

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
«КУРГАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

Кафедра «Безопасность информационных
и автоматизированных систем»

**ТЕОРИЯ ВЕРОЯТНОСТИ
И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА**

Методические рекомендации
к выполнению лабораторных работ
по дисциплине «Теория вероятности и математическая статистика»
для студентов очной формы обучения
10.05.03 «Информационная безопасность автоматизированных систем»

Кафедра: «Безопасность информационных и автоматизированных систем»
Дисциплина: «Теория вероятности и математическая статистика»
10.05.03 «Информационная безопасность автоматизированных систем»

Составил: канд. биол. наук, доцент А. В. Человечкова
старший преподаватель Д. С. Иванов

Печатается в соответствии с планом издания, утвержденным методическим советом университета 13 декабря 2024 г.

Утверждены на заседании кафедры 16 декабря 2025 г.

ВВЕДЕНИЕ

Почти во всех областях человеческой деятельности существуют ситуации, когда те или иные эксперименты или наблюдения могут повторяться большое число раз в одинаковых условиях. Теорию вероятностей интересуют те эксперименты, результат которых, выраженный каким-либо образом, может меняться от опыта к опыту. События, относящиеся к результату эксперимента, которые при этом могут происходить или не происходить, называют обычно случайными событиями. Формально-логическое изучение закономерностей случайных явлений имеет дело с их математическими моделями, представленными в виде случайных событий. Исследование построенных теоретико-вероятностных моделей позволяет изучить закономерности этих явлений.

Лабораторные работы по теории вероятности нужны для практического закрепления теории, развития вероятностного мышления и навыков работы с неопределенностью в реальных задачах, а также для экспериментального подтверждения теоретических положений и освоения прикладного инструментария (например, критериев проверки гипотез) с помощью программного обеспечения. Они помогают понять, как применять формулы комбинаторики, теории вероятности и статистики для анализа данных и принятия решений.

ЛАБОРАТОРНАЯ РАБОТА № 1

Тема. Опыт Пирсона (подбрасывание монеты).

Цель работы. Смоделировать процесс проведения опыта по подбрасыванию монеты.

Теоретическое введение

При выполнении некоторых опытов в естественных условиях проявляется одно из свойств вероятности. Пусть вероятность одного исхода некоторого явления равна p , и пусть это явление повторяется многократно N раз. Тогда этот исход произойдет примерно pN раз. Если вместо выражений «естественные условия», «примерно» и так далее использовать точные формулы, то получаем закон больших чисел. Такое название связано с тем, что чем больше N , тем точнее закон будет выполняться. Этот закон позволяет предсказывать результаты событий с «непредсказуемым» исходом. Суть его заключается в следующем: если явление подчиняется вероятностным законам и повторяется много раз, можно заранее достаточно точно предсказать, сколько раз произойдет каждый из возможных его исходов, хотя невозможно заранее предсказать результат в каждом конкретном случае.

Рассмотрим простейший опыт – подбрасывание монеты. В опыте возможны два исхода: монета может упасть орлом или решкой. Обоим исходам естественно приписать равные вероятности 0,5. Поэтому, опираясь на закон больших чисел, можно утверждать, что, если монету бросить 1000 раз, орел выпадет примерно 500 раз. И такие опыты проводились. Американский статистик А. К. Пирсон, бросив монету 24000 раз, получил 12012 исходов выпадения герба.

Вероятность выпадения орла $P(O)$ можно рассчитать по классической формуле определения вероятности:

$$P(O) = m/n,$$

где m – количество результатов броска монеты с выпадением орла (благоприятные события),

n – общее количество бросков.

Выполнение работы

В ходе выполнения лабораторной работы необходимо проверить закон больших чисел на примере смоделированного опыта Пирсона.

Задания:

1) Выберите среду и способы моделирования процесса подбрасывания монеты. Пример интерфейса, компонентов программы и ее выполнения приведен на рисунке 1.

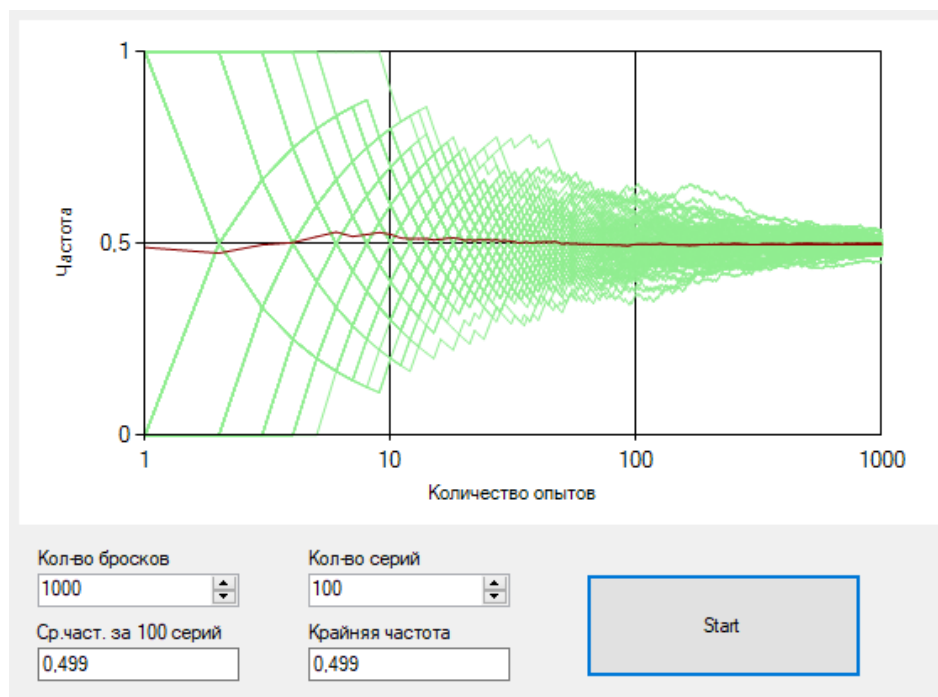


Рисунок 1 – Пример интерфейса

2) В программе необходимо предусмотреть возможность изменения входных параметров эксперимента (количество подбрасываний – до 1000, количество повторений опыта – до 100).

3) Смоделировать графическое отображение вероятности выпадения орла $P(O) = m/n$.

4) Найти и отобразить среднее значение вероятности для каждого броска. Пример работы программы с пояснениями представлен на рисунке 2.

