§15. Гранични теореми.

В **§13** получихме разпредението на сумата на няколко величини с еднакво показателно разпределение. При изследване на графиките на плътностите им с нарастване на броя на събираемите се забелязва изместване на графиките надясно и все по-голяма прилика с графиката на нормално разпределена величина. Сега ще разгледаме една важна теорема, наречена <u>*централна гранична теорема*</u>, която обяснява това явление и която доказва фундаменталната роля на величините с нормално разпределение.

1. Математическо очакване и дисперсия на независими величини с еднакво разпределение. Да разгледаме редицата $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ от независими случайни величини с ***еднакво разпределение***, които имат математическо очакване μ и средно квадратично отклонение σ. Образуваме величината $\xi = \xi_1 + \ldots + \xi_n + \ldots$ От свойствата на математическото очакване и дисперсията (виж **§§8, 10**) получаваме

$$E\xi = E(\xi_1 + \xi_2 + \ldots + \xi_n) = E\xi_1 + E\xi_2 + \ldots + E\xi_n = n\mu\,, \qquad (15.1)$$

$$D\xi = D(\xi_1 + \xi_2 + \ldots + \xi_n) = D\xi_1 + D\xi_2 + \ldots + D\xi_n = n\sigma^2\,, \qquad (15.2)$$

От тук следва, че за средно аритметичното $\bar{\xi} = \dfrac{\xi_1 + \ldots + \xi_n}{n}$ на величините $\xi_1, \xi_2, \ldots, \xi_n$ имаме

$$E\bar{\xi} = \frac{1}{n}E(\xi_1 + \ldots + \xi_n) = \frac{n\mu}{n} = \mu$$
$$D\bar{\xi} = \frac{1}{n^2}D(\xi_1 + \ldots + \xi_n) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}, \quad \sigma\xi = \frac{\sigma}{\sqrt{n}}, \qquad (15.3)$$

Пример 15.1. Автомат произвежда детайли, дължината ξ на които се контролира и която е подчинена на закона $N(3{,}5;\ 0{,}1)$. Избрани са случайно 100 детайла (случайна извадка с обем 100) и е намерено средното им аритметично $\bar{\xi}$. Да се определи какъв е законът на разпределение на случайната величина $\bar{\xi}$ и намерят математическото очакване и средно квадратичното й отклонение.

Решение. Въвеждаме величините $\xi_i \sim N(3{,}5;\ 0{,}1)$ - дължината на i-тия детайл. Тъй като величините са независими и имат еднакво разпределение, то по формула (15.3) имаме $E\bar{\xi} = 3{,}5$, $\sigma\bar{\xi} = 0{,}1/10 = 0{,}01$. Освен това, както гласи теорема **13.1**, величината $\bar{\xi}$ има нормално разпределение. Следователно, $\bar{\xi} \sim N(3{,}5;\ 0{,}01)$. ♦

Важен практически извод от примера е, че *средно аритметичното на величини с еднакво нормално разпределение е нормално разпределена със същото математическо очакване и с $\sqrt{n}$ пъти по-малко средно квадратично отклонение. Например, по-добре да се направят няколко измервания на една физична величина и да се вземе тяхното средноаритметично, тъй като с него подобрява се точността на измерването.*

2. Гранични теореми. От формули (15.1) и (15.2) за величината

$$Z = \frac{\xi_1 + \ldots + \xi_n - n\mu}{\sigma\sqrt{n}} \qquad (15.4)$$

имаме $EZ = \dfrac{1}{\sigma\sqrt{n}}(n\mu - n\mu) = 0$, т.е. Z е центрирана, а дисперсията й е равна на $DZ = \left(\dfrac{1}{\sigma\sqrt{n}}\right)^2 D(\xi_1 + \ldots + \xi_n - n\mu) = \dfrac{D\xi_1 + \ldots + D\xi_n - 0}{\sigma^2 n} = \dfrac{\sigma^2 n}{\sigma^2 n} = 1.$

Теоремата, която ще изкажем, гласи, че с нарастването на броя на независимите и еднакво разпределени величини $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ величината (15.4), не само има математическо очакване $EZ = 0$ и дисперсия $DZ = 1$, но законът на разпределението й се доближава до закона на нормално разпределена величина с същите математическо очакване и средно квадратично отклонение, т.е. $Z \sim N(0{,}1)$.

Теорема 15.1 (<u>***Централна гранична теорема***</u>)**.** *Ако $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ са независими случайни величини с еднакво разпределение, които имат математическо очакване μ и средно квадратично отклонение σ и $Z = \dfrac{\xi_1 + \ldots + \xi_n - n\mu}{\sigma\sqrt{n}}$, то за произволни $\alpha < \beta$*

$$\lim_{n\to\infty} P(\alpha < Z < \beta) = \frac{1}{\sqrt{2\pi}}\int_\alpha^\beta e^{-\frac{x^2}{2}}dx\,.$$

Забележка 15.1. Интегралът в последната формула е вероятността стандартна нормална величина да приеме стойност в интервала (α, β), следователно, Z *е величина с разпределение, близко до* $N(0{,}1)$.

Условията за прилагане на теорема **15.1** са твърде общи – иска се само *1) величините да са независими,*

2) да имат едно и също разпределение,

3) да имат математическо очакване и дисперсия.

А много често една случайна величина приема една или друга стойност именно под действието на множество други случайни и независими събития и затова се представя като сума от множество случайни величини с еднакви разпределения, приемащи малки стойности. Например:

- Налягането на газа върху стените на съда се дължи на бомбардирането му от молекулите на газа (на брой около 10^{22} за един литър), което зависи от скоростта на молекулите.

- Смущенията в много технически системи се дължат на голям брой малки независими случайни фактори и опитно е установено, че характерът на смущенията се описва с нормален закон за разпределение.

Като имаме предвид, че величината (15.4) може да се представи като $Z=\dfrac{\xi_1+\ldots+\xi_n)-n\mu}{\sigma\sqrt{n}}=\dfrac{\frac{1}{n}(\xi_1+\ldots+\xi_n)-EX}{\sigma/\sqrt{n}}=\dfrac{\bar{\xi}-\mu}{\sigma/\sqrt{n}}$, може да обобщим:

> **Забележка 15.2.** *Ако $\xi_1,\xi_2,\ldots,\xi_n$ са независими случайни величини с еднакво разпределение, $E\xi_i=\mu$, $D\xi_i=\sigma^2$, $i=1,\ldots,n$, то:*
>
> - *Сумата им $\xi=\xi_1+\ldots+\xi_n$ е случайна величина, за която $E\xi=n\mu$ и $\sigma\xi=\sqrt{D\xi}=\sqrt{n\sigma^2}=\sigma\sqrt{n}$. При големи стойности на n разпределението на ξ е близко до разпределението $N(n\mu,\sigma\sqrt{n})$;*
> - *Средноаритметичното им $\bar{\xi}=(\xi_1+\ldots+\xi_n)/n$ е случайна величина, за която $E\bar{\xi}=\mu$ и $\sigma\bar{\xi}=\sigma/\sqrt{n}$. При големи стойности на n разпределението на $\bar{\xi}$ е близко до разпределението $N\left(\mu,\sigma/\sqrt{n}\right)$.*

От гледна точка на практиката теоремата *премахва необходимостта от прецизен вероятностен модел* и дава възможност пресмятането на вероятност да се сведе до прилагане на формули и таблици за нормалното разпределение.

