Теория вероятностей и математическая статистика

Кракашова Ольга Анатольевна

доцент, канд. экон. наук,

доцент кафедры Статистики, эконометрики и оценки рисков РГЭУ (РИНХ)

Лекция № 5

Случайные величины и их характеристики

Понятие случайной величины

- **Случайная величина (СВ) - это величина, которая в результате эксперимента (опыта, испытания) принимает одно из своих возможных значений, причем заранее неизвестно, какое именно.**

Примеры случайных величин:

Эксперимент	Случайная величина	Возможные значения случайной величины
Контроль качества 70 деталей	Число дефектных деталей	0, 1, 2, ..., 70
Строительство жилого дома	Процент завершенного строительства спустя 6 месяцев	$0 \le x \le 100$
Проверка степени загрузки операционного отдела банка	Число клиентов в течение рабочего дня	0, 1, 2, ..., n
Торговля автомобилями	Число продаж в течение месяца	0, 1, 2, ..., n

Случайные величины обозначаются заглавными латинскими буквами: $\boldsymbol{X}$, $\boldsymbol{Y}$, $\boldsymbol{Z}$ и т.п. Прописные буквы используются для обозначения определенных значений случайной величины. Например, случайная величина $\boldsymbol{X}$ принимает значения $x_1, x_2, \dots x_n$

Различают случайные величины дискретные и непрерывные.

Дискретной (прерывной) называют случайную величину (ДСВ), которая принимает конечное или бесконечное (но счетное) число отдельных, изолированных возможных значений с определенными вероятностям.

Примеры дискретных СВ:

1) численный состав семьи;

2) число студентов в группе;

3) число зерен в колосе.

Характерной особенностью дискретных случайных величин (**ДСВ**) является тот факт, что **все возможные значения ДСВ можно заранее перечислить и представить в виде отдельных точек на числовой оси**.

Непрерывной СВ называется такая НСВ, которая может принимать **любые значения некоторого промежутка** (своего для каждой СВ).

Примеры непрерывных СВ :

1) расход горючего на единицу расстояния;

2) вес зерна пшеницы;

3) время безаварийной работы устройства;

4) количество осадков, выпавших в сутки.

Характерным для непрерывной СВ является то, что все возможные значения НСВ заполняют некоторый интервал, **перечислить все возможные значения непрерывной СВ невозможно, а можно только указать границы, в которых они заключены.**

Закон распределения случайных величин. Ряд распределения

Для дискретных случайных величин мы можем перечислить все возможные значения, которые эта величина может принять, и их можно представить в виде отдельных точек на числовой оси. Эта совокупность значений может быть задана таблицей, функцией или графиком.

Соотношение, устанавливающее связь между отдельными возможными значениями случайной величины и соответствующими им вероятностями, называется <u>законом распределения</u> дискретной случайной величины.

Простейшей формой закона распределения для дискретных случайных величин является **<u>ряд распределения</u>.**

Рядом распределения ДСВ X называется таблица, в которой перечислены возможные (различные) значения этой случайной величины $x_1, x_2, \dots x_n$ с соответствующими им вероятностями

$p_1, p_2, \dots p_n$:

x_i	x_1	x_2	...	x_n
p_i	p_1	p_2	...	p_n

или

X	x_1	x_2	...	x_n
p_i	p_1	p_2	...	p_n

Таким образом, случайная величина X в результате испытания может принять одно из возможных значений $x_1, x_2, \dots x_n$ с вероятностями

$P(X=x_1) = p_1, P(X=x_2) = p_2, \dots, P(X=x_n) = p_n$

Запись $P(X=x)$ для дискретных случайных величин означает вероятность того, что случайная величина X примет ***определенное значение*** x. Например, запись $P(X=5) = 0{,}2$ означает, что вероятность того, что случайная величина X примет значение, равное 5, есть 0,2. Мы можем также использовать более короткую запись: $P(x)$ вместо $P(X=x)$, или $P(5) = 0{,}2$.

