}}.$$

Reemplazando la variable de integración según la fórmula $y = \sigma(u - C)/\sigma_1\sigma_2$, obtenemos

$$p_\xi(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}}, \tag{16}$$

ya que

$$\frac{\sigma}{\sqrt{2\pi}\sigma_1\sigma_2} \int_{-\infty}^{\infty} e^{-\frac{\sigma^2(u-C)^2}{2\sigma_1^2\sigma_2^2}} du = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} du = 1.$$

La fórmula obtenida (16) define la densidad de la distribución normal con parámetros (a, σ). Ahora bien, *la suma de variables aleatorias independientes normalmente distribuidas tiene una distribución normal.*

EJEMPLO 5. Supongamos que las variables aleatorias ξ_1 y ξ_2 son independientes y están distribuidas según la ley de Poisson:

$$\mathbf{P}\{\xi_1 = k\} = \frac{\lambda_1^k}{k!} e^{-\lambda_1}, \quad \mathbf{P}\{\xi_2 = l\} = \frac{\lambda_2^l}{l!} e^{-\lambda_2}. \tag{17}$$

Sustituyendo en la fórmula (15) las probabilidades (17), encontramos

$$\mathbf{P}\{\xi_1 + \xi_2 = m\} = \sum_{n=0}^{m} \frac{\lambda_1^n}{n!} e^{-\lambda_1} \cdot \frac{\lambda_2^{m-n}}{(m-n)!} e^{-\lambda_2} =$$

$$= \frac{1}{m!} e^{-(\lambda_1 + \lambda_2)} \sum_{n=0}^{m} \frac{m!}{n!(m-n)!} \lambda_1^n \lambda_2^{m-n} =$$

$$= \frac{1}{m!} e^{-(\lambda_1 + \lambda_2)} \cdot (\lambda_1 + \lambda_2)^m.$$

De aquí que, suponiendo $\lambda = \lambda_1 + \lambda_2$, obtenemos

$$\mathbf{P}\{\xi_1 + \xi_2 = m\} = \frac{\lambda^m}{m!} e^{-\lambda},$$

o sea, *la suma de variables aleatorias independientes distribuidas según la ley de Poisson está distribuida asimismo según esta ley.*

Problemas

1. La densidad bidimensional $p_{\xi\eta}(x, y)$ para los puntos que satisfacen las condiciones $x + y \leqslant 1, x \geqslant 0, y \geqslant 0$ es igual a 2.

a) Hallar la probabilidad del suceso $\xi^2 + \eta^2 \leqslant 1/2$.

b) Mostrar que ξ y η son dependientes. (*Indicación*. Hallar la posibilidad de los sucesos $\{1/2 \leqslant \xi \leqslant 1, 1/2 \leqslant \eta \leqslant 1\}$, $\{1/2 \leqslant \xi \leqslant 1\}$, $\{1/2 \leqslant \eta \leqslant 1\}$.)

2. Las caras de un dado llevan los números de orden de modo que la suma de los puntos marcados en las caras contrarias siempre sea igual a 7. Sea ξ el número de los puntos salido en la cara superior y η, el número de los puntos en la cara inferior. Hallar la ley de distribución de $\zeta = \xi\eta$.

3. Hallar la función y la densidad de distribución de $\theta = \max(\xi, \eta)$, si ξ y η son independientes y están distribuidas uniformemente sobre $[0, 1]$.

4. Las variables aleatorias ξ y η son independientes y tienen la misma distribución exponencial con la densidad $p_\xi(x) = p_\eta(x) = \lambda e^{-\lambda x}, x \geqslant 0$.

Hallar la densidad $p_{\xi+\eta}(x)$.

§ 7. Esperanza matemática

7.1. Esperanza matemática en el esquema finito. Supongamos que sobre un espacio probabilístico finito $(\Omega, \mathscr{A}, \mathbf{P})$ se asigna una variable aleatoria $\xi = \xi(\omega)$. Una de las características numéricas más importantes de la variable aleatoria ξ es su esperanza matemática.

Se llama esperanza matemática de la variable aleatoria ξ en el esquema finito el número designado con $\mathbf{M}\xi$ *e igual a*

$$\mathbf{M}\xi = \sum_{\omega \in \Omega} \xi(\omega) p(\omega), \tag{1}$$

donde $p(\omega)$ son las probabilidades elementales que prefijan la probabilidad $\mathbf{P}$.

La fórmula (1) muestra que la esperanza matemática $\mathbf{M}\xi$ se define como número igual a la suma de los productos de los valores $\xi(\omega)$ de la variable aleatoria ξ por las probabilidades elementales respectivas $p(\omega)$. Esta suma se toma teniendo en cuenta todos los sucesos elementales ω del espacio Ω. En particular, si se atribuye un número de orden a cada suceso elemental y se pone $\Omega = \{\omega_1, \ldots, \omega_n\}$, entonces la fórmula (1) se puede escribir en forma

$$\mathbf{M}\xi = \xi(\omega_1)p(\omega_1) + \xi(\omega_2)p(\omega_2) = + \ldots + \xi(\omega_n)p(\omega_n).$$

En vez del término esperanza matemática de la variable aleatoria ξ se utiliza, a veces, el término *valor medio* de ξ o bien, todavía más brevemente, *media* de ξ.

EJEMPLO 1. Supongamos que en una lotería se sortean N billetes, con ello N_1 billetes tienen, cada uno, el premio a_1; N_2 billetes, el premio a_2, ..., N_r billetes el premio a_r. (Por ejemplo, en una lotería de libros compuesta de 1000 billetes la mayor parte de billetes tienen el premio $a_1 = 0$ rublos y los demás billetes dan los premios siguientes: $a_2 = 0{,}5$ rublos, $a_3 = 1$ rublo, $a_4 = 5$ rublos, etc.) Examinemos la variable aleatoria ξ, igual al premio que recibe un participante que ha comprado un billete. Existen diferentes sistemas de sorteo de una lotería. Por ejemplo, en una lotería con premios en dinero y cosas se venden billetes que llevan un número de orden respectivo y luego, con sorteos especiales, se determinan billetes premiados. Para sencillez, vamos a examinar un otro sistema adoptado en una lotería de libros y en la lotería "Sprint" en las cuales los billetes se venden cerrados y de antemano llevan marcado un premio. Puesto que exteriormente todos los billetes parecen iguales, en este caso podemos definir el espacio probabilístico del modo siguiente. Supongamos que el suceso elemental ω es un billete de lotería y el conjunto de todos los sucesos elementales $\Omega = \{\omega\}$ tiene la potencia $|\Omega| = N$. En virtud de la identidad exterior de los billetes de lotería las probabilidades elementales $p(\omega)$ son todas iguales entre sí $p(\omega) = 1/N$, por eso llegamos a la definición clásica de la probabilidad $\mathbf{P}$ según la cual $\mathbf{P}(A) = \dfrac{|A|}{|\Omega|}$. Sea que en el conjunto A_i de los billetes de lotería hay el premio a_i y $|A_i| = N_i$. Entonces la variable aleatoria $\xi = \xi(\omega)$ se puede prefijar, con ayuda de la partición

$$\Omega = A_1 + A_2 + \dots + A_r,$$

mediante las igualdades

$$\xi(\omega) = a_i, \quad \text{si} \quad \omega \in A_i.$$

En este caso la esperanza matemática $\mathbf{M}\xi$ se determina por la fórmula (1) del modo siguiente:

$$\begin{aligned}\mathbf{M}\xi = \sum_{\omega\in\Omega} \xi(\omega)p(\omega) &= \sum_{\omega\in A_1} \xi(\omega)p(\omega) + \\ &+ \sum_{\omega\in A_2} \xi(\omega)p(\omega) + \dots + \sum_{\omega\in A_r} \xi(\omega)p(\omega) = \\ &= a_1 \sum_{\omega\in A_1} p(\omega) + a_2 \sum_{\omega\in A_2} p(\omega) + \dots + a_r \sum_{\omega\in A_r} p(\omega) = \\ &\qquad = a_1\frac{N_1}{N} + a_2\frac{N_2}{N} + \dots + a_r\frac{N_r}{N},\end{aligned}$$

ya que $\xi(\omega) = a_i$ para todos los valores de $\omega \in A_i$. Del ejemplo citado llega a ser claro el origen del término *valor medio de* ξ puesto que la suma $\sum_{k=1}^{r} a_k N_k$ es la colección de premios de toda la lotería y su relación al número total de los billetes N es el premio medio para un billete.

De la definición de la esperanza matemática se deducen fácilmente sus propiedades siguientes.

1. *Aditividad*: $\mathbf{M}(\xi + \eta) = \mathbf{M}\xi + \mathbf{M}\eta$, o sea, *la esperanza matemática de la suma de las variables aleatorias es igual a la suma de las esperanzas matemáticas de los sumandos.*

DEMOSTRACIÓN. De la definición de la esperanza matemática tenemos

$$\mathbf{M}(\xi + \eta) = \sum_{\omega \in \Omega} (\xi(\omega) + \eta(\omega))p(\omega) =$$

$$= \sum_{\omega \in \Omega} \xi(\omega)p(\omega) + \sum_{\omega \in \Omega} \eta(\omega)p(\omega) = \mathbf{M}\xi + \mathbf{M}\eta.$$

Esta propiedad se extiende a todo número finito de los sumandos:

$$\mathbf{M}(\xi_1 + \xi_2 + \ldots + \xi_k) = \mathbf{M}\xi_1 + \mathbf{M}\xi_2 + \ldots + \mathbf{M}\xi_k.$$

2. *Para todo número c*

$$\mathbf{M}(c\xi) = c\mathbf{M}\xi,$$

o sea, *el factor constante c se puede sacar del signo de la esperanza matemática.*

DEMOSTRACIÓN. De la definición (1) tenemos

$$\mathbf{M}(c\xi) = \sum_{\omega \in \Omega} c\xi(\omega)p(\omega) = c \sum_{\omega \in \Omega} \xi(\omega)p(\omega) = c\mathbf{M}\xi.$$

El conjunto de las propiedades 1ª y 2ª se llama *propiedad de linealidad* de la esperanza matemática y se expresa por la igualdad siguiente:

$$\mathbf{M}(c_1\xi_1 + c_2\xi_2 + \ldots + c_k\xi_k) =$$

$$= c_1\mathbf{M}\xi_1 + c_2\mathbf{M}\xi_2 + \ldots + c_k\mathbf{M}\xi_k \qquad (2)$$

que es válida para todas variables aleatorias ξ_i y para todos números c_i.

3. *La esperanza matemática del indicador χ_A del suceso A es igual a la probabilidad de este suceso:*

$$\mathbf{M}\chi_A = \mathbf{P}(A).$$

DEMOSTRACIÓN. Como $\mathbf{P}(A) = \sum_{\omega \in A} p(\omega)$, entonces

$$\mathbf{M}\chi_A = \sum_{\omega \in \Omega} \chi_A(\omega) p(\omega) = \sum_{\omega \in A} 1 \cdot p(\omega) + \\ + \sum_{\omega \in \bar{A}} 0 \cdot p(\omega) = \sum_{\omega \in A} p(\omega) = \mathbf{P}(A).$$

4. *Propiedad de monotonía. Si* $\xi \geqslant \eta$, *entonces* $\mathbf{M}\xi \geqslant \mathbf{M}\eta$.

DEMOSTRACIÓN. Demostremos primeramente que de $\xi \geqslant 0$ resulta $\mathbf{M}\xi \geqslant 0$. En efecto, como para cada valor de $\omega \in \Omega$ $\xi(\omega) \geqslant 0$ y $p(\omega) \geqslant 0$, entonces la suma $\mathbf{M}\xi = \sum_{\omega \in \Omega} \xi(\omega) p(\omega)$ también será no negativa. Aplicando la propiedad demostrada a la diferencia no negativa $\xi - \eta \geqslant 0$, obtenemos $\mathbf{M}(\xi - \eta) = \mathbf{M}\xi - \mathbf{M}\eta \geqslant 0$ que es lo que se necesitaba demostrar.

Fórmulas para calcular la esperanza matemática

La anteriormente citada definición de la esperanza matemática $\mathbf{M}\xi$ de una variable aleatoria ξ no siempre es cómoda para el cálculo. Como hemos visto en el § 6, para calcular las probabilidades $\mathbf{P}\{\xi \in B\}$ no debemos obligatoriamente conocer el espacio probabilístico $(\Omega, \mathscr{A}, \mathbf{P})$ y la función $\xi(\omega)$ que asigna la variable aleatoria ξ, sino basta conocer sólo la ley de distribución de ξ, o sea, la colección de todos los valores $x_1, \ldots, x_k$ que toma la variable aleatoria ξ y las probabilidades respectivas $\mathbf{P}\{\xi = x_i\}$, $i = 1, 2, \ldots, k$. La esperanza matemática $\mathbf{M}\xi$ se calcula con ayuda de la ley de distribución de ξ por la fórmula siguiente:

$$\mathbf{M}\xi = \sum_{i=1}^{k} x_i \mathbf{P}\{\xi = x_i\}. \tag{3}$$

Para demostrar la fórmula (3) hagámonos uso de la representación de la variable aleatoria ξ por la partición $A_1 + A_2 + \ldots + A_k = \Omega$, donde $A_i = \{\omega : \xi(\omega) = x_i\}$ (véase el § 6):

$$\xi = \sum_{i=1}^{k} x_i \chi_{A_i}. \tag{4}$$

Aplicando a la suma (4) la propiedad de linealidad y la propiedad 3, tenemos

$$\mathbf{M}\xi = \mathbf{M}\left(\sum_{i=1}^{k} x_i \chi_{A_i}\right) = \sum_{i=1}^{k} x_i \mathbf{M}\chi_{A_i} = \sum_{i=1}^{k} x_i \mathbf{P}\{\xi = x_i\}.$$

Así, pues, hemos demostrado que la esperanza matemática de la variable aleatoria ξ es igual a la suma de los productos de sus valores x_i por las probabilidades $\mathbf{P}\{\xi = x_i\}$ de que ξ tome el valor respectivo.

Razonando análogamente, no es difícil obtener las siguientes fórmulas para calcular la esperanza matemática de las variables aleatorias que tienen la forma $g(\xi)$, $g(\xi, \eta)$, donde $g(x)$ y $g(x, y)$ son las funciones numéricas y ξ, η son las variables aleatorias:

$$\mathbf{M}g(\xi) = \sum_{i=1}^{k} g(x_i)\mathbf{P}\{\xi = x_i\}, \tag{5}$$

$$\mathbf{M}g(\xi, \eta) = \sum_{i=1}^{k} \sum_{j=1}^{m} g(x_i, y_i)\mathbf{P}\{\xi = x_i, \eta = y_j\}. \tag{6}$$

Para obtener las fórmulas (5) y (6) es necesario utilizar las representaciones de $g(\xi)$ y $g(\xi, \eta)$ en forma de las sumas siguientes:

$$g(\xi) = \sum_{i=1}^{k} g(x_i)\chi_{A_i},\ g(\xi, \eta) = \sum_{i=1}^{k} \sum_{j=1}^{m} g(x_i, y_j)\chi_{A_{ij}},$$

donde $\{A_i\}$ es la partición anteriormente introducida y $\{A_{ij}\}$ es la partición formada por los conjuntos $A_{ij} = \{\omega: \xi(\omega) = x_i,\ \eta(\omega) = y_j\}$. Una fórmula análoga a (6) tampoco es difícil obtenerla para $\mathbf{M}g(\xi_1, \dots, \xi_r)$. Puesto que según la fórmula (3) la esperanza matemática $\mathbf{M}\xi$ de la variable aleatoria ξ se determina unívocamente por su ley de distribución, en vez del término esperanza matemática de la variable aleatoria ξ se usa el término *esperanza matemática de la ley de distribución de* ξ.

EJEMPLO 2. Hallar la esperanza matemática de la ley binomial de distribución $\mathbf{P}\{\mu = k\} = C_n^k p^k q^{n-k}$, $k = 0, 1, \dots, n$, $q = 1 - p$, $0 \leqslant p \leqslant 1$. Conforme a (3), tenemos

$$\mathbf{M}\mu = \sum_{k=0}^{n} k\mathbf{P}\{\mu = k\} = \sum_{k=0}^{n} kC_n^k p^k q^{n-k}.$$

En virtud de la igualdad $kC_n^k = k\dfrac{n!}{k!(n-k)!} = nC_{n-1}^{k-1}$, la suma que está a la derecha se calcula del modo siguiente:

$$\sum_{k=0}^{n} kC_n^k p^k q^{n-k} = np \sum_{k=1}^{n} C_{n-1}^{k-1} p^{k-1} q^{n-k} = np,$$

ya que $\sum_{k=1}^{n} C_{n-1}^{k-1} p^{k-1} q^{n-k} = (p+q)^{n-1} = 1$. Ahora bien, $\mathbf{M}\mu = np$.

Propiedad multiplicativa. Para variables aleatorias independientes ξ y η tiene lugar la propiedad multiplicativa

$$\mathbf{M}\xi\eta = \mathbf{M}\xi \cdot \mathbf{M}\eta, \tag{7}$$

o sea, *la esperanza matemática del producto de variables aleatorias independientes es igual al producto de sus esperanzas matemáticas.*

DEMOSTRACIÓN. Hagamos uso de la fórmula (6):

$$\mathbf{M}\xi\eta = \sum_{i=1}^{k} \sum_{j=1}^{m} x_i y_j \mathbf{P}\{\xi = x_i, \eta = y_j\}.$$

Para las variables independientes ξ y η tiene lugar la igualdad $\mathbf{P}\{\xi = x_i, \eta = y_j\} = \mathbf{P}\{\xi = x_i\}\mathbf{P}\{\eta = y_j\}$, por eso

$$\mathbf{M}\xi\eta = \sum_{i=1}^{k} \sum_{j=1}^{m} x_i y_j \mathbf{P}\{\xi = x_i\}\mathbf{P}\{\eta = y_j\} =$$

$$= \sum_{i=1}^{k} x_i \mathbf{P}\{\xi = x_i\} \sum_{j=1}^{m} y_j \mathbf{P}\{\eta = y_j\} = \mathbf{M}\xi\mathbf{M}\eta.$$

La propiedad de multiplicatividad se extiende, naturalmente, al caso de un número finito arbitrario de variables aleatorias $\xi_1, \xi_2, \dots, \xi_r$:

$$\mathbf{M}\xi_1\xi_2 \dots \xi_r = \mathbf{M}\xi_1\mathbf{M}\xi_2 \dots \mathbf{M}\xi_r. \tag{8}$$

Cabe señalar que si la propiedad de aditividad de la esperanza matemática es válida para variables aleatorias cualesquiera, la de multiplicatividad de la misma es justa para variables aleatorias independientes.

Métodos de calcular las esperanzas matemáticas. Al calcular $\mathbf{M}\xi$ o $\mathbf{M}g(\xi)$ se puede utilizar la fórmula (1) que da la definición de la esperanza matemática o bien las fórmulas (3), (5) y (6) que expresan la esperanza matemática por medio de la ley de distribución. Sin embargo, a veces la variable aleatoria ξ o la función de ésta $g(\xi)$ se asignan de modo que la ley respectiva de distribución quede desconocida o sea muy complicada. Algunas veces la ley de distribución $\mathbf{P}\{\xi = x_i\}$ se prefija mediante una fórmula buena, pero la suma en (3), por cuyo medio se define $\mathbf{M}\xi$, es muy complicada y difícilmente se presta a la simplificación. En estos casos con frecuencia se puede calcular rápida y fácilmente $\mathbf{M}\xi$ con ayuda de la propiedad de aditividad. Para esto es necesario representar ξ en forma de la suma de sumandos más simples $\xi_1 + \xi_2 + \dots + \xi_r$, por ejemplo en forma de los indicadores de ciertos sucesos, cuyas probabilidades se conocen. Entonces $\mathbf{M}\xi = \mathbf{M}\xi_1 + \mathbf{M}\xi_2 + \dots + \mathbf{M}\xi_r$.

EJEMPLO 3. Calcular la esperanza matemática de la distribución binomial por el método propuesto anteriormente. Para esto valgámonos del hecho de que la distribución binomial la tiene la variable aleatoria μ, igual al número de los casos favorables en n pruebas según el esquema de Bernoulli con la probabilidad de un caso favorable p en cada experimento. Designemos con A_i el suceso consistente en que en la i-ésima prueba se ha producido el caso favorable. Entonces

$$\mu = \chi_{A_1} + \chi_{A_2} + \dots + \chi_{A_n}$$

y

$$\mathbf{M}\mu = \sum_{i=1}^{n} \mathbf{M}\chi_{A_i} = \sum_{i=1}^{n} \mathbf{P}(A_i) = np,$$

o sea, hemos obtenido el mismo resultado que en el ejemplo 2, pero de un modo más simple y evidente.

EJEMPLO 4. Calcular $\mathbf{M}\xi$ en la distribución hipergeométrica

$$\mathbf{P}\{\xi = k\} = \frac{C_M^k C_{N-M}^{n-k}}{C_N^n}, \quad k = 0, 1, \dots, n.$$

Hagamos uso del mismo método que en el ejemplo 3. La distribución hipergeométrica aparece en un esquema con urnas en el cual de una urna que contiene M bolas blancas y $(N - M)$ negras se realiza un muestreo sin reemplazamiento del volumen n. La variable aleatoria ξ es el número de las bolas blancas en la muestra. Supongamos que las bolas se sacan de la urna de un modo sucesivo. Designemos el suceso $A_i = \{$la i-ésima bola de la muestra es blanca$\}$. Entonces

$$\xi = \chi_{A_1} + \chi_{A_2} + \dots + \chi_{A_n}$$

y conforme a la propiedad de aditividad $\mathbf{M}\xi = \sum_{i=1}^{n} \mathbf{P}(A_i)$. Como $\mathbf{P}(A_i) = \dfrac{M}{N}$ para todo valor de i (véase el ejemplo 5 en el § 1), entonces

$$\mathbf{M}\xi = n\frac{M}{N}.$$

Al calcular las esperanzas matemáticas es útil, a veces, usar también la propiedad de multiplicatividad.

EJEMPLO 5. En el esquema de Bernoulli con n pruebas y con la probabilidad de un caso favorable p vamos a suponer que en la i-ésima prueba se produce un caso favorable doble, si tales casos han aparecido en la prueba $(i - 1)$ y en la i-ésima prueba. Designemos con ν_n el número de casos favorables dobles en toda la serie de experimentos. Hallemos $\mathbf{M}\nu_n$. Representemos $\nu_n = \eta_2 + \eta_3 + \ldots + \eta_n$, donde $\eta_i = 1$, si en la i-ésima prueba se ha producido un caso favorable doble y $\eta_i = 0$ en la ocasión contraria. No es difícil de ver que $\eta_i = \chi_{A_{i-1}} \cdot \chi_{A_i}$, donde A_i son los sucesos del ejemplo 3. De la propiedad de aditividad tenemos

$$\mathbf{M}\nu_n = \sum_{i=2}^{n} \mathbf{M}\chi_{A_{i-1}}\chi_{A_i}.$$

Como $\chi_{A_{i-1}}$ y χ_{A_i} son independientes, de la propiedad de multiplicatividad $\mathbf{M}\chi_{A_{i-1}}\chi_{A_i} = \mathbf{M}\chi_{A_{i-1}} \cdot \mathbf{M}\chi_{A_i} = p^2$. Finalmente tenemos $\mathbf{M}\nu_n = (n - 1)p^2$.

