

ЧИСЛОВИ ХАРАКТЕРИСТИКИ

Нека имаме множество от реални числа $x_1, x_2, x_3, \dots, x_n$. Абстрактно погледнато, те могат да бъдат произволни (формулите по-долу работят с всякакви реални числа), но за да има практическа полза от сметките, дадените числа трябва да бъдат еднородни, тоест операцията събиране да има смисъл. Например сборът “3 м + 2 кг” няма смисъл. С други думи, тези n величини трябва да имат една и съща мерна единица. Обикновено те са получени от n измервания на една случайна величина (например ръста или дохода на случайно избран човек). Когато множеството съдържа всички нейни стойности, то се нарича генерална съвкупност, а когато съдържа само някои от тях, се нарича извадка. Броят n на стойностите се нарича обем на генералната съвкупност или на извадката.

При статистическа обработка пресмятаме различни числови характеристики на данните. Тези характеристики се делят най-общо на мерки за средна стойност и мерки за отклонение. Тук ще разгледаме накратко най-често използваните от тях.

Мерки за средна стойност

- 1) Средно аритметично: $\bar{x} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n}$.
- 2) Средно геометрично: $(x_1 x_2 \dots x_n)^{1/n}$.
- 3) Средно хармонично: $\frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} + \dots + \frac{1}{x_n}}$.
- 4) Средно квадратично: $\sqrt{\frac{(x_1)^2 + (x_2)^2 + (x_3)^2 + \dots + (x_n)^2}{n}}$.
- 5) Средно степенно: $\left(\frac{(x_1)^p + (x_2)^p + (x_3)^p + \dots + (x_n)^p}{n} \right)^{1/p}$.

Някои от тези числови характеристики са дефинирани не за произволни реални числа. Например средното геометрично е дефинирано, когато n е нечетно число или всички данни са неотрицателни числа. При средното хармонично трябва да се избягва делението на нула. При средното степенно множеството от допустими стойности зависи от показателя p .

Неравенства между средните: Ако $x_1, x_2, x_3, \dots, x_n$ са неотрицателни числа, то ср. харм. $\leq$ ср. геом. $\leq$ ср. аритм. $\leq$ ср. кв.; равенства има само ако всички хиксове са равни, в противен случай неравенствата са строги. За неравенствата, които съдържат ср. харм., се изисква всички хиксове да са положителни. Изискването за неотрицателност не може да бъде освободено.

Неравенствата между средните могат да се обобщят с помощта на средното степенно. Разглеждано като функция на p , средното степенно е константа, когато хиксовете са равни, и строго растяща функция на p , когато поне два хикса са различни. Това твърдение е в сила, когато всички хиксове са неотрицателни (при $p > 0$), съответно положителни (при $p < 0$). То се явява обобщение на неравенствата от предишния абзац, защото:

- средното квадратично е средно степенно при $p = 2$;
- средното аритметично е средно степенно при $p = 1$;
- средното геометрично е средно степенно при $p \rightarrow 0$;
- средното хармонично е средно степенно при $p = -1$;
- максимумът и минимумът се получават съответно при $p \rightarrow +\infty$ и при $p \rightarrow -\infty$.

б) Средно претеглено: $\frac{\omega_1 x_1 + \omega_2 x_2 + \dots + \omega_n x_n}{\omega_1 + \omega_2 + \dots + \omega_n}$. Коефициентите $\omega_1, \omega_2, \omega_3, \dots, \omega_n$

се наричат тегла и обикновено са положителни числа. Често те се заменят с отношенията си към своя сбор; новият сбор се получава равен на единица.

Средното претеглено е подобно на средното аритметично, но не съвпада напълно с него. По-точно, средното аритметично е частен случай на средното претеглено — при равни тегла. В общия случай двете средни имат различна стойност, например

$$\frac{2}{5} \cdot 10 + \frac{3}{5} \cdot 20 = 16 \neq 15 = \frac{10 + 20}{2}.$$

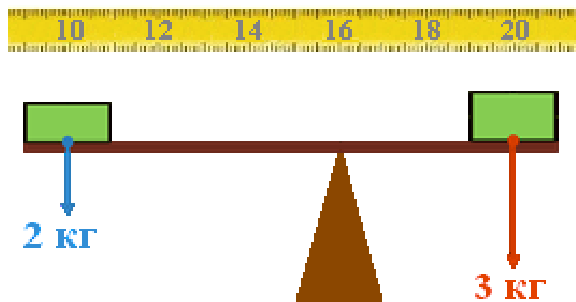
Обаче в определен смисъл (почти) всяко средно претеглено може да бъде изтълкувано като средно аритметично. Ако отхвърлим нетипичния (и неизползван на практика) случай, когато някои тегла са отрицателни, и ако отчетем, че нулевите тегла са излишни (те значат, че съответните хиксове не взимат участие), то остава да разгледаме само случая, когато всички тегла са положителни. Поначало се допуска теглата да бъдат ирационални числа, но в практиката обикновено работим със закръглени стойности, затова можем да приемем, че всички тегла са (положителни) рационални числа. От израза за средното претеглено се вижда, че той зависи от отношенията на теглата, а не от индивидуалните им стойности. А тъй като чрез умножение на няколко рационални числа по някой техен общ знаменател получаваме само цели числа, то можем без ограничение да предположим, че всички тегла са цели положителни числа. В този случай е лесно да изтълкуваме средното претеглено като средно аритметично. Например

