СЛУЧАЙНЫЕ ВЕЛИЧИНЫ. ЗАКОНЫ РАСПРЕДЕЛЕНИЯ СЛУЧАЙНОЙ ВЕЛИЧИНЫ

1. Понятие случайной величины.

Современная теория вероятностей предпочитает, где только возможно, оперировать не случайными событиями, а случайными величинами, для которых был разработан более гибкий и универсальный математический аппарат.

Случайная величина – *это величина, которая в результате опыта может принимать то или иное значение, заранее не известно, какое именно.*

Случайными величинами являются, например, количество очков, выпадающих при бросании игрального кубика, число посетителей аптеки в течение случайно взятого дня, температура больного в наугад выбранное время суток, рост случайно выбранного студента и тому подобное.

Случайные величины принято обозначать прописными буквами латинского алфавита – X, У, Z и т.д., а их возможные значения – соответствующими строчными буквами с числовыми индексами. Например, значения случайной величины X обозначают следующим образом: x_1, x_2, x_3, ….

Пример: Если X - количество очков, выпадающих при бросании игрального кубика, тогда данная случайная величина принимает следующие значения $X = \{1,2,3,4,5,6\}$, где $x_1 = 1$, $x_2 = 2$ и т.д. Таким образом, значения случайной величины образуют полную группу событий.

Случайные величины бывают:

а) **непрерывные** – *значения которых непрерывно заполняют какой-либо промежуток* (например: давление крови человека, температура его тела или состав крови);

б) **дискретные** – *принимающие отдельные друг от друга значения* (например: число звонков на станцию скорой помощи в течение часа или количество очков, выпадающих при бросании игрального кубика).

Каждое свое значение случайная величина может принимать с разной вероятностью.

Основная задача теории вероятностей, оперирующей случайными величинами, – это определение **закона распределения случайной величины**, то есть установление соответствия между возможными значениями случайной величины и вероятностью наблюдения этих значений.

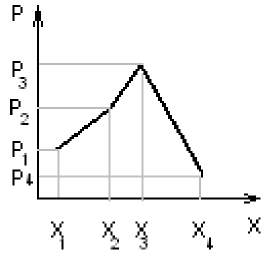
Формы закона распределения случайной величины

1) **Ряд распределения** – это таблица, где перечислены возможные значения случайной величины и соответствующие им вероятности.

Данная форма закона распределения используется только для дискретных случайных величин, так как перечислить все значения непрерывной случайной величины просто невозможно, да и вероятность наблюдения каждого из ее значений близка к нулю.

X	x_1	x_2	x_3	x_4
P	p_1	p_2	p_3	p_4

Ряд распределения



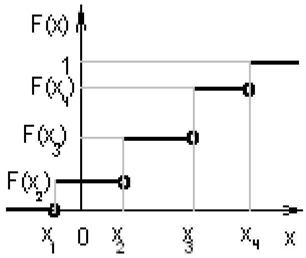
Многоугольник распределения

Графическим изображением ряда является многоугольник распределения.

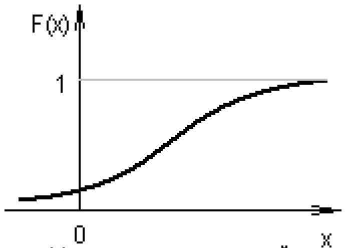
2) **Функция распределения случайной величины F(x)**, которая равна вероятности того, что случайная величина X в результате эксперимента примет значение, меньшее x. То есть $\underline{F(x) = P(X < x)}$.

Данную форму закона распределения случайной величины можно использовать как для непрерывной, так и для дискретной случайной величины.