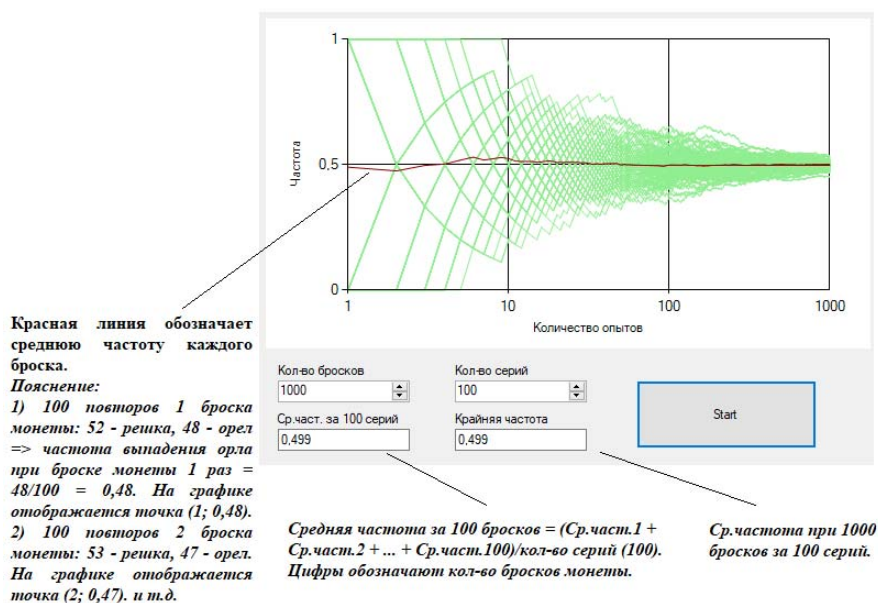


Рисунок 2 – Пояснения по отображаемым данным

- 5) Провести эксперимент по различным входным данным (от 50 бросков до 1000 с сериями от 10 до 100) и составить отчет со сравнительной таблицей.
6) Сделать вывод о результате экспериментов.

ЛАБОРАТОРНАЯ РАБОТА № 2

Тема. Моделирование задачи о встрече.

Цель работы. Смоделировать картину встречи.

Теоретическое введение

Классическое определение вероятности оказывается эффективным для решения целого спектра задач, но обладает и рядом ограничений. Одним из них является неприменимость к испытаниям с бесконечным количеством исходов.

На отрезке $[0;1]$ наудачу выбирается точка. Необходимо определить вероятность того, что она попадет в интервал, например, $[0,4;0,7]$.

Применение формулы классической вероятности $P(A) = \frac{m}{n}$ не является возможным, т. к. $n \Rightarrow \infty$. Поэтому для решения этой проблемы применяется формула **геометрической вероятности**.

Геометрическая вероятность – вероятность, с которой наступление некоторого события A в испытании определяется отношением

$$P(G) = \frac{g}{G},$$

где G – *геометрическая мера*, выражающая общее число всех возможных и равновозможных исходов,

g – *мера*, выражающая количество благоприятствующих событию G .

В качестве геометрической меры чаще всего выступает длина, площадь или объем.

Говорят, что эксперимент удовлетворяет условиям геометрического определения вероятности, если его исходы можно изобразить точками некоторой области так, что вероятность попадания точки в любую часть не зависит от формы или расположения внутри, а зависит лишь от меры области (и, следовательно, пропорциональна этой мере).

Пример 1:

Определить $P(G)$, что брошенная на отрезок $[0;1]$ точка попала в промежуток $[0,4;0,7]$. В этом случае общее число исходов выражается длиной всего доступного отрезка: $L=1-0=1$ ед., а благоприятствующие событию A исходы – длиной вложенного отрезка: $l = 0,7-0,4=0,3$ ед. По геометрическому определению вероятность наступления события A равна:

$$P(G) = \frac{l}{L} = \frac{0,3 \text{ ед}}{1 \text{ ед}} = 0,3.$$

Пример 2:

Двухметровую ленту случайным образом разрезают ножницами. Найти вероятность того, что длина обрезка составит не менее 160 см. Обозначим: $P(G)$ – вероятность, что длина обрезка составит не менее 1,6 м. Ленту можно разрезать где угодно. Общему числу исходов соответствует ее длина: $L= 2$ м.

Благоприятствующие событию А участки будут получаться при условии, что произведется разрез в 40 или менее сантиметров с одного или другого конца. Следовательно, расчет длины будет представлять из себя сумму длин отрезков всех вариантов (*Пояснение: благоприятными событиями являются два разных – отрезать от ленты 40 см слева или справа, следовательно, $l = 0,4м + 0,4м = 0,8м$*). По геометрическому определению: $P(G) = \frac{l}{L} = \frac{0,8м}{2м} = 0,4$.

Выполнение работы

Условие задачи. Два студента собрались встретиться перед парами в столовой и пообедать. Они условились, что каждый подойдет в некоторый общий интервал времени Т (например: Т = 1 час = 60 минут (с 12:00–13:00)). Из-за огромного количества посетителей столовой, по приходу любого из студентов первым, он начинает обед не дожидаясь второго. При этом каждый кушает одинаковое время – t минут. Необходимо найти вероятность того, что студенты смогут вместе пообедать.

Задания:

1) Выбрать среду и способы моделирования. Пример интерфейса и компонентов программы и ее выполнения (рисунок 3).

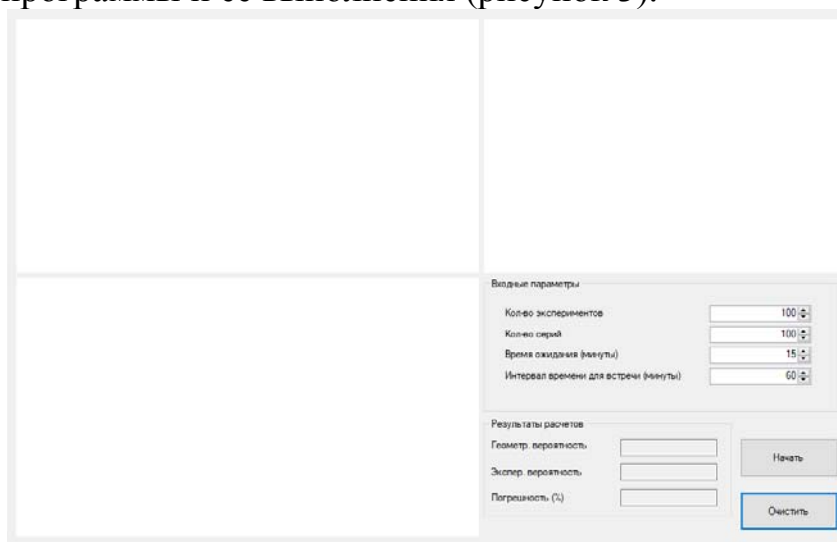


Рисунок 3 – Пример интерфейса

2) Предусмотреть возможность изменения входных параметров эксперимента (количество попыток встретиться – до 1000 (n), количество повторов каждой попытки с теми же условиями (серии) (s) – до 100, время ожидания (обеда) – от 1 до 60 минут, интервал встречи – от 1 часа).

