Случайные величины
и функции распределения

(Ахметов С.К.)

Основные задачи и темы курса

- *Цели и задачи курса* ***«Математические методы обработки гидрологической информации»***
- **Случайные величины и функции распределения**
- **Аналитические функции распределения, используемые в гидрологии**
- **Построение кривых обеспеченностей и оценка параметров распределения по эмпирическим данным**
- **Интервальное оценивание параметров и проверка статистических гипотез**
- **Статистический анализ зависимостей между гидрологическими переменными**
- **Случайные процессы**

Случайные величины

- ***Большое число факторов,*** влияющих на гидрологические характеристики – одно из обоснований для обработки гидрологических данных с использованием ***аппарата теории вероятностей***

- ***Случайная величина (СВ)*** – это величина, значение которой меняется от опыта к опыту

- ***Неслучайные или детерминированные величины*** - это величины, значения которых от опыта к опыту не меняются

Закон распределения случайной величины

Закон распределения СВ задан, если:

- *указано множество* возможных значений СВ
- *указан способ* количественного определения вероятности попадания СВ в любую область из множества возможных значений

Вероятность – ***P*** попадания СВ в интервал [a,b] можно определить следующим образом:

$$P_{(a,b)} = \lim_{n\to\infty}\left(\frac{m}{n}\right)$$

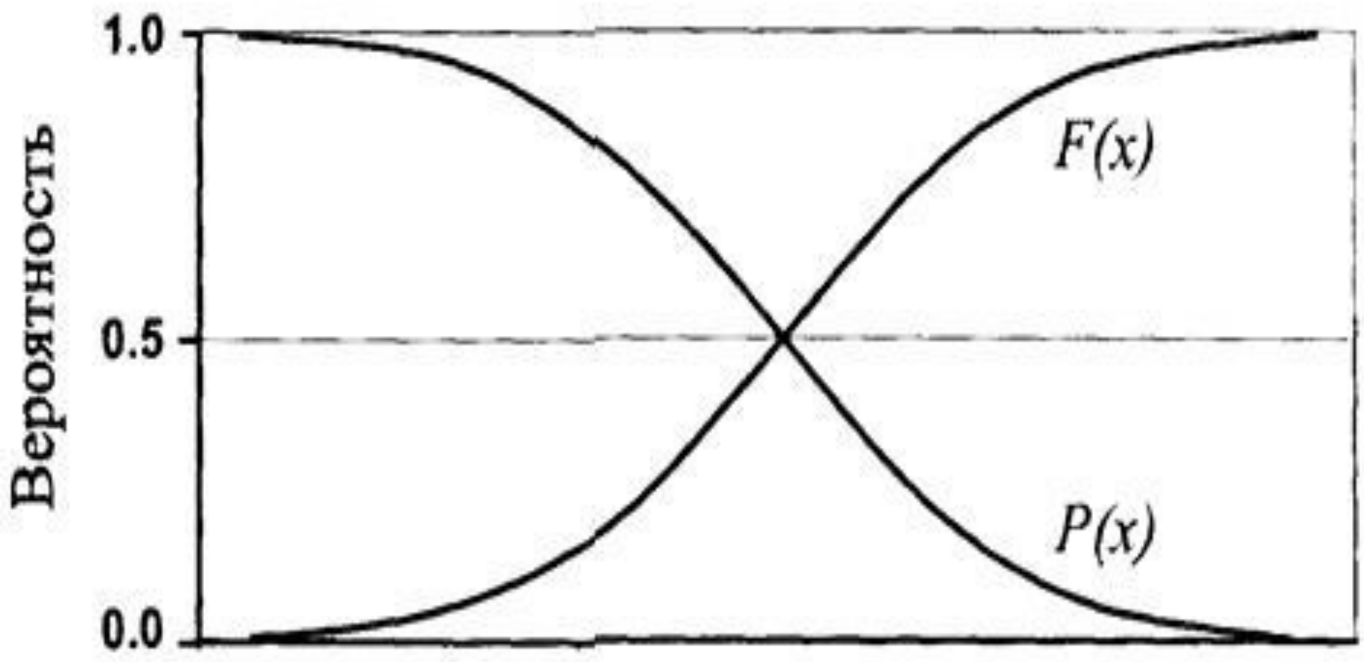
где ***m*** – число наблюдений СВ, оказавшихся в заданной области; N – общее число наблюдений.

