***Случайная величина*** -

**Генеральная совокупность X (сл. вел. X)** – множество возможных значений случайной величины X

**Закон распределения ген. сов. X** – закон распределения сл. вел. X

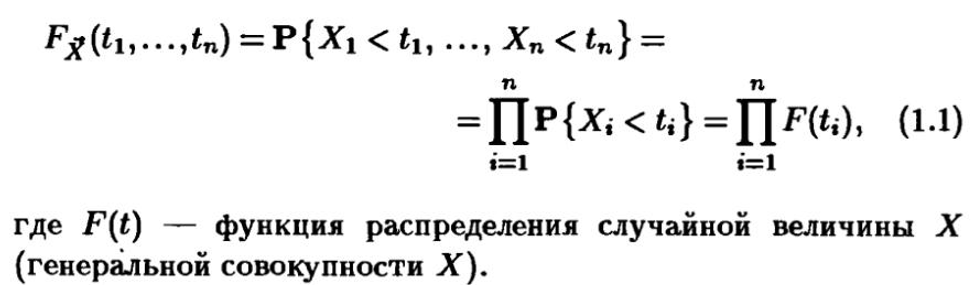
**Случайная выборка из ген. сов. X** – совокупность независимых сл. вел. , каждая из которых имеет то же распределение, что и сл. вел. X. Записывается

**Выборка из ген. сов. X (реализация сл. выборки )** – это любое возможное значение сл. выборки . Интерпретируется как результаты n независимых наблюдений над сл. величиной X.

**Выборочный метод** – свойства сл. вел. X устанавливаются путем изучения тех же свойств на случайной выборке.

**Выборочное пространство** - \chi \_n – множество значений сл. выборки .

Как выражается функция распределения сл. выборки через функцию распределения генеральной совокупности (т.е. через ф-цию распр. X).



стр 21. *Статистическая модель* – выборочное пространство, на котором задан класс распределений сл. выборки (если мы знаем тип функции распределения, но не знаем ее параметры).

*Параметрическая модель*

**Статистика (выборочная характеристика)** – любая функция случайной выборки – она является случайной величиной с распределением, называемым **выборочным распределением**

**Выборочное распределение** выборочной характеристики

**Выборочное значение** выборочной характеристики – значение выборочной характеристики , определенное по реализации случайной выборки .

стр. 24 ***Сходимость по вероятности***

***Сходимость по распределению (слабая)***

**Задачи мат. статистики:** оценка неизв. параметров, проверка стат. гипотез, установление формы и степени связи между сл. вел.

**Два подхода к оценке неизвестный параметров функции распределения генеральной совокупности** – точечная оценка и интервальная оценка.

**Точечная оценка** (или просто **оценка**) неизвестного параметра ф-ции распр. генеральной совокупности – это статистка (функция) , выборочное значение которой для любой реализации принимают за приближенное значение неизвестного параметра .

**Значение точечной оценки** - .

**Интервальная оценка** с коэффициентом доверия неизвестного параметра ф-ции распр. генеральной совокупности – это пара статистик (функций) и таких, что с вероятностью выполняется неравенство (то есть таких, что ).

**Доверительный интервал для с коэффициентом доверия**  - .

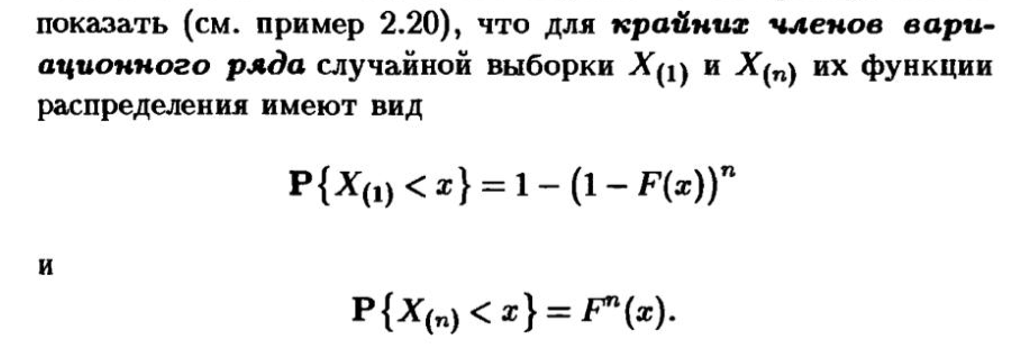
**Статистическая гипотеза** – любое предположение о распределении вероятностей (о вероятностных свойствах) наблюдаемой сл.вел. (гипотеза о величине м.о., об однородности (т.е. равенстве) дисперсий, о виде распределения и т.д.).