3) Смоделировать графическое отображение вероятности $P(A) = m/n$, где m – кол-во «состоявшихся» встреч (рисунок 4) и картины встречи, обозначая состоявшиеся – зеленым, остальные – красным (рисунок 5).

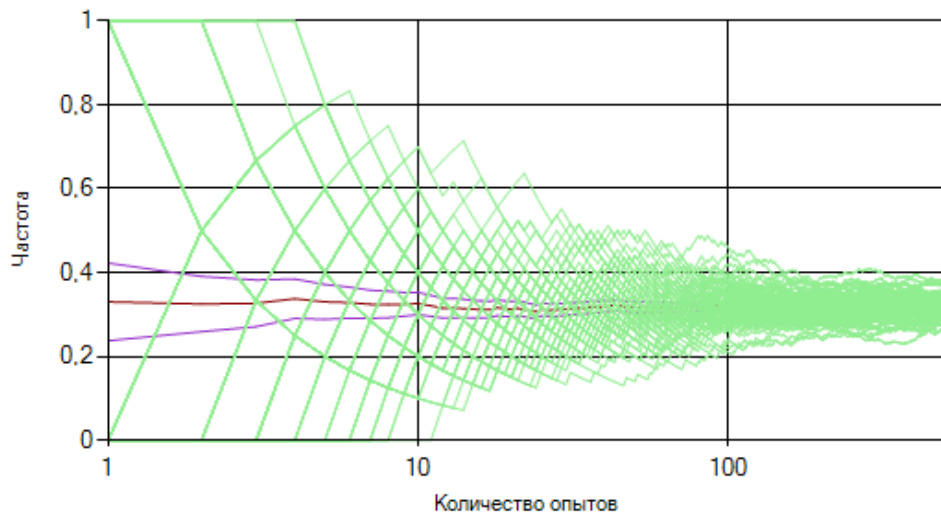


Рисунок 4 – Графическое отображение вероятности

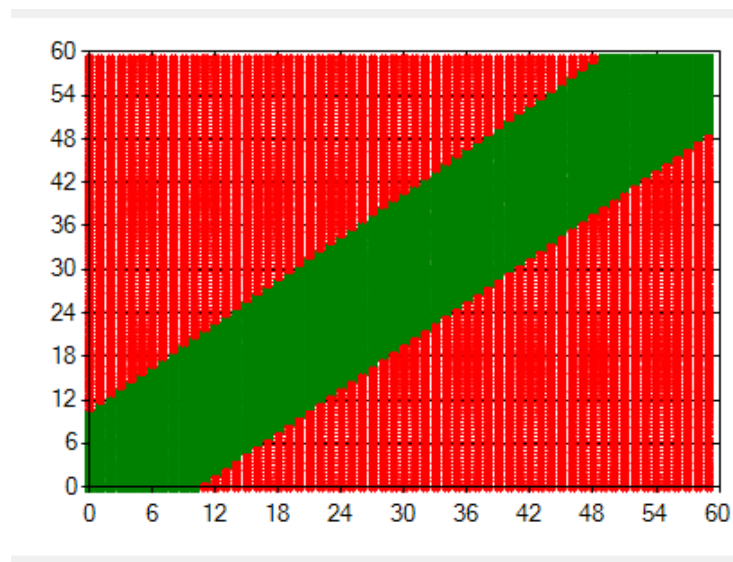


Рисунок 5 – Картина встречи

4) Найти и отобразить математическое ожидание для каждой серии и отобразить его (рисунок 4 – **красная линия**) - $M(y) = \sum_{i=1}^s \frac{y_i}{s}$. Отобразить доверительный интервал с заданной вероятностью ($P = 0.95$) по математическому ожиданию (max и min – границы):

$$DOV = M(y) \pm \frac{\Phi(P)}{\sqrt{s}} * \sqrt{\frac{\sum_{i=1}^s (y_i - M(y))^2}{s-1}}, \Phi(P) = 1,96.$$

Рассчитать отклонение математического ожидания от среднего значения доверительного интервала $\varepsilon = \left| \frac{\max + \min}{2\sigma} - 1 \right|$ для каждого эксперимента и графически отобразить его (третий график на рисунке 6).