Так как события $(X=x_1)$, $(X=x_2)$ ... $(X=x_n)$ составляют полную группу событий, то сумма вероятностей p_1, p_2, ... p_n равна единице :

$$\sum_{i=1}^{n} p_i = 1$$

Ряд распределения дискретной случайной величины должен удовлетворять следующим условиям:

1. $P(x) \geq 0$
2. $\sum_{i=1}^{n} P(x) = 1$ (или $\sum_{i=1}^{n} p_i = 1$) (1)

Пример 1. Каждый день местная газета получает заказы на новые рекламные объявления, которые будут напечатаны на следующий день. Число рекламных объявлений в газете зависит от многих факторов: дня недели, сезона, общего состояния экономики, активности местного бизнеса и т.д. Пусть X - число новых рекламных объявлений, напечатанных в местной газете в определенный день.

X - случайная величина, которая может быть только целым числом.

Пусть в нашем примере случайная величина X принимает значения 0, 1, 2, 3, 4, 5 с вероятностями 0,1; 0,2; 0,3; 0,2; 0,1; 0,1 соответственно.

Таблица 1

Ряд распределения случайной величины X – числа рекламных объявлений

x_i	0	1	2	3	4	5
$P(X)=p_i$	0,1	0,2	0,3	0,2	0,1	0,1

Таблица 1 говорит нам о многом. Например, мы видим: вероятность того, что в определенный день будут напечатаны 3 объявления, равна 0,2, а 2 объявления - 0,3 и т.д. Поскольку появления различных значений случайной величины X - несовместные события, то вероятность того, что в газету будут помещены или 2 или 3 рекламных объявления, - равна сумме вероятностей $\boldsymbol{P(2) + P(3) = 0{,}3 + 0{,}2 = 0{,}5}$. Вероятность же того, что их число будет находиться в пределах от 1 до 4 (включая 1 и 4), равна 0,8, то есть $\boldsymbol{P(1 \leq X \leq 4) = 0{,}8}$; а $\boldsymbol{P(X = 0) = 0{,}1}$.

Ряд распределения можно изобразить графически.

Для этого по оси абсцисс откладывают возможные значения случайной величины, а по оси ординат соответствующие им вероятности. Если точки $(x_i;p_i)$ соединить отрезками прямых, то полученная ломаная линия есть ***многоугольник (или полигон)*** распределения.