7.2. Esperanza matemática en el esquema numerable. Examinemos el espacio probabilístico $(\Omega, \mathscr{A}, \mathbf{P})$ con el conjunto numerable Ω de sucesos elementales. Asignemos los números de orden a los sucesos elementales ω_n, $n = 1, 2, \ldots$ Supongamos que la propiedad $\mathbf{P}$ se define por las propiedades elementales $p(\omega_n)$.

Se llama esperanza matemática $\mathbf{M}\xi$ *de la variable aleatoria* ξ *en el esquema finito la suma de la serie*

$$\mathbf{M}\xi = \sum_{n=1}^{\infty} \xi(\omega_n)p(\omega_n), \tag{9}$$

si esta serie converge absolutamente, o sea, si

$$\sum_{n=1}^{\infty} |\xi(\omega_n)| \cdot p(\omega_n) < \infty. \tag{10}$$

En el caso contrario se dice que $\mathbf{M}\xi$ no existe.

La condición (10) es muy importante. En efecto, si la serie (9) converge no absolutamente, entonces, como es sabido del análisis matemático, una serie obtenida por la reordenación de la serie inicial puede converger hacia un número cualquiera y puesto que la numeración de los sucesos elementales ω_n es arbitraria y no tiene ninguna relación con la esencia del modelo en cuestión, entonces en este caso la suma (9) tampoco expresa nada y pierde su significación objetiva. Por eso creemos que en este caso la esperanza matemática no existe.

La suma de la serie absolutamente convergente (9) es análoga, por sus propiedades, a la suma finita (1), por eso las propiedades de la esperanza matemática $\mathbf{M}\xi$ determinadas en el punto 7.1 para las variables aleatorias ξ en el esquema finito quedan válidas también para el esquema numerable. En particular, es justa la fórmula, semejante a (3):

$$\mathbf{M}\xi = \sum_{n=1}^{\infty} x_n \mathbf{P}\{\xi = x_n\}, \tag{11}$$

si la serie a la derecha converge absolutamente. La fórmula (11) expresa $\mathbf{M}\xi$ por medio de la ley de distribución de ξ.

EJEMPLO 6. Calculemos $\mathbf{M}\xi$ de una variable aleatoria ξ distribuida por la ley de Poisson $\mathbf{P}\{\xi = k\} = \dfrac{a^k}{k!} e^{-a}$:

$$\mathbf{M}\xi = \sum_{k=0}^{\infty} k\mathbf{P}\{\xi = k\} = \sum_{k=1}^{\infty} k \frac{a^k e^{-a}}{k!} = a \sum_{k=1}^{\infty} \frac{a^{k-1}}{(k-1)!} e^{-a} = a.$$

Las fórmulas (5) y (6) se extienden asimismo para el esquema numerable:

$$\mathbf{M}g(\xi) = \sum_{n=1}^{\infty} g(x_n)\mathbf{P}\{\xi = x_n\}, \tag{12}$$

$$\mathbf{M}g(\xi, \eta) = \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} g(x_n, y_m)\mathbf{P}\{\xi = x_n, \eta = y_m\}; \tag{13}$$

pero para su validez se necesita que las series converjan absolutamente.

7.3. Esperanza matemática en el caso general. Las anteriormente dadas definiciones de la esperanza matemática $\mathbf{M}\xi$ para las variables aleatorias en el esquema finito o numerable se extienden fácilmente a variables aleatorias que tomen un número finito o numerable de valores y estén definidas en un espacio probabilístico cualquiera $(\Omega, \mathscr{A}, \mathbf{P})$. Sea $A_1, \ldots, A_k$ la partición finita, es decir, los sucesos A_i son incompatibles dos a dos y $\sum_{i=1}^{k} A_i = \Omega$. Entonces la variable aleatoria

$$\xi = \sum_{i=1}^{k} x_i \chi_{A_i} \tag{14}$$

toma solamente un número finito de valores. La esperanza matemática $\mathbf{M}\xi$ de una tal variable aleatoria se determina como suma

$$\mathbf{M}\xi = \sum_{i=1}^{k} x_i \mathbf{P}(A_i). \tag{15}$$

En las fórmulas (14) y (15) no impongamos ningunas limitaciones para los valores de x_i. Si todos los valores de x_i son diferentes, entonces la fórmula (15) se convierte en (3), ya que en este caso $A_i = \{\xi = x_i\}$.

Análogamente, toda variable aleatoria discreta se escribe en forma

$$\xi = \sum_{i=1}^{\infty} x_i \chi_{A_i}, \tag{16}$$

donde $\{A_i\}$ es la partición numerable, o sea, A_i son incompatibles dos a dos y $\sum_{i=1}^{\infty} A_i = \Omega$. En este caso la esperanza matemática $\mathbf{M}\xi$ se determina como la suma de la serie

$$\mathbf{M}\xi = \sum_{i=1}^{\infty} x_i \mathbf{P}(A_i), \tag{17}$$

si esta serie converge absolutamente. En el caso contrario la esperanza matemática no existe. Si todos los valores de x_i son diferentes, la definición (17) se transforma en definición (11) que permite calcular $\mathbf{M}\xi$ con ayuda de la ley de distribución. Análogamente, se puede mostrar que para variables aleatorias discretas son válidas las fórmulas (12) y (13). Todas las propiedades de las esperanzas matemáticas, determinadas en el punto 7.1, quedan válidas asimismo en el caso general cuando $\mathbf{M}\xi$ se definen por las fórmulas (15) y (17).

Si una variable aleatoria ξ es continua y su ley de distribución se asigna por la densidad $p_\xi(x)$, entonces su esperanza matemática se calcula con ayuda de la fórmula

$$\mathbf{M}\xi = \int_{-\infty}^{\infty} x p_\xi(x) dx, \tag{18}$$

si la integral converge absolutamente. Esta fórmula es semejante a la fórmula (11) para $\mathbf{M}\xi$ en un caso discreto. Las fórmulas (12) y (13) para calcular $\mathbf{M}g(\xi)$ y $\mathbf{M}g(\xi, \eta)$ se convierten, siempre que se trate de las variables aleatorias continuas, en fórmulas

$$\mathbf{M}g(\xi) = \int_{-\infty}^{\infty} g(x)p_{\xi}(x)dx, \quad (19)$$

$$\mathbf{M}g(\xi, \eta) = \iint_{-\infty}^{\infty} g(x, y)p_{\xi\eta}(x, y)dx\,dy, \quad (20)$$

si las integrales del segundo miembro convergen absolutamente. En (20) $p_{\xi\eta}(x, y)$ es la densidad compatible de las variables aleatorias ξ y η.

OBSERVACIÓN. Las fórmulas (6), (13) y (20) escritas para las distribuciones bidimensionales tienen análogos para las distribuciones multidimensionales.

La definición de $\mathbf{M}\xi$ para una variable aleatoria continua valiéndose de la fórmula (18) no es cómoda, ya que en este caso no se puede demostrar las propiedades de la esperanza matemática. Es más lógico definir $\mathbf{M}\xi$ del modo siguiente. Con cada variable aleatoria ξ se puede relacionar la sucesión de variables aleatorias discretas ξ_n, suponiendo

$$\xi_n = \frac{k}{2^n} \quad \text{si} \quad \frac{k}{2^n} < \xi \leqslant \frac{k+1}{2^n};$$

es fácil de ver que $\xi \leqslant \xi_n \leqslant \xi + \dfrac{1}{2^n}$, por eso $\lim\limits_{n \to \infty} \xi_n = \xi$. Las variables aleatorias ξ_n se pueden escribir en la forma semejante a (16):

$$\xi_n = \sum_{k=-\infty}^{\infty} \frac{k}{2^n} \chi_{\left\{\frac{k}{2^n} < \xi \leqslant \frac{k+1}{2^n}\right\}}$$

Aplicando a ξ_n la fórmula (17), obtenemos

$$\mathbf{M}\xi_n = \sum_{k=-\infty}^{\infty} \frac{k}{2^n} P\left\{\frac{k}{2^n} < \xi \leqslant \frac{k+1}{2^n}\right\}, \quad (21)$$

si la serie del segundo miembro converge absolutamente. Es fácil de ver que $\xi_n \leqslant \xi_{n+1}$, para todo valor de n, por eso $\mathbf{M}\xi_n \leqslant \mathbf{M}\xi_{n+1}$ y existe el límite de $\mathbf{M}\xi_n$, cuyo valor tomamos por definición de la esperanza matemática:

$$\mathbf{M}\xi = \lim_{n \to \infty} \mathbf{M}\xi_n. \quad (22)$$

De la definición (22) no es difícil demostrar la fórmula (18). En efecto, suponiendo $x_{nk} = \dfrac{k}{2^n}$, se puede escribir (21) del modo siguiente:

$$\mathbf{M}\xi_n = \sum_{k=-\infty}^{\infty} \frac{k}{2^n} \int_{x_{nk}}^{x_{n,k+1}} p_{\xi}(u)du,$$

de donde

$$0 \leqslant \int_{-\infty}^{\infty} up_\xi(u)du - \mathbf{M}\xi_n =$$

$$= \sum_{k=-\infty}^{\infty} \int_{x_{nk}}^{x_{n,k+1}} \left(u - \frac{k}{2^n}\right) p_\xi(u)du \leqslant \frac{1}{2^n} \int_{-\infty}^{\infty} p_\xi(u)du = \frac{1}{2^n}$$

y $\lim\limits_{n\to\infty} \mathbf{M}\xi_n = \int\limits_{-\infty}^{\infty} up_\xi(u)du$. Tal definición de la esperanza matemática $\mathbf{M}\xi$ permite extender todas sus propiedades demostradas en el punto 7.1 también para el caso de variables aleatorias continuas, puesto que la validez de estas propiedades para las variables aleatorias discretas ξ_n se determina fácilmente y el paso al límite cuando $n \to \infty$ no altera estas propiedades.

EJEMPLO 7. Calculemos la esperanza matemática para una variable aleatoria ξ distribuida uniformemente en $[a, b]$:

$$\mathbf{M}\xi = \int_{-\infty}^{\infty} xp_\xi(x)dx = \frac{1}{b-a}\int_a^b x\,dx = \frac{a+b}{2}.$$

EJEMPLO 8. La esperanza matemática de la variable aleatoria ξ que tiene distribución normal con la densidad $p_\xi(x) = \dfrac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}}$ se calcula del modo siguiente:

$$\mathbf{M}\xi = \int_{-\infty}^{\infty} xp_\xi(x)dx = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} (x - a + a)e^{-\frac{(x-a)^2}{2\sigma^2}} dx =$$

$$= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} (x-a)e^{-\frac{(x-a)^2}{2\sigma^2}} dx + a\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} e^{-\frac{(x-a)^2}{2\sigma^2}} dx = a.$$

EJEMPLO 9. Supongamos que la variable aleatoria τ es el tiempo del buen funcionamiento de una pieza: $p(x)$ es la densidad de distribución de τ. Calcular el tiempo medio de trabajo de la pieza si se sabe que en todo caso la pieza se reemplaza al pasar un tiempo T. El tiempo de trabajo de la pieza se determina por el valor de la función $g(\tau)$, donde

$$g(x) = \begin{cases} x & \text{cuando } x \leqslant T, \\ T & \text{cuando } x > T. \end{cases}$$

Apliquemos la fórmula (19):

$$\mathbf{M}g(\tau) = \int_0^\infty g(x)p(x)dx = \int_0^T g(x)p(x)dx + T\int_T^\infty p(x)dx =$$

$$= \int_0^T g(x)p(x)dx + T\cdot\mathbf{P}\{\tau > T\}.$$

7.4. Desigualdad de Chébishev. Demostraremos la primera forma de la desigualdad de Chébishev para variables aleatorias no negativas ξ y para $x > 0$:

$$\mathbf{P}\{\xi > x\} \leqslant \frac{\mathbf{M}\xi}{x}. \tag{23}$$

Para esto representemos la variable aleatoria no negativa ξ en forma de la suma de dos variables aleatorias no negativas

$$\xi = \xi\cdot\chi_{\{\xi > x\}} + \xi\chi_{\{\xi \leqslant x\}}.$$

Según las propiedades de aditividad y de monotonía tenemos

$$\mathbf{M}\xi = \mathbf{M}\xi\chi_{\{\xi > x\}} + \mathbf{M}\xi\chi_{\{\xi \leqslant x\}} \geqslant \mathbf{M}\xi\chi_{\{\xi > x\}}.$$

Como $\xi\chi_{\{\xi > x\}} \geqslant x\chi_{\{\xi > x\}}$, entonces, aplicando una vez más la propiedad de monotonía, obtenemos

$$\mathbf{M}\xi \geqslant x\mathbf{M}\chi_{\{\xi > x\}} = x\cdot\mathbf{P}\{\xi > x\}$$

que es lo que demuestra (23).

Con ayuda de la desigualdad de Chébishev se puede ampliar la propiedad 4 del punto 7.1. Es justa la siguiente afirmación.

TEOREMA 1. *Si* $\mathbf{P}\{\xi \geqslant 0\} = 1$ *y* $\mathbf{M}\xi = 0$, *entonces* $\mathbf{P}\{\xi = 0\} = 1$.

DEMOSTRACIÓN. Representemos el evento $\{\xi > 0\}$ en forma de suma de sucesos incompatibles dos a dos

$$\{\xi > 0\} = \{\xi > 1\} + \sum_{n=1}^{\infty}\left\{\frac{1}{n+1} < \xi \leqslant \frac{1}{n}\right\}.$$

Según el axioma A4

$$\mathbf{P}\{\xi > 0\} = \mathbf{P}\{\xi > 1\} + \sum_{n=1}^{\infty} \mathbf{P}\left\{\frac{1}{n+1} < \xi \leqslant \frac{1}{n}\right\}.$$

Según la desigualdad de Chébishev para $\mathbf{M}\xi = 0$ tenemos

$$0 \leqslant \mathbf{P}\{\xi > 1\} \leqslant \mathbf{M}\xi = 0$$

y

$$0 \leqslant \mathbf{P}\left\{\frac{1}{n+1} < \xi \leqslant \frac{1}{n}\right\} \leqslant \mathbf{P}\left\{\frac{1}{n+1} < \xi\right\} \leqslant (n+1)\cdot\mathbf{M}\xi = 0,$$

de donde resulta que $\mathbf{P}\{\xi > 1\} = \mathbf{P}\left\{\frac{1}{n+1} < \xi \leqslant \frac{1}{n}\right\} = 0$ y $\mathbf{P}\{\xi > 0\} = 0$. Por eso $1 = \mathbf{P}\{\xi = 0\} + \mathbf{P}\{\xi > 0\} = \mathbf{P}\{\xi = 0\}$ que es lo que se necesitaba demostrar.

Problemas

1. Hallar la esperanza matemática de ξ, o sea, del número de puntos que salen en un dado.

2. Se toma al azar una ficha de dominó. Hallar la esperanza matemática de la suma de los puntos marcados en ambos cuadrados ξ y η de la misma. Hallar $\mathbf{M}(\xi + \eta)$.

3. El tiempo de trabajo τ de una bombilla eléctrica hasta que se queme posee una distribución exponencial con densidad $p(x) = \lambda e^{-\lambda x}$ cuando $x \geqslant 0$ y $p(x) = 0$ cuando $x < 0$. Hallar $\mathbf{M}\tau$.

4. Hallar el tiempo medio de trabajo de la bombilla eléctrica si, en adición de las condiciones del problema 3, se supone que con la probabilidad p la bombilla puede quemarse al conectarla (la bombilla se conecta una vez).

Indicación. Representar el tiempo de ardimiento de la bombilla en forma del producto $\tau\cdot\xi$ de dos variables aleatorias independientes, donde τ es lo mismo que en el problema 3; $\xi = 0$ con probabilidad p y $\xi = 1$ con probabilidad $1 - p$.

5. El punto de impacto en un blanco circular de radio R tiene la distribución uniforme. Hallar la esperanza matemática de la distancia ρ entre el punto de impacto y el centro.

6. En un grupo hay 25 estudiantes. Suponiendo que sus cumpleaños son independientes y están distribuidos uniformemente en 12 meses del año, hallar la esperanza matemática del número de los meses en que no cae ningún cumpleaños.

7. Una pieza tiene la forma de un cilindro de radio r y de altura h. Al fabricar el cilindro con estas dimensiones se permite una tolerancia de modo que el radio sea igual a $r + \xi$ y la altura $h + \eta$, con ello ξ y η son independientes y están distribuidas uniformemente en los intervalos respectivos $[-\Delta, \Delta]$ y $[-\delta, \delta]$. Hallar la esperanza matemática del volumen del cilindro.

§ 8. Varianza. Momentos

8.1. Definición de la varianza. Se llama *varianza* de una variable aleatoria $\xi = \xi(\omega)$, $\omega \in \Omega$ al número

$$\mathbf{D}\xi = \mathbf{M}(\xi - \mathbf{M}\xi)^2, \tag{1}$$

o sea, la varianza es igual a la esperanza matemática del cuadrado de la desviación de una variable aleatoria, referida a la esperanza matemática de ésta.

En la práctica como medida de dispersión de los valores de una variable aleatoria se usa, a veces, la magnitud $\sqrt{\mathbf{D}\xi}$ denominada *desviación estándar* (típica). Nótese que la varianza $\mathbf{D}\xi$ se define por la fórmula (1) si $\mathbf{M}(\xi - \mathbf{M}\xi)^2$ existe; en el caso contrario se dice que la varianza no existe.

De las propiedades de linealidad de la esperanza matemática se deduce que

$$\mathbf{M}(\xi - \mathbf{M}\xi)^2 = \mathbf{M}[\xi^2 - 2\xi\cdot\mathbf{M}\xi + (\mathbf{M}\xi)^2] =$$
$$= \mathbf{M}\xi^2 - 2\mathbf{M}\xi\cdot\mathbf{M}\xi + (\mathbf{M}\xi)^2 = \mathbf{M}\xi^2 - (\mathbf{M}\xi)^2.$$

De aquí y de (1) se obtiene una fórmula más para calcular la varianza

$$\mathbf{D}\xi = \mathbf{M}\xi^2 - (\mathbf{M}\xi)^2. \tag{2}$$

Si una variable aleatoria ξ es discreta con ley de distribución de las probabilidades $\mathbf{P}\{\xi = x_k\}, k = 1, 2, \dots, \sum_{k=1}^{\infty} \mathbf{P}\{\xi = x_k\} = 1$, entonces su varianza se calcula por la fórmula

$$\mathbf{D}\xi = \sum_{k=1}^{\infty} (x_k - \mathbf{M}\xi)^2\mathbf{P}\{\xi = x_k\}; \tag{3}$$

para la varianza de una variable aleatoria continua ξ con densidad de distribución $p_\xi(x)$ tenemos la fórmula siguiente:

$$\mathbf{D}\xi = \int_{-\infty}^{+\infty} (x - \mathbf{M}\xi)^2 p_\xi(x)dx. \tag{4}$$

Las fórmulas (3) y (4) se deducen de las fórmulas (5) y (12) del § 7. Demos una interpretación mecánica de la esperanza matemática y de la varianza. Interpretaremos la ley de distribución de las probabilidades $p_k = \mathbf{P}\{\xi = x_k\}, k = 1, 2, \dots, n, \sum_{k=1}^{n} p_k = 1$, de una variable aleatoria ξ como ley de distribución de una masa unitaria sobre una recta: en los puntos x_k están concentradas las masas p_k, $k = 1, 2, \dots, n$. Entonces $\mathbf{M}\xi = \sum_{k=1}^{n} x_k \mathbf{P}\{\xi = x_k\} = \sum_{k=1}^{n} x_k p_k$ es el centro de gravedad y $\mathbf{D}\xi = \sum_{k=1}^{n} (x_k - \mathbf{M}\xi)^2 \mathbf{P}\{\xi = x_k\} = \sum_{k=1}^{n} (x_k - \mathbf{M}\xi)^2 p_k$ es el momento de inercia.

