Федеральное государственное автономное
образовательное учреждение
высшего образования
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»
Институт космических и информационных технологий
Базовая кафедра «Интеллектуальные системы управления»

УТВЕРЖДАЮ
Заведующий кафедрой
______ Ю.Ю. Якунин
подпись инициалы, фамилия
« » июня 2018 г.

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

Топологическая идентификация закона распределения вероятности одномерной выборки

27.04.03 Системный анализ и управление

27.04.03.02 Системный анализ данных и технологий принятия решений

Научный руководитель	08.06.2018 подпись, дата	доцент, канд. техн. наук должность, ученая степень	А.А. Даничев инициалы, фамилия
Выпускник	08.06.2018 подпись, дата		А.И. Ярум инициалы, фамилия
Рецензент	08.06.2018 подпись, дата	доцент, канд. физ.-мат. наук должность, ученая степень	И.М. Федотова инициалы, фамилия

Красноярск 2018

РЕФЕРАТ

Выпускная магистерская диссертация по теме «Идентификация закона распределения вероятности одномерной выборки по параметрам формы». Данная работа содержит 87 страницы текстового документа, 57 иллюстраций, 119 формул, 11 использованных источников.

ВЫБОРКА, СЛУЧАЙНАЯ ВЕЛИЧИНА, ЗАКОН РАСПРЕДЕЛЕНИЯ, ИДЕНТИФИКАЦИЯ ЗАКОНА РАСПРЕДЕЛЕНИЯ, АСИММЕТРИЯ, ЭКСЦЕСС.

Существует много критериев идентификации распределения, но каждый из них не идеален и допускаются ошибки. Поэтому необходимо разрабатывать новые критерии нахождения закона распределения вероятности.

Объектом данной дипломной работы является закон распределения вероятностей.

Предмет представляет собой одномерную случайную выборку

Целью является идентификация закона распределения одномерной выборки по параметрам формы.

Задачи исследования

1) Анализ теоретических источников об идентификации законов распределения.
2) Теоретическое обоснование топологического метода для произвольного закона распределения случайной величины.
3) Выбор программной среды для выполнения статистических расчетов.
4) Разработка алгоритма топологической идентификации закона распределения вероятности случайной величины.
5) Разработка программной реализации алгоритма и апробация эффективности его работы на множестве выборок случайных величин.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ

Человек в своей повседневной жизни непрерывно сталкивается со случайными явлениями, т.е. явлениями, которые точно предсказать невозможно. Эти явления имеют количественную характеристику, которая называется случайной величиной. Закон распределения вероятностей случайной величины – это некоторая функция, которая позволяет определить вероятность того, что случайная величина принимает определенное значение или попадает в некоторый интервал. Благодаря закону распределения можно исследовать разнообразные величины, даже те, среди которых существует разброс. Множество физических процессов имеет случайную природу (к примеру, траектория движения молекулы). Каждую из этих величин необходимо описать. Для этого наиболее подходящим представляется использование закона распределения вероятностей.

Идентификация закона распределения вероятности представляет собой сложную, но необходимую задачу. Качество решения которой, напрямую зависит не только от применения конкретного метода, но и от объема имеющихся экспериментальных данных.

Для решения задачи определения закона распределения вероятности в основном применяются классические методы, основанные на математические статистики, и топологические методы.

Далее рассмотрим топологический и классический подходы.

При использовании классического метода обязательным условием является прохождение трех этапов: первое - определения модели закона распределения на основе имеющихся экспериментальных данных, нахождение оценок параметров этого закона и проверка этой модели на соответствие экспериментальным данным с помощью критериев согласия. Если критерий согласия дает отрицательный результат, то все начинается сначала, т.е. следует перейти на первый этап. Успех данного подхода определяется двумя факторами: множеством моделей законов распределения, которые рассматриваются в задаче, и используемых методов статистики. Следует отметить, что разные критерии согласия дают различные степени отклонения фактического закона распределения вероятности от теоретического. Поэтому целесообразно применять несколько критериев согласия одновременно. Это позволить получить наиболее достоверные результаты, т.е. достичь максимальный уровень правдоподобия и сократить вероятность ошибки. Однако для рассмотрения большого объема моделей закона распределения вероятности необходимо использовать определенное программное обеспечение.

Не следует забывать о выбросах, так как некоторые методы оценивания параметров распределения чувствительны к ним. Во избежание ошибочных результатов необходимо очищать выборку от аномальных измерений.

Топологический метод основан на использовании оценок моментов, например таких как: асимметрия и эксцесс. Эти оценки получаю по имеющимся экспериментальным данным. Если рассматривать значения асимметрии и эксцесса, как координаты точки, можно, исходя из близости ее к одной из линий, представляющих некоторые распределения, судить о принадлежности данной случайной величины этому закону. Данный метод следует применять при больших объемах исходной выборки, очищенной от аномальных измерений. Его можно использовать как предварительный, позволяющий выбрать из всего множества моделей закона распределения некоторую совокупность наиболее вероятных, которую в дальнейшем целесообразно обрабатывать классическим методом.

У большинства исследователей сложилось мнение, что топологический метод идентификации закона распределения случайной величины применять не следует. Это связано с тем, что данный подход требует большой вычислительной мощности, и ранее не было возможности проверить его работоспособность на практике. Сейчас же вычислительная способность компьютеров возросла, это позволяет создать программный продукт, реализующий топологический метод идентификации закона распределения случайной величины.

Объектом данной дипломной работы является закон распределения вероятностей.

Предмет представляет собой случайную выборку размера n.

Целью является идентификация закона распределения выборки по параметрам формы.

Задачи исследования.

1) Анализ теоретических источников об идентификации законов распределения.
2) Теоретически обосновать топологический метод для произвольного закона распределения случайной величины.
3) Выбор программной среды для выполнения статистических расчетов.
4) Разработка алгоритма топологической идентификации закона распределения вероятности случайной величины.
5) Разработка программной реализации алгоритма и апробация эффективности его работы.

1 Теоретический обзор

1.1 Случайная величина

Случайная величина - величина, которая в результате испытаний принимает то или иное возможное значение, зависящее от случая, т.е. от элементарного события. Причем она должна принимать только одно значение, которое ранее неизвестно, и изменяющееся от испытания к испытанию. Случайная величина характеризует результат испытания количественно, этим она отличается от случайного события, являющегося качественной характеристикой случайного результата испытания. Примерами случайной величины может служить число дефектных единиц продукции среди проверенных; число, выпавшее на верхней грани игрального кубика; размер обрабатываемой детали, погрешность результата измерения какого-либо параметра изделия или среды. Случайные величины, встречающиеся на практике, делятся на два основных типа: дискретные и непрерывные.

Дискретной называется случайная величина, принимающая бесконечное или конечное счетное множество значений. Например: частота попаданий при трех выстрелах; число отказов элементов прибора за определенный промежуток времени при испытании его надежности; число выстрелов до первого попадания в цель и т. д.

Непрерывная случайная величина - величина, принимающая любые возможные значения из некоторого конечного или бесконечного интервала. Следовательно, число возможных значений такой случайной величины бесконечно. К непрерывной случайной величине относятся: ошибка при измерении дальности радиолокатора; время безотказной работы микросхемы; погрешность изготовления деталей; концентрация соли в морской воде и т. д.

На практике случайные величины принято обозначать заглавными буквами X, Y и т. д., а их возможные значения строчными — x, y и т. д. Обычно случайная величина задается не только всеми принимаемыми ею значениями, но и частотой появления этих значений в результате испытаний при одних и тех же условиях. Эти частоты называются вероятностями. Совокупность всех возможных значений случайной величины и соответствующих им вероятностей составляет распределение случайной величины.

1.2 Закон распределения случайной величины

Законом распределения случайной величины называется соответствие между возможными значениями случайной величины и соответствующими им вероятностями. Про случайную величину говорят, что она подчиняется данному закону распределения. Две случайные величины называются ***независимыми***, если закон распределения одной из них не зависит то того, какие возможные значения приняла другая величина. В противном случае случайные величины называются ***зависимыми***. Несколько случайных величин называются ***взаимно независимыми***, если законы распределения любого числа из них не зависят от того, какие возможные значения приняли остальные величины.

Закон распределения случайной величины может быть задан в виде таблицы, функции распределения либо плотности распределения. Таблица, содержащая возможные значения случайной величины и соответствующие вероятности, является простейшей формой задания закона распределения случайной величины.

Табличное задание закона распределения можно использовать только для дискретной случайной величины с конечным числом возможных значений. Табличная форма задания закона случайной величины называется также рядом распределения.

Для наглядности ряд распределения представляют графически. При графическом изображении в прямоугольной системе координат по оси абсцисс откладывают все возможные значения случайной величины, а по оси ординат

— соответствующие вероятности. Точки (x_i, p_i), соединенные прямолинейными отрезками, представленные на рисунке 1, называют ***многоугольником распределения***. Следует помнить, что соединение точек (x_i, p_i) выполняется только с целью наглядности, так как в промежутках между x_1 и x_2, x_2 и x_3 и т. д. не существует значений, которые может принимать случайная величина X, поэтому вероятности её появления в этих промежутках равны нулю.

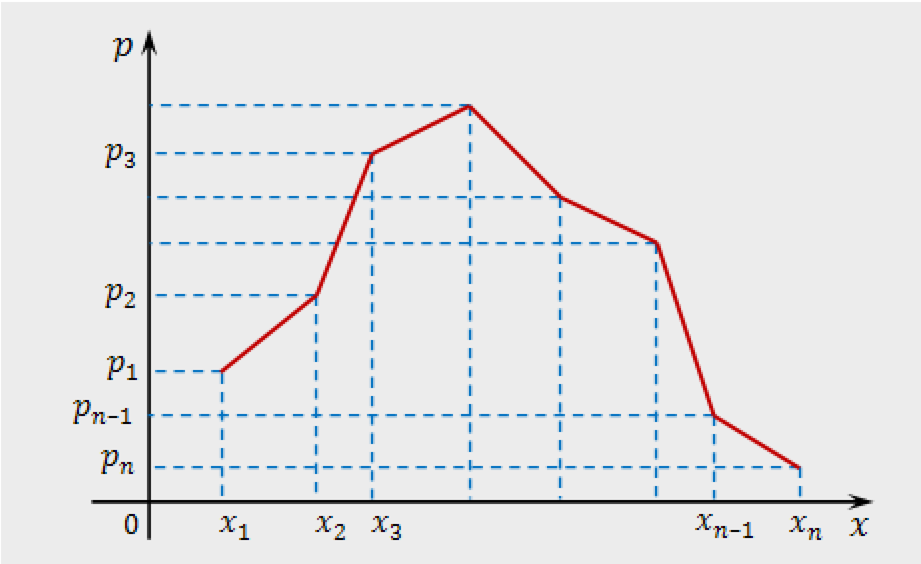


Рисунок 1 – Многоугольник распределения

Многоугольник распределения, как и ряд распределения, является одной из форм задания закона распределения дискретной случайной величины. Они могут иметь различную форму, однако все обладают одним общим свойством: сумма ординат вершин многоугольника распределения, представляющая собой сумму вероятностей всех возможных значений случайной величины, всегда равна единице. Это свойство следует из того, что все возможные значения случайной величины X образуют полную группу несовместных событий, сумма вероятностей которых равна единице.

1.3 Функция распределения вероятностей и ее свойства

Функция распределения является наиболее общей формой задания закона распределения. Она используется для задания как дискретных, так и непрерывных случайных величин. Обычно ее обозначают $F(x)$. ***Функция распределения*** определяет вероятность того, что случайная величина X принимает значения, меньшие фиксированного действительного числа x, т. е. $F(x) = P\{X < x\}$. Функция распределения полностью характеризует случайную величину с вероятностной точки зрения. Ее еще называют ***интегральной функцией распределения.***

Геометрическая интерпретация функции распределения очень проста. Если случайную величину рассматривать как случайную точку X на оси O_X, которая в результате испытания может занять то или иное положение на оси, то функция распределения $F(x)$ - это вероятность того, что случайная точка X в результате испытания попадет левее точки x, представленна на рисунке 2.

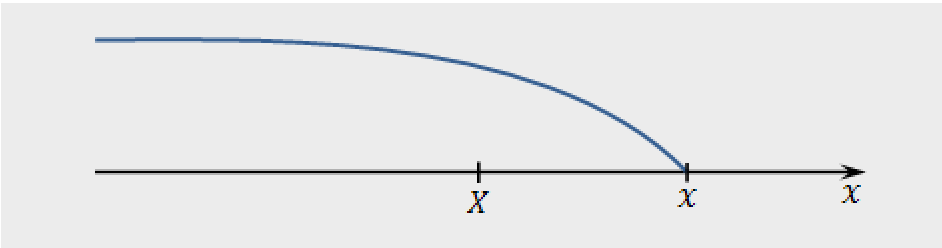


Рисунок 2 – Случайная точка на оси X

Для дискретной случайной величины X, которая может принимать значения $x_1, x_2, \dots, x_n$, функция распределения имеет вид

$$F(x) = \sum_{x_i < x} P\{ X = x_i \}, \tag{1.3.1}$$

где неравенство $x_i < x$ означает, что суммирование распространяется на все значения x_i, меньше x. Из этой формулы следует, что функция распределения дискретной случайной величины представляет собой ступенчатую ломаную линию, представленную на рисунке 3. При каждом новом значении случайной величины ступень поднимается выше на величину, равную вероятности этого значения. Сумма всех скачков функции распределения равна единице.

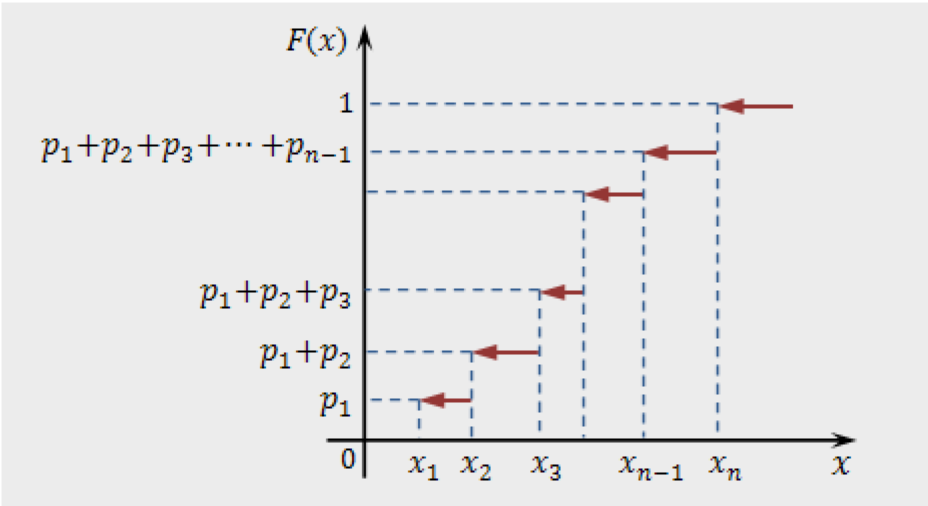


Рисунок 3 – Ломанная распределения

Непрерывная случайная величина имеет непрерывную функцию распределения, график этой функции имеет форму плавной кривой, представленной на рисунке 4.

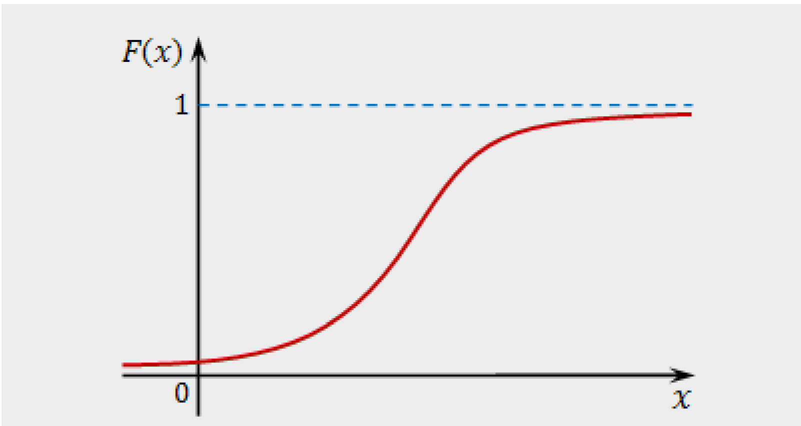


Рисунок 4 – График непрерывной функции распределения

Рассмотрим общие свойства функций распределения.

Свойство 1. Функция распределения — неотрицательная, функция, заключенная между нулем и единицей:

$$0 \leq F(x) \leq 1 \qquad (1.3.2)$$

Справедливость этого свойства вытекает из того, что функция распределения $F(x)$ определена как вероятность случайного события, состоящего в том, что $X < x$.

Свойство 2. Вероятность попадания случайной величины в интервал $(a,\ b)$ равна разности значений функции распределения на концах этого интервала, т. е.

$$P\{\, a \leq X \ \leq b \,\} = F(b) - \ F(a) \qquad (1.3.3)$$

Отсюда следует, что вероятность любого отдельного значения непрерывной случайной величины равна нулю.

Свойство 3. Функция распределения случайной величины есть неубывающая функция, т. е.:

$$F(b) \geq F(a). \qquad (1.3.4)$$

Свойство 4. На минус бесконечности функция распределения равна нулю, а на плюс бесконечности — единице, т. е.

$$\lim_{x \to -\infty} F(x) = 0 \quad (1.3.5)$$

$$\lim_{x \to +\infty} F(x) = 1 \quad (1.3.6)$$

1.4 Плотность распределения вероятностей и ее свойства

Функция распределения непрерывной случайной величины является ее вероятностной характеристикой. Но она имеет недостаток, заключающийся в том, что по ней трудно судить о характере распределения случайной величины в небольшой окрестности той или другой точки числовой оси. Более наглядное представление о характере распределения непрерывной случайной величины дает функция, которая называется плотностью распределения вероятности, или дифференциальной функцией распределения случайной величины.

Плотность распределения $f(x)$ равна производной от функции распределения $F(x)$, т. е.

$$f(x) = \frac{d}{dx}F(x) \quad (1.4.1)$$

Смысл плотности распределения $f(x)$ состоит в том, что она указывает на то, как часто случайная величина X появляется в некоторой окрестности точки x при повторении опытов.

Рассмотрим ***свойства плотности распределения:***

Свойство 1. Плотность распределения неотрицательна, т. е.

$$f(x) \geq 0. \tag{1.4.1}$$

Свойство 2. Функция распределения случайной величины равна интегралу от плотности в интервале от $-\infty$ до x, т. е.

$$F(x) = \int_{-\infty}^{x} f(t)\, dt \tag{1.4.2}$$

Свойство 3. Вероятность попадания непрерывной случайной величины на участок (a, b) равна интегралу от плотности распределения, взятому по этому участку, т. е.

$$P\{ a \leq X \leq b \} = \int_{a}^{b} f(x)\, dx \tag{1.4.3}$$

Свойство 4. Интеграл в бесконечных пределах от плотности распределения равен единице:

$$\int_{-\infty}^{+\infty} f(x)\, dx = 1 \tag{1.4.4}$$

1.5 Числовые характеристики случайных величин

Закон распределения полностью характеризует случайную величину с вероятностной точки зрения. Но при решении ряда практических задач нет необходимости знать все возможные значения случайной величины и соответствующие им вероятности, а удобнее пользоваться некоторыми количественными показателями. Такие показатели называются числовыми

характеристиками случайной величины. Основными из них являются математическое ожидание, дисперсия, моменты различных порядков, мода и медиана.

Математическое ожидание иногда называют средним значением случайной величины. Рассмотрим дискретную случайную величину X, принимающую значения $x_1, x_2, \dots, x_n$ с вероятностями соответственно $p_1, p_2, \dots, p_n$.

Определим среднюю арифметическую значений случайной величины, взвешенных по вероятностям их появлений. Таким образом, вычислим среднее значение случайной величины, или ее математическое ожидание $M(x)$:

$$M(x) = \frac{x_1 p_1 + x_2 p_2 + \ldots + x_n p_n}{p_1 + p_2 + \ldots + p_n} = \frac{\sum_{i=1}^{n} x_i p_i}{\sum_{i=1}^{n} p_i} \quad (1.5.1)$$

Учитывая, что $\sum_{i=1}^{n} p_i = 1$ получаем:

$$M(x) = \sum_{i=1}^{n} x_i p_i \quad (1.5.2)$$

Итак, ***математическим ожиданием*** дискретной случайной величины называется сумма произведений всех ее возможных значений на соответствующие вероятности.