Пример 15.2. През последните години са регистрирани средно по 49 рекламации годишно. Каква е вероятността, че тази година броят рекламациите ще бъде: а) 30: б) между 47 и 50.

Решение. Величината ξ - брой на рекламациите за една година е поасоново разпределена с параметър $\lambda=49$ (среден брой рекламации за година). Тъй като $\lambda=49$ е голямо число (e^{-49} е много малко), се използва като приближение величина η с нормално разпределение, за която $E\eta=E\xi=\lambda=49$ и $\sigma\eta=\sigma\xi=\sqrt{\lambda}=\sqrt{49}=7$. Тъй като величината ξ е дискретна, а величнаta $\eta\sim N(49,7)$ е непрекъсната, то един от начините за пресмятане на вероятности на събития, касаещи ξ, е следният:

а) $P(\xi=30)=P(29{,}5\le\eta\le30{,}5)=F\left(\dfrac{30{,}5-49}{7}\right)-F\left(\dfrac{29{,}5-49}{7}\right)=$

$=1-F(2{,}64)-1+F(2{,}78)=0{,}9973-0{,}9958=0{,}0015$.

б) $P(47\le\xi\le50)=P(46{,}5\le\eta\le50{,}5)=F\left(\dfrac{50{,}5-49}{7}\right)-F\left(\dfrac{46{,}5-49}{7}\right)=$

$=F(0{,}21)-1+F(0{,}36)=0{,}5832-1+0{,}6406=0{,}2238$. ♦

Ще приложим теорема **15.1** в следния частен случай. Нека $P(A)=p$ във всеки опит (т.е. извършваме опити по схемата на Бернули) и нека ξ_i е индикаторът на случайното събитие A за i-тия опит. Както знаем, за тези величини имаме $E\xi_i=p$, $D\xi_i=qp$, а сумата им $\xi=\xi_1+\ldots+\xi_n$ приема стойност, равна на броя k на настъпване на A в n опита. При това $E\xi=np$, $\sigma\xi=\sqrt{npq}$ (виж **§11** за биномно разпределение). Като вземем пред вид забележка **15.2**, получаваме, че $\xi\sim N(np,\sqrt{npq})$, откъдето следва

> **Теорема 15.2** ***<u>(Централна (интегрална) гранична теорема на Моавър-Лаплас)</u>*** *Ако $\xi\sim B(n,p)$ е биномно разпределена величина, то при големи стойности на n вероятността ξ да приеме стойност в интервала $[a,b]$ се пресмята по формулата*
>
> $$P(a\le\xi\le b)\approx F\left(\frac{b-np}{\sqrt{npq}}\right)-F\left(\frac{a-np}{\sqrt{npq}}\right). \qquad (15.4)$$

Следователно, *при големи стойности на n разпределението на биномна случайна величина се доближава до разпределението на нормално разпределена величина със същото математическо очакване и дисперсия*. От тук следва и формулата

> $$P_n(k)=C_n^k p^k q^{n-k}\approx\frac{1}{\sqrt{2\pi npq}}e^{-\frac{(k-np)^2}{2npq}},$$
>
> *която се доказва в така наречената **<u>локална гранична теорема на Моавър-Лаплас.</u>***

Забележка 15.3. За биномното разпределение може да използваме две приближения*:*

> *при голямо n и малко p - разпределение на Поасон $Po(np)$.*
>
> *при голямо n и $np>5$ - нормално разпределение $N(np,\sqrt{npq})$.*

Пример 15.3. Нека ξ е биномно разпределена величина с параметри $p=0{,}5$ и $n=36$ (тук $p=0{,}5$ е сравнително голямо). Ще сравним изчисленията по точните и приближените формули.

Решение. а) За $P(\xi\le21)$ имаме:

по точната формула: $P(\xi\le21)=\sum\limits_{k=0}^{21}C_{36}^k p^k q^{36-k}=\sum\limits_{k=0}^{21}\binom{36}{k}(0{,}5)^{36}=0{,}8785$, по формула (15.4): $P(\xi\le 21)=F\left(\frac{21-36.0{,}5}{\sqrt{36.0{,}5.0{,}5}}\right)=0{,}8413$.

б) За $P(\xi=16)$ аналогично: $P(\xi=16)=C_{36}^{16}.0{,}5^{16}.0{,}5^{20}=0{,}1063$ и

$$P(\xi=16)=\frac{1}{\sqrt{2\pi.36.0{,}5.0{,}5}}e^{-\frac{(16-18)^2}{2.36.0{,}5.0{,}5}}=\frac{1}{\sqrt{56{,}5488}}e^{-\frac{2}{9}}=0{,}1065.$$

Пример 15.4. Селище A има 2500 жители. Всеки от тях, примерно, 6 пъти месечно пътува до селище B по случайни, независещи от останалите жители причини. Колко места трябва да има влакът от A до B, че той да е препълнен не повече от 1 на 1000 дни, ако се движи веднъж дневно.

Решение. Нека C={ жител пътува за селище B}. $P(C)=\frac{6}{30}=0{,}2$. Величината ξ - брой на пътуващите за B на ден е биномно разпределена с параметри $p=0{,}2$ и $n=2500$. Трябва да се определи броят N на местата във вагоните така, че $P(\xi>N)\le 0{,}001$. Прилагаме формулата на Моавър-Лаплас: $P(N<\xi)=1-F\left(\frac{N-2500.0{,}2}{\sqrt{2500.0{,}2.0{,}8}}\right)=1-F\left(\frac{N-500}{20}\right)\le 0{,}001$.

$\Rightarrow F\left(\frac{N-500}{20}\right)\ge 0{,}999$. Тъй като $F(4)=0{,}999$, то $\frac{N-500}{20}\ge 4\Rightarrow N\ge 580$. Следователно, необходимо е влакът да има 580 места.♦

Упражнения.

1. Да се сравнят намерените вероятности в задача 8, **§11**, с приближените им стойности, пресметнати по формулите на Моавър-Лаплас и на Поасон.

2. Да се намери вероятността от 100 хвърляния да се падне 50 пъти лицето на монетата.

3. Нека $\xi_1,...,\xi_{25}$ са 25 независими случайни величини с експоненциално разпределение с параметър $\lambda=2$. Разглеждаме величината $\bar{\xi}=(\xi_1+\ldots+\xi_{25})/25$. Да се изчислят $P(\bar{\xi}<1{,}25)$ и $P\left(|\bar{\xi}-2|<0{,}25\right)$.

4. Частици еднакво вероятно попадат в произволни точки от интервала $[-10{,}10]$. Каква е вероятността средного разстояние от началото до мястото на падане на 100 частици да е по-голямо от 1?