Аналитическими выражениями законов распределения случайной величины являются ***функции распределения*** – *интегральная* и *дифференциальная.*

Интегральная функция распределения F(x)

Интегральная функция распределения ***F(x)*** СВ ***X*** показывает вероятность того, что СВ не превысит некоторого заданного числа ***x***, т.е.

$$F(x) = P\{X \leq x\}$$



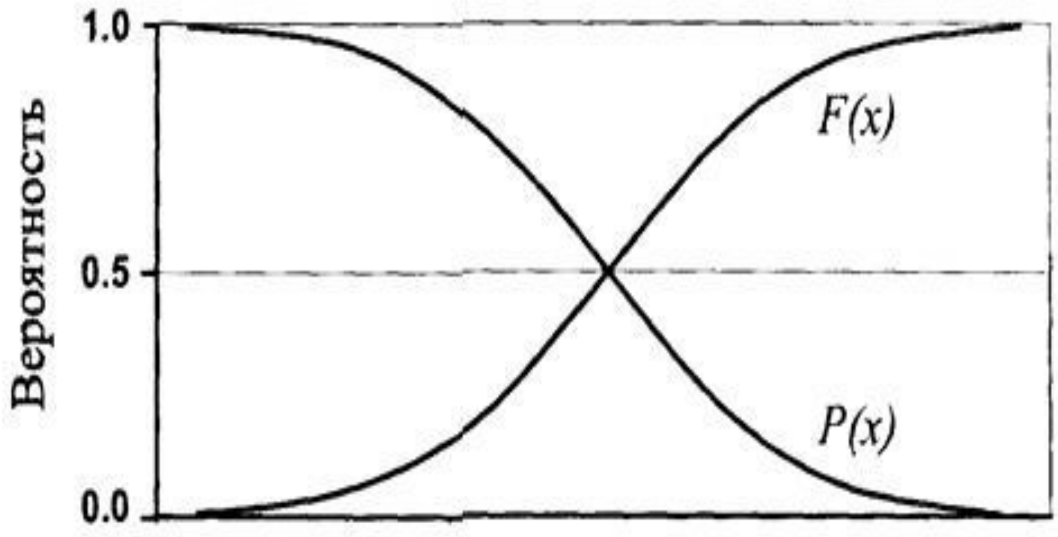
Интегральная функция распределения F(x)

Вероятность того, что значение СВ X заключено между x_1 и x_2 равно разности значений функций распределения, вычисленных в двух точках:

$$\mathbf{P}\ \{x_1 < X \le x_2\} = F(x_2) - F(x_1)$$

аналогично

$$\mathbf{P}\ \{X > x\} = \mathbf{P}\ \{+\infty > X > x\} = 1 - F(x)$$



Функция обеспеченности P(x)

В гидрологической практике вместо функции ***F(x)*** часто используется функция обеспеченности ***P(x)***, но с включением в интервал изменений значения ***x***

$$P(x) = 1 - F(x) = \mathrm{P}\{X \geq x\}$$

То есть функция обеспеченности ***P(x)*** СВ ***X*** показывает вероятность превышения некоторого заданного числа ***x***

Свойства интегральной функции распределения *F(x)* и функция обеспеченности *P(x)*

1. $\lim\limits_{x \to -\infty} F(x) = 0; \quad \lim\limits_{x \to -\infty} P(x) = 1.$

2. $\lim\limits_{x \to +\infty} F(x) = 1; \quad \lim\limits_{x \to +\infty} P(x) = 0.$

3. $F(x) \geq 0$ для всех x; $P(x) \geq 0$ для всех x.

