Лекция 16

§16. Случайни величини

1. Определения и примери. Числова величина, която в резултат от даден експеримент може да приема различни стойности се нарича ***случайна величина***. Случайната величина X се разглежда като числова функция $X(\omega):\Omega \to \mathbf{R}$, определена върху множеството Ω от елементарните изходи на даден експеримент, което се нарича ***извадъчно пространство*** на експеримента. Тук винаги се предполага, че в Ω е зададена подходяща алгебра от събития $\aleph$ с определена за всяко събитие вероятностна мярка P.

Пример 16.1. В опита за хвърляне на зар можем да определим случайната величина X, равна на показанието на зара. Възможните стойности на тази величина са 1, 2, 3, 4, 5 и 6, при която всяка от тях се приема с вероятност $P(X=k)=p_k=\dfrac{1}{6}$, $1\le k\le 6$. Извадъчното пространство се състои от всичките шест възможни показания на зара.

Пример 16.2. При схемата на Бернули с n повторения на единичен опит с вероятност на сбъдване на единичното събитие p, можем да определим случайната величина X, равна на броя на успехите в серията. Възможните стойности на X са 0, 1, 2, ..., n, които X приема с вероятности

$$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k},\ 0\le k\le n.$$

В този случай казваме, че случайната величина X има ***биномно разпределение*** и пишем $X\in B(n,p)$.

Пример 16.3. При серия от опити до първо сбъдване на дадено събитие A, което при единичен опит се сбъдва с вероятност p, можем да определим величината X, равна на поредния номер на сбъдването. Тук всяко естествено число $k=1,2,\ldots$ представлява възможна стойност на X, при което $P(X=k)=q^{k-1}p$. В този случай казваме, че величината X има ***геометрично разпределение*** и пишем $X\in G(p)$.

Изобщо ако имаме дискретно вероятностно пространство с множество на елементарни изходи $\Omega=\{\omega_1,\omega_2,\ldots\}$ (крайно или безкрайно), можем да определим една случайна величина X, задавайки нейните стойности x_k за всеки елементарен изход ω_k, $k=1,2,\ldots$. В този случай говорим за ***дискретна случайна величина*** (с извадъчно пространство Ω). Дискретните случайни величини се описват от ***таблицата на разпределение*** $p_k=P(X=x_k)$

X	x_1	x_2	$\cdots$
P	p_1	p_2	$\cdots$

Табл. 16.1.

Пример 16.4. Ако $X\in B(5,0.2)$, то X има следната таблица на разпределение

X	0	1	2	3	4	5
P	$\binom{5}{0}(0.2)^0(0.8)^5$	$\binom{5}{1}(0.2)^1(0.8)^4$	$\binom{5}{2}(0.2)^2(0.8)^3$	$\binom{5}{3}(0.2)^3(0.8)^2$	$\binom{5}{4}(0.2)^1(0.8)^4$	$\binom{5}{5}(0.2)^0(0.8)^5$

Табл. 16.2.

Приемането на някаква определена стойност x_k от страна на дадена дискретна случайна величина X представлява случайно събитие. По този начин случайната величина X определя пълна система от несъвместими събития $A_k=\{X=x_k\}$, при което $P(A_k)=p_k$. Очевидно над едно дискретно вероятностно пространство могат да се задават различни случайни величини.

В повечето случаи вероятностните модели позволяват да се абстрахираме от конкретния вид на пространството Ω, от което произлиза дискретната случайна величина X и да разглеждаме основно нейното разпределение. В този смисъл всяка таблица от вида 16.1 задава разпределение на някаква случайна величина стига да бъде налице ***условието за нормировка*** $p_1+p_2+\cdots=1$ $(p_k\ge 0)$.

Пример 16.5. Таблицата

X	-1	2	5
P	0.2	0.3	0.5

Табл. 16.3.

задава разпределение на някаква дискретна случайна величина X.

Нека X е случайна величина. Тогава функцията $F(x)=P(X<x)$ се нарича ***функция на разпределение*** на X. Ако X е дискретна случайна величина, то $F(x)$ има вида

$$F(x)=\sum_{x_k<x} p_k .$$

Непосредствено от определението се получават следните основни свойства на функцията на разпределение.

1) $F(x)$ е монотонно растяща.

2) $\lim\limits_{x\to-\infty} F(x)=F(-\infty)=0$ и $\lim\limits_{x\to\infty} F(x)=F(\infty)=1$.

3) $F(x)$ е непрекъсната отляво, $F(x)=F(x-0)$.

Ако X е дискретна случайна величина, то $F(x)$ представлява стъпаловидна функция, прекъсната в точките $x=x_k$, където x_k е някоя стойност на X с $p_k>0$.

Освен дискретни ще разглеждаме и непрекъснати случайни величини. Една случайна величина X се нарича ***непрекъсната***, когато нейната функция на разпределение може да се представи във вида

$$F(x)=P(X<x)=\int_{-\infty}^{x} f(t)dt$$

където $f(x)$ е някаква ***неотрицателна*** функция, определена за всяко $x\in\mathrm{R}$ и интегруема над всеки числов интервал. Функцията $f(x)$ се нарича ***плътност на разпределение*** на непрекъснатата случайна величина X. Една типична форма на плътност е показана на рис. 16.1.

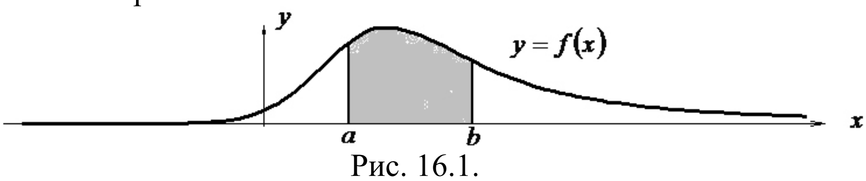

Рис. 16.1.

Съгласно геометричната интерпретация на определения интеграл, вероятността

$$P(a\le X<b)=F(b)-F(a)$$

е равна на лицето на съответния криволинеен трапец (рис. 16.1). Ако X е непрекъсната, то функцията на разпределение $F(x)$ е непрекъсната за всяко $x\in\mathrm{R}$, а в тези точки x, в които плътността $f(x)$ е непрекъсната функция, $F(x)$ е диференцируема и $F'(x)=f(x)$. В частност, ако $f(x)$ е непрекъсната навсякъде, $F(x)$ представлява една примитивна на $f(x)$. Вероятността $P(a\le X<b)$, на събитието

стойността на разглежданата величина да се помества в интервала $[a,b)$ се дава по формулата

(16.1) $$P(a \le X < b) = \int_a^b f(x)dx.$$

Трябва да се помни обаче, че в теорията съществуват случайни величини с много по-сложен характер от определените дотук. Понеже събитието $\{-\infty < X < \infty\}$ е достоверно, от формулата (16.1) намираме, че

(16.2) $$\int_{-\infty}^{\infty} f(x)dx = 1.$$

В повечето случаи можем да се абстрахираме от произхода на дадена непрекъсната случайна величина и да разглеждаме основно нейното разпределение. По този начин можем да приемем, че информацията за величината X се изчерпва с формулата (16.1), което означава, че една непрекъсната случайна величина идентифицираме (при направените уговорки) посредством нейната плътност $f(x)$. В този смисъл всяка неотрицателна функция $f(x)$, за която е изпълнено ***условието за нормировка*** (16.2), поражда някаква непрекъсната случайна величина X.

Съществуват много непрекъснати случайни величини, които имат принципно значение в теорията на вероятностите и математическата статистика. Определено най-важният пример се получава, ако разгледаме непрекъснатата случайна величина X, която има плътност

(16.3) $$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

където $\mu \in \mathrm{R}$ и $\sigma > 0$ са параметри, чиито смисъл ще бъде изяснен по-нататък. Не е трудно да се покаже, че така определената плътност удовлетворява условието за нормировка (16.2) и следователно наистина представлява вероятностна плътност. В този случай се казва, че случайната величина е ***разпределена нормално*** с параметри μ и σ^2, което се записва $X \in N(\mu, \sigma^2)$. При $\mu = 0$ и $\sigma = 1$ разпределението се нарича ***нормално стандартно***. Нормалните разпределения имат водещо значение в цялата теория. Вида на нормалното разпределение е показан на рис. 16.2.

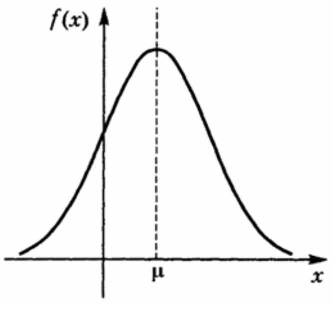

Рис. 16.2.

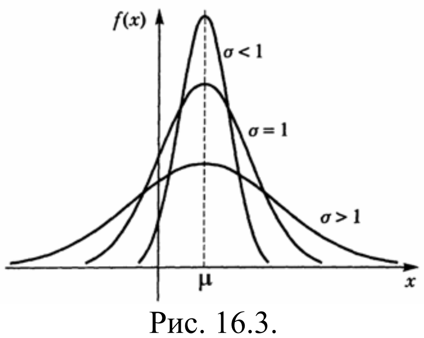

Рис. 16.3.

Функцията на плътност за нормалното разпределение е симетрична относно $x = \mu$, а в тази точка има локален максимум, равен на $\frac{1}{\sigma\sqrt{2\pi}}$. На рис. 16.3 е показано как се мени формата на плътността, когато се мени параметърът σ.

Пример 16.6. Казва се, че непрекъсната случайна величина X е разпределена ***равномерно*** в интервала $[a,b]$, $(b > a)$, когато има плътност (Рис. 16.4)

$$f(x)=\begin{cases} 0 \;\; за \;\; x<a \\ \dfrac{1}{b-a} \;\; за \;\; x\in[a,b] \\ 0 \;\; за \;\; x>b \end{cases}$$

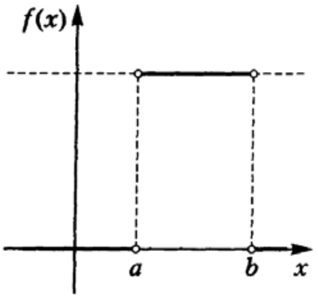

Рис. 16.4.

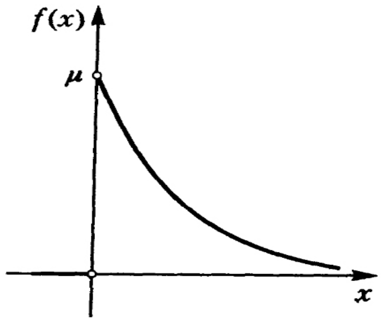

Рис. 16.5.

Да проверим условието за нормировка,

$$\int_{-\infty}^{\infty} f(x)dx=\int_a^b \frac{dx}{b-a}=\frac{1}{b-a}\int_a^b dx=\frac{1}{b-a}(b-a)=1 .$$

В този случай пишем $X\in U(a,b)$.

Пример 16.7. Казва се, че непрекъсната случайна величина X е разпределена ***експоненциално*** с параметър μ $(\mu>0)$, когато има плътност

$$f(x)=\begin{cases} 0 \;\; за \;\; x<0 \\ \mu e^{-\mu x} \;\; за \;\; x\ge 0 \end{cases}$$

Да проверим условието за нормировка,

$$\int_{-\infty}^{\infty} f(x)dx=\lambda\int_0^{\infty} e^{-\lambda x}dx=\int_0^{\infty} e^{-\lambda x}d(\lambda x)=-e^{-\lambda x}\Big|_0^{\infty}=-\lim_{x\to\infty}e^{-\lambda x}+1=0+1=1 .$$

Графиката на тази плътност е показана на фиг. 16.5. В този случай пишем $X\in E(\mu)$.

Непрекъснатите случайни величини представляват абстрактни математически понятия, доказали практически ползата от тяхното въвеждане и удобството за боравене с тях. Елементарен пример за пояснение на понятието непрекъсната случайна величина може да се получи въз основа на внимателен анализ на равномерно разпределена случайна величина от пример 16.6 с помощта на въображаем опит за ***случаен избор на точка от интервала*** $[a,b]$. В този случай понякога се казва, че "хвърляме" по случаен начин точка в интервала $[a,b]$ и отчитаме нейното попадение $\omega\in[a,b]$. Ако положим $X(\omega)=\omega$, то се получава случайна величина, която е разпределена равномерно в $[a,b]$. Тук извадъчното пространство Ω представлява целия интервал $[a,b]$. Множеството от събитията $\aleph$ обаче има сложна природа, чието пълно описание надхвърля съдържанието на настоящия лекционен курс. Едно типично събитие A представлява попадение на точката в зададен интервал $\langle\alpha,\beta\rangle$, $a\le\alpha\le\beta\le b$, при което

$$P(A)=P(X\in\langle\alpha,\beta\rangle)=\frac{\beta-\alpha}{b-a} .$$

Изобщо задаването на непрекъснати случайни величини става посредством идеята за геометрична вероятност.

Пример 16.8. Ако $X\in N(0,1)$ (нормално стандартно разпределение), то

$$F(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}dt .$$

Интегралът в дясната страна на последното равенство не може да бъде приведен в по-елементарен вид, понеже съответният неопределен интеграл не може да се запише като краен израз от основните елементарни функции и основните алгебрични операции. Поради честата си употреба на много места стойностите на функцията на разпределение за нормалното стандартно разпределение са табулирани.

2. Съвместни, маргинални и условни разпределения. Дискретните случайни величини се представят чрез стойности и техните вероятности, които за прегледност могат да бъдат подредени в таблица. Непрекъснатите случайни величини се представят посредством техните плътности. В резултат на даден експеримент може да бъдат наблюдавани няколко случайни величини едновременно. В такъв случай можем да говорим за ***случайни вектори***. Ако искаме да анализираме вероятностните връзки между две (или повече на брой) случайни величини, трябва да разполагаме с информация за тяхното съвместно разпределение.