5. Малка частица скача с еднаква вероятност наляво и надясно по права линия. Всеки скок е с дължина 1 мм и се извършва за 1 милисек. а) След 10 секунди каква е вероятността частицата да е на разстояние от началното си положение по-голямо от 5 см. б) След колко време с вероятност 0,5 частицата се намира по-далеч от 5 см?

6. Случайната величина ξ приема стойностите 1, 2 и 3 с вероятности съответно 0,25, 0,5, 0,25. Каква е вероятността средно аритметичното от 72 независими измервания на ξ да е по-малко от 2,1?

§16. Неравенство на Чебишев. Закон за големите числа.

В началото бе отбелязан фактът, че при дълги серии от опити относителната честота на поява на случайно събитие се колебае около определено число, затова то се приема като мярка за това колко често настъпва събитието (виж **§3** - статистическа вероятност). Сега ще докажем това свойство на относителната честота като по този начин се убеждаваме, че избраните математически модели на случаен експеримент и на вероятност на събитие се съгласуват с действителността. Ще обосновем също, защо при достатъчно много опити може да твърдим, че средно аритметичното на получените стойности на която и да е случайна величина не се различава съществено от математическото й очакване.

Групата теореми, които ще разгледаме и които установяват и обясняват устойчивостта на средните стойности на голям клас случайни величини, се обединяват под името ***закон за големите числа***.

Теорема 16.1 (неравенство на Чебишев). *Ако случайната величина ξ има дисперсия и $\varepsilon>0$ е произволно число, то в сила е неравенството*

$$P(|\xi-E\xi|<\varepsilon)\ge 1-\frac{D\xi}{\varepsilon^2}. \tag{16.1}$$

Доказателство. Очевидно, неравенството на Чебишев може да се запише във вида

$$P(|\xi-E\xi|\ge\varepsilon)\le\frac{D\xi}{\varepsilon^2}. \tag{16.2}$$

Ще докажем последното неравенство за дискретна случайна величина като оценим дисперсията й. Съгласно формула (8.2) представяме $D\xi$ във вида ($p_i=p_\xi(x_i)$)

$$D\xi=\sum_i(x_i-E\xi)^2p_i=\sum_{x_i:|x_i-E\xi|<\varepsilon}(x_i-E\xi)^2p_i+\sum_{x_i:|x_i-E\xi|\ge\varepsilon}(x_i-E\xi)^2p_i.$$

Първата сума е неотрицателна, а за всяко събираемо на втората имаме $|x_i-E\xi|\ge\varepsilon$, откъдето $\sum_{x_i:|x_i-E\xi|\ge\varepsilon}(x_i-E\xi)^2p_i\ge\varepsilon^2\left(\sum_{x_i:|x_i-E\xi|\ge\varepsilon}p_i\right)$.

Но $\sum_{x_i:|x_i-E\xi|\ge\varepsilon}p_i=P(|\xi-E\xi|\ge\varepsilon)$. Следователно, $D\xi\ge\varepsilon^2P(|\xi-E\xi|\ge\varepsilon)$.

От тук $P(|\xi-E\xi|\ge\varepsilon)\le\frac{D\xi}{\varepsilon^2}\Rightarrow P(|\xi-E\xi|<\varepsilon)\ge 1-\frac{D\xi}{\varepsilon^2}$.♦

Пример 16.1. Вероятността за настъпване на събитието A във всеки от 1500 опита е 0,2. Нека ξ е броят на поява събитието.

Използувайки неравенството на Чебишев, да се оцени вероятността за това, че отклонението на величината ξ от математическото й очакване да бъде по-голямо от 40.

Решение. Величината ξ има биномно разпределение с математическо очакване $E\xi = 1500.0{,}2 = 300$ и дисперсия $D\xi = 240$.

Прилагаме (16.1) за $\varepsilon = 40$: $P(|\xi - 300| < 40) \geq 1 - \dfrac{240}{40^2} = 0{,}15$.♦

Пример 16.2. При подхвърляне на монета 1000 пъти наблюдаваме събитието A={поява на герб}. Да се оцени вероятността отклонението на относителната честота $w(A)$ от негова вероятност $P(A)$ да бъде по-малко от 0,1.

Решение. Ако ξ_i е индикатора на A в i-тия опит, то $\xi = \dfrac{\xi_1 + \ldots + \xi_n}{n} = \dfrac{m}{n}$ е относителната честота на A в n опита. За да приложим неравенството, пресмятаме

$$E\xi = \frac{1}{n}(E\xi_1 + \ldots + E\xi_n) = \frac{1}{n}.n.\frac{1}{2} = \frac{1}{2}, \quad D\xi = \frac{1}{n^2}(D\xi_1 + \ldots + D\xi_n) = \frac{1}{n}.\frac{1}{2}.\frac{1}{2} = \frac{1}{4n}.$$

При $n = 1000$ получаваме $P\left(\left|\dfrac{m}{1000} - \dfrac{1}{2}\right| < 0{,}1\right) \geq 1 - \dfrac{1}{4000.0{,}01} = \dfrac{39}{40}$.♦

Неравенство (16.1) оценява вероятността за отклоняване на случайната величина от математическото й очакване. Но не бива да се преувеличава приложението на (16.1) и (16.2), тъй като тези оценки са доста груби. Например, ако $\varepsilon^2 < D\xi$, то неравенство (16.1) дава единствено $P(|\xi - E\xi| < \varepsilon) > 0$, което е предварително известно. Главното приложение на (16.1) е, че с него се доказва теорема, наричана ***<u>закон за големите числа.</u>*** Тази теорема "действа" в много физични явления. Например, знаем, че да се предскаже поведението на една от частиците на газ е невъзможно, но сумарното поведение на всички частици е предсказуемо. Смисълът на теоремата е именно в това – да се предскаже стойността, която ще приеме една от много случайни величини $\xi_1, \xi_2, \ldots, \xi_n$ е невъзможно, но средноаритметичното от стойностите им е относително постоянно и това е изпълнено за широк клас случайни величини. Теоремата гласи:

Теорема 16.2. *Ако случайните величини* $\xi_1, \ \xi_2, \ldots, \xi_n, \ldots$ *са независими помежду си и дисперсиите им са ограничени от едно и също число* C, *т.е.* $D\xi_1 \leq C$,…, $D\xi_n \leq C$, *то за всяко* $\varepsilon > 0$

$$\lim_{n\to\infty} P\left(\left|\frac{\xi_1 + \xi_2 + \ldots + \xi_n}{n} - \frac{E\xi_1 + E\xi_2 + \ldots + E\xi_n}{n}\right| < \varepsilon\right) = 1.$$

Доказателство. Разглеждаме величината $\bar{\xi} = \dfrac{\xi_1 + \ldots + \xi_n}{n}$, за която пресмятаме математическото очакване и дисперсията

$$E\bar{\xi} = \frac{E\xi_1 + E\xi_2 + \ldots + E\xi_n}{n}, \quad D\bar{\xi} = \frac{1}{n^2}(D\xi_1 + \ldots + D\xi_n) \leq \frac{1}{n^2} n.C = \frac{C}{n},$$

Прилагаме неравенството на Чебишев (16.1), т.е. $P(|\bar{\xi} - E\bar{\xi}| < \varepsilon) \geq 1 - \dfrac{D\bar{\xi}}{\varepsilon^2}$.