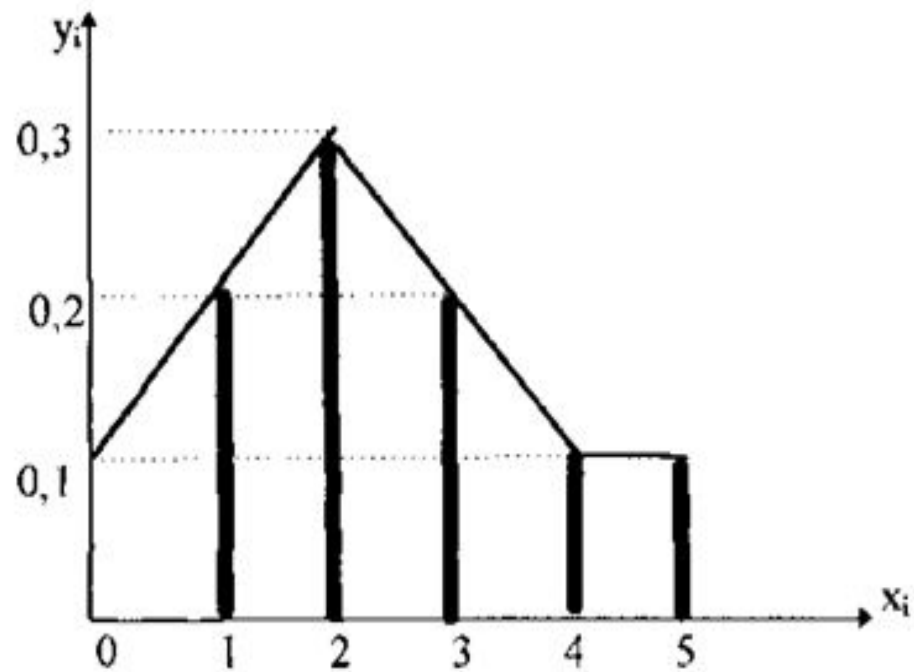

Рис.1. Полигон распределения для данных примера 1

Пример 2. В книжном магазине организована лотерея. Разыгрываются две книги стоимостью по 10 рублей и одна - стоимостью в 30 рублей. Составить закон распределения случайной величины *X* - суммы чистого (возможного) выигрыша для того, кто приобрел один билет за 1 рубль, если всего продано 50 билетов.

Случайная величина *X* может принимать три значения: - 1 руб. (если владелец билета не выиграет, а фактически проиграет 1 руб., уплаченные им за билет); 9 руб.; 29 руб. (фактический выигрыш уменьшается на стоимость билета- 1 руб.) Первому результату благоприятствуют 47 исходов из 50, второму - два, а третьему - один.

Поэтому их вероятности таковы:

$$P(X=-1)=\frac{47}{50}=0{,}94;\quad P(X=9)=\frac{2}{50}=0{,}04;\quad P(X=29)=\frac{1}{50}=0{,}02$$

Закон распределения случайной величины X имеет вид

Сумма выигрыша (X)	- 1	9	29
Вероятность (P)	0,94	0,04	0,02

Контроль: $\sum_{i=1}^{3} p_i = \mathbf{0{,}94 + 0{,}04 + 0{,}02 = 1}$

Пример 3. В условиях лотереи, описанной в примере 2., ***два*** посетителя магазина приобрели по одному билету стоимостью по 1 руб. Составить закон распределения суммы выигрыша для второго посетителя, если первый выиграл книгу стоимостью в 30 рублей.

Если первый посетитель выиграл книгу стоимостью в 30 руб., то второй посетитель может или не выиграть, или выиграть книгу стоимостью в 10 руб. Первому событию благоприятствуют 47 из оставшихся 49 исходов, второму - 2 исхода.

Следовательно, закон распределения случайной величины У (сумма выигрыша второго посетителя) при условии, что случайная величина Х (сумма выигрыша первого посетителя) приняла значение 29 (Х = 29), имеет вид;

Сумма выигрыша (Х)	- 1	9
Вероятность (Р)	47/49	2/49

Функция распределения (интегральная функция распределения)

При анализе экономических явлений определенный смысл имеют ***кумулятивные*** (накопленные) вероятности случайных величин. Нас может интересовать вероятность того, что число проданных единиц некоторого товара окажется не меньше некоторого определенного числа, гарантирующего прибыль продавцу, вероятность того, что суммы возможных убытков от рискованных инвестиций окажутся не выше (или только меньше) некоторого определенного значения и т.д.

Знаю закон распределения дискретной случайной величины, можно составить функцию накопленных вероятностей. Определим ***интегральную (кумулятивную) функцию распределения***.

Функцией распределения дискретной случайной величины называется функция $F(x)$, определяющая для каждого значения x вероятность того, что случайная величина X примет значение, меньшее x, то есть

$$F(x) = P(X < x) \qquad (2)$$

геометрически:

ДСВ *X* можно рассматривать как случайную точку на числовой оси.

Пусть на оси выбрана конкретная точка *x*, тогда в результате опыта случайная точка X может оказаться левее или правее выбранной нами точки. Очевидно, что вероятность того, что случайная точка *X* окажется левее точки *x* будет зависеть от положения точки *x*, то есть являться функцией аргумента *x*.