Да предположим за простота и определеност, че разглеждаме две дискретни величини X и Y с краен брой стойности. Тогава тяхното съвместно разпределение се дава с помощта на ***таблицата на съвместното разпределение*** както следва.

	y_1	y_2	$\cdots$	y_n
x_1	p_{11}	p_{12}	$\cdots$	p_{1n}
x_2	p_{21}	p_{22}	$\cdots$	p_{2n}
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
x_m	p_{m1}	p_{m2}	$\cdots$	p_{mn}

Табл. 16.4.

Условието за нормировка в този случай е

$$(16.4)\quad \sum_{i=1}^{m}\sum_{j=1}^{n} p_{ij} = 1.$$

Тук x_1, x_2, ..., x_m са възможните стойности на случайната величина X, а y_1, y_2, ..., y_n са възможните стойности на Y. Числата p_{ij} задават вероятността величината X да има стойност x_i, а величината Y да има стойност y_j, $1 \le i \le m$, $1 \le j \le n$, което записваме $p_{ij} = P(X = x_i, Y = y_j)$. Случайните събития $A_{ij} = \{X = x_i, Y = y_j\}$ образуват пълна група несъвместими събития, които характеризират вероятностната схема. По този начин $p_{ij} = P(A_{ij})$. Всяка от групите събития $A_{i\bullet} = \{X = x_i\}$, $1 \le i \le m$, и $A_{\bullet j} = \{Y = y_j\}$, $1 \le j \le n$, също образува пълна група от несъвместими събития, които характеризират индивидуалното поведение съответно на величините X и Y. По формулата за пълната вероятност имаме

$$P(A_{i\bullet}) = P(X = x_i) = \sum_{j=1}^{n} P(X = x_i \mid Y = y_j) P(Y = y_j),$$

следователно

$$(16.5)\quad P(A_{i\bullet}) = P(X = x_i) = \sum_{j=1}^{n} P(X = x_i, Y = y_j) = \sum_{j=1}^{n} p_{ij} = p_{i\bullet},\ 1 \le i \le m.$$

Аналогично

$$(16.6)\quad P(A_{\bullet j}) = P(Y = y_j) = \sum_{i=1}^{m} P(Y = y_j, X = x_i) = \sum_{i=1}^{m} p_{ij} = p_{\bullet j},\ 1 \le j \le n.$$

По този начин от таблицата на съвместното разпределение получихме индивидуалните разпределения на X и Y, които се наричат още ***маргинални разпределения***.

	y_1	y_2	$\cdots$	y_n		
x_1	p_{11}	p_{12}	$\cdots$	p_{1n}	$p_{1\bullet}$	X
x_2	p_{21}	p_{22}	$\cdots$	p_{2n}	$p_{2\bullet}$	
...	...	...	...	...	...	
x_m	p_{m1}	p_{m2}	$\cdots$	p_{mn}	$p_{m\bullet}$	
	$p_{\bullet 1}$	$p_{\bullet 2}$	$\cdots$	$p_{\bullet n}$		
	Y					

Табл. 16.5.

Съответните вероятности се получават посредством сумиране редовете и стълбовете на вероятностите в таблица 16.4, което е показано в таблица 16.5. При получаване маргиналното разпределение на X фактически се елиминира присъствието на другата величина Y и обратно.

По формулата за умножение на вероятностите, при всяко фиксирано j имаме

(16.7) $P(X = x_i \mid Y = y_j) = \dfrac{P(X = x_i, Y = y_j)}{P(Y = y_j)} = \dfrac{p_{ij}}{p_{\bullet j}}$, $1 \le i \le m$,

което задава ***условното разпределение*** на величината X, при условие $Y = y_j$, $1 \le j \le n$. Формулата (16.7) изпълнява условието за нормировка, понеже

$$\sum_{i=1}^{m} \frac{p_{ij}}{p_{\bullet j}} = \frac{1}{p_{\bullet j}} \sum_{i=1}^{m} p_{ij} = \frac{p_{\bullet j}}{p_{\bullet j}} = 1,\ 1 \le i \le m.$$

Аналогично, условното разпределение на величината Y при условие $X = x_i$ се дава по формулите

(16.8) $P(Y = y_j \mid X = x_i) = \dfrac{P(X = x_i, Y = y_j)}{P(X = x_i)} = \dfrac{p_{ij}}{p_{i\bullet}}$, $1 \le j \le n$.

Таблицата на съвместното разпределение ни позволява да елиминираме напълно едната величина, получавайки маргиналното разпределение на другата, а също така и да разглеждаме условното разпределение на едната величина, при условие, че другата има някаква определена стойност.

Величините X и Y се наричат ***независими***, когато всяко от събитията $A_{i\bullet}$ е независимо спрямо всяко от събитията $A_{\bullet j}$, т.е. когато приемат стойности по независим една от друга начин.

Твърдение 16.1. Дискретните случайни величини X и Y са независими тогава и само тогава, когато за всяка стойност x_i на X и за всяка стойност y_j на Y е изпълнено равенството $p_{ij} = p_{i\bullet} p_{\bullet j}$.

Доказателство. По определение X и Y са независими тогава и само тогава, когато за всяко i и всяко j е изпълнено $P(A_{i\bullet} A_{\bullet j}) = P(A_{i\bullet}) P(A_{\bullet j})$, което написано по друг начин означава $p_{ij} = p_{i\bullet} p_{\bullet j}$. □

Нека X и Y са независими. Тогава от (16.7) и твърдение 16.1 получаваме

(16.9) $P(X = x_i \mid Y = y_j) = \dfrac{p_{i\bullet} p_{\bullet j}}{p_{\bullet j}} = p_{i\bullet} = P(X = x_i)$, $1 \le i \le m$,

което означава, че условното разпределение на X при всяка стойност на Y съвпада с неговото безусловно (маргинално) разпределение. Аналогично е положението и при условното разпределение на Y. Направеният извод напълно съответства на интуитивните представи за независимост между случайни величини имайки предвид, че понятието разпределение се явява първично в теорията на случайните величини. Всъщност тази независимост на условните разпределения може да се възприеме и като определение за независимост между величините X и Y.

При две дискретни случайни величини X и Y, тяхната съвместна функция на разпределение $F(x,y)$ се дава по формулата

(16.10) $F(x,y)=\sum_{x_i<x}\sum_{y_j<y}p_{ij}$,

откъдето лесно се установява, че $F(x,\infty)=\lim_{y\to\infty}F(x,y)$, представлява индивидуалната (маргиналната) функция на разпределение за случайната величина X, понеже

(16.11) $F(x,\infty)=\sum_{x_i<x}\sum_{j=1}^{n}p_{ij}=\sum_{x_i<x}p_{i\bullet}$.

Аналогично, $F(\infty,y)=\lim_{x\to\infty}F(x,y)$,

(16.12) $F(\infty,y)=\sum_{i=1}^{m}\sum_{y_j<y}p_{ij}=\sum_{y_j<y}p_{\bullet j}$

представлява индивидуалната функция на разпределение за случайната величина Y. От формулите (16.10-12) и твърдение 16.1 следва, че величините X и Y са независими тогава и само тогава, когато $F(x,y)=F(x,\infty)F(\infty,y)$.

Пример 16.9. Нека таблицата на вероятностите на съвместното разпределение на величините X и Y има вида

	y_1	y_2	y_3	
x_1	0.1	0.1	0.3	**0.5**
x_2	0.2	0.2	0.1	**0.5**
	0.3	**0.3**	**0.4**	

Табл. 16.6.

В този случай величините X и Y са зависими, понеже не е изпълнено условието за независимост $p_{ij}=p_{i\bullet}p_{\bullet j}$, например $p_{11}=0.1\neq p_{1\bullet}p_{\bullet 1}=0.5*0.3=0.15$.

При непрекъснатите величини нещата се описват на езика на плътностите. Съвместното разпределение на две непрекъснати случайни величини се характеризира посредством тяхната съвместна функция на плътност $f(x,y)$, която е неотрицателна и интегруема над R^2 при изпълнение на следното условие за нормировка

$$\iint_{\mathrm{R}^2} f(x,y)dxdy=1.$$

Ако $D\subseteq\mathrm{R}^2$ е някакво измеримо множество в равнината, то вероятността на събитието $\{(x,y)\in D\}$ се определя от формулата

$$P((x,y)\in D)=\iint_D f(x,y)dxdy,$$

която представлява двумерен аналог на (16.1).

За ***индивидуалните (маргиналните) плътности*** на величините X имаме

$$f_X(x)=\int_{-\infty}^{\infty}f(x,y)dy \text{ и } f_Y(y)=\int_{-\infty}^{\infty}f(x,y)dx,$$

които са непрекъснати аналози на "дискретните" формули (16.5) и (16.6). За ***условната плътност*** на X при условие $Y=y$, която се бележи с $f(x\mid y)$, имаме

(16.13) $f(x \mid y) = \dfrac{f(x,y)}{f_Y(y)} = \dfrac{f(x,y)}{\int\limits_{-\infty}^{\infty} f(x,y)dx}$,

което е непрекъснат аналог на "дискретната" формула (16.7). Веднага се вижда, че е налице условието за нормировка

$$\int\limits_{-\infty}^{\infty} f(x \mid y)dx = 1.$$

Аналогично, за условната плътност на Y при условие $X = x$ имаме

$$f(y \mid x) = \frac{f(x,y)}{f_X(x)} = \frac{f(x,y)}{\int\limits_{-\infty}^{\infty} f(x,y)dy}.$$

Непрекъснатите случайни величини X и Y се наричат ***независими***, когато всевъзможните събития $P(a < X < b)$ и $P(c < Y < d)$ са независими, т.е. когато двете величини приемат стойности независимо една от друга.

Следното твърдение представлява аналог на твърдение 16.1 за случая на непрекъснати величини.

Твърдение 16.2. Непрекъснатите случайни величини X и Y са независими тогава и само тогава, когато съвместната им плътност се явява произведение на двете маргинални плътности, т.е. когато е изпълнено $f(x,y) = f_X(x) f_Y(y)$. □

В съдържанието на последното твърдение следва да се има предвид, че ако $f(x,y)$ и $\hat{f}(x,y)$ се различават върху множество с мярка нула, то за всяко измеримо множество $D \subseteq \mathbb{R}^2$ е изпълнено

$$P((x,y) \in D) = \iint\limits_D f(x,y)dxdy = \iint\limits_D \hat{f}(x,y)dxdy,$$

следователно двете функции са неразличими в ролята им на съвместна плътност за величините X и Y.

Ако X и Y са независими, то според (16.13) и твърдение 16.2 имаме

$$f(x \mid y) = \frac{f(x,y)}{f_Y(y)} = \frac{f_X(x) f_Y(y)}{f_Y(y)} = f_X(x),$$

което представлява непрекъснат аналог на формулата (16.9) и означава, че всичките условни плътности на X са равни на безусловната плътност. Аналогично се получава и за условната плътност на Y. Последният факт може, както и втората част на твърдение 16.2 могат да се разглеждат поотделно в качеството на определение за независимост на непрекъснати случайни величини.

При непрекъснати случайни величини X и Y, тяхната съвместна функция на разпределение $F(x,y)$ се дава по формулата

$$F(x,y) = \iint\limits_{t<x,\tau<y} f(t,\tau)dtd\tau,$$

което е непрекъснат аналог на (16.10). Оттук лесно се установява, че

$$F(x,\infty) = \lim_{y\to\infty} F(x,y) \text{ и } F(\infty,y) = \lim_{x\to\infty} F(x,y)$$

представляват индивидуалните функции на разпределения за случайните величини X и Y. От последните заключения и от твърдение 16.2 следва, че величините X и Y са независими тогава и само тогава, когато $F(x,y) = F(x,\infty)F(\infty,y)$, което е напълно аналогично с дискретния случай.

3. Действия със случайни величини. Нека X е дискретна случайна величина, зададена чрез своите стойности и техните вероятности, $p_k = P(X = x_k)$, $k = 1,2,\dots$ и нека $g(x)$ е някаква функция, определена над стойностите на X. Тогава под $Y = g(X)$ ще разбираме дискретна случайна величина със стойности $y_k = g(x_k)$, които приема със същите вероятности $P(Y = y_k = g(x_k)) = p_k$.

Пример 16.10. Ако X е зададена чрез таблицата

X	-1	2	5
P	0.1	0.4	0.5

Табл. 16.7.

Тогава случайната величина $Y = X^2$ $\left(g(x) = x^2\right)$ има следната таблица на разпределение

X	1	4	25
P	0.1	0.4	0.5

Табл. 16.8.

Нека сега X е непрекъсната случайна величина с плътност $f_X(x)$ и нека $g(x)$ е някаква непрекъсната функция. Тогава $Y = g(X)$ представлява непрекъсната случайна величина, чиято плътност подлежи на определяне според дефинициите.