$$\frac{2 \cdot 10 + 3 \cdot 20}{5} = 16 = \frac{10 + 10 + 20 + 20 + 20}{5}.$$

И тъй, ако изтълкуваме теглата (цели положителни числа) като бройки (абсолютни честоти) на съответните събираеми, тогава средното претеглено се превръща в средно аритметично.

Средното претеглено има различни приложения. Например в теорията на вероятностите така нареченото математическо очакване е именно средно претеглено: ролята на тегла играят вероятностите от съответното разпределение. Ако смесим разтвори на едно вещество с различни концентрации, концентрацията на сместа е средно претеглено на концентрациите на отделните разтвори, като за тегла служат масовите или обемните части на разтворите (в зависимост от мерната единица на концентрацията). При топлообмен между няколко тела температурите им се изравняват и крайната температура представлява средно претеглено на началните температури на телата, като теглата са топлинните капацитети на телата. Центърът на тежестта на система от материални точки е средно претеглено на точките, като математическите тегла съвпадат с физичните.

Пример: Ако в точката 10 от числовата ос сложим тяло с тегло 2 кг, а в точката 20 поставим тяло с тегло 3 кг, то центърът на тежестта ще бъде в точката 16, а не 15.



Неравенството между средното аритметично и средното геометрично се нарича още неравенство на Коши. То се използва при търсене на екстремуми на прости функции.

Пример: Търсим най-голямата стойност на израза $A = (x + 1)^5 \cdot (3 - x)^2$ при $0 < x < 3$. Можем да избегнем използването на производни, ако запишем израза така:

$$A = (x + 1)(x + 1)(x + 1)(x + 1)(x + 1)(3 - x)(3 - x).$$

За съжаление, сборът $3x + 11$ на множителите зависи от x . Ще поправим този недостатък, като умножим първите пет множителя по 2, а последните два множителя по 5:

$$800A = (2x + 2)(2x + 2)(2x + 2)(2x + 2)(2x + 2)(15 - 5x)(15 - 5x).$$

Изразите A и $800A$ едновременно достигат най-голяма стойност, така че е достатъчно да изследваме последния израз. Сборът 40 на неговите множители не зависи от x , ето защо средното им аритметично е константа (не е нужно да я пресмятаме). Средното геометрично, тоест седмият корен от израза $800A$, достига максимум (равен на средното аритметично) тогава и само тогава, когато всички множители са равни: $15 - 5x = 2x + 2$, т.е. $x = 13/7$. При тази стойност на x изразът A достига най-голямата си стойност: 248,6816...

Не е противоречие това, че при $x = 9$ се получава по-голяма стойност: $A = 3\,600\,000$. При $x = 9$ множителят $3 - x$ е отрицателен, а тогава неравенството на Коши не важи. В решението по-горе няма проблем, защото при $0 < x < 3$ множителите са неотрицателни.

Коя средна стойност да се използва, зависи от конкретната задача.

Примери:

— Ако в три поредни месеца сме получили съответно по 1000, 1500 и 1100 лева, то средният ни месечен доход е равен на $\frac{1000+1500+1100}{3} = 1200$ лева. Тук използваме средно аритметично, защото доходите са събираеми (те се събират със сумата, която имаме).

— Ако в поредни години цената на една стока е нараснала съответно 2, 3 и 4 пъти, то на година цената е растяла средно $\sqrt[3]{2 \cdot 3 \cdot 4} \approx 2,8845$ пъти. В този случай използваме средно геометрично, защото темповете на нарастване са множители.

— Ако лихвените проценти за три поредни години са съответно 5 %, 2 % и 1 %, средният годишен лихвен процент не е нито средно аритметично, нито средно геометрично на отделните проценти. Пресмятаме темповете на нарастване на капитала: 1,05; 1,02; 1,01. Понеже това са множители, средният темп на нарастване е тяхното средно геометрично: $\sqrt[3]{1,05 \cdot 1,02 \cdot 1,01} \approx 1,0265$. Следователно средният лихвен процент е около 2,65 %.