Для ДСВ **X**, которая может принимать значения $x_1, x_2, \ldots, x_n$, функция распределения примет вид

$$\boxed{F(X) = \sum_{x_i < x} P(X = x_i)} \qquad (3)$$

Где неравенство $x_i < x$ под знаком суммы означает, что суммирование касается тех значений x_i величина которых меньше ***x***. Поясним эту формулу, исходя из определения *F(x)*. Предположим, что аргумент ***x*** принял какое-то определенное значение, но такое, что выполняется неравенство $x_i < x < x_{i+1}$. Тогда левее числа ***x*** на числовой оси окажутся только те значения, которые имеют индекс 1,2,3,…,*i*. Поэтому неравенство *X*<*x* выполняется, если величина *X* примет значение x_k где *k*=1,2,3,…,*i* . Таким образом, событие *X*<*x* наступит, если наступит любое, неважно какое, из событий $X=x_1$, $X=x_2$, $X=x_3$, …, $X=x_i$. Так как эти события несовместные, то по теореме сложения вероятностей имеем

$$P(X < x) = P(X = x_1) + P(X = x_2) + P(X = x_3) + \ldots + P(X = x_i) = \sum_{x_i < x} P(X = x_i).$$

Пример 4. Для примера 1 найти функцию распределения случайной величины X - числа рекламных объявлений, построить её график.

Случайная величина X не принимает значений, меньших 0. Следовательно, если $x \leq 0$, то событие $X < x$ - невозможно, а вероятность его равна нулю. Поэтому функция распределения случайной величины X для всех значений $x \leq 0$ также равна 0. Для всех $\boldsymbol{x}$, удовлетворяющих двойному неравенству $0 < \boldsymbol{x} \leq 1$, функция $\boldsymbol{F(x)}$ означает вероятность события $\boldsymbol{X} < 0,2$. Но случайная величина X имеет вероятность меньшую 0,2, лишь в одном случае: значение 0 с вероятностью 0,1.

Покажем, что для всех $\boldsymbol{x}$, удовлетворяющих двойному неравенству $1<\boldsymbol{x}\leq 2$, $F(x) = 0,1 + 0,2 = 0,3$. Пусть, например, $x = 2$. Тогда $F(2)$ выражает вероятность события $X < 2$. Это возможно в двух случаях: или случайная величина X принимает значение 0 (с вероятностью 0,1), или 1 (с вероятностью 0,2). Применяя теорему сложения вероятностей, мы и получим указанное значение функции $F(x)$ при $x=2$.

Аналогичные рассуждения позволяют найти функцию распределения. Запишем ее в табличной форме.

Таблица 3

Функция распределения (интегральная функция распределения) для примера 3

x	x≤0	0<x≤1	1<x≤2	2<x≤3	3<x≤4	4<x≤5	x>5
F(x)	0	0,1	0,3	0,6	0,8	0,9	1

или F(x) можно записать так:

$$\mathbf{F(x)}=\begin{cases} 0 & \text{при} \quad x \le 0, \\ 0{,}1 & \text{при} \quad 0 < x \le 1, \\ 0{,}3 & \text{при} \quad 1 < x \le 2, \\ 0{,}6 & \text{при} \quad 2 < x \le 3, \\ 0{,}8 & \text{при} \quad 3 < x \le 4, \\ 0{,}9 & \text{при} \quad 4 < x \le 5, \\ 1 & \text{при} \quad x > 5 \end{cases}$$

Построим график интегральной функции распределения (вероятностную гистограмму) - Рис.2, который показывает, что интегральная функция - неубывающая и равна единице при x большем наибольшего возможного значения случайной величины. В нашем примере график $F(x)$ имеет ступенчатый вид.

Функция распределения каждой дискретной случайной величины постоянна на интервалах, на которых нет ее значений и имеет скачки в точках, соответствующих ее значениям. Скачки равны вероятностям, с которыми случайная величина принимает свои значения. Сумма всех скачков равна 1.