Пример 16.11. Нека величината X, следва нормално стандартно разпределение, $X \in N(0,1)$, което означава, че X има плътност

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

и да разгледаме величината $Y = X^2$. Да потърсим израз за функцията на разпределение $F_Y(y) = P(Y < y)$. Ще предполагаме $y \geq 0$, понеже при $y < 0$ очевидно $F_Y(y) = 0$. Имаме

$$F_Y(y) = P(Y < y) = P(X^2 < y) = P(-\sqrt{y} < X < \sqrt{y}) = \frac{1}{\sqrt{2\pi}} \int_{-\sqrt{y}}^{\sqrt{y}} e^{-\frac{t^2}{2}} dt = \frac{2}{\sqrt{2\pi}} \int_0^{\sqrt{y}} e^{-\frac{t^2}{2}} dt,$$

откъдето, след смяна на променливата $t^2 = \tau$, намираме

$$F_Y(y) = \frac{1}{\sqrt{2\pi}} \int_0^y \frac{1}{\sqrt{\tau}} e^{-\frac{\tau}{2}} dt,$$

което означава, че Y има плътност

$$f_Y(y) = \frac{1}{\sqrt{2\pi y}} e^{-\frac{y}{2}},\ y > 0,$$

за което полагаме $f_Y(y) = 0$ при $y \leq 0$. В този случай се казва, че Y има "хи квадрат" разпределение с една степен на свобода, $Y \in \chi^2(1)$.

Най-прости са линейните преобразования. Нека X е случайна величина с функция на разпределение $F_X(x)$ и нека $Y = ax + b$, където a и b са константи, $a > 0$. Тогава за функцията на разпределение $F_Y(y)$ на случайната величина Y получаваме

$$(16.14)\quad F_Y(y) = P(Y < y) = P(aX + b < y) = P\left(X < \frac{y-b}{a}\right) = F_X\left(\frac{y-b}{a}\right).$$

При $a < 0$ по аналогичен начин се получава

$$F_Y(y) = 1 - F_X\left(\frac{y-b}{a}\right).$$

Твърдение 16.3. Нека $X \in N(\mu, \sigma^2)$ има нормално разпределение с параметри μ и σ . Тогава случайната величина $Y = aX + b \quad (a \neq 0)$ има нормално разпределение с параметри $a\mu + b$ и $a\sigma$, $Y \in N(a\mu + b, a^2\sigma^2)$.

Доказателство.По условие X има плътност

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

и функция на разпределение $F_X(x) = \int_{-\infty}^{x} f_X(t)dt$. За определеност да предположим, че $a > 0$. Сега от (16.14) намираме

$$F_Y(y) = F_X\left(\frac{y-b}{a}\right) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\frac{y-b}{a}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt .$$

Извършваме смяна на променливите $at + b = \tau$, след което получаваме

$$F_Y(y) = \frac{1}{\sqrt{2\pi}\sigma a} \int_{-\infty}^{y} e^{-\frac{(\tau - a\mu - b)^2}{2\sigma^2 a^2}} d\tau ,$$

което доказва твърдението при $a > 0$. Случаят $a < 0$ се изследва аналогично. □

От твърдение 16.3 веднага следва, че ако $X \in N(\mu, \sigma^2)$, то величината

(16.15) $Y = \dfrac{X - \mu}{\sigma} \in N(0,1)$,

т.е. Y следва нормално стандартно разпределение. Действията, чрез които от X се получава Y във формулата (16.15), се наричат ***центриране*** – изваждането на μ и ***нормиране*** – разделяне на σ .

Нека X_1 и X_2 са две нормално разпределени величини $X_1 \in N(\mu_1, \sigma_1^2)$ и $X_2 \in N(\mu_2, \sigma_2^2)$. Тогава пак от твърдение (16.3) следва, че величината

$$\sigma_2 \frac{X_1 - \mu_1}{\sigma_1} + \mu_2$$

има нормално разпределение с параметри μ_2 и σ_2, т.е. е равна на X_2. В този смисъл можем да напишем

$$X_2 = \sigma_2 \frac{X_1 - \mu_1}{\sigma_1} + \mu_2,$$

приемайки случайните величини с едно и също разпределение за равни. Такава свобода на преминаване от едно нормално разпределение към друго има важно значение в теорията и практиката.

Нека X и Y са случайни величини с известно съвместно разпределение. Тогава можем да разглеждаме величината $Z = g(X, Y)$. Ако и двете величини са дискретни, зададени както в таблица 16.4, то Z представлява дискретна случайна величина със стойности $z_{ij} = g(x_i, y_j)$, които се приема с вероятности $p_{ij} = P(X = x_i, Y = y_j)$, $1 \leq i \leq m$, $1 \leq j \leq n$. В частност по този начин се определя ***сумата*** $X + Y$ и ***произведението*** XY на двете случайни величини X и Y. Когато X и Y се предполагат независими, тяхното съвместно разпределение се определя изцяло от двете индивидуални разпределения, понеже в този случай $p_{ij} = P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j)$.

Пример 16.12. Да разгледаме независимите случайните величини X и Y, зададени чрез следните таблици

X	2	3	-1
P	0.2	0.3	0.5

Табл. 16.9.

Y	1	-1
P	0.6	0.4

Табл. 16.10.

и да намерим разпределението на тяхната сума и тяхното произведение. Пресмятанията ще подредим в таблица. Имаме

(i,j)	$x_i + x_j$	$x_i x_j$	p_{ij}
$(1,1)$	$2+1=3$	$2*1=2$	$0.2*0.6=0.12$
$(1,2)$	$2-1=1$	$2*(-1)=-2$	$0.2*0.4=0.08$
$(2,1)$	$3+1=4$	$3*1=3$	$0.3*0.6=0.18$
$(2,2)$	$3-1=2$	$3*(-1)=-3$	$0.3*0.4=0.12$
$(3,1)$	$-1+1=0$	$(-1)*1=-1$	$0.5*0.6=0.30$
$(3,2)$	$-1-1=-2$	$(-1)*(-1)=1$	$0.5*0.4=0.20$

Табл. 16.11.

За сумата получаваме следната таблица на разпределение

$X+Y$	-2	0	1	2	3	4
P	0.20	0.30	0.08	0.12	0.12	0.18

Табл. 16.12.

За произведението имаме

XY	-3	-2	-1	1	2	3
P	0.12	0.08	0.30	0.20	0.12	0.18

Табл. 16.13.

Когато X и Y са непрекъснати, то разпределението $F_Z(z) = P(Z < z)$ на величината $Z = g(X,Y)$ трябва да намерим въз основа на представянето

$$F_Z(z) = P(Z < z) = P(g(X,Y) < z) = \iint\limits_{g(x,y)<z} f(x,y)\,dxdy\,,$$

където $f(x,y)$ е съвместната плътност на X и Y. Тук двойният интеграл е над областта от R^2_{xy}, зададена от условието $g(x,y) < z$. В частност за сумата $Z = X + Y$ намираме

(16.16) $F_Z(z) = P(X + Y < z) = \iint\limits_{x+y<z} f(x,y)\,dxdy\,.$

Да извършим смяна на променливите $x = u$ и $y = v - u$, която има якобиан

$$J(u,v) = \begin{vmatrix} x_u & x_v \\ y_u & y_v \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ -1 & 1 \end{vmatrix} = 1\,.$$

Тогава от (16.16) намираме

$$F_Z(z) = \iint\limits_{v<z} f(u,v-u)\,dudv = \int_{-\infty}^{z}\left[\int_{-\infty}^{\infty} f(u,v-u)\,du\right]dv\,,$$

което означава, че Z има плътност $f_Z(z)$, която се получава посредством формулата

(16.17) $f_Z(z) = \int_{-\infty}^{\infty} f(u,z-u)\,du\,.$

Нека X и Y са независими. Тогава за тяхната съвместна плътност имаме $f(x,y) = f_X(x) f_Y(y)$, което заедно със (16.17) дава верността на

Теорема 16.1. Нека X и Y са две непрекъснати и независими случайни величини. Тогава тяхната сума $Z = X + Y$ представлява непрекъсната случайна величина с плътност

(16.18) $f_Z(z) = \int\limits_{-\infty}^{\infty} f_X(u) f_Y(z-u)du$,

където $f_X(x)$ и $f_Y(y)$ са плътностите на X и Y. □

Интегралът в дясната страна на (16.18) се нарича ***конволюция*** на функциите $f_X(x)$ и $f_Y(y)$ и се среща в много раздели на математическия анализ. Конволюцията на две функции е ***симетрична*** операция и интегралът (16.18) може да се запише във вида

$$f_Z(z) = f_X * f_Y(z) = \int\limits_{-\infty}^{\infty} f_X(z-u) f_Y(u)du .$$

Пример 16.13. Нека X_1 и X_2 са независими величини с равномерно разпределение в интервала $[0,1]$, $X \in U(0,1)$, $Y \in U(0,1)$. Всяка от тях има плътност $f_X(x) = f_Y(x) = f(x)$, с графика показана на рис. 16.6.

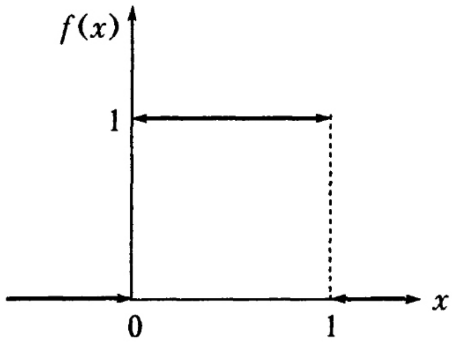

Рис. 16.6.

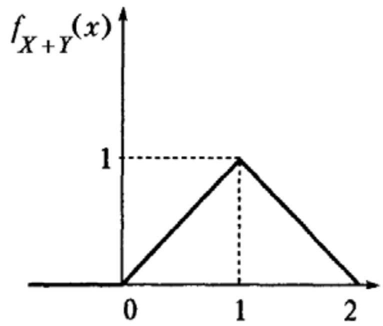

Рис. 16.7.

Съгласно (16.18) за плътността $f_{X+Y}(x)$ на сумата $X + Y$ намираме

$$f_{X+Y}(x) = \int\limits_{-\infty}^{\infty} f_X(t) f_Y(x-t)dt = \begin{cases} \int\limits_{0}^{x} dt = x \ \text{за} \ 0 \le x \le 1 \\ \int\limits_{1-x}^{1} dt = 2 - x \ \text{за} \ 1 \le t \le 2 \end{cases}$$

и $f_{X+Y}(x) = 0$ за $x \le 0$ и $x \ge 2$. Нейната графика е показана на рис. 16.7.

Могат да се разглеждат съвместни разпределения на ***повече от две*** случайни величини. Основните определения и свойства остават по същество идентични на случая за две величини. Всъщност важните за практиката заключения на теорията на вероятностите и математическата статистика се получават именно в случая когато се разглеждат много на брой случайни величини посредством тяхното съвместно разпределение.

В един от най-важните за статистиката случай, когато X_1, X_2, ..., X_n са ***независими по съвкупност непрекъснати случайни величини***, тяхната ***съвместна плътност*** $f(\mathbf{x})$ представлява функция на n променливи $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ и се явява произведение от индивидуалните плътности на отделните величини,

$$f(\mathbf{x}) = f_{X_1}(x_1) f_{X_2}(x_2) \cdots f_{X_n}(x_n).$$

Тази плътност описва разпределението на ***случайния вектор*** $(X_1, X_2, \ldots, X_n)$ по характерния за две величини начин.

Вярна е следната

Теорема 16.2. Нека X_1, X_2, ..., X_n са независими случайни величини, при което X_k е разпределена нормално с параметри μ_k и σ_k^2, $k = 1,2,\ldots,n$. Тогава тяхната линейна функция $X = \alpha_1 X_1 + \alpha_2 X_2 + \cdots + \alpha_n X_n + \beta$ е случайна величина, която има нормално разпределение с параметри

$\mu = \alpha_1\mu_1 + \alpha_2\mu_2 + \cdots + \alpha_n\mu_n + \beta$ и $\sigma^2 = \alpha_1^2\mu_1^2 + \alpha_2^2\mu_2^2 + \cdots + \alpha_n^2\mu_n^2$. □

4. Числови характеристики на случайни величини. В този раздел ще разгледаме някои основни числови характеристики на случайните величини. ***Математическо очакване (средно)*** на дискретната случайна величина X, $P(X = x_k) = p_k$, $k = 1,2,\ldots$, се нарича числото (когато съществува)

$$\mathbf{E}[X] = \sum_k x_k p_k .$$

Ако X е непрекъсната с плътност $f(x)$, то нейното математическо очакване се определя от интеграла (когато съществува)

$$\mathbf{E}[X] = \int_{-\infty}^{\infty} x f(x) dx .$$

Математическото очакване представлява най-важната мярка за ***централна тенденция*** за дадена случайна величина. Ако си поставим задачата да се намери някакво единствено число, което по най-добър начин да представя поведението на величината, то въпросното число се явява нейното математическо очакване. Математическото очакване се бележи още с $\mathbf{E}[X] = \mathbf{M}[X] = \mu_X$.

Ако X е непрекъсната случайна величина, а $g(x)$ е някаква непрекъсната функция, то намирането вида на разпределение на случайната величина $Y = g(X)$ в общия случай представлява трудна задача. Независимо от това, за математическото очакване на Y е валидна сравнително елементарна формула, която се съдържа в следното твърдение.

Твърдение 16.4. Нека X е непрекъсната случайна величина с плътност $f(x)$. Тогава за математическото очакване на случайната величина $Y = g(X)$ е изпълнено

$$\mathbf{E}[Y] = \int_{-\infty}^{\infty} g(x) f(x) dx . \ \square$$

Горното твърдение може да се обобщи за две и повече случайни величини. Ако $Z = g(X,Y)$, то

$$\mathbf{E}[Z] = \iint_{\mathrm{R}^2} g(x,y) f(x,y) dxdy ,$$

където $f(x,y)$ е съвместната плътност на X и Y.

