Теория Вероятностей и Математическая Статистика
ФИИТ, 2 курс, 4 семестр

Лекция 5

31 марта 2021 г.

1 Характеристики распределения случайной величины

На этой лекции мы продолжим рассмотрение характеристик распределения случайной величины.

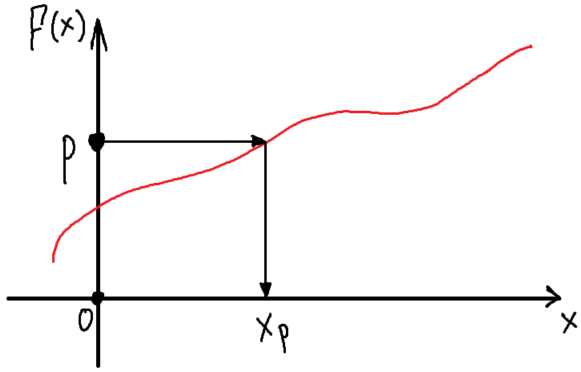
1.1 Квантиль распределения

Начнем мы с непрерывной случайной величины (НСВ). Для неё квантиль определяется следующим образом:

Определение. Число x_p – это квантиль распределения $F(x)$ представляет собой решение уравнения:

$$F(x_p) = p,$$

и называется **квантиль уровня** p $(0 < p < 1)$.



Замечание 1. В случае дискретной случайной величины (ДСВ) квантиль определяется в зависимости от задачи.

Замечание 2. В электронном учебнике квантиль относится к §5, 6, 7, там же посмотреть определение квантиля в теоретической части учебника, чтобы не сделать ошибку для задач с ДСВ.

1.1.1 Квартиль

Еще одно понятие, связанное скорее с мат. статистикой.
Квартиль – это частный случай квантиля, рассматриваемого для величин $p = 0.25, 0.5, 0.75$.
$p = 0.25, 0.75$ – нижний и верхний квартили.
$p = 0.5$ – медиана распределения

1.1.2 Пример

Сравним данные за два года, идущих друг за другом. В первом и во втором году были посчитаны средние зарплаты. Причем оказалось, во втором году средняя зарплата стала выше, чем в предыдущем году. При этом будем считать, что инфляция при этом отсутствует. Можно ли сказать, что благосостояние/уровень жизни людей выросло?

Можно рассмотреть много случаев, одни из них:

1. У людей с низкой зарплатой зарплата стала еще ниже, а у людей с высокой зарплатой наоборот выросла. Разрыв в зарплатах стал наоборот больше, хотя средняя могла повыситься/остаться той же.

2. У всех выросла зарплата, а значит и благосостояние.

Для нас этот вопрос важен с позиции числовых характеристик случайной величины. В общем случае, говорить об однозначном увеличении уровня жизни не приходится. То есть средняя величина, конечно, характеризует что-то, но для полного описания ситуации ее не хватит.

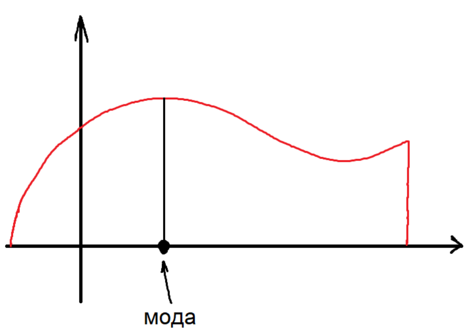
1.1.3 Пример

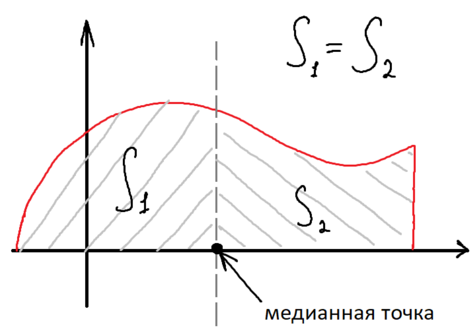
Если рассматривать пример с зарплатами в той же Москве, то мы сможем отметить, что средняя зарплата – 80к. Но кроме средней зарплаты, есть еще две характеристики:

1. **Медианная** – 50к (47к) – показывает центральную зарплату. То есть ровно половина населения получает меньше этой суммы, и ровно половина больше.

2. А так же **модальная** – 30к (33к) – попросту самая распространенная.

