1.3. Случайные величины и их характеристики

Случайной величиной (переменной) называется величина, которая под воздействием случайных факторов может с определенными вероятностями принимать те или иные значения из некоторого множества чисел.

Случайные величины обозначаются заглавными буквами латинского алфавита: X, Y, Z и т.д., а их возможные значения – строчными: $x_1, x_2, \ldots, x_n$, $y_1, y_2, \ldots, y_n$, $z_1, z_2, \ldots, z_n$ и т д..

Для полной характеристики случайной величины должны быть указанны не только все её значения, но и их вероятности.

Универсальным способом задания случайной величины X является задание её с помощью функции распределения.

Функцией распределения $F(x)$ случайной величины X называется вероятность того, что величина X принимает значения меньшие, чем x, т.е.

$$F(x) = P(X < x), \quad x \in R.$$

Различают *дискретные и непрерывные* случайные величины.

Дискретной называется случайная величина, которая может принимать лишь отдельные, причем изолированные друг от друга, значения из своего диапазона изменения. Число возможных значений дискретной случайной величины *конечно* или *счетно.*

Дискретную случайную величину удобнее задавать не в виде функции распределения, а в виде ряда распределения.

При табличном задании ряда распределения первая строка таблицы содержит возможные значения случайной величины, а вторая – соответствующие им вероятности, т.е.:

$$\begin{pmatrix} x_1 & x_2 & \ldots x_n \\ p_1 & p_2 & \ldots p_n \end{pmatrix}, \text{ где } p_i = P(X < x_i),$$

$$\sum p_i = 1.$$

Графическое изображение ряда распределения называется полигоном распределения.

Непрерывной называется случайная величина, которая может принимать любые значения из своего диапазона изменения. Число возможных значений непрерывной случайной величины бесконечно.

Задать непрерывную случайную величину рядом распределения невозможно, потому что она принимает бесконечное множество значений, поэтому её задают с помощью функции распределения $F(x)$.

Вместо функции распределения *F(x)* при задании непрерывной случайной величины часто используют функцию плотности распределения вероятностей *f(x)*.

Функцией плотности распределения вероятностей *f(x)* непрерывной случайной величины называется производная от функции распределения, т.е. $f(x)=F'(x)$.

Из определения производной вытекает вероятностный смысл функции плотности распределения вероятностей:

$$f(x)=F'(x)=\lim_{\Delta x\to 0}\frac{F(x+\Delta x)-F(x)}{\Delta x}=\lim_{\Delta x\to 0}\frac{p(x<X<x+\Delta x)}{\Delta x},$$

то есть предел отношения вероятности попадания случайной величины X в интервал $(x, x+\Delta x)$, к длине этого интервала, при $\Delta x\to 0$, равен значению плотности распределения вероятностей *f(x)*.

Из определения функции плотности распределения вероятностей следует, что функция распределения *F(x)* является первообразной для плотности распределения вероятностей *f(x)*.

Свойства функции плотности распределения вероятностей:

1. $f(x)\geq 0$ при любых $x\in R$.
2. $P(x_1\leq X\leq x_2)=\int\limits_{x_1}^{x_2} f(x)dx$.
3. $\int\limits_{-\infty}^{\infty} f(x)dx=1$.

1.4. Генеральная и выборочная совокупности и их характеристики

В основе математической статистики лежат понятия генеральной и выборочной совокупностей.

Генеральная совокупность – это множество всех значений (исходов) случайной величины, которые она может принять в процессе наблюдения. Например, данные о доходах всех жителей страны.

Для любой случайной величины важную роль помимо функции распределения играют числовые характеристики её распределения.

Числовая характеристика случайной величины – это число, выражающее одно из важных ее характеристик. В курсе эконометрики наиболее часто используются такие числовые характеристики, как **математическое ожидание** и **дисперсия**.

Математическим ожиданием дискретной случайной величины называется сумма произведений всех ее значений на соответствующие им вероятности, т.е.

$$M(X) = \sum x_i \cdot p_i ,$$

где суммирование осуществляется по всем возможным значениям случайной величины.

Математическое ожидание непрерывной случайной величины определяется выражением:

$$M(X) = \int_{-\infty}^{\infty} x \cdot f(x)dx,$$

где интегрирование осуществляется по интервалу, на котором определена $f(x)$.

Математическое ожидание случайной величины - это *среднее* ее значение по генеральной совокупности, обозначается $M(X) = \mu_x$.

Геометрически математическое ожидание случайной величины - это центр ее распределения.

Свойства математического ожидания (a,b-константы; X, Y - случайные величины):

1. $M(a) = a$.
2. $M(b{\cdot}X) = b\,M(X)$.
3. $M(a + b{\cdot}X) = a + b\,M(X)$.
4. $M(X + Y) = M(X) + M(Y)$.
5. $M(X - \mu_x) = 0$.