Дисперсията $\mathbf{D}[X]$ на дискретната случайна величина X се нарича числото (когато съществува)

$$\mathbf{D}[X] = \mathbf{E}\left[(X - \mu_X)^2\right] = \sum_k (x_k - \mu_X)^2 p_k .$$

Ако X е непрекъсната с плътност $f(x)$, то нейната дисперсия се определя от интеграла (когато съществува)

$$\mathbf{D}[X] = \mathbf{E}\left[(X - \mu_X)^2\right] = \int_{-\infty}^{\infty} (x - \mu_X)^2 f(x) dx .$$

Числото $\sigma_X = \sigma[X] = \sqrt{\mathbf{D}[X]}$ се нарича ***стандартно отклонение*** на X. Очевидно $\mathbf{D}[X] = \sigma_X^2$, което задава σ_X^2 като друго означение за дисперсия.

Пример 16.14. Нека X е дискретната величина, определена от таблицата

X	5	15	25
P	0.2	0.4	0.4

Табл. 16.14.

Тогава по определение

$$\mathbf{E}[X] = 5*0.2 + 15*0.4 + 25*0.4 = 17,$$

$$\mathbf{D}[X] = (5-17)^2 * 0.2 + (15-17)^2 * 0.4 + (25-17)^2 * 0.4 = 17 = 56.$$

Пример 16.15. Нека X е непрекъсната величина, която е разпределена равномерно в интервала $[a,b]$. Тогава

$$\mathbf{E}[X] = \int_{-\infty}^{\infty} x f(x)dx = \int_a^b x dx = \frac{a+b}{2},$$

$$\mathbf{D}[X] = \int_{-\infty}^{\infty} (x - \mathbf{E}[X])^2 f(x)dx = \int_a^b \left(x - \frac{a+b}{2}\right)^2 dx = \frac{(b-a)^2}{12}.$$

Дисперсията представлява мярка за изменчивост на случайната величина, и се разглежда като мярка за отклонение от математическото очакване. Случайните величини с малка дисперсия са групирани около своето математическо очакване. От определенията се вижда, че дисперсията винаги е неотрицателно число. Ако дисперсията на една случайна величина е равна на нула, то тази величина е по същество константа, т.е. приема една единствена стойност с вероятност 1 (определението не изключва възможността тя да приема и някои други стойности с вероятност нула).

Твърдение 16.5. Нека X е разпределена нормално с параметри μ и σ. Тогава за очакването и дисперсията имаме $\mathbf{E}[X] = \mu$ и $\mathbf{D}[X] = \sigma^2$. □

С други думи нормалното разпределение се определя напълно от своето математическо очакване и от своята дисперсия.

Математическото очакване има следните основни свойства, чиито доказателства ще проведем само за случая на дискретни величини.

1) Ако X е константа, т.е. ако X приема една единствена стойност $X = C$ с вероятност едно (и възможно някои други стойности с вероятност нула), то очевидно $\mathbf{E}[X] = C$.

2) Ако λ е константа, то $\mathbf{E}[\lambda X] = \lambda \mathbf{E}[X]$. Имаме

$$\mathbf{E}[\lambda X] = \sum_k (\lambda x_k) p_k = \lambda \sum_k x_k p_k = \lambda \mathbf{E}[X].$$

3) За всеки две величини X и Y е изпълнено $\mathbf{E}[X+Y] = \mathbf{E}[X] + \mathbf{E}[Y]$. Имаме

$$\mathbf{E}[X+Y] = \sum_i \sum_j (x_i + y_j) p_{ij} = \sum_i \sum_j x_i p_{ij} + \sum_i \sum_j y_j p_{ij},$$

$$\mathbf{E}[X+Y] = \sum_i x_i \left(\sum_j p_{ij} \right) + \sum_j y_j \left(\sum_i p_{ij} \right) = \sum_i x_i p_{i\bullet} + \sum_j y_j p_{\bullet j} = \mathbf{E}[X] + \mathbf{E}[Y].$$

От тези три свойства непосредствено следва верността на

Теорема 16.3. Математическото очакване представлява ***линеен функционал***. Ако X_1, X_2, ..., X_n са някакви случайни величини, а λ_0, λ_1, λ_2, ..., λ_n са някакви константи, то за математическото очакване на линейна функция е изпълнено

$$\mathbf{E}[\lambda_0+\lambda_1X_1+\lambda_2X_2+\cdots+\lambda_nX_n]=\lambda_0+\lambda_1\mathbf{E}[X_1]+\lambda_2\mathbf{E}[X_2]+\cdots+\lambda_n\mathbf{E}[X_n].\ \square$$

Да отбележим, че в горното твърдение не се предполага независимост на величините.

Теорема 16.4. Ако X и Y са независими случайни величини, то математическото очакване на тяхното произведение е равно на произведението на двете математически очаквания, $\mathbf{E}[XY]=\mathbf{E}[X]\mathbf{E}[Y]$.

Доказателство. При независими дискретни величини имаме $p_{ij}=p_{i\bullet}p_{\bullet j}$, следователно

$$\mathbf{E}[XY]=\sum_i\sum_j(x_iy_j)p_{ij}=\sum_i\sum_j x_iy_jp_{i\bullet}p_{\bullet j}=\left(\sum_i x_ip_{i\bullet}\right)\left(\sum_j y_jp_{\bullet j}\right)=\mathbf{E}[X]\mathbf{E}[Y].\ \square$$

За случайната величина X могат да се разглеждат различни ***начални моменти*** $\mathbf{E}[X^k]$, $k=1,2,\ldots$, и ***централни моменти*** $\mathbf{E}[(X-\mu_X)^k]$, $k=1,2,\ldots$. Математическото очакване представлява начален момент от ред $k=1$, а дисперсията представлява централен момент от ред $k=2$. С помощта на централните моменти от трети и четвърти ред се определят други две важни числови характеристики,

$$skewness=\frac{E[(X-\mu_X)^3]}{\sigma_X^3},$$

която се нарича ***асиметрия*** и ***ексцес***, който се определя по формулата

$$kurtosis=\frac{E[(X-\mu_X)^4]}{\sigma_X^4}-3.$$

Ковариацията $\operatorname{cov}[X,Y]=\sigma_{XY}$ на двете случайни величини X и Y се определя като

$$\operatorname{cov}[X,Y]=\mathbf{E}[(X-\mu_X)(Y-\mu_Y)].$$

Ако X и Y са дискретни, то

$$\operatorname{cov}[X,Y]=\sum_i\sum_j(x_i-\mathbf{E}[X])(y_j-\mathbf{E}[Y])p_{ij}.$$

За непрекъснати величини X и Y имаме

$$\operatorname{cov}[X,Y]=\iint_{\mathrm{R}^2}(x\text{-}\mu_x)(y\text{-}\mu_y)f(x,y)\,dxdy,$$

където $f(x,y)$ е тяхната плътност. Очевидно $\operatorname{cov}[X,X]=\mathbf{D}[X]$.

Твърдение 16.6. Нека величините X и Y са независими. Тогава $\operatorname{cov}[X,Y]=0$.

Доказателство. По условие X и Y са независими, следователно $X-\mu_X$ и $Y-\mu_Y$ също са независими. Сега от теорема 16.4 намираме

$$\operatorname{cov}[X,Y]=\mathbf{E}[(X-\mu_X)(Y-\mu_Y)]=\mathbf{E}[X-\mu_X]\mathbf{E}[Y-\mu_Y]=0,$$

понеже

$$\mathbf{E}[X-\mu_X]=\mathbf{E}[X]-\mu_X=\mu_X-\mu_X=0.\ \square$$

В общия случай от $\operatorname{cov}[X,Y]=0$ не следва, че X и Y са независими.

Дисперсията може да се пресмята по следната формула

$$\mathbf{D}[X]=\mathbf{E}[X]-(\mathbf{E}[X])^2.$$

Наистина,

$$\mathbf{D}[X]=\mathbf{E}[(X-\mu_x)^2]=\mathbf{E}[X^2-2\mu X+\mu^2],$$

откъдето според основните свойства на очакването намираме

$$\mathbf{D}[X]=\mathbf{E}[X^2]-2\mu\mathbf{E}[X]+\mu^2=\mathbf{E}[X^2]-2\mu\mu+\mu^2=\mathbf{E}[X^2]-\mu^2.$$

По аналогичен начин за ковариацията се доказва формулата

(16.19) $\mathrm{cov}[X,Y]=\mathbf{E}[XY]-\mathbf{E}[X]\mathbf{E}[Y]$,

откъдето в частност се вижда, че ковариацията е симетрична и линейна по двата си аргумента, $\mathrm{cov}[\alpha X,\beta Y]=\alpha\beta\,\mathrm{cov}[X,Y]$.

Дисперсията има следните основни свойства.

1) Ако X е константа, то $\mathbf{D}[X]=0$.

2) Ако C е константа, то $\mathbf{D}[X+C]=\mathbf{D}[X]$. Имаме

$$\mathbf{D}[X+C]=\mathbf{E}\left[(X+C-\mathbf{E}[X+C])^2\right]=\mathbf{E}\left[(X+C-\mathbf{E}[X]-\mathbf{E}[C])^2\right],$$

$$\mathbf{D}[X+C]=\mathbf{E}\left[(X+C-\mathbf{E}[X]-C)^2\right]=\mathbf{E}\left[(X-\mathbf{E}[X])^2\right]=\mathbf{D}[X].$$

3) Ако λ е константа, то $\mathbf{D}[\lambda X]=\lambda^2\mathbf{D}[X]$. Имаме

$$\mathbf{D}[\lambda X]=\sum_k(\lambda x_k)^2p_k=\lambda^2\sum_k x_k^2p_k=\lambda^2\mathbf{D}[X].$$

4) За всеки две величини X и Y, $\mathbf{D}[X+Y]=\mathbf{D}[X]+\mathbf{D}[Y]+2\,\mathrm{cov}[X,Y]$. Имаме

$$\mathbf{D}[X+Y]=\mathbf{E}\left[(X+Y-\mathbf{E}[X+Y])^2\right]=\mathbf{E}\left[(X+Y-\mathbf{E}[X]-\mathbf{E}[Y])^2\right],$$

$$\mathbf{D}[X+Y]=\mathbf{E}\left[(X-\mathbf{E}[X]+Y-\mathbf{E}[Y])^2\right]=$$

$$=\mathbf{E}\left[(X-\mathbf{E}[X])^2+2(X-\mathbf{E}[X])(Y-\mathbf{E}[Y])+(Y-\mathbf{E}[Y])^2\right]$$

$$\mathbf{D}[X+Y]=\mathbf{E}\left[(X-\mathbf{E}[X])^2\right]+2\mathbf{E}[(X-\mathbf{E}[X])(Y-\mathbf{E}[Y])]+\mathbf{E}\left[(X-\mathbf{E}[X])^2\right],$$

$$\mathbf{D}[X+Y]=\mathbf{D}[X]+\mathbf{D}[Y]+2\,\mathrm{cov}[X,Y],$$

което доказва свойството. В частност дисперсия на сума от две независими величини е равна на сумата от техните дисперсии.

Разсъждавайки по същия начин получаваме верността на

Теорема 16.5. Нека X_1, X_2, ..., X_n са някакви случайни величини, а λ_1, λ_2, ..., λ_n са някакви константи. Тогава за дисперсията на линейната комбинация е изпълнено

$$\mathbf{D}[\lambda_1X_1+\lambda_2X_2+\cdots+\lambda_nX_n]=\lambda_1^2\mathbf{D}[X_1]+\lambda_2^2\mathbf{D}[X_2]+\cdots+\lambda_n^2\mathbf{D}[X_n]+ \\ +2\sum_{i<j}\lambda_i\lambda_j\,\mathrm{cov}[X_i,X_j] \qquad (19.20)$$

Ако величините са две по две независими, то формулата (19.20) приема вида

$$\mathbf{D}[\lambda_1X_1+\lambda_2X_2+\cdots+\lambda_nX_n]=\lambda_1^2\mathbf{D}[X_1]+\lambda_2^2\mathbf{D}[X_2]+\cdots+\lambda_n^2\mathbf{D}[X_n].\ \square$$

Пример 16.16. При независими X и Y имаме $\mathbf{D}[2X+3Y]=4\mathbf{D}[X]+9\mathbf{D}[Y]$ и $\mathbf{D}[X-2Y]=\mathbf{D}[X]+4\mathbf{D}[Y]$.

В частност дисперсията на разликата на две независими случайни величини е равна на сумата от техните дисперсии.

5. Условно математическо очакване. Въведените по-горе числови характеристики на случайните величини могат да бъдат отнесени и към техните условни разпределения. Нека X и Y са дискретни величини със съвестно разпределение $P(X=x_i,Y=y_j)=p_{ij}$, $1\le i\le m$, $1\le j\le n$ (таблица 16.5). Тогава съгласно (16.7) при всяко фиксирана стойност $Y=y_j$, условното разпределение на X се дава по формулата

$$P(X=x_i\mid Y=y_j)=\frac{p_{ij}}{p_{\bullet j}},\ 1\le i\le m,$$

която открива възможност за дефиниция на ***условното математическо очакване***

(16.21) $\mathbf{E}\left[X \mid Y = y_j\right] = \sum_{i=1}^{m} x_i P\left(X = x_i \mid Y = y_j\right) = \sum_{i=1}^{m} x_i \frac{p_{ij}}{p_{\bullet j}} = \frac{1}{p_{\bullet j}} \sum_{i=1}^{m} x_i p_{ij}$.

Аналогично за условното математическо очакване на Y при условие $X = x_i$ имаме

(16.22) $\mathbf{E}\left[Y \mid X = x_i\right] = \sum_{j=1}^{n} y_j P\left(Y = y_j \mid X = x_i\right) = \sum_{j=1}^{n} y_j \frac{p_{ij}}{p_{i\bullet}} = \frac{1}{p_{i\bullet}} \sum_{j=1}^{n} y_j p_{ij}$.