Попробуем показать эти величины на графике плотности какой-нибудь СВ:





Модальная точка (мода) соответствует максимальному значению плотности.
Медианная точка соответствует линии, делящей площадь под графиком пополам.
Средняя точка в данном случае будет считаться как координаты центра тяжести фигуры, как в механике. На этом примерочном графике ее рассчитать достаточно трудно, но нам и так понятно, что значит среднее значение.

2 Дискретные случайные величины (ДСВ)

Напомним, что ДСВ характеризует кусочно-постоянная функция распределения, и ее можно задать рядом распределения:

$$\xi \sim \begin{pmatrix} x_1 & x_2 & \dots & x_n & \dots \\ p_0 & p_1 & \dots & p_n & \dots \end{pmatrix}$$

То есть перечислить точки разрыва функции распределения и величины скачков в этих точках — вероятности, отнесенных к этих числовым значениям. Условие нормировки для такой величины — сходимость ряда: $\sum\limits_{i=1}^{\infty} p_i = 1$.

Также напомним себе и о числовых характеристиках ДСВ:

- **Математическое ожидание** — сводится к сумме ряда (если ряд абсолютно сходящийся – иначе говорят, что величина не имеет мат. ожидания):

$$M\xi = \int\limits_{\mathbb{R}} x dF(x) = \sum_{i=1}^{\infty} x_i p_i$$

- **Дисперсия** — *второй центральный момент* (гугл):

$$D\xi = \int\limits_{\mathbb{R}} (x - M\xi)^2 dF(x) = \sum_{i=1}^{\infty} (x - M\xi)^2 p_i$$

Тут же можно отметить мат. ожидание квадрата СВ – *второй начальный момент* СВ:

$$M\xi^2 = \int\limits_{\mathbb{R}} x^2 dF(x) = \sum_{i=1}^{\infty} x_i^2 p_i$$

И тогда **дисперсию** мы можем выразить как разность между *вторым начальным моментом* и *квадратом первого момента*:

$$D\xi = M\xi^2 - (M\xi)^2 = \sum_{i=1}^{\infty} \xi_i^2 p_i - \left(\sum_{i=1}^{\infty} x_i p_i \right)^2$$

2.1 Целочисленные случайные величины

Вполне понятно, что во всем этом многообразии ДСВ мы захотим выделить некоторую группу наиболее частых – это **целочисленные СВ**.

Для работы с этими целочисленными СВ мы введем дополнительную характеристику – **производящая функция неотрицательной целочисленной СВ**.

2.1.1 Производящая функция неотрицательной целочисленной случайной величины

Пусть наша случайная величина принимает значения с разрывами в целочисленных точках:

$$\xi \sim \begin{pmatrix} 0 & 1 & 5 & \dots & x_n & \dots \\ p_0 & p_1 & p_2 & \dots & p_n & \dots \end{pmatrix}$$

Обозначим функцию как $\quad \psi(z) = Mz^{\xi}, \quad$ где $\begin{cases} z\text{– комплексная переменная} \\ |z| \leq 1 \end{cases}$

Условие $|z| \leq 1$ также можно понимать как "значения z берутся в круге радиуса 1".

Чтобы понять, что нам дает подобная функция, **найдем ее производную** (в записи штрихи ставить не будем, так как мы ее еще в степень будем возводить, так что будем писать нижний индекс, мол, дифференцируем по z: $\psi_z(z)$).

Как мы уже выяснили, мат. ожидание представляет собой интеграл, значит когда мы выполняем действие по дифференцированию у нас возникают две линейные операции: *вычисление производной* и *вычисление интеграла*. Понятно, что их можно менять местами:

$$\psi_z(z) = \frac{d}{dz}(Mz^{\xi}) = M(\frac{d}{dz}z^{\xi})$$

Тогда мы можем вполне получить запись такого вида:

$$\psi_z(z) = M(\xi \cdot z^{\xi-1})$$

При этом значение нашей функции в точке 1 в точности дает мат. ожидание: $\boxed{\psi_z(1) = M\xi}$.