*Корреляционный* и *дисперсионный* анализ – наличие связи между величинами и ее существенность. *Регрессионный* анализ – построение регрессионной модели (т.е. зависимости ср. знач. сл. величины от знач. других сл. величин).

**Вариационный ряд выборки ­** – упорядоченная последовательность элементов выборки .

**Вариационный ряд случайной выборки ­** – последовательность случайных величин , где – сл. величина, которая при каждой реализации случайной выборки принимает значение, равное i-му члену вариационного ряда выборки .

***Функции распр. крайних членов вариационного ряда ()***. Вывод



**Статистический ряд** – таблица, которая в первой строчке содержит уникальные отсортированные значения элементов выборки, а во второй – количество их повторений.

**Частота** – количество раз, которое встречается элемент в выборке.

**Относительная частота (частость)** – отношение частоты значения элемента выборке к общему количеству элементов в выборке.

**Интервальный статистический ряд** – отрезок, содержащий все значения выборки, делят на равные части и составляют статистический ряд, в котором количество элементов подсчитывается на интервале.

**Оптимальное число интервалов** для гистограммы – по правилу Стёрджеса – .

Стр 32 **Выборочная функция распределения**

**Эмпирическая функция распределения**

**Теоретическая функция распределения**

**Эмпирич. плотность распр.**

**Гистограмма, Полигон частот**

**Выборочные числовые моменты**

**Теоретические (генеральные) числовые характеристики**

**Выборочный начальный момент k-го порядка,**

**Выборочный центральный момент k-го порядка**

**Выборочное среднее, Выборочная дисперсия,**

**Выборочное ср/квадр. отклонение**

**Выборочный корреляционный момент,**

**Выборочный коэффициент корреляции**

**Кор. момент выборки,**

**Коэф. кор. выборки**

из лабы 1 **Коэффициент вариации**

**Стандартное отклонение**

**Стандартизованная асимметрия**

**Стандартизованный эксцесс**

**! Актуальные критерии нормальности распределения (асимметрия и эксцесс и что-то еще?)**

из лабы 2 **Состоятельность оценки**

**! Правила для определения достаточного объема выборки**

**Законы больших чисел**

**Теорема Бернулли – закон больших чисел** <https://studopedia.ru/12_163342_reshenie.html>

При неограниченном увеличении числа однородных независимых опытов частота события будет сколь угодно мало отличаться от вероятности события в отдельном опыте.