Ако X и Y са непрекъснати със съвместна функция на разпределение, то ***условното математическо очакване*** на X при условие $Y = y$ се определя по обичайния начин за непрекъснати величини посредством условната плътност $f(x \mid y)$ от (16.13) намираме

(16.23) $\mathbf{E}\left[X \mid Y = y\right] = \int_{-\infty}^{\infty} x f(x \mid y) dx = \int_{-\infty}^{\infty} x \frac{f(x,y)}{f_Y(y)} dx = \frac{\int_{-\infty}^{\infty} x f(x,y) dx}{\int_{-\infty}^{\infty} f(x,y) dx}$.

Аналогично за условното математическо очакване на Y при условие $X = x$ имаме

(16.24) $\mathbf{E}\left[Y \mid X = x\right] = \int_{-\infty}^{\infty} y f(y \mid x) dx = \int_{-\infty}^{\infty} y \frac{f(x,y)}{f_X(x)} dy = \frac{\int_{-\infty}^{\infty} y f(x,y) dy}{\int_{-\infty}^{\infty} f(x,y) dy}$.

Аналогично се определя и ***условната дисперсия***. Условните числови характеристики имат същите основни свойства както безусловните.

Функционалната зависимост $y = \mathbf{E}\left[Y \mid X = x\right]$, се нарича ***уравнение на регресия*** на случайната величина Y върху случайната величина X. Аналогично $x = \mathbf{E}\left[X \mid Y = y\right]$ представлява уравнението на регресия на X върху Y. Когато X и Y са непрекъснати, уравнението на регресия задава типична функция на променливата x. Ако обаче X и Y са дискретни, то функцията $y = \mathbf{E}\left[Y \mid X = x\right]$ ***е зададена само в стойностите на*** X .

Типична задача в математиката при разглеждане на две величини е дефиниране на функционална връзка между тях. Когато тези величини са случайни, тази задача се решава посредством уравненията на регресия (16.21-24), които предлагат описаните своеобразни функционални връзки. Регресионните уравнения са абстракция от вероятностния модел и значително опростяват неговото разбиране без да могат да го заместят по същество. Уравнението на регресия $y = \mathbf{E}\left[Y \mid X = x\right]$ дава възможност да правим ***прогнози*** за стойността на случайната величина Y въз основа поведението на случайната величина X и разбира се преди всичко въз основа на тяхното взаимно разпределение.

Уравненията на регресия дават значима информация само при наличието на зависимост между разглежданите величини. Ако X и Y са независими, то $p_{ij} = p_{i\bullet} p_{\bullet j}$ и от формулата (16.22) получаваме

$$\mathbf{E}\left[Y \mid X = x_i\right] = \frac{1}{p_{i\bullet}} \sum_{j=1}^{n} y_j p_{i\bullet} p_{\bullet j} = \sum_{j=1}^{n} y_j p_{\bullet j} = \mathbf{E}[Y],$$

което означава, че уравнението на регресия има тривиален вид

$$y = \mathbf{E}[Y] = const ,$$

което съответства на интуитивните представи, понеже в този случай разпределението на X не носи информация за разпределението на Y и обратно. В общия случай уравнението на регресия може да има твърде сложен вид.

Уравнението на ***линейна регресия*** $y = \alpha x + \beta$ на Y спрямо X се определя от условието за оптималност

(16.25) $\mathbf{E}\left[(Y - \alpha X - \beta)^2\right] \Rightarrow \min$.

Да положим $\overline{Y} = Y - \mu_Y$ и $\overline{X} = X - \mu_X$. Тогава $\mathbf{E}\left[\overline{Y}\right] = \mathbf{E}\left[\overline{X}\right] = 0$, а

$$\mathbf{E}\left[\overline{Y}^2\right] = \mathbf{E}\left[(Y - \mu_Y)^2\right] = \mathbf{D}[Y] = \sigma_Y^2 \text{ и } \mathbf{E}\left[\overline{X}^2\right] = \mathbf{E}\left[(X - \mu_X)^2\right] = \mathbf{D}[X] = \sigma_X^2 .$$

Сега (16.25) приема вида

$$\mathbf{E}\left[\left(\overline{Y} - \alpha\overline{X} - \gamma\right)^2\right] \Rightarrow \min,\ \gamma = \beta - \mu_Y + \alpha\mu_X,$$

което преобразуваме въз основа свойствата на математическото очакване

$$\varphi(\alpha,\gamma) = \mathbf{E}\left[\left(\overline{Y} - \alpha\overline{X} - \gamma\right)^2\right] = \mathbf{E}\left[\overline{Y}^2 + \alpha^2\overline{X}^2 + \gamma^2 - 2\alpha\overline{Y}\overline{X} - 2\gamma\overline{Y} + 2\alpha\gamma\overline{X}\right],$$

$$\varphi(\alpha,\gamma) = \mathbf{E}\left[\overline{Y}^2\right] + \alpha^2\mathbf{E}\left[\overline{X}^2\right] + \gamma^2 - 2\alpha\mathbf{E}\left[\overline{X}\overline{Y}\right] - 2\beta\mathbf{E}\left[\overline{Y}\right] + 2\alpha\beta\mathbf{E}\left[\overline{X}\right],$$

(16.26) $\varphi(\alpha,\gamma) = \sigma_Y^2 + \alpha^2\sigma_X^2 + \gamma^2 - 2\alpha\operatorname{cov}[X,Y]$.

За да намерим минимума на (16.26) пресмятаме и приравняваме на нула частните производни на $\varphi(\alpha,\gamma)$,

$$\varphi_\alpha(\alpha,\gamma) = 2\alpha\sigma_X^2 - 2\operatorname{cov}[X,Y] = 0,\ \varphi_\gamma(\alpha,\gamma) = 2\gamma = 0,$$

откъдето намираме

$$\alpha = \frac{\operatorname{cov}[X,Y]}{\sigma_X^2} \text{ и } \gamma = 0 .$$

Следователно $\beta = \mu_X - \alpha\mu_X$ и уравнението на линейна регресия има вида

(16.27) $y = \dfrac{\operatorname{cov}[X,Y]}{\sigma_X^2}x + \left(\mu_X - \dfrac{\operatorname{cov}[X,Y]}{\sigma_X^2}\mu_X\right)$.

Коефициентът на корелация ρ между X и Y се определя

$$\rho = \rho_{XY} = \frac{\operatorname{cov}[X,Y]}{\sigma_X\sigma_Y},\ \sigma_X > 0,\ \sigma_Y > 0 .$$

По този начин (16.27) приема вида

$$y = \rho\frac{\sigma_Y}{\sigma_X}x + \left(\mu_Y - \rho\frac{\sigma_Y}{\sigma_X}\mu_X\right),$$

което накратко се записва

(16.28) $\dfrac{y - \mu_Y}{\sigma_Y} = \rho\dfrac{x - \mu_X}{\sigma_X}$.

Аналогично, уравнението на линейна регресия на X спрямо Y се записва

(16.29) $\dfrac{x - \mu_X}{\sigma_X} = \rho\dfrac{y - \mu_Y}{\sigma_Y}$.

Теорема 16.6 (*Коши*). За ковариацията е в сила неравенството

$$\left|\operatorname{cov}[X,Y]\right| \le \sqrt{\mathbf{D}[X]}\sqrt{\mathbf{D}[Y]},$$

следователно за корелационния коефициент ρ винаги е изпълнено $-1 \le \rho \le 1$.

Доказателство. При дискретни величини X и Y имаме

$$\operatorname{cov}[X,Y] = \sum_{i,j}(x_i - \mathbf{E}[X])(y_j - \mathbf{E}[Y])p_{ij} = \sum_{i,j}\left[(x_i - \mathbf{E}[X])\sqrt{p_{ij}}\right]\left[(y_j - \mathbf{E}[Y])\sqrt{p_{ij}}\right].$$

От известното неравенство на Коши намираме

$$\operatorname{cov}[X,Y] \le \sqrt{\sum_{i,j}\left[(x_i - \mathbf{E}[X])\sqrt{p_{ij}}\right]^2}\sqrt{\sum_{i,j}\left[(y_j - \mathbf{E}[Y])\sqrt{p_{ij}}\right]^2},$$

$$\operatorname{cov}[X,Y] \le \sqrt{\sum_{i,j}(x_i - \mathbf{E}[X])^2 p_{ij}}\sqrt{\sum_{i,j}(y_j - \mathbf{E}[Y])^2 p_{ij}}\ ,$$

$$\operatorname{cov}[X,Y] \le \sqrt{\sum_{i}(x_i - \mathbf{E}[X])^2 p_{i\bullet}}\sqrt{\sum_{j}(y_j - \mathbf{E}[Y])^2 p_{\bullet j}} = \mathbf{D}[X]\mathbf{D}[Y].$$

Сега за корелационния коефициент получаваме

$$|\rho| = \frac{|\operatorname{cov}[X,Y]|}{\sqrt{\mathbf{D}[X]}\sqrt{\mathbf{D}[Y]}} \le 1.\ \square$$

Корелационният коефициент ρ може да приеме екстремални стойности $\rho = 1$ или $\rho = -1$ единствено когато между X и Y има ***линейна функционална зависимост***.

Пример 16.17. Да илюстрираме определените по-горе понятия върху следния пример на две дискретни случайни величини X и Y със следната таблица на взаимно разпределение

	$y_1 = 1$	$y_2 = 5$	$y_3 = 7$	
$x_1 = 2$	0.1	0.1	0.2	**0.4**
$x_2 = 3$	0.2	0.1	0.3	**0.6**
	0.3	**0.2**	**0.5**	

Табл. 16.15.

За условните очаквания на Y имаме

$$\mathbf{E}[Y \mid X = 2] = 1 * \frac{0.1}{0.4} + 5 * \frac{0.1}{0.4} + 7 * \frac{0.2}{0.4} = 5\ ,$$

$$\mathbf{E}[Y \mid X = 3] = 1 * \frac{0.2}{0.6} + 5 * \frac{0.1}{0.6} + 7 * \frac{0.3}{0.6} = 4.67\ .$$

Тук функцията на регресия $y = \varphi(x)$ е определена само в двете точки $x = x_1 = 2$ и $x = x_2 = 3$, при което приема стойности $\hat{y}_1 = \varphi(x_1) = \varphi(2) = 5$ и $\hat{y}_2 = \varphi(x_2) = \varphi(3) = 4.67$. Тази функция позволява да задаваме обективни прогнози за стойностите величината Y според стойностите на величината X.

За безусловните математическите очаквания и дисперсиите имаме

$$\mathbf{E}[X] = 2 * 0.4 + 3 * 0.6 = 2.6\ ,$$

$$\mathbf{D}[X] = (2 - 2.6)^2 * 0.4 + (3 - 2.6)^2 * 0.6 = 0.24\ ,\ \sigma_X = \sqrt{\mathbf{D}[X]} = 0.49\ ,$$

$$\mathbf{E}[Y] = 1 * 0.3 + 5 * 0.2 + 7 * 0.5 = 4.8\ ,$$

$$\mathbf{D}[Y] = (1 - 4.8)^2 * 0.3 + (5 - 4.8)^2 * 0.2 + (7 - 4.8)^2 * 0.5 = 6.76\ ,\ \sigma_Y = \sqrt{\mathbf{D}[Y]} = 2.6\ .$$

За ковариацията и коефициента на корелация имаме

$$\operatorname{cov}[X,Y] = (2 - 2.6)(1 - 4.8) * 0.1 + (2 - 2.6)(5 - 4.8) * 0.1 + (2 - 2.6)(7 - 4.8) * 0.2 +$$
$$+ (3 - 2.6)(1 - 4.8) * 0.2 + (3 - 2.6)(5 - 4.8) * 0.1 + (3 - 2.6)(7 - 4.8) * 0.3$$

$$\operatorname{cov}[X,Y] = -0.08\ ,$$

$$\rho = \frac{\operatorname{cov}[X,Y]}{\sigma_X \sigma_Y} = \frac{-0.08}{0.48 * 2.6} = -0.06\ .$$

В следващия раздел ще обсъдим пример с непрекъснато разпределени величини.

6. Многомерно нормално разпределение. Многомерните нормални разпределения имат ***водеща роля*** в цялата теория и представляват по същество основния апарат на приложната статистика. Да разгледаме отначало случая на две величини. Казва се, че случайните величини X и Y имат двумерно нормално разпределение, когато тяхната съвместна плътност се дава по формулата

(16.30) $$f(x,y)=\frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}e^{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_X)^2}{\sigma_X^2}-2\rho\frac{(x-\mu_X)}{\sigma_X}\frac{(y-\mu_Y)}{\sigma_Y}+\frac{(y-\mu_Y)^2}{\sigma_Y^2}\right]},$$

където μ_X, μ_Y, σ_X^2, σ_Y^2 и ρ са параметри, $\sigma_X>0$, $\sigma_Y>0$, $-1<\rho<1$. Две възможни форми на повърхнината $z=f(x,y)$ са дадени на следващите рисунки.

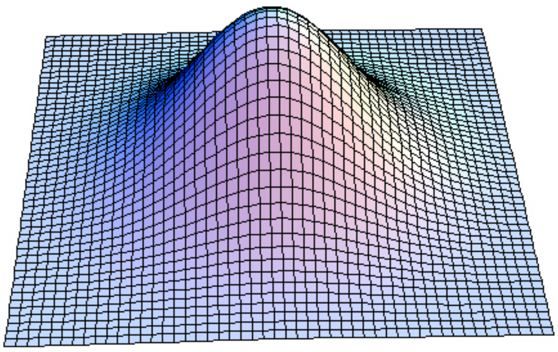

Рис. 16.8.