— Ако лека кола се движи еднакво време с различни скорости, напр. 30 км/ч. и 60 км/ч., то средната ѝ скорост е средно аритметично на отделните скорости, тоест тя е 45 км/ч. Ако обаче леката кола изминава един и същи път с различни скорости (30 км/ч. и 60 км/ч.), то средната ѝ скорост е средно хармонично на отделните скорости, тоест тя е 40 км/ч.

7) Медиана — средната по големина стойност при нечетен обем; ако той е четно число, медианата е средното аритметично на двете стойности, средни по големина.

Пример: Медианата на $\{3; 10; 15; 17; 18\}$ е 15, а на $\{4; 20; 22; 57\}$ е 21.

За разлика от средните по-горе, медианата не е чувствителна към екстремални стойности. Ако в примера заменим 18 с 1800, ср. аритм. ще нарасне много, а медианата ще остане 15.

8) Мода се нарича най-често срещаната стойност (може да съществуват няколко такива). Например модата на множеството $\{8; 12; 8; 8; 23\}$ е равна на 8.

Мерки за отклонение

1) Размах — най-голямата минус най-малката стойност.

2) Максимално отклонение — половината от размаха.

3) Средно абсолютно отклонение: $\frac{|x_1 - \bar{x}| + |x_2 - \bar{x}| + \dots + |x_n - \bar{x}|}{n}$.

4) Средноквадратично (стандартно) отклонение:

$$\sigma = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}} \text{ за генерална съвкупност;}$$

$$s = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n - 1}} \text{ за извадка.}$$

5) Дисперсия — квадратът на стандартното отклонение (т.е. изразът под корена).

Всяка от тези мерки може да бъде представена относително — спрямо някое средно.

Пример: Нека имаме следните числови данни: 10 ; 14 ; 20 ; 25 ; 90. Следователно размахът е равен на $90 - 10 = 80$, а максималното отклонение е половината, тоест 40; средното абсолютно отклонение е 23,28; $\sigma^2 = 872,96$; $\sigma \approx 29,55$; $s^2 = 1091,2$; $s \approx 33,03$. Относителното представяне на s спрямо средното аритметично 31,8 е около 104 %. Обикновено този процент е много по-малък от 100 %. Нетипичната му стойност тук се дължи на големите разлики между данните: най-голямото число е девет пъти по-голямо от най-малкото.

Размахът и максималното отклонение имат този недостатък, че зависят единствено от най-голямата и най-малката стойност, поради което не отразяват разпръскването на другите стойности около средната.

Средното абсолютно отклонение зависи от всички данни и е симетричен израз, освен това се изчислява по-просто от средноквадратичното отклонение. Но ако искаме да намерим екстремум на някое от тях (разглеждани като функции), трудно ще приложим операцията диференциране към средното абсолютно отклонение: абсолютната стойност си сменя аналитичния израз, преминавайки през нулата, затова трябва да разглеждаме твърде много случаи, за да разкрием всички модули.

Тази пречка не възниква при работа със средноквадратичното отклонение, ето защо именно то се предпочита в статистиката (откъдето идва и названието му “стандартно”).

Дисперсията се използва също толкова често, особено когато коренуването е излишно (например при сравняване на отклоненията на две числови съвкупности). Наложително е да използваме коренуване, ако искаме мярка за отклонение със същата мерна единица като обработваните числови данни. Ако например $x_1, x_2, x_3, \dots, x_n$ са измерени в метри, то мерната единица на стандартното отклонение ще бъде също метър, а мерната единица на дисперсията ще бъде квадратен метър.

При многократно измерване на един непроменящ се обект (напр. дължината на стая) е възможно да получим леко различаващи се стойности. В такъв случай мерките за средни са приближения на истинската, но неизвестна стойност, а мерките за отклонение са оценки на грешката на измерването.

СЛУЧАЙНИ ВЕЛИЧИНИ

Вероятност

Вероятност на събитие се нарича частта на изходите, благоприятни за това събитие, спрямо броя на всички изходи, когато всички изходи са равновероятни:

$$P(A) = \frac{k_A}{k_\Omega} \quad (\text{класическа вероятност}).$$

Тук $P(A)$ е вероятността на случайното събитие A ;

k_A е броят на изходите, благоприятни за събитието A ;

k_Ω е броят на всички изходи — благоприятни и неблагоприятни.

Благоприятни са тези изходи, при които събитието се сбъдва (настъпва).

Понеже $0 \leq k_A \leq k_\Omega$, то $0 \leq P(A) \leq 1$. В проценти: $0\% \leq P(A) \leq 100\%$.

Събитията се разглеждат като множества от елементарни изходи.

Невъзможно събитие: $\emptyset$ (не съдържа изходи).