Для непрерывной случайной величины математическое ожидание:

$$M(x) = \int_{-\infty}^{+\infty} x\, f(x)\, dx \quad (1.5.3)$$

1.5.1 Математическое ожидание случайной величины

Пусть X – неприрывная случайная величина, возможные значения которой принадлежат отрезку (a, b):

$$M(x) = \int_a^b x\, f(x)\, dx \tag{1.5.1.1}$$

Используя функцию распределения вероятностей $F(x)$, математическое ожидание случайной величины можно выразить так:

$$M(x) = \int_{-\infty}^{+\infty} x\, d(F(x)) \tag{1.5.1.2}$$

Некоторые свойства математического ожидания:

Свойство 1. Математическое ожидание суммы двух случайных величин равно сумме их математических ожиданий:

$$M(X+Y) = M(X) + M(Y) \tag{1.5.1.3}$$

Свойство 2. Математическое ожидание произведения двух независимых случайных величин равно произведению их математических ожиданий:

$$M(XY) = M(X)M(Y) \tag{1.5.1.4}$$

Свойство 3. Математическое ожидание постоянной величины равно самой постоянной:

$$M(c) = c \tag{1.5.1.5}$$

Свойство 4. Постоянный множитель случайной величины можно вынести за знак математического ожидания:

$$M(cX) = cM(X) \tag{1.5.1.6}$$

Свойство 5. Математическое ожидание отклонения случайной величины от ее математического ожидания равно нулю:

$$M(X - M(X)) = 0 \tag{1.5.1.7}$$

Модой M_0 дискретной случайной величины называется наиболее вероятное ее значение.

Модой M_0 непрерывной случайной величины называется такое ее значение, которому соответствует наибольшее значение плотности распределения. Геометрически моду интерпретируют как абсциссу точки глобального максимума кривой распределения, которую можно увидеть на рисунке 5.

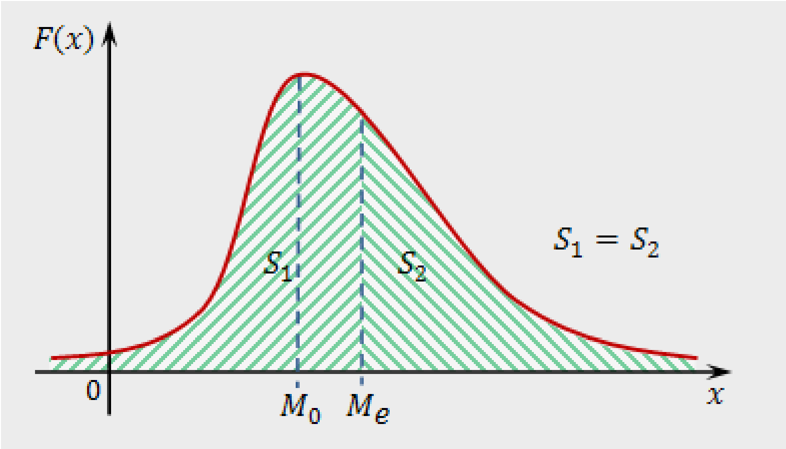


Рисунок 5 – Мода и медиана

Медианой M_e случайной величины называется такое ее значение, для которого справедливо равенство

$$P\{X < M_e\} = P\{X > M_e\} \quad (1.5.1.8)$$

С геометрической точки зрения медиана — это абсцисса точки, в которой площадь фигуры, ограниченной кривой распределения вероятностей и осью абсцисс, делится пополам (Рисунок 5). Так как вся площадь, ограниченная кривой распределения и осью абсцисс, равна единице, то функция распределения в точке, соответствующей медиане, равна 0,5, т. е.

$$F(M_e) = P\{X < M_e\} = 0.5 \quad (1.5.1.9)$$

С помощью дисперсии и среднеквадратического отклонения можно судить о рассеивании случайной величины вокруг математического ожидания. В качестве меры рассеивания случайной величины используют математическое ожидание квадрата отклонения случайной величины от ее математического ожидания, которое называют ***дисперсией случайной величины*** X и обозначают $D(X)$:

$$D(X) = M((X - M(X))^2) \tag{1.5.1.10}$$

Для дискретной случайной величины дисперсия равна сумме произведений квадратов отклонений значений случайной величины от ее математического ожидания на соответствующие вероятности:

$$D(X) = \sum_{i=1}^{n}(x_i - M(X))^2 \, p_i \tag{1.5.1.11}$$

Для непрерывной случайной величины, закон распределения которой задан плотностью распределения вероятности $f(x)$, дисперсия

$$D(X) = \int_{-\infty}^{+\infty} (x - M(x))^2 \, f(x) \, dx \tag{1.5.1.12}$$

Размерность дисперсии равна квадрату размерности случайной величины и поэтому ее нельзя интерпретировать геометрически. Этих недостатков лишено среднее квадратическое отклонение случайной величины, которое вычисляется по формуле

$$\sigma = \sqrt{D(X)} \tag{1.5.1.13}$$

Некоторые свойства дисперсии случайной величины:

Свойство 1. Дисперсия суммы двух независимых случайных величин равна сумме дисперсий этих величин:

$$D(X+Y) = D(X) + D(Y) \quad (1.5.1.14)$$

Свойство 2. Дисперсия случайной величины равна разности между математическим ожиданием квадрата случайной величины Xи квадратом ее математического ожидания:

$$D(X) = M(X^2) - (M(X))^2 \quad (1.5.1.15)$$

Свойство 3. Дисперсия постоянной величины равна нулю:

$$D(c) = 0 \quad (1.5.1.16)$$

Свойство 4. Постоянный множитель случайной величины, можно выносить за знак дисперсии, предварительно возведя его в квадрат:

$$D(cX) = c^2 D(X) \quad (1.5.1.17)$$

Свойство 5. Дисперсия произведения двух независимых случайных величин X и Y определяется по формуле

$$D(XY) = D(X)D(Y) + \left(M(X)\right)^2 D(Y) + \left(M(Y)\right)^2 D(X) \quad (1.5.1.18)$$

Обобщением основных числовых характеристик случайной величины является понятие моментов случайной величины.

Начальным моментом q-го порядка случайной величины называют математическое ожидание величины X^q:

$$\nu_q = M(X^q) \tag{1.5.1.19}$$

Начальный момент дискретной случайной величины

$$\nu_q = \sum_{i=1}^{n} x_i^q p_i, \tag{1.5.1.20}$$

Начальный момент непрерывной случайной величины

$$\nu_q = \int_{-\infty}^{+\infty} x^q f(x) dx \tag{1.5.1.21}$$

Центральным моментом q-го порядка случайной величины называют математическое ожидание величины $(X - M(X))^q$:

$$\nu_q = M((X - M(X))^q) \tag{1.5.1.22}$$

Центральный момент дискретной случайной величины

$$\nu_q = \sum_{i=1}^{n} (x_i - M(X))^q p_i, \tag{1.5.1.23}$$

Центральный момент непрерывной случайной величины

$$\nu_q = \int_{-\infty}^{+\infty} \left(x - M(X)\right)^q f(x)dx \qquad (1.5.1.24)$$

Начальный момент первого порядка представляет собой математическое ожидание, а центральный момент второго порядка — дисперсию случайной величины.

Нормированный центральный момент третьего порядка служит характеристикой скошенности или асимметрии распределения (***коэффициент асимметрии***), представленный на рисунке 6:

$$A_s = \frac{\mu_3}{\sigma^3} \qquad (1.5.1.25)$$

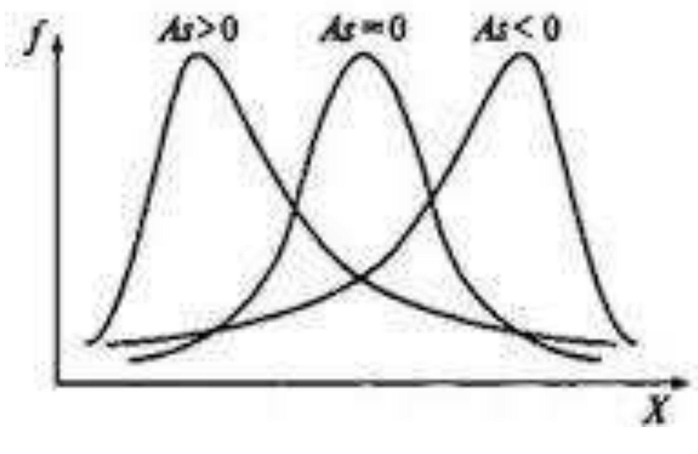


Рисунок 6 – Асимметрия

Асимметрия – свойство распределения выборки, характеризующее несимметричность распределения случайной величины. На практике симметричные распределения встречаются достаточно редко.

Асимметрия может быть положительной, равной нулю или отрицательной. На рисунке 6 видно, что при положительном значении график плотности распределения сдвигается влево (левосторонняя асимметрия), т.е. длинная часть кривой распределения расположена справа от моды, а отрицательный, в свою очередь, вправо (правосторонняя асимметрия) – длинная часть кривой расположена слева от моды.

Если $A_s < 0.25$, то это слабая асимметрия; если она находится в пределах от 0.25 и до 0.5, то умеренная асимметрия; а если $A_s > 0.5$, то данное распределение считается крайне асимметричным.

Нормированный центральный момент четвертого порядка служит характеристикой островершинности или плосковершинности распределения (***эксцесс***), представленный на рисунке 7:

$$E = \frac{\mu_4}{\sigma^4} - 3 \tag{1.5.1.26}$$

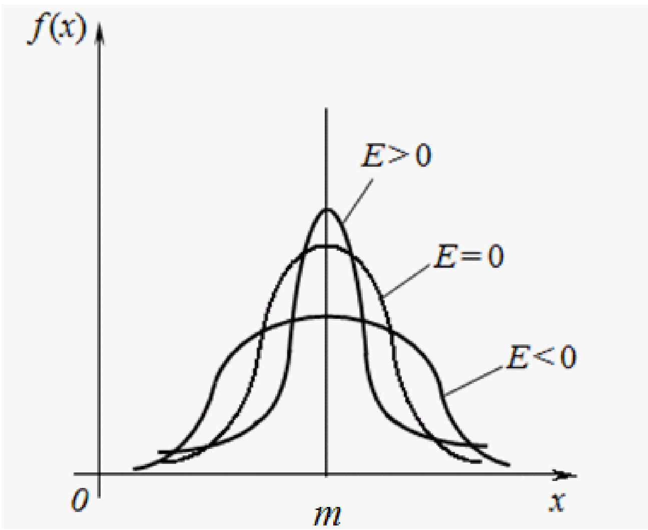


Рисунок 7 – Эксцесс

Эксцесс – мера крутости кривой распределения случайной величины. Данная кривая может быть островершинной, средне вершинной и плосковершинной (рисунок 7).

В качестве эталонного значения – своеобразного начала отсчета в измерении асимметрии и эксцесса служит нормальное (гауссовское) распределение, для которого значения этих параметров соответственно равны 1 и 0.

1.6 Законы распределения дискретных случайных величин

1.6.1 Биномиальный закон распределения

В общей форме биномиальный закон, представленный на рисунке 8, описывает осуществление признака в n испытаниях с возвратом. Наглядной схемой таких испытаний является последовательный выбор с возвращением шаров из урны, содержащей m_1 белых и m_2 чёрных шаров. Если X - число появления белых шаров в выборке из $n \leq m_1 + m_2$ шаров, то

$$P\{ X = k \} = C_n^k p^k (1-p)^{n-k} = C_n^k p^k q^{n-k}, \tag{1.6.1.1}$$

где p, q - вероятность появления при одном извлечении соответственно белого и чёрного:

$$p = \frac{m_1}{m_1 + m_2} \tag{1.6.1.2}$$

$$q = 1 - p \tag{1.6.1.3}$$

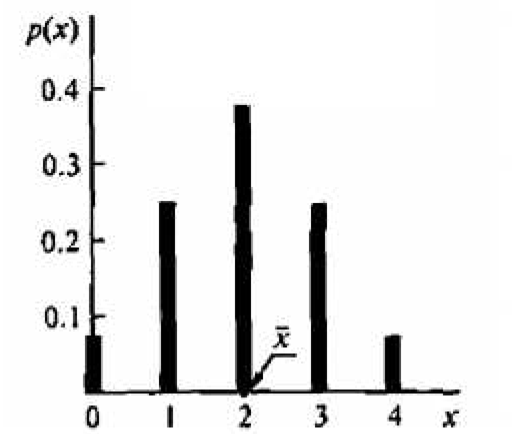


Рисунок 8 – Биномиальное распределение с параметрами p = 0.5 и n = 4

Некоторые характеристики биномиального распределения:

$$M(X) = np \qquad (1.6.1.4)$$

$$D(X) = npq \qquad (1.6.1.5)$$

$$A_s = \frac{q-p}{\sqrt{npq}} \qquad (1.6.1.6)$$

$$E = \frac{1-6pq}{npq} \qquad (1.6.1.7)$$

1.6.2 Закон распределения Пуассона

Случайная величина X называется распределённой по закону Пуассона, представленного на рисунке 10, с параметром $\lambda > 0$, если

$$P\{ X = k \} = \frac{\lambda^k}{k!} e^{-\lambda}, k = 0, 1, 2, \dots \quad (1.6.2.1)$$

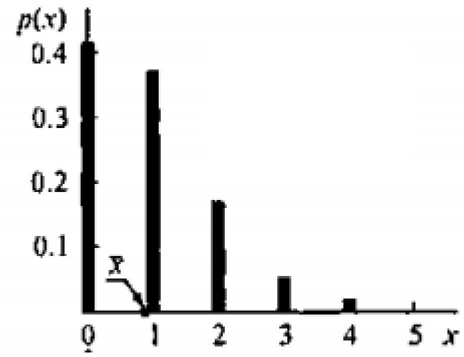


Рисунок 9 – Распределение Пуассона с параметром λ = 0.9

Характерной особенностью распределения Пуассона являются совпадения математического ожидания и дисперсии, причём

$$M(X) = D(X) = \lambda \quad (1.6.2.2)$$

Распределение Пуассона можно получить из биномиального распределения путём предельного перехода при $n \to \infty, p \to 0$ при условии $np = \lambda = const$ и в этом случае интерпретируется как закон “редких” явлений. Если n достаточно велико, а p мало, то формулу Пуассона часто используют в качестве приближения вместо точных биномиальных формул для вероятностей k успехов в n испытаниях.

К случайным величинам, подчинённым закону Пуассона, приводит большое количество задач, относящих к вопросам массового обслуживания. В качестве примера рассмотрим работу телефонной станции. Можно доказать, что при выполнении некоторых условий вероятность k вызовов за промежуток времени t определяется формулой

$$P\{ X = k \} = \frac{at^k}{k!} e^{-at} \tag{1.6.2.3}$$

где X – количество вызовов.

Если положить $at = \lambda$, то последняя формула означает, что случайная величина распределена по закону Пуассона.

Асимметрия и эксцесс распределения Пуассона:

$$A_s = \frac{1}{\sqrt{\mu}} \tag{1.6.2.4}$$

$$E = \frac{1}{\mu} \tag{1.6.2.5}$$

1.6.3 Геометрический закон распределения

Последовательно проводится несколько независимых испытаний до появления некоторого события A, вероятность которого в каждом испытании одна и та же и равна p. Тогда число X произведённых испытаний есть дискретная случайная величина с геометрическим распределением вероятности. Примером может служить стрельба по некоторой цели до первого попадания, причём вероятность попадания при каждом выстреле не зависит от результатов предыдущих выстрелов и сохраняет постоянное значение p. Число X произведённых выстрелов будет случайной величиной, возможные значения которой являются все натуральные числа. Геометрический закон распределения, представленный на рисунке 10, задаётся формулой

$$P\{ X = k \} = q^{k-1}p, k = 1, 2, \dots \tag{1.6.3.1}$$

где $q = 1 - p$.

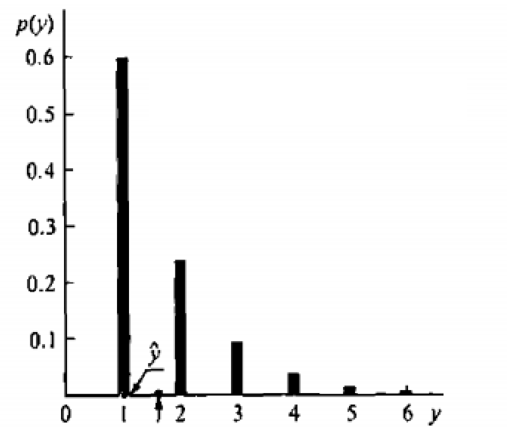


Рисунок 10 – Геометрическое распределение с параметром p = 0.6

Некоторые характеристики геометрического закона распределения:

$$M(x) = \frac{1}{p} \tag{1.6.3.2}$$

$$D(X) = \frac{q}{p^2} \tag{1.6.3.3}$$

$$A_s = \frac{1+q}{\sqrt{q}} \tag{1.6.3.4}$$

$$E = \frac{p^2}{q} + 6 \tag{1.6.3.5}$$

1.6.4 Равномерный закон распределения

Равномерное распределение задаётся следующим законом:

$$P\{X = k\} = \frac{1}{n}, k = 1, 2, \ldots, n \tag{1.6.4.1}$$

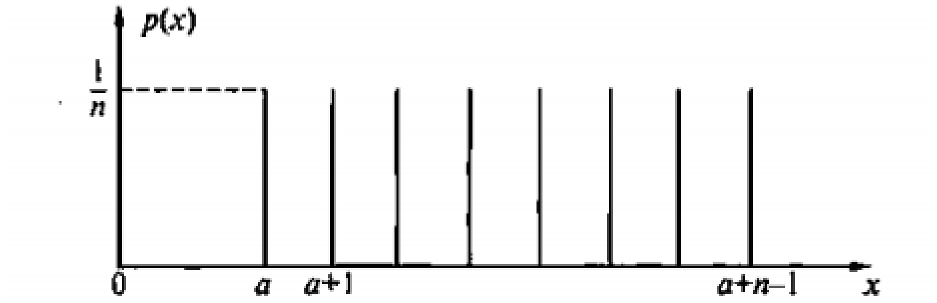


Рисунок 11 – Равномерное дискретное распределение

Этот закон имеет место в случае, когда n возможных исходов испытания равновероятны. Примером целочисленной случайной величины, распределённой по равномерному закону, может служить число очков, выпадающих при бросании симметричной кости (любое из значений $k = 1, 2, \ldots, 6$ выпадает с одинаковой вероятностью 1 / 6).

Некоторые числовые характеристики равномерного закона распределения:

$$M(x) = \frac{n+1}{2} \tag{1.6.4.2}$$

$$D(X) = \frac{n^2 - 1}{12} \tag{1.6.4.3}$$

$$A_s = 0 \tag{1.6.4.4}$$

$$E = -1.2 - \frac{2.4}{n^2-1} \tag{1.6.4.5}$$

1.6.5 Гипергеометрический закон распределения

Случайная величина X имеет гипергеометрическое распределение с параметрами n_1, n_2 и $n \leq n_1 + n_2$ (n, n_1, n_2— натуральные числа), представленное на рисунке 12, если она принимает конечное множество натуральных значений $\{k_1, \ k_1 + 1, \ldots, \ k_2\}$ соответственно с вероятностями

$$P\{X = k\} = \frac{C_{n_1}^{k} C_{n_2}^{n-k}}{C_{n_1+n_2}^{n}}, \qquad k_1 \leq k \leq k_2 \tag{1.6.5.1}$$

причём $k_1 = \max\{0, \ n - n_1\}, k_2 = \min\{n, \ n_1\}$.

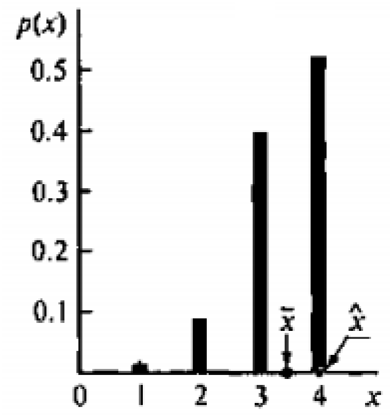


Рисунок 12 – Гипергеометрическое распределение с параметрами n = 4, $n_1 = n_2$ = 14

Гипергеометрическое распределение возникает в экспериментах по выбору без возвращения n шаров из урны, содержащей $n_1 + n_2$ шаров, из которых n_1 белых и n_2 чёрных. Таким образом, это распределение описывает осуществление признака в выборке без возврата (в отличии от биномиального распределения). На практике к гипергеометрическому распределению приводят задачи, где изделия из партии отбирают случайно (обеспечивая для каждого изделия равную возможность быть отобранным), но отобранные изделия не возвращают в партию. Такой отбор особенно важен в тех задачах, где проверка изделия связана с его разрушением (например, проверка изделия на срок службы).