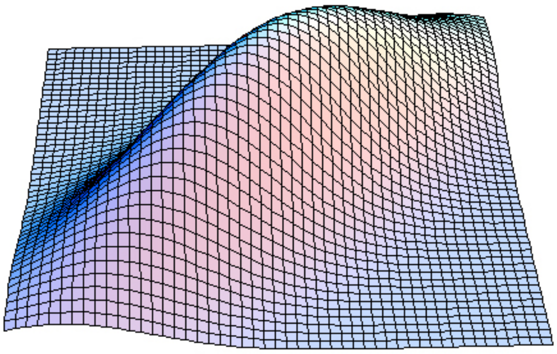

Рис. 16.9.

В първия случай корелационният коефициент е равен на нула, което означава фактически независимост на двете величини, а във втория е равен на 0.9, което е близо до неговата максимално възможна стойност и означава много силна правопропорционална зависимост.

С непосредствени макар и не много кратки пресмятания може да се покаже верността на следната

Теорема 16.7. Нека X и Y имат съвместно нормално разпределение с параметри μ_X, μ_Y, σ_X^2, σ_Y^2 и ρ, плътността $f(x,y)$ на което е определена от формулата (16.30). Тогава

1) Маргиналното разпределение $f_X(x)$ на X е нормално с параметри μ_X и σ_X^2, което означава

$$f_X(x)=\int_{-\infty}^{\infty}f(x,y)dy=\frac{1}{\sigma_X\sqrt{2\pi}}e^{-\frac{(x-\mu_X)^2}{2\sigma_X^2}}.$$

2) Маргиналното разпределение $f_Y(y)$ на Y е нормално с параметри μ_Y и σ_Y^2, което означава

$$f_Y(y)=\int_{-\infty}^{\infty}f(x,y)dx=\frac{1}{\sigma_Y\sqrt{2\pi}}e^{-\frac{(y-\mu_Y)^2}{2\sigma_Y^2}}.$$

3) Корелационният коефициент между X и Y е равен на ρ, т.е.

$$\frac{\iint_{R^2}(x-\mu_X)(y-\mu_Y)f(x,y)dxdy}{\sigma_X\sigma_Y}=\rho.$$

4) Условното разпределение на Y при условие $X=x$ има вида

$$f(y\,|\,x)=\frac{1}{\sigma_Y\sqrt{2\pi}\sqrt{1-\rho^2}}e^{-\frac{1}{2(1-\rho^2)}\left[\frac{y-\mu_Y}{\sigma_Y}-\rho\frac{x-\mu_X}{\sigma_X}\right]^2},$$

а за условното математическо очакване на Y имаме

$$\mathbf{E}[Y\,|\,X=x]=\mu_Y+\rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X),$$

с условна дисперсия $\mathbf{D}[Y \mid X] = \sigma_Y^2(1-\rho^2)$, което определя следното уравнение на регресия

$$y = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X).$$

5) Условното разпределение на X при условие $Y = y$ има вида

$$f(x \mid y) = \frac{1}{\sigma_X\sqrt{2\pi}\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[\frac{x-\mu_X}{\sigma_X} - \rho\frac{y-\mu_Y}{\sigma_Y}\right]^2},$$

а за условното математическо очакване на X имаме

$$\mathbf{E}[X \mid Y = y] = \mu_X + \rho\frac{\sigma_X}{\sigma_Y}(y - \mu_Y),$$

с условна дисперсия $\mathbf{D}[X \mid Y] = \sigma_X^2(1-\rho^2)$, което определя следното уравнение на регресия

$$x = \mu_X + \rho\frac{\sigma_X}{\sigma_Y}(y - \mu_Y). \quad \square$$

От теорема 16.7 следват няколко важни извода.

Двумерното нормално разпределение можем да разглеждаме като сглобено от две отделни нормални разпределения, при отчитане на зависимостта между величините. Особено забележително в този случай се явява обстоятелството, че тази зависимост се описва посредством едно единствено число – коефициентът на корелация ρ. Условието за независимост

$$f(x, y) = f_X(x) f_Y(y)$$

е изпълнено тогава и само тогава, когато $\rho = 0$. При $\rho = 0$ формата на $f(x, y)$ е симетрична, както е показано на рис. 16.8. При $\rho \neq 0$ тази симетрия се нарушава и плътността се групира по направление на регресионните прави, което е показано на рис. 16.9.

Уравненията на регресия в този случай съвпадат с уравненията на линейна регресия, както добре се вижда при съпоставяне на формулите (16.28) и (16.29) с формулите от точки 4 и 5 от теорема 16.7. Това е и едно от оправданията за това, че в общия случай се предпочита регресионната идея да се реализира посредством линейна регресия. Линейната регресия има в определен смисъл най-добри обобщаващи свойства макар, че от гледна точка на прогнозирането "вътре в данните" могат да се използват по-сложни регресионни модели.

Нека X_1, X_2, ..., X_n са случайни величини със средни $\mu_k = \mathbf{E}[X_k]$ и дисперсии $\mathbf{D}[X_k]$ и стандартни отклонения $\sigma_k = \sqrt{\mathbf{D}[X_k]}$, $k = 1, 2, \ldots, n$. Тогава можем да разгледаме ***вектора на средните*** $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_n)^T$ и ***ковариационната матрица***

$$\Sigma = \mathbf{cov}[\mathbf{X}] = \begin{pmatrix} \sigma_1^2 & \mathrm{cov}[X_1, X_2] & \cdots & \mathrm{cov}[X_1, X_n] \\ \mathrm{cov}[X_2, X_1] & \sigma_2^2 & \cdots & \mathrm{cov}[X_2, X_n] \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{cov}[X_n, X_1] & \mathrm{cov}[X_n, X_2] & \cdots & \sigma_n^2 \end{pmatrix}.$$

Ковариационната матрица представлява симетрична квадратна матрица от ред n. Когато между дадените величини няма линейна зависимост, тази матрица освен това е ***положително определена*** (в общия случай е неотрицателно определена). По този начин квадратичната форма

$$\left\langle (\mathbf{x}-\boldsymbol{\mu})\Sigma^{-1}, (\mathbf{x}-\boldsymbol{\mu})\right\rangle = (\mathbf{x}-\boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x}-\boldsymbol{\mu}),\ \mathbf{x} = (x_1, x_2, \ldots, x_n)^T,$$

също е положително определена. Корелационната матрица има вида

$$\boldsymbol{\rho} = \begin{pmatrix} 1 & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{21} & 1 & \cdots & \rho_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n1} & \rho_{n2} & \cdots & 1 \end{pmatrix},$$

където $\rho_{ij} = \rho_{X_i X_j}$ е корелационният коефициент на величините X_i и X_j, $1 \le i, j \le n$.

Нека сега Σ е произволна симетрична и положително определена квадратна матрица от ред n и нека $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_n)^T$ е някакъв вектор. Казва се, че случайните величини X_1, X_2, ..., X_n имат ***съвместно (многомерно) нормално разпределение***, когато имат съвместна плътност

$$(16.31)\ f(\mathbf{x}) = \frac{1}{\sqrt{\det(2\pi\Sigma)}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x}-\boldsymbol{\mu})}.$$

В този случай се пише $X \in N(\boldsymbol{\mu}, \Sigma)$, където $X = (X_1, X_2, \ldots, X_n)$. Може да се докаже, че всяка от величините X_k има нормално маргинално разпределение със средно μ_k, $1 \le k \le n$, при което зададената матрица Σ се явява точно ковариационната матрица за величините X_1, X_2, ..., X_n.

Многомерното нормално разпределение играе основна роля във вероятностните модели. Когато задачата се състои в определяне на взаимните зависимости между известен брой величини, то обикновено се предполага, че въпросните величини споделят взаимно многомерно нормално разпределение.

На пръв поглед формулата (16.31) изглежда сложна, но следващите разсъждения показват, че всъщност тази формула в известен смисъл е възможно най-простата имайки предвид целта която изпълнява. Ако искаме даден тип едномерно разпределение да бъде достатъчно гъвкаво, че да може да се приспособява за достатъчно разнообразни случайни величини от практиката, то напълно разумно е да предположим, че такова разпределение се управлява поне от два параметъра – един за отразяване на някаква централна тенденция и друг – за измерване разсейването около тази централна тенденция, т.е. параметрите е добре да бъдат поне два. Точно това е и случаят на едномерното нормално разпределение, което предлага допълнително удобство, формулата за неговата плътност да се записва просто именно посредством средното и дисперсията. Когато трябва да опишем връзката между две случайни величини, непременно ще използваме някакви допълнителни параметри. В случая на нормално разпределение този параметър е точно едни – коефициентът на корелация и в този смисъл ситуацията е максимално икономична. Ако анализираме внимателно формулата (16.31) ще видим, че там има следните параметри – n на брой параметри за индивидуалните средни, n на брой параметри за индивидуалните дисперсии и $\frac{n(n-1)}{2}$ на брой параметри за различните взаимни корелационни коефициенти. С други думи формулата (16.31) е максимално икономична без от това да пострада фатално нейната гъвкавост. От такава гледна точка формулата (16.31) трябва да изглежда вече съмнително елементарна и изисква повече аргументация за нейната универсална употреба. Тази аргументация ще бъде приведена в следващия раздел, където ще бъде формулирана така наречената централна гранична теорема, според която типичното разпределение на една случайна величина трябва да бъде нормално.

Такава концептуална простота съчетана с богата математическа структура използваща интензивно практически всичките основни раздели на математиката, определя теорията на вероятностите и математическата статистика като един от най-важните (ако не и най-важният) раздел на приложната математика, където резултатите получени от впечатляващо сложни математически преобразувания се прилагат непосредствено в многобройни практически задачи.

7. Гранични теореми. ***Закон за големите числа.*** Твърденията в този раздел имат мирогледно значение в теорията на вероятностите. Отначало ще докажем неравенството на Чебишов, което оценява отклонението на дадена случайна величина X от нейното математическо очакване посредством оценка вероятността на събития от вида $|X-\mathbf{E}[X]|\geq\varepsilon$, $\varepsilon>0$.

Теорема 16.8 ***(Неравенство на Чебишов).*** Нека X е случайна величина, $\mathbf{D}[X]>0$. Тогава за всяко $\varepsilon>0$ е изпълнено неравенството

(16.32) $P(|X-\mathbf{E}[X]|\geq\varepsilon)\leq\dfrac{\mathbf{D}[X]}{\varepsilon^2}$.

Доказателство. Нека X е дискретна случайна величина с разпределение $p_k=P(X=x_k)$, $k=1,2,\ldots,n$. Тук имаме вероятностно пространство с елементарни изходи $\omega_k=(X=x_k)$ и $P(\omega_k)=p_k$, $k=1,2,\ldots,n$. Да изберем някакво $\varepsilon>0$. Тогава събитието $(|X-\mathbf{E}[X]|\geq\varepsilon)$ включва онези елементарни изходи ω_k, за които е изпълнено $|x_k-\mathbf{E}[X]|\geq\varepsilon$, което означава, че

$$P(|X-\mathbf{E}[X]|\geq\varepsilon)=\sum_{|x_k-\mathbf{E}[X]|\geq\varepsilon}p_k .$$

За всичките събираеми в сумата отдясно е изпълнено $|x_k-\mathbf{E}[X]|\geq\varepsilon$, следователно можем да напишем

$$P(|X-\mathbf{E}[X]|\geq\varepsilon)\leq\sum_{|x_k-\mathbf{E}[X]|\geq\varepsilon}\left(\frac{|x_k-\mathbf{E}[X]|}{\varepsilon}\right)^2p_k=\frac{1}{\varepsilon^2}\sum_{|x_k-\mathbf{E}[X]|\geq\varepsilon}|x_k-\mathbf{E}[X]|^2p_k .$$

Добавяйки и (евентуално) липсващите събираеми в горната сума, получаваме

$$P(|X-\mathbf{E}[X]|\geq\varepsilon)\leq\frac{1}{\varepsilon^2}\sum_{k=1}^{n}|x_k-\mathbf{E}[X]|^2p_k=\frac{1}{\varepsilon^2}\mathbf{D}[X],$$

което доказва неравенството (16.32) в този случай.

Нека сега X е непрекъсната с плътност $f(x)$. Тогава от основното свойство на плътността имаме

$$P(|X-\mathbf{E}[X]|\geq\varepsilon)=\int_{|x-\mathbf{E}[X]|\geq\varepsilon}f(x)dx .$$

Тук интегрирането е върху множеството, определено от условието $|x-\mathbf{E}[X]|\geq\varepsilon$, следователно

$$P(|X-\mathbf{E}[X]|\geq\varepsilon)\leq\int_{|x-\mathbf{E}[X]|\geq\varepsilon}\left(\frac{|x-\mathbf{E}[X]|}{\varepsilon}\right)^2f(x)dx=\frac{1}{\varepsilon^2}\int_{|x-\mathbf{E}[X]|\geq\varepsilon}|x-\mathbf{E}[X]|^2f(x)dx .$$

Заменяйки интегрирането над множеството $|x-\mathbf{E}[X]|\geq\varepsilon$ с интегриране над цялата числова ос получаваме

$$P(|X-\mathbf{E}[X]|\geq\varepsilon)\leq\frac{1}{\varepsilon^2}\int_{-\infty}^{\infty}|x-\mathbf{E}[X]|^2f(x)dx=\frac{1}{\varepsilon^2}\mathbf{D}[X],$$

което доказва неравенството (16.32) и в този случай. □

Сходството между доказателствата на двата случая в теорема 16.8 не е случайно. В рамките на една по-сложна теория на интеграла – ***интеграл на Стилтес***, обикновеното интегриране и сумирането се обединяват в единни означения.

Неравенството на Чебишов е в основата на доказателството на ***закона за големите числа***.