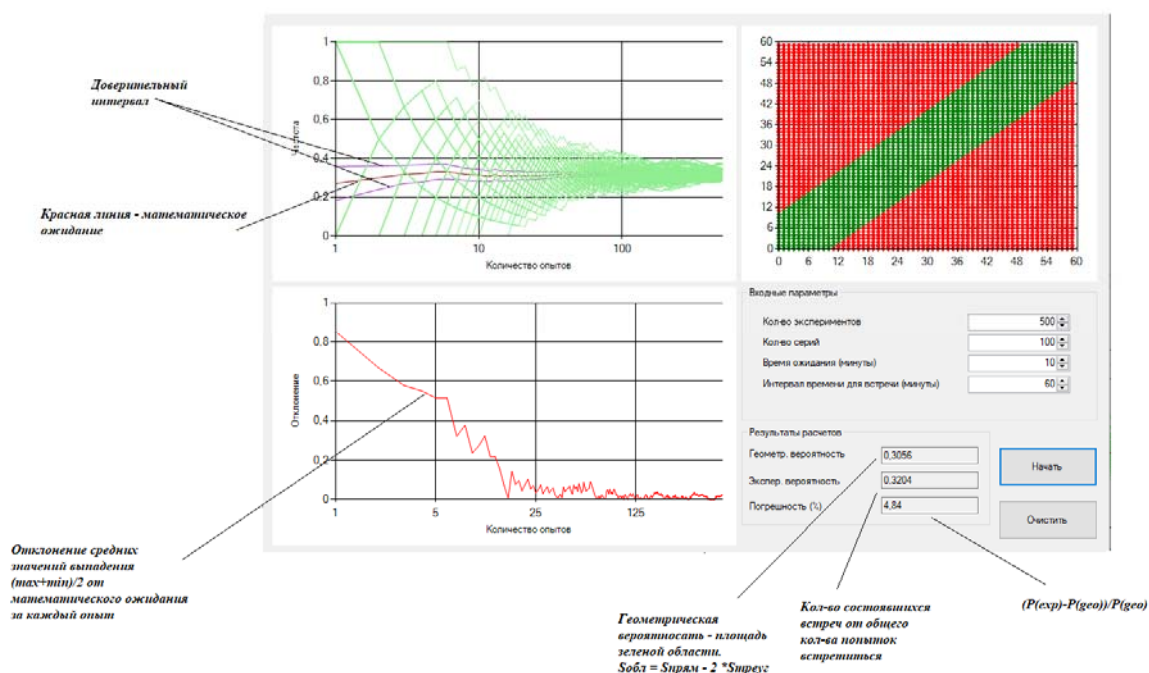


Рисунок 6 – Пример работы с пояснения по отображаемым данным

5) Найти геометрическую вероятность $(P(G) = \frac{g}{G})$, по графику картины встречи и сравнить с экспериментальной (классической $P(A) = \frac{m}{n}$), вывести значение ошибки $\epsilon = \left| \frac{P(A) - P(G)}{P(G)} \right|$.

6) Провести эксперимент по различным входным данным (от 10 до 1000 попыток встреч с сериями от 10 до 100) и составить отчет со сравнительной таблицей. Время и интервал встречи выбираются студентом на собственное усмотрение.

7) Сделать вывод о результате экспериментов.

ЛАБОРАТОРНАЯ РАБОТА № 3

Тема. Случайные блуждания (задача о блуждании частицы).

Цель работы. Смоделировать картину блужданий студента.

Теоретическое введение

Случайное блуждание – это недетерминированное передвижение по решетке: стартуя из нуля, мы делаем один шаг в секунду, переходя в одну из соседних вершин к той вершине, в которой мы находились в предыдущий момент. Описать случайное блуждание можно математической моделью случайного процесса, который описывает путь, состоящий из последовательных случайных шагов. При этом решение, в какую вершину шагнуть, принимается случайным образом – мы подкидываем «монетку» с

четырьмя сторонами, при этом вероятность каждой из сторон равна 0,25 (или, если так проще думать, две обычных монетки подряд – первая выбирает направление: вертикаль или горизонталь, вторая смещение: увеличение или уменьшение координаты).

Случайное блуждание есть случайный процесс, а именно последовательность так называемых случайных величин $v_0, v_1, v_2, \dots$. То есть, находясь в вершине v_n в момент времени n , мы можем попасть в одну из четырех соседних вершин с одинаковыми вероятностями.

Основными характеристиками являются:

1 *Стохастический процесс*. Случайное блуждание – это случайный процесс, состояние которого изменяется со временем случайным образом.

2 *Случайные шаги*. Перемещение из одной точки в другую происходит непредсказуемо, но подчиняется законам вероятности.

3 *Независимость шагов (марковское свойство)*. Вероятность следующего шага зависит только от текущего положения, а не от всей истории перемещений.

4 *Симметричность*. Симметричное блуждание на целой прямой, где частица движется вправо (+1) или влево (–1) с вероятностью 1/2.

Выполнение работы

Условие задачи. Студент «S» второго курса после очередного учебного дня возвращается домой, чтобы подготовиться к сдаче лабораторной работы по теории вероятности. Он идет по центру города, схема которого приведена на рисунке 7.

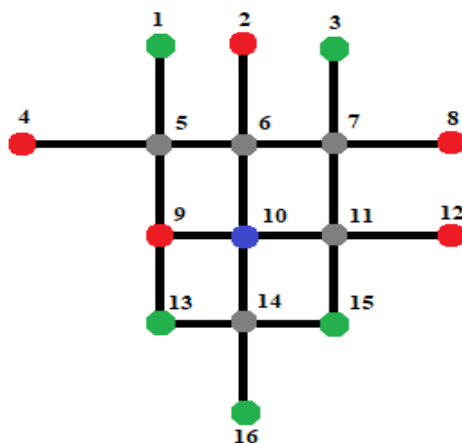


Рисунок 7 – Схема города

Находясь на перекрестке, первокурсник случайным образом решает, куда ему идти дальше (каждое из возможных направлений он выбирает с равной вероятностью). Если он окажется рядом с кофейней (**красные** кружочки на рисунке – 2,4,8,9,12), то непременно зайдет в нее, а если окажется рядом со автобусной остановкой (**зеленые** кружочки на рисунке – 1,3,13,15,16), то решит, что это знак судьбы и ему пора заканчивать развлекаться и надо заняться

учебой, – едет домой. Как найти вероятность того, что студент «S» **не** **доберется** до кофейни, а попадет на остановку и поедет домой делать лабораторные работы по теории вероятности и математической статистике, если он стоит на перекрестке, помеченном синим кружочком?

Задания:

- 1) Составить и решить систему линейных уравнений.
- 2) Выбрать среду и способы моделирования.
- 3) Предусмотреть возможность изменения входных параметров эксперимента (кол-во попыток «взяться за ум» – до 1000 (n)).
- 4) Смоделировать графическое отображение маршрута движения (каждый – линия определенного цвета). Достижение «контрольных точек» (кофейни или автобусной остановки) не изменяет место расположения студента, до окончания числа шагов (т. е. достигнув точки 9 из начальной позиции за 1 шаг, остальные 999 шагов при $n = 1000$ дублируют значение 9). Пример отображения приведен на диаграмме (рисунок 8).