Некоторые числовые характеристики гипергеометрического распределения:

$$M(x) = \frac{nn_1}{n_1+n_2} \qquad (1.6.5.2)$$

$$D(X) = \frac{nn_1n_2}{(n_1+n_2)^2}\frac{n_1+n_2-n}{n_1+n_2-1} \qquad (1.6.5.3)$$

Следует заметить, что если $n_1 + n_2$ очень велико по сравнению с n, то не имеет существенного значения, возвращаются шары обратно или нет, и формула гипергеометрического распределения может быть приближённо заменена формулой биномиального распределения.

1.7 Законы распределения непрерывных случайных величин

1.7.1 Нормальный закон распределения

Плотность вероятности нормально распределённой случайной величины X выражается формулой

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \qquad (1.7.1.1)$$

Кривая распределения изображена на рисунке 13. Она симметрична относительно точки $x = \mu$ (точка максимума). При уменьшении σ ордината точки максимума неограниченно возрастает, при этом кривая пропорционально сплющивается вдоль оси абсцисс, так что площадь под её графиком остаётся равной единицы.

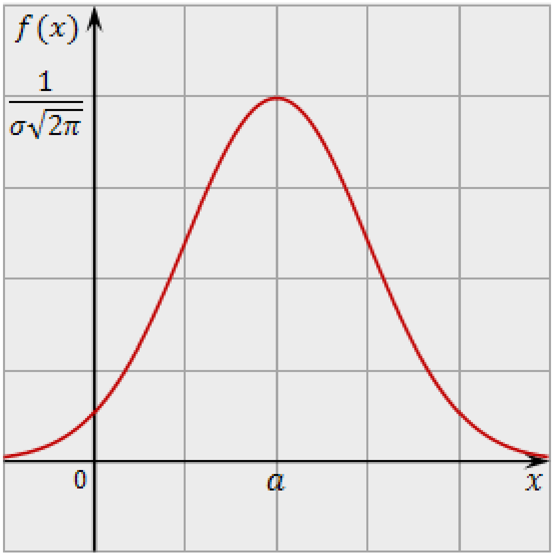


Рисунок 13 – Нормальное распределение

Нормальный закон распределения широко применяется в задачах практики. Объяснить причины этого впервые удалось Ляпунову. Он показал, что если случайная величина может рассматриваться как сумма большого числа малых слагаемых, то при достаточно общих условиях закон распределения этой случайной величины близок к нормальному независимо от того, каковы законы распределения отдельных слагаемых. А так как практически случайные величины в большинстве случаев бывают результатом действия множества причин, то нормальный закон оказывается наиболее распространённым законом распределения. Укажем некоторые числовые характеристики нормально распределённой случайной величины:

$$M(X) = \mu \tag{1.7.1.2}$$

$$D(X) = \sigma^2 \tag{1.7.1.3}$$

$$A_s = E = 0 \tag{1.7.1.4}$$

Таким образом, параметры μ и σ в выражении нормального закона распределения представляют собой математическое ожидание и среднее квадратическое отклонение случайной величины. Математическое ожидание и дисперсия полностью характеризуют нормально распределённую случайную величину. Разумеется, что в общем случае, когда характер закона распределения неизвестен, знание математического ожидания и дисперсии недостаточно для определения этого закона распределения.

1.7.2 Логарифмическое нормальное распределение

Говорят, что случайная величина Y имеет логарифмически нормальное распределение (сокращённо **логнормальное распределение**), если её логарифм $\ln Y = X$ распределён нормально, то есть если

$$Y = e^X \tag{1.7.2.1}$$

где величина X имеет нормальное распределение с параметрами μ, σ.

Плотность логнормального распределения задаётся формулой

$$f(y) = \frac{1}{y\sigma\sqrt{2\pi}} e^{\frac{-(\ln y - \mu)^2}{2\sigma^2}}, y > 0 \tag{1.7.2.2}$$

Математическое ожидание и дисперсию логнормального распределения определяют по формулам

$$M(Y) = e^{\mu + \frac{\sigma^2}{2}} \tag{1.7.2.3}$$

$$D(Y) = e^{2(2\sigma^2 + \mu)^2 - \mu^2} - e^{2\mu + \sigma^2} \tag{1.7.2.4}$$

Кривая этого распределения изображена на рисунке 14.

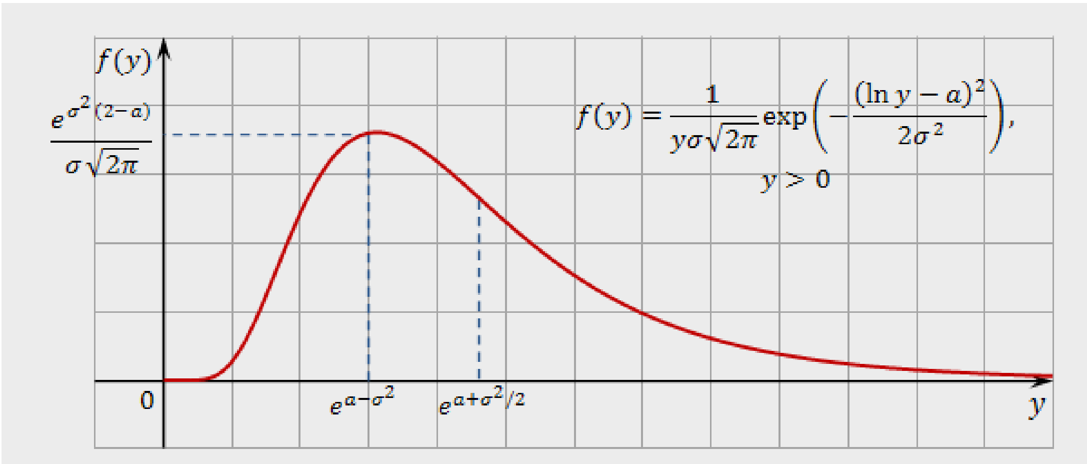


Рисунок 14 – Логнормальное распределение

Логарифмически нормальное распределение встречается в ряде технических задач. Оно даёт распределение размеров частиц при дроблении, содержаний элементов в минералах в изверженных горных пародах, численности рыб в море и т.д. Встречается такое распределение во всех задачах, где логарифм рассматриваемой величины можно представить в виде суммы большого количества независимых равномерно малых величин:

$$\ln Y = X = X_1 + X_2 + \ldots + X_n = \sum_{k=1}^{n} X_k \,, \tag{1.7.2.5}$$

то есть $Y = \prod_{k=1}^{n} e^{X_k}$, где e^{X_k} независимы.

Асимметрия и эксцесс логнормального распределения:

$$A_s = \left(e^{\sigma^2} + 2\right)\sqrt{e^{\sigma^2} - 1} > 0 \tag{1.7.2.6}$$

$$E = e^{4\sigma^2} + 2e^{3\sigma^2} + 3e^{2\sigma^2} - 6 \tag{1.7.2.7}$$

1.7.3 Гамма-распределение

Говорят, что случайная величина X имеет гамма-распределение с параметрами $a > 0$ и $b > 0$, если её плотность распределения вероятностей имеет вид

$$f(x) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx}, x > 0 \tag{1.7.3.1}$$

где $\Gamma(a) = \int_0^\infty t^{a-1}e^{-t}dt$ - гамма-функция Эйлера.

На рисунке 15 показаны кривые распределения вероятностей при значениях параметра $a > 1$ и $a < 1$ (при $a = 1$ получаем экспоненциальное распределение).

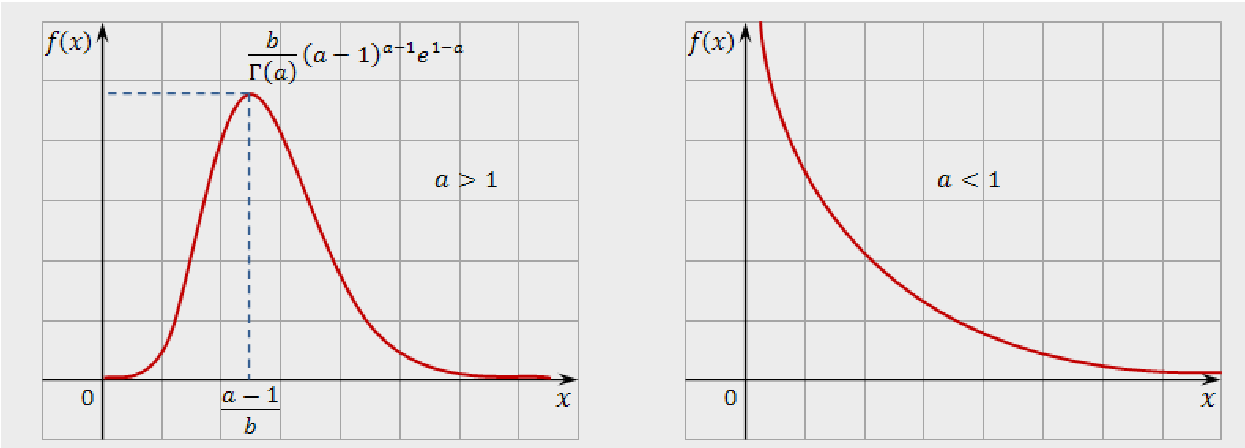


Рисунок 15 – Гамма-распределение

Математическое ожидание и дисперсия, подчинённые гамма-распределению, задаются формулами

$$M(X) = \frac{a}{b} \tag{1.7.3.2}$$

$$D(X) = \frac{a}{b^2} \tag{1.7.3.3}$$

Отметим, что при $a > 1$ гамма-распределение имеет моду

$$M_o = \frac{a-1}{b} \tag{1.7.3.4}$$

(графически это означает, что кривая распределения имеет точку максимума $x = M_o$, рисунок 14).

Асимметрия и эксцесс гамма-распределения:

$$A_s = \frac{1}{\sqrt{a}} \tag{1.7.3.5}$$

$$E = \frac{6}{a} \tag{1.7.3.6}$$

1.7.4 Экспоненциальный закон

Экспоненциальным распределением называется частный случай гамма-распределения с параметрами $a = 1, b = \lambda > 0$, то есть плотность вероятности в этом случае

$$f(x) = \lambda e^{-\lambda x}, x > 0 \tag{1.7.4.1}$$

Используя свойство два плотности распределения, можно найти функцию распределения $F(x)$ экспоненциального закона:

$$F(x) = 1 - e^{-\lambda x}, x \geq 0 \tag{1.7.4.2}$$

Основные характеристики (математическое ожидание и дисперсия) случайной величины X, распределённой по экспоненциальному, имеют вид

$$M(X) = \frac{1}{\lambda} \quad (1.7.4.3)$$

$$D(X) = \frac{1}{\lambda^2} \quad (1.7.4.4)$$

Кривая экспоненциального распределения вероятностей показана на рисунке 16.

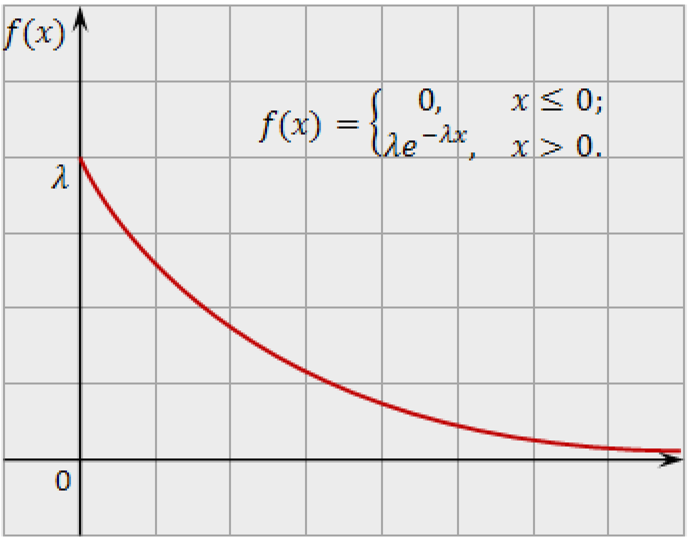


Рисунок 16 – Экспоненциальное распределение

Статистический смысл параметра λ состоит в следующем: λ есть среднее число событий на единицу времени, то есть $\frac{1}{\lambda}$ есть средний промежуток времени между двумя последовательными событиями.

Экспоненциальное (показательное) распределение часто встречается в теории массового обслуживания (например, X - время ожидания при техническом обслуживании или X - продолжительность телефонных разговоров, ежедневно регистрируемых на телефонной станции) и теории надёжности (например, X - срок службы радиоэлектронной аппаратуры).

Асимметрия и эксцесс экспоненциального распределения:

$$A_s = 2 \quad (1.7.4.5)$$

$$E = 6 \quad (1.7.4.6)$$

1.7.5 Распределение Вейбулла

Случайная величина X подчиняется закону распределения Вейбулла, представленное на рисунке 17, с параметрами $n \in \mathbb{N}, a \in \mathbb{R}, b > 0$, если её плотность распределения вероятностей записывается в виде

$$f(x) = \frac{n}{b}\left(\frac{x-a}{b}\right)^{n-1} e^{-\left(\frac{x-a}{b}\right)^n}, x > a \quad (1.7.5.1)$$

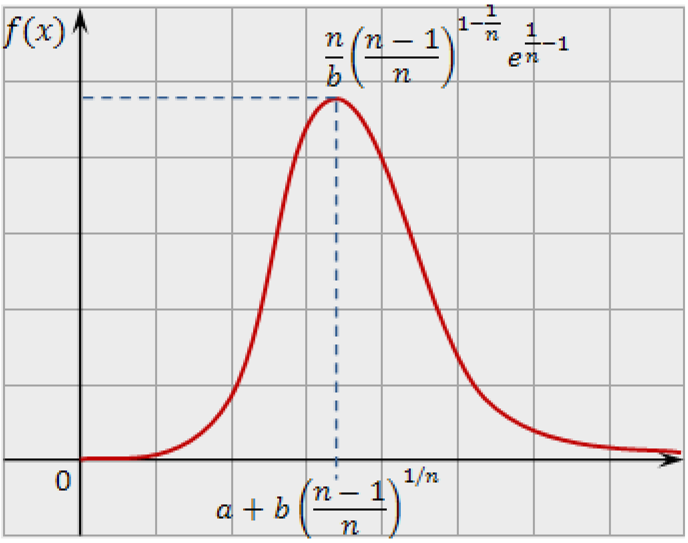


Рисунок 17 – Распределение Вейбула

Математическое ожидание и мода случайной величины, распределённые по закону Вейбулла, имеют следующий вид:

$$M(X) = a + b\Gamma(1 + \frac{1}{n}) \qquad (1.7.5.2)$$

$$M_o = a + \sqrt[n]{\frac{n-1}{n}} \qquad (1.7.5.3)$$

Распределение Вейбулла в ряде случаев характеризует срок службы радиоэлектронной аппаратуры и, кроме того, применяется для аппроксимации различных несимметричных распределений в математической статистике.

1.7.6 Равномерное распределение

Случайная величина X, представленная на рисунке 18, называется распределённой равномерно на отрезке $[a, b]$, если её плотность распределения вероятностей постоянна на данном отрезке:

$$f(x) = \frac{1}{b-a}, \qquad a \le x \le b \qquad (1.7.6.1)$$

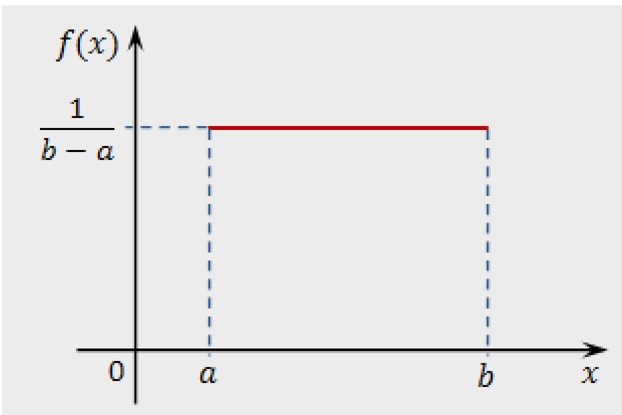


Рисунок 18 – Равномерное распределение

Все возможные значения равномерно распределённой случайной величины лежат в пределах некоторого интервала; кроме того. в пределах этого интервала все значения случайной величины одинаково вероятны (обладаю одной и той же плотностью вероятности). Равномерно распределение реализуется в экспериментах, где наудачу ставиться точка на отрезке $[a, b]$ (X - абсцисса поставленной точки). Равномерно распределённая случайная величина встречается также в измерительной практике при округлении отчётов измерительных приборов до целых делений шкал. Ошибка при округлении отчётов до ближайшего целого деления является случайной величиной X, которая может принимать с постоянной плотностью вероятности любое значение между двумя соседними целыми делениями.

Математическое ожидание и дисперсия равномерно распределённой случайной величины

$$M(X) = \frac{a+b}{2} \tag{1.7.6.2}$$

$$D(X) = \frac{(b-a)^2}{12} \tag{1.7.6.3}$$

Асимметрия и эксцесс равномерного распределения

$$A_s = 0 \tag{1.7.6.4}$$

$$E = -1.2 \tag{1.7.6.5}$$

1.7.7 Распределение хи-квадрат

Частный случай гамма-распределения с параметрами $a = \frac{n}{2}, n \in \mathbb{N}, b = 0.5$ называется распределением хи-квадрат с n степенями свободы (пишут $\chi^2(n)$). Если случайная величина X подчиняется закону $\chi^2(n)$, то её плотность распределения вероятностей есть

$$f(x) = \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, x > 0 \qquad (1.7.7.1)$$

Основные характеристики распределение хи-квадрат:

$$M(x) = n \qquad (1.7.7.2)$$

$$D(x) = 2n \qquad (1.7.7.3)$$

$$A_s = \sqrt{\frac{8}{n}} \qquad (1.7.7.4)$$

$$E = \frac{12}{n} \qquad (1.7.7.5)$$

Кривые распределения (для различных значений n) изображены на рисунке 19.

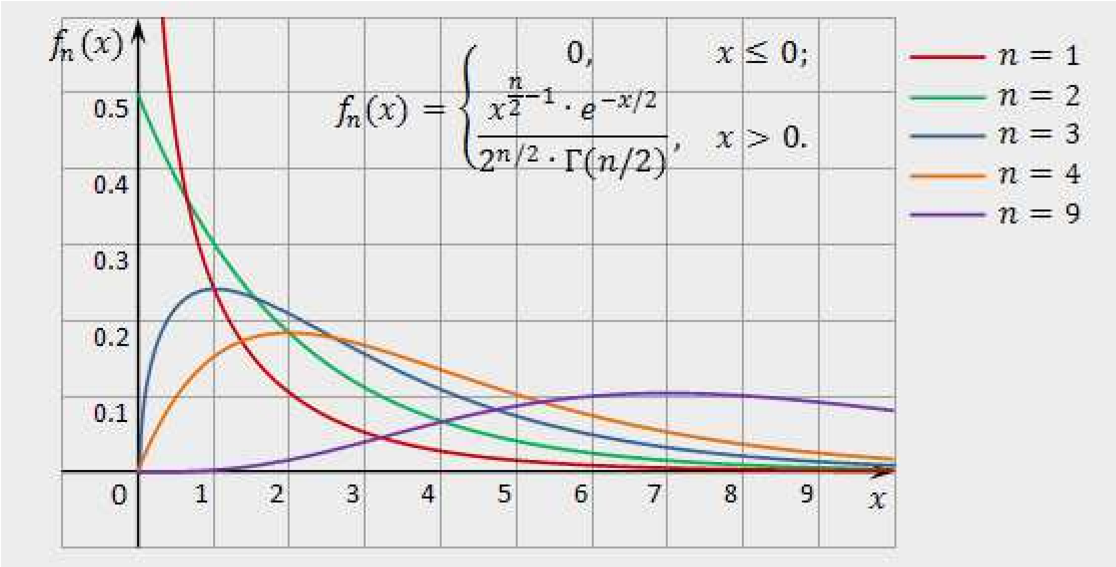


Рисунок 19 – Распределение хи-квадрат

Случайная величина $X = \chi^2(n)$, подчиняющаяся хи-квадрат распределению, равна сумме квадратов n независимых случайных величин $U_j, j \in \mathbb{N}$, каждая из которых имеет стандартизированное нормальное распределение, то есть

$$\chi^2(n) = U_1^2 + U_2^2 + \dots + U_n^2 \tag{1.7.7.6}$$

Пусть $\chi^2(n_1)$ и $\chi^2(n_2)$ - независимые случайные величины, имеющие хи-квадрат распределение со степенью свободы соответственно n_1 и n_2. Сумма этих случайных величин имеет также хи-квадрат распределение с $n_1 + n_2$ степенями свободы:

$$\chi^2(n_1) + \chi^2(n_2) = \chi^2(n_1 + n_2) \tag{1.7.7.7}$$

Заметим, что распределение $\chi^2(n)$ при больших значениях n $(n > 30)$ с достаточной для практических расчётов точностью аппроксимируется нормальным распределением с математическим ожиданием n и дисперсией $2n$. Поэтому при больших значениях n вероятности рассчитываются по нормальному закону.

