Курс: Статистические Методы Обработки Данных

Целью специального курса является овладение основами математической статистики и теории вероятностей и их использование при обработке и анализе экспериментальных данных.

Задачами специального курса являются:

-приобретение навыков и изучение способов, приёмов работы по анализу и статистической обработке информации на персональном компьютере, наиболее часто используемых в физических исследованиях (обратив внимание на смысл понятий критериев, возможности алгоритмизации)

-овладение новыми методами теоретических и экспериментальных исследований (решение практических задач, методы получения и обработки результатов опытов).

Рэпазіторый ГДУ ім. Ф. Скарыны

Специальность: 1-53 01 02 – Автоматизированные системы обработки информации

УО «ГГУ им. Ф. Скорины»

Преподаватель: Бабич К.С, ст. преподаватель, 2016

Курс: Статистические Методы Обработки Данных

Лекция 1. Основные понятия Теории Вероятностей и Математической Статистики

Рэпазіторый ГДУ ім. Ф. Скарыны

Специальность: 1-53 01 02 – Автоматизированные системы обработки информации

УО «ГГУ им. Ф. Скорины»

Преподаватель: Бабич К.С, ст. преподаватель, 2016

Литература

1. В.Н. Лавренчик Постановка физического эксперимента и статистическая обработка его результатов", М. Энергоатомиздат, 1986г., 272с.

2. Статистические методы в экспериментальной физике. Пер. с англ. Под ред. А. А. Тяпкина. М., Атомиздат, 1976, 335 с. (Авт.: Идье В., Драйард Д., Джеймс Ф., Рус М., Садуле Б.)

3. П.В. Новицкий, И.А. Заграф Оценка погрешностей результатов измерений, Л. Энергоатомиздат., 1991г., 304с., Лен. отделение

4. Тейлор Дж. Введение в теорию ошибок/ Дж.Тейлор. - М.: Мир, 1985. - 45 с.

5. Гмурман, В. Е. Теория вероятностей и математическая статистика: Учеб. пособие для вузов/В. Е. Гмурман. - 9-е изд., стер. - М.: Высш. шк., 2003. - 479 с.

6. Дж. Бендат, А. Пирсол Прикладной анализ случайных данных, Москва, Мир, 1989г., 504с.

Рэпазіторый ГДУ ім. Ф. Скарыны

1. Введение

Цель (любого) анализа информации – дать количественные характеристики свойств какого-либо объекта.

А для этого делают ограниченное число измерений (например ***n*** – штук измерений какого-либо параметра).

В таком случае говорят, что имеет место

выборка объёма n

рэпазіторый ГДУ ім. Ф. Скарыны

?: Как производится обработка информации?

(выборки n)

Существуют «инструменты»:

- Математическая Статистика (МатСт),
- Теория Вероятностей (ТВ).

Однако,

вводя основные понятия ТВ оперирует с бесконечными выборками $n \to \infty$

а в МатСтат говорят о т.н. ***генеральной совокупности***

(т.о. оперируют абстрактными понятиями)

Рэпазіторый ГДУ ім. Ф. Скарыны

?: Какие основные задачи, и какие усилия приходится предпринимать, чтобы получить количественную информацию о реальном наборе данных?

1) Группировка Данных (**среднее** выборки),
2) Рассеяние данных (**дисперсия**),
3) Связь различных данных между собой (**корреляция**)

м.б. двойная задача - изучить еще и форму этой взаимосвязи (**регрессия**)

Рэпазіторый ГДУ ім. Ф. Скарыны

2. Некоторые сведения из Теории Вероятностей

Частота попадания случайной величины.

Для исследования свойств объекта проводят измерения, позволяющие количественно дать характеристики свойств этого объекта. На практике производится ограниченное число измерений n при одинаковых условиях. В результате получаем множество событий (значений) исследуемой величины X: $\{x_1, x_2, \ldots x_n\}$, которое представляет собой выборку объема n. Таким образом, возможные исходы некоторого эксперимента (измерения) представляют собой множество точек, которые можно сочетать разными способами. **Такие сочетания называют событиями**.

В том случае,(т.е. когда одна и та же характеристика объекта принимает различные значения) говорят о том, что величина X является ***случайной величиной***. Случайная величина -это действительное число (или набор действительных чисел), которое заключено между $-\infty$ и ∞, которое сопоставляется каждой возможной точке из числа значений этой характеристики. При соответствующих условиях для каждое событие можно характеризовать **частотой появления** ν_n.

