Дискретные случайные величины

Содержание

Понятие дискретной случайной величины

Понятие случайной величины является одним из основных в теории вероятностей.

Под ***случайной величиной*** *понимается переменная, которая в результате испытания в зависимости от случая принимает одно из возможного множества своих значений* (*заранее неизвестно, какое именно*). Рассмотрим несколько примеров, выбранных из разных областей знаний:

1) число студентов, присутствующих на лекции;
2) число родившихся детей в течение года в городе Кирове;
3) число нераспавшихся ядер радиоактивного элемента на некоторый момент времени.

Если данные примеры рассматривать с точки зрения содержания, все они разнородны, так как описывают различные явления. Однако с точки зрения теории вероятностей всё это – заранее непредсказуемые вследствие влияния случайных, неподдающихся учёту обстоятельств, числовые значения. Например, если при случайном отборе выбор бы пал на другое животное, то никто не сможет оспорить, что прирост веса у него может быть другим.

Случайные величины принято обозначать прописными буквами латинского алфавита: X, Y, Z, ..., а возможные их значения (значения, которые они могут принимать при проведении испытаний) – строчными: x, y, z, ... Различают непрерывные и дискретные случайные величины.

Случайную величину, принимающую в процессе испытаний отдельные значения x_i, *которые можно пронумеровать, называют* ***дискретной (прерывной) случайной величиной***.

Вероятности *дискретных случайных* величин принято обозначать $P(x_i)$ или $P(X=x_i)$, что означает вероятность того, что случайная величина X примет определённое значение x_i (**примеры**).

Исследования показывают, что, хотя разнообразие случайных величин велико, однако принимаемые ими значения могут быть представлены в виде счётного или несчётного множества.

Например, множество фруктовых деревьев на садовом участке можно пронумеровать, а множество всех действительных чисел, принадлежащих отрезку [0, 1], – нельзя.

Дискретная случайная величина *X называется* ***заданной****, если перечислены все возможные значения, которые она может принимать в ходе испытания, и их вероятности.*

Если x_1, x_2, ..., x_n – попарно несовместные единственно возможные значения дискретной случайной величины X и $p_i=P(x_i)$ (где $i=1, 2, \ldots, n$) – вероятности этих значений, то, на основании определения полной группы событий и теоремы теории вероятностей имеем:

$$p_1 + p_2 + \ldots + p_n = \sum_{i-1}^{n} p_i = 1. \qquad (1)$$

Выражение (1) называется ***условием нормировки***.

Соответствие между всеми возможными значениями x_i дискретной случайной величины X и их вероятностями p_i называется ***законом распределения*** *случайной величины X***.**

Закон распределения дискретной случайной величины в простейшем случае может быть задан:

- в виде таблицы:

X	x_1	x_2	x_3	…	x_i	…	x_n
P	p_1	p_2	p_3	…	p_i	…	p_n

где $p_1 + p_2 + \ldots + p_n = \sum_{i-1}^{n} p_i = 1$;

- графически:

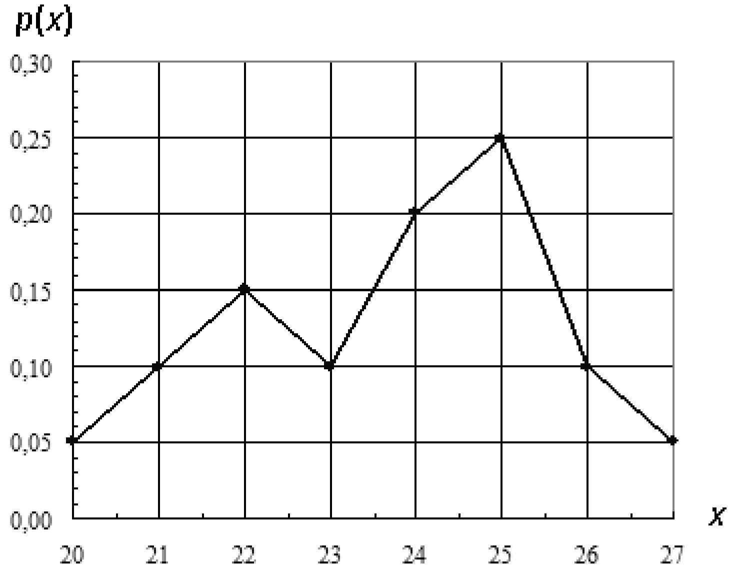

Рис. 1

Приведённая на рис. 1 фигура называется *многоугольником* (или *полигоном*) *распределения*.

Пример 1. Вероятности того, что студент сдаст экзамены в сессию (всего два экзамена) по дисциплинам A и B, равны соответственно 0,8 и 0,6. Составить закон распределения числа экзаменов, которые сдаст студент в сессию.

Решение. Случайная величина X – число экзаменов, которые сдаст студент, её возможные значения: 0, 1 и 2.