Сигурно събитие: Ω (съдържа всички изходи).

Върху събития може да се извършват операции от теорията на множествата:

— допълнение: съответства на отрицание (“не”), т.е. допълването събитие не настъпва;

— сечение: съответства на конюнкция (“и”), т.е. настъпват всички събития;

— обединение: съответства на дизюнкция (“или”), т.е. настъпва поне едно събитие.

Условна вероятност $P(A|B)$ е вероятността да настъпи събитието A , при условие че е настъпило събитието B . Изисква се предположението да е осъществимо, т.е. $B \neq \emptyset$.

Свойства на вероятността:

- 1) $P(\emptyset) = 0$.
- 2) $P(\Omega) = 1 = 100\%$.
- 3) $P(\overline{A}) = 1 - P(A)$.
- 4) $P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1) P(A_2) \dots P(A_n)$, ако събитията са независими в съвкупност (т.е. сбъдването на кое да е от тях не влияе върху останалите събития).
 $P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1) P(A_2|A_1) P(A_3|A_1 \cap A_2) \dots P(A_n|A_1 \cap \dots \cap A_{n-1})$ при зависими събития (т.е. за всяко събитие се взимат предвид предишните събития).
- 5) $P(A_1 \cup A_2 \cup \dots \cup A_n) = P(A_1) + P(A_2) + \dots + P(A_n)$, ако събитията са две по две несъвместими (т.е. съвместното сбъдване на кои да е две от тях е невъзможно). Ако пак има съвместими събития, прилага се принципът за включване и изключване:

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = \sum_{\substack{X \\ \emptyset \neq X \subseteq \{1; 2; \dots; n\}}} \left((-1)^{|X|+1} P\left(\bigcap_{k \in X} A_k\right) \right).$$

Непрекъснати случайни величини

Техните стойности запълват цял интервал. Има различни непрекъснати разпределения, обаче най-важно за теорията и практиката се явява т. нар. нормално разпределение $N(\mu; \sigma)$. Казваме, че случайната величина ζ има **нормално разпределение** $N(\mu; \sigma)$, ако нейната функция на разпределение $F(t)$ се задава със следната формула:

$$F(t) := P(\zeta < t) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \int_{-\infty}^t \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx.$$

Свойства на нормалното разпределение $N(\mu; \sigma)$

$$P(\zeta = t) = 0;$$

$$P(\zeta < t) = P(\zeta \leq t) = F(t);$$

$$P(\zeta > t) = P(\zeta \geq t) = 1 - F(t);$$

$$P(a < \zeta < b) = P(a \leq \zeta < b) = P(a < \zeta \leq b) = P(a \leq \zeta \leq b) = F(b) - F(a);$$

$$P(\zeta < \mu) = P(\zeta \leq \mu) = F(\mu) = 0,5;$$

$$P(\zeta > \mu) = P(\zeta \geq \mu) = 1 - F(\mu) = 1 - 0,5 = 0,5;$$

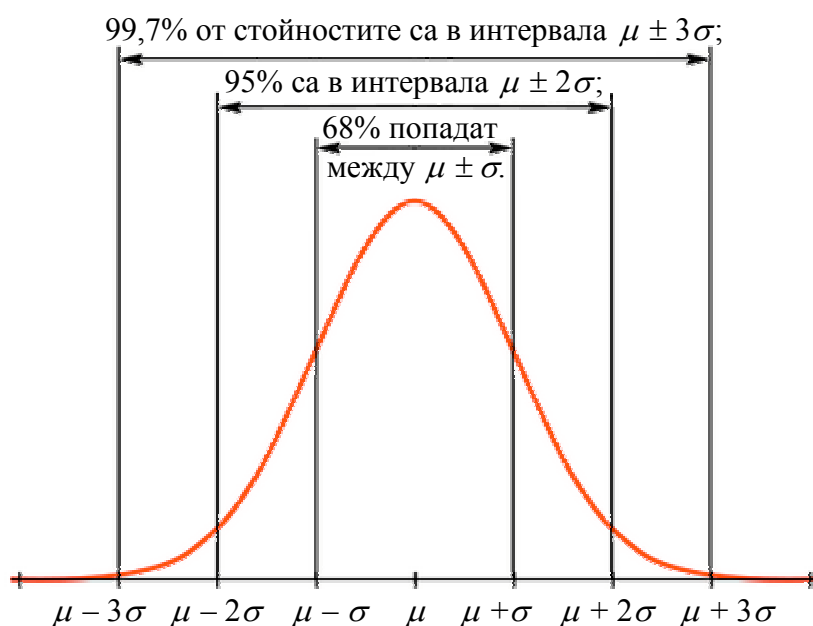
$$F(t) \rightarrow 1 \text{ при } t \rightarrow +\infty;$$

$$F(t) \rightarrow 0 \text{ при } t \rightarrow -\infty.$$

Параметрите на нормалното разпределение имат следните значения:

1) μ е средна стойност в няколко смисъла: математическо очакване (средно претеглено), медиана и мода;

2) σ^2 е дисперсията, а σ е стандартното (средноквадратичното) отклонение.