4. $F(x_2) \geq F(x_1)$, если $x_2 > x_1$; $P(x_2) \leq P(x_1)$, если $x_2 > x_1$.

Дифференциальная функция распределения вероятностей

Если функция распределения ***F(x)*** дифференцируема для всех значений СВ ***X*,** то закон распределения вероятностей может быть выражен и в виде дифференциальной функции распределения вероятностей

$$f(x)=\frac{dF(x)}{dx}=\lim_{\Delta x\to 0}\frac{\mathbf{P}\{x<X\le x+\Delta x\}}{\Delta x},$$

где $\Delta x > 0$

f(x) называют также функцией плотности распределения вероятностей или ***функцией плотности вероятности***

Свойства
функции плотности вероятности f(x)

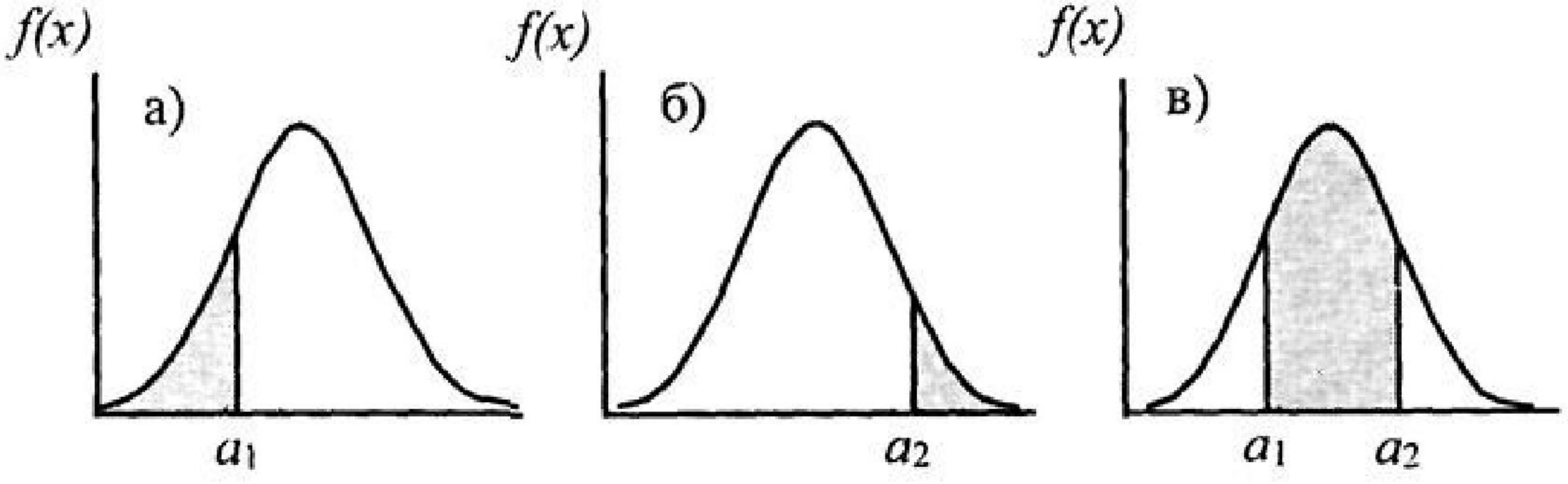
$$1)\ f(x) \geq 0\ ;\quad 2)\ \lim_{|x|\to\infty} f(x) = 0;$$

$$3)\ \int_{-\infty}^{+\infty} f(x)dx = 1;\quad 4)\ \int_{-\infty}^{x} f(z)dz = F(x)$$

С помощью дифференциальной функции распределения можно вычислить вероятность попадания СВ с любую заданную область из множества возможных значений, в частности:

Вычисление вероятности попадания СВ в заданную область с помощью дифференциальной функции распределения

а) $\mathbf{P}\{X \le a_1\} = \int_{-\infty}^{a_1} f(x)dx$ б) $\mathbf{P}\{X > a_2\} = \int_{a_2}^{\infty} f(x)dx$ в) $\mathbf{P}\{a_1 < X \le a_2\} = \int_{a_1}^{a_2} f(x)dx$