Рэпазіторый ГДУ ім. Ф. Скарыны

2. Некоторые сведения из Теории Вероятностей

Частота попадания случайной величины.

Частотой появления ν_n события A называется отношение числа m появления данного события к общему числу проведенных одинаковых испытаний, в каждом из которых могло появиться или не появиться данное событие:

$$\nu_n = \frac{m}{n} \tag{1}$$

Если число испытаний n велико, то, как правило, частоты появления данного события A в различных сериях измерений отличаются мало друг от друга. Это утверждение записывают следующим образом:

$$\lim_{n\to\infty} \nu_n = p\ . \tag{2}$$

Число p называются вероятностью (англ.-probability) случайного события A . Отметим, что существуют такие события у которых частота появления может сильно отличаться от вероятности, даже при большом числе испытаний.

Рэпазіторый ГДУ ім. Ф. Скарыны

2. Некоторые сведения из Теории Вероятностей

Как видим из вышеприведенного примера, для описания случайного поведения величины необходима совокупность, содержащая неограниченное число значений измеряемой величины ($n = \infty$). Такая выборка называется генеральной совокупностью. Генеральная совокупность часто используется как важное абстрактное понятие, необходимое в теоретических расчетах, связанных с исследованием поведения физической величины как случайной величины.

Различают два основных типа случайных величин: дискретные случайные величины и непрерывные случайные величины. Если величина X имеет **конечное** число (счетное множество) из последовательности возможных значений $\{x_1, x_2, \ldots x_k, \ldots, \}$ то такая величина называется *дискретной случайной величиной*. Если случайная величина может принимать любое значение из интервала возможных значений, то такая величина называется *непрерывной случайной величиной*.

Рэпазіторый ГДУ ім. Ф. Скарыны

2. Некоторые сведения из Теории Вероятностей

Закон распределения вероятности.

Для характеристики частоты появления различных значений случайной величины X теория вероятностей предлагает пользоваться указанием **закона распределения вероятностей** различных значений этой величины.

При этом различают два вида описания законов распределения:
1) **интегральный**
2) **дифференциальный.**

Рэпазіторый ГДУ ім. Ф. Скарыны

2. Некоторые сведения из Теории Вероятностей

Интегральный закон распределения вероятности.

Интегральным законом, или функцией распределения вероятностей $F(x)$ случайной величины , называют функцию, значения которой представляют вероятность того, что значения случайной величины x_k величина X меньше некоторого значения x (x - некоторое произвольное число).

Данное утверждение символически записывается в виде

$$F(x) = \mathrm{Prob}\,(x_k < x)\ , \tag{3}$$

где $\mathrm{Prob}\,(x_k < x)$ и представляет собой вероятность события в вышеприведенном определении.

Очевидно, что

$$F\,(a) < (b)\,, \quad \text{при} \quad a \leq b \quad \text{(неубывающая функция)} \tag{4}$$

$$F\,(-\infty) = 0\,, \quad F\,(\infty) = 1\,. \tag{5}$$

Рэпазіторый ГДУ ім. Ф. Скарыны

2. Некоторые сведения из Теории Вероятностей

Дифференциальный закон распределения вероятностей

Для случайной величины с непрерывной и дифференцируемой функцией распределения $F(x)$ можно найти **дифференциальный закон распределения вероятностей**:

$$p(x) = \frac{\mathrm{d}F(x)}{\mathrm{d}x}. \tag{6}$$

$p(x)$ называют кривой плотности распределения вероятностей (или просто **плотность вероятности**).

Свойства $p(x)$: 1) $p(x) \geq 0$

2) Из (6) следует, что

$$F(x) = \int_{-\infty}^{x} p(\xi)d\xi \,. \tag{7}$$

3) Условие нормировки:

$$\int_{-\infty}^{\infty} p(x)\,\mathrm{d}x = 1 \,. \tag{8}$$

Рэпазіторый ГДУ ім. Ф. Скарыны

2. Некоторые сведения из Теории Вероятностей

Замечания:

Зная законы распределения имеем информацию о случайной величине с помощью мат. аппарата МатСт.

Но т.к. все понятия в теории основаны на понятии «генеральной совокупности», то нам для использования этого теоретического аппарата на практике придется искать «обходной путь».