Пусть A – событие, состоящее в том, что студент сдаст экзамен по дисциплине A, B – событие, состоящее в том, что студент сдаст экзамен по дисциплине B. Здесь события A и B независимые. Тогда искомые вероятности будут равны

$$p_1 = P(X=0) = P(\overline{A}\cdot\overline{B}) = P(\overline{A})\cdot P(\overline{B}) = (1-0{,}8)\cdot(1-0{,}6) = 0{,}08;$$

$$p_2 = P(X=1) = P(\overline{A}\cdot B + A\cdot\overline{B}) = P(\overline{A})\cdot P(B) + P(A)\cdot P(\overline{B}) =$$

$$= (1-0{,}8)\cdot 0{,}6 + 0{,}8\cdot(1-0{,}6) = 0{,}44;$$

$$p_3 = P(X=2) = P(A\cdot B) = P(A)\cdot P(B) = 0{,}8\cdot 0{,}6 = 0{,}48.$$

Закон распределения (соответствие между значениями дискретной случайной величины и их вероятностями) представим в виде таблицы, составленной на основании полученных данных:

x_i	0	1	2
p_i	0,08	0,44	0,48

Проверим условие нормировки: 0,08+0,44+0,48=1.

Случайные величины *X и Y называются* ***независимыми****, если на закон распределения одной из них не влияет, какие значения в ходе испытания принимает другая.*

Числовые характеристики дискретных случайных величин

Полную характеристику случайной величины даёт закон её распределения. Однако, как показывает практика, он в силу различного рода причин (например, если вся изучаемая совокупность не доступна или слишком велика) в большинстве случаев бывает неизвестен. Поэтому, наряду с законом распределения (а также вместо него), для её представления могут быть использованы определённые числовые параметры, которые обобщают то, что свойственно всем значениям данной случайной величины, получившие название *числовые характеристики случайной величины*.

К ним относятся:

- математическое ожидание $M(X)$;
- дисперсия $D(X)$ или σ^2;
- среднее квадратическое отклонение $\sigma(X) = \sqrt{D(X)}$.

Математическое ожидание дискретной случайной величины

Математическим ожиданием дискретной случайной величины X называется число, характеризующее среднее значение этой случайной величины и ***являющееся для неё центром рассеяния***. Математическое ожидание обозначают $M(X)$ или μ.

Математическое ожидание случайной величины равно сумме парных произведений всех возможных её значений x_i на соответствующие им вероятности p_i:

$$M(X) = x_1 \cdot p_1 + x_2 \cdot p_2 + \ldots + x_n \cdot p_n = \sum_{i=1}^{n} x_i \cdot p_i . \qquad (2)$$

Пример 2. Рассмотрим пример о сдаче студентом семестровых экзаменов. Найти математическое ожидание числа сданных экзаменов.

Решение. На основании найденного закона распределения

x_i	0	1	2
p_i	0,08	0,44	0,48

и определения математического ожидания (формула (2)) вычислим:

$$M(X) = 0 \cdot 0{,}08 + 1 \cdot 0{,}44 + 2 \cdot 0{,}48 = 1{,}4 \approx 1 \text{ экзамен.}$$

Таким образом, среднее число сданных студентом семестровых экзаменов составляет 1 экзамен.

Как видно из определения, математическое ожидание $M(X)$ – величина постоянная, поэтому оно представляет собой числовую характеристику случайной величины X.

Основные свойства математического ожидания

Теорема 1

Математическое ожидание постоянной величины равно этой постоянной, то есть, если C – постоянная, то:

$$M(C)=C. \qquad (3)$$

Доказательство. Постоянную величину C можно рассматривать как величину, принимающую значение C с вероятностью 1. Поэтому $M(C)=C\cdot 1=1$ (см. (2)).

Теорема 2

Математическое ожидание алгебраической суммы конечного числа случайных величин равно такой же сумме их математических ожиданий, то есть:

$M(X\pm Y)=M(X)\pm M(Y)$

или

$$M(X\pm Y\pm\ldots\pm Z)=M(X)\pm M(Y)\pm\ldots\pm M(Z). \qquad (4)$$

Теорема 3

Математическое ожидание произведения конечного числа независимых случайных величин равно произведению их математических ожиданий, то есть:

$M(X\cdot Y)=M(X)\cdot M(Y)$

или

$$M(X\cdot Y\cdot\ldots\cdot Z)=M(X)\cdot M(Y)\cdot\ldots\cdot M(Z). \qquad (5)$$

Теорема 4

Если каждое значение x_i случайной величины X умножить (*разделить*) *на одну и ту же постоянную величину, то и математическое ожидание данной случайной величины увеличится* (*уменьшится*) *во столько же раз*.

Доказательство. Пусть C – постоянная величина. Действительно, на основании определения математического ожидания имеем:

$$M(C\cdot X)=\sum_{i=1}^{n}\left(C\cdot x_i\cdot p_i\right)=C\cdot\sum_{i=1}^{n}\left(x_i\cdot p_i\right)=C\cdot M(X) \qquad (6)$$

или

$$M\left(\frac{1}{C}\cdot X\right)=\sum_{i=1}^{n}\left(\frac{1}{C}\cdot x_i\cdot p_i\right)=\frac{1}{C}\cdot\sum_{i=1}^{n}\left(x_i\cdot p_i\right)=\frac{1}{C}\cdot M\left(X\right). \qquad (7)$$

Теорема 5

Если каждое значение x_i случайной величины X увеличить (*уменьшить*) *на одну и ту же постоянную величину, то и её математическое ожидание увеличится* (*уменьшится*) *на эту же постоянную величину*.