Теорема 16.9 ***(закон за големите числа).*** Нека X_1, X_2, ..., X_n, ... са редица от две по две независими случайни величини с равномерно ограничени дисперсии, $\mathbf{D}[X_n] \le C$, $n = 1, 2, \ldots$. Тогава за всяко $\varepsilon > 0$ е изпълнено

(16.33) $$\lim_{n\to\infty} P\left(\left|\frac{X_1 + X_1 + \cdots + X_n}{n} - \frac{\mathbf{E}[X_1] + \mathbf{E}[X_2] + \cdots + \mathbf{E}[X_n]}{n}\right| \ge \varepsilon\right) = 0.$$

Доказателство. Да изберем едно $\varepsilon > 0$ и да положим

$$Y_n = \frac{X_1 + X_2 + \cdots + X_n}{n} = \frac{1}{n}X_1 + \frac{1}{n}X_2 + \cdots + \frac{1}{n}X_n, \ n = 1, 2, \ldots.$$

Тогава от свойствата на очакването и дисперсията намираме

$$\mathbf{E}[Y_n] = \frac{\mathbf{E}[X_1] + \mathbf{E}[X_2] + \cdots + \mathbf{E}[X_n]}{n},$$

$$\mathbf{D}[Y_n] = \frac{\mathbf{D}[X_1] + \mathbf{D}[X_2] + \cdots + \mathbf{D}[X_n]}{n^2},$$

следователно от условието за равномерна ограниченост на дисперсиите имаме

$$\mathbf{D}[Y_n] \le \frac{C + C + \cdots + C}{n^2} = \frac{C}{n}.$$

Съотношението (16.33) може да се препише във вида

(16.34) $$\lim_{n\to\infty} P\left(\left|Y_n - \mathbf{E}[Y_n]\right| \ge \varepsilon\right) = 0.$$

Сега неравенството на Чебишов заедно с оценката за $\mathbf{D}[Y_n]$ ни дават

$$P\left(\left|Y_n - \mathbf{E}[Y_n]\right| \ge \varepsilon\right) \le \frac{\mathbf{D}[Y_n]}{\varepsilon^2} \le \frac{C}{n\varepsilon^2},$$

което веднага доказва граничното съотношение (16.34), а следователно и (16.33). □

С други думи средното аритметично на сбор от случайни величини с нарастването на техния брой все по малко се отклонява от собственото си математическо очакване.

Пример 16.18. Да разгледаме схема на Бернули със серия от повторения с вероятност за успех при единичен опит p. Всеки пореден опит с номер k определя една случайна величина X_k приемаща две стойност $X_k = 1$ при успех и $X_k = 0$ в противен случай с вероятности $P(X_k = 1) = p$ и $P(X_k = 0) = 1 - p = q$. Таблицата на нейното разпределение има вида

X_k	0	1
P	p	q

Табл. 16.16.

Следвайки определенията пресмятаме

$$\mathbf{E}[X_k] = 1 * p + 0 * q = p \text{ и } \mathbf{D}[X_k] = (1-p)^2 * p + (0-p)^2 * q = pq.$$

Нека случайната величина U_n представя броя на успехите в схемата на Бернули, т.е. U_n има биномно разпределение с параметри n и p, $U_n \in B(n, p)$. Тогава

(16.35) $$U_n = X_1 + X_2 + \cdots + X_n,$$

където по условие величините X_1, X_2, ..., X_n са независими (условие за независимост на отделните елементарни опити).

Твърдение. 16.7. Нека $U_n \in B(n,p)$. Тогава $\mathbf{E}[U_n] = np$ и $\mathbf{D}[U_n] = npq$.

Доказателство. От (16.35) и от основните свойства на очакването и дисперсията получаваме

$$\mathbf{E}[U_n] = \mathbf{E}[X_1] + \mathbf{E}[X_2] + \cdots + \mathbf{E}[X_n] = p + p + \cdots + p = np,$$
$$\mathbf{D}[U_n] = \mathbf{D}[X_1] + \mathbf{D}[X_2] + \cdots + \mathbf{D}[X_n] = pq + pq + \cdots + pq = npq. \ \square$$

Сега от неравенството на Бернули приложено за величината $\frac{U_n}{n}$ представляваща относителния брой на успехите, която има очакване и дисперсия съответно

$$\mathbf{E}\left[\frac{U_n}{n}\right] = \frac{1}{n}\mathbf{E}[U_n] = \frac{1}{n}np = p \text{ и } \mathbf{D}\left[\frac{U_n}{n}\right] = \frac{1}{n^2}\mathbf{D}[U_n] = \frac{1}{n^2}npq = \frac{pq}{n},$$

получаваме оценката

$$(16.36)\quad P\left(\left|\frac{U_n}{n} - p\right| \geq \varepsilon\right) \leq \frac{pq}{n\varepsilon^2},$$

откъдето непосредствено следва верността на

Теорема 16.10 *(Бернули)*. За броя на успехите в схема с независими повторения на опитите е изпълнено

$$\lim_{n\to\infty} P\left(\left|\frac{U_n}{n} - p\right| \geq \varepsilon\right) = 0. \ \square$$

Теоремата на Бернули следва веднага и от закона за големите числа, понеже тук е изпълнено условието за равномерна ограниченост на дисперсиите на величините X_k.

Теоремата на Бернули има много ясно тълкуване. Емпиричната честота на сбъдване на дадено събитие при повторения на опита се отклонява малко от вероятността за поява на събитието, когато броят на повторенията нараства неограничено. Този резултат може да се съпостави с най-вероятния брой на успехи при схемата на Бернули, от който направихме заключение със сходен характер, че най-вероятният брой успехи се отклонява малко от стойността на p. Тези два резултата насочват недвусмислено към здравия фундамент, върху който почива теорията на вероятностите. Не бива обаче да се забравя, че теорията на вероятностите, както и всяка друга теория (математическа или не), се явява преди всичко плод на въображението на изследователите с всички плюсове и минуси от това.

Централна гранична теорема. Отначало ще дадем сведения за аналитичния апарат, необходим при формулировката и доказването на централната гранична теорема, макар че самото доказателство ще бъде само скицирано. 0-

Казва се, че ***редицата от случайните величини $\{X_n\}$ клони към случайната величина X по разпределение***, когато редицата от функциите на разпределение $\{F_n(x)\}$ клони към функцията на разпределение $F(x)$, във всяка точка $x \in \mathbb{R}$, в която $F(x)$ е непрекъсната. Тук $F_n(x)$ е функцията на разпределение на X_n, а $F(x)$ е функцията на разпределение на X. Пишем $X_n \xrightarrow{d} X$.

Ако X е непрекъсната, то нейната функция на разпределение също е непрекъсната. Следователно, ако $X_n \xrightarrow{d} X$, то

$$\lim_{n\to\infty} P(a < X_n < b) = \lim_{n\to\infty}[F_n(b) - F_n(a)] = F(b) - F(a).$$

В този случай $F(x)$ се нарича ***гранично разпределение*** за редицата $\{X_n\}$.

Съгласно формулата на Ойлер $e^{i\alpha}=\cos\alpha+i\sin\alpha$, където $i=\sqrt{-1}$ е имагинерната единица, при което за всяко реално α имаме $\left|e^{i\alpha}\right|=1$.

Нека X е случайна величина. Тогава $\varphi(t)=\mathbf{E}\left[e^{iXt}\right]$ се нарича ***характеристична функция*** на X. Характеристичната функция представлява комплекснозначна функция на реалната променлива $t\in\mathrm{R}$. Ако X е дискретна с разпределение $P(X=x_k)=p_k$, $k=1,2,\ldots$, то $\varphi(t)=\sum_k e^{ix_k t}$. Ако X е непрекъсната с плътност $f(x)$, то

(16.37) $\varphi(t)=\int_{-\infty}^{\infty}e^{ixt}f(x)dx=\int_{-\infty}^{\infty}\cos xt\, f(x)dx+i\int_{-\infty}^{\infty}\sin xt\, f(x)dx$.

Ако $f(x)$ е четна функция, то вторият интеграл в горната формула е нула, следователно в този случай характеристичната функция е реална. Очевидно $\varphi(0)=1$.

Пример 16.19. Да намерим характеристичната функция на нормалното стандартно разпределение. По определение

$$\varphi(t)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-\frac{x^2}{2}}e^{ixt}dx.$$

Тук е възможно диференциране под знака на интеграла. След това диференциране получаваме

$$\varphi'(t)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-\frac{x^2}{2}}(ix)e^{ixt}dx=\frac{i}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-\frac{x^2}{2}}xe^{ixt}dx=\frac{i}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-\frac{x^2}{2}}e^{ixt}d\frac{x^2}{2}.$$

Сега интегрираме по части,

$$\varphi'(t)=\frac{-i}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{ixt}de^{-\frac{x^2}{2}}=\frac{-i}{\sqrt{2\pi}}e^{ixt}e^{-\frac{x^2}{2}}\Bigg|_{-\infty}^{\infty}+\frac{i}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-\frac{x^2}{2}}de^{ixt},$$

$$\varphi'(t)=\frac{i}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-\frac{x^2}{2}}(it)e^{ixt}dx=\frac{i^2t}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-\frac{x^2}{2}}e^{ixt}dx=-t\varphi(t),$$

понеже

$$\frac{-i}{\sqrt{2\pi}}e^{ixt}e^{-\frac{x^2}{2}}\Bigg|_{-\infty}^{\infty}=\lim_{x\to\infty}\left[\frac{-i}{\sqrt{2\pi}}e^{ixt}e^{-\frac{x^2}{2}}\right]-\lim_{x\to-\infty}\left[\frac{-i}{\sqrt{2\pi}}e^{ixt}e^{-\frac{x^2}{2}}\right]=0-0=0.$$

Получихме, че характеристичната функция удовлетворява диференциалното уравнение $\varphi'(t)=-t\varphi(t)$, чиито решения имат вида $\varphi(t)=Ce^{-\frac{t^2}{2}}$ с някаква константа $C=\varphi(0)$. От друга страна винаги е изпълнено равенството $\varphi(0)=1$, откъдето окончателно за характеристичната функция на нормалното стандартно разпределение получаваме

(16.38) $\varphi(t)=e^{-\frac{t^2}{2}}$.

Твърдение 16.8. Нека X_k са независими по съвкупност случайни величини с характеристични функции $\varphi_k(t)$, $k=1,2,\ldots,n$, и нека $X=\alpha_0+\alpha_1X_1+\alpha_2X_2+\cdots+\alpha_nX_n$, с някакви константи α_k, $k=0,1,\ldots,n$. Тогава за характеристичната функция $\varphi(t)$ на величината X е изпълнено

$$\varphi(t)=e^{it\alpha_0}\varphi_1(\alpha_1t)\varphi_2(\alpha_2t)\cdots\varphi_n(\alpha_nt).$$

В частност характеристичната функция на сума от независими случайни величини се равнява на произведението на отделните характеристични функции.

Доказателство. Когато величините са независими, всякакви функции от тези величини представляват също независими случайни величини. От друга страна математическото очакване от произведение на независими величини е равно на произведението на отделните очаквания, следователно

$$\mathbf{E}\left[e^{it(\alpha_0+\alpha_1X_1+\alpha_2X_2+\cdots+\alpha_nX_n)}\right]=\mathbf{E}\left[e^{it\alpha_0}e^{it\alpha_1X_1}e^{it\alpha_2X_2}\cdots e^{it\alpha_nX_n}\right],$$

$$\mathbf{E}\left[e^{it(\alpha_0+\alpha_1X_1+\alpha_2X_2+\cdots+\alpha_nX_n)}\right]=e^{it\alpha_0}\mathbf{E}\left[e^{it\alpha_1X_1}\right]\mathbf{E}\left[e^{it\alpha_2X_2}\right]\cdots\mathbf{E}\left[e^{it\alpha_nX_n}\right],$$

$$\mathbf{E}\left[e^{it(\alpha_0+\alpha_1X_1+\alpha_2X_2+\cdots+\alpha_nX_n)}\right]=e^{it\alpha_0}\mathbf{E}\left[e^{i(t\alpha_1)X_1}\right]\mathbf{E}\left[e^{i(t\alpha_2)X_2}\right]\cdots\mathbf{E}\left[e^{i(t\alpha_n)X_n}\right],$$

$$\mathbf{E}\left[e^{it(\alpha_0+\alpha_1X_1+\alpha_2X_2+\cdots+\alpha_nX_n)}\right]=e^{it\alpha_0}\varphi_1(t\alpha_1)\varphi_2(t\alpha_2)\cdots\varphi_n(t\alpha_n),$$

което доказва твърдението. □

Нека $X\in N(\mu,\sigma^2)$ има нормално разпределение с параметри μ и σ^2. Тогава както знаем $X=\sigma Z+\mu$, където $Z\in N(0,1)$ има нормално стандартно разпределение. Сега от твърдение 16.8 и (16.7) за характеристичната функция на X намираме

(16.39) $\varphi(t)=e^{it\mu-\frac{\sigma^2t^2}{2}}$.

Следните две теореми оправдават въвеждането на понятието характеристична функция. Първата от тях ни учи, че характеристичната функция определя разпределението по единствен начин.

Теорема 16.11 ***(за единственост).*** Характеристичната функция определя разпределението еднозначно, т.е. на всяка характеристична функция съответства само едно разпределение. □

Освен това е валидна

Теорема 16.12 ***(за непрекъснатост).*** Нека $\{X_n\}$ е редица от случайни величини с функции на разпределение $\{F_n(x)\}$ и характеристични функции $\{\varphi_n(t)\}$. Тогава редицата $\{X_n\}$ клони по разпределение към случайната величина X с функция на разпределение $F(x)$ и характеристична функция $\varphi(t)$ тогава и само тогава, когато $\lim\limits_{n\to\infty}\varphi_n(t)=\varphi(t)$ при всяко $t\in\mathrm{R}$. □

Поточковата сходимост на характеристичните функции се оказа еквивалентна на сходимостта по разпределение.