Следователно, $P\left(\left|\dfrac{\xi_1 + \ldots + \xi_n}{n} - \dfrac{E\xi_1 + \ldots + E\xi_n}{n}\right| < \varepsilon\right) \geq 1 - \dfrac{D\bar{\xi}}{\varepsilon^2} \geq 1 - \dfrac{C}{n\varepsilon^2}$.

Равенството се получава при $n \to \infty$.♦

Ще приложим теоремата в случаите:

Теорема 16.3. (Чебишев) *Ако величините* $\xi_1, \ \xi_2, \ldots, \xi_n, \ldots$ *са еднакво разпределени, независими са помежду си и имат крайна дисперсия, то за всяко* $\varepsilon > 0$

$$\lim_{n\to\infty} P\left(\left|\frac{\xi_1 + \xi_2 + \ldots + \xi_n}{n} - E\xi\right| < \varepsilon\right) = 1,$$

където $E\xi$ *е математическото очакване на величините.*

Доказателството следва от това, че в случая $E\xi_1 = \ldots = E\xi_n = E\xi$.

Тази теорема служи за обосноваване на *<u>основния принцип на математическата статистика</u>, според който по случайна извадка на изучаваната съвкупност могат да се правят изводи за цялата съвкупност от изследвани обекти.*

Да приложим теорема **16.3** за схемата на Бернули. Нека в серия от n опита наблюдаваме броя на поява на събитието A с една и съща вероятност за настъпване $P(A) = p$. Нека ξ_k е индикаторът на събитието A за k-тия опит. Тогава $\dfrac{\xi_1 + \xi_2 + \ldots + \xi_n}{n} = \dfrac{m}{n}$, където m е броят на поява на събитието A в серията. От друга страна $\xi_1, \ \xi_2, \ldots, \xi_n$ са еднакво разпределени величини като $E\xi_k = p$ (виж **§11**). Като приложим теорема **16.3** получаваме твърдението

Теорема 16.4. (Бернули) *Нека* m *е броят на поява на* A *в* n *опита и* $P(A) = p$ *във всеки опит. Тогава, ако* $\varepsilon > 0$, *то* $\lim_{n\to\infty} P\left(\left|\dfrac{m}{n} - p\right| < \varepsilon\right) = 1$

По този начин, изхождайки от основните аксиоми и определения, чрез теорема **16.4** се доказва устойчивостта на относителната честота $w(A) = m/n$ на случайно събитие, което е практически установен факт.

ДВУМЕРНИ СЛУЧАЙНИ ВЕЛИЧИНИ

§17. Двумерна случайна величина - закон за разпределение. Закони за разпределение на съставните величини.

В теорията на вероятностите често се разглеждат съвместно няколко случайни величини. Например, при изследване на няколко признака на някакво физично явление, на въздействието на дадено лекарство върху различни функции на човешкия организъм и т.н. Едни от основните въпроси са:

- какви са законите на разпределение на тези величини;
- дали величините са зависими едни от други;
- какъв е характерът на зависимостта между тях.

Нека ξ_1, $\xi_2,\ldots,\xi_n$ са случайни величини. Величината $(\xi_1,\xi_2,\ldots,\xi_n)$ се нарича n-<u>мерна случайна величина</u>, а величините ξ_1, $\xi_2,\ldots,\xi_n$ - нейни съставни.

Геометрически n-мерната случайна величина се интерпретира като случайна точка в n-мерното пространство.

Ще разгледаме частния случай на двумерна случайна величина (ξ,η). Областта на възможните стойности на величината (ξ,η) е област от координатната равнина Oxy, а за да бъде величината (ξ,η) напълно определена, трябва да знаем как да намираме вероятността $P((\xi,\eta)\in D)$ за произволна двумерна област D.

<u>Закон за разпределение на вероятностите</u> на (ξ,η) (<u>съвместен закон на разпределение на съставните</u> ξ и η) се нарича всяко правило, с помощта на което може да се намери вероятността на произволно случайно събитие, отнасящо се за тази величина.

Дискретни двумерни величини.

Ако ξ и η са дискретни, възможните стойности на които са x_i и y_j, $i=1,2,\ldots,n,\ j=1,2,\ldots,m$, то величината (ξ,η) се нарича <u>дискретна</u>.

Законът на разпределението на дискретна двумерна величина най-често се задава с *<u>плътността на разпределението</u> – функция $p_{\xi,\eta}(x_i,y_j)$, дефинирана за възможните стойности на величината и установяваща връзка между възможните стойности (x_i,y_j) и вероятността p_{ij} за приемането им, а именно:*

$$p_{\xi,\eta}(x_i,y_j)=P(\xi=x_i\cap\eta=y_j)=p_{ij}\ .$$

Събитията $\{\xi=x_i\cap\eta=y_j\}$ образуват пълна група събития, затова

$$\sum_i\sum_j p_{ij}=1\ .$$

Обикновено възможните стойности и вероятностите p_{ij} се представят с таблица от вида (таблица на разпределението)

$\eta \backslash \xi$	x_1	$\ldots$	x_n
y_1	p_{11}	$\ldots$	p_{n1}
$\vdots$	$\vdots$		
y_m	p_{1m}	$\ldots$	p_{nm}

, $\sum_i\sum_j p_{ij}=1$.

Сумата $\sum_{j=1}^{m} p_{ij}$ е безусловната вероятност $p_\xi(x_i)$ на събитието $\{\xi=x_i\}$, тъй като не зависи от това каква стойност е приела величината η. По този начин, събирайки вероятностите p_{ij} по стълбове получаваме безусловния закон на разпределение на съставната случайна величина ξ. Ако съберем вероятностите по редове, получаваме безусловния закон на величината η. Получените таблици на разпределение имат вида

ξ	x_1	$\ldots$	x_n
P	$p_\xi(x_1)$	$\ldots$	$p_\xi(x_n)$

η	y_1	$\ldots$	y_m
P	$p_\eta(y_1)$	$\ldots$	$p_\eta(y_m)$

.

Пример 17.1. Зарче се подхвърля един път. Разглеждат се събитията A={пада се четно число} с индикатор случайната величина ξ и B={пада се число, кратно на 3 } с индикатор случайната величина η. Да се състави таблицата на разпределение на величината (ξ,η). Да се намерят законите на разпределение на величините ξ и η.

Решение. Например, $\xi=0$ и $\eta=0$ означава,че са се паднали 1 или 5 точки, следователно, $P(\xi=0\cap\eta=0)=\dfrac{2}{6}=\dfrac{1}{3}$. По аналогичен начин изчисляваме и другите вероятности. Така получаваме

$\eta \backslash \xi$	0	1	$P_\eta(y_i)$
0	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{2}{3}$
1	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{3}$
$P_\xi(x_i)$	$\frac{1}{2}$	$\frac{1}{2}$	1

.