Распределение $\chi^2(n)$ играет большую роль в математической статистике.

1.7.8 Распределение Стьюдента

Случайная величина $T(n)$ есть отношение двух независимых случайных величин U и $\sqrt{\frac{\chi^2(n)}{n}}$, то есть

$$T(n) = \frac{U}{\sqrt{\frac{\chi^2(n)}{n}}} \tag{1.7.8.1}$$

Распределение случайной величины $T(n)$ называется распределением Стьюдента с $n-1$ степенями свободы. Его плотность задаётся формулой

$$f(x) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{\pi n}\Gamma(\frac{n}{2})}\,(1+\frac{x}{n})^{-\frac{n+1}{2}} \tag{1.7.8.2}$$

Математическое ожидание и дисперсия случайной величины, подчинённой распределению Стьюдента $X = T(n)$, есть

$$M(x) = 0 \tag{1.7.8.3}$$

$$D(X) = \frac{n}{n-2} \qquad (1.7.8.4)$$

Кривые распределения Стьюдента (для различных значений n) изображены на рисунке 20.

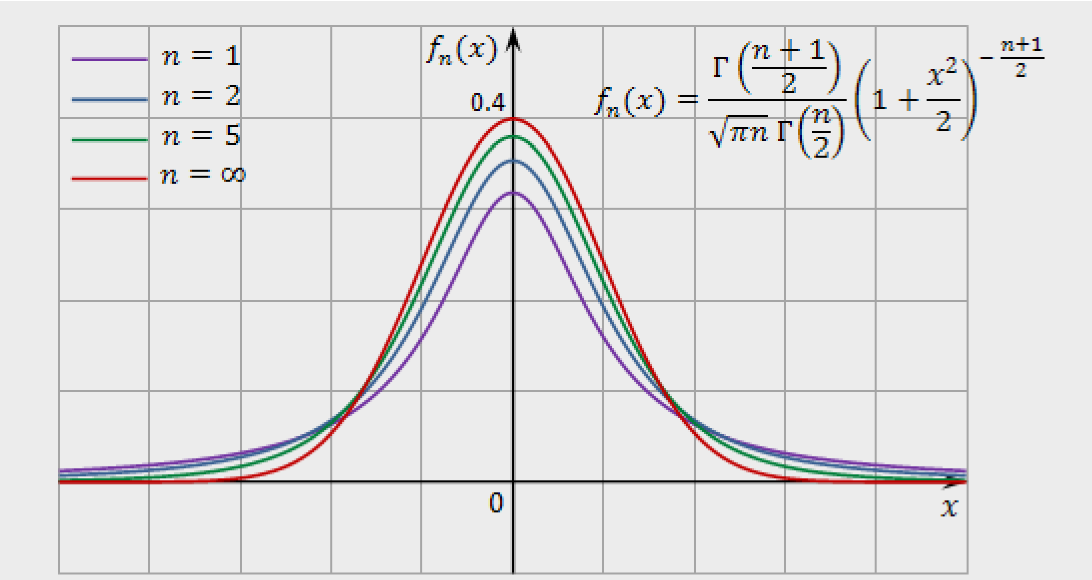


Рисунок 20 – Распределение Стьюдента

Как и в случае и хи-квадрат распределением, при увеличении n распределение Стьюдента стремиться к нормальному, более того, стандартизованному нормальному (то есть с нулевым математическим ожиданием и единичной дисперсией). Распределение Стьюдента, как хи-квадрат распределение, широко применяется в задачах математической обработки измерений.

Асимметрия и эксцесс распределения Стьюдента:

$$A_s = 0 \qquad (1.7.8.4)$$

$$E = \frac{6}{n-4} \qquad (1.7.8.5)$$

1.7.9 Распределение Фишера

Пусть случайная величина $F(n_1,\ n_2)$ равна отношению двух независимых случайных величин $\frac{\chi^2(n_1)}{n_1}$ и $\frac{\chi^2(n_2)}{n_2}$, то есть

$$F(n_1,\ n_2) = \frac{\frac{\chi^2(n_1)}{n_1}}{\frac{\chi^2(n_2)}{n_2}} \qquad (1.7.9.1)$$

Распределение случайной величины $F(n_1, n_2)$ называется распределением Фишера с n_1и n_2 степенями свободы. Оно имеет следующую плотность вероятности

$$f(x) = \frac{\Gamma\left(\frac{n_1+n_2}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)}\left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}}\frac{x^{\frac{n_1}{2}-1}}{\left(1+\frac{n_1}{n_2}x\right)^{\frac{n_1+n_2}{2}}}, x > 0 \qquad (1.7.9.2)$$

Математическое ожидание случайной величины, подчинённой распределению Фишера, $X = F(n_1, n_2)$ определяется по формуле

$$M(X) = \frac{n_2}{n_2 - 2}, n_2 > 2 \qquad (1.7.9.3)$$

Между случайными величинами, имеющими нормальное распределение: хи-квадрат, Стьюдента и Фишера, имеют место соотношения

$$T^2(n) = F(1,n) \quad (1.7.9.4)$$

$$F(n,\infty) = \frac{\chi^2(n)}{n} \quad (1.7.9.5)$$

$$\chi^2(1) = U^2 \quad (1.7.9.6)$$

1.8 Выбор программной среды для статистических расчетов

Существует множество программных решения для проведения статистических вычислений и расчетов. Каждый из них по-своему хорош, уникален и применим к определенной задаче.

Язык программирования и среда статистических расчетов ***R*** применима для статистической обработки данных, математического моделирования и работы с графикой. Но лучше всего он справляется с задачами статистики: от вычисления средних величин и до вейвлет – преобразований временных рядов. Среда R – среда с открытым исходным кодом, это означает, что она абсолютна бесплатна. Так же язык R отлично интегрируется с другими программными продуктами в частности с Visual Studio 2017 Community. Он интуитивно понятен и удобен в использовании, а синтаксис достаточно прост и легок в использовании

Высокоуровневый язык и интерактивная среда для программирования численных расчетов и отображения результатов ***MATLAB*** применяется для анализа данных, разработки алгоритмов, моделирования объектов, разработки систем управления, проектирования коммуникационных систем, создания приложений и многого другого. Он содержит в себе множество математических функций, например, преобразование Фурье. Этот продукт лицензирован и имеет пробную версию.

Программный пакет, система компьютерной алгебры ***Maple*** ориентирована на сложные математические вычисления, визуализацию полученных данных и моделирование. Содержит в себе множество алгоритмов и правил преобразования математических выражений. Эта программа легко интегрируется с Ecxel и MATLAB.

Для решения поставленной задачи следует остановить свой выбор на языке статистических расчетов R, так как он распространяется в свободном доступе, интегрируется со средой разработки Visual Studio. Также для проведения необходимых исследований нам не потребуется графической интерпретации результатов в R так как, программа проверяющая работоспособность алгоритма идентификации закона распределения написана на языке программирования C# и вся необходимая информация отображается на форме – окне приложения, которое, в свою очередь, подключается к среде статистических расчетов R.

2 Топологическая идентификация закона распределения вероятности

Разработанный алгоритм определения закона распределения вероятности случайной величины по параметрам формы, представленный в данной работе основан на оценках моментов высших порядков (3 и 4 моменты): асимметрии и эксцесса. Которые получены по генеральной совокупности измерений случайной величины.

Для достоверного определения «нормальности» высчитываются коэффициенты асимметрии и эксцесса (формулы 2.1.1 и 2.1.2.), где в знаменатель подставлены их ошибки репрезентативности. Причем эти коэффициенты должны удовлетворять системе 2.1.3 – это означает, что распределение достоверно нормально. В противном случае достоверно ненормально.

Для принятия решения выдвигается непараметрическая гипотеза о виде закона распределения случайной величины и осуществляется проверка выборочных данных выдвинутой (предполагаемой) гипотезе. Параллельно ей исследователь рассматривает и другую противоречащую ей, альтернативную (конкурирующую) гипотезу. Эта гипотеза считается справедливой, если основная гипотеза отвергается. При таком подходе могут возникнуть две ситуации: основная гипотеза может подтвердиться, а может быть отвергнута. Если главная гипотеза верна, а исследователь принял иное решение, отвергнув ее, то данная ошибка будет являться ошибкой первого рода. Если основная гипотеза неверна, а лицо, принимающее решение (ЛПР) делает упор на ее достоверности, то такая ошибка является ошибкой второго рода.

2.1 Алгоритм идентификации

Пусть дана выборка некоторых случайных величин: $X_1, X_2, \dots, X_n$ размера n; список теоретических распределений $Distributions$. В дальнейшем из этих распределений будет выбираться то, которое по параметрам формы (полигона) наилучшим образом описывает закон распределения, из которого взята эта выборка.

Шаг 1: отбраковка аномальный измерений (выбросов) по критерию Грабса. Оценивание параметров по выборке каждого теоретического распределения ($F_1 \dots F_k$) в массиве $Distributions$ методом максимального правдоподобия или методом моментов (например, для нормального распределения – параметры положения и масштаба или математическое ожидание и дисперсия и т.д.).

Шаг 2: выдвижение гипотезы h_1 о законе распределения вероятности случайной величины.

Шаг 3: преобразование всех точек выборки таким образом, чтобы они стали равномерно распределенными. Для этого необходимо подставить значения случайной величины в формулу предполагаемого закона распределения вероятности F_1 из гипотезы h_1. При этом следует использовать параметры, вычисленные на шаге 1.

Это позволяет нам нормировать наши данные, т.е. привести их к одинаковому типу, к одной шкале. Чтобы в дальнейшим погрешность результата работы была минимальна.

Шаг 4: приведение точек выборки, полученных на шаге 3 к нормальному закону распределения вероятности с параметрами, полученными на шаге 1.

Шаг 5: нахождение асимметрии A'_s и эксцесса E', также необходимо вычислить их коэффициенты по формулам 2.1.1 и 2.1.2. Эти параметры определяются по точкам выборки, полученным на шаге 4.

Используя оценки параметров асимметрии и эксцесса необходимо выполнить преобразования этих оценок по формулам 2.1.5 и 2.1.6.

Переход на шаг 6 осуществляется после того как найдены параметры всех распределений из массива $Distributions$.

$$\left|K_{A'_s}\right| = \left|\frac{A'_s}{\sqrt{\frac{6}{n}}}\right| \tag{2.1.1}$$

$$|K_{E'}| = \left|\frac{E'}{\sqrt{\frac{24}{n}}}\right| \tag{2.1.2}$$

$$\begin{cases} \left|K_{A'_s}\right| \leq 3 \\ |K_{E'}| \leq 3 \end{cases} \tag{2.1.3}$$

Шаг 6: сравнение результатов. Для этого целесообразно использовать формулу 2.1.4, т.е. вычислять расстояние между A_s и E (нормального распределения) и полученными значениями. Также по формуле Евклида вычисляется расстояние от теоретической точки с координатами b_{1_T}, b_{2_T} до всех остальных значений b_{1_H}, b_{2_H}. Т.е. для предполагаемого распределения (гипотезы).

$$d = \sqrt{\left(A_s - A'_s\right)^2 + (E - E')^2} \tag{2.1.4}$$

Если при сравнении полученные значения A_s и E сильно отличаются от табличных или теоретических (для нормального распределения $A_s = E = 0$), то расстояние d будет слишком велико и K_{A_s}, K_E выходят за границы, тогда гипотеза о том, что выборка имеет закон распределения F_1 отвергается. Следовательно у того распределения, которое наилучшим образом будет описывать закон распределения вероятности случайной величины по параметрам формы расстояние d минимально и коэффициенты находятся в рамках допустимых значений.

На рисунке 21 представлена схема алгоритма определения закона распределения.

Рисунок 21 – Схема алгоритма определения закона распределения

Для наглядного представления отобразим оценки асимметрии и эксцесса на графике, представленном на рисунке 22.

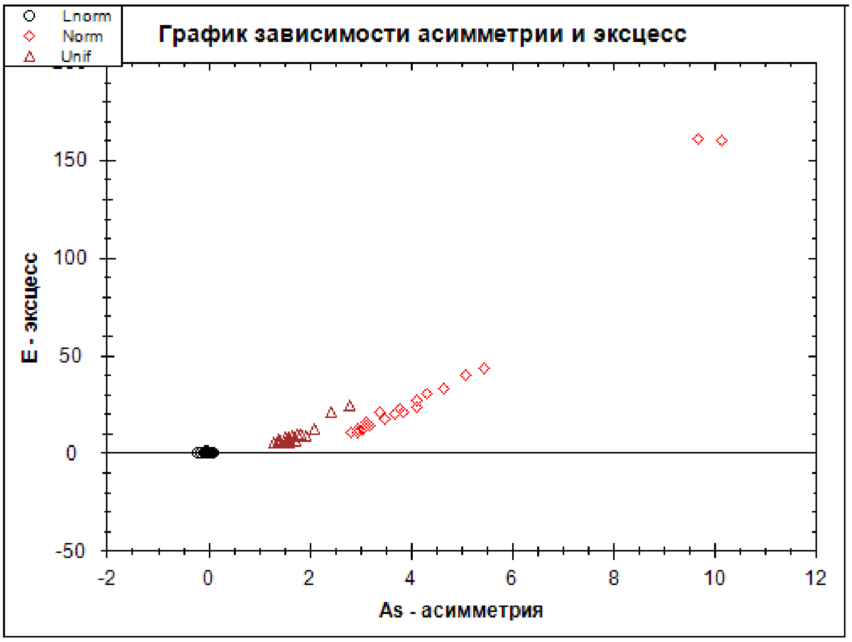


Рисунок 22 – Оценки параметров асимметрии и эксцесс

$$b_1 = {A_s'}^2 \tag{2.1.5}$$

$$b_2 = E' - {A_s'}^2 + 2 \tag{2.1.6}$$

Полученные значения изображаем в виде точек на плоскости Ob_1b_2 на рисунке 23.

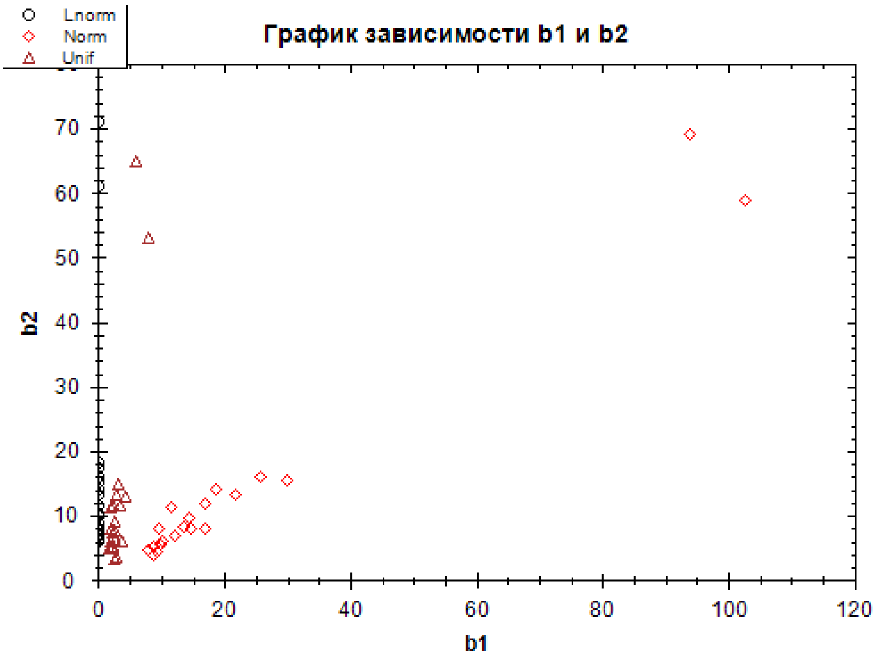


Рисунок 23 – Преобразованные оценки параметров асимметрии и эксцесс

2.2 Определение доверительного интервала

Для определения доверительного интервала необходимо определить, как данный алгоритм справляется с примером, когда исходная выборка сгенерирована по нормальному (Гауссовскому) закону распределения и выдвигается гипотеза о том, что это распределение нормально.

Пусть дана выборка случайных величин: $X_1, X_2, \ldots, X_n$ размера n; в списке теоретических распределений $Distributions$ находится только нормальное распределение; точки b_{1_T}, b_{2_T} для теоретического нормального распределения, которые соответственно равны 0 и 2.

Шаг 1: генерируем случайные величины по нормальному закону распределения со стандартными параметрами математического ожидания и дисперсии.

Шаг 2: выдвигаем гипотезу h_1 о законе распределения вероятности случайной величины. Т.е. предполагаем, что исследуемые случайные величины (сгенерированная выборка) распределены по нормальному закону.

Шаг 3: «прогоняем» сгенерированную выборку по алгоритму идентификации закона распределения, описанного в данной работе.

Шаг 4: определяем значения b_{1_H}, b_{2_H} для преобразованной выборки по оценкам асимметрии и эксцесса. Данные значения представляем в виде точек на плоскости Ob_1b_2.

Шаг 5: определяем расстояния от теоретической точки $[b_{1_T};\ b_{2_T}]$ до практических, выбираем наибольшее из них - d и определяем доверительный интервал в 95%. Определение расстояния для точек представлено на рисунке 24.

Этот интервал используем при отборе значений b_{1_H}, b_{2_H} всех рассматриваемых в работе распределений, отбраковывая все то, что не входит в 95%.

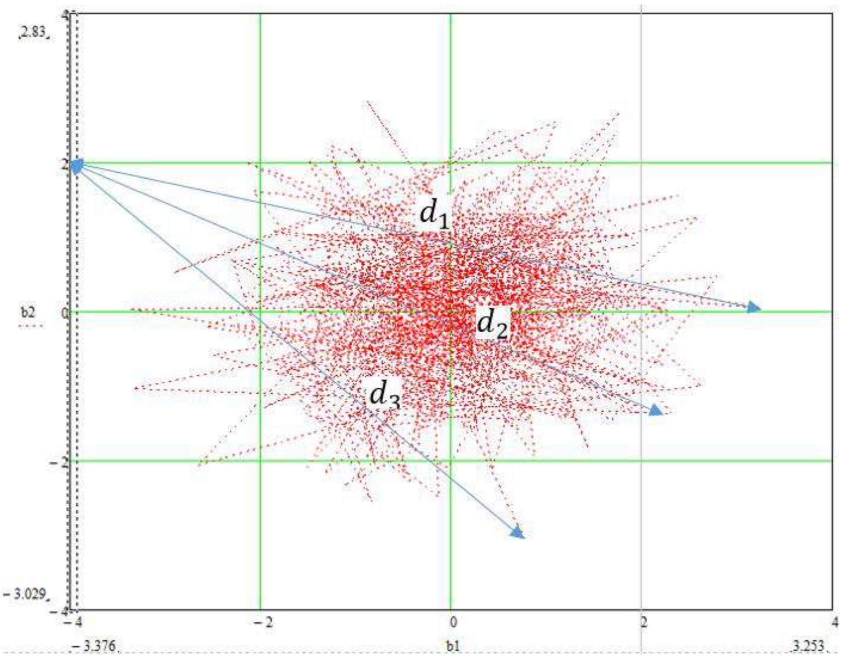


Рисунок 24 – Определение расстояний для точек $[b_1 b_2]$

2.3 Отбраковка аномальных измерений

Для повышения робастности необходимо очистить выборку от аномальных измерений (выбросов). Если этого не сделать, то топологический метод будет давать плохой результат. Для этих целей применяется множество различных способов, но в данной работе использован критерий типа Грабса, для отбраковки одновременно трех минимальных или трех максимальных значений в выборке.

Пусть $X_1, X_2 \dots X_n$ – наблюдаемая выборка, а $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ – построенный по ней вариационный ряд. Метод отсеивания аномальный измерений заключается в проверки того, что точки ряда $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n-1)}$ принадлежат одному закону распределения, а $X_{(n)}$ некоторому другому, который сдвинут вправо. Статистика критерия Грабса имеет вид:

$$G_n = (X_{(n)} - \bar{X})/S, \tag{2.3.1}$$

где:

$$\bar{X} = \frac{1}{n}\sum_{j=1}^{n} X_j, \tag{2.3.2}$$

$$S^2 = \frac{1}{n-1}\sum_{j=1}^{n} (X_j - \bar{X})^2, \tag{2.3.3}$$

$$S = \sqrt{S^2} \tag{2.3.4}$$

Элемент выборки максимальный или минимальный будет считаться выбросом, если значение статистики превышает критическое:

$$G_n \geq G_{n,1-\alpha} \tag{2.3.5}$$

$$G_1 \geq G_{1,1-\alpha} \tag{2.3.6}$$

2.4 Работа программы

При запуске программы открывается окно, изображенное на рисунке 25, состоящее из области управления, где можно генерировать выборку и области с результатом работы программы: график, таблица с количеством принятых решений в пользу того или иного закона распределения.