Доказательство. Пусть C – постоянная величина. Тогда, используя формулы (2), (3) и (4), имеем:

$$\begin{aligned}M\left(X\pm C\right)&=\sum_{i=1}^{n}\left(\left(x_i\pm C\right)\cdot p_i\right)=\sum_{i=1}^{n}\left(x_i p_i\pm C\cdot p_i\right)=\\&=\sum_{i=1}^{n}\left(x_i\cdot p_i\right)\pm C\cdot\sum_{i=1}^{n}p_i=\sum_{i=1}^{n}\left(x_i\cdot p_i\right)\pm C=M\left(X\right)\pm C.\end{aligned} \qquad (8)$$

Замечание. Следует помнить, что использование средних требует строгого соблюдения принципа однородности значений случайной величины. Несоблюдение этого ведет к искажённому представлению о реальных процессах.

Дисперсия

Среднее далеко не всегда позволяет исчерпывающе охарактеризовать изменяющуюся случайную величину, и тем менее точна эта характеристика, чем больше разброс её значений.

Пример 3. Пусть случайные величины X и Y имеют законы распределения:

X	x_i	-10	-5	-2	2	5	10
P_x	p_{xi}	0,1	0,2	0,2	0,2	0,2	0,1

Y	y_i	-1,0	-0,5	-0,2	0,2	0,5	1,0
P_y	p_{yi}	0,1	0,2	0,2	0,2	0,2	0,1

Тогда:

$$M(X)=\sum_{i=1}^{n_x}(x_i\cdot p_{xi})=0,$$

$$M(Y)=\sum_{i=1}^{n_y}(y_i\cdot p_{yi})=0.$$

Из данного примера мы видим, что математические ожидания случайных величин могут быть одинаковыми или близкими, а варьирование значений отличаться существенным образом. Поэтому, для более полного представления данной случайной величины, недостаточно определить только среднюю, нужно оценивать ещё и показатели, учитывающие степень вариации (рассеяния) её значений.

Дисперсия (от латинского *dispersio* – рассеяние) является одним из показателей вариации всех значений случайной величины, которые могут появиться в результате испытаний. Дисперсия – числовая характеристика случайной величины.

Пусть случайная величина X, принимающая в результате испытаний значения x_i, где i=1, 2, ..., n, имеет математическое ожидание μ. Тогда, очевидно, и разности (x_i–μ) также являются случайными величинами, – их называют *центрированными случайными величинами* или случайным отклонением индивидуального значения x_i от математического ожидания μ (центра рассеяния). В общем случае они обозначаются разностью (X–μ).

Тогда, если случайная величина X задана известным законом распределения:

X	x_1	x_2	…	x_i	…	x_{n-1}	x_n
P	p_1	p_2	…	p_i	…	p_{n-1}	p_n

центрированная случайная величина будет иметь закон распределения:

$X-\mu$	$x_1-\mu$	$x_2-\mu$	…	$x_i-\mu$	…	$x_{n-1}-\mu$	$x_n-\mu$
P	p_1	p_2	…	p_i	…	p_{n-1}	p_n

Теорема 6

Для любой случайной величины X, имеющей математическое ожидание μ, математическое ожидание отклонений её значений от центра рассеяния μ равно нулю:

$$M\left(X - M(X)\right) = \sum_{i=1}^{n}\left((x_i-\mu)\cdot p_i\right) = 0. \qquad (9)$$

Доказательство. Доказательство основано на том, что μ – числовая характеристика, то есть постоянная величина.

Тогда на основании свойств математического ожидания (теоремы 1 и 2) имеем:

$$M\,(X-M(X))=M(X)-M\,(M(X))=M(X)-M(X)=0.$$

На основании этой теоремы мы можем заключить, что учёт *простых* отклонений типа ($x_i-\mu$) не даёт нам достоверной информации о степени рассеяния значений случайной величины.

Дисперсией $D(X)$ называется числовой показатель, позволяющий оценить степень рассеяния значений, принимаемых случайной величиной, относительно её математического ожидания μ.

Дисперсия *случайной дискретной величины равна математическому ожиданию квадрата отклонения значений случайной величины X от её математического ожидания*.

На основании определения математического ожидания имеем:

$$D(X) = \sigma^2 = M\left((X-\mu)^2\right) = \sum_{i=1}^{n}\left((x_i-\mu)^2\cdot p_i\right). \qquad (10)$$

Другими словами: **дисперсия** – это среднее значение квадрата отклонения случайной величины от её среднего значения.