В тази таблица, освен съвместното разпределение на двете величини ξ и η са получени и безусловните им закони – последният стълб (получен чрез събиране по редове) е съставен от вероятностите $P(\eta=y_j)=p_\eta(y_j)$, а последният ред (получен чрез събиране по стълбове) – от вероятностите $P(\xi=x_i)=p_\xi(x_i)$.♦

Непрекъснати двумерни величини.

Непрекъснати случайни величини (ξ,η) *са тези, за които съставните* ξ *и* η *са непрекъснати. Те се задават с помощта на функцията* *плътност* $p_{\xi,\eta}(x,y)$ *на съвместното разпределение, дефинирана в цялата равнина* Oxy *, която има свойствата .*

1) $p_{\xi,\eta}(x,y)\geq 0,\ (x,y)\in R^2$, 2) $\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} p_{\xi,\eta}(x,y)dxdy=1$;

3) $P((x,y)\in G)=\iint\limits_G p_{\xi,\eta}(x,y)dxdy$.

Плътностите на съставните ξ и η се намират по формулите:

$$p_\xi(x)=\int\limits_{-\infty}^{\infty} p_{\xi,\eta}(x,y)dy \text{ -плътност на } \xi,$$

$$p_\eta(x)=\int\limits_{-\infty}^{\infty} p_{\xi,\eta}(x,y)dx \text{ - плътност на } \eta,$$

И тук се дефинира

функция на разпределението : $F_{\xi,\eta}(x,y)=P(\xi\leq x\cap\eta\leq y),\ (x,y)\in R$.

За дискретни случайни величини фунцията на разпределение е прекъсната във всяка точка (x_i,y_j) , $i=1,\ldots,n,\ \ j=1,\ldots,m$.

За непрекъснати случайни величини предполагаме, че функцията $F_{\xi,\eta}(x,y)$ е непрекъсната. Имаме следните връзки между плътността и функцията на разпределение на двумерна случайна величина:

$$p_{\xi,\eta}(x,y)=\frac{\partial^2 F}{\partial x\partial y}, \qquad F_{\xi,\eta}(x,y)=\int\limits_{-\infty}^{x}\int\limits_{-\infty}^{y} p_{\xi,\eta}(x,y,)dxdy.$$

При зададена функция $F_{\xi,\eta}(x,y)$ функцията на разпределение на компонентата ξ на непрекъснатата величина (ξ,η), може да намерим чрез граничния преход $F_\xi(x)=\lim\limits_{y\to\infty}F_{\xi,\eta}(x,y)$.

Допълвайки с останалите свойства на функцията на разпределение на двумерна случайна величина, получаваме:

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

$$\lim_{x\to\infty}F_{\xi,\eta}(x,y)=F_\eta(y),\ \lim_{y\to\infty}F_{\xi,\eta}(x,y)=F_\xi(x),\ \lim_{x\to\infty\ y\to\infty}F_{\xi,\eta}(x,y)=1$$

$$P(a\leq\xi<b,\ c\leq\eta<d)\in D)=F_{\xi,\eta}(b,d)-F_{\xi,\eta}(a,d)-[F_{\xi,\eta}(b,c)-F_{\xi,\eta}(a,c)].$$

Може да обобщим:

Безусловни закони на съставните величини :

$p_\xi(x_i)=\sum\limits_j p_{ij}$ *и* $p_\eta(y_j)=\sum\limits_i p_{ij}$, *ако* (ξ,η) *е дискретна.*

$p_\xi(x)=\int\limits_{-\infty}^{\infty}p_{\xi,\eta}(x,y)dy$ *и* $p_\eta(y)=\int\limits_{-\infty}^{\infty}p_{\xi,\eta}(x,y)dx$, *ако* (ξ,η) *е непрекъсната*

От тези безусловни закони може да получим всички числени характеристики на ξ и η като $E\xi$ и $E\eta$, $D\xi$ и $D\eta$ и т.н. Например,

$$E\xi=\begin{cases}\sum\limits_i p_\xi(x_i)x_i, & \xi-\text{дискр.}\\ \int\limits_{-\infty}^{\infty}p_\xi(x)xdx, & \xi-\text{непр.}\end{cases},\qquad E\eta=\begin{cases}\sum\limits_i p_\eta(y_i)y_i, & \eta-\text{дискр.}\\ \int\limits_{-\infty}^{\infty}p_\eta(y)ydy, & \eta-\text{непр.}\end{cases},$$

Ще отбележим също и следното свойство на независимите случайни величини:

Величините ξ *и* η *са независими* $\Leftrightarrow$ $\begin{matrix}p_{\xi,\eta}(x,y)=p_\xi(x)p_\eta(y)\\ F_{\xi,\eta}(x,y)=F_\xi(x)F_\eta(y)\end{matrix}$

Пример 17.2. Да се намерят безусловните закони на величините, ξ и η, ако е дадена плътността на съвместното им разпределение $p_{\xi,\eta}(x,y)=6x^2y$, за $0<x<1,\ 0<y<1$ и $p_{\xi,\eta}(x,y)=0$ във всички останали случаи.

Решение. Съгласно формулите

$p_\xi(x)=\int\limits_{-\infty}^{\infty}p_{\xi,\eta}(x,y)dy=6\int\limits_0^1 x^2ydy=3x^2$ за $x\in(0,1)$ и $p_\xi(x)=0$ за $x\notin(0,1)$.

$p_\eta(y)=\int\limits_{-\infty}^{\infty}p_{\xi,\eta}(x,y)dx=6\int\limits_0^1 x^2ydx=2y$ за $y\in(0,1)$ и $p_\eta(y)=0$ за $y\notin(0,1)$.

Очевидно, $p_\xi(x)p_\eta(y)=3x^2.2y=6x^2y=p_{\xi,\eta}(x,y)$, откъдето получаваме, че величините ξ и η са независими.♦

Пример 17.3. Нека $p_{\xi,\eta}(x,y)=x+y$ за $(0<x<1,\ 0<y<1)$ и $p_{\xi,\eta}(x,y)=0$ в останалите случаи. Да се докаже, че величините ξ и η са зависими.

Решение. Безусловните закони на величините са

$$p_\xi(x)=\int_0^1(x+y)dy=x+\frac{1}{2},\ x\in(0,1),\ p_\eta(y)=\int_0^1(x+y)dx=y+\frac{1}{2},\ y\in(0,1).$$

Очевидно, $p_{\xi,\eta}(x,y)\neq p_\xi(x).p_\eta(y)$, т.е. величините са зависими.♦

Пример 17.4. За величината (ξ,η) със закон на разпределение

$\eta \backslash \xi$	-1	0	1
1	$0{,}15$	$0{,}3$	$0{,}35$
2	$0{,}05$	$0{,}05$	$0{,}1$

да се намерят $E\xi$, $E\eta$, $D\xi$ и $D\eta$.