EJEMPLO 1. *Distribución de Poisson.* Para una variable aleatoria ξ distribuida por la ley de Poisson con el parámetro a

$$\mathbf{P}\{\xi = k\} = e^{-a}\frac{a^k}{k!}, \quad k = 0, 1, 2, \dots, a > 0,$$

se sabe (véase el ejemplo 6 del § 7) que $\mathbf{M}\xi = a$. Calculemos ahora $\mathbf{M}\xi^2$. Utilizando la igualdad $\xi = \xi(\xi - 1) + \xi$ y la propiedad de linealidad de la esperanza matemática, obtenemos

$$\mathbf{M}\xi^2 = \mathbf{M}[\xi(\xi - 1) + \xi] = \mathbf{M}[\xi(\xi - 1)] + \mathbf{M}\xi = a^2 + a,$$

ya que $\mathbf{M}\xi = a$ y

$$\mathbf{M}\xi(\xi - 1) = \sum_{k=0}^{\infty} k(k-1)e^{-a}\frac{a^k}{k!} = e^{-a}a^2 \sum_{k=2}^{\infty} \frac{a^{k-2}}{(k-2)!} = a^2$$

(aquí hemos hecho uso de la igualdad

$$\sum_{l=2}^{\infty} \frac{a^{l-2}}{(l-2)!} = \sum_{l=0} \frac{a^l}{l!} = e^a).$$

Sustituyendo $\mathbf{M}\xi^2 = a^2 + a$ y $\mathbf{M}\xi = a$ en la fórmula (2), obtenemos $\mathbf{D}\xi = a^2 + a - a^2 = a$. De este modo, para una variable aleatoria distribuida por la ley de Poisson con el parámetro a, tenemos

$$\mathbf{M}\xi = \mathbf{D}\xi = a.$$

EJEMPLO 2. *Distribución uniforme.* La densidad de distribución de una variable aleatoria ξ, distribuida uniformemente en el segmento $[a, b]$, tiene la forma

$$p_\xi(x) = \begin{cases} \dfrac{1}{b - a} & \text{si} \quad x \in [a, b], \\ 0 & \text{si} \quad x \notin [a, b]. \end{cases}$$

En este caso (véase el ejemplo 7 del § 7) $\mathbf{M}\xi = (a + b)/2$. Por la fórmula (19) del § 7, suponiendo $g(x) = x^2$, encontramos

$$\mathbf{M}\xi^2 = \int_a^b \frac{x^2dx}{b-a} = \frac{x^3}{3(b-a)}\Bigg|_a^b = \frac{b^3 - a^3}{3(b-a)} = \frac{b^2 + ab + a^2}{3}.$$

Sustituyendo los valores de $\mathbf{M}\xi$ y $\mathbf{M}\xi^2$ en (2), obtenemos

$$\mathbf{D}\xi = \frac{a^2 + ab + b^2}{3} - \frac{(a+b)^2}{4} = \frac{(b-a)^2}{12}.$$

EJEMPLO 3. *Distribución normal.* La densidad de distribución de una variable aleatoria ξ, distribuida normalmente con los parámetros (a, σ), tiene la forma

$$p_\xi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}, \quad -\infty < a < +\infty, \quad \sigma > 0.$$

En este caso (véase el ejemplo 9 del § 7) $M\xi = a$ y $\mathbf{D}\xi = \mathbf{M}(\xi - \mathbf{M}\xi)^2 =$

$$= \int_{-\infty}^{+\infty} (x-a)^2 p_\xi(x)dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} (x-a)^2 e^{-\frac{(x-a)^2}{2\sigma^2}} dx.$$

De aquí, sustituyendo la variable de integración $(x - a)/\sigma = y$, $x = \sigma y + a$, obtenemos

$$\mathbf{D}\xi = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} y^2 e^{-\frac{y^2}{2}} dy = \frac{\sigma^2}{\sqrt{2\pi}} \left(- \int_{-\infty}^{+\infty} y de^{-\frac{y^2}{2}} \right) =$$

$$= \frac{\sigma^2}{\sqrt{2\pi}} \left(-ye^{-\frac{y^2}{2}} \Bigg|_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} e^{-\frac{y^2}{2}} dy \right) = \sigma^2 \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{y^2}{2}} dy = \sigma^2.$$

De suerte que la esperanza matemática y la varianza de una variable aleatoria ξ distribuida normalmente con los parámetros (a, σ^2) están relacionadas con estos parámetros del modo siguiente:

$$\mathbf{M}\xi = a, \quad \mathbf{D}\xi = \sigma^2.$$

8.2. Propiedades de la varianza. Señalemos las propiedades principales de la varianza.

1. *La varianza de toda variable aleatoria* ξ *es no negativa, además* $\mathbf{D}\xi = 0$ *si y sólo si* ξ *es constante.*

La propiedad de no negatividad se deduce de la desigualdad $(\xi - \mathbf{M}\xi)^2 \geqslant 0$ y de la propiedad de monotonía de la esperanza matemática: $\mathbf{D}\xi = \mathbf{M}(\xi - \mathbf{M}\xi)^2 \geqslant 0$. Si $\xi = c$, entonces $\mathbf{D}c = \mathbf{M}(c - \mathbf{M}c)^2 = 0$. En virtud del teorema 1 del § 7, de $\mathbf{D}\xi = \mathbf{M}(\xi - \mathbf{M}\xi)^2 = 0$ resulta que $(\xi - \mathbf{M}\xi)^2 = 0$ y, por consiguiente, también $\xi = \mathbf{M}\xi$ con probabilidad 1.

2. *Si* a *es constante, entonces*

$$\mathbf{D}(a\xi) = a^2\mathbf{D}\xi. \tag{5}$$

En efecto, $\mathbf{D}(a\xi) = \mathbf{M}(a\xi - \mathbf{M}(a\xi))^2 = \mathbf{M}[a(\xi - \mathbf{M}\xi)]^2 = a^2\mathbf{M}(\xi - \mathbf{M}\xi)^2 = a^2\mathbf{D}\xi$.

3. *Si las variables aleatorias* ξ y η *son independientes, entonces*

$$\mathbf{D}(\xi + \eta) = \mathbf{D}\xi + \mathbf{D}\eta. \tag{6}$$

DEMOSTRACIÓN. Utilizando la definición de la varianza (1) y la propiedad de linealidad de la esperanza matemática, obtenemos

$$\mathbf{D}(\xi + \eta) = \mathbf{M}[(\xi + \eta) - \mathbf{M}(\xi + \eta)]^2 = \mathbf{M}[(\xi - \mathbf{M}\xi) + (\eta - \mathbf{M}\eta)]^2 = \mathbf{M}(\xi - \mathbf{M}\xi)^2 + 2\mathbf{M}(\xi - \mathbf{M}\xi)(\eta - \mathbf{M}\eta) + \mathbf{M}(\eta - \mathbf{M}\eta)^2.$$

De aquí se deduce la fórmula (6), ya que según la propiedad de multiplicatividad de la esperanza matemática

$$\mathbf{M}(\xi - \mathbf{M}\xi)(\eta - \mathbf{M}\eta) = \mathbf{M}(\xi - \mathbf{M}\xi)\mathbf{M}(\eta - \mathbf{M}\eta) = (\mathbf{M}\xi - \mathbf{M}\xi)(\mathbf{M}\eta - \mathbf{M}\eta) = 0.$$

Por inducción, la fórmula (6) se extiende a la suma de n variables aleatorias independientes. Si $\xi_1, \xi_2, \ldots, \xi_n$ son independientes, entonces

$$\mathbf{D}(\xi_1 + \xi_2 + \ldots + \xi_n) = \mathbf{D}\xi_1 + \mathbf{D}\xi_2 + \ldots + \mathbf{D}\xi_n. \tag{7}$$

Calculemos las varianzas de algunas variables aleatorias, utilizando las propiedades demostradas de este estadístico.

EJEMPLO 4. *Distribución binomial.* El número de casos favorables μ_n en n pruebas de Bernoulli tiene la distribución binomial (véase (7) del § 3).

Utilizamos la representación de μ_n en forma de la suma $\mu_n = \sum_{k=1}^{n} \xi_k$, donde ξ_k, $k = 1, 2, \dots, n$, son los indicadores siguientes: $\xi_k = 1$ si en la k-ésima prueba ha sido favorable y $\xi_k = 0$ si se ha sido desfavorable. De este modo, $\mathbf{P}\{\xi_k = 1\} = p$, $\mathbf{P}\{\xi_k = 0\} = q$, $p + q = 1$, $k = 1, 2, \dots, n$. Los indicadores ξ_k son independientes en virtud de la independencia de las pruebas. Para variables aleatorias independientes, según la fórmula (7), hallamos

$$\mathbf{D}\mu_n = \mathbf{D}\left(\sum_{k=1}^{n} \xi_k\right) = \sum_{k=1}^{n} \mathbf{D}\xi_k.$$

De aquí, puesto que $\mathbf{M}\xi_k = 1\cdot p + 0\cdot q = p$, $\mathbf{M}\xi_k^2 = p$, $\mathbf{D}\xi_k = \mathbf{M}\xi_k^2 - (\mathbf{M}\xi_k)^2 = p - p^2 = p(1 - p) = pq$, obtenemos

$$\mathbf{D}\mu_n = npq.$$

EJEMPLO 5. *Distribución hipergeométrica*. Supongamos que de una urna que contiene M bolas blancas y $N - M$ bolas negras se han sacado n bolas según el esquema de selección aleatoria sin reposición (véase el ejemplo 4 del § 1). La variable aleatoria η_n, igual al número de bolas blancas en la muestra, tiene la distribución hipergeométrica (véase (5) del § 1). En el § 1 la fórmula (5) ha sido deducida para el modelo en el cual como sucesos elementales se consideran colecciones no ordenadas de las bolas extraídas. Aquí, al calcular $\mathbf{M}\eta_n$ y $\mathbf{D}\eta_n$ es cómodo examinar un espacio probabilístico más detallado, con colecciones ordenadas, introducido en el ejemplo 5 del § 1. Con tal selección del modelo probabilístico la fórmula (5) del § 1 se conserva. Introduzcamos los indicadores ξ_k, $k = 1, \dots, n$, haciendo $\xi_k = 1$ si la k-ésima bola es blanca y $\xi_k = 0$ en el caso contrario. Es evidente que $\eta_n = \xi_1 + \xi_2 + \dots + \xi_n$. Nótese que los indicadores ξ_k en la representación μ_n (ejemplo 4) son independientes. Aquí en la representación η_n los indicadores ξ_k no son independientes ni siquiera de par en par y, por lo tanto, para calcular $\mathbf{D}\eta_n$ no se puede utilizar la fórmula (7).

Valgámonos de la fórmula (2). Hallemos primeramente $\mathbf{M}\xi_k$ y $\mathbf{M}\xi_k\xi_l$. Según la fórmula (3) del § 7

$$\mathbf{M}\xi_k = 1\cdot\mathbf{P}\{\xi_k = 1\} + 0\cdot\mathbf{P}\{\xi_k = 0\} = \mathbf{P}\{\xi_k = 1\} = \frac{M}{N},$$

$$\mathbf{M}\xi_k^2 = \mathbf{M}\xi_k.$$

Para $k \neq l$, puesto que

$$\mathbf{P}\{\xi_k\xi_l = 1\} = \frac{M(M-1)}{N(N-1)},$$

$$\mathbf{P}\{\xi_k\xi_l = 0\} = 1 - \frac{M(M-1)}{N(N-1)},$$

obtenemos

$$\mathbf{M}\xi_k\xi_l = \frac{M(M-1)}{N(N-1)}.$$

De aquí, utilizando la propiedad de linealidad de la esperanza matemática, encontramos

$$\mathbf{M}\eta_n = \mathbf{M}\left(\sum_{k=1}^{n} \xi_k\right) = \sum_{k=1}^{n} \mathbf{M}\xi_k = n\frac{M}{N},$$

$$\mathbf{M}\eta_n^2 = \mathbf{M}\left(\sum_{k=1}^{n} \xi_k\right)^2 = \mathbf{M}\left[\sum_{k=1}^{n} \xi_k^2 + \sum_{k \neq l} \xi_k\xi_l\right] =$$

$$= \sum_{k=1}^{n} \mathbf{M}\xi_k^2 + \sum_{k \neq l} \mathbf{M}\xi_k\xi_l = n\frac{M}{N} + n(n-1)\frac{M(M-1)}{N(N-1)}$$

y, por consiguiente,

$$\mathbf{D}\eta_n = \mathbf{M}\eta_n^2 - (\mathbf{M}\eta_n)^2 = n\frac{M}{N} + n(n-1)\frac{M(M-1)}{N(N-1)} - n^2\frac{M^2}{N^2}.$$

Después de algunas transformaciones elementales, llegamos a la fórmula final

$$\mathbf{D}\eta_n = n\frac{M}{N}\left(1 - \frac{M}{N}\right)\frac{N-n}{N-1}.$$

El número de bolas blancas (designémoslo con μ_n) en la muestra obtenida según el esquema de selección aleatoria con reposición, posee distribución binomial con el parámetro $p = \dfrac{M}{N}$. Nótese que

$$\mathbf{M}\mu_n = \mathbf{M}\eta_n = \frac{M}{N}; \quad \mathbf{D}\eta_n = \frac{N-n}{N-1}\mathbf{D}\mu_n.$$

8.3. Momentos de orden superior. Además de las características numéricas anteriormente examinadas de las variables aleatorias (la esperanza matemática y la varianza) se utilizan con frecuencia también otras características llamadas momentos. Se llama *momento de orden k* o *k-ésimo momento de una variable aleatoria* ξ al número $m_k = \mathbf{M}\xi^k$, $k = 1, 2, \dots$ El número $\mathbf{M}|\xi|^k$ se denomina *momento absoluto de orden k*, $k = 1, 2, \dots$; los números $\mathbf{M}(\xi - \mathbf{M}\xi)^k$ y $\mathbf{M}|\xi - \mathbf{M}\xi|^k$ se denominan, respectivamente, *momento central* y *momento central absoluto de orden k*, $k = 1, 2, \dots$ (o *k*-ésimos momentos respectivos).

Hagamos notar que de la existencia del momento $\mathbf{M}\xi^m$ se deduce la existencia de los momentos de órdenes más bajos $\mathbf{M}\xi^k$, $k = 1, 2, \dots, m - 1$. Esta afirmación resulta de las desigualdades

$$|\xi(\omega)|^k \leqslant |\xi(\omega)|^m + 1, \quad \omega \in \Omega, \ k = 1, 2, \dots, m - 1.$$

Observemos que el momento de primer orden $m_1 = \mathbf{M}\xi$ es la esperanza matemática y el momento central de segundo orden $\mathbf{M}(\xi - \mathbf{M}\xi)^2$ es la varianza.

En las aplicaciones se usan con frecuencia dos características más, o sea, *la asimetría S y el exceso E*. Ellas se definen del modo siguiente:

$$S = \frac{1}{2}\frac{\mathbf{M}(\xi - \mathbf{M}\xi)^3}{(\mathbf{D}\xi)^{3/2}}, \quad E = \frac{1}{8}\left[\frac{\mathbf{M}(\xi - \mathbf{M}\xi)^4}{(\mathbf{D}\xi)^2} - 3\right]. \tag{8}$$

La asimetría S y el exceso E de una variable distribuida normalmente son iguales a cero. El hecho de que la asimetría y el exceso sean distintos de cero muestra que la distribución de la variable aleatoria no es normal.

EJEMPLO 6. Hallar los momentos de todos los órdenes para una variable aleatoria ξ distribuida normalmente con parámetros $(0, \sigma^2)$. En este caso la densidad de distribución de las probabilidades $p_\xi(x)$ tiene la forma

$$p_\xi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}, \quad -\infty < x < +\infty.$$

Todos los momentos de orden impar, como integrales de funcionales impares, son iguales a cero:

$$m_{2k+1} = \mathbf{M}\xi^{2k+1} = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{+\infty} x^{2k+1} e^{-\frac{x^2}{2\sigma^2}}\, dx = 0,$$

$$k = 0, 1, 2, \dots$$

Calculemos ahora los momentos de orden par. Integrando por partes, obtenemos

$$m_{2k} = M\xi^{2k} = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^{2k} e^{-\frac{x^2}{2\sigma^2}} dx =$$

$$= -\frac{\sigma}{\sqrt{2\pi}} x^{2k-1} e^{-\frac{x^2}{2\sigma^2}} \Big|_{-\infty}^{+\infty} + (2k-1)\frac{\sigma^2}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^{2k-2} e^{-\frac{x^2}{2\sigma^2}} dx =$$

$$= (2k-1)\sigma^2 \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^{2k-2} e^{-\frac{x^2}{2\sigma^2}} dx = (2k-1)\sigma^2 m_{2k-2}.$$

Así, pues, hemos determinado la siguiente fórmula recurrente:

$$m_{2k} = (2k-1)\sigma^2 m_{2k-2}.$$

De aquí, puesto que

$$m_0 = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2\sigma^2}} dx = 1,$$

hallamos que $m_2 = \sigma^2 m_0 = \sigma^2$, $m_4 = 3\sigma^2 m_2$, etc.

Por inducción obtenemos

$$m_{2k} = (2k-1)(2k-3)\ldots 3\cdot 1\cdot \sigma^{2k} = (2k-1)!!\sigma^{2k}.$$

De suerte que los momentos de una variable aleatoria ξ, distribuida normalmente con parámetros $(0, \sigma^2)$, se definen por las fórmulas

$$\mathbf{M}\xi^{2k-1} = 0, \quad \mathbf{M}\xi^{2k} = (2k-1)!!\sigma^{2k}, \quad k = 1, 2, \ldots$$

Problemas

1. La longitud del diámetro de un círculo está distribuida uniformemente en el segmento $[0, 1]$. Hallar la esperanza matemática y la varianza del área del círculo.

2. La densidad de distribución de ξ es igual a

$$p_\xi(x) = \begin{cases} \dfrac{c}{x}, & x \in [1, 2], \\ 0, & x \notin [1, 2]. \end{cases}$$

Hallar c, $\mathbf{M}\xi$, $\mathbf{D}\xi$.

3. Las coordenadas de dos puntos aleatorios marcados sobre una recta son independientes y están distribuidas uniformemente en el segmento $[0, 1]$. Hallar la esperanza matemática y la varianza de la distancia entre ellas.

4. De una urna que contiene m bolas blancas y n negras se extraen bolas, según el esquema de selección aleatoria con reposición, hasta que salga una bola blanca. Hallar la esperanza matemática y la varianza del número de las bolas extraídas.

5. Demostrar que para la distribución normal la asimetría y el exceso son iguales a cero.

6. En la k-ésima prueba del esquema de Bernoulli se produce un caso favorable doble si el caso favorable ha sucedido en la k-ésima prueba y en la prueba $k + 1$. Designemos con η_n el número de casos favorables dobles en n pruebas del esquema de Bernoulli. Hallar $\lim\limits_{n\to\infty} \frac{\mathbf{D}\eta_n}{n}$ si la probabilidad del caso favorable en cada prueba es igual a $p(0 < p < 1)$.

§ 9. Covarianza. Coeficiente de correlación

Se llama *covarianza* de las variables aleatorias $\xi_1 = \xi_1(\omega)$ y $\xi_2 = \xi_2(\omega)$ al número

$$\operatorname{cov}(\xi_1, \xi_2) = \mathbf{M}[(\xi_1 - \mathbf{M}\xi_1)(\xi_2 - \mathbf{M}\xi_2)]. \tag{1}$$

De la definición de covarianza y de las propiedades de linealidad de la esperanza matemática se deduce que

$$\operatorname{cov}(\xi_1, \xi_2) = \mathbf{M}[(\xi_1 - \mathbf{M}\xi_1)(\xi_2 - \mathbf{M}\xi_2)] =$$
$$= \mathbf{M}[\xi_1\xi_2 - \xi_1\mathbf{M}\xi_2 - \xi_2\mathbf{M}\xi_1 + \mathbf{M}\xi_1\cdot\mathbf{M}\xi_2] = \mathbf{M}\xi_1\xi_2 - \mathbf{M}\xi_1\cdot\mathbf{M}\xi_2.$$

Así, pues, la covarianza se puede calcular asimismo por la fórmula siguiente:

$$\operatorname{cov}(\xi_1, \xi_2) = \mathbf{M}\xi_1\xi_2 - \mathbf{M}\xi_1\cdot\mathbf{M}\xi_2. \tag{2}$$

Si ξ_1 y ξ_2 son independientes, entonces

$$\operatorname{cov}(\xi_1, \xi_2) = 0. \tag{3}$$

Esta igualdad se deduce de la definición (1) y de la propiedad de multiplicatividad de la esperanza matemática del producto de las variables independientes $\xi_1 - \mathbf{M}\xi_1$ y $\xi_2 - \mathbf{M}\xi_2$. En virtud de las propiedades de la esperanza matemática se comprueba también fácilmente que

$$\begin{gathered}\operatorname{cov}(\xi, \xi) = \mathbf{D}\xi, \quad \operatorname{cov}(\xi_1, \xi_2) = \operatorname{cov}(\xi_2, \xi_1), \\ \operatorname{cov}(c_1\xi_1 + c_2\xi_2, \xi_3) = c_1\operatorname{cov}(\xi_1, \xi_3) + c_2\operatorname{cov}(\xi_2, \xi_3),\end{gathered} \tag{4}$$

donde c_1 y c_2 son constantes.

TEOREMA 1. *Si para las variables aleatorias $\xi_1, \xi_2, \dots, \xi_n$ existen $\sigma_{ij} = \mathrm{cov}(\xi_i, \xi_j)$, $i, j = 1, 2, \dots, n$, entonces para constantes $c_1, c_2 \dots \dots, c_n$ cualesquiera existe*

$$\mathbf{D}\left(\sum_{i=1}^{n} c_i\xi_i\right) \quad \text{y} \quad \mathbf{D}\left(\sum_{i=1}^{n} c_i\xi_i\right) = \sum_{i,j=1}^{n} c_ic_j\sigma_{ij}. \tag{5}$$

En efecto, utilizando la propiedad de linealidad de la esperanza matemática, obtenemos

$$\mathbf{D}\left(\sum_{k=1}^{n} c_k\xi_k\right) = \mathbf{M}\left[\sum_{k=1}^{n} c_k\xi_k - \mathbf{M}\left(\sum_{k=1}^{n} c_k\xi_k\right)\right]^2 =$$

$$= \mathbf{M}\left[\sum_{k=1}^{n} c_k(\xi_k - \mathbf{M}\xi_k)\right]^2 =$$

$$= \mathbf{M}\left[\sum_{i,j=1}^{n} c_ic_j(\xi_i - \mathbf{M}\xi_i)(\xi_j - \mathbf{M}\xi_j)\right] =$$

$$= \sum_{i,j=1}^{n} c_ic_j\mathrm{cov}(\xi_i, \xi_j) = \sum_{i,j=1}^{n} c_ic_j\sigma_{ij}.$$

OBSERVACIÓN 1. Se las variables aleatorias ξ_i, $i = 1, 2, \dots, n$, son independientes de par en par entonces de la igualdad (3) resulta que $\sigma_{ij} = \mathrm{cov}(\xi_i, \xi_j) = 0$ cuando $i \neq j$. De aquí y de la igualdad (5) obtenemos la fórmula

$$\mathbf{D}\left(\sum_{k=1}^{n} c_k\xi_k\right) = \sum_{k=1}^{n} c_k^2\mathbf{D}\xi_k$$

que no es más que la generalización de la fórmula (7) del 8.

OBSERVACIÓN 2. Utilizando la propiedad de no negatividad de la varianza de la fórmula (5), poniendo en ella $n = 2$, $c_1 = X$ y $c_2 = 1$, obtenemos que

$$f(X) = \mathbf{D}(X\xi_1 + \xi_2) = X^2\sigma_{11} + 2X\sigma_{12} + \sigma_{22} \geqslant 0$$

para todo valor real de X. Por lo tanto, $\sigma_{12}^2 - \sigma_{11}\cdot\sigma_{22} \leqslant 0$ ó $|\sigma_{12}| < \sqrt{\sigma_{11}\sigma_{12}}$. De aquí, sustituyendo σ_{12}, σ_{11}, σ_{22} por $\mathrm{cov}(\xi_1, \xi_2)$, $\mathbf{D}\xi_1$, $\mathbf{D}\xi_2$, obtenemos la desigualdad

$$|\mathrm{cov}(\xi_1, \xi_2)| \leqslant \sqrt{\mathbf{D}\xi_1\cdot\mathbf{D}\xi_2}. \tag{6}$$

TEOREMA 2. *Si para las variables aleatorias $\xi_1, \xi_2, \dots, \xi_n$ existen $\sigma_{ij} = \operatorname{cov}(\xi_i, \xi_j)$, $i, j = 1, 2, \dots, n$, entonces, cualesquiera que sean las constantes c_{ij}, $i, j = 1, 2, \dots, n$, para las variables aleatorias $\eta_k = c_{k1}\xi_1 + c_{k2}\xi_2 + \dots + c_{kn}\xi_n$, $k = 1, 2, \dots, n$, existen las covarianzas $h_{ij} = \operatorname{cov}(\eta_i, \eta_j)$ y las matrices de covarianza $H = \|h_{ij}\|$ y $S = \|\sigma_{ij}\|$ están relacionadas por la igualdad*

$$H = CSC',$$

donde $C = \|c_{ij}\|$ y C' es la matriz transpuesta a C.