Для дискретной случайной величины $F(x_i)$ находится следующим образом:



Дискретная случайная величина



Непрерывная случайная величина

$F(x_1) = P(X < x_1) = 0$

$F(x_2) = P(X < x_2) = P(x_1)$

$F(x_3) = P(X < x_3) = P(X = x_1) + P(X = x_2)$

$F(x_4) = P(X = x_1) + P(X = x_2) + P(X = x_3)$

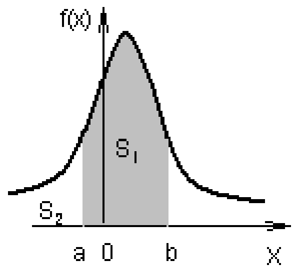
$F(x_5) = P(X = x_1) + P(X = x_2) + P(X = x_3)$
$+ P(X = x_4) = 1$

Свойства функции распределения случайной величины:

1) Функция распределения удовлетворяет неравенству: $0 \le F(x) \le 1$.
2) Функция распределения является неубывающей функцией, то есть если $x_1 < x_2$, то $F(x_1) \le F(x_2)$.
3) Вероятность того, что случайная величина в результате испытания примет значение, лежащее в интервале (a,b), равна приращению функции распределения на этом интервале, то есть $P(\beta < X < \alpha) = F(\alpha) - F(\beta)$.

Следует отметить, что функция распределения случайной величины не позволяет (как многоугольник распределения) наглядно представить, какие из своих значений непрерывная случайная величина принимает с большей вероятностью, а какие с меньшей. Для этого используется **функция плотности распределения случайной величины f(x),** являющейся дифференциальной кривой от функции F(x), то есть $f(x) = F'(x)$.

График функции *f(x)* называется *кривой распределения.* Из графика следует, что свои значения, лежащие в интервале (a,b), случайная величина принимает с большей вероятностью, чем какие-либо другие значения.



Свойства плотности распределения случайной величины *f(x):*

1) Плотность распределения случайной величины является неотрицательной функцией, так как несет смысл вероятности, то есть $f(x) \ge 0$.
2) Вероятность того, что в результате испытания непрерывная случайная величина примет значение, лежащее в интервале (a,b), равна определенному интегралу в пределах от a до b, от плотности распределения этой случайной величины, и равна площади криволинейной трапеции S_1 то есть

$$P(a < X < b) = \int_a^b f(x)dx = S_1 .$$

3) Определенный интеграл в пределах от -∞ до а от плотности распределения случайной величины равен функции распределения этой величины или площади криволинейной трапеции S_2, то есть $F(a)=\int\limits_{-\infty}^{a} f(x)dx = S_2$.

4) Условие нормировки, то есть определенный интеграл от -∞ до +∞ от плотности распределения случайной величины равен единице, то есть $\int\limits_{-\infty}^{+\infty} f(x)dx = 1$.

Следует отметить, что функция распределения случайной величины или ее плотность распределения полностью характеризуют данную случайную величину.

2. Числовые характеристики случайных величин.

Однако, в ряде случаев нет необходимости характеризовать случайную величину полностью, исчерпывающим образом. Поэтому были разработаны числовые характеристики случайной величины, отражающие основные особенности ее распределения.

1. *Математическое ожидание M(X)* – это центральная точка, вокруг которой рассеяны все значения случайной величины X.
   Математическое ожидание дает представление о среднем значении случайной величины. Размерность математического ожидания совпадает с размерностью самой случайной величины: $[M(X)] = [X]$.

   Найти математическое ожидание случайной величины X можно по следующим формулам:

   а) если X -- дискретная случайная величина, то $M(X) = \sum\limits_{i=1}^{n} x_i p_i$;

   б) если X - непрерывная случайная величина, то $M(X) = \int\limits_{-\infty}^{+\infty} x f(x)dx$.

   Свойства математического ожидания:
   1) $M(X+Y) = M(X) + M(Y)$, где X и $У$ – случайные величины.
   2) $M(X \cdot Y) = M(X) \cdot M(Y)$, где X и $У$ – случайные величины.
   3) $M(CX) = CM(X)$, где C – постоянная величина.

2. *Дисперсия D(X)* – это характеристика степени разброса случайной величины относительно ее среднего значения. Она находится как математическое ожидание квадрата отклонения случайной величины от ее математического ожидания: $D(X) = M(X - M(X))^2$.
   Размерность дисперсии D(X) равна размерности случайной величины, возведенной в квадрат: $[D(X)] = [X^2]$.