Решение. Намираме безусловните закони (може да допълним таблицата на съвместното разпределение както в пример **17.1.**)

ξ	-1	0	1
P	$0{,}2$	$0{,}35$	$0{,}45$

η	1	2
P	$0{,}8$	$0{,}2$

откъдето: $E\xi=-0{,}2+0{,}45=0{,}25$, $E\eta=1.0{,}8+2.0{,}2=1{,}2$

$E\xi^2=(-1)^2.0{,}2+1^2.0{,}45=0{,}65$, $E\eta^2=1^2.0{,}8+2^2.0{,}2=1{,}6$

$D\xi=0{,}65-0{,}25^2=0{,}5875$ $D\eta=1{,}6-1{,}2^2=0{,}16$.♦

Сума на две случайни величини. Съвместното разпределение на дискретните величини ξ и η е удобно за намиране на сумата им $\xi+\eta$ (случаят на непрекъснати величини е разгледан в **§12**).

Пример 17.5. (продължение на пример **17.1**) Да се намери разпределението на величината $\zeta=\xi+\eta$.

Решение. Сумата на величините, дадени с таблицата от пример **17.1**, има възможни стойности са 0, 1 и 2, като

$P(\zeta=0)=P(\xi=0\cap\eta=0)=\frac{1}{3}$, $P(\xi=2)=P(\xi=1\cap\eta=1)=\frac{1}{6}$,

$P(\zeta=1)=P(\xi=0\cap\eta=1)+P(\xi=1\cap\eta=0)=\frac{1}{2}$ (събиране по диагоналите на матрицата от вероятности на двумерното разпределение).

Следователно, за величината ζ получаваме таблицата

ζ	0	1	2
P	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{6}$

.♦

Пример 17.6. За да приложим същия метод (чрез събиране на вероятностите, разположени диагонално) за сумата от величините, чието съвместно разпределение е дадено с таблица 17.2,

$\eta \backslash \xi$	-1	0	1
1	$0{,}15$	$0{,}3$	$0{,}35$
3	$0{,}05$	$0{,}05$	$0{,}1$

,

Табл.17.2.

$\eta \backslash \xi$	-1	0	1
1	$0{,}15$	$0{,}3$	$0{,}35$
2	0	0	0
3	$0{,}05$	$0{,}05$	$0{,}1$

Табл.17.3.

прибавяме стойност $\eta=2$ с нулеви вероятности (табл.17.3). По този начин при фиксирано k сумите x_1+y_k, $x_2+y_{k-1},\ldots,\ x_k+y_1$ имат една и съща стойност σ_k и затова $P(\zeta=\sigma_k)=p_{1,k}+p_{2,k-1}+\ldots+p_{k,1}$. Така чрез събиране по диагоналите получаваме таблицата

$\xi+\eta$	0	1	2	3	4
P	$0{,}15$	$0{,}3$	$0{,}4$	$0{,}05$	$0{,}1$

.♦

Упражнения.

1. Дадено е съвместното разпределение на величините ξ и η:

а)

$y_j \backslash x_i$	1	2	3	4	5	6
0	$1/8$	$1/16$	$1/16$	$1/16$	$1/16$	$1/8$
1	$1/12$	$1/12$	$1/12$	$1/12$	$1/12$	$1/12$

б)

$y_j \backslash x_i$	0	1	2
$1/2$	$5/48$	$10/48$	$5/48$
$1/3$	$1/27$	$4/27$	$4/27$
1	0	0	$1/4$

Да се намерят безусловните разпределения на ξ и η.

2. Имаме една правилна монета, едно правилно зарче и едно зарче, за което вероятността да се паднат 1 или 6 точки е $1/4$, а 2, 3, 4 или 5 точки е 1/8. Извършва се следният опит: Първо се подхвърля монетата. Ако се падне ези, се хвърля правилното зарче, ако се падне тура, хвърля се неправилното зарче. Да се докаже, че съвместното разпределение на величините ξ - брой на падналите точки и η - брой на падане на ези е дадено с таблицата в задача 1а).

3. В урна има 4 големи и 1 малка червени топки, 2 големи и 3 малки черни топки. Избираме една топка. Величината ξ приема стойност 1, ако изтеглената топка е червена и -1, ако топката е черна. Величината η приема стойност 1, ако изтеглената топка е голяма и -1, ако е малка. Да се намерят:

а) разпределенията на величините (ξ,η), ξ и η;

б) разпределенията на величините $\zeta_1=\xi+\eta$ и $\zeta_2=\xi\eta$.

4. Двумерната случайна величина е равномерно разпределена в квадрата $D=\{-6<x<6;-6<y<6\}$, т.е. $p_{\xi,\eta}(x,y)=a$ за $(x,y,)\in D$. Да се намери: а) константата a: б) да се докаже, че ξ и η са независими.

§18. Условни разпределения. Условно математическо очакване. Регресия и линия на регресията.

От съвместния закон на две случайни величини може да намерим безусловните закони на двете съставни величини (но не и обратното). Тези закони не дават информация дали има зависимост между двете величини. Затова при всяка възможна стойност на едната съставна се съставя закон за разпределение на другата, който се нарича условен.

Изхождайки от формулата за условна вероятност

$$P(A|B)=\frac{P(A\bigcap B)}{P(B)},$$

получаваме следните формули за условни вероятности на ξ и η на дискретна случайна величина (ξ,η)

$$P(\xi=x_i\,|\eta=y_j)=p_{\xi|y_j}(x_i\,|y_j)=\frac{p_{\xi,\eta}(x_i,y_j)}{p_\eta(y_j)},\ \ i=1,\ldots,n$$ *- плътност на величината ξ, ако $\eta=y_j$.*

$$P(\eta=y_j\,|\xi=x_i)=p_{\eta|x_i}(y_j\,|x_i)=\frac{p_{\xi,\eta}(x_i y_j)}{p_\xi(x_i)},\ \ j=1,\ldots,m$$ *- плътност на величината η при $\xi=x_i$.*

Ако (ξ,η) е непрекъсната, то условните плътности на компонентите ξ и η са $\boxed{p_{\xi|y}(x|y)=\frac{p_{\xi,\eta}(x,y)}{p_\eta(y)},\ p_{\eta|x}(y|x)=\frac{p_{\xi,\eta}(x,y)}{p_\xi(x)}}$.

За плътността $p_{\xi|y}(x|y)$ на величината $\xi|\eta=y$ x е променлива, а y е параметър, който показва при каква стойност на η е получено условното разпределение на ξ. Аналогично в $p_{\eta|x}(y|x)$ y е променлива, а x е параметър.

Случайните величини ξ и η се наричат ***независими***, ако условният закон на едната величина не зависи от това каква стойност е приела другата, т.е.

$p_{\xi|y_j}(x_i\,|y_j)=p_\xi(x_i),\ p_{\eta|x_i}(y_j\,|x_i)=p_\eta(y_j),$ *ако (ξ,η) е дискретна.*
$p_{\xi|y}(x|y)=p_\xi(x),\ p_{\eta|x}(y|x)=p_\eta(y),$ *ако (ξ,η) е непрекъсната*

Пример 18.1.. Да се намерят условните закони $\xi|\eta=0$ и $\eta|\xi=1$ и на величината от пример **17.1**. Зависими ли са съставните ξ и η.