Математическое ожидание функции $g(X)$ определяется выражением:

$$M(g(X)) = \sum g(x_i)p_i ,$$

где суммирование осуществляется по всем возможным значением x_i.

В частности, если $g(X) = X^2$, то $M(X^2) = \sum x_i^2 \cdot p_i$.

Пример 1.1. Известно, что математическое ожидание случайной величины X равно $M(X)=2$. Найти математическое ожидание случайной величины $Y = 3{\cdot}X+5$.

Используя свойства математического ожидания, получаем:

$$M(Y) = M(3{\cdot}X+5)=M(3{\cdot}X)+M(5)=3{\cdot}M(X)+5=3{\cdot}2+5=11.$$

Пример 1.2. Вычислить математическое ожидание дискретной случайной величины X, заданной рядом распределения (таблица 1.1):

Таблица 1.1

x_i	-1	2	5	10	20
p_i	0,1	0,2	0,3	0,3	0,1

Используем формулу для вычисления математического ожидания дискретной случайной величины:

$$M(X)=\sum x_i \cdot p_i \,,$$

получаем: $M(X)=\sum x_i \cdot p_i$ = $-1\cdot 0{,}1+2\cdot 0{,}2+5\cdot 0{,}3+10\cdot 0{,}3+20\cdot 0{,}1=6{,}8$.

Пример 1.3. Найти математическое ожидание случайной величины X, распределенной непрерывно с плотностью распределения вероятностей: $f(x)=12\cdot(x^2-x^3)$ когда $x\in(0,1)$ и $f(x)=0$ при остальных значениях x.

Используем для нахождения математического ожидания формулу:

$$M(X)=\int_{-\infty}^{+\infty} f(x)\cdot x dx,$$

Подставляем в эту формулу из условия задачи функцию плотности распределения вероятностей и вычисляем значение интеграла:

$$M(X)=\int_{-\infty}^{+\infty} f(x)\cdot x dx=\int_{0}^{1} 12\cdot(x^2-x^3)\cdot x dx \ =\int_{0}^{1} 12\cdot(x^3-x^4)\ dx=(3\cdot x^4-\frac{12}{5}\cdot x^5)\Big|_0^1 \ =$$

$$=(3\cdot 1^4-\frac{12}{5}\cdot 1^5)-(3\cdot 0^4\,\frac{12}{5}\cdot 0^5)=3\ -\frac{12}{5}=\frac{15-12}{5}=\frac{3}{5}=0{,}6.$$

Определение. Случайные величины X, Y называются **независимыми**, если $P(X=x;\ Y=y)=P(X=x)\cdot P(Y=y)$ для любых значений x, y.

Следствие. Если случайные величины X, Y независимы, то

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

$M[(X-\mu_x)(Y-\mu_x)]=0$.

Теоретическая (генеральная) дисперсия случайной величины определяется как математическое ожидание квадрата отклонения случайной величины относительно ее средней, т.е.:

$$\sigma_x^2=D(X)=M(X-\mu_x)^2.$$

Замечание. *Если ясно, о какой переменной идет речь, нижний индекс в* μ_x *или* σ_x^2 *можно не указывать.*

Для вычисления дисперсии часто используется другое выражение, получаемое из определения дисперсии:

$$D(X)=M(X^2)-\mu_x^2.$$

Дисперсия является *мерой рассеяния значений* случайной величины относительно ее среднего значения (центра). Размерность дисперсии равняется квадрату размерности случайной величины.

Стандартным отклонением случайной величины называется корень квадратный из ее дисперсии, т.е.

$$\sigma_x = \sqrt{D(X)}\,.$$

Стандартное отклонение показывает, насколько в среднем отклоняется случайная величина в совокупности относительно своего среднего значения.

Свойства дисперсии (*a*, *b* - константы; *X*, *Y* - случайные величины):

1. $D(a) = 0$.
2. $D(b{\cdot}X) = b^2{\cdot}D(X)$.
3. $D(a + b{\cdot}X) = b^2{\cdot}D(X)$.

Следствие. Если случайные величины *X*, *Y* независимы, то
$D(X + Y) = D(X) + D(Y)$.

Заметим, что $M(X)$ и $D(X)$ - это числовые характеристики генеральной совокупности (числа), а не функции.