Рэпазіторый ГДУ ім. Ф. Скарыны

2. Некоторые сведения из Теории Вероятностей

Примеры законов распределений

Нормальное распределение:

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-a)^2}{2\sigma^2}\right], \tag{9}$$

a и σ -параметры нормального распределения.

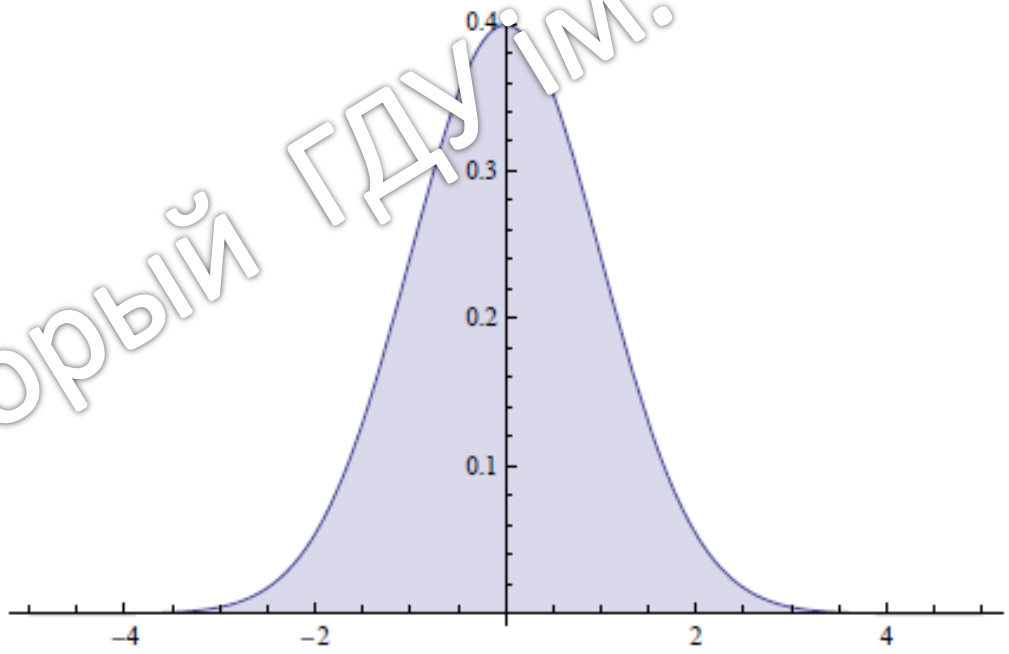


Рис.: Нормальное распределение с $a = 0$ и $\sigma = 1$

Рэпазіторый ГДУ ім. Ф. Скарыны

2. Некоторые сведения из Теории Вероятностей

Примеры законов распределений

Распределение χ^2 с n степенями свободы:

$$p(\chi^2) = \frac{(\chi^2)^{(\frac{n}{2}-1)} e^{-\frac{\chi^2}{2}}}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}, \quad (\chi^2 > 0); \tag{10}$$

n -параметры нормального распределения.