Свойства дисперсии

Теорема 7

Дисперсия случайной величины равна разности между математическим ожиданием квадрата случайной величины $M(X^2)$ *и квадратом её математического ожидания* $[M(X)]^2$:

$$D(X)=\mathrm{M}(X^2)-[M(X)]^2.$$

Доказательство. При доказательстве воспользуемся основными свойствами математического ожидания:

$$D(X)=M[(X-M(X))]^2=M[X^2-2X\cdot M(X)+M(X)^2]=$$
$$=M(X^2)-2M(X)\cdot M(X)+[M(X)]^2=$$
$$=M(X^2)-2[M(X)]^2+[M(X)]^2=\mathrm{M}(X^2)-[M(X)]^2. \qquad (11)$$

Формула (11) облегчает задачу нахождения дисперсии случайной величины.

Теорема 8

Дисперсия постоянной величины равна нулю: $D(C)=0$ (постоянная величина рассеяния не имеет).

Доказательство. Пусть C – постоянная величина. Тогда, на основании свойства математического ожидания (теорема 1), имеем: $M(C)=C$. Следовательно:

$$D(C)=M\,[C-M(C)]^2=M(C-C)^2=M(0)=0. \qquad (12)$$

Теорема 9

Дисперсия суммы двух независимых случайных величин X и Y равна сумме дисперсий каждой из них

$$D(X+Y)=D(X)+D(Y). \qquad (13)$$

Доказательство. При доказательстве, как и в предыдущем случае, будем пользоваться как свойствами математического ожидания, так и формулой (10).

Действительно:

$$D(X+Y)=M\,[(X+Y)^2]-[M(X+Y)]^2=M\,(X^2+2XY+Y^2)-[M(X)+M(Y)]^2=$$

$$=M(X^2)+2M(X)\cdot M(Y)+M(Y2)-[M(X)]2-2M(X)\cdot M(Y)-[M(Y)]2=$$

$$=M(X\,2)-[M(X)]^2+M(Y^2)-[M(Y)]^2=D(X)+D(Y).$$

Следствие. *Дисперсия суммы конечного числа независимых случайных величин равна сумме дисперсий этих величин*:

$$D(X+Y+\ldots+Z)=D(X)+D(Y)+\ldots+D(Z). \qquad (14)$$

Теорема 10

Постоянный множитель можно выносить за знак дисперсии, возводя его в квадрат:

$$D(C\cdot X)=C^2\cdot D(X).$$

Доказательство. Доказательство основано на соответствующем свойстве математического ожидания (следствие 1) и дисперсии (теорема 8). Действительно, если C – постоянный множитель, то:

$$D(C\cdot X)=M(C^2\cdot X^2)-[M(C\cdot X)]^2=C^2\cdot M(X^2)-[C\cdot M(X)]^2=$$

$$=C^2\cdot M(X^2)-C^2\cdot[M(X)]^2=C^2\cdot\{M(X^2)-[M(X)]^2\}=C^2\cdot D(X). \qquad (15)$$

Следствие 1. *Дисперсия разности двух независимых случайных величин равна сумме их дисперсий*, то есть, если случайные величины X и Y независимы, то:

$$D(X-Y)=D(X)+D(Y). \qquad (16)$$

Доказательство. Доказательство основано на использовании ранее рассмотренных нами свойств дисперсии (теоремы 2.7 и 2.8):

$$D(X-Y)=D[X+(-Y)]=D(X)+D(-Y)=D(X)+(-1)^2\cdot D(Y)=D(X)+D(Y).$$

Следствие 2. *Если C – постоянная, то случайные величины X и $X\pm C$ имеют одинаковые дисперсии*. Этот вывод основан на данных теорем 8 и 9 и следствия 1.

Доказательство. Действительно, если C – постоянная, то:

$$D(X\pm C)=D(X)+D(C)=D(X)+0=D(X). \qquad (17)$$

На основании этого следствия можно заключить, что, если каждое значение случайной величины X увеличить или уменьшить на одно и то же постоянное число C, то дисперсия не изменится. Это свойство позволяет упростить процесс вычисления дисперсии случайной величины.

Среднее квадратическое отклонение

Дисперсия $D(X)$ является важной числовой характеристикой случайной величины, отражающей степень разброса её значений относительно центра её рассеяния – математического ожидания. Она имеет размерность квадрата размерности самой случайной величины. В тех случаях, когда оценку разброса значений случайной величины необходимо иметь в виде числа той же размерности, что и характеризуемая величина, пользуются *средним квадратическим отклонением*.

Среднее квадратическое отклонение случайной величины X принято обозначать $\sqrt{D(X)}$ или σ.

Среднее квадратическое отклонение *– число, характеризующее среднюю величину рассеяния отдельных значений случайной величины относительно её математического ожидания, размерность которого совпадает с размерностью самой случайной величины. Среднее квадратическое отклонение равно квадратному корню из дисперсии:*

$$\sigma=\sqrt{D(X)}=\sqrt{\sum_{i=1}^{n}\left(x_i-\mu\right)^2\cdot p_i}\,. \qquad (18)$$

Пример 4. Закон распределения случайной величины X, отражающей рост человека в сантиметрах, задан таблицей:

X	150	155	160	165	170	175	180	185	190	200
P	0,03	0,05	0,10	0,12	0,16	0,20	0,15	0,09	0,06	0,04

Определить математическое ожидание, дисперсию и среднее квадратическое отклонение.