Повечето стойности на нормалното разпределение са близки до средната му стойност. По-точно, важат следните формули:

$$P(|\zeta - \mu| < \sigma) = P(|\zeta - \mu| \leq \sigma) \approx 68\%;$$

$$P(|\zeta - \mu| < 2\sigma) = P(|\zeta - \mu| \leq 2\sigma) \approx 95\%;$$

$$P(|\zeta - \mu| < 3\sigma) = P(|\zeta - \mu| \leq 3\sigma) \approx 99,7\%.$$

Тъй като интегралът във функцията на разпределение не е елементарна функция, той се пресмята числено. Преди изобретяването на компютъра пресметнатите стойности са се пазели във вид на печатни таблици, които са били използвани наготово при нужда. Не е удобно да се съставя отделна таблица за различните стойности на параметрите μ и σ , затова от многото възможни стойности се отделят две (по една за всеки от двата параметъра) и се съставя таблица само за това нормално разпределение, а другите се свеждат до него.

Стандартно нормално разпределение $N(0; 1)$

То има средна стойност $\mu = 0$, стандартно отклонение $\sigma = 1$, функция на разпределение

$$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t \exp\left(-\frac{x^2}{2}\right) dx.$$

t	0,0000	0,5000	1,0000	1,2816	1,5000	1,6449	1,9600	2,0000	2,3263	2,5000	3,0000
$\Phi(t)$	0,5000	0,6915	0,8413	0,9000	0,9332	0,9500	0,9750	0,9772	0,9900	0,9938	0,9987

Като всяка функция на разпределение, $\Phi(t) \rightarrow 1$ при $t \rightarrow +\infty$, $\Phi(t) \rightarrow 0$ при $t \rightarrow -\infty$. Стандартното нормално разпределение е симетрично спрямо нулата, затова важи формулата

$$\Phi(-t) = 1 - \Phi(t).$$

Например $\Phi(-2) = 1 - \Phi(2) = 1 - 0,9772 = 0,0228$.

Нестандартните нормални разпределения $N(\mu; \sigma)$ се изразяват чрез стандартното така:

$$F(t) = \Phi\left(\frac{t - \mu}{\sigma}\right).$$

Изваждането на μ се нарича центриране, а делението на σ — нормиране.

Приложения на нормалното разпределение

Нормалното разпределение е най-често използваното разпределение в статистиката: повечето непрекъснати случайни величини от практиката имат нормално разпределение. Това се дължи на **централната гранична теорема**, чиято интуитивна формулировка гласи:

Сборът на много случайни величини, независими в съвкупност (или слабо зависими) и от еднакъв порядък, е приблизително нормално разпределен.

От практическа гледна точка използването на нормалното разпределение е оправдано, защото повечето случайни величини, срещани в ежедневието, зависят от множество фактори, които действат независимо един от друг и всеки има малък принос към общата стойност.

Пример: Автомат пълни кутии с прясно мляко, като количеството мляко в кутия е случайна величина със средна стойност 1 литър и стандартно отклонение 15 милилитра.

а) Каква част от кутиите съдържат по-малко от 970 милилитра?

б) В какви граници около 1 л попада с 95 % сигурност количеството в случайна кутия?

Решение: Нека ζ е количеството мляко, налято в случайно избрана кутия. По условие $\zeta \sim N(\mu; \sigma)$, където $\mu = 1000$ мл, $\sigma = 15$ мл.

а) Случайно избрана кутия съдържа по-малко от 970 мл мляко с вероятност $P(\zeta < 970) = F(970) = \Phi\left(\frac{970 - 1000}{15}\right) = \Phi(-2) = 1 - \Phi(2) \approx 1 - 0,9772 = 0,0228 = 2,28 \%$.

Отговор: 2,28 % от кутиите съдържат по-малко от 970 мл мляко.

б) Търсим такова число $d > 0$, че $P(1000 - d \leq \zeta \leq 1000 + d) = 95 \%$. Следователно $0,95 = F(1000 + d) - F(1000 - d) = \Phi\left(\frac{d}{15}\right) - \Phi\left(-\frac{d}{15}\right) = 2\Phi\left(\frac{d}{15}\right) - 1 \Leftrightarrow \Phi\left(\frac{d}{15}\right) = 0,975$.

Оттук $d / 15 = 1,96 \Leftrightarrow d = 1,96 \cdot 15 = 29,4$ мл, затова търсеният интервал е $1000 \pm 29,4$ мл.