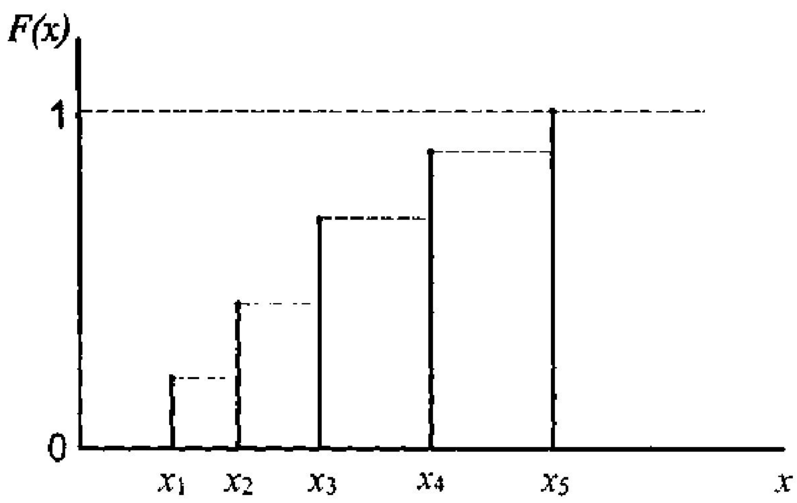


Дискретные и непрерывные случайные величины

Дискретная СВ – это СВ, которая принимает только конечные или счетное множество значений: $x_1, x_2, x_3 \ldots$

Непрерывная СВ может принимать любые значения из некоторого замкнутого или открытого интервала, в том числе и бесконечного.

Интегральная функция распределения дискретной СВ ***X*** в практических ситуациях представляет собой ступенчатую функцию со скачками в точках $x_1, x_2, x_3 \ldots$



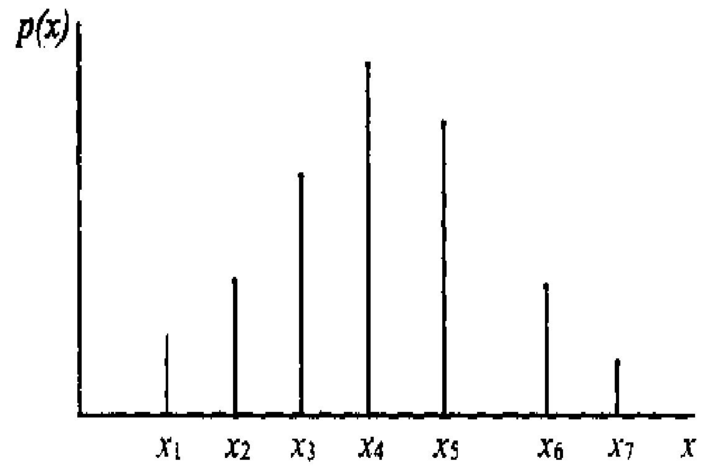
Ряд распределения СВ

Интегральная функция распределения ***F(x)*** дискретной СВ не дифференцируема. Поэтому вместо функции плотности вероятности используется ее дискретный аналог, который называется ***рядом распределения*** и может представляться в виде таблицы

Ряд распределения случайной величины X

X ...	x_1	x_2	x_3	...	x_n
p_x ...	p_1	p_2	p_3	...	p_n

На основании такой таблицы можно построить гистограмму распределения вероятностей дискретной СВ. Для ряда распределения дискретной СВ должно выполняться равенство