Пример 1.4. Вычислить и сравнить дисперсии двух случайных величин *X* и *У* со следующими законами распределения (таблица 1.2):

Таблица 1.2

x_i	1	2
p_i	0,5	0,5
y_i	-10	10
p_i	0,5	0,5

Для наглядности расчетов взяты простые распределения с двумя значениями и одинаковыми вероятностями. В первом случае значения случайной величины расположены рядом (*1* и *2*), а во втором – дальше друг от друга (*-10* и *10*). А теперь посмотрим, насколько различаются дисперсии:

$$D(X) = \sum_{i=1}^{n} x_i^2 \cdot p_i - \left(\sum_{i=1}^{n} x_i \cdot p_i\right)^2 = 1^2 \cdot 0{,}5 + 2^2 \cdot 0{,}5 - (1 \cdot 0{,}5 + 2 \cdot 0{,}5)^2 = 2{,}5 - 1{,}5^2 = 0{,}25\,;$$

$$D(Y) = \sum_{i=1}^{n} y_i^2 \cdot p_i - \left(\sum_{i=1}^{n} y_i \cdot p_i\right)^2 = (-10)^2 \cdot 0{,}5 + 10^2 \cdot 0{,}5 - (-10 \cdot 0{,}5 + 10 \cdot 0{,}5)^2 = 100 - 0^2 = 100\,.$$

Итак, значения случайных величин различались на *1* и *20* единиц, тогда как дисперсии показывают меру разброса в *0,25* и *100*. Если перейти к среднеквадратическому отклонению, получим $\sigma(X)=0{,}5$; $\sigma(Y)=10$, то есть вполне ожидаемые величины: в первом случае значения отстоят в обе стороны на *0,5* от среднего *1,5;* а во втором – на *10* единиц от среднего *0*.

Ясно, что для других распределений, где число значений больше и вероятности не одинаковы, картина будет другой, прямой зависимости от значений случайной величины уже не будет (но будет как раз оценка разброса).

Пример 1.5. Найти дисперсию дискретной случайной величины X, заданной рядом распределения, приведенным в таблице 1.3.

Таблица 1.3

x_i	-1	2	5	10	20
p_i	0,1	0,2	0,3	0,3	0,1

Используем формулу для вычисления дисперсии дискретной случайной величины:

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

В случае, когда вычислительных действий много, удобно выполнять их по шагам. Сначала найдем математическое ожидание случайной величины X:

$$M(X)=\sum_{i=1}^{n} x_i \cdot p_i = -1\cdot 0{,}1+2\cdot 0{,}2+5\cdot 0{,}3+10\cdot 0{,}3+20\cdot 0{,}1=6{,}8\,.$$

Потом математическое ожидание случайной величины X^2:

$$M(X^2)=\sum_{i=1}^{n} x_i^2 \cdot p_i = (-1)^2\cdot 0{,}1+2^2\cdot 0{,}2+5^2\cdot 0{,}3+10^2\cdot 0{,}3+20^2\cdot 0{,}1=78{,}4.$$

А затем подставим вычисленные значения в формулу для дисперсии и получим:

$$D(X)=M(X^2)-M(X)^2=78{,}4-6{,}8^2=32{,}16.$$

Пример 1.6. Найти дисперсию непрерывной случайной величины X, имеющей функцию плотности распределения вероятностей: $f(x)=x/18$ при $x\in(0,\,6)$ и $f(x)=0$ при остальных значениях x.

Используем для расчета формулу дисперсии непрерывной случайной величины:

$$D(X)=\int_{-\infty}^{+\infty} f(x)\cdot x^2 dx-\left(\int_{-\infty}^{+\infty} f(x)\cdot x dx\right)^2.$$

Вычислим сначала математическое ожидание:

$$M(X)=\int_{-\infty}^{+\infty} f(x)\cdot x dx=\int_0^6 \frac{x}{18}\cdot x dx=\int_0^6 \frac{x^2}{18} dx=\left.\frac{x^3}{54}\right|_0^6=\frac{6^3}{54}=4.$$

Далее вычислим

$$M(X^2)=\int_{-\infty}^{+\infty} f(x)\cdot x^2 dx=\int_{-\infty}^{+\infty} \frac{x}{18}\cdot x^2 dx=\int_0^6 \frac{x^3}{18} dx=\left.\frac{x^4}{72}\right|_0^6=\frac{6^4}{72}=18.$$

Подставляем вычисленные значения $M(X)$ и $M(X^2)$ в формулу для расчета дисперсии и получаем:

$D(X) = M(X^2) - M(X)^2 = 18 - 4^2 = 2.$

Пример 1.7. Известно, что дисперсия случайной величины X равна $D(X)=2$. Найти дисперсию случайной величины $Y = 4\cdot X+1$.

Используя свойства дисперсии, находим:

$D(Y) = D(4\cdot X+1)=D(4\cdot X) + D(1)=4^2\cdot D(X) + 0 = 16\cdot D(X) = 16\cdot 2 = 32.$

Наиболее часто используемые в эконометрике законы распределения случайных величин

Нормальный закон («закон Гаусса» или нормальное распределение) играет исключительную роль в теории вероятностей и математической статистике. Главная особенность закона Гаусса состоит в том, что он является предельным законом, к которому приближаются, при определенных условиях, другие законы распределения случайных величин.

Определение: непрерывная случайная величина X распределена **по нормальному закону** с параметрами μ и σ >= 0, если ее функция плотности распределения вероятностей имеет вид:

$$f(x)=\frac{1}{\sigma\cdot\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}},\quad x\in R.$$