Решение. $p_{\xi|0}(1|\eta=0)=p_{\xi|y_1}(1|0)=\frac{p_{\xi,\eta}(1,0)}{p_\eta(0)}=\frac{1/3}{2/3}=\frac{1}{2}$, и т.н. (всички числа от първия ред на съвместното разпределение се разделят на 2/3. По аналогичен начин пресмятаме и останалите условни вероятности. Така получаваме условните разпределения.

$x_i\|\eta=0$	0	1
$p_{\xi\|0}(x_i\|0)$	$\frac{1}{2}$	$\frac{1}{2}$

$y_i\|\xi=1$	0	1
$p_{\eta\|1}(y_i\|1)$	$\frac{2}{3}$	$\frac{1}{3}$

.

Очевидно, безусловните и условните закони съвпадат, т.е. величините са независими. ♦

Пример 18.2. За величината (ξ,η) със закон на разпределение

η \\ ξ	-1	0	1
1	0,15	0,3	0,35
2	0,05	0,05	0,1

(виж пример **17.4**) да се намери условният закон на ξ при условието $\eta=y_2=2$.

Решение. Безусловните закони са:

ξ	-1	0	1
P	0,2	0,35	0,45

и

η	1	2
P	0,8	0,2

.

За закона на ξ при условието $\eta=y_2=2$ разделяме втори ред на таблицата на безусловната вероятност $p_\eta(y_2)=0{,}2$ и получаваме

$\xi\|2$	-1	0	1
P	0,25	0,25	0,5

.

Очевидно, ξ и η са зависими.♦

Пример 18.3. Да се намерят условните закони на величините от пример **17.3.**

Решение. Плътността на условното разпределение на величината ξ е $p_{\xi|}y(x|y)=\frac{x+y}{y+0{,}5}$, където $0<x<1$ е променлива, а $0<y<1$ е параметър. Например, при $y=0{,}5$ плътността на величината $\xi|\eta=0{,}5$ е $p_{\xi|0{,}5}(x|0{,}5)=x+0{,}5$.

Аналогично, $p_{\eta|}x(y|x)=\frac{x+y}{x+0{,}5}$, където $x\in(0{,}1)$ - параметър, а $y\in(0{,}1)$ е променлива.♦

Пример 18.4. Нека съвместното разпределение на две зависими случайни величини се задава с плътността

$$p_{\xi,\eta}(x,y)=\frac{1}{2\pi\sqrt{(1-\rho^2)}}e^{-\frac{x^2-2\rho xy+y^2}{2(1-\rho^2)}}.$$

а) Да се докаже, че всяка от съставните величини има стандартно нормално разпределение. б) Да се намери условният закон на η при условие $\xi=x$, където x е произволно число.

Решение. а) Пресмятаме $p_\xi(x)=\int_{-\infty}^{\infty}p_{\xi,\eta}(x,y)dy=$

$$\frac{1}{2\pi\sqrt{1-\rho^2}}\int_{-\infty}^{\infty} e^{-\frac{x^2-2\rho x+y^2}{2(1-\rho^2)}}dy=\frac{1}{2\pi\sqrt{1-\rho^2}}\int_{-\infty}^{\infty} e^{-\frac{x^2(1-\rho^2)+(x^2\rho^2-2\rho x+y^2)}{2(1-\rho^2)}}dy=$$

$$=\frac{e^{-\frac{x^2(1-\rho^2)}{2(1-\rho^2)}}}{2\pi\sqrt{1-\rho^2}}\sqrt{1-\rho^2}\int_{-\infty}^{\infty} e^{-\frac{(y-\rho x)^2}{2\left(\sqrt{(1-\rho^2)}\right)^2}}d\frac{y-x\rho}{\sqrt{1-\rho^2}}=\frac{e^{-\frac{x^2}{2}}}{2\pi}\sqrt{2\pi}=\frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}.$$

Тук е използван *__интегралът на Поасон__* $\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}dx=\sqrt{2\pi}$.

По същия начин намираме $p_\eta(y)=\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}$.

б) По формулите за условно разпределение имаме

$$p_{\eta|x}(y|x)=\frac{p_{\xi,\eta}(x,y)}{p_\xi(x)}=$$

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

следователно, при фиксирано $\xi=x$ величината $\eta|x$ е нормално разпределена с математическо очакване $E(\eta|x)=x\rho$ и средно квадратично отклонение $\sigma=\sqrt{1-\rho^2}$, т.е. със закон $N(x\rho,\ \sqrt{1-\rho^2})$. Аналогично законът на величината $\xi|y$ е $N(\rho y,\ \sqrt{1-\rho^2})$. Величините са зависими.♦

Условно математическо очакване. Често в практиката е необходимо да се оцени характерът на зависимостта между две случайни величини. Например е важно да се намери връзката между количеството валежи и средния добив от единица площ, между параметрите на технологията и качеството на готовата продукция и т.н. Обикновено едната от двете величини, се счита за дадена, а се търси средната стойност на другата. Както знаем, “носител” на средната стойност на една случайна величина е нейното математическо очакване. Затова, ако $p_{\eta|x}(y|x)$ е плътността на величината $\eta|\xi=x$, то математическото очакване на величината η при зададена стойност $\xi=x$ се пресмята по формулата

$$E(\eta|x)=\int_{-\infty}^{\infty} p_{\eta|x}(y|x)\,y\,dy$$

и зависи от стойността x, която величината ξ е приела.

Например, $E(\eta|x)=\rho x$ от пример **18.4** е линейна функция на x.

Аналогично, $E(\xi|y)=\rho y$.

Следователно, ако величините ξ и η са зависими, то $E(\eta|x)$ е функция на x и *отразява характера на зависимостта на средната стойност на величината η от стойността, приета от величината ξ.*

Функцията $g(x)=E(\eta|x)$ се нарича __регресия__ на η от ξ, функцията $h(y)=E(\xi|y)$ се нарича __регресия__ на ξ от η. Графиките на тези функции се наричат съответно __линия на регресия__ на η от ξ и линия на регресия на ξ от η.

Очевидно, условното математическо очакване е неслучайна функция на случайна величина и затова също се явява случайна величина, означението за която е $E(\eta|x)$. Ще отбележим само свойството

$$E(E(\eta|x))=E(\eta)\ ,$$

(математическото очакване на величината $E(\eta|x)$ е равно на математическото очакване на величината η).

Упражнения.

1. За разпределенията от. задача 1, **§17** да се намерят условните закони на $\eta|\xi=k$, к=1,2, и $\xi|\eta=1$ и математическите им очаквания.
2. За величината с плътност $p_{\xi,\eta}(x,y)=6x^2y$ за $(x,y)\in D=(0<x<1, 0<y<1)$ да се намерят условните закони на съставните и регресиите на η от ξ и на ξ от η. Да се провери, че ξ и η са независими.
3. За величината с плътност $p_{\xi,\eta}(x,y)=x+y$ за $(x,y)\in D=(0<x<1, 0<y<1)$. Да се намерят условните закони на съставните и регресиите на η от ξ и на ξ от η и да се начертаят графиките им. Зависими ли са ξ и η?

§19. Ковариация. Коефициент на корелация.