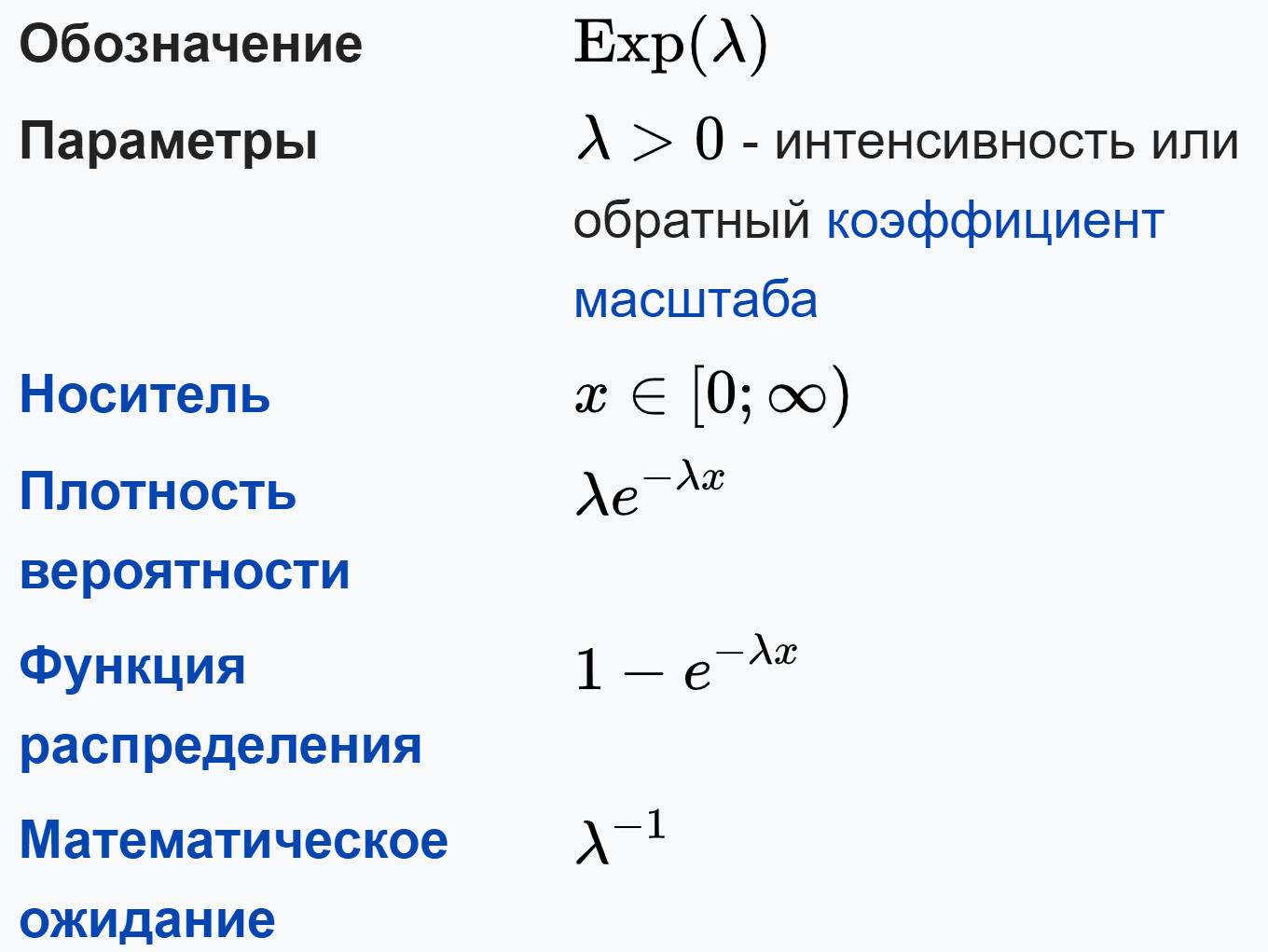
Иначе, вероятность того, что отклонение относительной частоты m/n наступления события А от постоянной вероятности p события А очень мало при n -> +oo, стремится к 1 при любом eps > 0

https://lh3.googleusercontent.com/wSw55d86wUTdlhc8RXc639KPHUWehkZvmiYL9f29qm31PEPxVOvAr5OEmcZjF_2peUUgg7YJ3StgOW9oeoFhXCMMvf19Cu252htRKipVDBmV-UUkauKYNzy9FlyjcmBkDssLXLw6

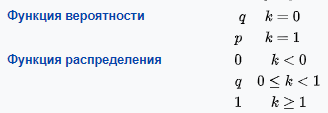
**Геометрическое распределение** - распределение вероятностей случайной величины равной количеству «неудач» до первого «успеха» в серии испытаний Бернулли и принимающей значения либо распределение вероятностей случайной величины равной номеру первого «успеха» и принимающей значения .



**Экспоненциальное распределение** - абсолютно непрерывное распределение, моделирующее время между двумя последовательными свершениями одного и того же события.

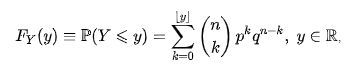
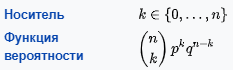


**Распределение Бернулли** – дискретное распределение вероятностей, моделирующее случайный эксперимент произвольной природы, при заранее известной вероятности успеха () или неудачи ().



**Биномиальное распределение** - распределение количества «успехов» в последовательности из n независимых случайных экспериментов, таких, что вероятность «успеха» в каждом из них постоянна и равна .

Пусть - конечная последовательность независимых случайных величин, имеющих одинаковое распределение Бернулли с параметром . Тогда сл. вел. имеет биномиальное распределение с параметрами и . .