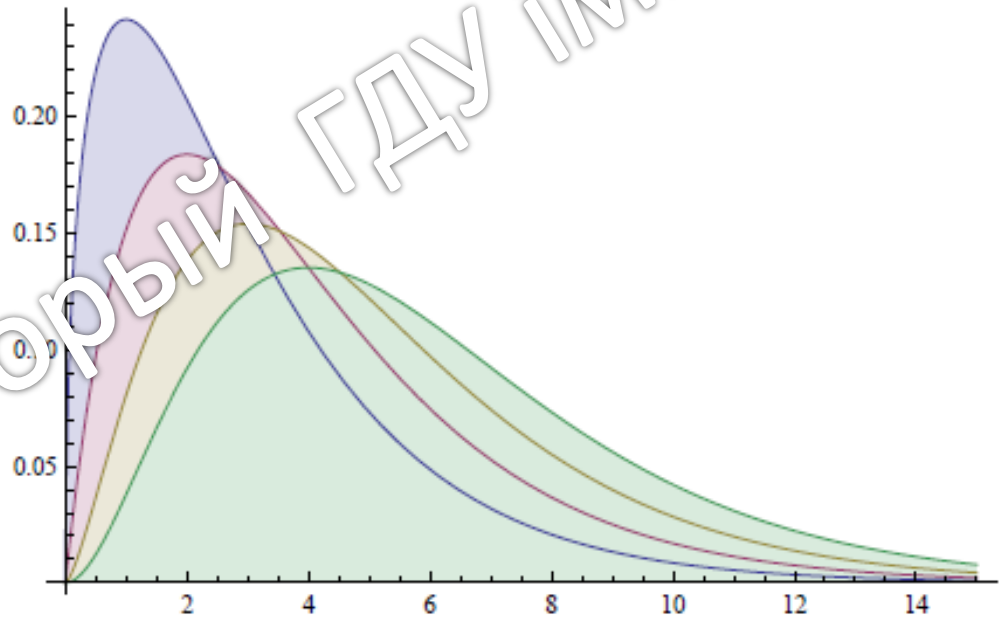


Рис.: Распределение χ^2 с $n = 3, 4, 5, 6$ степенями свободы

Рэпазіторый ГДУ ім. Ф. Скарыны

2. Некоторые сведения из Теории Вероятностей

Примеры законов распределений

Равномерное распределение:

$$p(x) = \frac{1}{b-a} \quad , (b > a) \, , \tag{11}$$

a, b -параметры распределения.

$$p(x) = \begin{cases} \dfrac{1}{b-a}, & a \le x \le b \\ 0, & для\ ост.\ x \end{cases}$$

Рэпазіторый ГДУ ім. Ф. Скарыны

2. Некоторые сведения из Теории Вероятностей

Примеры законов распределений

Равномерное распределение:

$$p(x) = \frac{1}{b-a} \quad , (b > a) \, , \tag{11}$$

a, b -параметры распределения.

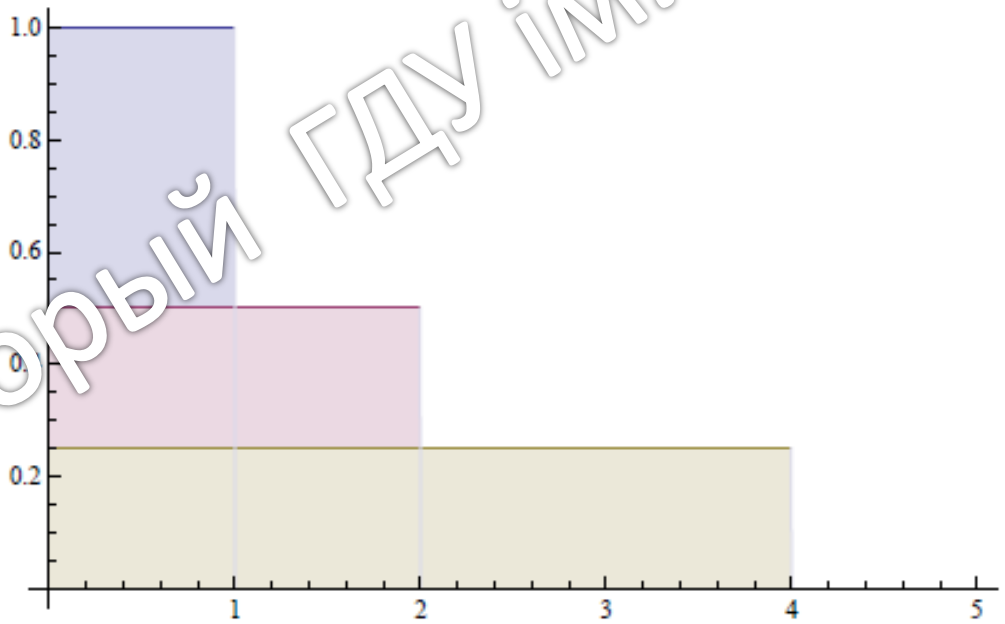


Рис.: Равномерное распределение с $a = 0$ и $b = 1, 2, 4$

Рэпазіторый ГДУ ім. Ф. Скарыны

2. Некоторые сведения из Теории Вероятностей

Примеры законов распределений

Равномерное распределение:

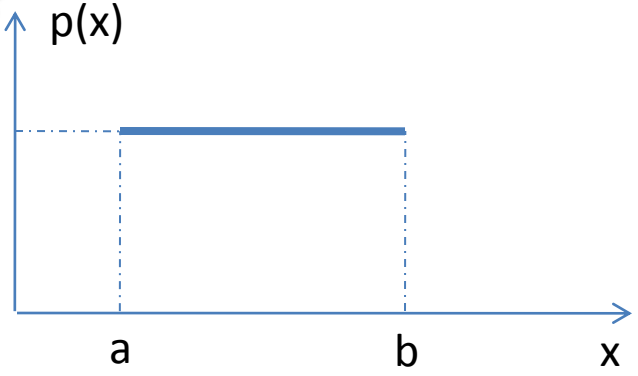
$$p(x) = \frac{1}{b-a} \quad ,(b > a) \, , \tag{11}$$