Решение. Математическое ожидание находим по формуле (2):

$$M(X)=\sum_{i=1}^{n}\left(x_i\cdot p_i\right)=173{,}3\,.$$

Дисперсию находим по формуле (34):

$$D(X)=M(X^2)-[M(X)]^2=30158{,}5-30032{,}9=125{,}6.$$

Среднее квадратическое отклонение находим по формуле (41):

$$\sigma(X)=\sqrt{D(X)}=\sqrt{\sum_{i=1}^{n}\left(x_i-M(X)\right)^2\cdot p_i}=\sqrt{125{,}6}=11{,}2\,.$$

Замечание. Помимо дисперсии и среднего квадратического отклонения существуют и другие показатели вариации значений случайной величины. К таким показателям относятся: ***лимиты, размах вариации.***

Лимиты. Обозначаются символом lim (от латинского *limites* – предел).

В биометрии под этим термином понимают минимальное x_{min} и максимальное x_{max} значения, между которыми находятся все остальные значения, принимаемые случайной величиной.

Размах вариации. Обозначается буквой R и определяется по разности максимального и минимального значения случайной величины

$$R = x_{max} - x_{min}. \qquad (19)$$

Пример 5. В результате испытаний случайные величины X и Y приняли ряд значений:

X: 1,25; 1,29; 1,32; 1,26; 1,21; 1,23; 1,16; 1,31

Y: 6,45; 6,74; 6,86; 6,21; 6,32; 6,12; 6,48; 6,56

Найти: 1) лимиты; 2) размах вариации.

Решение.

1) Согласно определению, лимиты:

$$\lim X = (x_{\min} \div x_{\max}) = (1{,}16 \div 1{,}32),$$
$$\lim Y = (y_{\min} \div y_{\max}) = (6{,}12 \div 6{,}86).$$

2) Размах вариации:

$$R_x = x_{max} - x_{min} = 1{,}32 - 1{,}16 = 0{,}16;$$
$$R_y = y_{max} - y_{min} = 6{,}85 - 6{,}12 = 0{,}73.$$

Рассмотренные показатели вариабельности случайной величины дополняют друг друга, они просты в понимании и позволяют быстро представить степень её рассеяния. Однако их нельзя наравне с математическим ожиданием, дисперсией и средним квадратическим отклонением отнести к числовым характеристикам. Например, для одной и той же случайной величины при повторении серии испытаний лимиты и размах вариации способны сильно изменяться, в то время как $M(X)$ постоянно. Кроме того, эти характеристики не позволяют оценить и характер варьирования значений случайной величины, так как для их расчетов используются только крайние значения и совершенно не учитываются остальные значения. Тем не менее они позволяют быстро составить представление об относительной величине и диапазоне варьирования значений случайной величины.

Пример 6. Пусть случайная величина X в результате проведённых двух серий испытаний приняла значения:

$$x_{i1} = 10;\ 15;\ 20;\ 25;\ 30;\ 35;\ 40;\ 45;\ 50.$$
$$x_{i2} = 15;\ 30;\ 28;\ 30;\ 10;\ 28;\ 32;\ 50;\ 32.$$

Мы видим, что при одинаковых средних $\overline{X_1} = 30$ и $\overline{X_2} = 30$ одинаковых лимитах – $x_{min1} = x_{min2} = 10$, $x_{max1} = x_{max2} = 50$ и размахе вариации $R_{x1} = R_{x2} = 40$ числовые значения (характер варьирования значений случайной величины) разные и это не отразилось на величине лимитов и размахе вариации.

Законы распределения дискретных случайных величин

В этом пункте рассмотрим наиболее распространённые законы распределения дискретных случайных величин.

Биномиальное распределение случайной величины

Биномиальное распределение является одним из самых распространённых законов распределения дискретной случайной величины. Оно может быть использовано, если необходимо оценить распределение вероятностей в зависимости от числа появлений случайного события при ограниченном числе независимых испытаний и при неизменных условиях опыта.

Рассмотрим дискретную случайную величину X, если её частные значения ($x_0=0$, $x_1=1$, $x_2=2$, …, $x_n=n$) соответствуют числу появлений некоторого события A в n независимых испытаниях, то вероятности отдельных значений случайной величины X могут быть определены при помощи формулы Бернулли:

$$P_n(k)=C_n^k\cdot p^k\cdot q^{n-k}=\frac{n!}{k!\cdot(n-k)!}\cdot p^k\cdot q^{n-k}, \qquad (20)$$

где $q=1-p$, $0<p<1$, $k=0, 1, 2, \ldots, n$.