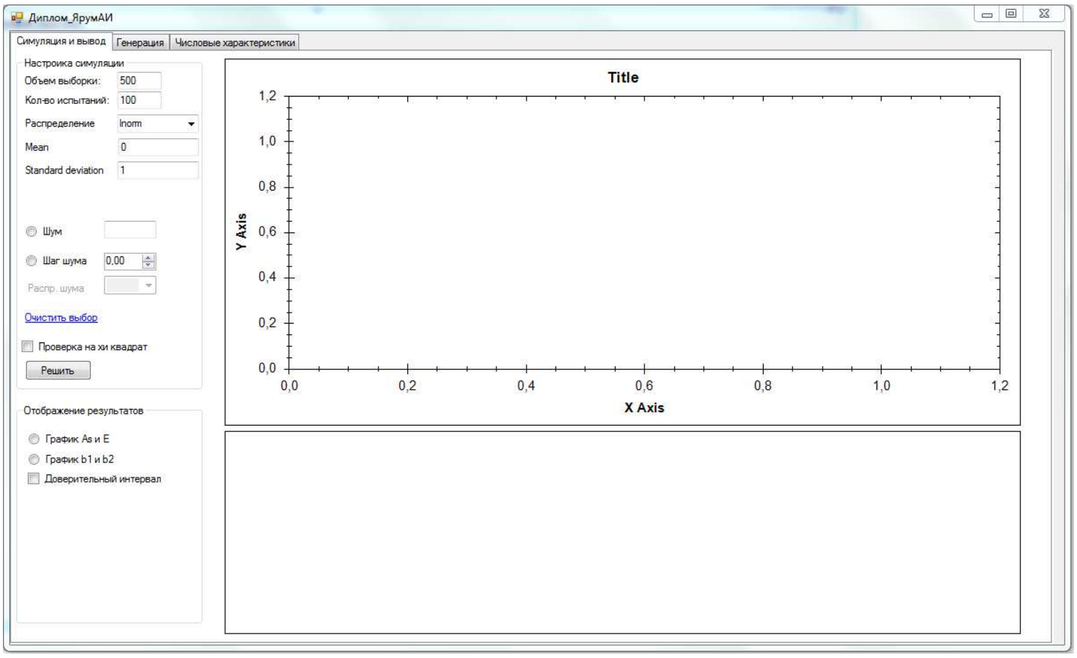


Рисунок 25 – Начало работы программы

На данной вкладке у пользователя есть возможность сгенерировать выборку, распределенную по одному из выбранных законов распределения в выпадающем списке – «Распределение». Также исследователь может выбрать объем данной выборки, количество экспериментов – т.е. сколько раз необходимо будет генерировать выборку по одним и тем же параметрам (математическое ожидание, дисперсия, среднеквадратичное отклонение и т.д.) и параметры конкретного распределения.

Для проверки работы алгоритма в программе присутствует возможность добавлять шум в выборку, что приближает ее к реальным данным. Шум также генерируется по выбранному закону распределения, т.е. мы можем «зашумлять» выборку, распределенную по одному закону значениями, распределенными по-другому. Причем можно задавать шаг шума, например, с шагом в 10% наша выборка с каждым прогоном будет «зашумляться» на 10% больше и так до 100%.

На вкладке генерация, изображенной на рисунке 26, есть возможность загрузить выборку из файла и увидеть ее гистограмму, справа от панели получения данных. Также есть возможность сгенерировать выборку значений случайной величины, используя возможности статистического языка программирования R. При загрузке из файла следует учитывать, что файл должен состоять из вещественных чисел, выраженных в десятичной форме (например, 13.593, -0.789, 67 и т.д.).

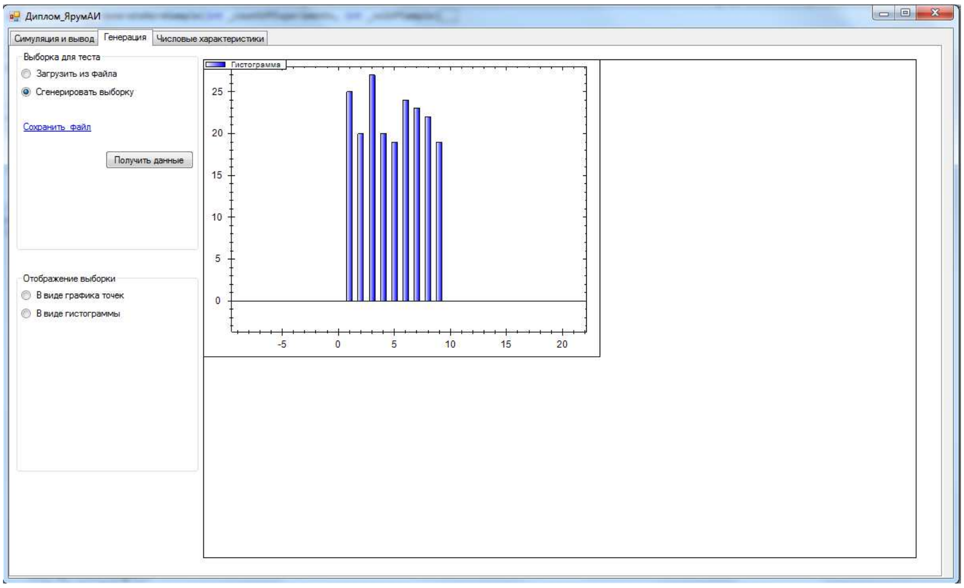


Рисунок 26 – Гистограмма выборки

Выбрав, какой - либо пункт на панели, необходимо кликнуть по кнопке «Получить данные». После чего произойдет автоматический переход на другую вкладку, изображенную на рисунке 27, где осуществляется генерация и идентификация закона распределения вероятности случайной величины. Если выборка загружена из внешнего источника, то программа на вкладке «Симуляция и вывод» сразу выдает результат. Результат работы программы представлен на рисунке 28.

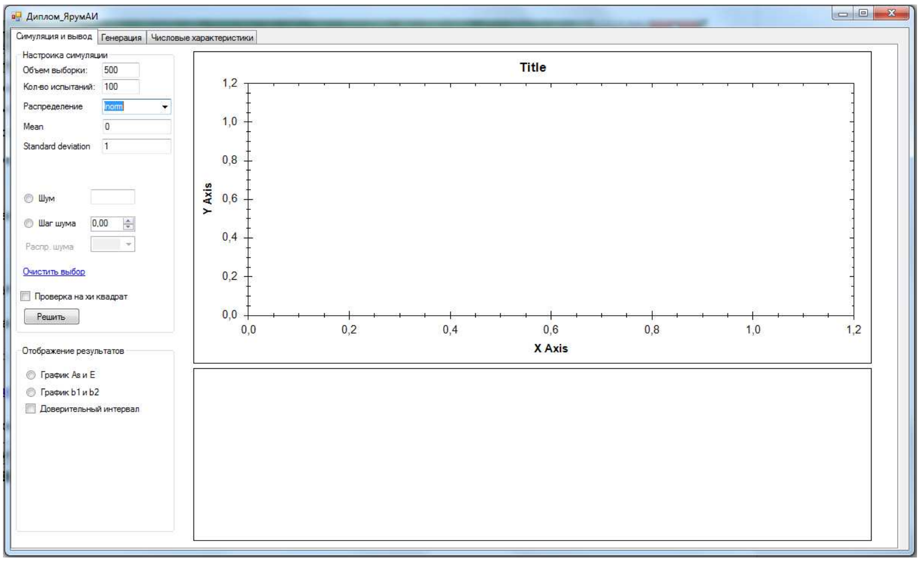


Рисунок 27 – Симуляция и вывод

При генерации имеется возможность указания размера выборки, закона распределения, из которого эта выборка будет взята, параметров этого распределения, а также процент зашумления.

Помеха может распределяться по нормальному, логнормальному и равномерному закону со стандартными параметрами (например, для нормального закона математическое ожидание равно 0, а дисперсия 1). Если выбрать пункт «Шаг шума», то выборка будет генерироваться несколько раз при разных значениях помехи, например, шаг шума 50%, следовательно, выборка будет сгенерирована три раза при нулевом, 50 и 100 процентном зашумлении.

Также в программе реализована возможность сохранить полученную выборку значений случайной величины во внешний файл, для дальнейшего анализа и использования.

По клику на кнопку «Решить», находящейся на панели «Настройка симуляции» на экране справа появляется таблица с результатами работы программы, а также график, построенный по точкам с координатами асимметрии и эксцесса для всех распределений из выпадающего списка.

В таблице, изображенной на рисунке 28, цифрами указывается, принят или отвергнут предполагаемый закон распределения вероятности.

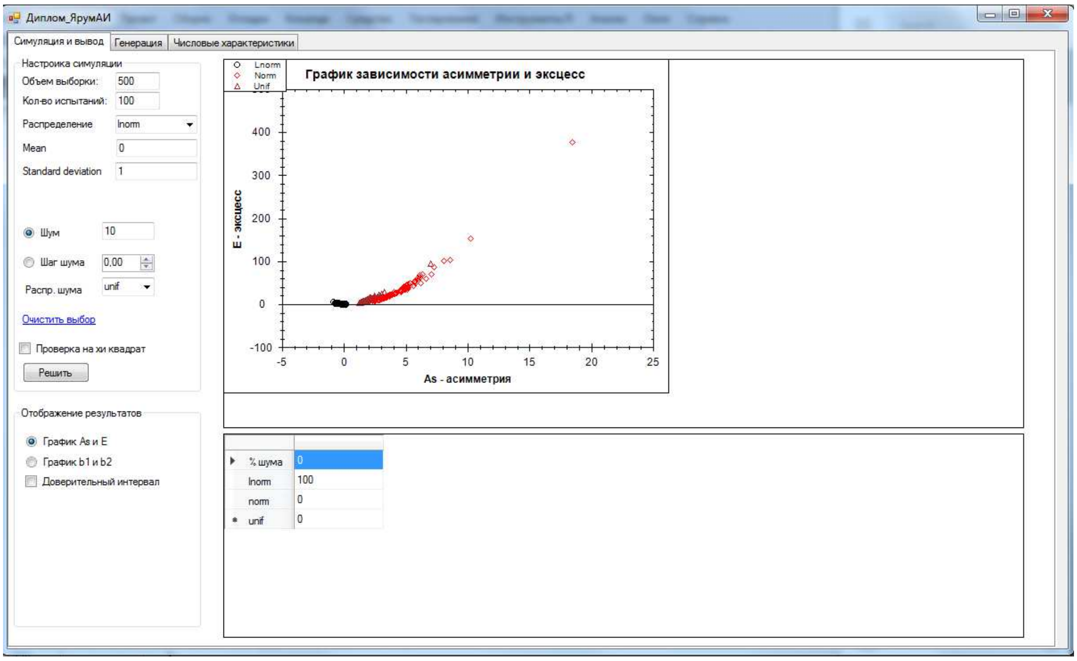


Рисунок 28 – Результат работы программы

У пользователя также существует возможность настраивать отображение результатов. Т.е. он может посмотреть график асимметрии и эксцесса, а также график, построенный на значениях b_1, b_2, представленный на рисунке 29, существующих распределений.

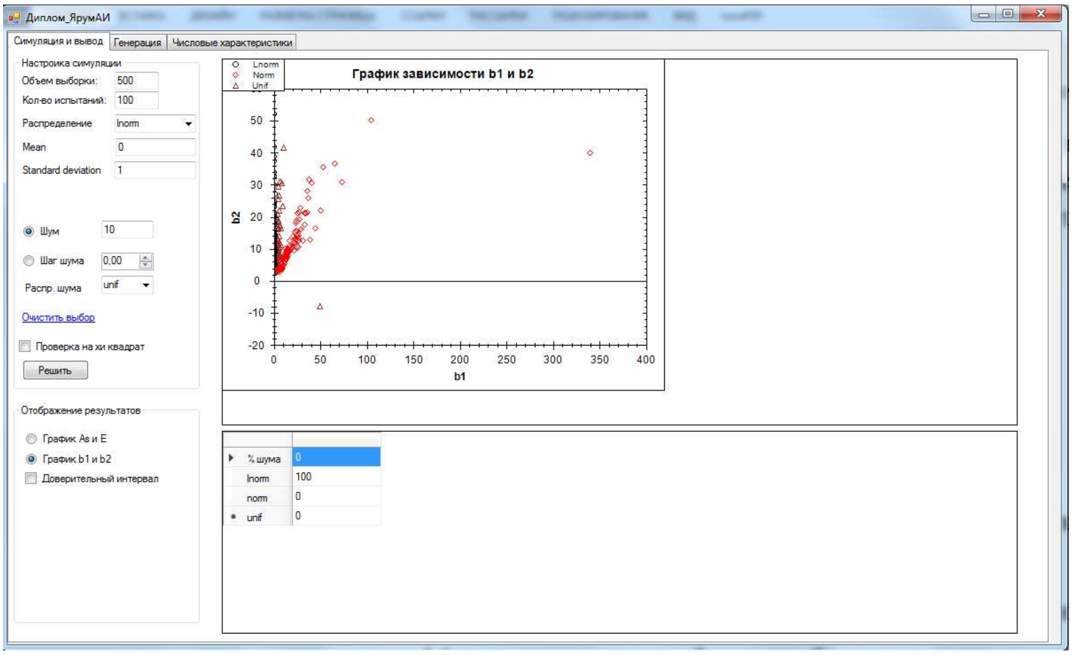


Рисунок 29 – График b_1, b_2

Также на этой панели присутствует элемент управления «CheckBox» - «Доверительный интервал», позволяющий отбросить все точки, которые не входят в интервал доверия, описанный в пункте 2.2.

3 Апробация алгоритма топологической идентификации

3.1 Определение минимального объема выборки

Для проверки разработанного алгоритма на практике и оценки эффективности его работы необходимо провести идентификацию выборок разного объема и сгенерированных по разным законам распределения.

Прежде всего, следует определить минимальный объем выборки, при котором метод будет работать. Логнормальное распределение объемом 30 представлено на рисунке 30. В теоретических источниках даются различные рекомендации по объемы выборки, для определения закона распределения по форме полигона. Они разбрасываются от 30 до 200. Поэтому генерируем нашу выборку 15 раз с начальным объемом равным 30 измерений. На рисунках 30 и 31 представлен результат работы алгоритма для логнормального распределения с объемом выборки 30.

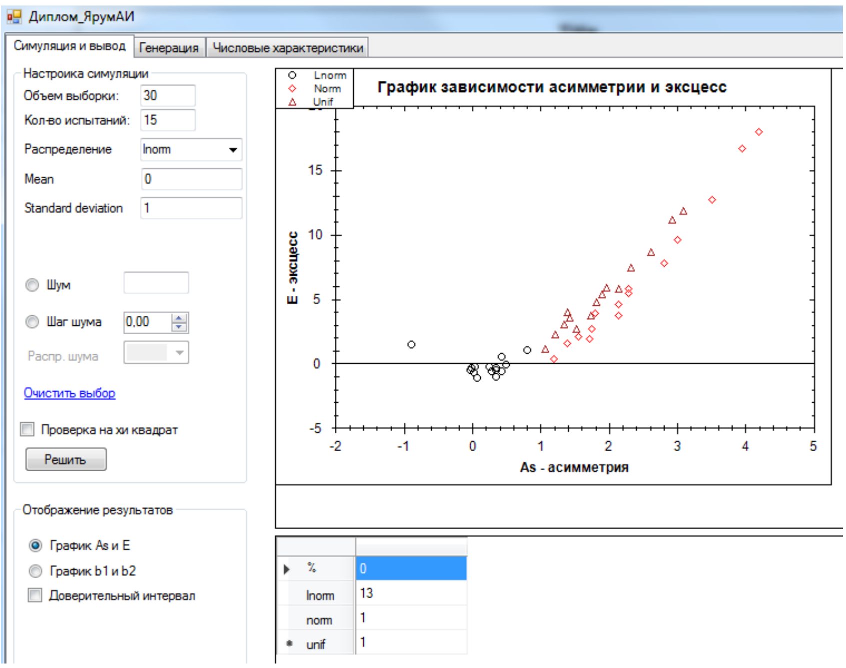


Рисунок 30 – Логнормальное распределение объемом 30

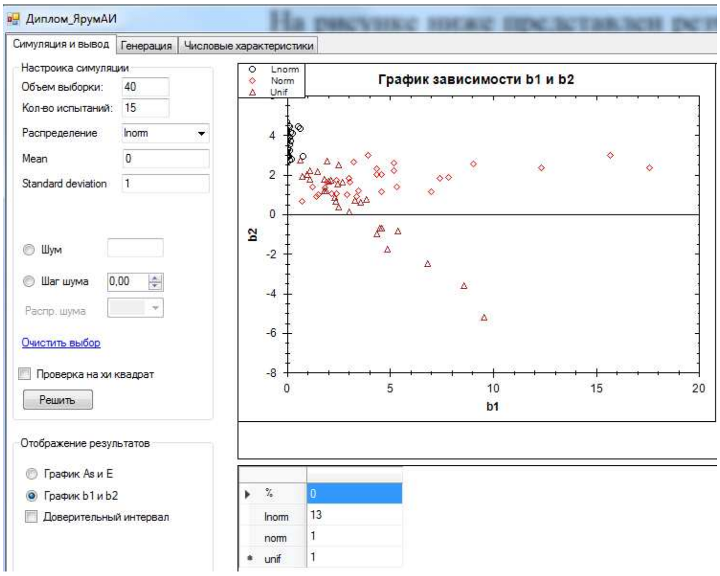


Рисунок 31 – Логнормальное распределение объемом 30 b_1, b_2

Далее увеличим объем выборки на 10 измерений и повторим эксперимент. Результат работы указан на рисунках 32 и 33.

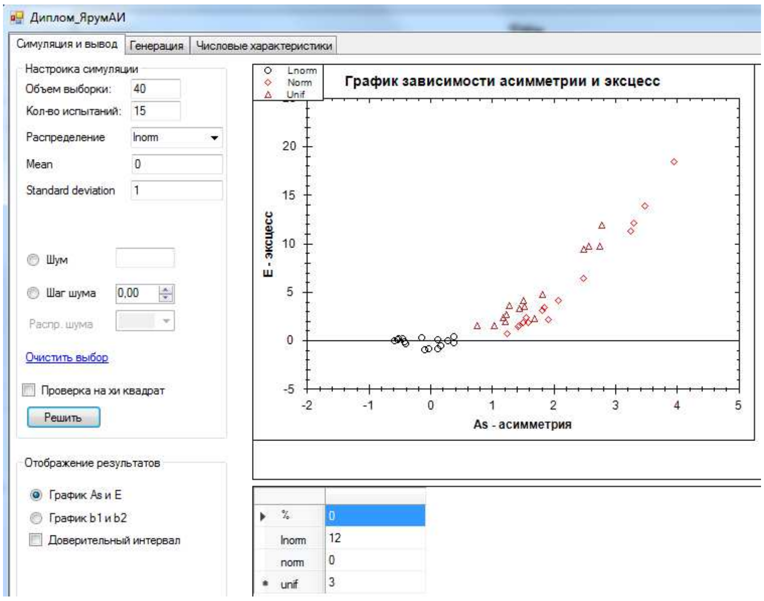


Рисунок 32 – Логнормальное распределение объемом 40

Для логнормальной выборки при 40 измерениях, программа три раза приняла неверное решение, совершив ошибку первого рода.