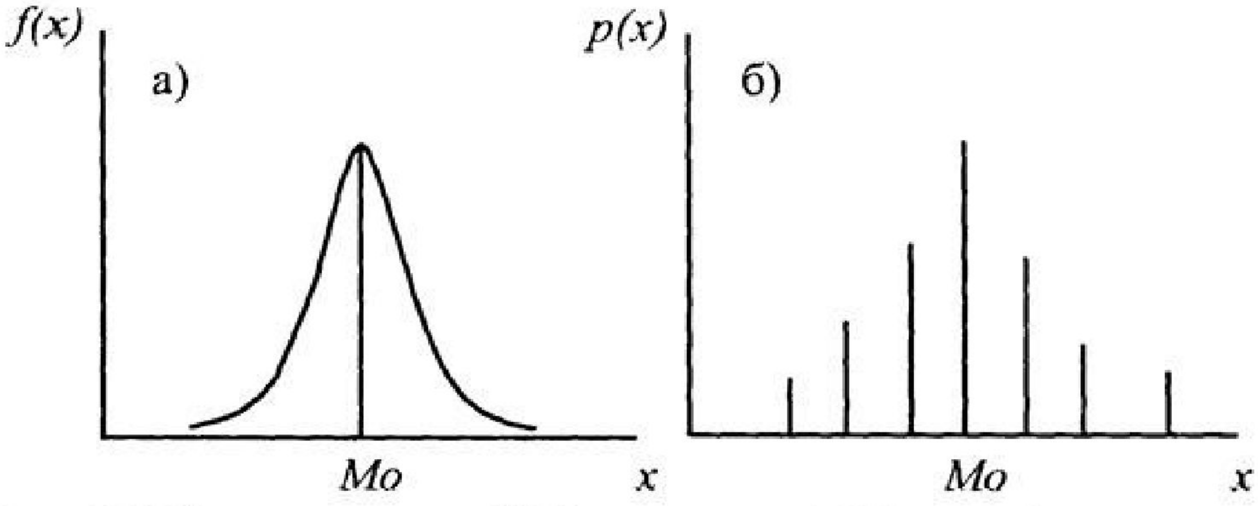
$$\sum_{i=1}^{n} p(x_i) = 1$$

Числовые характеристики случайных величин. ***Мода***

Мода, медиана, математическое ожидание - это параметры, характеризующие положение центра распределения.

Модой *Mo* ***непрерывной*** СВ ***X*** называется такое ее значение, которому соответствует максимум плотности вероятности

Модой *Mo* ***дискретной*** СВ ***X*** называется наиболее вероятное значение СВ



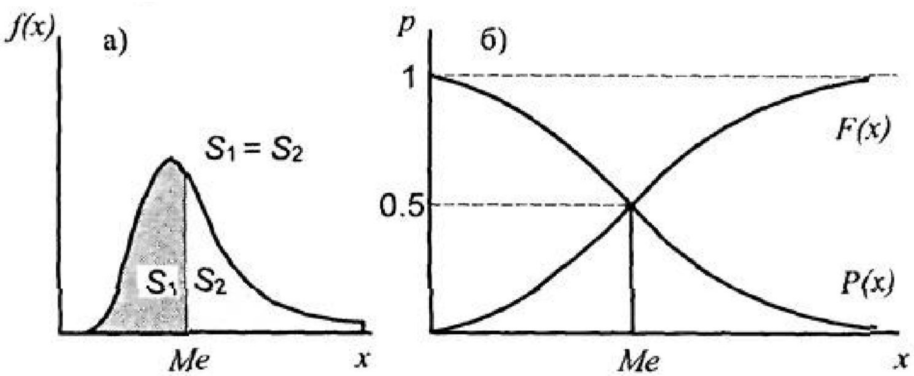
Медиана

Медианой Ме непрерывной СВ ***X*** называется такое ее значение, при котором

$$\mathbf{P}\{X < Me\} = \mathbf{P}\{X > Me\} = 0{,}5\,.$$

Можно сказать, что *Me* – это такое значение СВ, при котором значение функции обеспеченностей равно значению интегральной функции распределения.

$$P(Me) = F(Me) = 0{,}5$$



Положение медианы на графиках дифференциальной (а) и интегральной (б) функций распределения.

Для дискретных СВ медиана определяется неоднозначно и практически не употребляется.