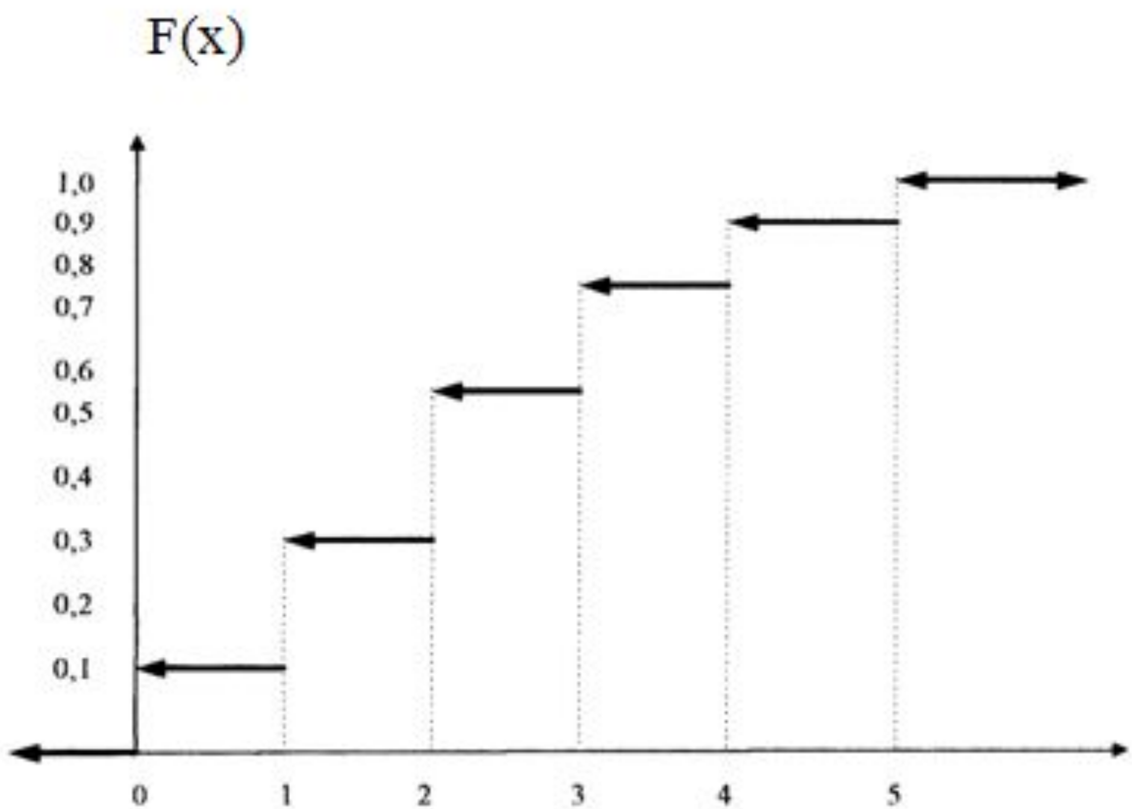

Рис.2. График интегральной функции для примера 1

Независимость случайных величин и математические операции над случайными величинами

Если рассматривать не одну, а две или более случайных величин (системы случайных величин), то необходимо знать, изменяется или не изменяется закон распределения одной из них в зависимости от того, какое значение принимают другие случайные величины.

Если закон распределения одной случайной величины не зависит от того, какие возможные значения приняли другие случайные величины, то такие случайные величины называются независимыми в совокупности.

Если закон распределения одной случайной величины зависит от того, какие возможные значения приняли другие случайные величины, то такие случайные величины называются зависимыми в совокупности.

Пусть случайная величина X принимает значения: x_1, x_2, ..., x_n с вероятностью $p_1, p_2, ..., p_n$, а случайная величина Y принимает значения y_1, y_2, ..., y_m с вероятностями q_1, q_2, ..., q_m.

Определим некоторые операции над случайными величинами.

1. Произведение случайной величины X на постоянную величину C есть случайная величина CX, которая принимает значения Cx_1, Cx_2, ... Cx_n с теми же вероятностями, что и случайная величина X.

2. Квадрат случайной величины X, то есть X^2 - это случайная величина, которая принимает свои значения $x_1^2, x_2^2, ..., x_n^2$ с теми же вероятностями, что и случайная величина X.