Отговор: 95 % от кутиите съдържат между 970,6 мл и 1029,4 мл мляко.

Дискретни случайни величини

Всяка случайна величина има закон за разпределение, който задава съвкупността от възможни стойности на величината и вероятността за всяка стойност. Тук разглеждаме само дискретните случайни величини, т.е. тези величини, чието множество от стойности е изброимо (крайно или безкрайно).

Нека случайната величина ξ има следното дискретно разпределение:

стойност	x_1	x_2	$\dots$	x_n
вероятност	p_1	p_2	$\dots$	p_n

Стойностите $x_1, x_2, \dots, x_n$ трябва да бъдат две по две различни.

Вероятностите $p_1, p_2, \dots, p_n$ трябва да бъдат неотрицателни числа, а сборът им трябва да бъде равен на единица: $p_1 + p_2 + \dots + p_n = 1$.

Математическо очакване: $E\xi = p_1 x_1 + p_2 x_2 + \dots + p_n x_n$. То има смисъл на средна стойност на случайната величина: ако извършим дълга редица от наблюдения върху стойностите на ξ и пресметнем средноаритметичната стойност на всички наблюдавани, то тя с голяма вероятност ще бъде близка до $E\xi$.

Дисперсия: $D\xi = p_1 (x_1 - E\xi)^2 + p_2 (x_2 - E\xi)^2 + \dots + p_n (x_n - E\xi)^2$. Дисперсията е мярка за средното отклонение на случайната величина от нейното математическо очакване. От определението се вижда, че дисперсията винаги е неотрицателна. Мерната единица на дисперсията е квадратът на мерната единица на самата случайна величина. Затова, въпреки че дисперсията е мярка за отклонението, тя не съвпада със самото отклонение: то се получава, като коренуваме дисперсията:

$\sigma = \sqrt{D\xi}$ (стандартно отклонение). То винаги е неотрицателно. По определение то е средно (по-точно, средноквадратично) отклонение, а не максимално отклонение.

Свойства на математическото очакване, дисперсията и стандартното отклонение:

- 1) $Ec = c$. 2) $E(c\xi) = c \cdot E\xi$. Тук c е произволна константа.
В частност, при $c = -1$ получаваме равенството $E(-\xi) = -E\xi$.
- 3) $E(\xi_1 \xi_2 \dots \xi_n) = (E\xi_1)(E\xi_2) \dots (E\xi_n)$. Тази формула важи само за случайни величини, които са независими в съвкупност.
- 4) $E(\xi_1 + \xi_2 + \dots + \xi_n) = E\xi_1 + E\xi_2 + \dots + E\xi_n$ за всякакви случайни величини.
В частност, $E(\xi_1 + \xi_2) = E\xi_1 + E\xi_2$ и $E(\xi_1 - \xi_2) = E\xi_1 - E\xi_2$.
- 5) $Dc = 0$. 6) $D(c\xi) = c^2 \cdot D\xi$. Тук c е произволна константа.
В частност, при $c = -1$ получаваме равенството $D(-\xi) = D\xi$.
- 7) $D(\xi_1 + \xi_2 + \dots + \xi_n) = D\xi_1 + D\xi_2 + \dots + D\xi_n$. Тази формула важи само за случайни величини, които са независими в съвкупност.
В частност, $D(\xi_1 \pm \xi_2) = D\xi_1 + D\xi_2$, ако ξ_1 и ξ_2 са независими величини.
- 8) $D\xi = E(\xi^2) - (E\xi)^2$. Така по-удобно се пресмята дисперсията.
- 9) $\sigma(c) = 0$. 10) $\sigma(c\xi) = |c| \cdot \sigma(\xi)$. Тук c е произволна константа.
- 11) $E(\xi + c) = c + E\xi$. 12) $D(\xi + c) = D\xi$. 13) $\sigma(\xi + c) = \sigma(\xi)$.

Тези формули могат да бъдат прилагани успешно и при анализ на по-сложни изрази, образувани от случайни величини. Например имаме n случайни величини $\xi_1, \xi_2, \dots, \xi_n$ с едно и също математическо очакване μ и с едно и също стандартно отклонение σ . Какви ще са математическото очакване и стандартното отклонение на средното аритметично?

Ще получим отговор на този въпрос с помощта на приведените формули. И тъй,

$$\mathbf{E}\left(\frac{\xi_1 + \xi_2 + \dots + \xi_n}{n}\right) = \frac{\mathbf{E} \xi_1 + \mathbf{E} \xi_2 + \dots + \mathbf{E} \xi_n}{n} = \frac{\mu + \mu + \dots + \mu}{n} = \frac{n\mu}{n} = \mu.$$

С други думи, средното аритметично има същото математическо очакване μ . Този извод важи за всякакви случайни величини — както независими, така и зависими.