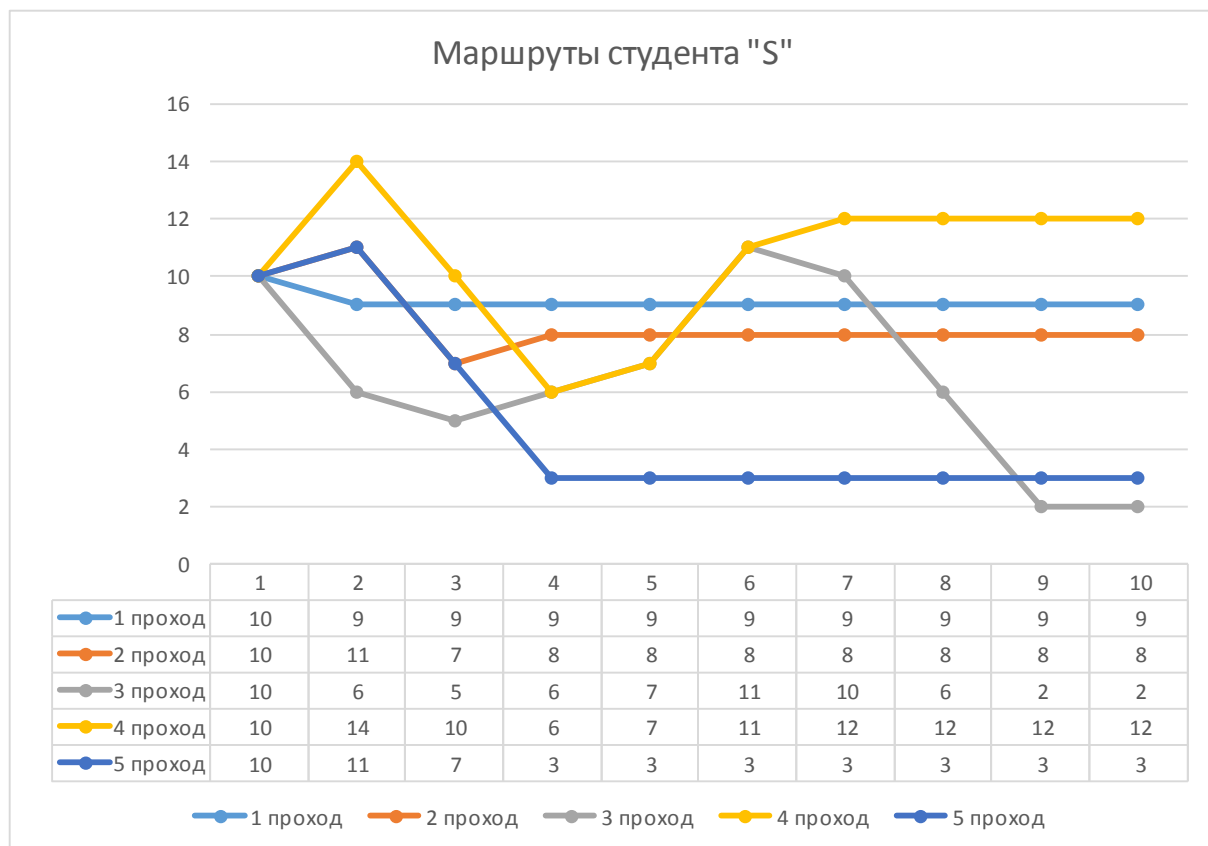


Рисунок 8 – Диаграмма построения 5 маршрутов студента «S»

- 5) Рассчитать вероятность достижения кофейни студентом «S» из начального положения: $P(b) = \text{количество окончаний «путешествия» в кофейне (BAR)} / n \Rightarrow \text{BAR}/n$.

6) Сравнить эмпирический результат с теоретическим и рассчитать погрешность.

7) Провести эксперимент по различным входным данным (50, 100, 250, 500, 1000 «путешествий» студента «S») с расчетом погрешности (%) для каждого количества походов в виде таблицы. *Погрешность* = $|\text{результат практики} - \text{результат решения системы}| / \text{результат решения системы}$. Результат решения системы – вероятность достижения цели из начальной точки (10).

8) Отобразить погрешность в виде графика по данным из таблицы.

ЛАБОРАТОРНАЯ РАБОТА № 4

Тема. Датчики случайных величин.

Цель работы. Научиться реализовать различные генераторы случайных чисел (ГСЧ) и оценивать вероятность попадания полученной случайной величины в заданный интервал, математическое ожидание и дисперсию по результатам наблюдения полученных значений при проведении опытов.

Теоретическое введение

Генераторами или датчиками случайных величин называют различные технические устройства, вырабатывающие случайные величины. Чаще всего для построения датчика используют «шумящие» радиоэлектронные приборы (диоды, тиратроны, газотроны и др.). Не вдаваясь в технические подробности, рассмотрим один из возможных способов построения датчика, вырабатывающего случайные двоичные цифры а.

Нетрудно представить себе счетчик, который подсчитывает количество v -флуктуаций напряжения шумящего прибора, превышающих заданный уровень зафиксированного времени. Еще проще устроить счетчик, который выдавал бы число $v \pmod{2}$, т. е. 0 при четном v и 1 при нечетном v . Если вероятности появления 0 и 1 в таком процессе равны между собой, то можно считать, что устройство вырабатывает случайную последовательность двоичных цифр.

Если вероятность появления нуля отлична от половины, то можно ввести какую-нибудь схему стабилизации вероятности. Например, можно группировать цифры парами и выдавать 0 при получении пары 01 и 1 – при получении пары 10, а пары 00 и 11 просто опускать. Так как-то в результате получим последовательность нулей и единиц с равными вероятностями.

Обычно датчики случайных чисел содержат генераторов описанного типа, работающих независимо, так что датчиком выдается приближенное случайное число, записанное в форме двоичной дроби. Для случайных чисел отведена специальная ячейка в накопителе, и скорость генерирования их столь велика, что на каждом такте работы ЭВМ в этой ячейке получается новое случайное число.

Применение датчиков случайных чисел свободно от тех недостатков, которые препятствуют широкому применению таблиц: не требуется места во внутреннем накопителе и запас чисел практически не ограничен. Тем не менее

подавляющее большинство задач, решенных методом Монте-Карло, сосчитано без применения датчиков, имеющих свои новые недостатки.

Во-первых, числа, выработанные датчиком, нельзя воспроизвести. Это затрудняет контроль расчетов и делает невозможным счет на таких ЭВМ, на которых двойной пересчет является правилом. Во-вторых, приходится содержать и эксплуатировать дополнительное устройство, которое требует ухода и регулярной проверки «качества» вырабатываемых чисел с помощью специальных тестов.