3. Суммой случайных величин X и Y называется случайная величина X+Y, возможные значения которой равны суммам каждого возможного значения X с каждым возможным значением Y, а вероятности возможных значений X+Y для независимых величин X и Y равны произведению вероятностей слагаемых; для зависимых величин - произведениям вероятности одного слагаемого на условную вероятность второго.

4. Произведением независимых случайных величин X и Y называется случайная величина XY, возможные значения которой равны произведениям каждого возможного значения X на каждое возможное значение Y, а вероятности возможных значений произведения XY равны произведениям вероятностей возможных значений сомножителей.

Ожидаемое значение дискретной случайной величины

Математическое ожидание дискретной случайной величины X равно сумме произведений значений случайной величины на соответствующие им вероятности

$$\boxed{M(x) = \sum_{i=1}^{n} x_i P(x_i) = \sum_{i=1}^{n} x_i p_i} \qquad (4)$$

Свойства математического ожидания дискретной случайной величины

1. Математическое ожидание постоянной величины равно этой постоянной

$$\boxed{M(c) = c} \qquad (5)$$

2. Постоянный множитель можно выносить за знак математического ожидания, то есть

$$\boxed{M(cX) = cM(X)} \qquad (6)$$

где c - постоянная величина

3. Математическое ожидание алгебраической суммы конечного числа n случайных величин равно алгебраической сумме их математических ожиданий, то есть

$$M(X_1 \pm X_2 \pm \cdots \pm X_n) = M(X_1) \pm M(X_2) \pm \cdots \pm M(X_n) \qquad (7)$$

Свойства математического ожидания дискретной случайной величины

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

$$M(X_1 \cdot X_2 \cdots X_n) = M(X_1) \cdot M(X_2) \cdots M(X_n) \qquad (8)$$

5. Если все значения случайной величины X уменьшить (увеличить) на одно и то же число c, то ее математическое ожидание уменьшится (увеличится) на то же число c, то есть

$$M(X - C) = M(X) - C \qquad (9)$$

Следствие. Математическое ожидание отклонений значений случайной величины X от ее математического ожидания равно нулю, то есть

$$M[X - M(X)] = 0 \qquad (10)$$

Свойства математического ожидания дискретной случайной величины

6. Математическое ожидание среднего арифметического значения n ***одинаково распределенных взаимно независимых***[1] случайных величин равно математическому ожиданию каждой из величин, то есть

$$M(X) = M(X_i) \qquad (11)$$

Пусть $X_1, X_2, \ldots, X_n$ - одинаково распределенные случайные величины, математические ожидания каждой из которых одинаковы и равны a. Тогда математическое ожидание их суммы равно $n \cdot a$ и математическое ожидание средней арифметической равно a

$$M(X) = 1/n \cdot M(X_1 + X_2 + \ldots + X_n) = n\,a/n = a$$

$$M(X) = a$$

Дисперсия и среднее квадратическое отклонение дискретной случайной величины

Дисперсия случайной величины есть математическое ожидание квадрата отклонения значений случайной величины от ее математического ожидания. Дисперсия дискретной случайной величины задается как:

$$\boxed{\sigma^2 = D(X) = M[(X - M(X))^2] = \sum_{i=1}^{n} [x_i - M(X)]^2 P(x_i)} \qquad (12)$$

По формуле (12) дисперсия вычисляется путем вычитания математического ожидания из каждого значения случайной величины, затем возведением в квадрат результатов, умножением их на вероятности $P(x_i)$ и сложением результатов для всех x_i. Для примера о числе рекламных объявлений, размещаемых в газете в определенный день, дисперсия вычисляется так:

$$\sigma^2 = \sum_{i=1}^{n} [x_i - M(X)]^2 P(x_i) = (0-2,3)^2 + (1-2,3)^2 + (2-2,3)^2 +$$

$$+(3-2,3)^2 + (4-2,3)^2 + (5-2,3)^2 = 2,01.$$

Свойства дисперсии дискретной случайной величины

1. Дисперсия постоянной величины равна нулю, то есть

$$D(c) = 0 \qquad (13)$$

2. Постоянный множитель можно выносить за знак дисперсии, возведя его при этом в квадрат, то есть

$$D(cX) = c^2 D(X), \qquad (14)$$

где c - постоянная величина

3. Дисперсия суммы (разности) конечного числа n независимых случайных величин равна сумме их дисперсий, то есть

$$D(X_1 \pm X_2 \pm \ldots \pm X_n) = D(X_1) \pm D(X_2) \pm \ldots \pm D(X_n) \qquad (15)$$