   Найти дисперсию случайной величины X можно по следующим формулам:

   а) если X -- дискретная случайная величина, то $D(X) = \sum\limits_{i=1}^{n} (x_i - M(x))^2 p_i$;

   б) если X - непрерывная случайная величина, то $D(X) = \int\limits_{-\infty}^{+\infty} (x - M(x))^2 f(x)dx$.

   Свойства дисперсии
   1) $D(X+Y) = D(X) + D(Y)$, где X и $У$ – случайные величины.

2) $D(CX) = C^2 D(X)$, где C – постоянная величина.

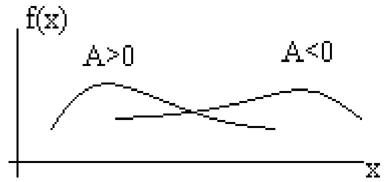
3) $D(X) = M(x^2) - (M(x))^2$.

Наряду с дисперсией в качестве числовой характеристики степени разброса возможных значений случайной величины относительно ее математического ожидания часто используют *среднее квадратическое отклонение* σ ($\sigma = \sqrt{D}$), размерность которого совпадает с размерностью случайной величины.

Математическое ожидание и дисперсия являются *основными числовыми характеристиками случайной величины*. Однако для уточнения вида распределения случайной величины могут использоваться и дополнительные характеристики. Такие как асимметрия, эксцесс, мода и медиана.

1) *Асимметрия* – характеристика «скошенности» распределения.

Она является безразмерной величиной и находится по формуле: $A = \dfrac{\mu_3}{\sigma^3}$,

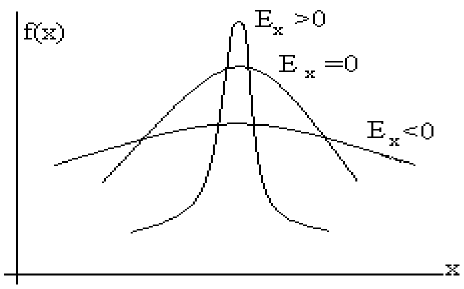


где $\mu_3 = \int\limits_{-\infty}^{+\infty} (x - M(X))^3 f(x)dx$ - для непрерывной случайной величины,

$\mu_3 = \sum\limits_{i=1}^{n} (x_i - M(X))^3 p(x_i)$ - для дискретной случайной величины.

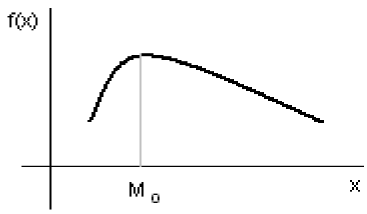
2) Эксцесс – характеристика «островершинности» распределения.

Она является безразмерной величиной и находится по формуле: $E_x = \dfrac{\mu_4}{\sigma^4} - 3$.



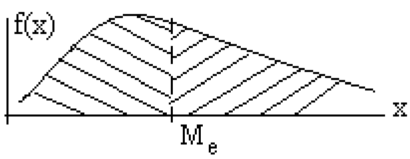
где $\mu_4 = \int\limits_{-\infty}^{+\infty} (x - M(x))^4 f(x)dx$ - для непрерывной случайной величины,

$\mu_4 = \sum\limits_{i=1}^{n} (x_i - M(x))^4 p(x_i)$ - для дискретной случайной величины.



3) *мода* – это наиболее вероятное значение случайной величины.

4) медиана – это значение случайной величины, которое делит площадь под функцией плотности распределения случайной величины *f(x)* пополам.



В механике формулы, через которые выражаются числовые характеристики случайных величин, называются начальными ($\alpha_k = \mu(X^k)$) или центральными

($\mu_k = M(X - M(X))^k = M(\dot{X}^k)$) моментами, а случайная величина $\dot{X} = X - M(X)$ - центрированной случайной величиной. Таким образом, математическое ожидание случайной величины является начальным моментом первого порядка (k=1), дисперсия – центральным моментом второго порядка (k=2), асимметрия находится через центральный момент третьего порядка (k=3), а эксцесс – через центральный момент четвертого порядка (k=4).