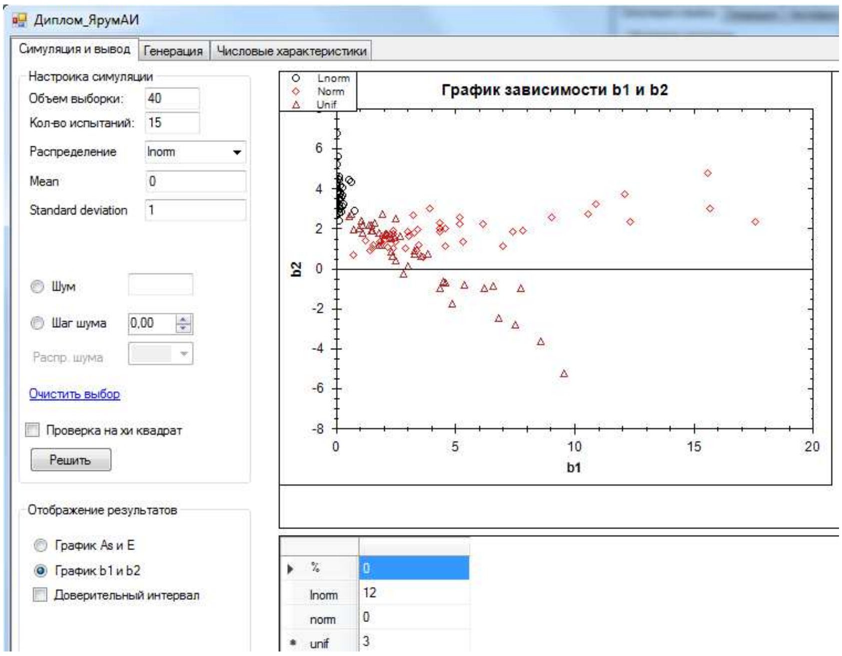


Рисунок 33 – Логнормальное распределение объемом 40 b_1, b_2

При дальнейшем увеличении объема выборки кол-во неверно принятых решений уменьшается, результаты представлены в таблице 1.

Таблица 1 – Результат работы программы. Логнормальное распределение

Объем выборки	Количество неверно принятых решений
30	2
40	3
50	1
60	0

Далее проведем несколько экспериментов с нормальным и равномерным законами распределения. Данные о нормальном законе распределения занесем в таблицу 2, а информацию о равномерном законе распределения вероятности занесем в таблице 3.

Таблица 2 – Результат работы программы. Нормальное распределение

Объем выборки	Количество неверно принятых решений
30	3
40	2
50	2
60	2
100	1

Таблица 3 – Результат работы программы. Равномерное распределение

Объем выборки	Количество неверно принятых решений
30	2
40	1
50	1
60	1
70	0

Проведя несколько экспериментов можно смело утверждать, что минимальный объем выборки равен 60 - 70 значений. Так как при объеме 50 иногда принимается неправильное решение. Полученные данные изображаем на рисунке 34.

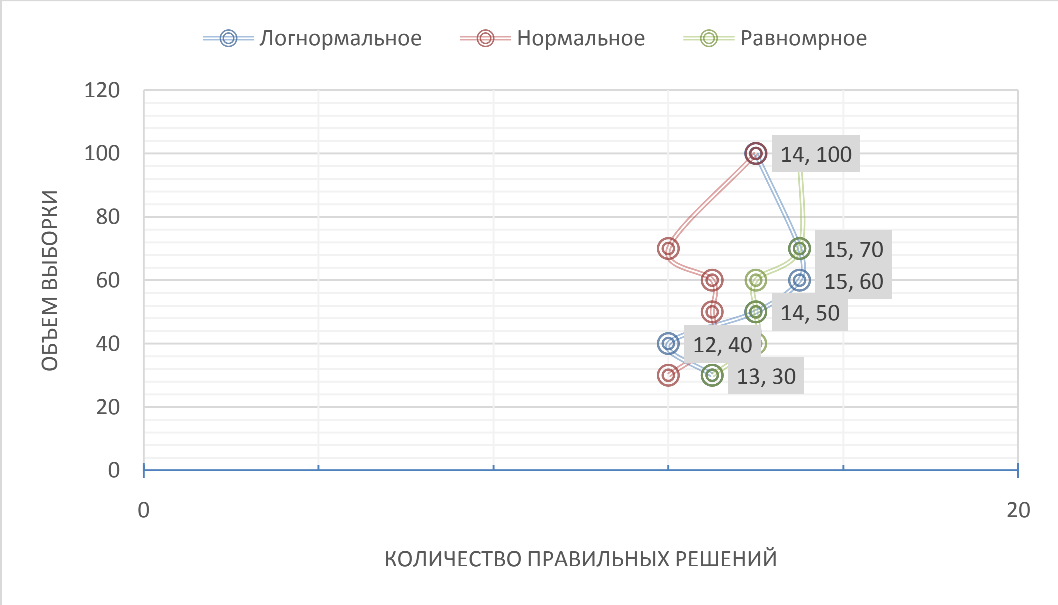


Рисунок 34 – Определение объема выборки

3.2 Логнормальное распределение случайной величины

3.2.1 Зашумление нормально распределенной случайной величиной

Опробуем метод идентификации по параметрам полигона на случайных выборках. Значения, которых, распределены по логнормальному закону. Генерируем 20 выборок, размер которых равен 500. С аддитивной помехой (0, 50 и 100 процентов) значения которой, распределены по нормальному закону, представленному на рисунке 35. Т.е. зашумляем логнормальную выборку нормально распределенной.

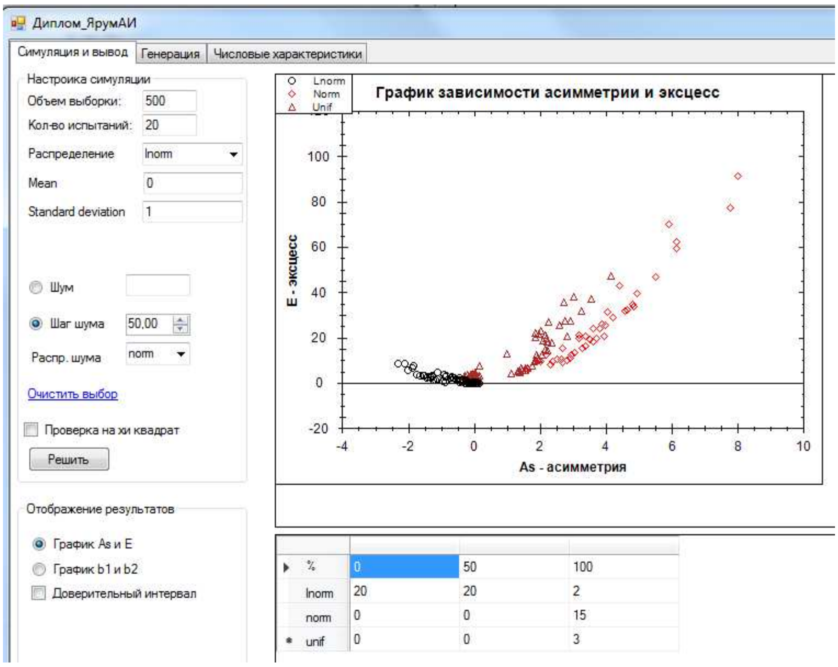


Рисунок 35 – Результат работы программы при Логнормальном распределении

При 50 процентном шуме метод идентификации все еще работает, а уже при 100 процентах программа принимает решение в пользу нормального закона распределения (ошибка первого рода). Поэтому имеет смысл рассмотреть, как поведет себя алгоритм при зашумлении от 50% до 100%.

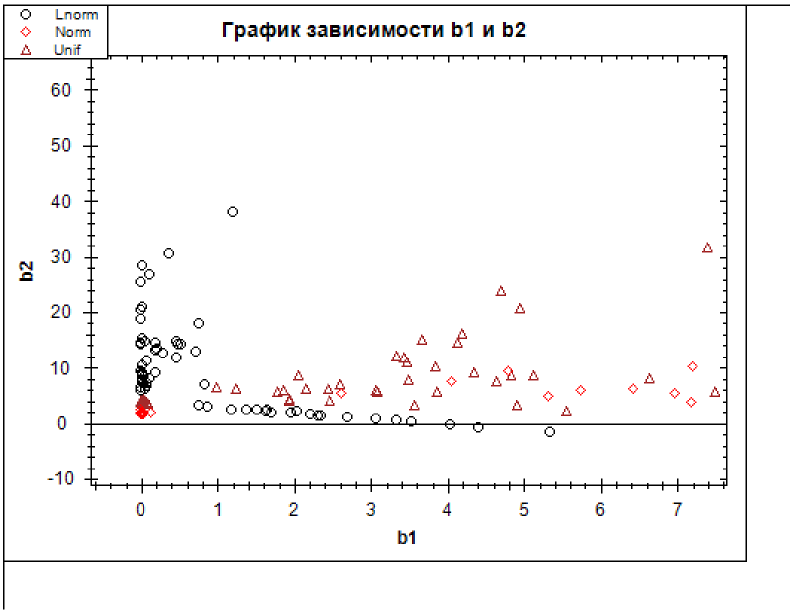
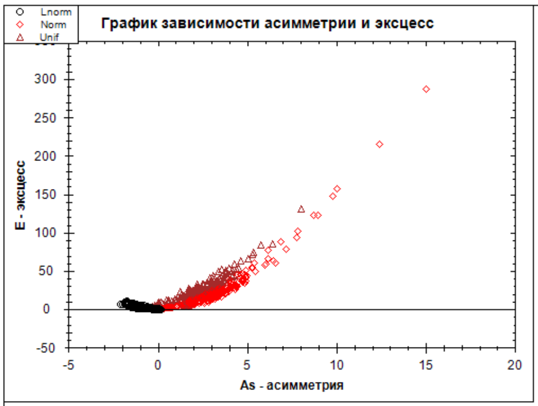


Рисунок 36 – График b_1, b_2 для Логнормального распределения

При исследовании графика зависимости b_1, b_2, так же получены результаты, представленные на рисунке 36, подтверждающие проверку статистических гипотез по асимметрии и эксцессу. Так как, теоретические (табличные) значения b_1, b_2 соответственно равны 0 и 2. А в данном диапазоне количество точек b_1, b_2 логнормального распределения больше всех остальных (расстояние измерялось по формуле Евклида).

Проверим, как изменяется результат работы, если помеха будет иметь значения от 50 до 100 процентов с шагом в 10%. Результат при Логнормальном распределении с шумом от 50 до 100% представлен на рисунке 37.



% шума	0	10	50	60	70	80	90	100
lnorm	20	20	20	20	18	18	13	1
norm	0	0	0	0	2	2	7	17
unif	0	0	0	0	0	0	0	2

Рисунок 37 – Результат при Логнормальном распределении с шумом от 50 до 100%

Можно заметить, что программа начинает выдавать неправильное решение при 70% зашумлении.

На рисунке 38 представлен график b_1, b_2 при зашумлении от 50% до 100%.

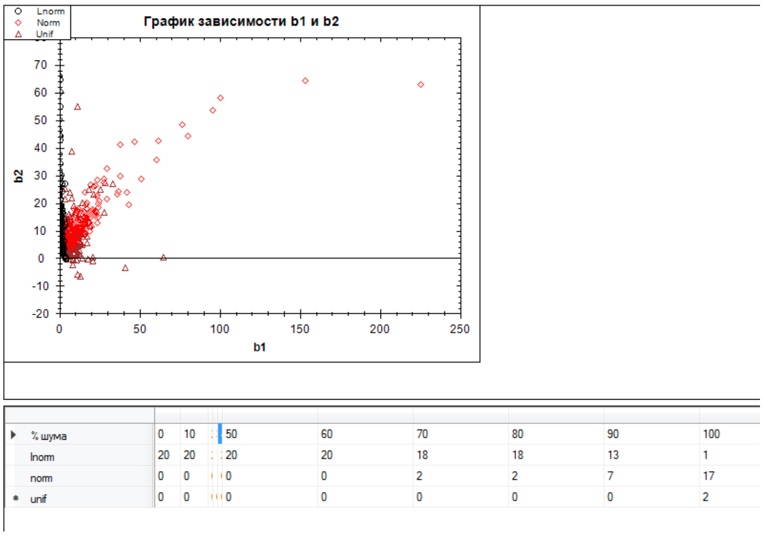


Рисунок 38 –Логнормальное распределении с шумом от 50 до 100%, b_1, b_2

На основании полученных результатов построим зависимость получения правильных ответов от процента шума, представленную на рисунке 39. Переступая 80% порог аддитивной помехи количество правильно принятых решений резко уменьшается, практически до 0.

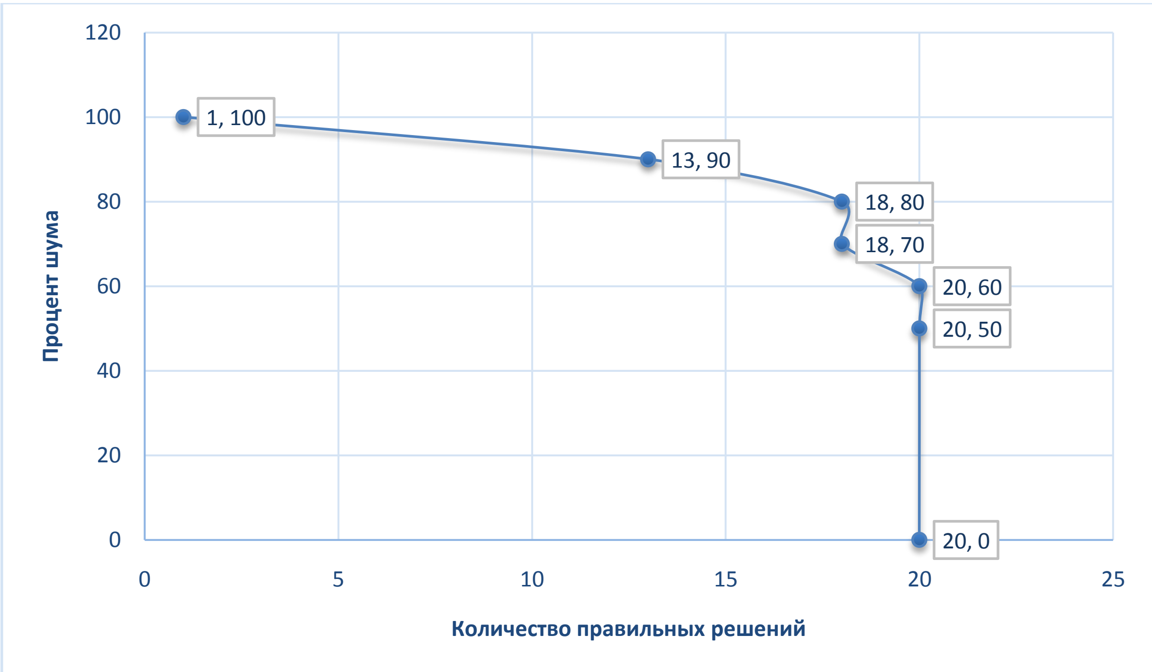


Рисунок 39 – Зависимость при Логнормальном распределении с нормальной помехой

3.2.2 Зашумление равномерно распределенной случайной величиной

Теперь рассмотрим, как поведет себя алгоритм при равномерном зашумлении. Параметры выборки точно такие же, как и в предыдущем пункте. Логнормальное распределение изображено на рисунке 40.

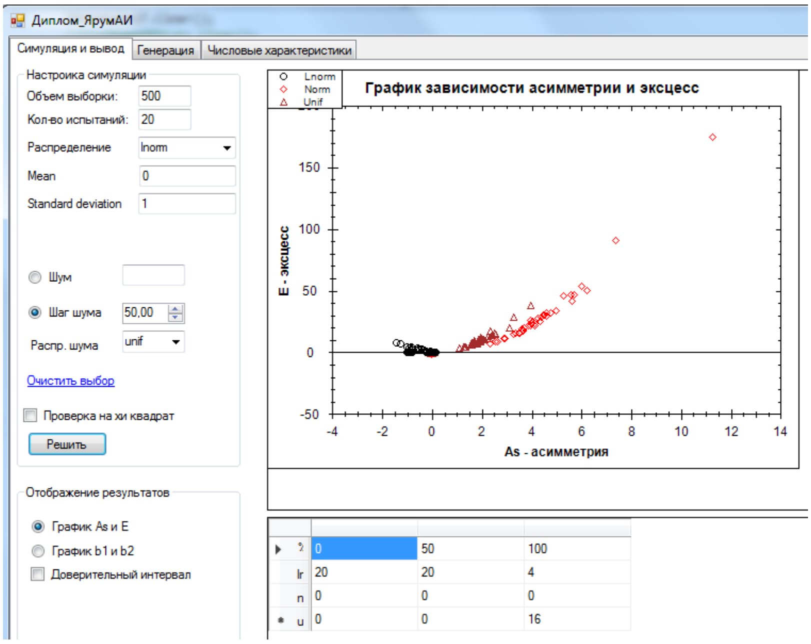


Рисунок 40 – Логнормальное распределение с шумом от 0 до 100% по равномерному

Программа выдает неверные результаты при аддитивной помехе от 50%. Отсюда следует, что необходимо рассмотреть, как поведет себя алгоритм при зашумлении от 50 до 100% с приращение помехи на 10%. Результаты работы алгоритма с аддитивной помехой изображены на рисунке 41.

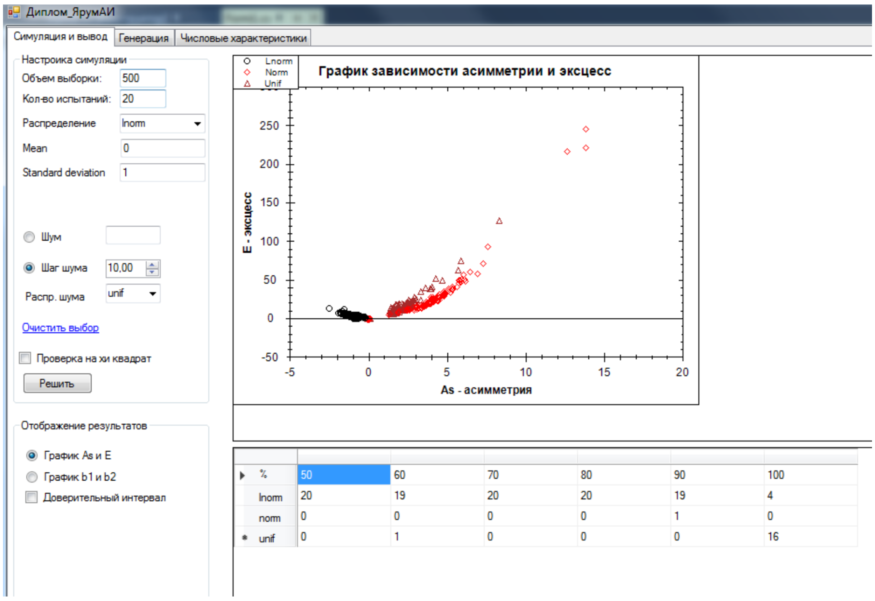


Рисунок 41 – Логнормальное распределение с шумом от 50 до 100% по равномерному

На рисунке 42 соискатель может увидеть изображение логнормального распределения точек с координатами b_1, b_2 с шумом от 50 до 100 % с 10 % приращением аддитивной помехи.

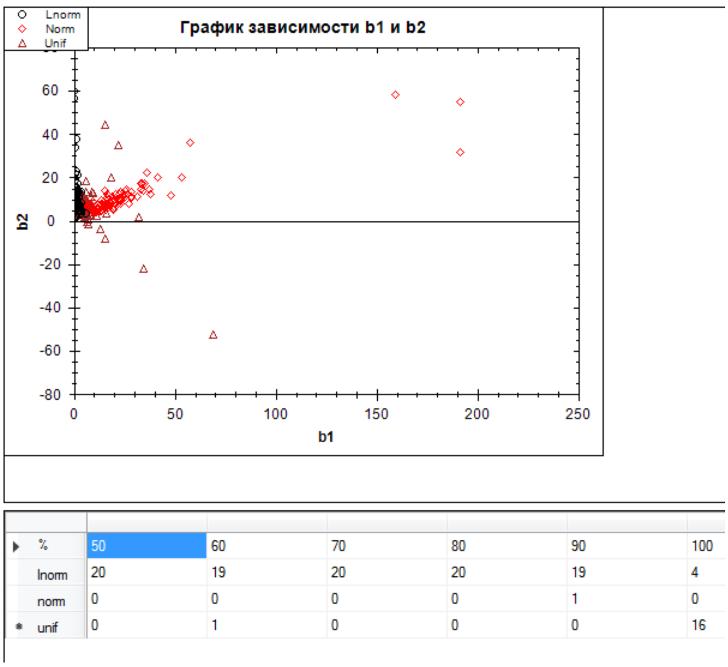


Рисунок 42 –Логнормальное распределение с шумом от 50 до 100%, b_1, b_2 по равномерному

На основании полученных данных построим график зависимости процента помехи от количества правильно принятых решений, представленный на рисунке 43.