Нека $X_k\in N(\mu_k,\sigma_k^2)$, $k=1,2,\ldots,n$, и нека $X=\alpha_0+\alpha_1X_1+\alpha_2X_2+\cdots+\alpha_nX_n$, с някакви константи α_k, $k=0,1,\ldots,n$. Тогава съгласно твърдение 16.2 и формулата (16.8) имаме

$$\varphi_X(t)=e^{it\alpha_0}e^{it\alpha_1\mu_1-\frac{\alpha_1^2\sigma_1^2t^2}{2}}e^{it\alpha_2\mu_2-\frac{\alpha_2^2\sigma_2^2t^2}{2}}\cdots e^{it\alpha_n\mu_n-\frac{\alpha_n^2\sigma_n^2t^2}{2}},$$

$$\varphi_X(t)=e^{it[\alpha_0+\alpha_1\mu_1+\alpha_2\mu_2+\cdots+\alpha_n\mu_n]-\frac{[\alpha_1^2\sigma_1^2+\alpha_2^2\sigma_2^2+\cdots+\alpha_n^2\sigma_n^2]t^2}{2}},$$

което представлява характеристична функция на нормално разпределение с параметри

$\mu=\alpha_0+\alpha_1\mu_1+\alpha_2\mu_2+\cdots+\alpha_n\mu_n$ и $\sigma^2=\alpha_1^2\sigma_1^2+\alpha_2^2\sigma_2^2+\cdots+\alpha_n^2\sigma_n^2$.

Сега от теоремата за единственост получаваме доказателство на известната теорема, че линейна комбинация на нормални разпределения представлява също нормално разпределение.

Характеристичната функция може да се използва за пресмятане числовите характеристики на величината. Да разгледаме за илюстрация случая на непрекъсната величина X с плътност $f(x)$. По определение имаме

$$\varphi(t)=\mathbf{E}\left[e^{itX}\right]=\int_{-\infty}^{\infty} e^{itx} f(x)dx,$$

откъдето след диференциране намираме

$$\varphi'(t)=\int_{-\infty}^{\infty}(ix)e^{itx} f(x)dx=i\int_{-\infty}^{\infty} xe^{itx} f(x)dx,$$

следователно

(16.40) $\varphi'(0)=i\int_{-\infty}^{\infty} xf(x)dx=i\mathbf{E}[X]$ или още $\mathbf{E}[X]=-i\varphi'(0)$.

Аналогично

$$\varphi''(t)=i\int_{-\infty}^{\infty} x(ix)e^{itx} f(x)dx=i^2\int_{-\infty}^{\infty} x^2e^{itx} f(x)dx,$$

следователно

(16.41) $\varphi''(0)=-\int_{-\infty}^{\infty} x^2 f(x)dx=-\mathbf{E}\left[X^2\right]$.

Сега имайки предвид, че $\mathbf{D}[X]=\mathbf{E}\left[X^2\right]-(\mathbf{E}[X])^2$, получаваме

$$\mathbf{D}[X]=-\varphi''(0)-(-i\varphi(0))^2=\varphi'^2(0)-\varphi''(0).$$

Нека $\{X_k\}$ е редица от независими случайни величини с очаквания $\mu_k=\mathbf{E}[X_k]$ и дисперсии $\sigma_k^2=\mathbf{D}[X_k]$ и да положим

$$Y_n=X_1+X_2+\cdots+X_n.$$

Тогава

$$\mathbf{E}[Y_n]=\mathbf{E}[X_1]+\mathbf{E}[X_2]+\cdots+\mathbf{E}[X_n]=\mu_1+\mu_2+\cdots+\mu_n,$$
$$\mathbf{D}[Y_n]=\mathbf{D}[X_1]+\mathbf{D}[X_2]+\cdots+\mathbf{D}[X_n]=\sigma_1^2+\sigma_2^2+\cdots+\sigma_n^2.$$

При централната гранична теорема се разглеждат величините

$$Z_n=\frac{Y_n-\mathbf{E}[Y_n]}{\sqrt{\mathbf{D}[Y_n]}}$$

и твърди, че (при определени условия) ***величините Z_n клонят по разпределение към нормално стандартно разпределение***, т.е.

(16.42) $\lim\limits_{n\to\infty} P\left(\dfrac{Y_n-\mathbf{E}[Y_n]}{\sqrt{\mathbf{D}[Y_n]}}<x\right)=\Phi(x)$,

където

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

е функцията на разпределение за нормалното стандартно разпределение. Формулата (16.11) открива възможност за приблизително пресмятане на вероятности както следва

$$P\left(a<\frac{Y_n-\mathbf{E}[Y_n]}{\sqrt{\mathbf{D}[Y_n]}}<b\right)\approx\Phi(b)-\Phi(a).$$

Ако величините $\{X_k\}$ са независими и еднакво разпределени с очаквания $\mathbf{E}[X_k]=\mu$ и дисперсии $\mathbf{D}[X_k]=\sigma^2$, то $\mathbf{E}[Y_n]=n\mu$ и $\mathbf{D}[Y_n]=n\sigma^2$. В този случай за величините Z_n имаме

$$Z_n = \frac{Y_n - n\mu}{\sigma\sqrt{n}},$$

а централната гранична теорема се формулира както следва.

Теорема 16.13 ***(централна гранична теорема за еднакво разпределени величини).*** Нека $\{X_n\}$ е редица от независими и еднакво разпределени случайни величини с очаквания $\mathbf{E}[X_n] = \mu$ и дисперсии $\mathbf{D}[X_n] = \sigma^2$. Тогава е изпълнено

$$(16.43)\quad \lim_{n\to\infty} P\left(\frac{Y_n - n\mu}{\sigma\sqrt{n}} < x\right) = \Phi(x).$$

Доказателство. Да запишем Z_n във вида

$$Z_n = \frac{Y_n - n\mu}{\sigma\sqrt{n}} = \frac{X_1 - \mu}{\sqrt{n}\sigma} + \frac{X_2 - \mu}{\sqrt{n}\sigma} + \cdots + \frac{X_n - \mu}{\sqrt{n}\sigma}.$$

Нека $\psi(t)$ е характеристичната функция на величините $\frac{X_n - \mu}{\sigma}$, която е една и съща за всички, понеже по условие те са разпределени еднакво. Имаме

$$\mathbf{E}\left[\frac{X_n - \mu}{\sigma}\right] = 0 \text{ и } \mathbf{D}\left[\frac{X_n - \mu}{\sigma}\right] = 1,\ n = 1,2,\ldots,$$

следователно съгласно (16.40) и (16.41) е изпълнено $\psi'(0) = 0$ и $\psi''(0) = -1$. В околност на точката $t = 0$ за функцията $\psi(t)$ е в сила представянето

$$\psi(t) = \psi(0) + \psi'(0)t + \frac{\psi''(t)}{2}t^2 + t^2\varepsilon(t),$$

където $\varepsilon(t) \to 0$ при $t \to 0$. По този начин за $\psi(t)$ намираме

$$\psi(t) = 1 - \frac{1}{2}t^2 + t^2\varepsilon(t).$$

Сега от твърдение 16.8 следва, че характеристичната функция $\hat{\psi}(t)$ на еднакво разпределените величини $\frac{X_n - \mu}{\sqrt{n}\sigma}$ има вида

$$\hat{\psi}(t) = 1 - \frac{1}{2n}t^2 + \frac{t^2}{n}\varepsilon\left(\frac{t}{\sqrt{n}}\right).$$

Нека $\varphi_n(t)$ е характеристичната функция за случайната величина Z_n. Тогава отново съгласно твърдение 16.8, функцията $\varphi_n(t)$ представлява произведението от характеристичните функции на величините $\frac{X_n - \mu}{\sqrt{n}\sigma}$, следователно

$$(16.44)\quad \varphi_n(t) = \prod_{k=1}^{n} \hat{\psi}(t) = \left[1 - \frac{1}{2n}t^2 + \frac{t^2}{n}\varepsilon\left(\frac{t}{\sqrt{n}}\right)\right]^n.$$

От правилата за пресмятане на граници знаем, че ако $\lim_{n\to\infty} \alpha_n = \alpha$, то

$$(16.45)\quad \lim_{n\to\infty}\left(1 + \frac{\alpha_n}{n}\right)^n = \lim_{n\to\infty}\left(1 + \frac{1}{(n/\alpha_n)}\right)^{(n/\alpha_n)\alpha_n} = e^{\alpha}.$$

Преобразуваме

$$1 - \frac{1}{2n}t^2 + \frac{t^2}{n}\varepsilon\left(\frac{t}{\sqrt{n}}\right) = 1 + \frac{1}{n}\left(-\frac{t^2}{2} + t^2\varepsilon\left(\frac{t}{\sqrt{n}}\right)\right) = 1 + \frac{\alpha_n}{n},$$

където

$$\alpha_n = -\frac{t^2}{2} + t^2\varepsilon\left(\frac{t}{\sqrt{n}}\right) \to -\frac{t^2}{2} \text{ при } n \to \infty .$$

Съгласно (16.44) и (16.45) намираме

$$\lim_{n\to\infty} \varphi_n(t) = \lim_{n\to\infty}\left[1 + \frac{\alpha_n}{n}\right]^n = e^{\lim\limits_{n\to\infty}\alpha_n} = e^{-\frac{t^2}{2}} .$$

Установихме, че редицата от характеристичните функции $\{\varphi_n(t)\}$ клони поточково към характеристичната функция на нормалното стандартно разпределение. Сега верността на централната гранична теорема следва непосредствено от теоремите за единственост и непрекъснатост. □

От централната гранична теорема веднага се получава верността на

Теорема 16.14 ***(Моавър-Лаплас).*** Нека $U_n \in B(n, p)$ е броят на успехите в схемата на Бернули. Тогава

(16.46) $\lim\limits_{n\to\infty} P\left(\frac{U_n - np}{\sqrt{npq}} < x\right) = \Phi(x)$,

където $\Phi(x)$ е функцията на нормално стандартно разпределение.

Доказателство. Както вече обсъдихме, при схемата на Бернули със серия от независими повторения с вероятност за успех при единичен опит p, всеки пореден опит с номер k определя случайна величина X_k приемаща стойност $X_k = 1$ при успех и $X_k = 0$ в противен случай с вероятности p и $q = 1 - p$. Освен това $\mu = \mathbf{E}[X_k] = p$ и $\sigma^2 = \mathbf{D}[X_k] = pq$. За броя на успехите U_n имаме $U_n = X_1 + X_2 + \cdots + X_n$. Сега става очевидно, че формулата (16.46) се получава като частен случай на (16.43). □

Формулата (16.46) се използва обикновено за приблизително пресмятане на вероятностите $P(a < U_n < b)$, понеже

$$P(a < U_n < b) = P\left(\frac{a - np}{\sqrt{npq}} < \frac{U_n - np}{\sqrt{npq}} < \frac{b - np}{\sqrt{npq}}\right),$$

откъдето съгласно (16.46) намираме

(16.47) $P(a < U_n < b) \approx \Phi\left(\frac{b - np}{\sqrt{npq}}\right) - \Phi\left(\frac{a - np}{\sqrt{npq}}\right)$.

Аналогично се получават формулите

(16.48) $P(U_n < b) \approx \Phi\left(\frac{b - np}{\sqrt{npq}}\right)$ и $P(a < U_n) \approx 1 - \Phi\left(\frac{a - np}{\sqrt{npq}}\right)$.

Пример 16.20. Да оценим вероятността при $n = 100$ броя хвърляне на монета, броят на показанията "герб" да бъде между 45 и 55. Тук $p = q = \frac{1}{2}$, $np = 50$, $npq = 25$. Търсим вероятността $P(45 < U_{100} < 55)$. Съгласно (16.47) имаме

$$P(45 < U_{100} < 55) \approx \Phi\left(\frac{55 - 50}{5}\right) - \Phi\left(\frac{45 - 50}{5}\right) = \Phi(1) - \Phi(-1),$$

$$P(45 < U_{100} < 55) \approx 0.841345 - 0.158655 = 0.682689 .$$

Да намерим сега вероятността броят на гербовете да бъде по-малък от 40. Съгласно (16.48) намираме

$$P(U_{100} < 40) \approx \Phi\left(\frac{40-50}{5}\right) = \Phi(-2) = 0.022750\,.$$

Поради симетрията в стандартното нормално разпределение, толкова е и вероятността броят на успехите да бъде по-голям от 60.

Заключението на централната гранична теорема с известни уговорки може да бъде прочетено по следния начин. Ако X_1, X_2, ..., X_n са независими случайни величини, то тяхната сума има приблизително нормално разпределение със средно и дисперсия

$$\mu = \mathbf{E}[X_1] + \mathbf{E}[X_2] + \cdots + \mathbf{E}[X_1] \text{ и } \sigma^2 = \mathbf{D}[X_1] + \mathbf{D}[X_2] + \cdots + \mathbf{D}[X_n]\,.$$

Тук по-важен е фактът, че сумата следва някакво нормално разпределение. При много величини от практиката, статистическите наблюдения показват, че те наистина се подчиняват на нормално разпределение със някакво конкретно математическо очакване и дисперсия. Това обстоятелство намира своето естествено обяснение именно посредством централната гранична теорема, понеже величините в динамично равновесие с околната среда е напълно разумно да се разглеждат като сума от голям брой случайни въздействия от необозрим характер. По този начин се оправдава предпочитането на нормалното разпределение като универсален модел в приложното статистическо моделиране. Разбира се има величини, за които вследствие тяхната специфика е ясно, че се подчиняват на други видове разпределения, например разпределение на Поасон, логонормално разпределение и т.н.