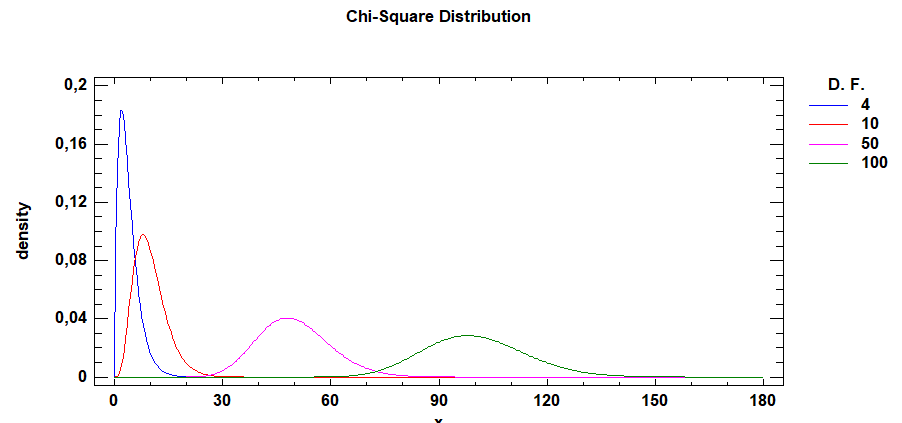


**Распределение Пуассона** – вероятностное распределение дискретного типа, моделирует случайную величину, представляющую собой число событий, произошедших за фиксированное время, при условии, что данные события происходят с некоторой фиксированной средней интенсивностью и независимо друг от друга. , где – м.о.

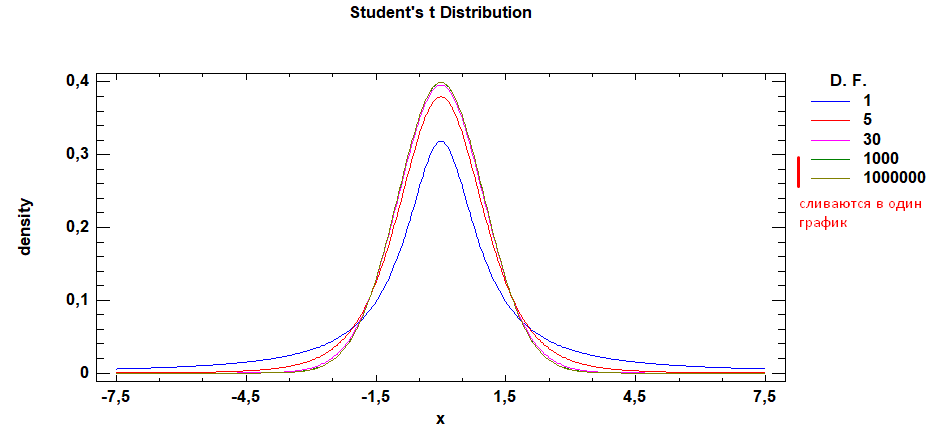
Функция вероятности Функция распределения .

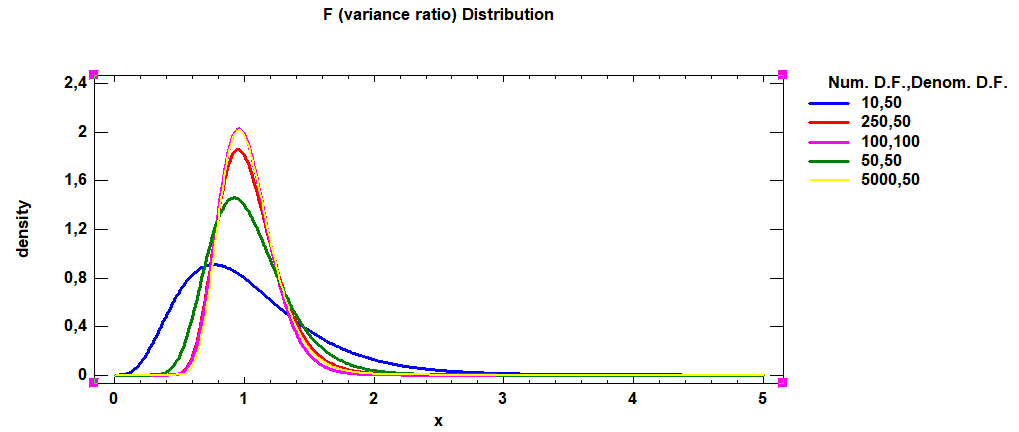
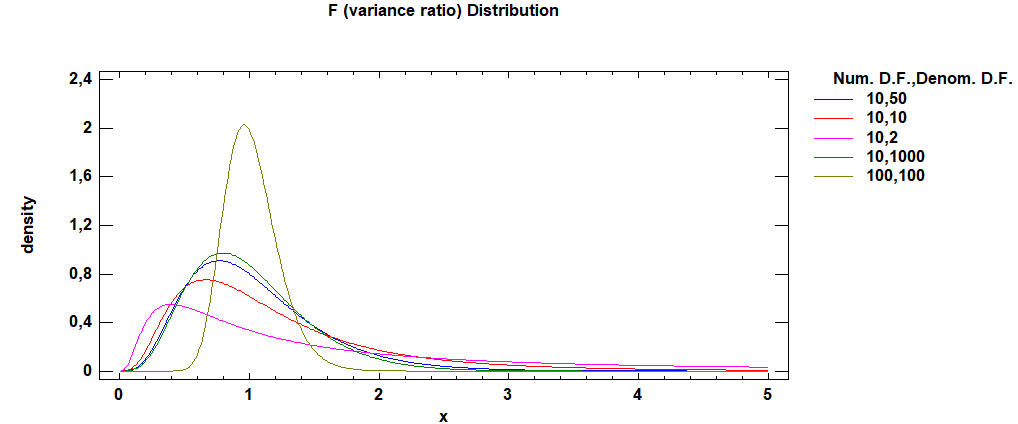
**Распределение (Chi-Squared)** с степенями свободы — это распределение суммы квадратов независимых стандартных нормальных случайных величин.