Свойства дисперсии дискретной случайной величины

4. Если $X_1, X_2, \ldots, X_n$ - одинаково распределенные независимые случайные величины, дисперсия каждой из которых равна σ^2, то дисперсия их суммы равна $n\sigma^2$, а дисперсия средней арифметической равна σ^2/n, то есть

$$D(X) = \sigma^2/n \qquad (16)$$

Для вычисления дисперсии проще пользоваться другой формулой, которая получается из формулы (12) путем несложных математических выкладок.

$$\mathbf{D(X)=M[(X-M(X))^2]=M[X^2-2M(X)X+(M(X))^2]=M(X)^2-2M(X)M(X)+}$$
$$\mathbf{+(M(X))^2=M(X^2)-[M(X)]^2=M(X^2)-M^2(X)}$$

Формула для упрощенного вычисления дисперсии дискретной случайной величины

$$\sigma^2 = D(X) = M(X^2) - M^2(X) \qquad (17)$$

Вычислим дисперсию случайной величины для примера 1, используя этот способ. Результаты оформим в виде рабочей таблицы.

x	P(x)	xP(x)	$x^2P(x)$
0	0,1	0	0
1	0,2	0,2	0,2
2	0,3	0,6	1,2
3	0,2	0,6	1,8
4	0,1	0,4	1,6
5	0,1	0,5	2,5
$\sum$	1,0	M(x)=2,3	$M(x^2)$= 7,3

Первая колонка в таблице - значения **X**, вторая колонка - вероятности этих значений, третья есть результат произведения первой колонки на вторую и четвертая есть результат произведения первой колонки на третью (потому что **x^2P(x)** получается умножением **x** на **x(P(x)).** Сумма значений третьей колонки дает ожидаемое среднее значение X, а сумма значений четвертой колонки - ожидаемое среднее значение **X^2**. Затем, чтобы получить дисперсию **X**, мы вычисляем разность ***$M(X^2)$ - [M(X)]*** ^{2}

$$\mathbf{D(X) = M(X^2) - [M(X)]^2 = 7,3 - (2,3)^2 = 2,01}$$

Результат совпал с тем, что мы получили, используя формулу (12).

- Среднее квадратическое отклонение (стандартное) отклонение дискретной случайной величины равно корню квадратному из дисперсии, обозначается как σ или **S(X)**

$$\sigma = \sqrt{D(X)} \qquad (18)$$

Для примера 1 среднее квадратическое отклонение есть:

$$\sigma = \sqrt{2{,}01} = \mathbf{1{,}418.}$$

По определению σ^2 - средний квадрат отклонения значений случайной величины от математического ожидания. Отсюда следует, что это мера ***рассеяния*** всех возможных значений случайной величины относительно среднего ожидаемого значения. Дисперсия характеризует колеблемость, изменчивость случайной величины: чем больше вариация, тем дальше от средней находятся возможные значения случайной величины. Для содержательной интерпретации зачастую полезно применять значение, которое дает корень квадратный из дисперсии - среднее квадратическое отклонение (стандартное отклонение). Если сравнивают две случайные величины, то та из них, которая имеет большую дисперсию и среднее квадратическое отклонение, более вариабельна. Риск, ассоциируемый с инвестициями, часто измеряют стандартным отклонением возврата инвестиций. Если сравниваются два типа инвестиций с одинаковой ожидаемой средней возврата, то инвестиции с более высоким средним квадратическим отклонением считаются более рискованными (хотя более высокое стандартное отклонение предполагает возврат более вариабельный с обеих сторон - как ниже, так и выше средней).