DEMOSTRACIÓN. Utilizando la propiedad (4), obtenemos

$$h_{ij} = \operatorname{cov}(\eta_i, \eta_j) = \operatorname{cov}\left(\sum_{k=1}^{n} c_{ik}\xi_k, \sum_{r=1}^{n} c_{jr}\xi_r\right) =$$

$$= \sum_{k,r=1}^{n} c_{ik}c_{jr}\sigma_{kr} = \sum_{k,r=1}^{n} c_{ik}\sigma_{kr}c'_{rj},$$

donde $c'_{rj} = c_{jr}$ es un elemento de la matriz C'. Ahora bien, el elemento h_{ij} coincide con (i, j)-ésimo elemento de la matriz CSC'. El teorema queda demostrado.

La igualdad (3) es justa para variables aleatorias independientes. Por consiguiente, si $\operatorname{cov}(\xi_1, \xi_2) \neq 0$, entonces las variables aleatorias ξ_1 y ξ_2 son dependientes. En calidad de característica cuantitativa del grado de dependencia de las variables aleatorias ξ_1 y ξ_2 se introduce *el coeficiente de correlación* $\rho(\xi_1, \xi_2)$ definido por la igualdad siguiente:

$$\rho(\xi_1, \xi_2) = \frac{\operatorname{cov}(\xi_1, \xi_2)}{\sqrt{\mathbf{D}\xi_1 \mathbf{D}\xi_2}}.$$

Las propiedades principales del coeficiente de correlación son las siguientes.

1. *El coeficiente de correlación satisface la desigualdad*

$$|\rho(\xi_1, \xi_2)| \leqslant 1; \tag{8}$$

con ello, si ξ_2 se expresa linealmente por ξ_1, $\xi_2 = a\xi_1 + b$, donde a y b son constantes, entonces

$$|\rho(\xi, \xi_2)| = 1. \tag{9}$$

DEMOSTRACIÓN. Dividiendo ambos miembros de la desigualdad (6) por $\sqrt{\mathbf{D}\xi_1 \cdot \mathbf{D}\xi_2}$, obtenemos la desigualdad (8). Demostremos la desigualdad (9). Sea $\xi_2 = a\xi_1 + b$. Entonces $\mathbf{D}\xi_2 = \mathbf{D}(a\xi_1) = a^2\mathbf{D}\xi_1$ y $\operatorname{cov}(\xi_1,$

$\xi_2) = \mathbf{M}[(\xi_1 - \mathbf{M}\xi_1)(\xi_2 - \mathbf{M}\xi_2)] = \mathbf{M}[(\xi_1 - \mathbf{M}\xi_1)(a\xi_1 + b - \mathbf{M}(a\xi_1 + b)] = \mathbf{M}[a(\xi_1 - \mathbf{M}\xi_1)(\xi_1 - \mathbf{M}\xi_1)] = a\mathbf{D}\xi_1$. De aquí y de la fórmula (7) obtenemos

$$\rho(\xi_1, \xi_2) = \rho(\xi_1, a\xi_1 + b) = \frac{a\mathbf{D}\xi_1}{\sqrt{\mathbf{D}\xi_1 \cdot a^2\mathbf{D}\xi_1}} = \frac{a}{|a|}.$$

Por lo tanto, $|\rho(\xi_1, \xi_2)| = 1$.

Tiene lugar la afirmación inversa: si el coeficiente de correlación $|\rho(\xi_1, \xi_2)| = 1$, entre ξ_1 y ξ_2 hay una dependencia lineal.

2. *Si las variables aleatorias ξ_1 y ξ_2 son independientes, entonces $\rho(\xi_1, \xi_2) = 0$.*

Esta afirmación se deduce de la definición del coeficiente de correlación (7) y del hecho de que $\mathrm{cov}(\xi_1, \xi_2) = 0$ para las variables aleatorias independientes ξ_1 y ξ_2.

La afirmación inversa no es justa. De la igualdad $\rho(\xi_1, \xi_2) = 0$ no se deduce la independencia de las variables aleatorias ξ_1 y ξ_2. Examinemos un ejemplo.

EJEMPLO. Supongamos que la variable aleatoria ξ_1 está distribuida normalmente con parámetros (0, 1) y sea $\xi_2 = \xi_1^2 - 1$. Las variables aleatorias ξ_1 y ξ_2 son dependientes. Mostremos que $\rho(\xi_1, \xi_2) = 0$. En efecto,

$$\mathbf{M}\xi_1 = 0, \mathbf{M}\xi_1^2 = \mathbf{D}\xi_1 = 1, \mathbf{M}\xi_2 = \mathbf{M}(\xi_1^2 - 1) = 0,$$

$$\mathbf{D}\xi_2 = \mathbf{M}\xi_2^2 = \mathbf{M}(\xi_1^2 - 1)^2 = \mathbf{M}(\xi_1^4 - 2\xi_1^2 + 1) = 3 - 2 + 1 = 2,$$

$$\mathrm{cov}(\xi_1, \xi_2) = \mathbf{M}\xi_1\xi_2 = \mathbf{M}\xi_1(\xi_1^2 - 1) = \mathbf{M}\xi_1^3 - \mathbf{M}\xi = 0.$$

Por lo tanto, $\rho(\xi_1, \xi_2) = (\mathrm{cov}(\xi_1, \xi_2))/\sqrt{\mathbf{D}\xi_1, \mathbf{D}\xi_2} = 0$, aunque ξ_1 y ξ_2 son dependientes.

Las variables aleatorias ξ_1 y ξ_2 no están correlacionadas si $\rho(\xi_1, \xi_2) = 0$.

De lo expuesto arriba se deduce que las variables aleatorias independientes no están correlacionadas, mientras que las variables aleatorias no correlacionadas no son obligatoriamente independientes. Examinemos las variables aleatorias ξ, η con la densidad de distribución

$$p_{\xi,\eta}(x, y) = \frac{1}{\sqrt{2\pi}} \cdot \frac{\varsigma}{\sqrt{1 - R^2}} e^{-(x^2 + y^2 - 2Rxy)/(2(1 - R^2))}, \qquad (10)$$

donde $|R| < 1$. En este caso se dice que las variables aleatorias ξ, η tienen una distribución normal bidimensional.

Determinemos las densidades unidimensionales y el coeficiente de correlación de las variables aleatorias ξ y η.

Haciendo uso de la igualdad

$$x^2 + y^2 - 2Rxy = (x - Ry)^2 + y^2(1 - R^2),$$

obtenemos

$$p_\eta(y) = \int_{-\infty}^{+\infty} \frac{1}{2\pi\sqrt{1 - R^2}} e^{-\frac{x^2 + y^2 - 2Rxy}{2(1 - R^2)}} dx =$$

$$= \frac{1}{2\pi\sqrt{1 - R^2}} e^{-\frac{y^2}{2}} \int_{-\infty}^{+\infty} e^{-\frac{(x - Ry)^2}{2(1 - R^2)}} dx.$$

De aquí, haciendo $\frac{x - Ry}{\sqrt{1 - R^2}} = z$, $dx = \sqrt{1 - R^2}dz$, encontramos

$$p_\eta(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} \cdot \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{z^2}{2}} dz = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}$$

o bien

$$p_\eta(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}, \quad -\infty < y < +\infty.$$

Análogamente

$$p_\xi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad -\infty < x < +\infty.$$

Vamos a calcular ahora el coeficiente de correlación $\rho(\xi, \eta)$ para las variables aleatorias ξ, η con la densidad de distribución normal (10). Como $\mathbf{M}\xi = \mathbf{M}\eta = 0$, $\mathbf{D}\xi = \mathbf{D}\eta = 1$, entonces

$$\operatorname{cov}(\xi, \eta) = \mathbf{M}\xi\eta = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xy p_{\xi\eta}(x, y)\,dx\,dy =$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} y e^{-\frac{y^2}{2}} \left(\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi(1 - R^2)}} x e^{-\frac{(x - Ry)^2}{2(1 - R^2)}} dx \right) dy.$$

La integral interior se puede considerar como esperanza matemática de la variable aleatoria distribuida normalmente con parámetros $a = Ry$ y $\sigma^2 = 1 - R^2$. Por eso es igual a Ry y

$$\mathbf{M}\xi\eta = \frac{R}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} y^2 e^{-\frac{y^2}{2}} dy = R.$$

Sustituyendo los valores hallados de $\mathbf{M}\xi$, $\mathbf{M}\eta$, $\mathbf{D}\xi$, $\mathbf{D}\eta$ y $\mathbf{M}\xi\eta$ en las fórmulas (2) y (7) obtenemos que el coefficiente de correlación de las variables aleatorias ξ, η con densidad normal de distribución compatible (10) es igual a R, o sea, $\rho(\xi, \eta) = R$.

Nótese que para las variables aleatorias ξ y η distribuidas normalmente con densidad (10) la independencia es equivalente a la no correlatividad.

En efecto, si ξ y η son independientes, entonces $\rho(\xi, \eta) = 0$ y, por lo tanto, las variables aleatorias ξ y η no están correlacionadas. Si las variables aleatorias ξ y η tienen una distribución normal compatible con densidad (10) y no están correlacionadas, o sea $\rho(\xi, \eta) = R = 0$, entonces

$$p_{\xi,\eta}(x, y) = \frac{1}{2\pi} e^{-(x^2+y^2)/2} = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} = p_\xi(x) p_\eta(y)$$

que es lo que significa la independencia de las variables aleatorias ξ y η.

Problemas

1. La distribución compatible de las variables aleatorias ξ_1, ξ_2 está definida por las fórmulas

$$\mathbf{P}\{\xi_1 = -1, \xi_2 = -1\} = \mathbf{P}\{\xi_1 = 0, \xi_2 = -1\} = \mathbf{P}\{\xi_1 = 1, \xi_2 = -1\} = \frac{1}{6},$$

$$\mathbf{P}\{\xi_1 = -1, \xi_2 = 1\} = \frac{1}{4}, \mathbf{P}\{\xi_1 = 0, \xi_2 = 1\} = \mathbf{P}\{\xi_1 = 1, \xi_2 = 1\} = \frac{1}{8}.$$

Calcular las esperanzas matemáticas y las varianzas de ξ_1 y ξ_2, así como la covarianza y el coeficiente de correlación.

2. La distribución conjunta de las variables aleatorias ξ_1 y ξ_2 se define por las fórmulas $\mathbf{P}\{\xi_1 = 0, \xi_2 = 1\} = \mathbf{P}\{\xi_1 = 0, \xi_2 = -1\} = \mathbf{P}\{\xi_1 = 1, \xi_2 = 0\} = \mathbf{P}\{\xi_1 = -1, \xi_2 = 0\} = \frac{1}{4}$. a) Hallar $\mathbf{M}\xi_1$, $\mathbf{M}\xi_2$, $\mathbf{D}\xi_1$ y $\mathbf{D}\xi_2$; b) calcular la covarianza y el coeficiente de correlación de ξ_1 y ξ_2. ¿Son independientes las variables aleatorias?

3. La densidad de distribución conjunta de las variables aleatorias ξ_1 y ξ_2 está definida por la fórmula

$$p_{\xi_1, \xi_2}(x, y) = \begin{cases} \dfrac{c}{x^3y^3} & \text{cuando } x \geqslant 1, y \geqslant 1, \\ 0 & \text{en los demás casos.} \end{cases}$$

Hallar la constante c y la covarianza de ξ_1, ξ_2.

4. La densidad de distribución conjunta de las variables aleatorias ξ_1 y ξ_2 está definida por la fórmula

$$p_{\xi_1, \xi_2}(x, y) = \begin{cases} c(1 - xy^3) & \text{cuando } |x| \leqslant 1, |y| \leqslant 1, \\ 0 & \text{en los demás casos.} \end{cases}$$

Hallar la constante c y el coeficiente de correlación de ξ_1, ξ_2.

5. Las variables aleatorias $\xi_1, \xi_2, \xi_3, \xi_4, \xi_5$ son independientes; $\mathbf{D}\xi_i = \sigma^2$. Hallar el coeficiente de correlación a) de las variables $\xi_1 + \xi_2$, $\xi_3 + \xi_4 + \xi_5$; b) de las variables $\xi_1 + \xi_2 + \xi_3$, $\xi_3 + \xi_4 + \xi_5$.

6. Se arrojan dos dados. Sea ξ el número de puntos en el primer dado y η el número máximo entre dos números de puntos salidos.

a) Escribir la distribución compatible de ξ y η.

b) Hallar las esperanzas matemáticas y las varianzas de ξ y η.

c) Calcular la covarianza y el coeficiente de correlación de ξ y η.

§ 10. Ley de los grandes números

10.1. Desigualdad de Chébishev. Las características numéricas de variables aleatorias permiten dar ciertas estimaciones de distribuciones de las probabilidades de las mismas.

TEOREMA 1 (**segunda forma de la desigualdad de Chébishev**). *Si una variable aleatoria tiene la varianza* $\mathbf{D}\xi$ *y* Δ *es un número positivo arbitrario, entonces*

$$\mathbf{P}\{|\xi - \mathbf{M}\xi| \geqslant \Delta\} \leqslant \frac{\mathbf{D}\xi}{\Delta^2}. \tag{1}$$

DEMOSTRACIÓN. En el § 7 hemos demostrado la primera forma de la desigualdad de Chébishev, $\mathbf{P}\{\eta \geqslant \varepsilon\} \leqslant \dfrac{\mathbf{M}\eta}{\varepsilon}$, para variables aleatorias no negativas η. Suponiendo $\eta = |\xi - \mathbf{M}\xi^2|$ y $\varepsilon = \Delta^2$, obtenemos (1).

Si en (1) la magnitud Δ se sustituye por $t\sqrt{\mathbf{D}\xi}$ y se pasa al evento contrario, la desigualdad de Chébishev (1) se puede escribir de la forma siguiente:

$$\mathbf{P}\{|\xi - \mathbf{M}\xi| < t\sqrt{\mathbf{D}\xi}\} \geqslant 1 - \frac{1}{t^2}. \tag{2}$$

La desigualdad de Chébishev permite estimar las probabilidades de que los valores de una magnitud aleatoria se desvíen de su esperanza matemática. Supongamos que se realizan n mediciones independientes de cierta magnitud a. Consideraremos como variables aleatorias los errores de la medición $\delta_1, \delta_2, \dots, \delta_n$. Supongamos que $\mathbf{M}\delta_k = 0$, $k = 1, 2, \dots, n$. Esta condición se puede estimar como ausencia del error sistemático. Supongamos también que $\mathbf{D}\delta_k = b^2$. Como valor de la magnitud desconocida a suele tomarse la media aritmética de los resultados de mediciones. Entonces el error cometido en la determinación del número a será igual a $\eta_n = (\delta_1 + \delta_2 + \dots + \delta_n)/n$ y

$$\mathbf{D}\eta_n = \frac{1}{n^2}(\mathbf{D}\delta_1 + \dots + \mathbf{D}\delta_n) = \frac{b^2}{n}, \quad \mathbf{M}\eta_n = 0.$$

Evaluaremos el número de mediciones n para el cual el error η_n no excede Δ con una probabilidad suficientemente grande. Por ejemplo, $\mathbf{P}\{|\eta_n| < \Delta\} > 0{,}99$ o bien

$$\mathbf{P}\{|\eta_n| \geqslant \Delta\} \leqslant 0{,}01. \tag{3}$$

Según la desigualdad de Chébishev (1) tenemos

$$\mathbf{P}\{|\eta_n| \geqslant \Delta\} \leqslant \frac{\mathbf{D}\eta_n}{\Delta^2} = \frac{b^2}{n\Delta^2}.$$

Por lo tanto, (3) será cumplida si

$$\frac{b^2}{n\Delta^2} \leqslant 0{,}01 \quad \text{o bien} \quad n > 100\frac{b^2}{\Delta^2}.$$

Así, pues, hemos obtenido la estimación del número de mediciones, necesario para obtener la precisión prefijada.

10.2. Ley de los grandes números. Se dice que a las variables aleatorias ξ_k, $k = 1, 2, \dots$ con esperanzas matemáticas $\mathbf{M}\xi_k$, $k = 1, 2, \dots$ les es aplicable la ley de los grandes números si para todo valor de $\varepsilon > 0$

$$\lim_{n\to\infty} \mathbf{P}\left\{\left|\frac{\xi_1 + \xi_2 + \dots + \xi_n}{n} - \right.\right.$$
$$\left.\left. - \frac{\mathbf{M}\xi_1 + \mathbf{M}\xi_2 + \dots + \mathbf{M}\xi_n}{n}\right| < \varepsilon\right\} = 1. \tag{4}$$

TEOREMA 7 (teorema de Márkov). *Si las variables aleatorias ξ_k, $k = 1, 2, \ldots$, tienen varianzas y si para $n \to \infty$*

$$\frac{1}{n^2}\mathbf{D}\left(\sum_{k=1}^{n}\xi_k\right) \to 0, \qquad (5)$$

entonces a las variables indicadas ξ_k, $k = 1, 2, \ldots$, les es aplicable la ley de los grandes números.

DEMOSTRACIÓN. Designemos $\eta_n = (\xi_1 + \ldots + \xi_n)/n$. Haciendo uso de (1) con $\Delta = \varepsilon$, $\xi = \eta_n$, obtenemos

$$\mathbf{P}\{|\eta_n - \mathbf{M}\eta_n| \geqslant \varepsilon\} \leqslant \frac{1}{\varepsilon^2}\mathbf{D}\eta_n.$$

De aquí, puesto que

$$\mathbf{M}\eta_n = \frac{\mathbf{M}\xi_1 + \ldots + \mathbf{M}\xi_n}{n}, \quad \mathbf{D}\eta_n = \frac{1}{n^2}\mathbf{D}\left(\sum_{k=1}^{n}\xi_k\right),$$

para la probabilidad del suceso contrario hallamos la estimación

$$\mathbf{P}\left\{\left|\frac{\xi_1 + \ldots + \xi_n}{n} - \frac{\mathbf{M}\xi_1 + \ldots + \mathbf{M}\xi_n}{n}\right| < \varepsilon\right\} \geqslant$$

$$\geqslant 1 - \frac{1}{n^2}\mathbf{D}\left(\sum_{k=1}^{n}\xi_k\right).$$

De esta desigualdad y de la condición (5) se deduce (4). El teorema queda demostrado.

Señalemos algunos casos particulares de este teorema.

TEOREMA 3 (teorema de Chébishev). *Si las variables aleatorias ξ_k, $k = 1, 2, \ldots$, son independientes de par en par y tienen las varianzas uniformemente acotadas (o sea, existe una constante c tal que $\mathbf{D}\xi_k < c$ para todos los valores de $k = 1, 2, \ldots$), entonces a las variables aleatorias ξ_k, $k = 1, 2, \ldots$, es aplicable la ley de los grandes números.*

En efecto, para demostrar el teorema basta comprobar la condición (5). De las desigualdades $\mathbf{D}\xi_k < c$, $k = 1, 2, \ldots$, y de la independencia de par en par de las variables aleatorias ξ_k, $k = 1, 2, \ldots$, resulta que

$$\mathbf{D}\left(\sum_{k=1}^{n}\xi_k\right) = \sum_{k=1}^{n}\mathbf{D}\xi_k < nc.$$

De aquí obtenemos la condición (5):

$$\frac{1}{n^2}\mathbf{D}\left(\sum_{k=1}^{n}\xi_k\right) < \frac{c}{n} \to 0$$

cuando $n \to \infty$.

TEOREMA 4. *Si las variables aleatorias ξ_k, $k = 1, 2, \ldots$, están igualmente distribuidas, son independientes de par en par y poseen varianzas finitas, entonces a estas variables les es aplicable la ley de los grandes números.*

El teorema 4 se deduce del teorema 3. En efecto, las varianzas $\mathbf{D}\xi_k$, $k = 1, 2, \ldots$, existen y son iguales entre sí. Por lo tanto, están uniformemente acotadas y nos encontramos en las condiciones del teorema 3.

La afirmación del teorema 4 significa que para cualesquiera valores de $\varepsilon > 0$, $\delta > 0$, existe tal N que cuando $n > N$ sea justa la desigualdad

$$\mathbf{P}\left\{\left|\frac{\xi_1 + \xi_2 + \ldots + \xi_n}{n} - a\right| < \varepsilon\right\} \geqslant 1 - \delta, \tag{6}$$

donde $a = \mathbf{M}\xi_k$, $k = 1, 2, \ldots$

TEOREMA 5 (teorema de Bernoulli). *Sea μ_n el número de los casos favorables en n pruebas de Bernoulli y p la probabilidad del caso favorable en cada prueba. Entonces para todo valor de $\varepsilon > 0$*

$$\lim_{n\to\infty} \mathbf{P}\left\{\left|\frac{\mu_n}{n} - p\right| < \varepsilon\right\} = 1. \tag{7}$$

Para demostrar este teorema valgámonos de la representación de μ_n en forma de la suma de n indicadores: $\mu_n = \xi_1 + \xi_2 + \ldots + \xi_n$, donde $\xi_k = 1$, si la k-ésima prueba es favorable y $\xi_k = 0$ si es desfavorable.

Como ξ_k, $k = 1, 2, \ldots, n$, son independientes, están distribuidas igualmente ($\mathbf{P}\{\xi_k = 1\} = p$, $\mathbf{P}\{\xi_k = 0\} = 1 - p = q$), las varianzas de las variables aleatorias ξ_k existen y $\mathbf{M}\xi_k = p$, entonces el teorema 5 se deduce directamente del teorema 4.

OBSERVACIÓN. La demostración del teorema 5 se puede realizar asimismo sin recurrir a este último teorema. En efecto, como $\mathbf{M}\frac{\mu_n}{n} = p$, $\mathbf{D}\xi_k = pq$, $\mathbf{D}\frac{\mu_n}{n} = \frac{pq}{n}$, entonces según la desigualdad de Chébishev (1)

$$\mathbf{P}\left\{\left|\frac{\mu_n}{n} - p\right| < \varepsilon \geqslant\right\} 1 - \frac{1}{\varepsilon^2}\frac{pq}{n}. \tag{8}$$

Esta desigualdad permite estimar la probabilidad de que en el esquema de Bernoulli la frecuencia $\frac{\mu_n}{n}$ se desvíe de la probabilidad p en no más que ε. Si se elige tal n_0 que cuando $n \geqslant n_0$, $\frac{pq}{\varepsilon^2 n} < \delta$, donde δ es un número pequeño, entonces según (8) cuando $n \geqslant n_0$ la frecuencia $\frac{\mu_n}{n}$ se encontrará, con una probabilidad no menor que $1 - \delta$, entre los límites

$$p - \varepsilon \leqslant \frac{\mu_n}{n} \leqslant p + \varepsilon. \tag{9}$$

Partiendo de la propiedad de estabilidad de las frecuencias, indicada en el § 1, se puede enunciar el principio de certeza práctica: los sucesos cuya probabilidad es próxima a la unidad son prácticamente ciertos, o sea, en una prueba aislada por lo general se producen. Según este principio en una serie de n pruebas independientes la desigualdad (9) debe cumplirse si $n \geqslant n_0$.