В основе метода Монте-Карло лежит генерация случайных чисел, которые должны быть равномерно распределены в интервале (0; 1).

Если генератор выдает числа, смещенные в какую-то часть интервала (одни числа выпадают чаще других), то результат решения задачи, решаемой статистическим методом, может оказаться неверным. Поэтому проблема использования хорошего генератора действительно случайных и действительно равномерно распределенных чисел стоит очень остро.

Математическое ожидание m_r и дисперсия D_r такой последовательности, состоящей из n случайных чисел r_i , должны соответствовать формулам (1 и 2), если это действительно равномерно распределенные случайные числа в интервале от 0 до 1:

$$m_r = \frac{\sum_{i=1}^n r_i}{n} = 0,5, \quad (1)$$

$$D_r = \frac{\sum_{i=1}^n (r_i - m_r)^2}{n} = \frac{1}{12}, \quad (2)$$

Если пользователю потребуется, чтобы случайное число x находилось в интервале $(a; b)$, отличном от $(0; 1)$, нужно воспользоваться формулой (3).

$$x = a + (b - a) \cdot r, \quad (3)$$

где r — случайное число из интервала $(0; 1)$. Законность данного преобразования демонстрируется на рисунке 9.

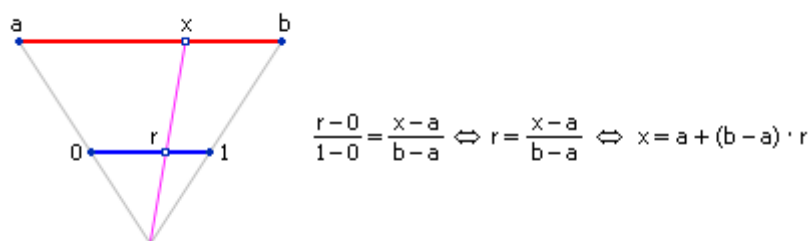


Рисунок 9 – Схема перевода числа из интервала $(0; 1)$ в интервал $(a; b)$

Теперь x — случайное число, равномерно распределенное в диапазоне от a до b .

Эталоном генератора случайных чисел (ГСЧ) принят такой генератор, который порождает последовательность случайных чисел с равномерным

законом распределения в интервале $(0; 1)$. За одно обращение данный генератор возвращает одно случайное число. Если наблюдать такой ГСЧ достаточно длительное время, то окажется, что, например, в каждый из десяти интервалов $(0; 0.1)$, $(0.1; 0.2)$, $(0.2; 0.3)$, ..., $(0.9; 1)$ попадет практически одинаковое количество случайных чисел – то есть они будут распределены равномерно по всему интервалу $(0; 1)$. Если изобразить на графике $k = 10$ интервалов и частоты N_i попаданий в них, то получится экспериментальная кривая плотности распределения случайных чисел, представленная на рисунке 10.

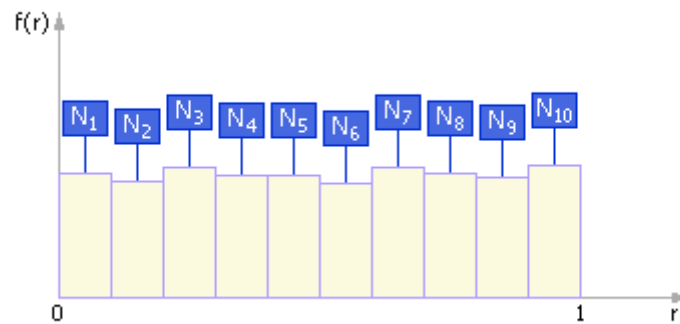


Рисунок 10 – Частотная диаграмма выпадения случайных чисел, порождаемых реальным генератором

В идеале кривая плотности распределения случайных чисел выглядела бы так, как показано на рисунке 11. То есть, в идеальном случае в каждый интервал попадает одинаковое число точек: $N_i = N/k$, где N – общее число точек, k – количество интервалов, $i = 1, \dots, k$.

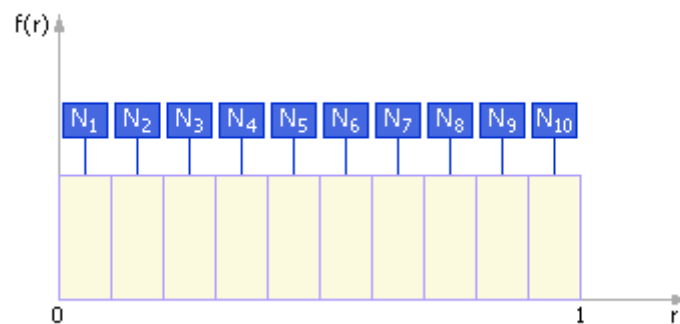


Рисунок 11 – Частотная диаграмма выпадения случайных чисел, порождаемых идеальным генератором теоретически

Следует помнить, что генерация произвольного случайного числа состоит из двух этапов:

1 генерация нормализованного случайного числа (то есть равномерно распределенного от 0 до 1);

2 преобразование нормализованных случайных чисел r_i в случайные числа x_i , которые распределены по необходимому пользователю (произвольному) закону распределения или в необходимом интервале.

Генераторы случайных чисел по способу получения чисел делятся на:

- 1 физические;
- 2 табличные;
- 3 алгоритмические.

Примеры работы генераторов случайных чисел

Физический

Примером физических ГСЧ могут служить: монета («орел» – 1, «решка» – 0); игральные кости; поделенный на секторы с цифрами барабан со стрелкой; аппаратный генератор шума (ГШ), в качестве которого используют шумящее тепловое устройство, например, транзистор (рисунки 12 и 13).