Биномиальный закон распределения .

Он применяется для дискретных случайных величин, если задача сводится к следующему типу: проведено *n* опытов, в каждом из которых появляется с одной и той же вероятностью *p* событие А. Тогда вероятность получить событие А в *m* опытах из *n* находится по формуле Бернулли: $P(x = m) = C_n^m p^m q^{n-m}$, где $q = 1 - p$ - вероятность не наступления события А в каждом из испытаний, а $C_n^m = \dfrac{n!}{m!(n-m)!}$. Математическое ожидание данного распределения и его дисперсию можно найти следующим образом: $M(X) = np$, $D(X) = npq$.

Пример: В поликлинике работает 7 участковых врачей. Вероятность заболеть гриппом для каждого из них составляет 0,6. Какова вероятность того, что во время эпидемии 5 из 7 заболеют?

Решение:

$$P(m = 5) = C_7^5 p^5 (1-p)^{7-5} = \frac{7!}{5! \cdot 2!} \cdot 0{,}6^5 \cdot 0{,}4^{7-5} \approx 0{,}26.$$

Распределение Пуассона.

В случаях, когда объем n серии независимых повторных испытаний велик, использование формулы Бернулли сопряжено со значительными трудностями (из-за необходимости оперировать большими числами). Однако, если объем испытании не менее нескольких десятков, а вероятность наступления случайного события А в каждом из испытаний мала (p<<1), причем математическое ожидание данного события $\mu = np$ не превышает 10, то используется формула Пуассона – «закона редких событий»: $P(x = m) = \dfrac{\mu^m}{m!} e^{-\mu}$. Распределение Пуассона является частным случаем биноминального закона.

Пример.

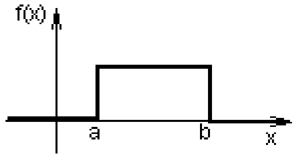
При перевозке 1000 стеклянных колб вероятность разбить 1 колбу равно 0,002. Какова вероятность, что будут разбиты 4 колбы?

Решение.

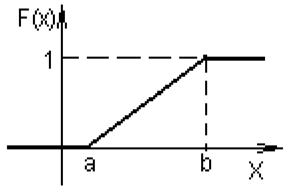
$$P(4) = \frac{(1000 \cdot 0{,}002)^4}{4! e^{1000 \cdot 0{,}002}} = \frac{2^4}{1 \cdot 2 \cdot 3 \cdot 4 \cdot e^2} \approx 0{,}08.$$

Равномерное распределение.

Если плотность распределения случайной величины не изменяется в определенном интервале ее значений и обращаются в ноль вне его, то говорят, что случайная величина распределена равномерно. Тогда ее плотность распределения и функция распределения будут иметь вид:



$$f(x)=\begin{cases}0, x>b, x<a\\ \dfrac{1}{b-a}, a\le x\le b\end{cases}.$$



$$F(x)=\int_{-\infty}^{x} f(x)dx=\int_{a}^{x}\frac{1}{b-a}dx=\frac{1}{b-a}x\Big|_{a}^{x}=\frac{x-a}{b-a}.$$

Найдем основные числовые характеристики данного распределения.

$$M(X)=\int_{a}^{b}\frac{x}{b-a}dx=\frac{1}{b-a}\cdot\frac{x^2}{2}\Big|_{a}^{b}=\frac{b^2-a^2}{2(b-a)}=\frac{b+a}{2}.$$

$$D(X)=\int_{a}^{b}\left(x-\frac{b+a}{2}\right)^2\frac{1}{b-a}dx=\frac{(b-a)^2}{12}.$$

<u>Пример.</u>

Случайная величина X равномерно распределена на [1, 6]. Требуется найти ее плотность распределения и основные числовые характеристики распределения.