При этом выполняются следующие условия (схема Бернулли):

• вероятность p появления события A в отдельном испытании является величиной постоянной, повторяющейся от одного испытания к другому (вероятность «успеха»);

• вероятность того, что событие A не появится в отдельном испытании, равна q (вероятность «неуспеха»);

• все испытания $x_0, x_1, x_2, \ldots, x_n$ независимы друг относительно друга.

Таким образом, суть биномиального закона распределения случайной величины X для события A заключается в соответствии числа его появлений в n испытаниях их вероятностям $P_n(k)$.

Пример 7. Найти вероятность появления в группе из четырёх особей трёх особей, обладающих некоторым признаком.

Решение. Предположим, что вероятность появления особи, обладающей некоторым признаком, равна 0,5. Тогда $p=q=0{,}5$. Воспользуемся формулой (20), где $n=4$, $x_3=3$ и $p=q=0{,}5$:

$$P_4(3)=\frac{4!}{3!\cdot(4-3)!}\cdot 0{,}5^3\cdot 0{,}5^{4-3}=\frac{1\cdot 2\cdot 3\cdot 4}{(1\cdot 2\cdot 3)\cdot 1}\cdot 0{,}5^3\cdot 0{,}5=0{,}25.$$

Таким образом, вероятность появления трёх особей (в группе из четырёх), обладающих одинаковым признаком, составляет 0,25 или 25%.

Закон биномиального распределения в общем виде может быть представлен в виде таблицы:

X	$x_i=k$	0	1	2	...	n-1	n
P	$P_n(k)$	$P_n(0)$	$P_n(1)$	$P_n(2)$	...	$P_n(n$-1$)$	$P_n(n)$

или графически в виде полигона, где ординаты $P_n(k)=C_n^{k_i}\cdot p^k\cdot q^{n-k}$ соответствуют членам разложения бинома $(p+q)^n$:

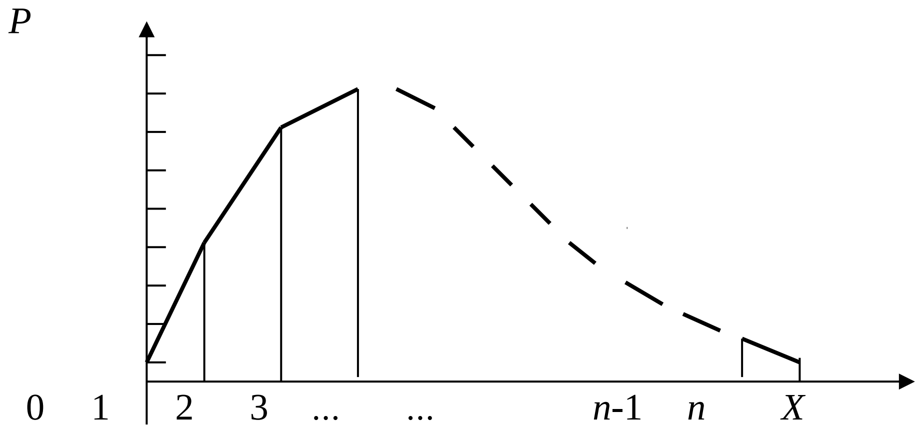

Рис. 2

Характер биномиального распределения полностью определяется параметрами n и p.

Числовые характеристики биномиального распределения:

- Математическое ожидание $M(X)=\mu=np$. (21)
- Дисперсия $D(X)=npq$. (22)
- Среднее квадратическое отклонение $\sigma=\sqrt{D(X)}=\sqrt{npq}$, (22*)

где n – число испытаний, p – вероятность события A, q=1–p – вероятность события $\bar{A}$.

Пример 8. Имеется шесть пробирок с питательной средой. Вероятность появления колонии микроорганизмов определённого вида в какой-либо из шести пробирок постоянна и равна p=0,7. Составить закон распределения появления колонии микроорганизмов.

Решение. Пусть событие A – появление колонии микроорганизмов в отдельно взятой пробирке. Вероятность события A постоянна и равна: p=0,7. Тогда вероятность того, что в отдельно взятой пробирке не появится колония микроорганизмов, равна q=1–p=0,3.

Пусть случайная величина X – число пробирок с появившейся колонией микроорганизмов. При этом определённое значение x_i случайной величины X заключается в том, что при проверке колония может быть не обнаружена – x_0=0; или обнаружена:

- в одной пробирке, то есть x_1=1;
- в двух пробирках – x_2=2;
- в трёх пробирках – x_3=3;
- в четырёх пробирках – x_4=4;
- в пяти пробирках – x_5=5;
- в шести пробирках, то есть – x_6=6.

Условие данной задачи полностью соответствует схеме Бернулли, следовательно, для нахождения вероятностей этих возможностей воспользуемся формулой (20):

$$P_n(k) = C_n^k \cdot p^k \cdot q^{n-k} = \frac{n!}{k! \cdot (n-k)!} \cdot p^k \cdot q^{n-k}.$$

Откуда:

- $P_6(0)=0{,}3^6=0{,}0007$;
- $P_6(1)=6\cdot 0{,}7^1\cdot 0{,}3^5=0{,}01$;
- $P_6(2)=15\cdot 0{,}7^2\cdot 0{,}3^4=0{,}057$;
- $P_6(3)=20\cdot 0{,}7^3\cdot 0{,}3^3=0{,}18$;
- $P_6(4)=15\cdot 0{,}7^4\cdot 0{,}3^2=0{,}32$;
- $P_6(5)=6\cdot 0{,}7^5\cdot 0{,}3^1=0{,}30$;
- $P_6(6)=0{,}7^6=0{,}11$.