Ну а теперь, продолжая исследование функции, **найдем вторую производную** функции. Тогда нам нужно найти производную от выражения, которое стоит внутри мат. ожидания, по той же самой причине, что обе операции линейны:

$$\psi_{zz}(z) = M(\xi(\xi - 1)z^{\xi-2})$$

И опять же, посмотрим на значение производной в точке 1:

$$\psi_{zz}(1) = M(\xi^2 - \xi) = M\xi^2 - M\xi$$

Отсюда мы можем выразить *математическое ожидание квадрата*:

$$M\xi^2 = \psi_{zz}(1) + \psi_z(1)$$

И далее наконец можем найти *дисперсию*:

$$\boxed{D\xi} = M\xi^2 - (M\xi)^2 = \boxed{\psi_{zz}(1) + \psi_z(1) - \psi_z^2(1)}$$

2.1.2 Виды неотрицательных целочисленных случайных величин

1. **Равновероятное распределение.**

   Величина принимает конечное количество значений, и каждое значение имеет равную вероятность:

   $$\xi \sim \begin{pmatrix} 1 & 2 & 3 & \dots & x_n \\ \frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} \end{pmatrix}$$

Математическое ожидание можно расписать следующим образом:

$$M\xi = 1\cdot\frac{1}{n} + 2\cdot\frac{1}{n} + \ldots + n\cdot\frac{1}{n} = \frac{1}{n}(1+2+\ldots+n) = \frac{1}{n}\cdot\frac{n+1}{2}\cdot n$$

Итого получаем следующую формулу для мат. ожидания:

$$\boxed{M\xi = \frac{n+1}{2}}$$

Математическое ожидание квадрата – распишем тем же образом, пользуясь формулой для суммы квадратов:

$$M\xi^2 = 1^2\cdot\frac{1}{n} + \ldots + n^2\cdot\frac{1}{n} = \frac{1}{n}\cdot(1^2+\ldots+n^2) = \frac{1}{n}\cdot\frac{n(n+1)(2n+1)}{6} = \frac{(n+1)(2n+1)}{6}$$

Дисперсия (расчеты можно сделать самостоятельно):

$$\boxed{D\xi} = M\xi^2 - (M\xi)^2 = \boxed{\frac{n^2-1}{12}}$$

Замечание: отметим ситуацию, когда величина принимает не обязательно целочисленные значения. но они все еще равновероятные:

$$\xi \sim \begin{pmatrix} x_1 & x_2 & x_3 & \ldots & x_n & \ldots \\ \frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \ldots & \frac{1}{n} & \ldots \end{pmatrix},$$

математическое ожидание тогда представляет собой среднее арифметическое $\overline{x}$:

$$M\xi = \frac{1}{n}x_1 + \ldots + \frac{1}{n}x_n = \frac{x_1+\ldots+x_n}{n} = \overline{x}$$

2. **Распределение Бернулли.**

   (Это распределение, и все остальные виды СВ, будет использовать схему Бернулли.)

   В данном случае, распределение описывает схему Бернулли, состоящую из одного единственного опыта: $n = 1$. То есть опыт оканчивается либо "успехом либо "неудачей". Таким образом случайную величину можем описать следующим образом:

$$\xi \sim \begin{pmatrix} 0 & 1 \\ q = 1-p & p \end{pmatrix}$$

Тогда наши характеристики можно расписать следующим образом:

$$\boxed{M\xi} = 0 \cdot q + 1 \cdot p = \boxed{p}$$

$$\boxed{D\xi} = p \cdot q = \boxed{p(1-p)}$$

3. **Биномиальное распределение.**

   Теперь мы рассматриваем схему Бернулли, в которой проведено n опытов (все еще конечное количество!). Тогда случайную величину можно описать как количество успехов:

   $$\xi \sim \begin{pmatrix} 0 & 1 & 2 & \dots & x_n \\ p_0 & p_1 & p_2 & \dots & p_n \end{pmatrix}$$

   Вероятность соответствующему значению определяется из формулы:

   $$p_k = C_n^k \cdot p^k \cdot q^{n-k}$$

   Однако теперь вопрос о вычислении **математического ожидания** становится более существенным, чем в предыдущих случаях. Попробуем подойти к решению с помощью ранее описанной производящей функции:

   $$\psi(z) = Mz^\xi = \sum_{k=0}^{n} z^k \cdot C_n^k \cdot p^k \cdot q^{n-k} =$$

   Сгруппируем выражения с одинаковой степенью и, заметив разложение Бинома Ньютона, получим:

   $$= \sum_{k=0}^{n} C_n^k \cdot (zp)^k \cdot q^{n-k} = (zp+q)^n$$

   Далее нам потребуется найти первую и вторую производные данной функции:

   $$\psi_z(z) = n(zp+q)^{n-1} \cdot p$$

   $$\psi_{zz}(z) = n(n-1)(zp+p)^{n-2} \cdot p^2$$

   С помощью них можем выразить **математическое ожидание** по раннее выраженной формуле:

   $$M\xi = \psi_z(1) = np(zp+q)^{n-1} = np(p+q)^{n-1} = np \qquad \textit{// p + q = 1}$$