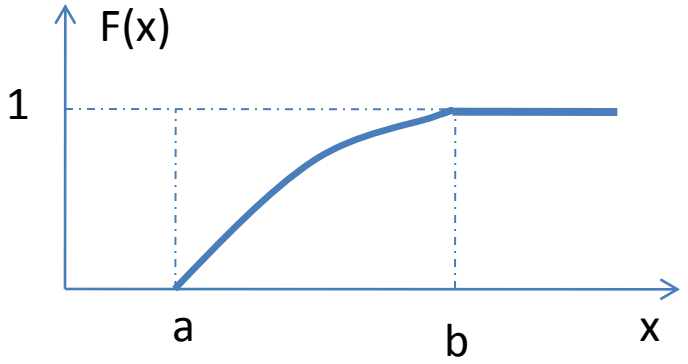
a, b -параметры распределения.

$$p(x) = \begin{cases} \frac{1}{b-a}, & a \le x \le b \\ 0, & для\ ост.\ x \end{cases}$$

$$F(x) = \int_{-\infty}^{x} p(\xi) d\xi$$

$$F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \le x \le b \\ 1, & x > b \end{cases}$$





Рэпазіторый ГДУ ім. Ф. Скарыны

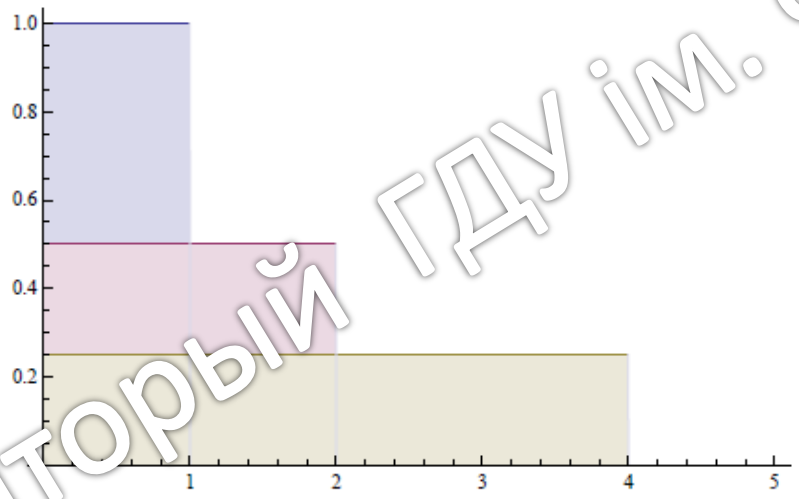
2. Некоторые сведения из Теории Вероятностей

Примеры законов распределений

Равномерное распределение:

$$p(x) = \frac{1}{b-a} \quad ,(b > a) \, , \qquad (1)$$

a, b -параметры распределения.



Замечание: *зная равномерное распределение можно смоделировать любое распределение.*

Рэпазіторый ГДУ ім. Ф. Скарыны

3. Моменты распределения

МатСтат. Предлагает использовать вместо распределений - **моменты.**

Если мы сможем найти моменты (все) распределения, то – следовательно – мы будем знать само распределение.

Рэпазіторый ГДУ ім. Ф. Скарыны

3. Моменты распределения

Алгебраический момент распределения k-го порядка (для непрерывной случайной величины)

$$\mu_k = \int_{-\infty}^{\infty} x^k p(x)dx \ , \qquad (12)$$

Центральный момент k-го порядка

$$\nu_k = \int_{-\infty}^{\infty} (x - \mu_1)^k p(x)dx \ . \qquad (13)$$

Замечание: *в случае дискретной сл. величины – суммы вместо интегралов.*

На практике: стоит задача найти (12) и (13), а уже по ним отыскать p(x).