Математическое ожидание $\mu=n\cdot p=6\cdot 0{,}7=4{,}2\approx 4$. Сведём данные в таблицу и получим закон распределения вероятностей появления колонии микроорганизмов в пробирках.

X	$x_i=k$	0	1	2	3	**4**	5	6
P	$P_6(k)$	0,0007	0,01	0,057	0,18	**0,32**	0,30	0,11

Отметим, что вероятности значений случайной величины X сначала возрастают, затем убывают, а значение случайно величины, приближённо равное μ, является наиболее вероятным.

Распределение Пуассона

Распределение Пуассона (закон распределения редких событий) рассматривается при тех же условиях, что и распределение Бернулли, однако оно даёт более точный результат, если вероятность события A в отдельном испытании мала ($p<0{,}1$), а число испытаний достаточно велико, например, $n>50$.

Распределение Пуассона может быть использовано при решении некоторых вероятностных задач в физике, теории информации, теории связи, теории надёжности, теории массового обслуживания, словом везде, где в течение определённого времени может происходить случайное число каких-то событий. Например, этому распределению удовлетворяют такие явления, как радиоактивные распады, телефонные вызовы, отказы оборудования, полиэмбриония в семенах растений, распределение микроэлементов в образце почвы.

Другими примерами пуассонова распределения могут также служить частота рождения троен и четверен у человека; количество сорных растений на делянках посевов; число вредных насекомых, попадающих в ловушки; частота Лангергенса в тканях поджелудочной железы; частота спонтанных мутаций у кишечной палочки; подсчёт клеток под микроскопом; количество семян, поражённых вредителями; число колоний определённого вида бактерий в поле зрения микроскопа и т. п.

Если случайная величина X представляет собой частоты появления маловероятного события A в n независимых испытаний (n велико), то вероятности отдельных значений случайной величины X могут быть найдены по формуле Пуассона:

$$P_n(k) = \frac{\mu^k}{k!} \cdot e^{-\mu}, \qquad (23)$$

где параметр μ является числовой характеристикой распределения Пуассона – его математическим ожиданием – и находится по одной из формул:

- $\mu = np$ при известном значении n;
- $\mu = \lambda t$ (λ – среднее число событий, появившихся за единицу времени).

Числовые характеристики пуассонова распределения равны

$$M(X) = \mu = D(X), \qquad (24)$$

$$\sigma = \sqrt{D(X)} = \sqrt{\mu} \qquad (24^*)$$

Пример 9. Предположим, что в воздухе классной комнаты содержится 10^4 бактерий, которые равномерно распределены по всему объёму. Вероятность нахождения бактерии в случайным образом отобранной пробе равна в среднем $p=10^{-3}$. Найти распределение вероятностей численности X бактерий в отобранной случайным образом пробе.

Решение. Событие, заключающееся в попадании в пробу хотя бы одной бактерии из большого количества, относится к разряду редких событий, поэтому данную задачу следует решать, воспользовавшись формулой Пуассона (23):

$$P_n(k) = \frac{\mu^k}{k!} \cdot e^{-\mu}.$$

При этом параметр μ равен

$$\mu = n \cdot p = 10^4 \cdot 10^{-3} = 10.$$

По условию случайная величина X может принимать значения $x_0=0$, $x_1=1$, $x_2=2$, …, $x_n=n$. Тогда вероятность того, что:

- в пробе не будет обнаружено бактерий: $P_n(0) = \frac{10^0}{0!} \cdot e^{-10} \approx 0{,}000045$;
- в пробе будет обнаружена одна бактерия: $P_n(1) = \frac{10^1}{1!} \cdot e^{-10} \approx 0{,}00045$;
- две бактерии: $P_n(2) = \frac{10^2}{2!} \cdot e^{-10} \approx 0{,}0023$;
- три: $P_n(3) = \frac{10^3}{3!} \cdot e^{-10} \approx 0{,}0076$ и т. д.

Таким образом, закон распределения между числом бактерий в пробе и их вероятностью имеет следующий вид:

X	P	X	P

0	0,000045	9	0,125
1	0,00045	10	0,125
2	0,0023	11	0,114
…	…	12	0,095
8	0,113	…	…

Заметим, что распределение Пуассона достигает максимума при значениях случайной величины, равных 9÷10, то есть приближённо равных μ.

Пример 10. Изучается распределение островков Лангерганса в поджелудочной железе обезьяны шимпанзе. Всего было просмотрено 900 квадратов. Распределение островков Лангерганса в поджелудочной железе обезьяны шимпанзе представлено в виде следующей таблицы.