Пусть – совместно независимые стандартные нормальные случайные величины, то есть . Тогда случайная величина имеет распределение хи-квадрат с степенями свободы, т.е. .

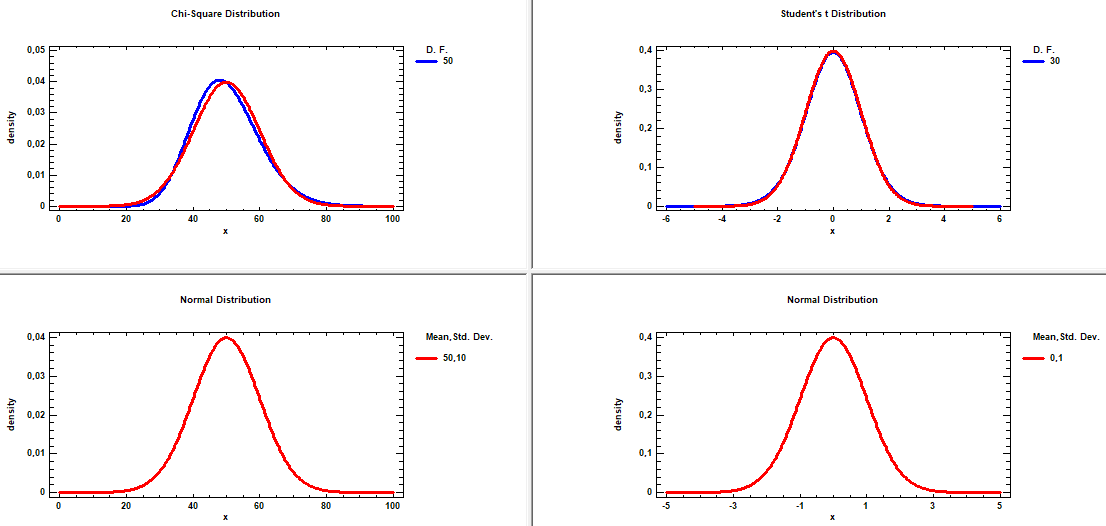
**Распределение Стьюдента (Student’s t)** – это однопараметрическое семейство абсолютно непрерывных распределений. Пусть – конечная последовательность независимых стандартных нормальных случайных величин, т.е. . Тогда распределение сл. вел.

имеет распределение Стьюдента с степенями свободы. .

**Распределение Фишера F (Variance Ratio)** – это двухпараметрическое семейство абсолютно непрерывных распределений. Пусть  — две независимые случайные величины, имеющие распределение хи-квадрат: , где . Тогда распределение случайной величины

называется распределением Фишера со степенями свободы и . Пишут 

**Асимптотическая нормальность распределений Стьюдента и**  – , а при .



Bin(0.5, 100), Bin(0.01, 100), Bin(0.99, 100)