Математическое ожидание (МО)

Математическое ожидание (МО) СВ определяется следующими формулами

для дискретной СВ $$m_x = \sum_i x_i p_i$$

для непрерывной СВ $$m_x = \int_{-\infty}^{+\infty} x f(x) dx$$

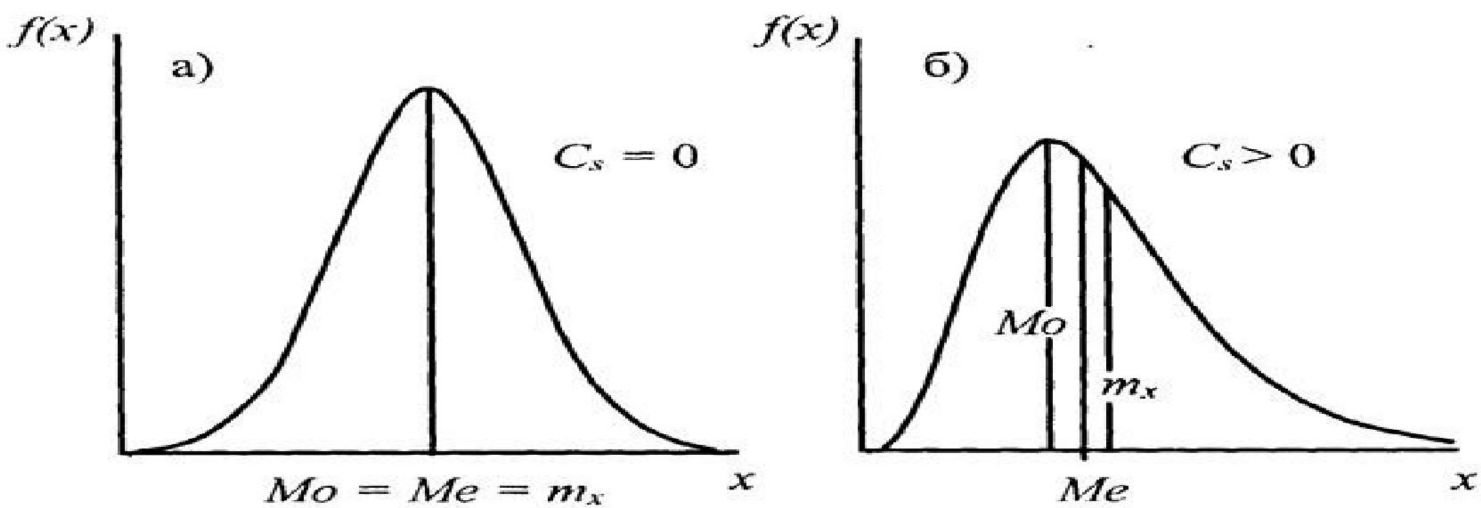
МО можно трактовать как центр тяжести плотности вероятности

В качестве символа ***МО*** используется обозначение **М [Х]**. Таким образом, для СВ ***X*** можно записать также $\boldsymbol{m_x}$ ~ **М [Х]**

Математическое ожидание (МО)

Математическим ожиданием может называться ***генеральное среднее,*** в этом случае для обозначения ***МО*** используется символ $\bar{x}_N$, где $N \to \infty$.

Если мода, медиана и математическое ожидание совпадают, то распределение является симметричным. Если ***МО*** расположено правее медианы, то распределение является положительным, в противном случае – отрицательным.



Моменты случайной величины

Различают ***начальные*** и ***центральные*** моменты СВ

Начальный момент S – го порядка СВ равен

$$a_s = M[X^s] \quad \text{или} \quad \alpha_s = \int_{-\infty}^{+\infty} x^s f(x)\,dx$$

Центральный момент S-го порядка СВ ***X*** определяется формулой

$$\mu_S = \mathbf{M}[(X - m_x)^S], \quad \text{или}$$

$$\mu_S = \int_{-\infty}^{+\infty} (x - m_x)^S f(x)\,dx$$

***МО* - первый начальный момент**, то есть

$$m_x = \mathbf{M}[\mathbf{X}^1] = \alpha_1$$

Дисперсия

Вторую группу наиболее часто используемых на практике параметров составляют параметры, характеризующие степень рассеяния СВ относительно центра распределения.