   А также **дисперсию**:

   $$D\xi = M\xi^2 - (M\xi)^2 = \psi_{zz}(1) + \psi_z(1) - \psi_z^2(1) =$$

$$= n(n-1)p^2 + np - (np)^2 = n^2p^2 - np^2 + np - n^2p^2 = np(1-p) = npq$$

Таким образом:

$$\boxed{M\xi = np} \qquad \boxed{D\xi = npq}$$

Замечание. Биномиальное распределение воспроизводимо:

$$\xi = \xi_1 + \xi_2 + \ldots + \xi_n,$$

где ξ задает кол-во успехов в биномиальном распределении в схеме Бернулли для n опытов, а ξ_i задает распределение в i опыте длинной 1, и может быть выражено как распределение Бернулли (п. 2).

Поэтому для биномиального распределения используется обозначение:

$$\boxed{\xi \sim Bi(n;p)},$$

где величины в этой записи называются *параметрами*: n – кол-во опытов в схеме Бернулли, p – вероятность однократного успеха.

А величина, имеющая распределение Бернулли, будет обозначена таким образом: $\xi \sim Bi(1;p)$

Наиболее вероятное значение (мода).

Оно определяется величиной $np - q$, есть два варианта:

(a) Величина **целая**. Тогда получаются два наиболее вероятных значения:
$$\begin{cases} m_0^{'} = np - q \\ m_0^{''} = np - q + 1 \end{cases}$$

(b) Величина **дробная** – значение единственное:
$m_0 = [np - q] + 1$, где [ ] – целая часть значения.

2.1.3 Пример

Из трехзначных чисел наугад выбрано одно.
Случайная величина ξ – количество попарно различных цифр в записи этого числа.
Требуется найти математическое ожидание $M\xi$.

У этой задачи есть два подхода:

- Первый подход (напрямую).

  У нас трехзначное число, тогда сколько попарно различных цифр может быть в записи этого числа? Одна (цифры совпадают), две (две цифры совпадают, но отличаются от третьей) и три (все три различны):

$$\text{3-х значное число} \rightarrow \quad 1, \quad 2, \quad 3 \quad \text{различных цифр.}$$

То есть случайная величина принимает всего 3 значения:

$\xi \sim \begin{pmatrix} 1 & 2 & 3 \\ p_1 & p_2 & p_3 \end{pmatrix}$

Далее в таком случае можно прорешать и найти мат. ожидание самостоятельно.

- Второй подход с другой стороны.

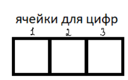
  Введем не одну, а много случайных величин:

$$\underbrace{\xi_0}_{\text{кол-во нулей}}, \quad \xi_1, \quad \ldots, \quad \xi_8, \quad \underbrace{\xi_9}_{\text{кол-во девяток}}$$

Тогда наша случайная величина – это сумма кол-ва нулей, единиц и т.д.:

$$\xi = \underline{\xi_1} + \underbrace{\xi_2 + \ldots + \xi_9}_{\text{рассмотрим сначала их}}$$

Мы рассмотрим сначала кол-во цифр от 1 до 9, по понятным причинам — число трехзначное и с нуля начинаться не может:



$\frac{8}{9} \cdot \frac{9}{10} \cdot \frac{9}{10}$

$$\xi_k \sim \begin{pmatrix} 0 & 1 \\ \frac{8}{9} \cdot \left(\frac{9}{10}\right)^2 & 1 - \frac{8}{9} \cdot \left(\frac{9}{10}\right)^2 \end{pmatrix} \quad k = 1, \ldots, 9$$

И отдельно рассмотрим цифру ноль:

$$\xi_0 \sim \begin{pmatrix} 0 & 1 \\ 1 \cdot \left(\frac{9}{10}\right)^2 & 1 - \left(\frac{9}{10}\right)^2 \end{pmatrix}$$

Таким образом, можем найти мат. ожидания:

$$M\xi_0 = 1 - \left(\frac{9}{10}\right)^2; \qquad M\xi_k = 1 - \frac{8}{9} \cdot \left(\frac{9}{10}\right)^2$$