Количество островков в квадратах, x_i	0	1	2	3	4	5	6
Количество квадратов, f_i	327	340	160	53	16	3	1
$p_i = f_i / n$	$\frac{327}{900}$	$\frac{340}{900}$	$\frac{160}{900}$	$\frac{53}{900}$	$\frac{16}{900}$	$\frac{3}{900}$	$\frac{1}{900}$

Решение. Вероятность нахождения островка Лангерганса (для этого гистологические срезы проецируют на экран, находят на них островки Лангерганса, учитывая количество квадратов) на любом квадрате одинакова и невелика, но распределение островков по отдельным квадратам неравномерно. Можно предположить, что оно соответствует распределению Пуассона. Найдём математическое ожидание и дисперсию по следующим формулам соответственно (2) и (10):

$$M(x) = \mu = 0\cdot\frac{327}{900} + 1\cdot\frac{340}{900} + 2\cdot\frac{160}{900} + 3\cdot\frac{53}{900} + 4\cdot\frac{16}{900} + 5\cdot\frac{3}{900} + 6\cdot\frac{1}{900} = 1{,}0.$$

$$D(x) = (0-1)^2\cdot\frac{327}{900} + (1-1)^2\cdot\frac{340}{900} + (2-1)^2\cdot\frac{160}{900} + (3-1)^2\cdot\frac{53}{900} +$$

$$+ (4-1)^2\cdot\frac{16}{900} + (5-1)^2\cdot\frac{3}{900} + (6-1)^2\cdot\frac{1}{900} = \frac{916}{900} \approx 1{,}02.$$

Близость значений математического ожидания и дисперсии (см. формулу (24)), а также очень малая величина вероятности появления островков Лангерганса $p = \frac{\mu}{n} = \frac{1}{900} \approx 0{,}001$ служит доказательством того, что распределение островков Лангерганса является пуассоновым.

Распределение геометрическое

Рассмотрим геометрическое распределение дискретной случайной величины. Пусть происходит серия независимых испытаний, в каждом из которых событие может появится с одной и той же вероятностью p. Тогда случайная величина X − количество испытаний до первого появления события, имеет геометрическое распределение вероятностей. Она может принимать всевозможные целые значения от 0 (событие произошло в первом испытании) и больше (счетное число значений).

Формула для вычисления соответствующих вероятностей легко выводится:

$$P(k)=q^k \cdot p,\ k = 0, 1, 2, ..., n, ... \qquad (25)$$

Для геометрического распределения известны готовые формулы для математического ожидания и дисперсии:

$$M(X) = 1/p,\ D(X) = q/p^2. \qquad (26)$$

Пример 11. Два орудия залпом, но при независимой наводке, стреляют в цель до первого попадания хотя бы одним орудием. Вероятность попадания в цель первым орудием при одном выстреле равна 0,2, вторым – 0,3. Найти: 1) закон распределения числа X сделанных залпов; 2) $M(X)$ с точностью до сотых.

*Решение***.** Сначала найдем вероятность p того, что залп успешный, то есть попало хотя бы одно орудие в цель:

$$p = 1 - q_1 q_2 = 1 – (1 – 0{,}2)\cdot(1 – 0{,}3) = 1 – 0{,}8\cdot 0{,}7 = 0{,}44.$$

Пусть X − число сделанных залпов. Найдем закон распределения X. X распределено по геометрическому закону с вероятностью $P(k) = 0{,}56^{k-1}0{,}44$ (первые $k-1$ неуспешных залпов с вероятностью 0,56 каждый, на k залпе попадание с вероятностью 0,44).

Закон распределения имеет вид (по формуле (25)):

X	1	2	…	k	…
P	0,44	0,2464	…	$0{,}56^{k-1}0{,}44$	…

Найдем математическое ожидание (по формуле (26)):

$$M(X) = 1/0{,}44=2{,}27$$

Распределение альтернативное

Рассмотрим альтернативное распределение дискретной случайной величины. Пусть происходит серия независимых испытаний, в каждом из которых событие может как появится (значение случайной величины берем за 1), так и не появится (значение случайной величины берем за 0) с одной и той же вероятностью p = 0,5; тогда закон распределения имеет вид:

X	0	1
P	0,5	0,5

Литература

1. Гмурман, В. Е. Теория вероятностей и математическая статистика : учебное пособие / В. Е. Гмурман. – М.: Высш. шк., 2000.
2. Гмурман, В. Е. Руководство к решению задач по теории вероятностей и математической статистике : учебное пособие / В. Е. Гмурман. – М.: Высш. шк., 1999.
3. Мхитарян, В. С. Теория вероятностей и математическая статистика : учебник / В. С. Мхитарян, В. Ф. Шишов, А. Ю. Козлов. – М.: Академия, 2012.
4. Тарасов, Л. В. Мир, построенный на вероятности / Л. В. Тарасов. – М.: Просвещение, 1984.
5. Феллер, В. Введение в теорию вероятностей и её приложения / В. Феллер. – М.: Мир, 1984.