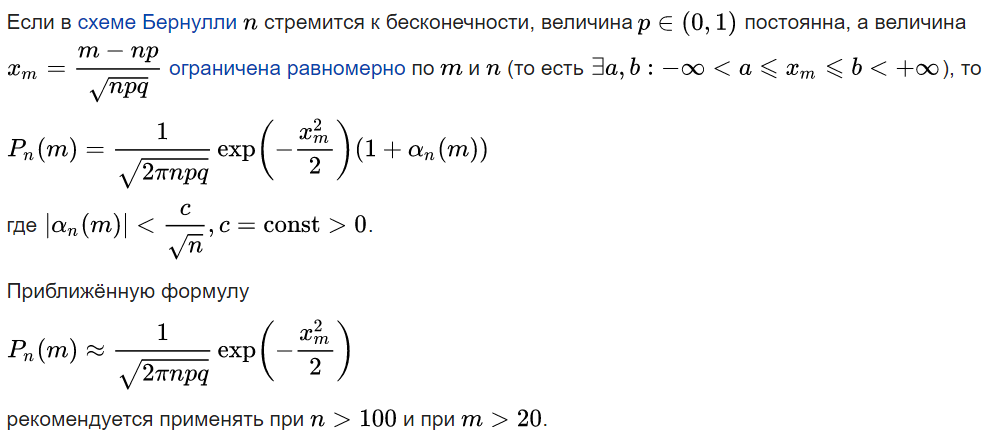
А) Для какой из выборок гистограмма «похожа» на нормальную кривую? Почему это можно было ожидать (вспомните предельные теоремы из теории вероятностей (какую???).

Б) На какое распределение должна быть «похожа» гистограмма для второго распределения? Наложите это распределение на гистограмму.

В) Почему нормальная аппроксимация дает плохой результат для третьей выборки?

**~~! Тест Колмогорова-Смирнова для проверки нормальности~~**

**~~! Критерий~~**

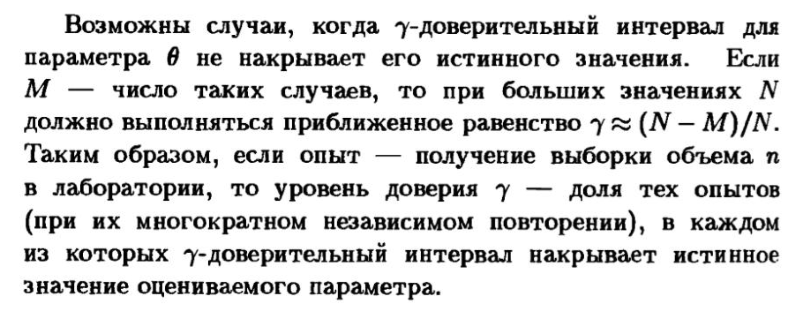
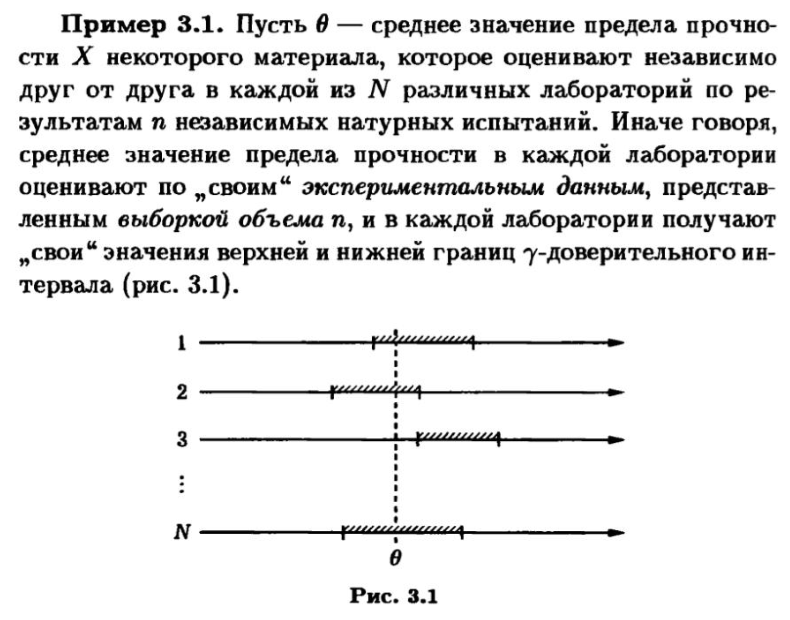
**Локальная теорема Муавра — Лапласа**