Осталось только просуммировать мат. ожидания (при расчетах значения ξ для цифр от 1 до 9 одинаковы, поэтому просто умножим одну величину на 9):

$$\xi = \xi_0 + \xi_1 + \ldots + \xi_9$$

$$M\xi = M\xi_0 + 9 \cdot M\xi_1 + = 1 - \left(\frac{9}{10}\right)^2 + 9 \cdot \left(1 - \frac{8}{9} \cdot \left(\frac{9}{10}\right)^2\right) = 2,71$$

2.2 Геометрическое распределение

В этом распределении вновь используется схема Бернулли. Проводится серия опытов, которая рассматривается до наступления первого "успеха". Случайная величина в данном случае – количество проведенных опытов до наступления "успеха".

$$\xi \sim \begin{pmatrix} 1 & 2 & 3 & \dots & x_n & \dots \\ p_0 & p_1 & p_2 & \dots & p_n & \dots \end{pmatrix}$$

Как несложно заметить, если случайная величина равна единице, то у нас сразу был успех (возьмем его вероятность за p): $P(\xi = 1) = p$.

Если равна двойке, то сначала был неуспех, потом успех: $P(\xi = 2) = q \cdot p$.

$$P(\xi = 3) = q^2 \cdot p \qquad\qquad P(\xi = 4) = q^3 \cdot p$$

$$\boxed{p_n = q^{n-1} \cdot p}$$

В качестве задания — производящую функцию построить самостоятельно, и с помощью нее найдете мат. ожидание и дисперсию:

$$\boxed{M\xi = \frac{1}{p}} \qquad\qquad \boxed{D\xi = \frac{q}{p^2}}$$

Замечание. Важно обратить внимание на слово "до [наступления успеха]". Иногда эту случайную величину записывают не включая сам "успех то есть случайная величина будет начинаться с нуля (ноль "неуспехов"до "успеха"):

$$\xi \sim \begin{pmatrix} 0 & 1 & 2 & \dots & x_n & \dots \\ p_0 & p_1 & p_2 & \dots & p_n & \dots \end{pmatrix}$$

Для такого случая дисперсия не изменится, а мат. ожидание получится со сдвигом на единицу:

$$M\xi = \frac{1}{p} - 1 = \frac{q}{p}$$

2.2.1 Пример

Данное распределение можно наблюдать в азартных играх. Случайная величина с таким распределением может дать ответ на вопрос, как наилучшим образом выиграть у казино и вообще, возможно ли это. Построим распределение:

$$\xi \sim \begin{pmatrix} 1 & 2 & 3 & \dots \\ p & qp & q^2p & \dots \end{pmatrix}$$

Посмотрим на моду этого распределения — ей будет являться единица ($\xi = 1$).

Буквально, если смотреть на различные вариации достижения «успеха» , то «успех» с первой игры более вероятен. Просто потому, что вероятности всех остальных значений гарантированно меньше первого (мы же домножаем на $q, q^2, \dots$).

То есть если все игроки придут и сыграют один раз и не более (без разницы, «успех» ли), таким образом их шансы выиграть будет наивысшим, а если человек будет достаточно много, казино можно вовсе разорить...

Вот такой вот математический взгляд на азартные игры.

2.2.2 Характеристическое свойство геометрического распределения

Нам нужно вычислить вероятность:

$$P(\xi = n + k \mid \xi > n)$$

В соответствии с формулой условной вероятности, мы получим:

$$= \frac{P(\xi = n+k,\ \xi > n)}{P(\xi > n)} = \frac{P(\xi = n+k)}{P(\xi > n)}^{//\text{ т.к. очевидно, что n + k > n}} =$$

$$= \frac{q^{n+k-1}p}{q^n p + q^{n+1}p + \ldots} = \frac{q^{n+k-1}p}{q^n p)(1 + q + q^2 + \ldots)}_{//\text{бесконечно убывающая прогрессия}} =$$

$$= \frac{q^{n+k-1}p}{q^n p \cdot \frac{1}{1-q}} = \boxed{q^{k-1}p}$$

Наше распределение вновь стало геометрическим. Это означает, что если в течение n опытов «успех» не наступает, и мы точно знаем о проведении этих опытов, то в целом о них можно просто забыть, и начинать все сначала — то есть все зависит от дальнейших k опытов.

Можно сказать, что это **распределение «не имеет памяти»**. Это и есть основное *характеристическое свойство* данного распределения.