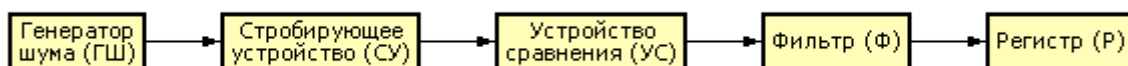


Рисунок 12 – Схема аппаратного метода генерации случайных чисел

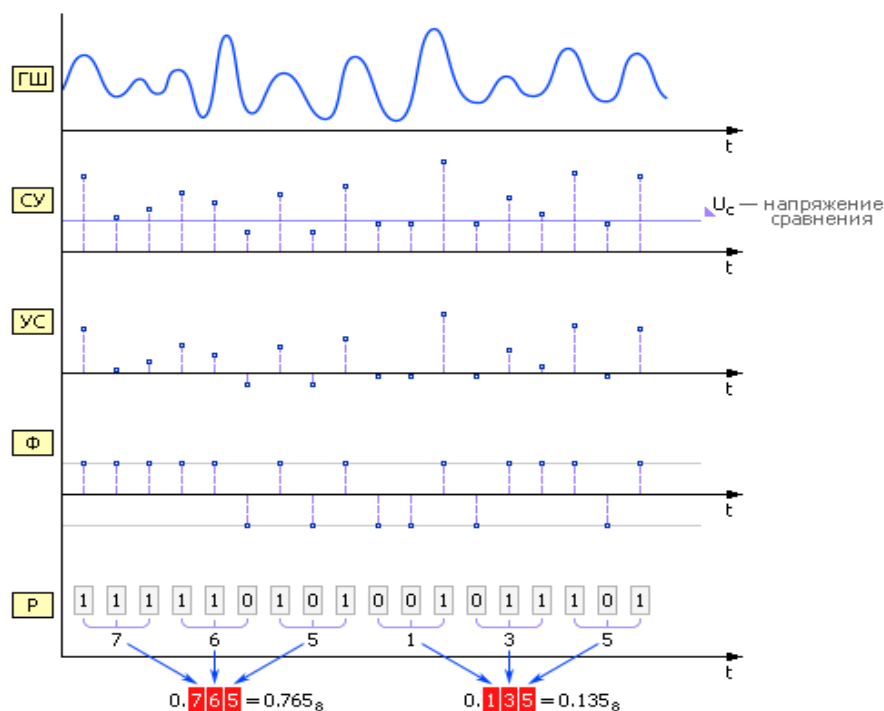


Рисунок 13 – Диаграмма получения случайных чисел аппаратным методом

Подбросьте монету 9 раз, и если монета упала решкой, то запишите «0», если орлом, то «1». Итак, допустим, что в результате эксперимента получили случайную последовательность 100110100.

Полученные случайные значения можно определить одним из двух способов.

Первый: Начертите интервал от 0 до 1. Считывая числа в последовательности слева направо, разбивайте интервал пополам и выбирайте каждый раз одну из частей очередного интервала (если выпал 0, то левую, если выпала 1, то правую). Таким образом, можно добраться до любой точки интервала, сколько угодно точно.

Итак, выпадает 1: интервал $[0; 1]$ делится пополам – $[0; 0.5]$ и $[0.5; 1]$, – выбирается правая половина, интервал сужается: $[0.5; 1]$. Следующее число 0: интервал $[0.5; 1]$ делится пополам – $[0.5; 0.75]$ и $[0.75; 1]$, – выбирается левая половина $[0.5; 0.75]$, интервал сужается: $[0.5; 0.75]$. Следующее число 0: интервал $[0.5; 0.75]$ делится пополам – $[0.5; 0.625]$ и $[0.625; 0.75]$, – выбирается левая половина $[0.5; 0.625]$, интервал сужается: $[0.5; 0.625]$. Следующее число 1: интервал $[0.5; 0.625]$ делится пополам – $[0.5; 0.5625]$ и $[0.5625; 0.625]$, – выбирается правая половина $[0.5625; 0.625]$, интервал сужается: $[0.5625; 0.625]$.

По условию точности задачи решение найдено: им является любое число из интервала $[0.5625; 0.625]$, например, 0.625.

В принципе, если подходить строго, то деление интервалов нужно продолжить до тех пор, пока левая и правая границы найденного интервала не СОВПАДУТ между собой с точностью до третьего знака после запятой. То есть с позиций точности сгенерированное число уже не будет отличным от любого числа из интервала, в котором оно находится.

Второй: разобьем полученную двоичную последовательность 100110100 на триады: 100, 110, 100. После перевода этих двоичных чисел в десятичные получаем: 4, 6, 4. Подставив спереди «0.», получим: 0.464. Таким методом могут получаться только числа от 0.000 до 0.777 (так как максимум, что можно «выжать» из трех двоичных разрядов – это $111_2 = 7_8$) – то есть, по сути, эти числа представлены в восьмеричной системе счисления. Для перевода *восьмеричного* числа в *десятичное* представление выполним: $0.464_8 = 4 \cdot 8^{-1} + 6 \cdot 8^{-2} + 4 \cdot 8^{-3} = 0.6015625_{10} = 0.602_{10}$. Итак, искомое число равно: 0.602.

Табличный

Табличные ГСЧ в качестве источника случайных чисел используют специальным образом составленные таблицы, содержащие проверенные некоррелированные, то есть никак не зависящие друг от друга, цифры. В таблице 1 приведен небольшой фрагмент такой таблицы. Обходя таблицу слева направо сверху вниз, можно получать равномерно распределенные от 0 до 1 случайные числа с нужным числом знаков после запятой. Так как цифры в таблице не зависят друг от друга, то таблицу можно обходить разными способами, например, сверху вниз или справа налево, или, скажем, можно выбирать цифры, находящиеся на четных позициях.

Таблица 1 – Случайные цифры, равномерно распределенные от 0 до 1

Случайные цифры								Равномерно распределенные от 0 до 1 случайные числа
9	2	9	2	0	4	2	6	0.929
9	5	7	3	4	9	0	3	0.204
5	9	1	6	6	5	7	6	0.269
...								...

Достоинство данного метода в том, что он дает действительно случайные числа, так как таблица содержит проверенные некоррелированные цифры. Недостатки метода: для хранения большого количества цифр требуется много памяти; большие трудности порождения и проверки такого рода таблиц, повторы при использовании таблицы уже не гарантируют случайности числовой последовательности, а значит и надежности результата.

Здесь находится таблица, содержащая 500 абсолютно случайных проверенных чисел (взято из книги И. Г. Венецкого, В. И. Венецкой «Основные математико-статистические понятия и формулы в экономическом анализе»).

Алгоритмический

Числа, генерируемые с помощью этих ГСЧ, всегда являются псевдослучайными (или квазислучайными), то есть каждое последующее сгенерированное число зависит от предыдущего.

Последовательности, составленные из таких чисел, образуют петли, то есть обязательно существует цикл, повторяющийся бесконечное число раз. Повторяющиеся циклы называются периодами.