К ним относится *дисперсия, среднеквадратическое отклонение и коэффициент вариации.*

Дисперсия СВ X представляет собой второй центральный момент, то есть

$$D_x = \mu_2 = \mathrm{M}[(X - m_x)^2]$$

Для *непрерывной* СВ X дисперсия определяется формулой

$$D_x = \int_{-\infty}^{+\infty} (x - m_x)^2 f(x)\,dx$$

Среднеквадратичное отклонение
Коэффициент вариации

Среднеквадратичное отклонение (СКО) СВ X ***(стандарт)*** это квадратный корень из дисперсии.

$$\sigma_x = \sqrt{D_x} \text{ или } \sigma_x^2 = D_x .$$

Для описания рассеяния положительных СВ можно использовать безразмерную характеристику – ***коэффициент вариации.***

Коэффициент вариации C_v СВ X это отношение СКО к МО.

$$C_v = \frac{\sigma_x}{m_x} = \frac{\sqrt{D_x}}{m_x}$$

Асимметрия

Коэффициент асимметрии С является безразмерным параметром и характеризует степень симметричности рассеяния относительно математического ожидания.

Коэффициент асимметрии определяется формулой

$$C_S = \frac{\mu_3}{\sigma_x^3} = \frac{\mathbf{M}\left[\left(X - m_x\right)^3\right]}{\sigma_x^3}$$

Для симметричных распределений коэффициент асимметрии равен нулю.