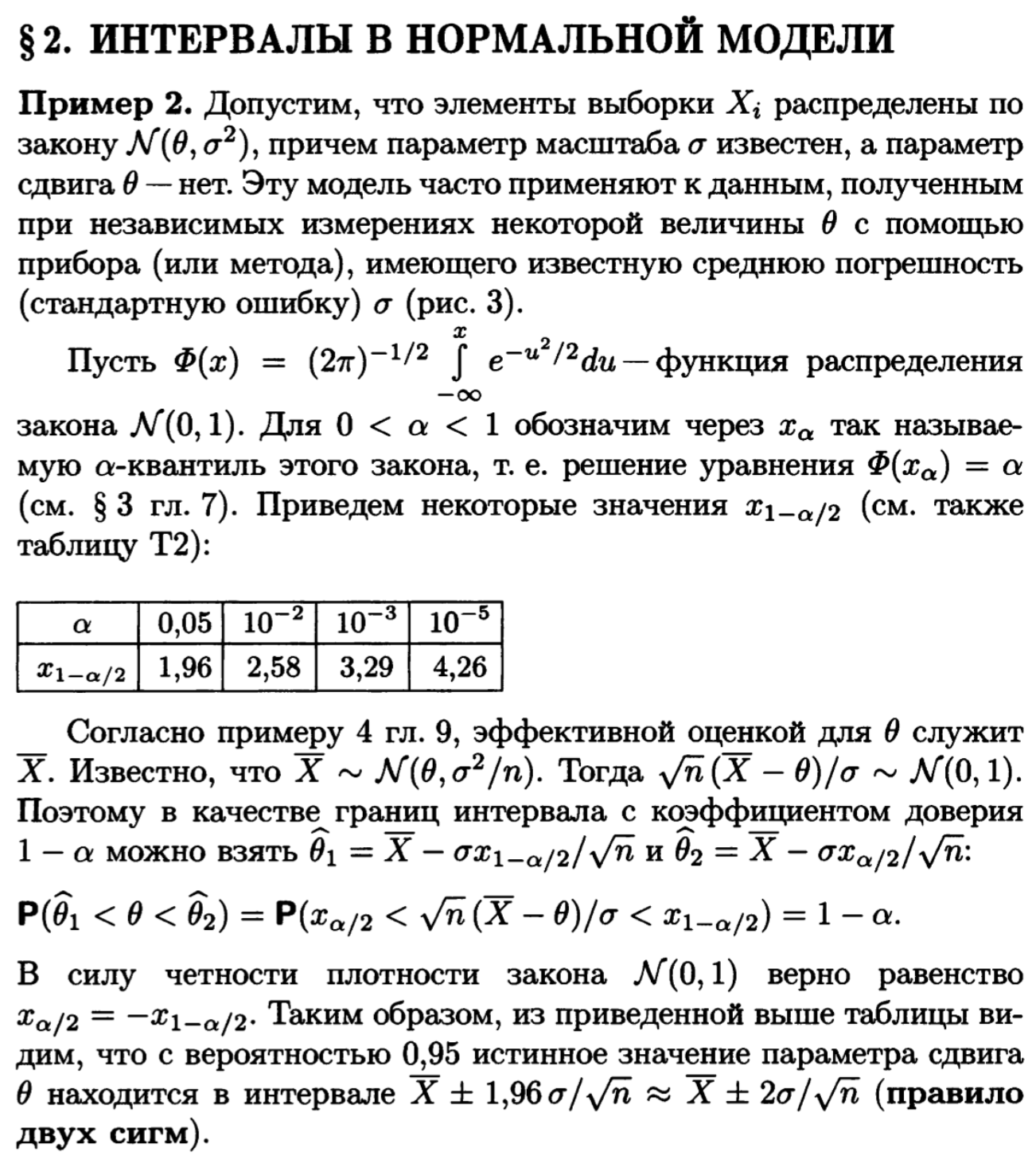
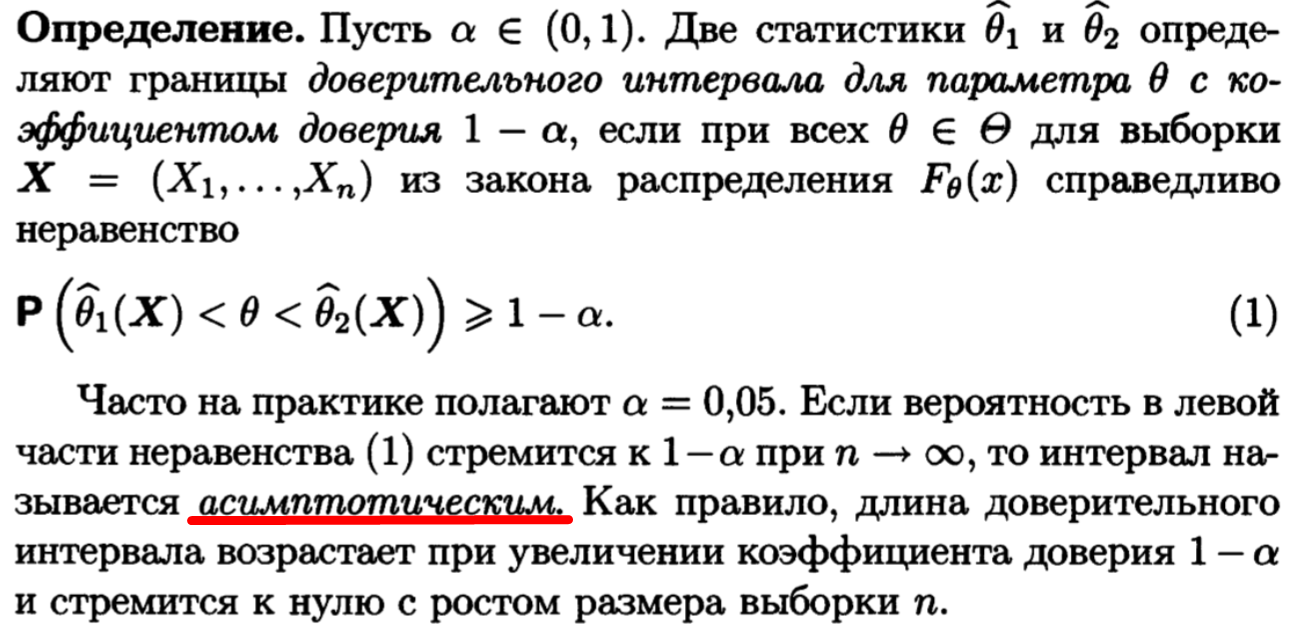
**-доверительная интервальная оценка** – .