Ако случайните величини са независими, можем да получим извод и за отклонението.

$$\mathbf{D}\left(\frac{\xi_1 + \xi_2 + \dots + \xi_n}{n}\right) = \frac{\mathbf{D} \xi_1 + \mathbf{D} \xi_2 + \dots + \mathbf{D} \xi_n}{n^2} = \frac{\sigma^2 + \dots + \sigma^2}{n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

Като коренуваме получения израз, стигаме до следния извод за стандартното отклонение:

$$\sigma\left(\frac{\xi_1 + \xi_2 + \dots + \xi_n}{n}\right) = \frac{\sigma}{\sqrt{n}}.$$

Тази формула показва, че средното аритметично на независими случайни величини е по-точно от всяка от тях, взета поотделно. Изводът е с важно практическо значение: случайните грешки имат склонност да се компенсират с нарастване на броя на измерванията, стига измерванията да са независими.

Пример: Да приемем, че всеки ръчен часовник показва времето с точност ± 6 минути. Колко души трябва да попитаме за часа, та като добавим и показанието на нашия часовник, да получим времето с точност ± 2 минути?

Решение: В условието на задачата се подразбира, че става въпрос за стандартното, а не за максималното отклонение. Тъй като стандартното отклонение намалява като $\sqrt{n}$, то за да намалее 3 пъти (от 6 минути на 2 минути), е нужно $\sqrt{n} = 3$, откъдето $n = 9$. С други думи, 9 измервания намаляват случайната грешка 3 пъти. Понеже в измерванията се брой и нашият собствен часовник, остава да попитаме още осем души за точния час.

Отговор: Трябва да попитаме осем души.

Забележка: Тъй като средното аритметично има същото математическо очакване μ (което е равно на точния час), то за приближение на неизвестния час трябва да вземем средното аритметично на показанията на деветте часовника.

Казаното не важи за т. нар. систематични грешки. С нарастване на броя на измерванията те не се унищожават, а напротив, натрупват се. Ако например няколко пъти измерим дължината на отсечка посредством невярна линейка, която мери с грешка +3 милиметра, то тази грешка е систематична и няма да се смали от много измервания: крайният резултат (средноаритметичен на измерванията) ще бъде завишен с около 3 милиметра.

И така, направеното сравнение между случайни и систематични грешки показва, че увеличаването на броя на измерванията води до намаляване на случайните грешки, но не и на систематичните грешки. Причината е, че случайните грешки са независими, а систематичните не са.

Свойствата на математическото очакване, дисперсията и стандартното отклонение важат както за дискретни, така и за независими случайни величини. Същото се отнася за техните следствия по-горе за числовите характеристики на средното аритметично.

Често срещани дискретни разпределения

Следните дискретни разпределения имат голямо теоретическо и практическо значение:

- 1) Биномно разпределение: $\xi \sim \text{Bi}(n, p) \Leftrightarrow \xi = \text{броят на успехите от } n \text{ независими опита, всеки от които има една и съща вероятност за успех } p$.
- 2) Геометрично разпределение, започващо от нула: $\xi \sim \text{Ge}(p) \Leftrightarrow \xi = \text{броят на неуспехите до първия успех, ако опитите са независими и имат една и съща вероятност за успех } p$.
- 3) Геометрично разпределение, започващо от единица: $\xi = \text{брой опити до първия успех включително, ако опитите са независими и имат една и съща вероятност за успех } p$.
Когато опитите са независими и вероятността за успех е една и съща, се казва, че опитите се провеждат по урновата схема “със връщане” (например трите разпределения по-горе).
- 4) Хипергеометрично разпределение: $\xi \sim \text{HG}(n, N, M) \Leftrightarrow \xi = \text{брой успехи от } n \text{ опита по урновата схема “без връщане”}; \text{отначало урната има } N \text{ елемента и } M \text{ от тях са успехи}$.
- 5) Поасоново разпределение: $\xi \sim \text{Po}(\lambda) \Leftrightarrow \xi = \text{брой събития в даден интервал от време, настъпващи поединично, случайно и независимо; } \lambda \text{ е средният брой събития, настъпващи в интервал от време със същата дължина}$.
- 6) Равномерно разпределение $U(n)$: стойностите са $1, 2, 3, \dots, n$ и са равновероятни.