__Ковариация__ $\mathrm{cov}(\xi,\eta)$ на две случайни величини ξ и η се нарича математическото очакване на величината $(\xi-E\xi)(\eta-E\eta)$, т.е.

$$\mathrm{cov}(\xi,\eta)=E[(\xi-E\xi)(\eta-E\eta)]\ . \qquad (19.1)$$

Ако $\mathrm{cov}(\xi,\eta)=0$, то величините ξ и η се наричат *__некорелирани.__*

Ако величините ξ и η са независими, то

$$\mathrm{cov}(\xi,\eta)=E[(\xi-E\xi)(\eta-E\eta)]=E[(\xi-E\xi)]E[(\eta-E\eta)]=(E\xi-E\xi)(E\eta-E\eta)=0$$

Следователно, ако величините са независими, то те са некорелирани. Обратното, обаче, не винаги е вярно, т.е. може $\mathrm{cov}(\xi,\eta)=0$, но величините да са зависими.

Ако $\operatorname{cov}(\xi,\eta)\neq 0$, то величините ξ и η се наричат <u>корелирани</u>.

Пример 19.1. Ще намерим ковариацията на двумерна случайна величина със съвместно разпределение

$y_i \backslash x_i$	-1	0	1
-1	0	0,25	0
0	0,25	0	0,25
1	0	0,25	0

Решение.
Тъй като $E\xi=E\eta=0$, то $\operatorname{cov}(\xi,\eta)=E[(\xi-0)(\eta-0)]=E(\xi\eta)$.

За всички двойки (x_i,y_j) или x_i, или y_j, или p_{ij} е равно на нула, откъдето

$$E(\xi\eta)=0 \text{ и } \operatorname{cov}(\xi,\eta)=E(\xi\eta)-E\xi E\eta=0,$$

т.е. величините са некорелирани. Въпреки това те са зависими - ако едната е приела стойност, различна от нула, то другата трябва да бъде равна нула.♦

За изчисляване на ковариацията вместо по формула (19.1) се използува формулата

$$\operatorname{cov}(\xi,\eta)=E(\xi.\eta)-E\xi E\eta, \tag{19.2}$$

където $E\xi$ и $E\eta$ са математическите очаквания на величините ξ и η

$$E(\xi\eta)=\begin{cases}\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{m}p_{\xi,\eta}(x_i,y_j)\,x_iy_j & (\xi,\eta) \text{ - дискретна}\\ \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}xy\,p_{\xi,\eta}(x,y)dxdy & (\xi,\eta) \text{ - непрекъсната}\end{cases}.$$

Доказателство. Тъй като $E\xi$ и $E\eta$ са константи, то

$$\operatorname{cov}(\xi,\eta)=E(\xi\eta-\xi.E\eta-\eta.E\xi+E\xi E\eta)=E(\xi\eta)-E\xi.E\eta-E\eta.E\xi+E\xi.E\eta,$$

Откъдето следва формула (19.2).♦

Други свойства на ковариацията са:

$$(\operatorname{cov}(\xi,\eta))^2\le D\xi D\eta.$$

Доказателство. Да използуваме, че за дисперсията на случайната величина $a\xi+b\eta$, където a и b са произволни реални числа, е изпълнено неравенството $D(a\xi+b\eta)\ge 0$. От свойствата на дисперсията имаме

$$\begin{aligned}D(a\xi+b\eta)&=E(a\xi+b\eta-(aE\xi+bE\eta))^2=E(a(\xi-E\xi)+b(\eta-E\eta))^2=\\&=a^2E(\xi-E\xi)^2+2abE((\xi-E\xi)(\eta-E\eta))+b^2E(\eta-E\eta))^2=\\&=a^2D\xi+b^2D\eta+2abE((\xi-E\xi)(\eta-E\eta))\end{aligned}$$

Следователно,

$$D(a\xi+b\eta)=a^2D\xi+b^2D\eta+2ab\operatorname{cov}(\xi,\eta). \tag{19.3}$$

Като разделим на b^2 и означим $t=\left(\frac{a}{b}\right)$, получаваме квадратното неравенство $D\xi.t^2+2\operatorname{cov}(\xi,\eta).t+D\eta\ge 0$. То е изпълнено за всяко t, ако дискриминантата му е неотрицателна, т.е. $\operatorname{cov}^2(\xi,\eta)-D\xi.D\eta\ge 0$.♦

Следствие19.1. От формула (19.3) при $a=1,\ b=1$ получаваме

*<u>формула за дисперсия на сума на две **зависими** величини</u>*

$$D(\xi+\eta)=D\xi+D\eta+2\operatorname{cov}(\xi,\eta) \tag{19.4}$$

Като характеристика на степента на зависимост на случайните величини се използува

<u>коефициентът на корелация</u> $\rho(\xi,\eta)=\dfrac{\operatorname{cov}(\xi,\eta)}{\sqrt{D\xi\,D\eta}}$.

<u>Свойства на коефициента на корелация:</u>

1) $|\rho(\xi,\eta)|\le 1$
2) $\rho(\xi,\eta)$ *е безразмерна величина*
3) *ако ξ и η са независими,* то $\rho(\xi,\eta)=0$
4) *ако $\eta=a\xi+b$, т.е. ξ и η са линейно зависими,* то $\rho(\xi,\eta)=1$.

В общия случай, ако $\rho(\xi,\eta)=0$, не може да се твърди, че ξ и η са независими, обаче за некорелирани случайни величини $\rho(\xi,\eta)\neq 0$.

Пример 19.2. Ще намерим ковариацията и коефициента на корелация на величината от пример **17.4.**

$$E\xi\eta=0{,}15.1.(-1)+0{,}3.0.1+0{,}35.1.1+0.05.1.2.+0{,}05.0.1+0{,}1.1.1=0{,}4,$$

$$\operatorname{cov}(\xi,\eta)=E\xi\eta-E\xi E\eta=0{,}4-0{,}5875.0{,}16=0{,}306,$$

$$\rho(\xi,\eta)=\frac{\operatorname{cov}(\xi,\eta)}{\sqrt{D\xi D\eta}}=\frac{0{,}1}{\sqrt{0{,}5875.0{,}16}}=0{,}326.\ ♦$$

Упражнения.

1. Дадена е $\operatorname{cov}(\xi,\eta)=3$. Да се намери $\operatorname{cov}(2\xi-5{,}4\eta+2)$.
2. Да се намерят ковариацията и коефициентът на корелация на величините, за които а) $p_{\xi,\eta}(x,y)=x+y$ за $(x,y)\in D=(0<x<1,0<y<1)$.
   б) $p_{\xi,\eta}(x,y)=6x^2y$ за $(x,y)\in D=(0<x<1,0<y<1)$
   в) $p_{\xi,\eta}(x,y)=15x^2y$ за $(x,y)\in D=(0<y<x<1)$
3. Да се намерят ковариациите и коефициентите на корелация на величините от задача **1**, **§17**.
4. Да се намерят математическите очаквания и дисперсиите на величините ζ_1 и ζ_2 от задача **3**, **§17**, и провери верността на формули (10.1) и (19.4).