Problemas

1. Una variable aleatoria ξ posee distribución normal con parámetros (a, σ). Estimar por la desigualdad de Chébishev $\mathbf{P}\{|\xi - a| \geqslant 2\sigma\}$. Comparar con el valor exacto de esta probabilidad.

2. La ley de distribución de una variable aleatoria ξ está definida por las fórmulas

$$\mathbf{P}\{\xi = 0\} = 1 - \frac{\sigma^2}{\Delta^2},\ \mathbf{P}\{\xi = -\Delta\} = \mathbf{P}\{\xi = \Delta\} = \frac{\sigma^2}{2\Delta^2}.$$

Comparar el valor exacto de la probabilidad $\mathbf{P}\{|\xi| \geqslant \Delta\}$ con la estimación obtenida con ayuda de la desigualdad de Chébishev.

3. Se supone realizar 10 mediciones $x_1, x_2, \ldots, x_{10}$ de una magnitud desconocida a. Considerando $x_1, \ldots, x_{10}$ variables aleatorias independientes, distribuidas normalmente, con $\mathbf{M}x_k = a$, $\mathbf{D}x_k = 0{,}01$, hallar Δ si

$$\mathbf{P}\left\{\left|\frac{x_1 + x_2 + \ldots + x_{10}}{10} - a\right| < \Delta\right\} = 0{,}99.$$

Estimar Δ, utilizando la desigualdad de Chébishev. Comparar los resultados obtenidos.

4. ¿Es aplicable la ley de los grandes números a la sucesión de variables aleatorias independientes $\xi_1, \xi_2, \ldots, \xi_k, \ldots$ si

$$\mathbf{P}\{\xi_k = \sqrt{k}\} = \mathbf{P}\{\xi_k = -\sqrt{k}\} = \frac{1}{2\sqrt{k}},\quad \mathbf{P}\{\xi_k = 0\} = 1 - \frac{1}{\sqrt{k}}?$$

§ 11. Teorema central del límite

En el § 4 hemos demostrado el teorema de Moivre — Laplace según el cual el número de casos favorables μ_n en n pruebas de Bernoulli tiene una distribución próxima a la normal. Representemos μ_n en forma de la suma de los indicadores independientes $\mu_n = \xi_1 + \xi_2 + \ldots + \xi_n$, donde $\xi_k = 1$ si la k-ésima prueba ha sido favorable y $\xi_k = 0$ si ha sido desfavorable.

El teorema de Moivre — Laplace se puede enunciar en la forma siguiente.

Si las variables aleatorias $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ *son independientes,* $\mathbf{P}\{\xi_k = 1\} = 1 - \mathbf{P}\{\xi_k = 0\} = p, k = 1, 2, \ldots$, *entonces para* $n \to \infty$, $0 < p < 1$,

$$\mathbf{P}\left\{\frac{\xi_1 + \xi_2 + \ldots + \xi_n - n\mathbf{M}\xi_1}{\sqrt{n\mathbf{D}\xi_1}} < x\right\} \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \cdot e^{-\frac{u^2}{2}} du \qquad (1)$$

es uniforme respecto a $x \in (-\infty, +\infty)$.

La afirmación (1) sigue siendo válida para supuestos suficientemente generales sobre la ley de distribución de las variables aleatorias ξ_k. Determinemos las densidades de distribuciones de las sumas $\eta_n = \xi_1 + \xi_2 + \ldots + \xi_n$, donde $n = 1, 2, 3$, $\xi_k (k = 1, 2, 3)$ son variables aleatorias independientes distribuidas uniformemente en el segmento $[-1, 1]$. De este modo, las densidades $p_{\xi_k}(x) = p(x)$ $(k = 1, 2, 3)$, donde

$$p(x) = \begin{cases} 1/2, & \text{si} \quad x \in [-1, 1], \\ 0, & \text{si} \quad x \notin [-1, 1]. \end{cases} \qquad (2)$$

La densidad de distribución de la suma $\eta_2 = \xi_1 + \xi_2$ se determina por la fórmula (véase (14), § 6)

$$p_{\eta_2}(x) = \int_{-\infty}^{+\infty} p(u)p(x - u)du = \frac{1}{2} \int_{-1}^{1} p(x - u)du.$$

De la fórmula (2) resulta que la expresión subintegral $p(x - u) = 1/2$ si $-1 < x - u < 1$ ó $x - 1 < u < x + 1$ y $p(x - u) = 0$ en los demás casos. La parte común del segmento de integración $[-1, 1]$ y del segmento $[x - 1, x + 1]$ en el cuál $p(x - u) = 1/2$ se puede escribir de la forma $[\text{máx}(-1,\ x-1),\ \text{mín}(1,\ x+1)]$. Por consiguiente, cuando $-2 < x < 0$

$$p_{\eta_2}(x) = \frac{1}{2}\int_{-1}^{x+1} p(x-u)du = \frac{1}{4}\int_{-1}^{x+1} du = \frac{x+2}{4},$$

y si $0 < x < 2$, entonces

$$p_{\eta_2}(x) = \frac{1}{2}\int_{x-1}^{1} p(x-u)du = \frac{1}{4}\int_{x-1}^{1} du = \frac{2-x}{4};$$

cuando $|x| > 2$ tenemos $p_{\eta_2}(x) = 0$. De este modo,

$$p_{\eta_2}(x) = \begin{cases} 0, & \text{si} \quad |x| \geqslant 2, \\ \dfrac{x+2}{4}, & \text{si} \quad -2 \leqslant x \leqslant 0, \\ \dfrac{2-x}{4}, & \text{si} \quad 0 \leqslant x \leqslant 2. \end{cases} \tag{3}$$

La distribución definida por la densidad (3) se llama *distribución de Simpson*.

La densidad de distribución de la suma de tres variables puede ser determinada como densidad de distribución de la suma de dos variables $\eta_3 = \eta_2 + \xi_3$:

$$p_{\eta_3}(x) = \int_{-\infty}^{+\infty} p_{\eta_2}(u)p(x-u)du =$$

$$= \int_{-2}^{0} \frac{u+2}{4} p(x-u)du + \int_{0}^{2} \frac{2-u}{4} p(x-u)du.$$

Una vez determinados los intervalos en los cuales las funciones subintegrales son positivas y efectuados los cálculos, análogos a los que se realizan para dos sumandos, obtenemos

$$p_{\eta_3}(x) = \begin{cases} 0, & \text{si} \quad |x| \geqslant 3, \\ (x+3)^2/16, & \text{si} \quad -3 \leqslant x \leqslant -1, \\ (3-x)^2/8, & \text{si} \quad -1 \leqslant x \leqslant 1, \\ (x-3)^2/16, & \text{si} \quad 1 \leqslant x \leqslant 3. \end{cases} \tag{4}$$

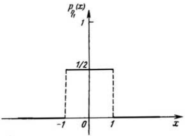

Fig. 13.

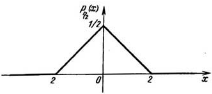

Fig. 14.

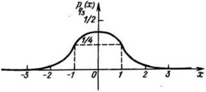

Fig. 15.

Los gráficos de las densidades $p_{\eta_1}(x) = p(x)$, $p_{\eta_2}(x)$ y $p_{\eta_3}(x)$ se muestran en las figs. 13, 14 y 15. Nótese que la densidad de distribución de la suma se hace cada vez más suave con el crecimiento del número de sumandos: $p_{\eta_1}(x)$ es discontinua; $p_{\eta_2}(x)$ es continua; $p_{\eta_3}(x)$ es derivable. Nótese asimismo que la densidad $p_{\eta_3}(x)$ se parece más a la de una densidad normal (en comparación con $p_{\eta_1}(x)$ y $p_{\eta_2}(x)$). Citemos, sin demostrar, las condiciones en que la ley de distribución de la suma de variables aleatorias independientes se aproxima a la distribución normal.

TEOREMA 1. *Si las variables aleatorias $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ son independientes, están igualmente distribuidas y tienen varianza finita, entonces cuando $n \to \infty$ uniformemente respecto a $x \in (-\infty; +\infty)$*

$$\mathbf{P}\left\{\frac{\xi_1 + \xi_2 + \ldots + \xi_n - na}{\sigma\sqrt{n}} < x\right\} \to \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{u^2}{2}}\,du,$$

donde $a = \mathbf{M}\xi_k$, $\sigma^2 = \mathbf{D}\xi_k$, $k = 1, 2, \ldots, n$.

Las condiciones de convergencia de las funciones de distribución de las variables aleatorias independientes que tienen diferentes distribuciones son contenidas en el teorema de Liapunov. Vamos a enunciarlo sin demostrarlo.

TEOREMA 2. *Sean $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ variables aleatorias independientes que poseen tercer momento absoluto finito. Hacemos*

$$a_k = \mathbf{M}\xi_k,\ b_k^2 = \mathbf{D}\xi_k,\ c_k^3 = \mathbf{M}|\xi_k - a_k|^3,\ k = 1, 2, \ldots, n,$$

$$A_n = \sum_{k=1}^{n} a_k,\quad B_n^2 = \sum_{k=1}^{n} b_k^2,\quad C_n^3 = \sum_{k=1}^{n} c_k^3.$$

Si

$$\lim_{n\to\infty} \frac{C_n}{B_n} = 0, \tag{5}$$

entonces cuando $n \to \infty$ uniformemente respecto a x $(-\infty < x < +\infty)$

$$\mathbf{P}\left\{\frac{\xi_1 + \xi_2 + \ldots + \xi_n - A_n}{B_n} < x\right\} \to \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{u^2}{2}}\,du.$$

Al referirse a los teoremas sobre la convergencia de distribuciones de las sumas de variables aleatorias hacia la ley normal, es cómodo utilizar el concepto de normalidad asintótica. Si las funciones de distribución de una sucesión de variables aleatorias $(\eta_n - A_n)/B_n$ convergen, cuando

$n \to \infty$, hacia la función $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{u^2}{2}} du$, entonces se dice que la variable aleatoria η_n es para $n \to \infty$ *asintóticamente normal con parámetros* (A_n, B_n).

Los teoremas 1 y 2 permiten explicar la frecuente aparición de la distribución normal. Por ejemplo, los errores de mediciones resultan a menudo normalmente distribuidos. Este hecho se puede aclarar al tomar en cuenta que un error se compone de un gran número de sumandos compuestos por factores independientes.

La afirmación suficientemente general sobre la convergencia de las distribuciones de las sumas de variables aleatorias hacia la ley normal se llama *teorema central del límite.*

A título de ejemplo de la aplicación de los teoremas 1 y 2 estimemos el número de pruebas que se necesita en el método de Montecarlo para calcular una integral múltiple con precisión prefijada. Supongamos que la función $f(x) = f(x_1, \dots, x_r)$ está definida sobre el cubo unitario r-dimensional V. Se requiere calcular la integral $a = \int_V \dots \int f(x)dx$. Sea que se conoce una constante C tal que $|f(x)| \leqslant C$, $x \in V$. Designemos con $\xi = (\xi_1, \dots, \xi_r)$ un vector aleatorio distribuido uniformemente sobre V. Entonces $p_\xi(x_1, \dots, x_r) = 1$, si $x \in V$ y $p_\xi(x_1, \dots, x_r) = 0$ en el caso contrario. Determinemos la esperanza matemática de la variable aleatoria $\eta = f(\xi)$ con ayuda de la fórmula

$$\mathbf{M}\eta = \int_V \dots \int f(x_1, \dots, x_r) p_\xi(x_1, \dots, x_r) \times$$
$$\times\, dx_1 \dots dx_r = \int_V \dots \int f(x)dx = a.$$

Así, pues, $\mathbf{M}\eta$ coincide con el valor de la integral que se calcula. Como $|f(x)| < C$, entonces

$$\sigma^2 = \mathbf{D}\eta = \int_V \dots \int (f(x) - a)^2 dx \leqslant 4C^2.$$

Supongamos ahora que los vectores aleatorios $\xi_k = (\xi_{k1}, \dots, \xi_{kr})$, $k = 1, \dots, n$, son independientes* y están distribuidas uniformemente

* Los vectores aleatorios ξ_k, $k = 1, \dots, n$, son independientes si para rectángulos r-dimensionales cualesquiera B_k, $k = 1, 2, \dots, n$,

$$\mathbf{P}\{\xi_1 \in B_1, \dots, \xi_n \in B_n\} = \mathbf{P}\{\xi_1 \in B_1\} \dots \mathbf{P}\{\xi_n \in B_n\}.$$

sobre el cubo unitario V. Entonces las variables aleatorias $\eta_k = f(\xi_k)$, $k = 1, \ldots, n$, son independientes y están igualmente distribuidas. Según la ley de los grandes números la variable aleatoria $\zeta_n = (\eta_1 + \ldots + \eta_n)/n$ para grandes valores de n es próxima a la constante $a = \mathbf{M}\eta_k$. Supongamos que se necesita calcular a con una precisión Δ. Estimemos la probabilidad

$$\mathbf{P}\{|\zeta_n - a| < \Delta\} = \mathbf{P}\left\{\left|\frac{\xi_n - a}{\sigma/\sqrt{n}}\right| < \frac{\Delta\sqrt{n}}{\sigma}\right\}.$$

Como $\sigma < 2C$, entonces

$$\mathbf{P}\{|\zeta_n - a| < \Delta\} \geqslant \mathbf{P}\left\{\left|\frac{\zeta_n - a}{\sigma/\sqrt{n}}\right| < \frac{\Delta\sqrt{n}}{2C}\right\},$$

y para grandes valores de n

$$\mathbf{P}\left\{\left|\frac{\zeta_n - a}{\sigma/\sqrt{n}}\right| < \frac{\Delta\sqrt{n}}{2C}\right\} \approx 2\Phi_0\left(\frac{\Delta\sqrt{n}}{2C}\right).$$

Siempre que se asigne una probabilidad pequeña al evento indeseable $|\zeta_n - a| \geqslant \Delta$, se puede, al igual que en el § 4, hallar n.

Problemas

1. Se suman 10^4 números cada uno de los cuales está redondeado con precisión hasta 10^{-m}. Suponiendo que los errores engendrados por el redondeo son independientes y están uniformemente distribuidos en el intervalo $\left(-\frac{1}{2}10^{-m}, \frac{1}{2}10^{m}\right)$, hallar los límites en los cuales, con una probabilidad no menor que 0,99, se encuentra el error total.

2. Obtener el teorema de Moivre — Laplace en calidad de corolario de los teoremas 1 y 2.

3. Supongamos que una variable aleatoria ξ_λ está distribuida uniformemente según la ley de Poisson con parámetro λ. Hallar

$$\lim_{\lambda \to \infty} \mathbf{P}\left\{\frac{\xi_\lambda - \lambda}{\sqrt{\lambda}} < x\right\}.$$

Indicación. Hacer uso del hecho de que la variable aleatoria $\xi_{\lambda_n}(\lambda_n = n \cdot h)$ para todo valor de $h > 0$ se representa de la forma $\xi_{\lambda_h} = \xi_h^{(1)} + \xi_h^{(2)} + \ldots + \xi_h^{(n)}$, donde $\xi_h^{(k)}$ son las variables aleatorias independientes, distribuidas según la ley de Poisson con parámetro h.

§ 12. Procesamiento de los resultados de las mediciones

Las mediciones de magnitudes desconocidas o parámetros se llevan a cabo en los más diversos campos de la actividad científica y técnica. Los resultados de mediciones de una misma magnitud se distinguen en cierto grado entre sí, puesto que es imposible mantener por completo las condiciones en que se realizan diferentes mediciones. Examinemos un modelo matemático elemental del proceso de medición.

12.1. Muestra. Por lo general se procura realizar mediciones independientes entre sí y, aproximadamente, en condiciones iguales. Sea que se realicen n mediciones. Como resultado se obtienen n números $x_1, x_2, \dots, x_n$. Si se repiten una vez más n mediciones, se obtienen otros n números que se distinguen de la primera colección. Es natural describir el proceso de n mediciones independientes como n variables aleatorias independientes cuyos valores numéricos para diferentes ω corresponden a los resultados de mediciones obtenidos en diferentes pruebas.

Se llama muestra de volumen n a n variables aleatorias independientes $X_1, X_2, \dots, X_n$ *cada una de las cuales está distribuida del mismo modo que cierta variable aleatoria* ξ *con función de distribución* $\mathbf{P}\{\xi \leqslant x\} = F_\xi(x)$.

De este modo, $\mathbf{P}\{X_i \leqslant x\} = F_\xi(x)$ para todo valor de $i = 1, 2, \dots, n$.

La muestra es un modelo matemático de n mediciones independientes que se llevan a cabo en las condiciones iguales. La variable aleatoria ξ es la característica del instrumento o del método de mediciones. Las variables $X_1, \dots, X_n$ se pueden considerar como n "ejemplares" independientes de la variable ξ.

12.2. Estimación. Con ayuda de los resultados de las mediciones se necesita hallar un número próximo al valor desconocido del parámetro que se mide. Supongamos, por ejemplo, que por los valores de la muestra del volumen n se requiere evaluar el parámetro desconocido θ de la ley de distribución $\mathbf{P}\{\xi \leqslant x\} = F_\xi(\theta, x)$. Se llama *estimación* del parámetro desconocido θ a una función arbitraria $\hat{\theta}_n = \hat{\theta}_n(X_1, X_2, \dots, X_n)$. Los valores de esta función $\hat{\theta}_n(x_1, x_2, \dots, x_n)$ referentes a los valores de $X_i = x_i$, $i = 1, \dots, n$, obtenidos como resultado de las mediciones, vamos a considerarlos como valor aproximado del parámetro θ. La definición citada de la estimación no refleja sino la exigencia más general consistente en que ésta debe determinarse por los valores de la muestra. Es evidente que cualquier estimación no será obligatoriamente próxima al parámetro θ que se evalúa. Introduzcamos dos propiedades de las estimaciones que aseguran su cercanía a los parámetros correspondientes.

La estimación $\hat{\theta}_n$ del parametro $\hat{\theta}$ se llama insesgada, si $\mathbf{M}\hat{\theta}_n = \theta$.

La estimación $\hat{\theta}_n$ del parámetro θ se llama conciliable si para $n \to \infty$

$$\mathbf{P}\{|\hat{\theta}_n - \theta| < \varepsilon\} \to 1$$

para todo valor de $\varepsilon > 0$.

La propiedad de insesgada significa que la estimación no tiene error sistemático. La propiedad de conciliabilidad asegura la aproximación de la estimación con el parámetro sujeto a medición al aumentar el número de mediciones.

EJEMPLO 1. Examinemos la muestra $X_1, X_2, \dots, X_n$ correspondiente a la variable aleatoria ξ con $F_\xi(x) = \mathbf{P}\{\xi \leqslant x\}$. Sea que nos interese el parámetro $a = \mathbf{M}\xi$. Supongamos que existe $\mathbf{D}\xi = \sigma^2$. Por lo general, en calidad de estimación $\hat{a}_n$ del parámetro a se elige la media aritmética $\bar{X}$ de los valores de la muestra, o sea se supone

$$\hat{a}_n = \bar{X} = \frac{X_1 + \dots + X_n}{n}. \tag{1}$$

De acuerdo con la definición de muestra, las variables $X_1, \dots, X_n$ son independientes y cada una posee distribución coincidente con ξ. Por eso

$$\mathbf{M}X_i = \mathbf{M}\xi = a, \quad \mathbf{D}X_i = \mathbf{D}\xi = \sigma^2.$$

Utilizando las propiedades de linealidad de la esperanza matemática y las propiedades de la varianza, de la fórmula (1) obtenemos

$$\mathbf{M}\hat{a}_n = \mathbf{M}\bar{X} = \frac{1}{n}(\mathbf{M}X_1 + \dots + \mathbf{M}X_n) = \frac{na}{n} = a, \tag{2}$$

$$\mathbf{D}\hat{a}_n = \frac{1}{n^2}(\mathbf{D}X_1 + \dots + \mathbf{D}X_n) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}. \tag{3}$$

La igualdad (2) significa que la estimación $\hat{a}_n = \bar{X}$ es la estimación insesgada del parámetro a. De (3) resulta que la desviación estándar de la estimación $\bar{X}$ a partir del parámetro desconocido a, o sea, la desviación igual a $\sigma/\sqrt{n}$ decrece al aumentar el número de mediciones n.

Haciendo uso de la desigualdad de Chébishev y de la igualdad (3), obtenemos que para todo valor de $\varepsilon > 0$

$$\mathbf{P}\{|\hat{a}_n - a| > \varepsilon\} \leqslant \frac{\mathbf{D}\hat{a}_n}{\varepsilon^2} = \frac{\sigma^2}{\varepsilon^2 n} \to 0$$

cuando $n \to \infty$. Por lo tanto, la probabilidad del evento contrario $\mathbf{P}\{|\hat{a}_n - a| < \varepsilon\} \to 1$, cuando $n \to \infty$, o sea, la estimación $\hat{a}_n = \bar{X}$ es válida. Así, pues, hemos mostrado que

la media aritmética $\bar{X}$ es una estimación insesgada y conciliable del parámetro que se mide si las mediciones de X_i tienen varianza finita.

Pueden existir varias estimaciones insesgadas y válidas distintas del parámetro a. Entre todas estas estimaciones se elige, naturalmente, la de menor varianza.