Как правило ищут несколько моментов…

Рэпазіторый ГДУ ім. Ф. Скарыны

3. Моменты распределения

Первый момент называют – Математическим Ожиданием (М.о.)

$$\mu_1 = E(x) = \int_{-\infty}^{\infty} xp(x)dx \qquad (14)$$

Он характеризует центр распределения.
Однако следует отметить, что не для всех распределений существует м.о. Например, распределение Коши. Интеграл (14) для него расходится.

т.о. 1-й момент как характеристика центра распределения не является универсальным и вводят другие величины.

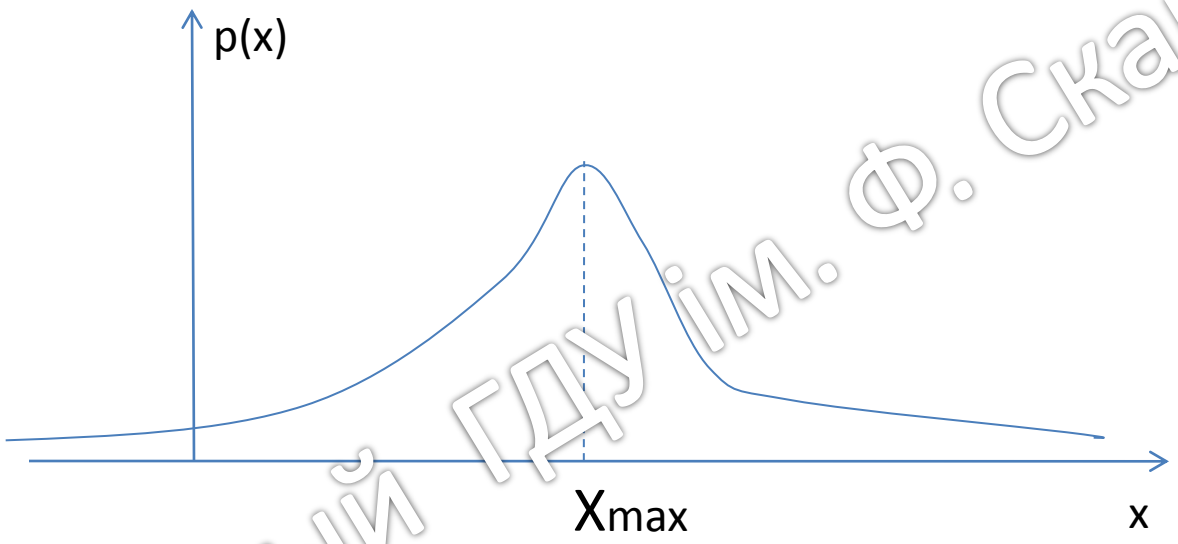
Наиболее общая характеристикой центра распределения – Медиана.

Медиана – прямая, параллельная оси Y, проходящая через точку на оси X, слева и справа от которой вероятности появления различных значений случайной величины равны между собой и составляют $P_1 = P_2 = 0.5$.

Рэпазіторый ГДУ ім. Ф. Скарыны

3. Моменты распределения

Мода распределения - точка max значения плотности вероятности



Если точек максимумов >=2, то говорят, что распределение ***многомодальное****.*

В этом случае может не совпадать с М.о.

Рэпазіторый ГДУ ім. Ф. Скарыны

3. Моменты распределения

Второй центральный момент называется – Дисперсия случайной величины.

$$\nu_2 = D = \int_{-\infty}^{\infty} (x - \mu_1)^2 p(x) dx \,. \tag{15}$$

Дисперсия характеризует рассеяние отдельных значений случайной величины от центра распределения.

Для дискретных значений имеем:

$$\nu_2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2 \,. \tag{16}$$

Дисперсия имеет размерность квадрата случайной величины и выражает как бы мощность рассеяния, поэтому для более наглядной характеристики используют:

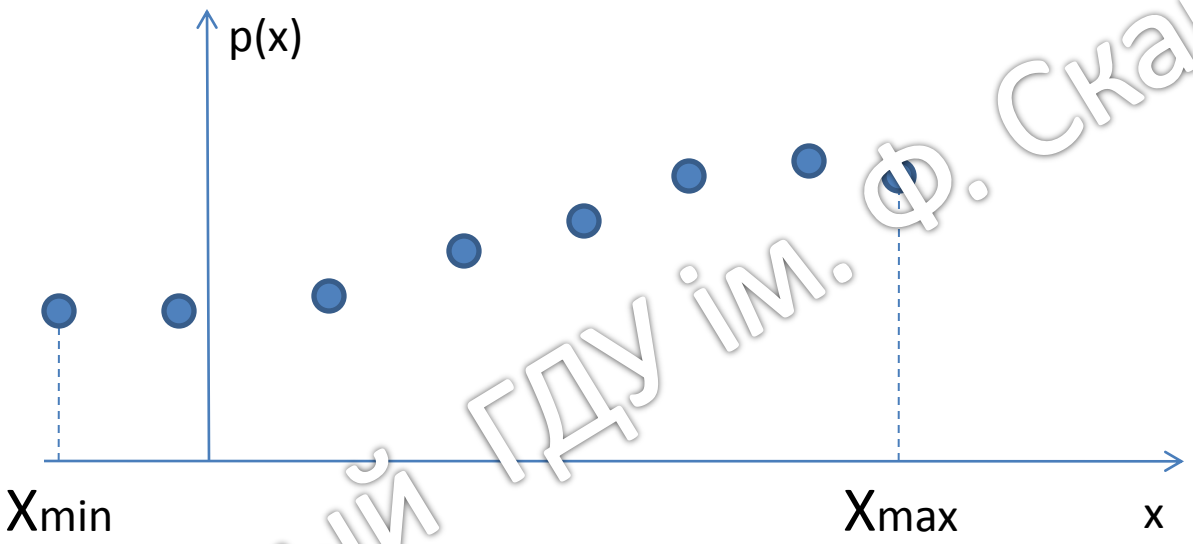
$$\sigma = \sqrt{D} \,, \tag{17}$$

которое называют среднеквадратичным отклонением (с.к.о.) и имеет размерность самой случайной величины.

Рэпазіторый ГДУ ім. Ф. Скарыны

3. Моменты распределения

Размах – интервал, где содержится выборка.



Размах = X min – X max.

Рэпазіторый ГДУ ім. Ф. Скарыны

3. Моменты распределения

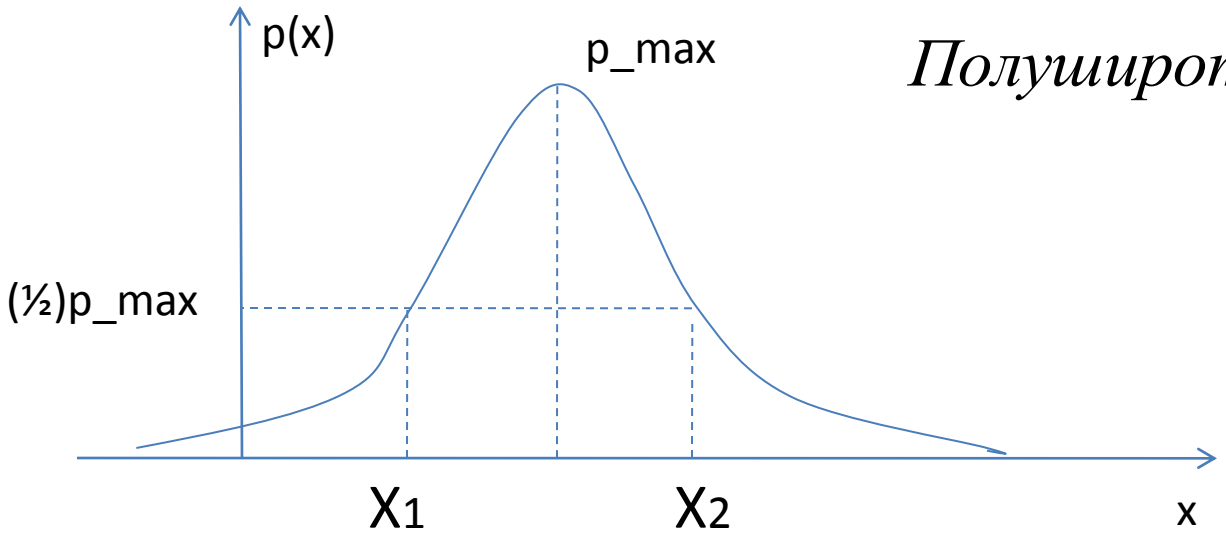
Коэффициент вариации (безразмерная величина).

$$Коэф.Вар. = \frac{\sigma}{M.o.[x]}$$

Рэпазіторый ГДУ ім. Ф. Скарыны

3. Моменты распределения

Полуширота (для однородных непрерывных распределений)