<u>Решение:</u>

$$f(x)=\begin{cases}0, x>6, x<1\\ \dfrac{1}{6-1}=\dfrac{1}{5}, 1\le x\le 6\end{cases},\qquad M(X)=\int_{1}^{6}\frac{x}{5}dx=\frac{1}{5}\cdot\frac{x^2}{2}\Big|_{1}^{6}=\frac{6^2-1}{2\cdot 5}=\frac{35}{10}=3{,}5,$$

$$D(X)=\int_{1}^{6}(x-3{,}5)^2\frac{1}{5}dx=\frac{25}{12}=2{,}08$$

<u>Показательный (экспоненциальный) закон распределения.</u>

Непрерывная случайная величина имеет показательный закон распределения с параметром λ, если ее плотность распределения имеет вид:

$$f(x)=\begin{cases}\lambda e^{-\lambda x}, x\ge 0\\ 0, x\prec 0\end{cases}$$

Математическое ожидание и дисперсию данного распределения можно найти следующим образом: $M(X)=\dfrac{1}{\lambda}$, $D(X)=\dfrac{1}{\lambda^2}$.

<u>Пример.</u>

Случайная величина X распределена по показательному закону с параметром 5. Записать ее плотность распределения и найти его основные числовые характеристики.

<u>Решение:</u>

$$f(x)=\begin{cases}5e^{-5x}, x\ge 0\\ 0, x\prec 0\end{cases},\ M(X)=\frac{1}{5},\ D(X)=\frac{1}{25}.$$

<u>Нормальный закон распределения (для непрерывных случайных величин).</u>

Если <u>непрерывная</u> случайная величина X является результатом действия большого числа разнообразных факторов, то она подчиняется нормальному закону распределения:

$$F(x) = P(X < x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(x-M(x))^2}{2\sigma^2}} dx \quad \text{или} \quad f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-M(x))^2}{2\sigma^2}} ,$$

где $\sigma = \sqrt{D}$ - среднеквадратическое отклонение случайной величины от ее математического ожидания M(X).

Подобную функцию очень трудно представить в виде таблицы, т.к. много изменяющихся данных: $P(X < x) = f(\sigma, x, M(x))$, поэтому заменой переменной $\frac{x - M(x)}{\sigma} = t$ ее приводят к функции Лапласа: $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{0}^{x} e^{-\frac{t^2}{2}} dt$, зависящей только от одной переменной, для которой построены таблицы ее значений.

$$F(x_i) = P(X < x_i) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x_i} e^{-\frac{(x-M(x))^2}{2\sigma^2}} dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\frac{x_i - M(x)}{\sigma}} e^{-\frac{t^2}{2}} \sigma \cdot dt =$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} e^{-\frac{t^2}{2}} dt + \frac{1}{\sqrt{2\pi}} \int_{0}^{\frac{x_i - M(x)}{\sigma}} e^{-\frac{t^2}{2}} dt = 0,5 + \Phi\left(\frac{x_i - M(x)}{\sigma}\right).$$

Определив математическое ожидание и среднеквадратическое отклонение исследуемой случайной величины, можно, используя таблицу функции Лапласа, построить функцию распределения этой случайной величины.

Если требуется найти вероятность того, что случайная величина принимает значения, лежащие в интервале (α;β), то используется формула:

$$P(\alpha < x < \beta) = F(\beta) - F(\alpha) = \Phi\left(\frac{\beta - M(x)}{\sigma}\right) - \Phi\left(\frac{\alpha - M(x)}{\sigma}\right)$$

Следствием такого представления нормальной величины является правило трех сигм: *для нормально распределенной случайной величины все распределение с точностью до долей процента укладывается на участок* $M(x) \pm 3\sigma$.

<u>Доказательство:</u> $P(M(x) - 3\sigma < x < M(x) + 3\sigma) = \Phi\left(\frac{M(x) + 3\sigma - M(x)}{\sigma}\right) - \Phi\left(\frac{M(x) - 3\sigma - M(x)}{\sigma}\right) =$

$\Phi(3) - \Phi(-3) = 2\Phi(3) \approx 0,9973$.