EJEMPLO 2. Supongamos ahora que por la muestra $X_1, \ldots, X_n$ que corresponde a la distribución de $F_\xi(x)$ se necesitan evaluar dos parámetros $a = \mathbf{M}\xi$ y $\sigma^2 = \mathbf{D}\xi$. En calidad de estimación de a escogemos $\bar{X}$. La estimación $\hat{\sigma}_n^2$ del parámetro σ^2 se define por la fórmula $\hat{\sigma}_n^2 = s^2$, donde

$$s^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \bar{X})^2. \tag{4}$$

Vamos a mostrar que la estimación $\hat{\sigma}_n^2 = s^2$ es una estimación insesgada de σ^2. Transformemos primeramente la fórmula (4), introduciendo las variables aleatorias $Y_k = X_k - c$. Como

$$X_k - \bar{X} = (Y_k + c) - \frac{1}{n}\sum_{k=1}^{n}(Y_k + c) = Y_k - \bar{Y},$$

donde $\bar{Y} = (Y_1 + \ldots + Y_n)/n$, de la fórmula (4) obtenemos

$$s^2 = \frac{1}{n-1}\sum_{k=1}^{n}(Y_k - \bar{Y})^2, \tag{5}$$

o sea, la magnitud s^2 no cambia si todas las magnitudes X_k se desplazan en un mismo valor. Pongamos ahora $c = a = \mathbf{M}X_k$. Entonces

$$\mathbf{M}Y_k = 0,\ \mathbf{D}Y_k = \mathbf{M}Y_k^2 = \mathbf{D}X_k = \sigma^2 \text{y}$$

$$\mathbf{M}(Y_k - \bar{Y})^2 = \mathbf{M}Y_k^2 - \frac{2}{n}\mathbf{M}\left(Y_k\sum_{l=1}^{n}Y_l\right) + \frac{1}{n^2}\sum_{l_1, l_2 = 1}^{n}\mathbf{M}Y_{l_1}Y_{l_2}.$$

Puesto que Y_{l_1} e Y_{l_2} son independientes cuando $l_1 \neq l_2$ y $\mathbf{M}Y_{l_1}Y_{l_2} = \mathbf{M}Y_{l_1}\mathbf{M}Y_{l_2} = 0 (l_1 \neq l_2)$, entonces

$$\mathbf{M}\left(Y_k \sum_{l=1}^{n} Y_l\right) = \mathbf{M}\left(Y_k^2 + \sum_{k:\, k \neq l} Y_k Y_l\right) =$$

$$= \mathbf{M}Y_k^2 + \sum_{k:\, k \neq l} \mathbf{M}Y_k Y_l = \sigma^2,$$

$$\sum_{l_1, l_2 = 1}^{n} \mathbf{M}Y_{l_1} Y_{l_2} = \sum_{l=1}^{n} \mathbf{M}Y_l^2 + \sum_{l_1 \neq l_2} \mathbf{M}Y_{l_1} Y_{l_2} = \sum_{l=1}^{n} \mathbf{M}Y_l^2 = n\sigma^2.$$

Por consiguiente,

$$\mathbf{M}(Y_k - \bar{Y})^2 = \sigma^2 - \frac{2}{n}\sigma^2 + \frac{n\sigma^2}{n^2} = \frac{n-1}{n}\sigma^2.$$

De aquí y de la fórmula (5) encontramos

$$\mathbf{M}s^2 = \frac{1}{n-1} \sum_{k=1}^{n} \mathbf{M}(Y_k - \bar{Y})^2 = \frac{1}{n-1}\left(n \cdot \frac{n-1}{n}\sigma^2\right) = \sigma^2.$$

De este modo, queda demostrado que la estimación s^2 es insesgada. Si se supone que $\mathbf{M}\xi^4$ existe, se puede mostrar que cuando $n \to \infty$

$$\mathbf{D}s^2 = O\left(\frac{1}{n}\right).$$

De aquí, utilizando la desigualdad de Chébishev

$$\mathbf{P}\{|s^2 - \sigma^2| > \varepsilon\} \leqslant \frac{\mathbf{D}s^2}{\varepsilon^2} = O\left(\frac{1}{n}\right),$$

obtenemos, al igual que en el ejemplo 1 para $\bar{X}$, la validez de la estimación s^2.

EJEMPLO 3. Sea que la muestra $X_1, \ldots, X_n$ corresponde a las pruebas del esquema de Bernoulli, es decir, que

$$\mathbf{P}\{X_i = 1\} = p, \quad \mathbf{P}\{X_i = 0\} = q = 1 - p.$$

Entonces la frecuencia del "caso favorable" $h_n = \frac{1}{n}(X_1 + \ldots + X_n)$.

Mostremos que h_n es una estimación válida e insesgada de p. Como $\mathbf{M}X_k = p$ y $\mathbf{D}X_k = pq$ es finita, el esquema en cuestión es un caso particular del esquema general dado en el ejemplo 1. De este modo, se pueden utilizar las deducciones del ejemplo 1 y, por lo tanto, la frecuencia h_n es una estimación válida e insesgada de p.

EJEMPLO 4. Hallemos por la muestra $X_1, \ldots, X_n$, que corresponde a la distribución de $F_\xi(x)$, la estimación $\hat{F}_n(x)$ de la función de distribución $F_\xi(x)$ para cada valor de x. Pongamos

$$\hat{F}_n(x) = \frac{1}{n}\mu_n(x),$$

donde $\mu_n(x)$ es el número de variables aleatorias entre $X_1, X_2, \ldots, X_n$ que no excedan x. La función $\hat{F}_n(x)$ se llama *función de distribución empírica*. Con cada magnitud de X_k se pueden ligar dos sucesos $\{X_k \leqslant x\}$ y $\{X_k > x\}$. Las probabilidades de estos sucesos son, evidentemente, iguales a $p = \mathbf{P}\{X_k \leqslant x\} = F_\xi(x)$ y $q = \mathbf{P}\{X_k > x\} = 1 - F(x)$. Si llamamos "caso favorable" al evento $\{X_k \leqslant x\}$, entonces $\mu_n(x)$ es el número de casos favorables sucedidos en la serie de n pruebas independientes y $\hat{F}_n(x)$ es la frecuencia del caso favorable. Por lo tanto, conforme al ejemplo precedente, $\hat{F}_n(x)$ es una estimación insesgada y válida del parámetro $p = F_\xi(x)$. En la fig. 16 se muestran el gráfico de la función $F_\xi(x) = \Phi(x)$ y el gráfico de $\hat{F}_n(x)$ (para $n = 100$) calculado según la realización tomada entre los números aleatorios normalmente distribuidos (véanse [1], [11], [14]).

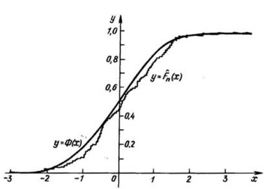

Fig. 16.

A veces, para representar de un modo evidente la muestra, se utiliza un *histograma* que se obtiene así. El eje numérico se divide en varios semiintervalos disjuntos: $(-\infty, \infty) = \bigcup_{k=0}^{r} (z_k, z_{k+1}]$, donde $-\infty = z_0 < z_1 < z_2 < \ldots < z_r < z_{r+1} < \infty$. Luego se determinan las frecuencias de $\hat{p}_k = \hat{F}_n(z_{k+1}) - \hat{F}_n(z_k)$ cuando los elementos de la muestra toman valores pertenecientes a estos intervalos. De segundo, sobre cada segmento $[z_k, z_{k+1}]$ se construye un rectángulo de área proporcional a la frecuencia de $\hat{p}_k$. Las frecuencias de $\hat{p}_k$, siendo grandes los valores de n, son próximas a $p_k = \int_{z_k}^{z_{k+1}} p(x)dx$, donde $p(x)$ es la densidad de distribución de $F_\xi(x)$. Así, pues, si la anchura de los intervalos se elige de un modo conveniente, el histograma puede parecerse al gráfico de la densidad de distribución $p(x)$.

El examen de los gráficos de función de distribución empírica y del histograma puede dar cierta idea previa sobre la función de distribución desconocida $F_\xi(x)$.

En la fig. 17 están representados el gráfico de la densidad de distribución normal con parámetros (0, 1) y el histograma, correspondiente a $\hat{F}_n(x)$ (véase la fig. 16).

EJEMPLO 5. De una urna que contiene M bolas blancas y $N - M$ negras se extraen, según el esquema de selección aleatoria sin reposición, n bolas. Determinemos los indicadores X_k, $k = 1, \ldots, n$, poniendo

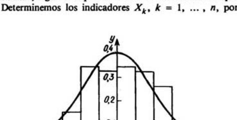

Fig. 17.

$X_k = 1$ si la k-ésima bola (según el orden de extracción) es blanca y $X_k = 0$ en el caso contrario. Las variables aleatorias $X_1, \ldots, X_n$ son, evidentemente, dependientes. En este caso, al igual que en el de las muestras independientes determinadas en el punto 12.1, se pueden buscar estimaciones de los parámetros desconocidos del esquema. Evaluemos, por ejemplo, la porción $p = M/N$ de las bolas blancas en la urna. Hagamos

$$\hat{p}_n = \frac{1}{n}(X_1 + X_2 + \ldots + X_n).$$

La esperanza matemática y la varianza del número de bolas blancas en la muestra $\eta_n = X_1 + \ldots + X_n$ han sido determinadas en el § 8 (ejemplo 5). Utilizando estas fórmulas, obtenemos

$$\mathbf{M}\hat{p}_n = p, \quad \mathbf{D}\hat{p}_n = \frac{1}{n}p(1-p)\cdot\frac{N-n}{N-1}.$$

Así, pues, $\hat{p}_n$ es una estimación insesgada de p. La validez de $\hat{p}_n$ se deduce de la desigualdad de Chébishev

$$\mathbf{P}\{|\hat{p}_n - p| \geqslant \varepsilon\} \leqslant \frac{1}{\varepsilon^2 n}p(1-p)\frac{N-n}{N-1}.$$

12.3. Estimaciones por intervalo. Si se conoce la ley de distribución de una estimación o su varianza, entonces se pueden indicar las fronteras dentro de las cuales se encuentra, con una gran probabilidad, un valor desconocido del parámetro.

Estas fronteras dependen de los valores del parámetro desconocido y entonces no pueden ser utilizadas.

Examinemos la muestra $X_1, X_2, \ldots, X_n$ con $\mathbf{P}\{X_i \leqslant x\} = F_\xi(\theta, x)$, donde θ es el parámetro desconocido. Supongamos que hemos logrado hallar dos funciones $\underline{\theta} = \underline{\theta}(X_1, X_2, \ldots, X_n)$ y $\bar{\theta} = \bar{\theta}(X_1, \ldots, X_n)$ tales que $\underline{\theta} < \bar{\theta}$ para todos los valores $X_1, \ldots, X_n$ y para cualesquiera valores de θ

$$\mathbf{P}\{\underline{\theta}(X_1, \ldots, X_n) < \theta < \bar{\theta}(X_1, \ldots, X_n)\} = 1 - 2\alpha,$$

o sea, la probabilidad de que el intervalo aleatorio $(\underline{\theta}, \bar{\theta})$ cubra el parámetro desconocido θ no depende del parámetro. En este caso el intervalo $(\underline{\theta}, \bar{\theta})$ se llama *intervalo de confianza* para el parámetro desconocido θ, intervalo correspondiente a *la probabilidad de confianza* $1 - 2\alpha$. En algunos casos se pueden hallar las funciones $\underline{\theta}$ y $\bar{\theta}$ que poseen las propiedades indicadas.

Supongamos que existe la muestra $X_1, \dots, X_n$, con ello las variables X_k están distribuidas normalmente con parámetros (a, σ) y el parámetro σ es conocido. Hallemos el intervalo de confianza para a. La variable aleatoria $\bar{X} = (X_1 + \dots + X_n)/n$ tiene distribución normal y $\mathbf{M}\bar{X} = a$, $\mathbf{D}\bar{X} = \sigma^2/n$. Por lo tanto, la magnitud $\dfrac{\bar{X} - a}{\sigma/\sqrt{n}}$ queda distribuida normalmente con parámetros (0, 1) y su distribución no depende de a. De aquí que

$$\mathbf{P}\left\{\left|\frac{\bar{X} - a}{\sigma/\sqrt{n}}\right| < u_\alpha\right\} = 1 - 2\alpha$$

o bien

$$\mathbf{P}\left\{\bar{X} - u_\alpha \frac{\sigma}{\sqrt{n}} < a < \bar{X} + u_\alpha \frac{\sigma}{\sqrt{n}}\right\} = 1 - 2\alpha, \tag{6}$$

donde u_α se determina como solución de la ecuación

$$\frac{1}{\sqrt{2\pi}} \int\limits_{u_\alpha}^{\infty} e^{-\frac{x^2}{2}}\, dx = \alpha. \tag{7}$$

De este modo, hemos encontrado el intervalo de confianza $\left(\bar{X} - u_\alpha \dfrac{\sigma}{\sqrt{n}}, \bar{X} + u_\alpha \dfrac{\sigma}{\sqrt{n}}\right)$ para el parámetro a.

Es más natural la situación cuando ambos parámetros a y σ son desconocidos. En este caso para construir el intervalo confidencial se utiliza la magnitud $\tau_{n-1} = \dfrac{\bar{X} - a}{s}\sqrt{n}$, donde s^2 está definida por la fórmula (4). En los manuales de la estadística matemática se demuestra (véase, por ejemplo, [7]) que la distribución de la magnitud τ_{n-1} no depende de los parámetros a y σ de la distribución normal correspondiente a la muestra; la densidad de distribución τ_{n-1} se define por la fórmula

$$p_{\tau_{n-1}}(x) = C \cdot \left(1 + \frac{x^2}{n-1}\right)^{-\frac{n}{2}}, \quad n \geqslant 2,$$

donde la constante normalizante $C = \Gamma\left(\dfrac{n}{2}\right) \Big/ \Gamma\left(\dfrac{n-1}{2}\right) \times \sqrt{\pi(n-1)}$. La distribución de la magnitud τ_{n-1} se llama *distribución* *de Student con $n - 1$ grados de libertad.* Determinemos la magnitud $t_{\alpha, n-1}$ como solución de la ecuación

$$\mathbf{P}\{|\tau_{n-1}| < t_{\alpha, n-1}\} = 1 - 2\alpha. \tag{8}$$

De aquí, tomando $\tau_{n-1}(\bar{X} - a)\sqrt{n}/s$, obtenemos el intervalo confidencial para a:

$$\mathbf{P}\left\{\bar{X} - t_{\alpha, n-1}\frac{s}{\sqrt{n}} < a < \bar{X} + t_{\alpha, n-1}\frac{s}{\sqrt{n}}\right\} = 1 - 2\alpha. \tag{9}$$

Citemos una pequeña tabla de valores de $t_{\alpha, n}$:

2α \ n	5	10	20	30	∞
0,05	2,571	2,228	2,086	2,042	1,960
0,01	4,032	3,169	2,845	2,750	2,576

Cuando $n \to \infty$ la densidad $p_{\tau_{n-1}}(x)$ converge hacia la densidad de distribución normal con parámetros (0, 1). Puesto que para un valor finito cualquiera de n la densidad $p_{\tau_{n-1}}(x)$, cuando $x \to \infty$, decrece más lentamente que la de distribución normal, entonces $t_{\alpha, n-1} > t_{\alpha, \infty} = u_\alpha$. Esto da lugar a que la longitud media del intervalo de confianza (9) sea mayor que la del intervalo de confianza (6).

Al construir los intervalos de confianza (6) y (9) hemos supuesto que los elementos de las muestras poseen distribución normal. Supongamos ahora que $X_1, X_2, \ldots, X_n$ es una muestra arbitraria. Cuando los valores de n son grandes, se pueden construir intervalos de confianza aproximados sin presuponer la normalidad de X_k. Determinemos, por ejemplo, el intervalo de confianza para el parámetro $a = \mathbf{M}X_k$. Sea $\sigma^2 = \mathbf{D}X_k$. Entonces, según el teorema central del límite, la distribución de la magnitud

$$\frac{X_1 + X_2 + \ldots + X_n - na}{\sqrt{\sigma^2 n}} = \frac{\bar{X} - a}{\sigma/\sqrt{n}}$$

es próxima a la normal con parámetros (0, 1). De aquí, utilizando la validez de la estimación s^2 del parámetro σ^2, se puede mostrar que

$$\lim_{n \to \infty} \mathbf{P}\left\{\left|\frac{\bar{X} - a}{s/\sqrt{n}}\right| < u_\alpha\right\} = 1 - 2\alpha,$$

o bien

$$\mathbf{P}\left\{\bar{X} - u_\alpha \frac{s}{\sqrt{n}} < a < \bar{X} + u_\alpha \frac{s}{\sqrt{n}}\right\} \approx 1 - 2\alpha \tag{10}$$

cuando $n \to \infty$. No vamos a demostrar esta afirmación.

12.4. Método de la máxima verosimilitud para encontrar las estimaciones de parámetros. Método de momentos. Si la función de distribución $F(\theta, x) = \mathbf{P}\{X_k \leqslant x\}$, $k = 1, \dots, n$, tiene la densidad $p(\theta, x)$, entonces la función

$$L = L(X_1, \dots, X_n, \theta) = p(\theta, X_1) \dots p(\theta, X_n) \tag{11}$$

se denomina *función de verosimilitud*. Para una muestra compuesta por las magnitudes discretas X_k, $k = 1, \dots, n$, con distribución $\mathbf{P}\{X_k = x\} = p_x(\theta)$ la función de verosimilitud se define por la igualdad

$$L = L(X_1, \dots, X_n, \theta) = p_{X_1}(\theta) \cdot p_{X_2}(\theta) \cdot \dots p_{X_n}(\theta). \tag{12}$$

Para $X_1, \dots, X_n$ fijas vamos a considerar la función L como función del parámetro θ.

Según el método de la máxima verosimilitud, por estimación del parámetro θ se toma el valor del argumento $\theta = \theta(X_1, \dots, X_n)$ para el cual L tiene el valor máximo, o sea, se elige el valor de θ para el cual la probabilidad de obtener los valores dados de la muestra es máxima.

Puesto que $\ln L$ para $(X_1, \dots, X_n)$ fijas alcanza su máximo en un mismo valor de θ que L, entonces para encontrar la estimación se puede resolver la ecuación de verosimilitud

$$\frac{\partial \ln L}{\partial \theta} = 0. \tag{13}$$

Llamaremos solución (13) sólo a las raíces que dependen de $X_1, X_2, \dots, X_n$. Cada solución de la ecuación de verosimilitud (13) la denominaremos *estimación de máxima verosimilitud para θ*.

En ciertas condiciones bastante generales (véase [7]) la ecuación (13) posee una solución $\hat{\theta}_n$ que es una estimación válida del parámetro θ; además, cuando $n \to \infty$ la estimación θ_n es asintóticamente normal.

Si $\theta = (\theta_1, \dots, \theta_s)$, las estimaciones de máxima verosimilitud de los parámetros $\theta_1, \dots, \theta_s$ son las soluciones, dependientes de $(X_1, \dots, X_n)$, del sistema de ecuaciones:

$$\frac{\partial \ln L}{\partial_k} = 0, \quad k = 1, 2, \dots, s.$$

EJEMPLO 6. Supongamos que las magnitudes X_k, $k = 1, 2, \dots, n$, tienen distribución normal. Los parámetros desconocidos son $a = \mathbf{M}X_k$ y $b = \sigma^2 = \mathbf{D}X_k$. Hallar sus estimaciones de máxima verosimilitud. Por la fórmula (11)

$$L = L(X_1, \dots, X_n, a, b) = \left(\frac{1}{\sqrt{2\pi b}}\right)^n \exp\left(-\sum_{k=1}^{n} \frac{(X_k - a)^2}{2b}\right)$$

y

$$\ln L = -\frac{n}{2}(\ln 2\pi + \ln b) - \frac{1}{2b}\sum_{k=1}^{n}(X_k - a)^2.$$

De aquí, para las estimaciones $\hat{a}$ y $\hat{b}$ obtenemos el sistema

$$\begin{cases} \dfrac{\partial \ln L}{\partial a} = \dfrac{1}{b}\displaystyle\sum_{k=1}^{n}(X_k - \hat{a}) = 0, \\[2ex] \dfrac{\partial \ln L}{\partial b} = -\dfrac{n}{2\hat{b}} + \dfrac{1}{2\hat{b}^2}\displaystyle\sum_{k=1}^{n}(X_k - \hat{a})^2 = 0. \end{cases}$$

De la primera ecuación $\hat{a} = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n} X_k = \bar{X}$. Sustituyendo este valor en la segunda ecuación, encontramos

$$\hat{b} = m_2 = \frac{1}{n}\sum_{k=1}^{n}(X_k - \bar{X})^2.$$

EJEMPLO 7. Hallar la estimación de máxima verosimilitud para la probabilidad del caso favorable p en el esquema de Bernoulli. Por la fórmula

(12) $L = L(X_1, \dots, X_n, p) = \prod_{k=1}^{n} p^{X_k}(1 - p)^{1 - X_k}$, $X_k = 0, 1$

($X_k = 1$ si en la prueba k ha resultado favorable y $X_k = 0$ si ha resultado desfavorable) y

$$\ln L = \sum_{k=1}^{n}(X_k \ln p + (1 - X_k)\ln(1 - p)).$$

De aquí que

$$\frac{\partial \ln L}{\partial p} = \sum_{k=1}^{n} \left(\frac{X_k}{p} - \frac{1 - X_k}{1 - p} \right) = \frac{\bar{X}n}{p} - \frac{n}{1 - \hat{p}} + \frac{n\bar{X}}{1 - \hat{p}} = 0$$

y $\hat{p} = \bar{X}$. Como para el número de casos favorables μ_n tenemos la igualdad $\mu_n = X_1 + X_2 + \ldots + X_n$, entonces $\hat{p} = \mu_n/n$.

§ 13. Método de los cuadrados mínimos

En la práctica las dependencias funcionales desconocidas se determinan con frecuencia por medio de las mediciones. Supongamos, por ejemplo, que se investiga la dependencia funcional

$$y = \varphi(x), \tag{1}$$

donde $\varphi(x)$ es desconocida. Para diferentes valores $x_1, x_2, \ldots, x_n$, son exactamente conocidos y están medidos con errores los valores $y_1, y_2, \ldots,$

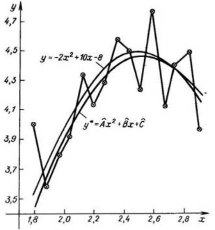

Fig. 18.

y_n que les corresponden. El modelo de tales mediciones se puede definir por las igualdades:

$$y_i = \varphi(x_i) + \delta_i, \quad i = 1, 2, \dots, n, \tag{2}$$

donde δ_i son variables aleatorias independientes, $\mathbf{M}\delta_i = 0$. Si falta por completo la información sobre el carácter de la función $\varphi(x)$, es imposible obtener una aproximación suficientemente satisfactoria de la misma. En este caso, uniendo los pares (x_i, y_i), $i = 1, 2, \dots, n$, por segmentos de rectas, obtenemos un gráfico que poco se asemeja a la función inicial $y = \varphi(x)$ (véanse las figs. 18, 19 y el ejemplo dado al final del § 13).