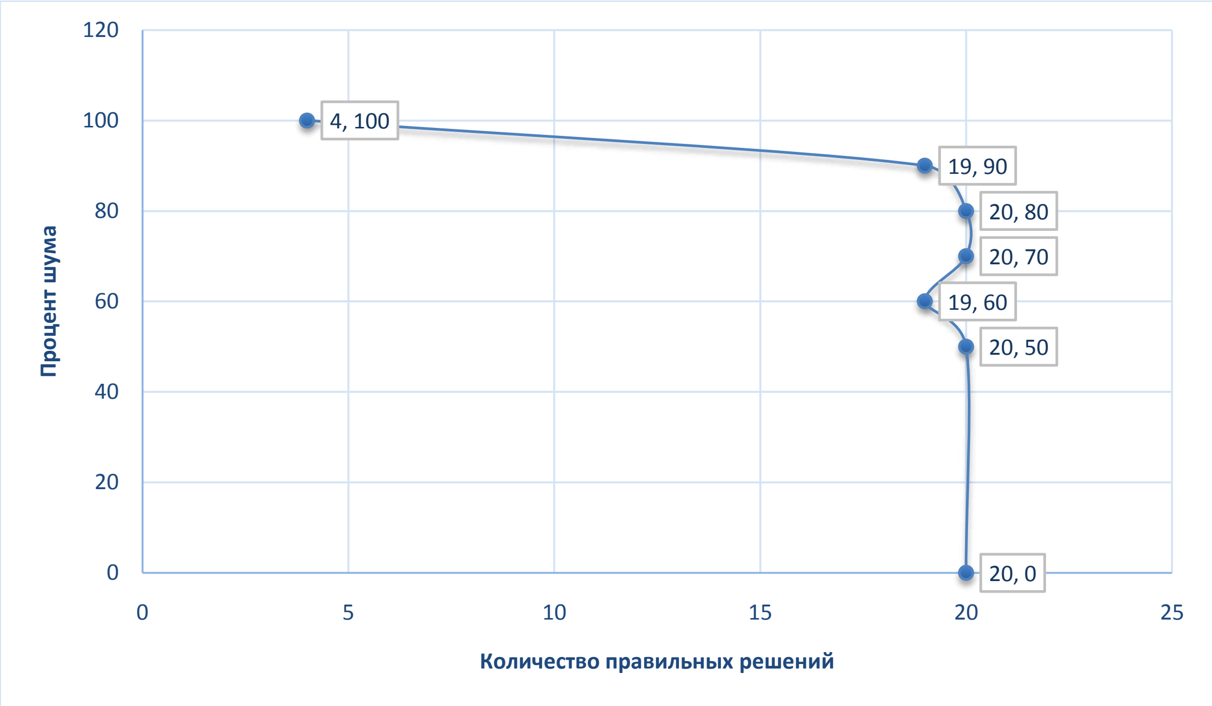


Рисунок 43 – Зависимость при Логнормальном распределении с равномерной помехой

3.3 Нормальное распределение случайной величины

3.3.1 Зашумление равномерно распределенной случайной величиной

Проверим топологический метод идентификации закона распределения случайной величины, распределенной по нормальному закону. Генерируем 20 выборок, размер которых равен 500 измерений. С аддитивной помехой (0, 50 и 100 процентов) значения которой, распределены по равномерному закону. Т.е.

зашумляем нормальную выборку равномерно распределенной. Результат представлен на рисунке 44.

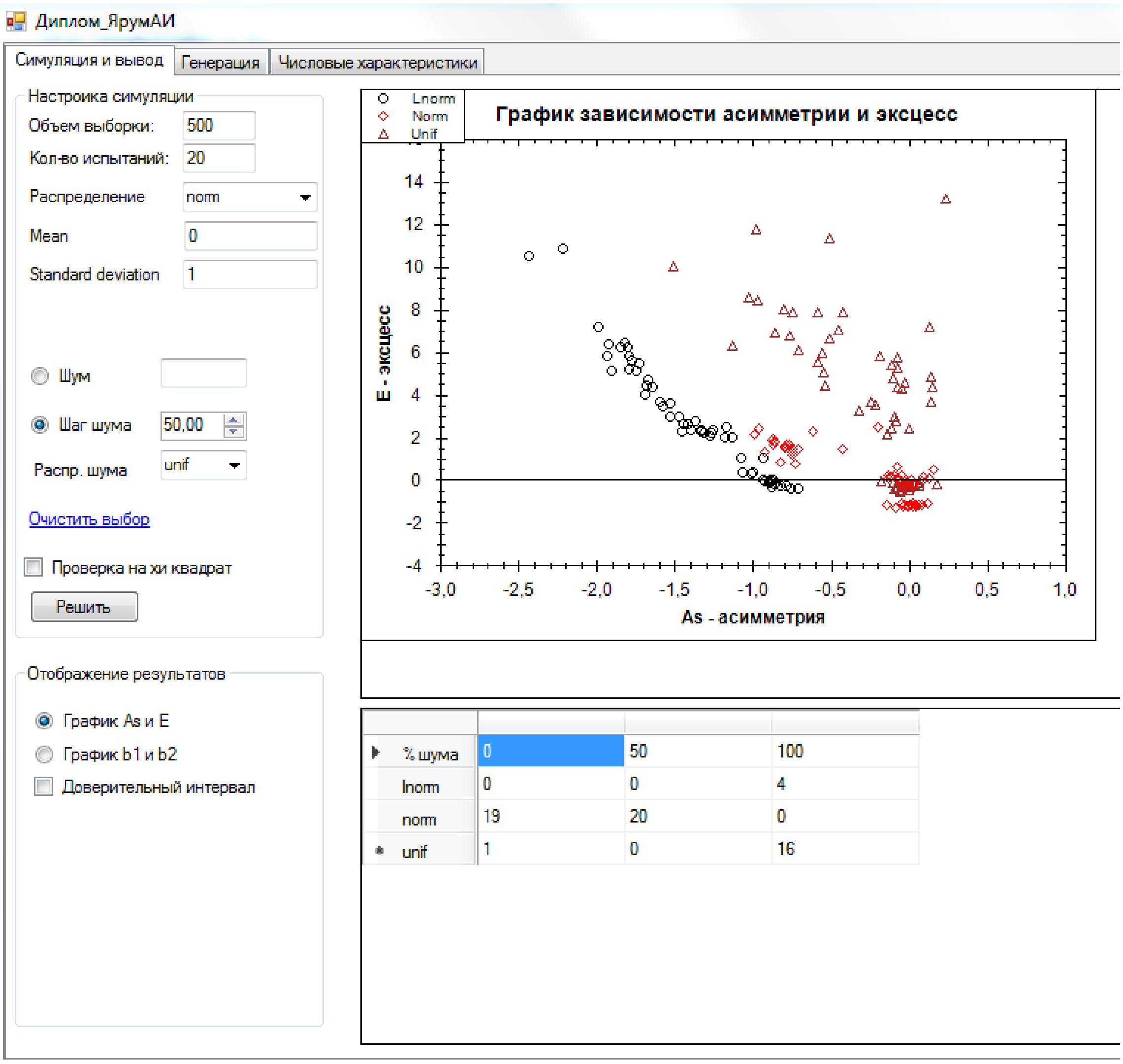


Рисунок 44 –Нормальное распределение с шумом от 0 до 100% по равномерному

На рисунке 45 изображен график с координатами точек b_1, b_2, с шумом от 0 до 100 %. С помехой через каждые 50 %.

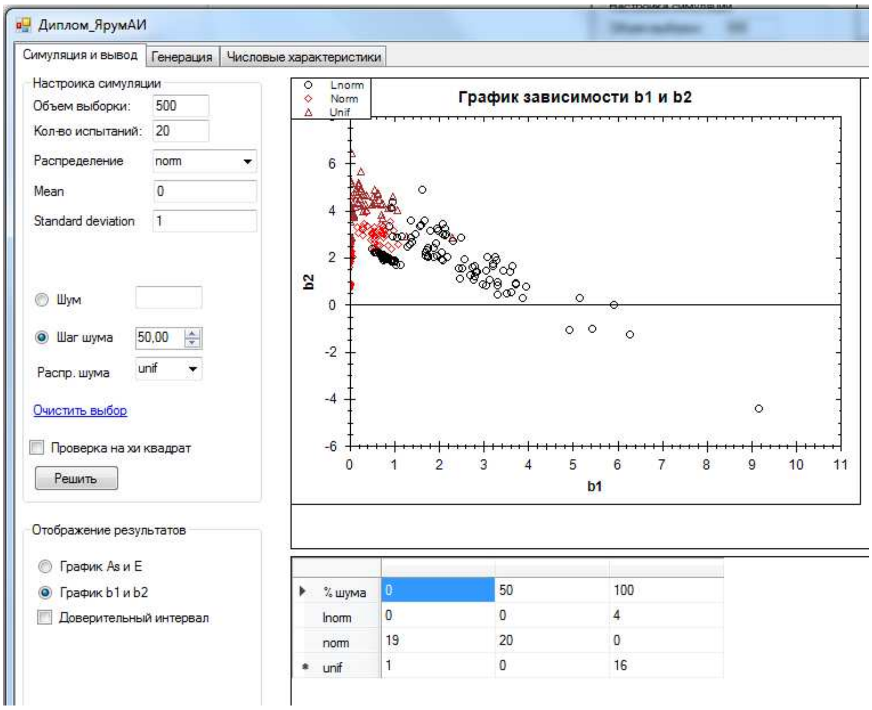


Рисунок 45 –Нормальное распределение с шумом от 0 до 100%, b_1, b_2 по равномерному

Программа выдает неверные результаты при аддитивной помехе от 50% соответственно необходимо рассмотреть, как поведет себя алгоритм при зашумлении от 50 до 100% с приращение помехи на 10%, результаты данного эксперимента представлены на рисунке 46.

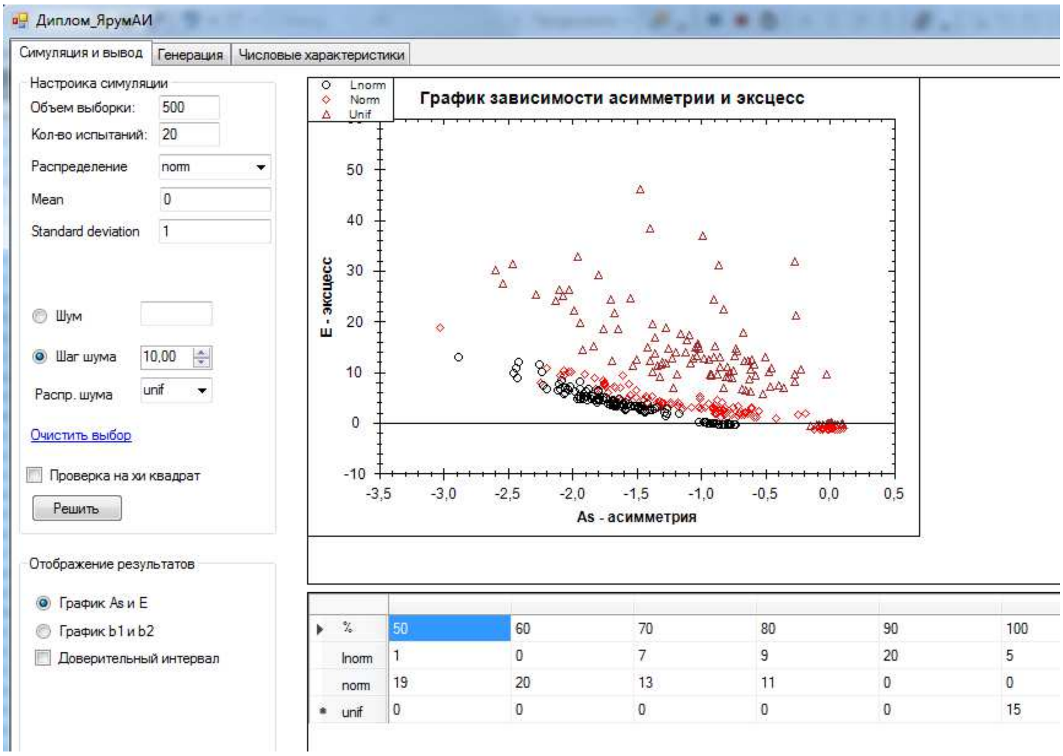


Рисунок 46 – Нормальное распределение с шумом от 50 до 100% по равномерному

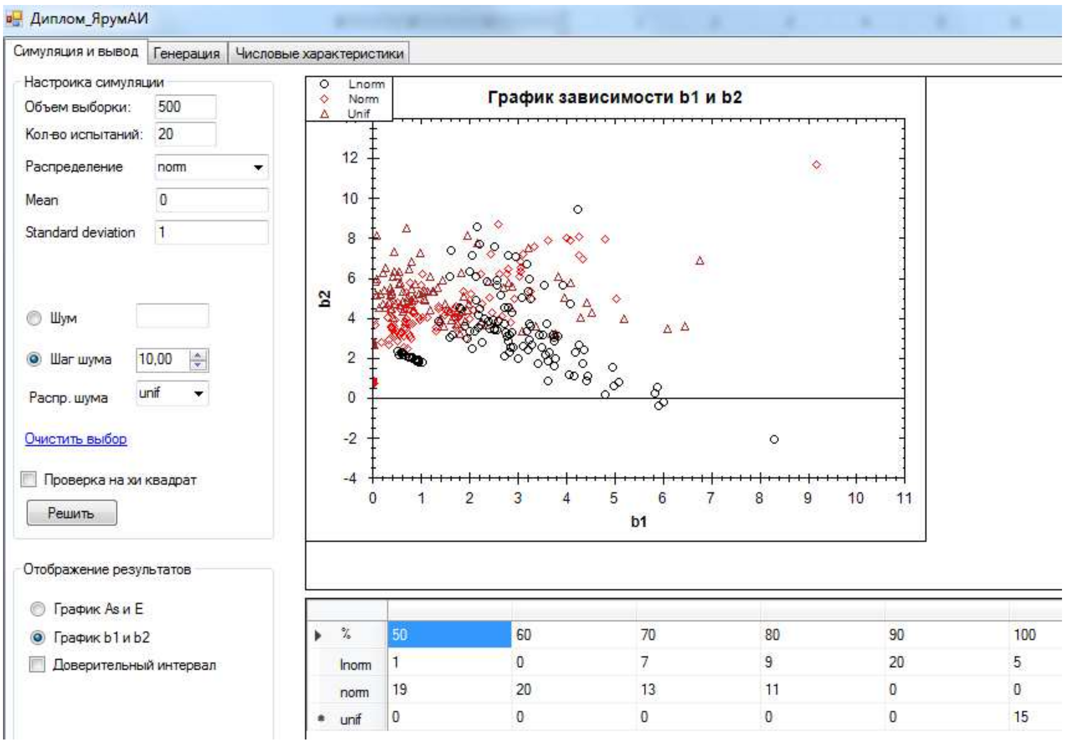


Рисунок 47 –Нормальное распределение с шумом от 50 до 100%, b_1, b_2 по равномерному

Построим зависимость получения правильных ответов от процента шума, которая представлена на рисунке 48.

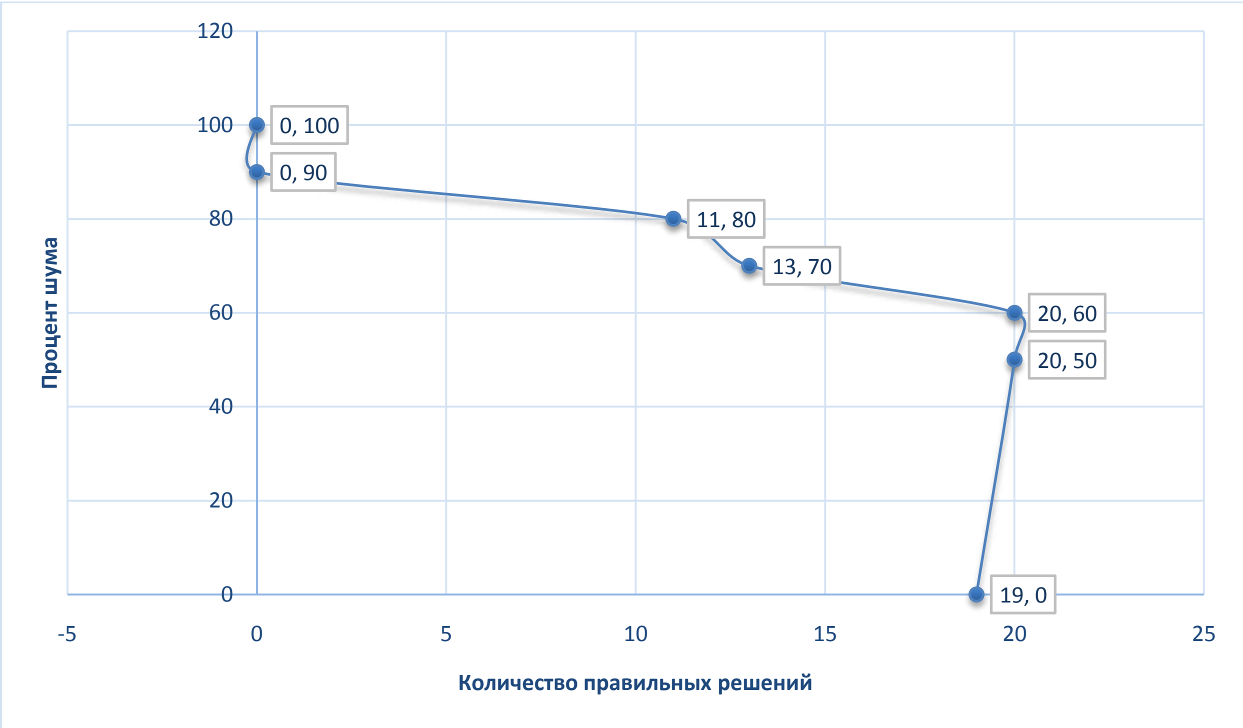


Рисунок 48 – Зависимость при нормальном распределении

3.4 Равномерное распределение случайной величины

3.4.1 Зашумление нормально распределенной случайной величиной

Проведем исследования алгоритма идентификации по полигону на случайных выборках значения, которых, распределены по равномерному закону. Объем выборки также равен 500, а помеха имеет нормальный закон распределения, проводим 20 испытаний с 0, 50 и 100% шумом. Результат которых, представлен на рисунке 49.

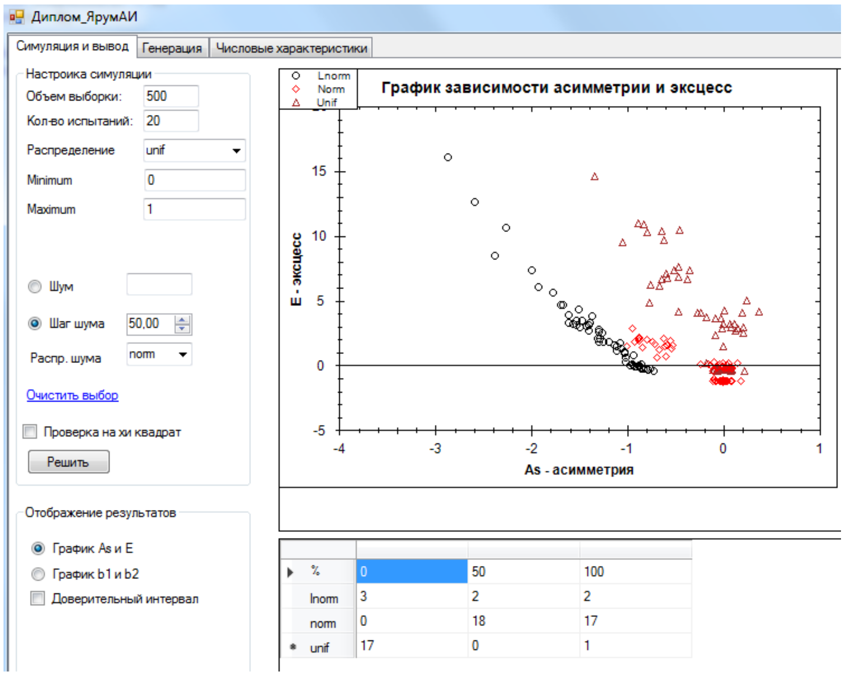


Рисунок 49 –Равномерное распределение с шумом от 0 до 100% по нормальному

На рисунке 50 представлен график построенный по точкам с координатами b_1, b_2.