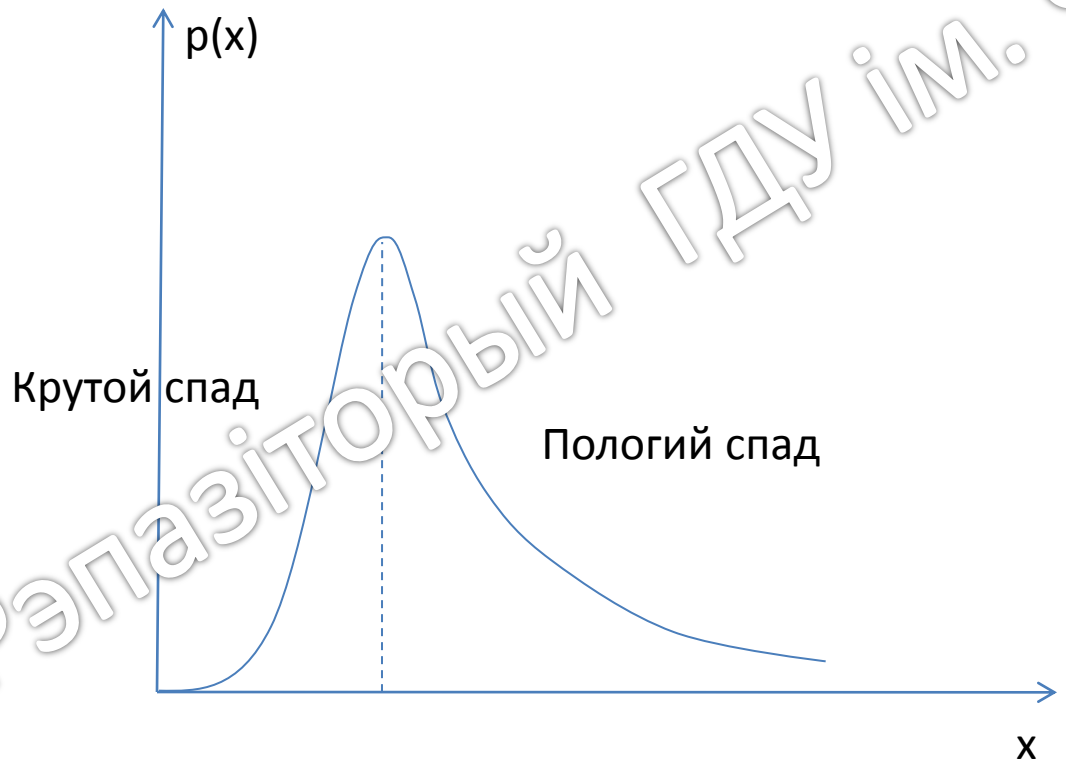
$$Полуширота = \frac{x_2 - x_1}{2}$$

Примечание: реальные спектры - модальные.

3. Моменты распределения

Третий центральный момент ν_3 характеризует асимметрию, т.е. скошенность распределения: когда один спад - крутой, а другой - пологий. Для симметричных относительно центра распределений он равен нулю. ν_3 имеет размерность куба случайной величины, поэтому для относительной характеристики используют безразмерный коэффициент:

$$S = \frac{\nu_3}{\sigma^3}$$ **Коэффициент Асимметрии**



Пр. Распределение Реллея

Рэпазіторый ГДУ ім. Ф. Скарыны

3. Моменты распределения

Четвертый центральный момент ν_4 характеризует протяженность распределения. Значение

$$\varepsilon = \frac{\nu_4}{\sigma^4}$$ **- Эксцесс**

называют эксцессом распределения.

Для Нормального распределения ε=3.

Иногда вводят коэффициент эксцесса γ_2 = (ε – 3).

для менее протяжённых распределений γ_2 = [-2; 0],

а для более протяжённых чем нормальное распределение γ_2 >0 [0, ∞]

для удобства часто используют контр-эксцесс

$$\kappa = \frac{1}{\sqrt{\varepsilon}}, \quad \kappa \in [0,1]$$

Рэпазіторый ГДУ ім. Ф. Скарыны

3. Моменты распределения

На практике используем:

1-й центральный момент

2-й центральный момент

3-й центральный момент

4-й центральный момент

и другие характеристики

?: Как их определять на эксперименте?

Рэпазіторый ГДУ ім. Ф. Скарыны