Por lo general, se suponen conocidas las propiedades de suavidad de la función $\varphi(x)$ o conocida su forma con una exactitud hasta ciertos parámetros. Sea, por ejemplo, $y = \varphi(x, A, B)$, donde la función $\varphi(x, A, B)$ es conocida, x es una variable independiente; A y B son parámetros desconocidos. Supongamos asimismo que en la igualdad (2) las variables δ_i tienen varianzas iguales $\mathbf{D}\delta_i = \sigma^2$. Con ayuda de los valores de los pares (x_i, y_i) se necesita determinar los coeficientes desconocidos A y B. Para

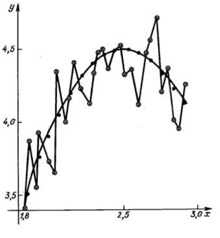

Fig. 19.

calcular las estimaciones de A y B se utiliza *el método de los cuadrados mínimos* consistente en que en calidad de estimaciones de A y B se eligen los valores de $\hat{A}$ y $\hat{B}$ que minimizan la suma de las desviaciones cuadráticas de y_i medidas con relación a los valores calculados respectivos de $\varphi(x_i, A, B)$:

$$Q(A, B) = \sum_{k=1}^{n} (y_k - \varphi(x_k, A, B))^2. \tag{3}$$

De este modo, en calidad de estimaciones $\hat{A}$, $\hat{B}$ de los parámetros A y B se propone tomar la solución del sistema de ecuaciones

$$\frac{\partial Q}{\partial A} = 0, \quad \frac{\partial Q}{\partial B} = 0. \tag{4}$$

Examinemos las propiedades de las estimaciones obtenidas por el método de los cuadrados mínimos cuando se trata de un caso elemental. Supongamos que se requiere encontrar los coeficientes A y B de la función lineal $y = A_x + B$. Consideremos también que en la muestra $(y_1, \ldots \ldots, y_n)$ los valores de y_i son representables de la forma

$$y_i = Ax_i + B + \delta_i, \tag{5}$$

donde δ_i son independientes, $\mathbf{M}\delta_i = 0$, $\mathbf{D}\delta_i = \sigma^2$, x_i son conocidas sin error. Las ecuaciones (4) para la función $\varphi(x) = Ax + B$ tienen la forma

$$\begin{cases} -\dfrac{\partial Q}{\partial A} = 2 \displaystyle\sum_{k=1}^{n} x_k(y_k - Ax_k - B) = 0, \\ -\dfrac{\partial Q}{\partial B} = 2 \displaystyle\sum_{k=1}^{n} (y_k - Ax_k - B) = 0. \end{cases} \tag{6}$$

Desplazando el punto de origen en el eje x, siempre se puede lograr que se cumpla la igualdad $\sum\limits_{k=1}^{n} x_k = 0$. En este caso el sistema (6) se reduce a la forma siguiente:

$$\begin{cases} \sum\limits_{k=1}^{n} x_k y_k - A \sum\limits_{k=1}^{n} x_k^2 = 0, \\ \sum\limits_{k=1}^{n} y_k - B \cdot n = 0. \end{cases}$$

De aquí obtenemos

$$\hat{A} = \left(\sum_{k=1}^{n} x_k y_k\right) \Big/ \left(\sum_{k=1}^{n} x_k^2\right), \qquad \hat{B} = \frac{1}{n} \sum_{k=1}^{n} y_k. \tag{7}$$

Las fórmulas (7) se pueden utilizar para calcular las estimaciones. Para investigar las propiedades de las estimaciones $\hat{A}$ y $\hat{B}$ usamos las fórmulas

$$\hat{A} = A + \frac{\sum\limits_{k=1}^{n} x_k \delta_k}{\sum\limits_{k=1}^{n} x_k^2}, \qquad \hat{B} = B + \frac{1}{n} \sum_{k=1}^{n} \delta_k, \tag{8}$$

que se deducen de las fórmulas (7) si en éstas y_k se sustituye por la fórmula (5). Como las variables δ_k son independientes y $\mathbf{M}\delta_k = 0$, $\mathbf{D}\delta_k = \sigma^2$, entonces

$$\mathbf{M}\left(\frac{\sum\limits_{k=1}^{n} x_k \delta_k}{\sum\limits_{k=1}^{n} x_k^2}\right) = \frac{1}{\sum\limits_{k=1}^{n} x_k^2} \sum_{k=1}^{n} x_k \mathbf{M}\delta_k = 0,$$

$$\mathbf{M}\left(\frac{1}{n} \sum_{k=1}^{n} \delta_k\right) = \frac{1}{n} \sum_{k=1}^{n} \mathbf{M}\delta_k = 0,$$

$$\mathbf{D}\left(\frac{\sum\limits_{k=1}^{n} x_k \delta_k}{\sum\limits_{k=1}^{n} x_k^2}\right) = \frac{1}{\left(\sum\limits_{k=1}^{n} x_k^2\right)^2} \sum_{k=1}^{n} x_k^2 \mathbf{D}\delta_k = \frac{\sigma^2}{\sum\limits_{k=1}^{n} x_k^2},$$

$$\mathbf{D}\left(\frac{1}{n} \sum_{k=1}^{n} \delta_k\right) = \frac{1}{n^2} \sum_{k=1}^{n} \mathbf{D}\delta_k = \frac{\sigma^2}{n}.$$

De aquí y de la fórmula (8) hallamos

$$\mathbf{M}\hat{A} = A, \quad \mathbf{M}\hat{B} = B, \quad \mathbf{D}\hat{A} = \frac{\sigma^2}{\sum_{k=1}^{n} x_k^2}, \quad \mathbf{D}\hat{B} = \frac{\sigma^2}{n}.$$

Así, pues, las estimaciones $\hat{A}$ y $\hat{B}$ son estimaciones insesgadas de los parámetros A y B. Las estimaciones $\hat{A}$ y $\hat{B}$ son válidas si $\sum_{k=1}^{n} x_k^2 \to \infty$ cuando $n \to \infty$, ya que en este caso $\mathbf{D}\hat{A} \to 0$, $\mathbf{D}\hat{B} \to 0$.

Si se supone que δ_i y, por lo tanto, también y_i están distribuidas normalmente, se puede mostrar que las estimaciones obtenidas por el método de los cuadrados mínimos coinciden con las obtenidas por el método de máxima verosimilitud.

En efecto, en este caso la función de verosimilitud de la muestra $y_1, \ldots, y_n$ tiene la forma

$$L(y, A, B) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}\sum_{k=1}^{n} (y_k - Ax_k - B)^2\right\}.$$

De aquí que para las estimaciones de la verosimilitud máxima de los parámetros A y B obtenemos el sistema de ecuaciones

$$\left\{\begin{aligned} -\frac{\partial \ln L}{\partial A} &= \frac{1}{\sigma^2}\sum_{k=1}^{n} x_k(y_k - Ax_k - B) = 0, \\ -\frac{\partial \ln L}{\partial B} &= \frac{1}{\sigma^2}\sum_{k=1}^{n} (y_k - Ax_k - B) = 0 \end{aligned}\right. \tag{9}$$

que no se distingue del sistema (6) sino por el factor constante positivo $1/2\sigma^2$. El método de los cuadrados mínimos se puede aplicar también para funciones no lineales.

EJEMPLO. Sea la función desconocida $y = Ax^2 + Bx + C$. Para los valores de $x_i = \dfrac{2i}{n}$, $i = 0, 1, \ldots, n - 1$, $A = -2$, $B = 10$, $C = -8$ están determinados los valores de $Ax_i^2 + Bx_i + C$ y con ayuda del captador de números aleatorios se han obtenido dos realizaciones ($n = 25$, $n = 100$) $\tilde{\delta}_1, \tilde{\delta}_2, \ldots, \tilde{\delta}_n$ de los errores aleatorios $\delta_1, \ldots, \delta_n$, suponiendo que $\tilde{\delta}_i$ son independientes y están distribuidas normalmente con $\mathbf{M}\delta_i = 0$, $\mathbf{D}\delta_i = 0{,}04$.

Así, pues, se simulan las mediciones

$$\tilde{y}_i = -2x_i^2 + 10x_i - 8 + \tilde{\delta}_i, \quad i = 1, \dots, n. \tag{10}$$

Utilizando el método de los cuadrados mínimos, vamos a evaluar los parámetros A, B y C cuyos valores reales conocemos. Minimizaremos la suma

$$Q(A, B, C) = \sum_{k=1}^{n} (\tilde{y}_k - Ax_k^2 - Bx_k - C)^2.$$

Las estimaciones $\hat{A}$, $\hat{B}$ y $\hat{C}$ satisfacen el sistema de ecuaciones

$$\begin{cases} -\dfrac{\partial Q}{\partial A} = 2\displaystyle\sum_{k=1}^{n} (\tilde{y}_k - Ax_k^2 - Bx_k - C)x_k^2 = 0, \\ -\dfrac{\partial Q}{\partial B} = 2\displaystyle\sum_{k=1}^{n} (\tilde{y}_k - Ax_k^2 - Bx_k - C)x_k = 0, \\ -\dfrac{\partial Q}{\partial C} = 2\displaystyle\sum_{k=1}^{n} (\tilde{y}_k - Ax_k^2 - Bx_k - C) = 0. \end{cases}$$

De aquí hallamos

a) $n = 25$; $\hat{A} = -1{,}871$, $\hat{B} = 9{,}510$, $\hat{C} = -7{,}627$,

b) $n = 100$; $\hat{A} = -2{,}046$, $\hat{B} = 10{,}200$, $\hat{C} = -8{,}218$.

En las figuras 18 y 19 que se refieren a los casos $n = 25$ y $n = 100$, respectivamente, se muestra el gráfico de dependencia funcional exacta $y = -2x^2 + 10x - 8$; con el signo ⊗ están marcados los "resultados de las mediciones" de $(x_i, \tilde{y}_i)$ (en la fig. 18 están marcados todos los puntos con $x_i \in (1{,}8; 3{,}0)$; en la fig. 19 están marcados los puntos con $x_i \in (1{,}8; 3{,}0)$ y $x_{i+1} - x_i = 0{,}04$)) y unidos por la línea quebrada. En la fig. 18 se representa el gráfico de la dependencia funcional $y^* = \hat{A}x^2 + \hat{B}x + \hat{C}$ determinada por las estimaciones para $n = 25$. El gráfico de la dependencia funcional determinada para 100 mediciones casi no se distingue de $y = -2x^2 + 10x - 8$ y por eso en la fig. 19 se muestran solamente los puntos (x_i, y_i^*), donde $y_i^* = \hat{A}x_i^2 + \hat{B}x_i + \hat{C}$ ($n = 100$). Estos puntos llevan el signo •. A continuación se da la tabla de valores de $y_i = -2x_i^2 + 10x_i - 8$, $y_i^* = \hat{A}x_i^2 + \hat{B}x_i + \hat{C}$:

x_i	1,800	1,880	1,960	2,040	2,120	2,200
y_i	3,520	3,731	3,917	4,077	4,211	4,320
$y_i^*(n = 25)$	3,430	3,640	3,826	3,988	4,126	4,241
$y_i^*(n = 100)$	3,511	3,725	3,912	4,073	4,208	4,317

x_i	2,280	2,360	2,440	2,520	2,600	2,680
y_i	4,403	4,461	4,493	4,499	4,480	4,435
$y_i^*(n = 25)$	4,331	4,397	4,440	4,458	4,453	4,423
$y_i^*(n = 100)$	4,399	4,456	4,486	4,490	4,467	4,419

§ 14. Verificación estadística de hipótesis

14.1, Criterio χ^2. En las aplicaciones aparece con frecuencia el problema siguiente. Supongamos que como resultado de cualquier experimento se obtiene la muestra

$$X_1, X_2, \dots, X_n \tag{1}$$

de dimensión n con función de distribución $F(x) = \mathbf{P}\{X_i \leqslant x\}$. Llamaremos *hipótesis estadísticas* o simplemente *hipótesis* a cualesquiera suposiciones sobre la forma de distribución de $F(x)$. A veces nos interesa la hipótesis consistente en que la función de distribución $F(x)$ coincida con cierta función dada $F_0(x)$. El problema de verificación de la hipótesis estadística $F(x) = F_0(x)$ reside en resolver si concuerdan o no con ésta los valores X_i de la muestra (1). Para resolver este problema procedemos del modo siguiente. Con los puntos $z_0 = -\infty < z_1 < z_2 < \dots < z_{r-1} < z_r = \infty$ dividimos toda la recta en r intervalos $(z_{k-1}, z_k]$. Designamos con $p_k = F_0(z_k) - F_0(z_{k-1}) = \mathbf{P}\{z_{k-1} < X_i \leqslant z_k\}$ la probabilidad de que X_i tome un valor perteneciente al intervalo $(z_{k-1}, z_k]$ en el caso en que nuestra hipótesis es justa. Por la muestra (1) determinemos los números ν_k, $k = 1, \dots, r$, donde ν_k es el número de los elementos X_i de la muestra (1) que han ido a parar al intervalo $(z_{k-1}, z_k]$. De este modo, hemos reducido el problema dado a otro, más fácil. Hay n pruebas independientes con r resultados. La probabilidad del k-ésimo resultado es igual a p_k. La colección de las probabilidades de los resultados

$$p_1, p_2, \dots, p_r, \quad \sum_{k=1}^{r} p_k = 1, \tag{2}$$

se determina por la hipótesis estadística inicial. Las variables aleatorias

$$\nu_1, \nu_2, \dots, \nu_r, \quad \sum_{k=1}^{r} \nu_k = n,$$

determinables por la muestra (1) tienen distribución polinomial (véase el § 3 (8)) con probabilidades de los resultados (2):

$$\mathbf{P}\{\nu_k = m_k, k = 1, \dots, r\} = \frac{n!}{m_1! \dots m_r!} p_1^{m_1} \dots p_r^{m_r}, \qquad \sum_{i=1}^{r} m_i = n. \tag{3}$$

Si los valores de ν_k corresponden a las probabilidades p_k, entonces las diferencias $\frac{\nu_k}{n} - p_k$ deben ser pequeñas. Examinemos la variable aleatoria

$$\chi^2 = \sum_{k=1}^{r} \frac{n}{p_k}\left(\frac{\nu_k}{n} - p_k\right)^2 = \sum_{k=1}^{r} \frac{(\nu_k - np_k)^2}{np_k} \tag{4}$$

que llamaremos χ^2 estadístico de Pearson. Si las variables aleatorias ν_k tienen una distribución polinomial (3), entonces es justa la afirmación siguiente: *para todo valor de* $x > 0$

$$\lim_{n \to \infty} \mathbf{P}\{\chi^2 \leqslant x\} = \int_0^x k_{r-1}(u)\,du; \tag{5}$$

la densidad de $k_{r-1}(x)$ *se llama densidad de la distribución* χ^2 *con grado de libertad* $r - 1$ *y tiene la forma siguiente:*

$$k_{r-1}(x) = \begin{cases} \dfrac{x^{\frac{r-3}{2}}}{2^{\frac{r-1}{2}}\Gamma\left(\frac{r-1}{2}\right)} e^{-\frac{x}{2}} & \text{cuando} \quad x \geqslant 0, \\ 0 & \text{cuando} \quad x < 0. \end{cases}$$

Este resultado se utiliza del modo siguiente. Atribuyamos un valor pequeño cualquiera a la probabilidad α que denominamos *nivel de significación* del criterio. Sustituyamos para grandes valores de n la relación límite (5) por la igualdad aproximada

$$\mathbf{P}\{\chi^2 \leqslant x\} \approx \int_0^k k_{r-1}(u)du. \tag{6}$$

Eligiendo $x = x_{\alpha,r-1}$ de modo que $\int\limits_0^{x_{\alpha,r-1}} k_{r-1}(u)du = 1 - \alpha$ ($x_{\alpha,r-1}$ se denomina *cuantil* α de la distribución χ^2), obtenemos que, cuando la hipótesis verificada es justa, el evento

$$\{\chi^2 > x_{\alpha,r-1}\} \tag{7}$$

puede producirse sólo con una pequeña probabilidad que es igual, aproximadamente a α. Por lo general, se supone que $\alpha = 0{,}05$ o bien $\alpha = 0{,}01$. Si la hipótesis es cierta, entonces el evento poco probable (7) es prácticamente imposible. Si este evento sucede, consideraremos que la hipótesis no es justa. Pero si $\chi^2 \leqslant x_{\alpha,r-1}$, decimos que la muestra (1) no contradice la hipótesis $F = F_0$.

En la práctica se considera que la igualdad aproximada (6) puede utilizarse siempre que todos los productos $np_i \geqslant 10$ (a excepción, quizás, de los extremos). Si la elección de los puntos de división z_k y el número de intervalos r depende de nosotros, debemos, por un lado, asegurar la desigualdad $np_i \geqslant 10$ y, por otro lado, no tomar intervalos $(z_{k-1}, z_k]$ muy grandes, para que las probabilidades $p_k = F_0(z_k) - F_0(z_{k-1})$ reflejen suficientemente bien la forma de la función de distribución $F_0(x)$.

Cabe señalar que sólo se pueden utilizar la relación límite (5) y la igualdad aproximada (6), deducida de ésta, en el caso cuando los puntos de división z_k se escogen independientes de la muestra (1).

Si la función de distribución $F(x; \theta_1, \ldots, \theta_s)$ de los elementos de la muestra (1) depende de los parámetros desconocidos $\theta_1, \ldots, \theta_s$, podemos escoger los valores de estos parámetros de modo que el estadístico (4) se convierta en mínimo. Designemos

$$\tilde{\chi}^2 = \min_{\theta_1, \dots, \theta_s} \sum_{k=1}^{s} \frac{(\nu_k - np_k(\theta_1, \dots, \theta_s))^2}{np_k(\theta_1, \dots, \theta_s)},$$

donde $p_k(\theta_1, \dots, \theta_s) = F(z_k; \theta_1, \dots, \theta_s) - F(z_{k-1}; \theta_1, \dots, \theta_s)$. Si en (8) el mínimo se alcanza cuando $\theta_i = \hat{\theta}_i$, entonces

$$\tilde{\chi}^2 = \sum_{k=1}^{r} \frac{(\nu_k - np_k(\hat{\theta}_1, \dots, \hat{\theta}_s))^2}{np_k(\hat{\theta}_1, \dots, \hat{\theta}_s)}, \tag{8}$$

donde $\hat{\theta}_i$ son las estimaciones de los parámetros θ_i, o sea, $\hat{\theta}_i$ son las funciones de la muestra (1), y se pueden tomar por valores aproximados de θ_i. En este caso en vez de la relación límite (5) tiene lugar la afirmación siguiente:

$$\lim_{n \to \infty} \mathbf{P}\{\tilde{\chi}^2 \leqslant x\} = \int_0^x k_{r-s-1}(u)du,$$

o sea, *la función de distribución* $\mathbf{P}\{\tilde{\chi}^2 \leqslant x\}$ *es aproximadamente igual a la función de distribución* χ^2 *con grado de libertad* $r - s - 1$. Esta misma afirmación es válida si en la igualdad (8) las estimaciones $\hat{\theta}_i$ se han obtenido por otro método cualquiera, pero cuando $n \to \infty$, se aproximan bastante rápidamente a θ_i.

EJEMPLO 1. Supongamos que la muestra (1) se ha obtenido de la distribución normal con parámetros (a, σ), o sea, sus elementos tienen la función de distribución

$$\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{x} e^{-\frac{(u-a)^2}{2\sigma^2}} du = \Phi\left(\frac{x-a}{\sigma}\right).$$

Vamos a evaluar los parámetros a y σ del modo siguiente:

$$\hat{a} = \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i, \quad \hat{\sigma}^2 = s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2.$$

Tomamos $F_0(x) = \Phi\left(\frac{x - \bar{X}}{s}\right)$ y luego procedemos del modo descrito anteriormente: determinamos por z_k las probabilidades p_k y calculamos la magnitud χ^2. Sólo ahora, puesto que por la muestra hemos determinado $s = 2$ parámetros, en vez de la igualdad aproximada (6) es necesario usar la igualdad aproximada

$$\mathbf{P}\{\chi^2 \leqslant x\} \approx \int_0^x k_{r-3}(u)du.$$

Al final del libro se muestra la tabla de valores $\chi_{\alpha, m}$ para $m = 1$ a 25, $\alpha = 0{,}05$ y $\alpha = 0{,}01$.

14.2. Elección entre dos hipótesis. A menudo surge el problema siguiente. Supongamos que se tiene una muestra (1). En lo que se refiere a la función de distribución $F(x)$ de esta muestra existen dos hipótesis: hipótesis principal H_0: $F(x) = F_0(x)$ e hipótesis alternativa H_1: $F(x) = F_1(x)$. Por la muestra (1) es necesario determinar qué hipótesis tiene lugar en realidad. Para resolver este problema se construye un *criterio estadístico* que consiste en lo siguiente: la hipótesis H_0 se rechaza (y, por lo tanto, se toma la hipótesis alternativa H_1) si la muestra $(X_1, \dots, X_n)$ va a parar a cierto conjunto *crítico* S, escogido de antemano, o sea si

$$(X_1, \dots, X_n) \in S;$$

la hipótesis H_0 se acepta si

$$(X_1, \dots, X_n) \notin S.$$

Designemos con $\mathbf{P}_i(S)$ la probabilidad del evento $(X_1, \dots, X_n) \in S$ si es justa la hipótesis H_i. Cada criterio se caracteriza por las probabilidades de los errores de primer y segundo género. *La probabilidad* α *del error de primer género* se define por la igualdad $\alpha = \mathbf{P}_0(S)$ y es igual a la probabilidad de rechazar la hipótesis principal si ésta es justa. *La probabilidad* β *del error de segundo género* se define por la igualdad $\beta = \mathbf{P}_1(\bar{S})$, donde $\bar{S} = R^n \setminus S$ es el complemento del conjunto S, e igual a la probabilidad de aceptar la hipótesis principal H_0 si es justa la hipótesis alternativa H_1. La probabilidad $1 - \beta = \mathbf{P}_1(S)$ se llama *potencia* del criterio.

Con ayuda de las probabilidades α y β de los errores de primer y segundo género, los criterios pueden ser comparados entre sí y se puede resolver el problema sobre el mejor criterio. Por lo general, se fija cualquier pequeño valor de la probabilidad del error de primer género (esta probabilidad se denomina frecuentemente *nivel de significación* del criterio) y se busca un criterio con el nivel de significación α y con la potencia máxima $1 - \beta$ (o, lo que es equivalente, con la probabilidad mínima β del error de segundo género). Tal criterio, si existe, se llama criterio *más potente* u *óptimo*.

Es óptimo el llamado *criterio de Neumann — Pearson* que consiste en lo siguiente. Sean $p_0(x) = F_0'(x)$, $p_1(x) = F_1'(x)$ las densidades de distribución de la muestra (1) para las hipótesis principal H_0 y alternativa H_1, respectivamente. Entonces el conjunto crítico del criterio óptimo es necesario buscarlo entre los conjuntos S de la forma siguiente:

$$(X_1, \dots, X_n) \in S \text{ si y sólo si}$$

$$\frac{p_1(X_1)p_1(X_2)\dots p_1(X_n)}{p_0(X_1)p_0(X_2)\dots p_0(X_n)} > C \tag{9}$$

para cierto número $C > 0$.