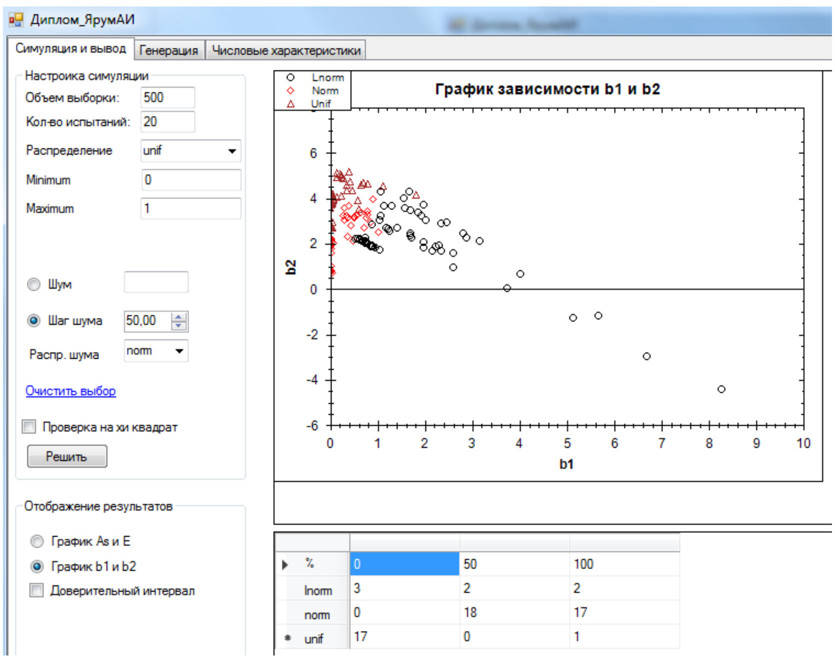


Рисунок 50 – Равномерное распределение с шумом от 0 до 100%, b_1, b_2 по нормальному

Посмотрим, как изменяется результат работы, если помеха будет иметь значения от 50 до 100 процентов, результаты представлены на рисунках 51 и 52.

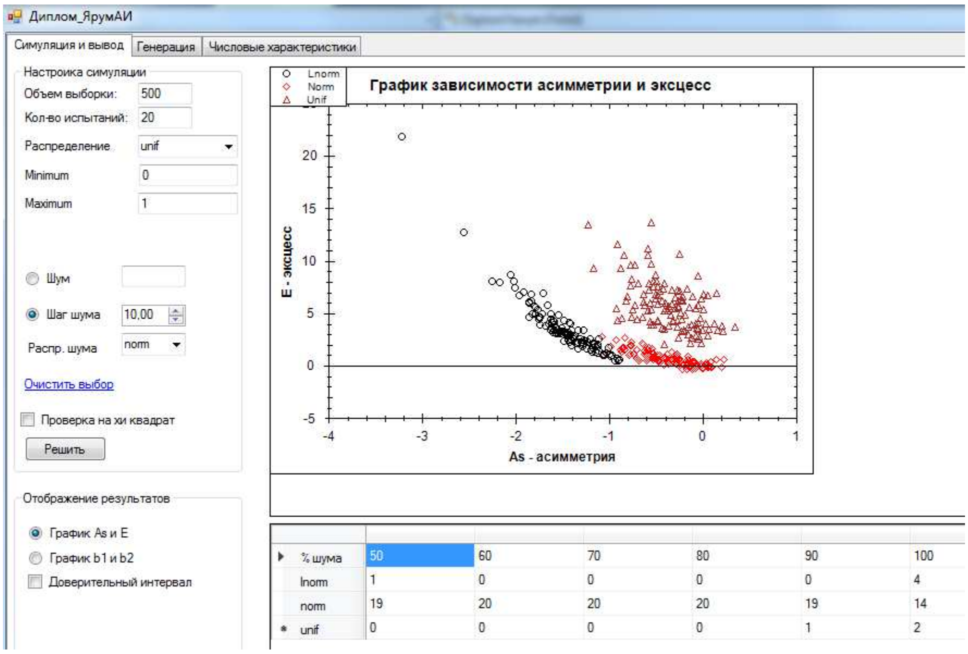


Рисунок 51 – Равномерное распределение с шумом от 50 до 100% по нормальному

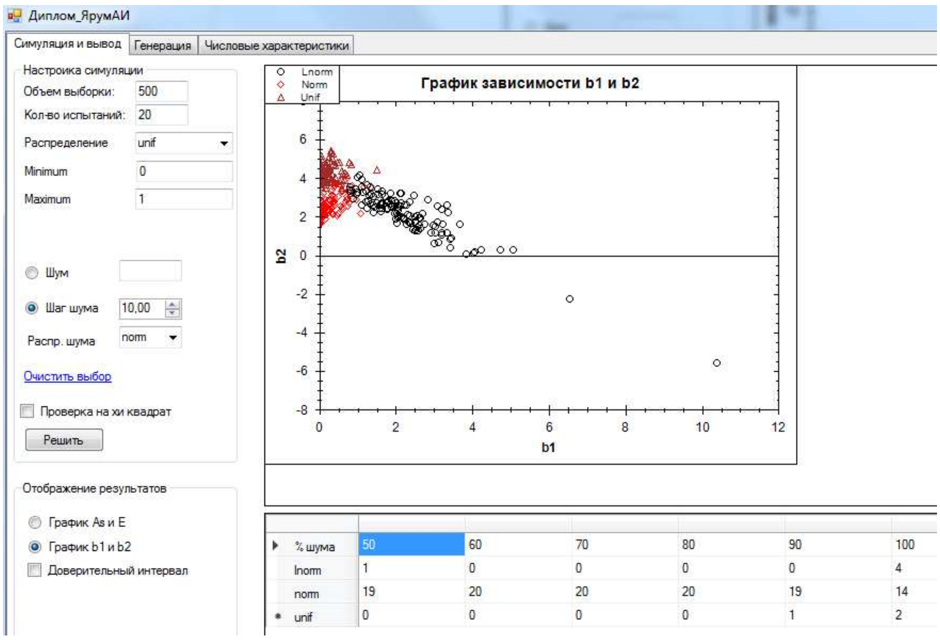


Рисунок 52 – Равномерное распределение с шумом от 50 до 100%, b_1, b_2 по нормальному

Исходя из данных необходимо просмотреть как ведет себя алгоритм при зашумлении от 0 до 50% так, как свыше 50% принятие правильного решения носит скорее случайный характер, а не закономерность, результаты на рисунках 53 и 54.

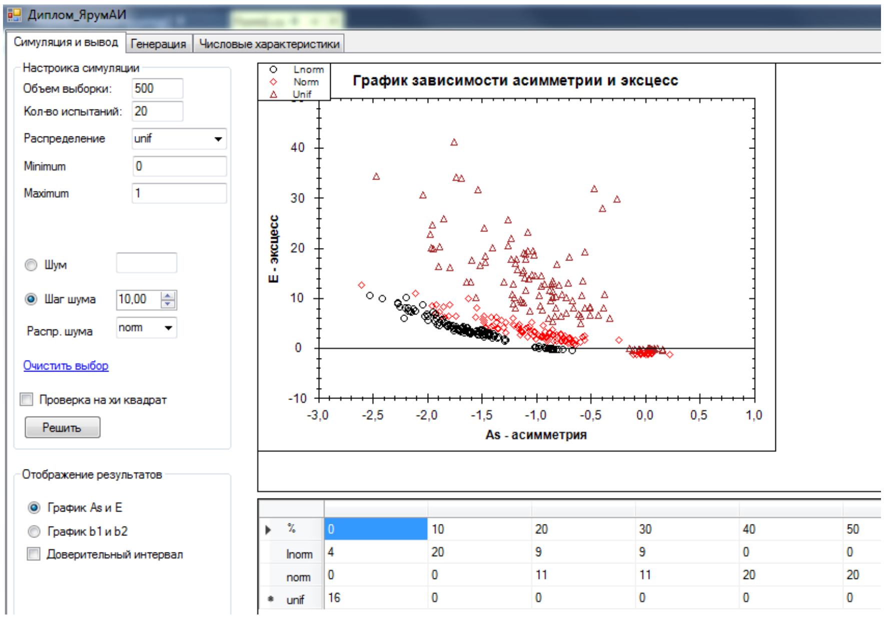


Рисунок 53 – Равномерное распределение с шумом от 0 до 50% по нормальному

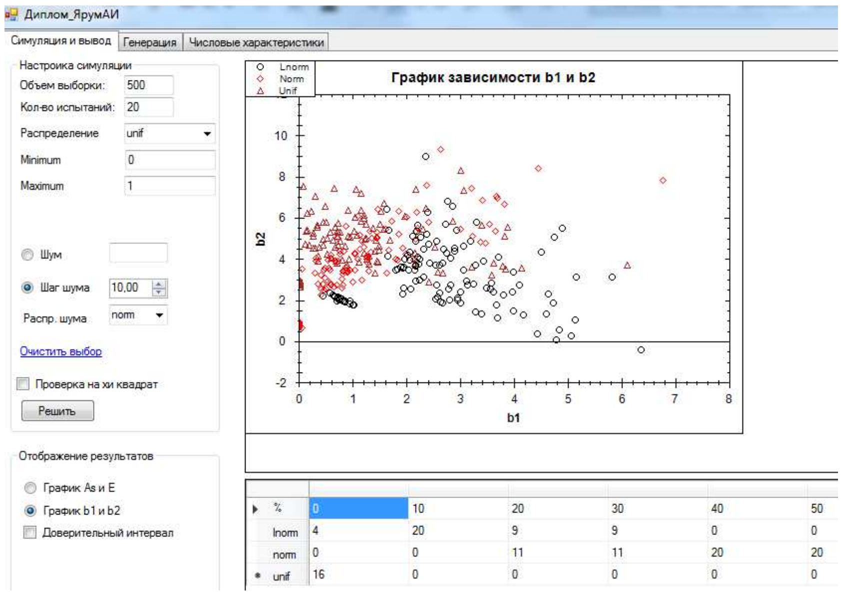


Рисунок 54 – Равномерное распределение с шумом от 0 до 50%, b_1, b_2 по нормальному

Далее рассмотрим случай с помехой от 0 до 10% с шагом в 1%, результат представлен на рисунке 55.

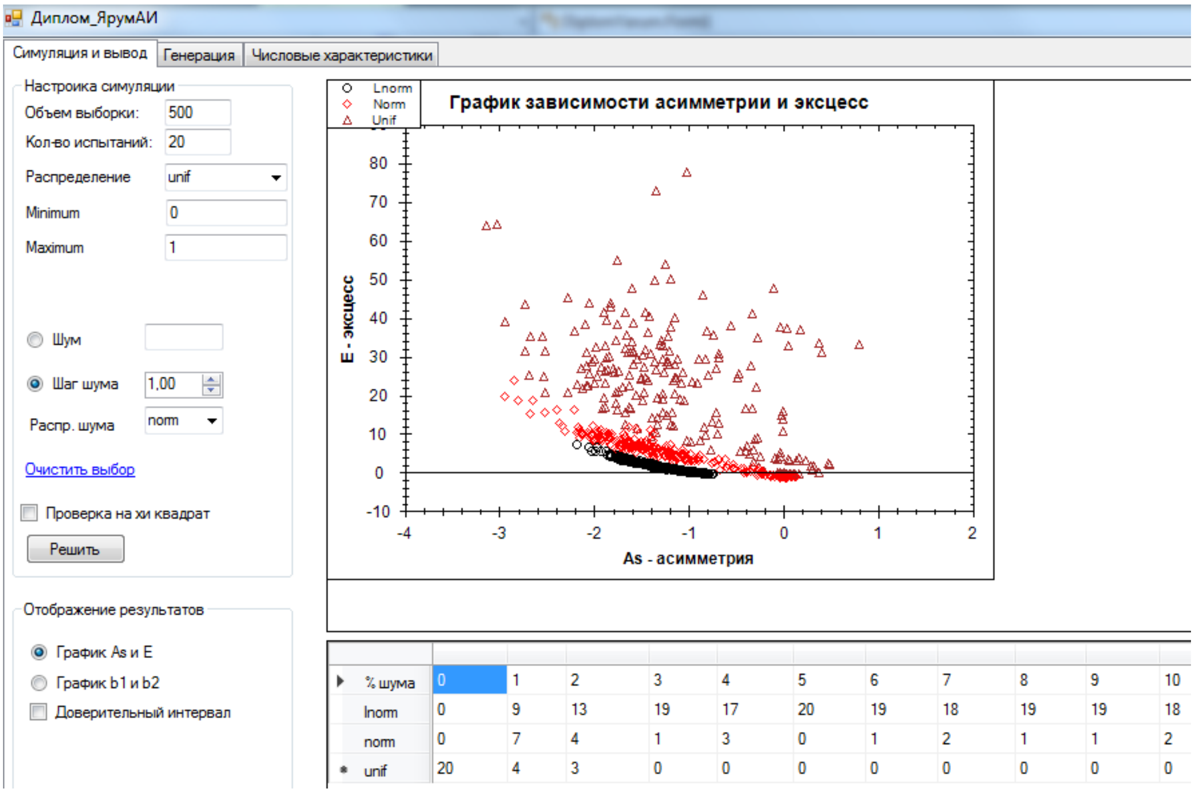


Рисунок 55 – Равномерное распределение с шумом от 0 до 10% по нормальному

Проведем эксперимент с помехой от 0 до 2% с шагом в 0,5, результат изображено на рисунке 56.

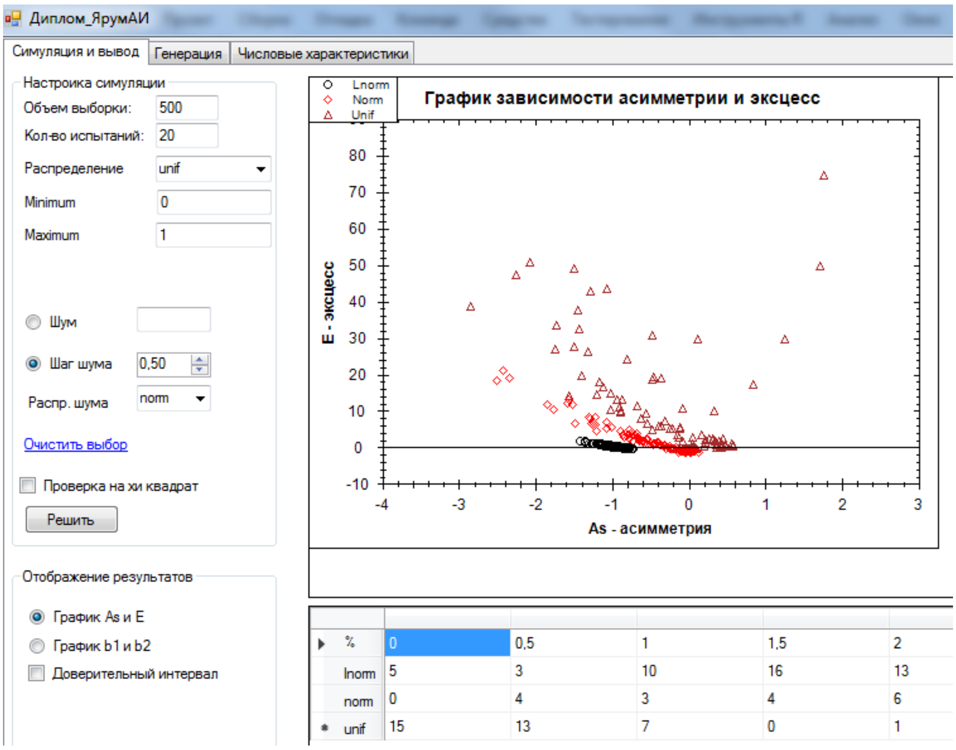


Рисунок 56 – Равномерное распределение с шумом от 0 до 2% по нормальному

На основании полученных результатов построим зависимость получения правильных ответов от процента шума и отметим их на рисунке 57.

Рисунок 57 – Зависимость при равномерном распределении

При сравнении топологического подхода с классическим критерием хи квадрат сделан вывод, что топологический метод немного уступает классическому. Это связано с объемом выборки. Для топологического подхода требуется большее количество экспериментальных данных, чем для критерия хи-квадрат.

ЗАКЛЮЧЕНИЕ

В процессе выполнения магистерской диссертации проведен анализ теоретических источников об определении закона распределения вероятности случайной величины классическими и топологическими методами. На основание чего можно утверждать, что необходим новый метод определения закона, основанный на форме полигона выборки, а точнее по значениям оценок асимметрии и эксцесса, значений b_1, b_2 и выполнения условий системы формулы 2.1.3. Т.е. для данного метода применяются три критерия определения закона распределения исследуемой выборки.

Также проведен небольшой анализ существующих программных решений для статистических расчетов, на основании которого сделан вывод о том, что наиболее подходящим является язык и среда статистических расчетов R, который позволил генерировать выборку по закону распределения, преобразовывать ее согласно алгоритму. Для удобства проверки эффективности разработанного алгоритма создан программный продукт на языке C# в среде разработки Visual Studio 2017 Community. В процессе исследования выявлено, что в некоторых случая язык R не поддерживает векторного расчета. Это существенно замедляет работу программы.

Данный метод следует применять, когда объем выборки превышает пороговое значение в 60 – 70 измерений. В противном случае, алгоритм будет выдавать неверный результат.

Для повышения робастности сгенерированная выборка очищается от аномальных измерений с помощью критерия Грабса. Причем осуществляется одновременная проверка на три максимальный или минимальных выбраса. Благодаря чему удается существенно повысит качество правильно принятых гипотез.

При сравнение топологического метода с классическим методом хи - квадрат сделан вывод, что при небольших объемах выборки (менее 50) и большой зашумлённости экспериментальные точки на столько сильно смешиваются друг с другом, что по топологическому методу невозможно дать однозначный ответ в пользу того или иного закона распределения, в то время как критерий хи – квадрат дает однозначные результаты. При выборке большого объема и малой зашумлённости алгоритм работает эффективнее. Следует учесть, что разработанный алгоритм опробован не на всех законах распределения случайной величины, и возможны ситуации, когда данный метод будет лучше справляться с поставленной задачей для какого-либо другого распределения, чем классические критерии.

Также необходимо отметить, что данный метод может быть использован вместе с классическими критериями при большом наборе данных, что позволит точнее определить по полигону выборки закон, путём выделения определенной группы распределений, которую далее проверяют на каком-либо другом классическом методе.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Айвазян, С. А. Прикладная статистика: Основы моделирования и первичная обработка данных : Справочное изд. / С. А. Айвазян, И. С. Енюков, Л. Д. Мешалкин. — М.: Финансы и статистика, 1983. — 471 с.
2. Андронов, А. М. Теория вероятностей и математическая статистика : Учебник для вузов. / А. М. Андронов, Е. А. Копытов, Л. Я. Гринглаз. — СПб. : Питер, 2004.
3. Вадзинский, Р.Н. Справочник по вероятностным распределениям. : / Р. Н. Вадзинский. — М.: Наука, 2001. — 295 с.
4. Вентцель, Е. С. Теория вероятностей : учебник / Е.С. Вентцель. — М. : Наука, 1969. — 576 с. с илл.
5. Гмурман, В. Е. Теория вероятностей и математическая статистика. : Учебник для прикладного бакалавриата. / В. Е. Гмурман. — М.: Юрайт, 2015. — 480 с.
6. Зорин, А.В., Введение в прикладной статистический анализ в пакете R. : Учебно-методическое пособие. / А. В. Зорин, М. А. Федоткин. — Нижний Новгород. : Нижегородский гос. ун-т им. Н. И. Лобачевского, 2010. — 50 с.
7. Кобзарь, А. И. Прикладная математическая статистика. Для инженеров и научных работников : / А. И. Кобзарь. — М. : Физматлит, 2006. — 816 с.
8. Коробейников, А.И. Анализ данных с R. : Методические указания к спецкурсу «Вычислительные методы и пакеты в статистическом исследовании». / А. И. Коробейников, С. В. Малов, И. В. Матвеева. — Санкт-Петербург, 2010. — 12 с.
9. Кремер, Н. Ш. Теория вероятностей и математическая статистика. : Учебник для студентов вузов / Н. Ш. Кремер. — М. : ЮНИТИ-ДАНА, 2012. — 551с.

10. Пустыльник, Е. И. Статистические методы анализа и обработки наблюдений. : / Е. И. Пустыльник. — М. : Наука, 1968.
11. Савельев, А. А., Использование языка R для статистической обработки данных. : Учебно-методическое пособие / А. А. Савельев, С. С. Мухарамова, А. Г. Пилюгин. — Казань. : Казанский гос. ун-т, 2007. — 28 с.

Федеральное государственное автономное
образовательное учреждение
высшего образования
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»
Институт космических и информационных технологий
Базовая кафедра «Интеллектуальные системы управления»

УТВЕРЖДАЮ
Заведующий кафедрой
_________ Ю.Ю. Якунин
подпись инициалы, фамилия
« » июня 2018 г.

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

Топологическая идентификация закона распределения вероятности одномерной выборки

27.04.03 Системный анализ и управление

27.04.03.02 Системный анализ данных и технологий принятия решений

Научный руководитель	08.06.2018 подпись, дата	доцент, канд. техн. наук должность, ученая степень	А.А. Даничев инициалы, фамилия
Выпускник	08.06.2018 подпись, дата		А.И. Ярум инициалы, фамилия
Рецензент	08.06.2018 подпись, дата	доцент, канд. физ.-мат. наук должность, ученая степень	И.М. Федотова инициалы, фамилия

Красноярск 2018