**Нижняя и верхняя границы** интервальной оценки - пара статистик (функций) и .

Коэффициент доверия (доверительная вероятность, уровень доверия).

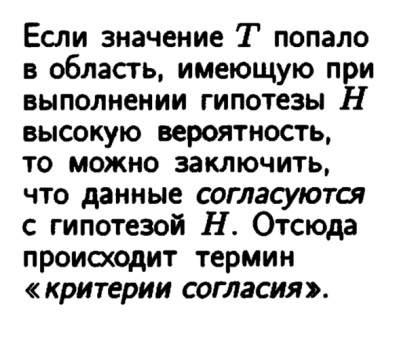
**Односторонняя нижняя (и соответственно верхняя) – доверительная граница** – , когда .





**Статистический критерий** – правило, позволяющее принять или отвергнуть гипотезу H на основе реализации выборки .

**Статистика критерия** – – статистика (функция), для которой типично принимать умеренные значения в случае, когда гипотеза H верна, и большие (малые), когда H не выполняется.

**Уровень значимости** – – вероятность, с которой мы можем позволить себе отвергнуть верную гипотезу H.

**Схема:**

Берем – статистика (какая-то функция),   
 – реализацию выборки (данные эксперимента).

Делаем гипотезу H, выбираем приемлемый уровень значимости.

Если H – верна, то у нас есть определенные ожидания от значения T. Мы находим , такое, что . То есть это максимальное значение для функции T такое, что её вероятность быть больше этого значения «равна» (точнее не больше , то есть маленькая). Потом вычисляем реальное и смотрим, (что более ожидаемо при выполнении H) или нет.

**Критическое значение** – – значение статистики критерия , при превышении которого мы должны отвергнуть гипотезу (так как по факту произошло маловероятное событие, наше предположение, наша гипотеза, скорее всего, не верна).

Фактический уровень значимости - .

**Статистическая гипотеза**

**Нулевая гипотеза**

**Альтернативная гипотеза**

**Двусторонняя гипотеза**

**Односторонняя**

**Простая гипотеза**

**Параметрическая гипотеза**

**Ошибка первого рода**

**Ошибка второго рода**

**Критическая область**

**Наилучшая критическая область (область принятия решений)**

**Непараметрическая гипотеза**

**Статистическая значимость** какого-либо значения

**Непараметрический критерий**

**Парные/непарные наблюдения/критерии**

**Для проверки гипотезы о типе распределения существуют специальные критерии — критерии согласия**

**! Тест Колмогорова-Смирнова для проверки нормальности**

**! Критерий**

**Однородность выборки**

**Равномерность выборки**

**---------------------------------**

**Таблица сопряженности**