Разпределение	Закон на разпределение $P\{\xi = k\}$	Математическо очакване $E\xi$	Дисперсия $D\xi$
Биномно $\xi \sim \text{Bi}(n, p)$ $n \in \mathbb{N}, 0 < p < 1$	$\binom{n}{k} p^k (1-p)^{n-k}$ $k = 0, 1, 2, \dots, n$	np	$np(1-p)$
Геометрично от 0 $\xi \sim \text{Ge}(p)$ $0 < p < 1$	$(1-p)^k p$ $k = 0, 1, 2, \dots$	$\frac{1-p}{p}$	$\frac{1-p}{p^2}$
Геометрично от 1 $0 < p < 1$	$(1-p)^{k-1} p$ $k = 1, 2, 3, \dots$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Хипергеометрично $\xi \sim \text{HG}(n, N, M)$ $n, N, M \in \mathbb{N};$ $n < N, M < N$	$\frac{C_M^k \cdot C_{N-M}^{n-k}}{C_N^n}$ $\max(0; n+M-N) \leq k \leq \min(n; M)$	$\frac{nM}{N}$	$\frac{nM(N-M)(N-n)}{N^2(N-1)}$
Поасоново $\xi \sim \text{Po}(\lambda)$ $\lambda > 0$	$\frac{\lambda^k}{k!} e^{-\lambda}$ $k = 0, 1, 2, \dots$	λ	λ
Равномерно $\xi \sim U(n)$ $n \in \mathbb{N}$	$\frac{1}{n}$ $k = 1, 2, \dots, n$	$\frac{n+1}{2}$	$\frac{n^2-1}{12}$

Примери:

1) Броят точки, хвърлени от зарче, притежава равномерно разпределение $U(n)$ с $n = 6$: всички изходи са равновероятни (с вероятност $1/6$). Ето защо можем да направим извод, че зарчето хвърля средно $\frac{6+1}{2} = 3,5$ точки със стандартно отклонение $\sqrt{\frac{6^2-1}{12}} \approx 1,7$ т.

2) Броят попадения от общо 24 изстрела, всеки от които улучва с вероятност $1/3$, има биномно разпределение $Bi(n, p)$ с параметри $n = 24$ и $p = 1/3$. Броят улучени цели е равен средно на $24 \cdot (1/3) = 8$ със стандартно отклонение $\sqrt{24 \cdot (1/3) \cdot (2/3)} \approx 2,3$. Вероятността точно 5 изстрела да улучат целта е равна на $C_{24}^5 \cdot (1/3)^5 \cdot (2/3)^{19} \approx 0,08 = 8 \%$.

3) Ако всеки изстрел улучва с вероятност $1/3$, то броят на неуспешните изстрели до първото улучване на целта има геометрично разпределение $Ge(p)$, започващо от нулата, с вероятност за успех $p = 1/3$. Средният брой неуспешни изстрели до първото улучване е равен на $(1/p) - 1 = 2$ и притежава дисперсия $[1 - (1/3)] / [(1/3)^2] = (2/3) / (1/9) = 6$. Затова стандартното отклонение е $\sqrt{6} \approx 2,4$. Вероятността да настъпят повече от два пропуска до първото улучване на целта е единица минус вероятността за нула, един или два пропуска: $1 - (2/3)^0 \cdot (1/3) - (2/3)^1 \cdot (1/3) - (2/3)^2 \cdot (1/3) = 1 - (1/3) - (2/9) - (4/27) = 8/27 \approx 0,30 = 30 \%$.

4) При условията на предишния пример броят на всички изстрели до първото улучване (като броим и самото него) има геометрично разпределение, започващо от единица; $p = 1/3$. Дисперсията и стандартното отклонение са същите, обаче средният брой е $1/p = 3$.

5) Ако в урна има 4 бели и 6 черни топки и извадим от нея 3 топки, без да ги връщаме, то броят на белите топки измежду извадените притежава хипергеометрично разпределение $HG(n, N, M)$ с параметри $n = 3$, $N = 10$, $M = 4$. Средният брой извадени бели топки е равен на $3 \cdot 4 / 10 = 1,2$ с дисперсия $(3 \cdot 4 \cdot 7 \cdot 6) / (10 \cdot 10 \cdot 9) = 0,56$, откъдето намираме стандартното отклонение: $\sqrt{0,56} \approx 0,7$. Вероятността да извадим точно две бели топки е равна на $C_4^2 \cdot C_6^1 / C_{10}^3 = 6 \cdot 6 / 120 = 0,30 = 30 \%$.

6) Ако през едно конкретно кръстовище минават средно по 10 леки коли в минута, то броят на колите, които ще минат оттам през следващата минута, е случайна величина с поасоново разпределение $Po(\lambda)$, където $\lambda = 10$. Това по условие е средният брой коли, а стандартното отклонение е $\sqrt{10} \approx 3,2$. Вероятността да преминат точно две коли през следващата минута е равна на $10^2 \cdot e^{-10} / 2! \approx 0,002 = 0,2 \%$.