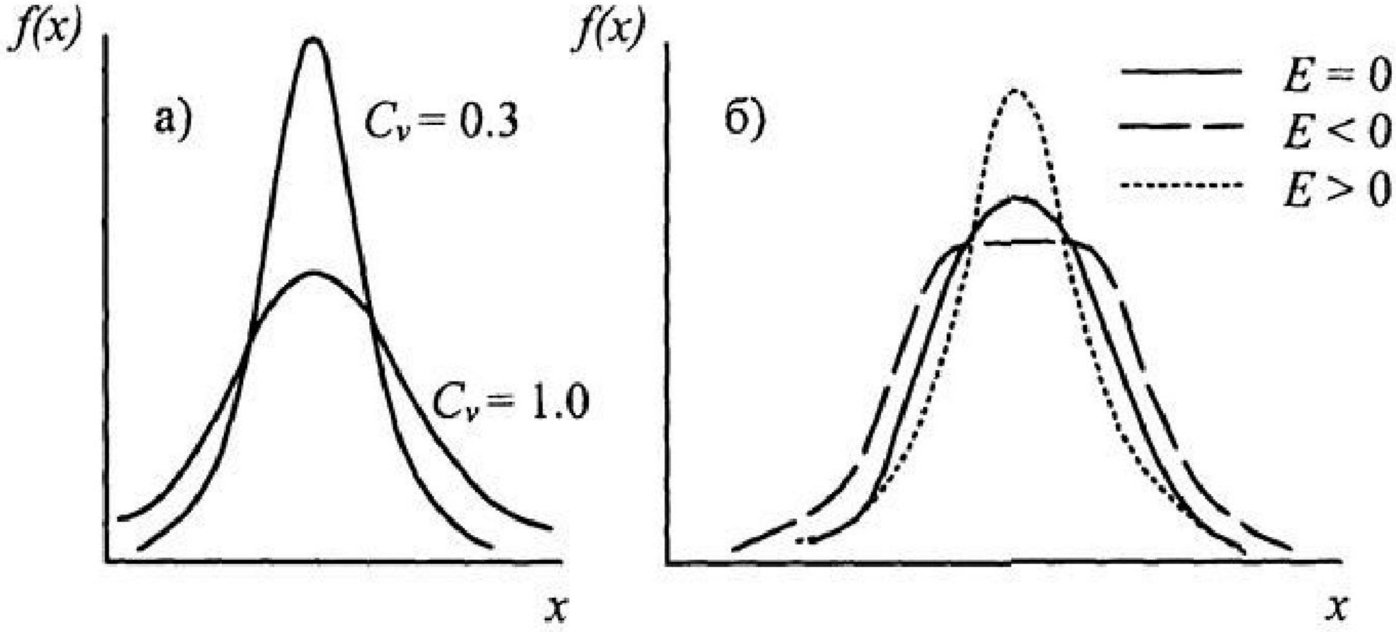
Эксцесс

Эксцесс E_x также является безразмерным параметром и определяется формулой

$$E_x = \frac{\mu_4}{\sigma_x^4} - 3 = \frac{\mathbf{M}\left[\left(x - m_x\right)^4\right]}{\sigma_x^4} - 3$$

Эксцесс позволяет оценить островершинность, или наоборот туповершинность, функции плотности вероятности СВ X относительно нормального закона распределения, для которого $E_x = 0$.

Влияние коэффициента вариации (а) и эксцесса (б) на форму функции плотности вероятности



Свойства математического ожидания

1. МО постоянной величины равно самой этой величине:

$$\mathbf{M[c] = c}, \quad \text{где } \mathbf{c} \text{ – константа}$$

2. Постоянный множитель можно выносить за знак **МО**:

$$\mathbf{M}[cX] = c\mathbf{M}[X]$$

3. МО суммы независимых СВ равно сумме их **МО**

$$\mathbf{M}\left[\sum_{i=1}^{N} X_i\right] = \sum_{i=1}^{N} \mathbf{M}[X_i],$$ так, например $$\mathbf{M}[X+Y] = \mathbf{M}[X] + \mathbf{M}[Y]$$

4. МО линейной функции от СВ выражается формулой

$$\mathbf{M}[aX+b] = a\mathbf{M}[X] + b$$

5. МО произведения независимых СВ равно произведению их **МО:**

$$\mathbf{M}\left[\prod_{i=1}^{N} X_i\right] = \prod_{i=1}^{N} \mathbf{M}[X_i]$$

Свойства дисперсии

1. Дисперсия постоянной величины равно нулю

D[c] = 0, где c = ***const.***

2. Постоянную величину можно вынести за знак дисперсии, возведя ее в квадрат

$$\mathbf{D}[cX] = c^2\,\mathbf{D}[X]$$

3. Дисперсия суммы случайных величин равна сумме их дисперсий

$$\mathbf{D}\left[\sum_{i=1}^{N} X_i\right] = \sum_{i=1}^{N} \mathbf{D}[X_i]$$

4. Дисперсия линейной функции СВ определяется выражением

$$\mathbf{D}[aX + b] = a^2\,\mathbf{D}[X]$$

Стандартные преобразования случайных величин.

В гидрологической практике наиболее часто используется замена СВ ***X*** модульными коэффициентами и замена СВ стандартной нормированной СВ.

Модульным коэффициентом называется соотношение СВ к ее математическому ожиданию

$$k_i = x_i / m_x$$

Стандартная нормированная величина может быть получена из СВ по формуле

$$t_i = (x_i - m_x)/\sigma_x$$

или с учетом формулы выше $t_i = (k_i - 1)/C_v$

Квантили распределения

Во многих практических случаях необходимо по заданной вероятности не превышения ***F(x) = p'*** определить величину x'_p. Для обозначения x'_p в этом случае в математической статистике используется специальный термин – ***квантиль***

p – квантилем называется значение случайной величины x'_p, соответствующее заданному значению вероятности непревышения ***F(x) = p'.***

По аналогии с квантилями в гидрологической практике используется *p – ординаты* кривой обеспеченности

Ординатой кривой обеспеченности называется такое значение СВ X (x_p), которое соответствует заданной вероятности превышения $P(x) = p$

То есть $P(x) = 1 - F(x)$, следовательно, p и p' связаны соотношением $p = 1 - p'$ или (если p в %) $p = 100 - p'$

СПАСИБО ЗА ВНИМАНИЕ!