EJEMPLO 2. Supongamos que la muestra (1) se ha tomado de la distribución normal con parámetros (a, σ); σ es conocido e igual para ambas hipótesis. Las hipótesis H_0 y H_1 se definen por las igualdades:

$$H_0 : a = a_0; \quad H_1 : a = a_1 > a_0.$$

En este caso la desigualdad (9) se transforma en la siguiente:

$$\frac{\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{1}{2\sigma^2}\sum\limits_{i=1}^{n}(X_i - a_1)^2}}{\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{1}{2\sigma^2}\sum\limits_{i=1}^{n}(X_i - a_0)^2}} > C,$$

que es equivalente a la desigualdad

$$\bar{X} > C_1 \tag{10}$$

para cierto valor de C_1 que depende de C, a_0, a_1 y σ. Puesto que para la hipótesis H_i la variable aleatoria $\bar{X}$ posee distribución normal con parámetros $(a_i, \sigma/\sqrt{n})$, los errores de primer y segundo género son iguales a

$$\alpha = \mathbf{P}_0\left(\frac{\bar{X} - a_0}{\sigma/\sqrt{n}} > \frac{C_1 - a_0}{\sigma/\sqrt{n}}\right) = \frac{1}{\sqrt{2\pi}}\int\limits_{\frac{C_1 - a_0}{\sigma/\sqrt{n}}}^{\infty} e^{-\frac{u^2}{2}}\,du = 1 - \Phi\left(\frac{C_1 - a_0}{\sigma/\sqrt{n}}\right),$$

$$\beta = \mathbf{P}_1\left(\frac{\bar{X} - a_1}{\sigma/\sqrt{n}} \leqslant \frac{C_1 - a_1}{\sigma/\sqrt{n}}\right) = \Phi\left(\frac{C_1 - a_1}{\sigma/\sqrt{n}}\right). \tag{11}$$

Determinemos u_γ como solución de la ecuación

$$\Phi(u_\gamma) = 1 - \gamma.$$

En virtud de la simetría de la densidad de la distribución normal con parámetros $(0, 1) u_{1-\gamma} = -u_\gamma$. Entonces de (11) resulta que

$$\frac{C_1 - a_0}{\sigma/\sqrt{n}} = u_\alpha, \quad \frac{C_1 - a_1}{\sigma/\sqrt{n}} = -u_\beta$$

o bien

$$C_1 = a_0 + u_\alpha \frac{\sigma}{\sqrt{n}}, C_1 = a_1 - u_\beta \frac{\sigma}{\sqrt{n}}. \tag{12}$$

De este modo, el criterio (10) con la constante C_1 definida por la primera igualdad (12) ofrece para el valor dado de α la menor probabilidad de error de segundo género β.

De la igualdad (12) podemos determinar el volumen de la muestra n para el cual el criterio óptimo tiene las probabilidades de los errores de primer y segundo género α y β:

$$n = \sigma^2 \frac{(u_\alpha + u_\beta)^2}{(a_1 - a_0)^2}.$$

Tablas

Distribución normal

Valores de la función $\Phi_0(x) = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_0^x e^{-\frac{u^2}{2}}\,du$

Tabla 1

x	Centésimas partes de x									
	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0200	0,0239	0,0279	0,0319	0,0359
0,1	398	438	478	517	557	596	636	675	714	753
0,2	793	832	871	910	948	987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	406	443	480	517
0,4	554	591	628	664	700	736	772	808	844	879
0,5	915	950	985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2257	0,2291	0,2324	357	389	422	454	486	517	549
0,7	580	611	642	673	703	734	764	794	823	852
0,8	881	910	939	967	995	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	289	315	340	365	389
1,0	413	437	461	485	508	531	554	577	599	621
1,1	643	665	686	708	729	749	770	790	810	830
1,2	849	869	888	907	925	944	962	980	997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	192	207	222	236	251	265	279	292	306	319
1,5	332	345	357	370	382	394	406	418	429	441
1,6	452	463	474	484	495	505	515	525	535	545
1,7	554	564	573	582	591	599	608	616	625	633
1,8	641	649	656	664	671	678	686	693	699	706
1,9	713	719	726	732	738	744	750	756	761	767
2,0	772	778	783	788	793	798	803	808	812	817
2,1	821	826	830	834	838	842	846	850	854	857
2,2	861	864	868	871	875	878	881	884	887	890
2,3	893	896	898	901	904	906	909	911	913	916
2,4	918	920	922	925	927	929	931	932	934	936
2,5	938	940	941	943	945	946	948	949	951	952
2,6	953	955	956	957	959	960	961	962	963	964
2,7	965	966	967	968	969	970	971	972	973	974
2,8	974	975	976	977	977	978	979	979	980	981
2,9	981	982	982	983	984	984	985	985	985	986
3,0	987	987	987	988	988	989	989	989	990	990

Valores de la función u_α

La función u_α se define por la igualdad $\alpha = \int_{u_\alpha}^{\infty} e^{-\frac{x^2}{2}} dx$

Tabla 2

α	0,001	0,005	0,010	0,015	0,020	0,025	0,030	0,035	0,040	0,045	0,050
u_α	3,0902	2,5758	2,3263	2,1701	2,0537	1,9600	1,8808	1,8119	1,7507	1,6954	1,6449

Distribución de Poisson

Valores de la función $p_k(\lambda) = \frac{\lambda^k}{k!} e^{-\lambda}$

Tabla 3

k \ λ	0,1	0,2	0,3	0,4	0,5
0	0,90484	0,81873	0,74082	0,67032	0,60653
1	0,09048	0,16375	0,22225	0,26813	0,30327
2	0,00452	0,01638	0,03334	0,05363	0,07582
3	0,00015	0,00109	0,00333	0,00715	0,01264
4		0,00006	0,00025	0,00072	0,00158
5			0,00002	0,00006	0,00016
6					0,00001

k \ λ	0,6	0,7	0,8	0,9
0	0,54881	0,49659	0,44933	0,40657
1	0,32929	0,34761	0,35946	0,36591
2	0,09879	0,12166	0,14379	0,16466
3	0,01976	0,02839	0,03834	0,04940
4	0,00296	0,00497	0,00767	0,01112
5	0,00036	0,00070	0,00123	0,00200
6	0,00004	0,00008	0,00016	0,00030
7		0,00001	0,00002	0,00004

Continuación de la tabla 3

k \ λ	1,0	2,0	3,0	4,0	5,0
0	0,36788	0,13534	0,04979	0,01832	0,00674
1	0,36788	0,27067	0,14936	0,07326	0,03369
2	0,18394	0,27067	0,22404	0,14653	0,08422
3	0,06131	0,18045	0,22404	0,19537	0,14037
4	0,01533	0,09022	0,16803	0,19537	0,17547
5	0,00307	0,03609	0,10082	0,15629	0,17547
6	0,00051	0,01203	0,05041	0,10419	0,14622
7	0,00007	0,00344	0,02160	0,05954	0,10445
8	0,00001	0,00086	0,00810	0,02977	0,06528
9		0,00019	0,00270	0,01323	0,03627
10		0,00014	0,00081	0,00529	0,01813
11		0,00001	0,00022	0,00193	0,00824
12			0,00006	0,00064	0,00343
13			0,00001	0,00020	0,00132
14				0,00006	0,00047
15				0,00002	0,00016
16					0,00005
17					0,00001

Distribución χ^2

Valores de la función $\chi^2_{\alpha,\, m}$

La función $\chi^2_{\alpha,\, m}$ se define por la igualdad $\mathbf{P}\{\chi^2_m > \chi^2_{\alpha,\, m}\} = \alpha$, donde la variable aleatoria χ^2_m posee distribución χ^2 con m grados de libertad. La densidad de distribución de χ^2_m es igual a

$$p_{\chi^2_m}(x) = \frac{1}{\Gamma\left(\frac{n}{2}\right) 2^{\frac{n}{2}}} x^{\frac{n}{2} - 1} e^{-\frac{x}{2}}, x > 0$$

Tabla 4

α \ m	1	2	3	4	5	6	7	8	9
0,05	3,8	6,0	7,8	9,5	11,1	12,6	14,1	15,5	16,9
0,01	6,6	9,2	11,3	13,3	15,J	16,8	18,5	20,1	21,7

Continuación de la tabla 4

α \ m	10	11	12	13	14	15	16	17	18
0,05	18,3	19,7	21,0	22,4	23,7	25,0	26,3	27,6	28,9
0,01	23,2	24,7	26,2	27,7	29,1	30,6	32,0	33,4	34,8

α \ m	19	20	21	22	23	24	25
0,05	30,1	31,4	32,7	33,9	35,2	36,4	37,7
0,01	36,2	37,6	38,9	40,3	41,6	43,0	44,3

Suplemento
Demostración del teorema local de Moivre — Laplace

Estimemo el logaritmo de probabilidad

$$P_n(m) = \mathbf{P}\{\mu_n = m\} = \frac{n!}{m!(n-m)!} p^m q^{n-m}$$

que es igual a

$$\ln P_n(m) = \ln n! - \ln m! - \ln(n-m)! + m\ln p + (n-m)\ln q.$$

Puesto que

$$m = np + x_m\sqrt{npq}, \quad n - m = nq - x_m\sqrt{npq},$$

entonces, haciendo uso de la fórmula de Stirling

$$\ln n! = \ln\sqrt{2\pi n} + n\ln n - n + O\left(\frac{1}{n}\right),$$

obtenemos

$$\ln m! = \frac{1}{2}\ln(2\pi m) +$$

$$+ (np + x_m\sqrt{npq})\ln(np + x_m\sqrt{npq}) - np - x_m\sqrt{npq} + O\frac{1}{n}$$

y

$$\ln(n-m)! = \frac{1}{2}\ln(2\pi(n-m)) +$$

$$+ (nq - x_m\sqrt{npq})\ln(nq - x_m\sqrt{npq}) - nq + x_m\sqrt{npq} + O\left(\frac{1}{n}\right).$$

Utilizando estas fórmulas, así como las fórmulas

$$\ln(np + x_m\sqrt{npq}) = \ln np + x_m\sqrt{\frac{q}{np}} - \frac{x_m^2}{2}\frac{q}{np} + O\left(\frac{q^3}{n\sqrt{n}p^3}\right),$$

$$\ln(nq - x_m\sqrt{npq}) = \ln nq - x_m\sqrt{\frac{p}{nq}} - \frac{x_m^2}{2}\frac{p}{nq} + O\left(\frac{p^3}{n\sqrt{n}q^3}\right),$$

no es difícil obtener la afirmación del teorema en cuestión.

Respuestas a los problemas

§ 1

1. $A(\bar{B} + C)$. **2.** 1) $A\overline{BC}$, 2) ABC, 3) ABC, 4) $A + B + C$, 5) $\overline{ABC}$. **3.** $(C_{95}^{10} + 5C_{95}^{9})/C_{100}^{10} = 0{,}923143 \ldots$ **4.** $\frac{12!}{12^{12}} = 0{,}0000537 \ldots$ **5.** $C_{37}^{6}/C_{49}^{6} = 0{,}1662481 \ldots$

6. $P(A) = 3!/3^3 = \frac{2}{9}$; $P(B) = 3 \cdot 2 \cdot 3/3^3 = 2/3$. **7.** $\pi/4$. **8.** a) $\min(x, 1), x \geqslant 1$. b) $4x(1 - x)$, $0 \leqslant x \leqslant 1/2$; $x \geqslant 1/2$. **9.** $\left(1 - 2\frac{r\sqrt{3}}{3a}\right)^2$. **10.** 1/4. **11.** $p_n = \sum_{k=1}^{n} (-1)^{k+1} \frac{1}{k!} \to 1 - e^{-1}, n \to \infty$.

§ 2

1. 1/12. **2.** 1/3. **4.** C_1, C_2, C_3 no son mutuamente independientes; C_1, C_2 y C_3 son dependientes. **5.** 1) $p_1(1 - p_3)p_4$, 2) $p_1 + p_2 - p_1p_2$, 3) $(p_1 + p_2 - p_1p_2)(p_3 + p_4 - p_3p_4)$. **6.** $(1 - p_3)(1 - p_1p_2)$. **7.** 1) $\alpha_1\alpha_2$, 2) $\beta_1 + \beta_2 - \beta_1\beta_2$. **8.** 11/18. **9.** 1) $p(1 - \beta_1)(1 - \beta_2) + (1 - p)\alpha_1\alpha_2$, 2) $p(1 - \beta_1)(1 - \beta_2)/[p(1 - \beta_1)(1 - \beta_2) + (1 - p)\alpha_1\alpha_2]$. **10.** 20/21. **11.** 9/16. **12.** $e^{-\lambda t}$. **13.** $\pi_0 = 1 - \theta$, $\pi_k = (1 - \theta)\theta^k$, $k \geqslant 1$.

§ 3

1. 3/5; 2/5. **2.** a) 0,95009 … b) 0,0480298 … c) 0,0009782 … **3.** 21; $Q(21) = 0{,}9907768 \ldots$ **4.** $C_{10}^{4}(5/72)^4(1 - 5/72)^6 = 0{,}00317 \ldots$

§ 4

1. 0,99. **2.** 0,001. **3.** 0,00015; 0,00016. **4.** a) 0,13534, b) 0,67668. **5.** $(0{,}9)^n \leqslant 0{,}1$; $n \geqslant 22$. **6.** 96.

§ 5

1. a) $\mathbf{P}\{\chi_A = 1\} = 1/2$; b) $\mathbf{P}\{\chi_B = 1\} = 1/3$; c) $\mathbf{P}\{\xi = 0\} = 1/3$, $\mathbf{P}\{\xi = 1\} = 1/2$, $\mathbf{P}\{\xi = 2\} = 1/6$. **2.** $\mathbf{P}\{\xi = 7\} = 7/16$, $\mathbf{P}\{\xi = 9\} = 5/16$, $\mathbf{P}\{\xi = 11\} = 3/16$, $\mathbf{P}\{\xi = 13\} = 1/16$. **3.** $\mathbf{P}\{\xi = k\} = 1/28$, $k = 0, 1, 11, 12$; $\mathbf{P} = \{\xi = k\} = 1/14$, $k = 2, 3, 9, 10$; $\mathbf{P}\{\xi = k\} = 3/28$, $k = 4, 5, 7, 8$; $\mathbf{P} = \{\xi = 6\} = 1/7$. **4.** $\mathbf{P}(A) = (1 + e^{-2a})/2$, $\mathbf{P}(B) = (1 - e^{-2a})/2$. **5.** $C = 1/2$; $F_\xi(x) = 0$ cuando $x \leqslant 1$;

$F_\xi(x) = 1 - \dfrac{1}{\sqrt{x}}$ cuando $x \geqslant 1$. **6.** $F_\xi(x) = x^2/r^2, 0 \leqslant x \leqslant r; p_\xi(x) = 2x/r^2, 0 \leqslant x \leqslant r; p_\xi(x) = 0$ en los demás casos. **7.** $\mathbf{P}\{\nu = k\} = (21 - 2k)/100, k = 1, 2, \ldots, 10$. **8.** $2 - \sqrt{2}$. **9.** $(1 + \sqrt{2})/4$. **10.** $F_\eta(x) = 0$ cuando $x < 0$, $F_\eta(x) = x$ cuando $0 \leqslant x \leqslant 1$; $F_\eta(x) = 1$ cuando $x \geqslant 1$, $p_\eta(x) = 1$ cuando $0 \leqslant x \leqslant 1$ y $p_\eta(x) = 0$ en los demás casos. **11.** $p_\eta(x) = 1/3\sqrt{x}$ cuando $0 \leqslant x \leqslant 1$, $p_\eta(x) = 1/6\sqrt{x}$ cuando $1 \leqslant x \leqslant 4$; $p_\eta(x) = 0$ en los demás casos.

§ 6

1. a) $\pi/4$; b) $0 = \mathbf{P}\{1/2 \leqslant \xi \leqslant 1, 1/2 \leqslant \eta \leqslant 1\} \neq \mathbf{P}\{1/2 \leqslant \xi \leqslant 1\}.\mathbf{P}\{1/2 \leqslant \eta \leqslant 1\} = \dfrac{1}{4^2}$. **2.** $\mathbf{P}\{\zeta = 6\} = 1/3, \mathbf{P}\{\zeta = 10\} = 1/3, \mathbf{P}\{\zeta = 12\} = 1/3$. **3.** $F_\theta(x) = x^2, 0 \leqslant x \leqslant 1; p_\theta(x) = 2x, 0 \leqslant x \leqslant 1$. **4.** $p_{\xi+\eta}(x) = \lambda^2 x e^{-\lambda x}, x \geqslant 0$.

§ 7

1. 3,5. **2.** 6. **3.** $1/\lambda$. **4.** $(1 - p)/\lambda$. **5.** $R/3$. **6.** $12(1 - 1/12)^{25} = 1{,}36292\ldots$ **7.** $\pi h r^2 + \dfrac{1}{3}\pi\delta^2 h$.

§ 8

1. $\pi/12$, $\pi^2/180$. **2.** $1/\ln 2$, $1/\ln 2$, $(3\ln 2 - 2)/\ln^2 2$. **3.** 1/3; 1/18. **4.** $(m + n)/n; n(m + n)/m^2$. **6.** $p^2(1 + 2p - 3p^2)$.

§ 9

1. $\mathbf{M}\xi_1 = -1/8$, $\mathbf{M}\xi_2 = 0$, $\mathbf{D}\xi_1 = 133/192$, $\mathbf{D}\xi_2 = 1$, $\operatorname{cov}(\xi_1, \xi_2) = -1/8$. **2.** a) 0; 0; 0,5; 0,5. b) 0; 0; las variables aleatorias son dependientes. **3.** $C = 4$; $\operatorname{cov}(\xi_1, \xi_2) = 0$; las variables aleatorias son independientes. **4.** $C = 1/4$, $\rho(\xi_1, \xi_2) = -1/5$. **5.** a) 0, b) 1/3. **6.** a) $\mathbf{P}\{\xi = \eta = i\} = i/36$; $\mathbf{P}\{\xi = i, \eta = j\} = 1/36$, $i < j$; $\mathbf{P}\{\xi = i, \eta = j\} = 0, i > j$, b) $\mathbf{M}\xi = 7/2$, $\mathbf{M}\eta = 161/36$, $\mathbf{D}\xi = 35/12$, $\mathbf{D}\eta = 2555/1296$; c) $\operatorname{cov}(\xi, \eta) = 105/72$.

§ 10

1. $\mathbf{P}\{|\xi - a| \geqslant 2\sigma\} \leqslant 0{,}25$; $\mathbf{P}\{|\xi - a| \geqslant 2\sigma\} = 0{,}0456$. **2.** $\mathbf{P}\{|\xi| \geqslant \Delta\} = \sigma^2/\Delta^2$; $\mathbf{P}\{|\xi| \geqslant \Delta\} \leqslant \mathbf{D}\xi/\Delta^2 = \sigma^2/\Delta^2$. **3.** $\Delta = 0{,}0813$; según la desigualdad de Chébishev $\Delta \leqslant 1/\sqrt{10} = 0{,}3162\ldots$ **4.** Sí.

§ 11

1. $0{,}75 \cdot 10^{-m+2}$. **3.** $\Phi(x)$.

Bibliografía

[1] **Большев Л. Н. Смирнов Н. В.** Таблицы математической статистики. — М.: Наука, 1965.
Bólshev L. N., Smirnov N. V. Tablas de la estadística matemática, *en ruso.*

[2] **Боровков А. А.** Теория вероятностей. — М.: Наука, 1976.
Borovkov A. A. Teoría de las probabilidades, *en ruso.*

[3] **Володин Б. Г.** и др. Сборник задач по теории вероятностей, математической статистике и теории случайных функций. Под ред. А. А. Свешникова. — М.: Наука, 1965.
Volodin B. G. y otros. Problemas de la teoría de las probabilidades, de la estadística matemática y de la teoría de las funciones aleatorias. Dirigido por A. A. Svéshnikov, *en ruso.*

[4] **Гнеденко Б. В.** Курс теории вероятностей. 5-е изд. — М.: Наука, 1971.
Gnedenko B. V. Manual de la teoría de las probabilidades, *en ruso.*

[5] **Ермаков С. М.** Метод Монте-Карло и смежные вопросы. — М.: Наука, 1971.
Ermakov S. M. Método de Montecarlo y cuestiones contiguas, *en ruso.*

[6] **Колмогоров А. Н.*** Основные понятия теории вероятностей. — М.: Наука, 1974.
Kolmogórov A. N.* Conceptos principales de la teoría de las probabilidades, *en ruso.*

[7] **Крамер Г.** Математические методы статистики. 2-е изд. Перевод с англ. — М.: Мир, 1975.
Cramer G. Métodos matemáticos de la estadística, *en ruso*, traducción del inglés.

[8] **Прохоров Ю. В., Розанов Ю. А.** Теория вероятностей. (Основные понятия, предельные теоремы, случайные процессы.) 2-е изд. — М.: Наука, 1973.
Prójorov Yu, V., Rozánov Yu. A. Teoría de las probabilidades. (Conceptos principales, teoremas del límite, procesos aleatorios), *en ruso.*

[9] **Розанов Ю. А.** Случайные процессы. 2-е изд. — М.: Наука, 1979.
Rozánov Yu. A. Procesos aleatorios, *en ruso.*

[10] **Севастьянов Б. А.** Курс теории вероятностей и математической статистики. — М.: Наука, 1982.
Sevastiánov B. A. Manual de la teoría de las probabilidades y de la estadística matemática, *en ruso.*

[11] **Севастьянов Б. А., Чистяков В. П., Зубков А. М.** Сборник задач по теории вероятностей. — М.: Наука, 1980.
Sevastiánov B. A., Chistiakov V. P., Zubkov A. M. Problemas de la teoría de las probabilidades. Editorial "Mir", Moscú, 1985

[12] **Смирнов Н. В., Дунин-Барковский И. В.** Курс теории вероятностей и математической статистики для технических приложений. 3-е изд. — М.: Наука, 1969.
Smirnov N. V., Dunin-Barkovski I. V. Manual de la teoría de las probabilidades y de la estadística matemática para aplicaciones técnicas, *en ruso.*

[13] **Феллер В.** Введение в теорию вероятностей и ее приложения, т. 1. Перевод с англ. — М.: Мир, 1967.
Feller W. Introdución a la teoría de las probabilidades y sus aplicaciones, t. 1, *en ruso*, traducción del inglés.

[14] **Чистяков В. П.** Курс теории вероятностей. 2-е изд. . М.: Наука, 1982.
Chistiakov V. P. Manual de la teoría de las probabilidades, *en ruso.*

---

* El libro de A. N. Kolmogórov "Conceptos principales de la Teoría de las probabilidades" fue editado por primera vez en 1933, *en alemán.*

Este manual corresponde al programa mínimo del curso de la teoría de probabilidades en los establecimentos de enseñanza superior de especialidades técnicas. La exposición en él se efectúa de la forma más simple. El examen de las cuestiones fundamentales de la teoría de las probabilidades va acompañado del análisis de ejemplos y ejercicios; se plantean problemas para ser resueltos por cuenta propia. Una serie de ejercicios responden a las exigencias de las clases prácticas, establecidas en el programa para los institutos superiores de la URSS. Esta obra está destinada a los estudiantes universitarios, y le será también de utilidad a los docentes.

EDITORIAL MIR