Достоинством данных ГСЧ является быстроедействие; генераторы практически не требуют ресурсов памяти, компактны. Недостатки: числа нельзя в полной мере назвать случайными, поскольку между ними имеется зависимость, а также наличие периодов в последовательности квазислучайных чисел.

Рассмотрим несколько алгоритмических методов получения ГСЧ:

- метод серединных квадратов;
- метод серединных произведений;
- метод перемешивания;

Выполнение работы

Задания:

- 1) Выбрать среду и язык программирования.
- 2) Реализовать ГСЧ встроенными средствами.
- 3) Построить график значений случайной величины от номера опыта.
- 4) Теоретически и по результатам наблюдений значений случайной величины оценить вероятность попадания случайной величины в указанные преподавателем интервалы для разного количества опытов (1000, 100, 50, 25, 10 и 5).
- 5) Сравнить и проанализировать полученные результаты: функцию

распределения и плотность распределения вероятностей случайной величины для разного количества опытов (1000, 100, 50, 25, 10 и 5)

6) Сравнить и проанализировать полученные результаты: математическое ожидание и дисперсию случайной величины для разного количества опытов (1000, 100, 50, 25, 10 и 5).

7) Реализовать ГСЧ алгоритмическими методами (метод серединных квадратов, метод серединных произведений и метод перемешивания), описание которых можно найти в приложении 1.

8) Повторить шаги 3-6 для реализованных ГСЧ.

9) Составить отчет.

10) Сделать вывод по работе.

Дополнительный материал к лабораторной работе

Описание алгоритмических методов получения генераторов случайных чисел

Метод серединных квадратов

Имеется некоторое четырехзначное число $R0$. Это число возводится в квадрат и заносится в $R1$. Далее из $R1$ берется середина (четыре средних цифры) – новое случайное число – и записывается в $R0$. Затем процедура повторяется (рисунок 14). Отметим, что на самом деле в качестве случайного числа необходимо брать не $ghij$, а $0.ghij$ – с приписанным слева нулем и десятичной точкой.

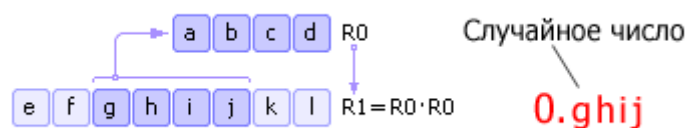


Рисунок 14 – Схема метода серединных квадратов

Недостатки метода:

1) если на некоторой итерации число $R0$ станет равным нулю, то генератор вырождается, поэтому важен правильный выбор начального значения $R0$;

2) генератор будет повторять последовательность через M^n шагов (в лучшем случае), где n – разрядность числа $R0$, M – основание системы счисления.

Для примера на рисунке 1, если число $R0$ будет представлено в двоичной системе счисления, то последовательность псевдослучайных чисел повторится через $2^4 = 16$ шагов. Заметим, что повторение последовательности может произойти и раньше, если начальное число будет выбрано неудачно.

Метод серединных произведений

Число $R0$ умножается на $R1$, из полученного результата $R2$ извлекается середина $R2^*$ (это очередное случайное число) и умножается на $R1$. По этой схеме вычисляются все последующие случайные числа (рисунок 15).

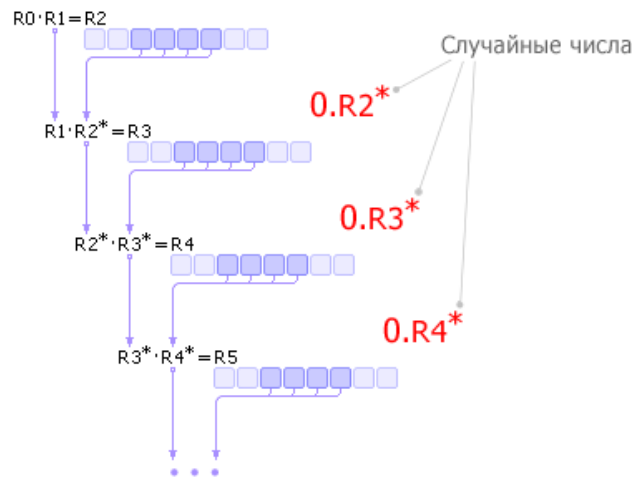


Рисунок 15 – Схема метода срединных произведений

Метод перемешивания

В методе перемешивания используются операции циклического сдвига содержимого ячейки влево и вправо. Идея метода состоит в следующем. Пусть в ячейке хранится начальное число $R0$. Циклически сдвигая содержимое ячейки влево на $1/4$ длины ячейки, получаем новое число $R0^*$. Точно так же, циклически сдвигая содержимое ячейки $R0$ вправо на $1/4$ длины ячейки, получаем второе число $R0^{**}$. Сумма чисел $R0^*$ и $R0^{**}$ дает новое случайное число $R1$. Далее $R1$ заносится в $R0$, и вся последовательность операций повторяется (рисунок 16).

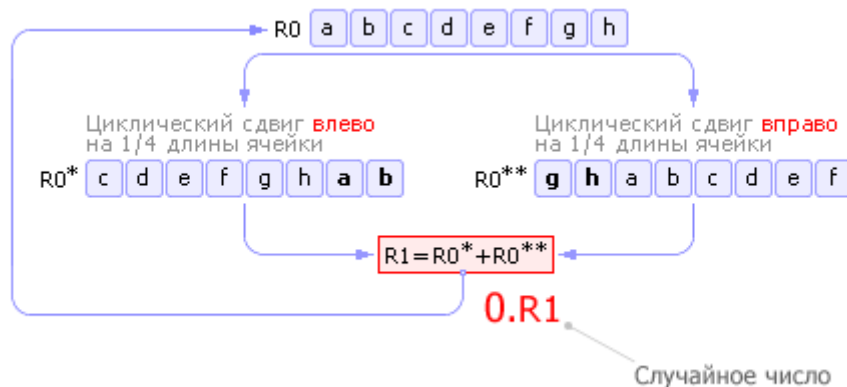


Рисунок 16 – Схема метода перемешивания

Обратите внимание, что число, полученное в результате суммирования $R0^*$ и $R0^{**}$, может не уместиться полностью в ячейке $R1$. В этом случае от полученного числа должны быть отброшены лишние разряды. Поясним это для рисунка 3, где все ячейки представлены восемью двоичными разрядами. Пусть $R0^* = 10010001_2 = 145_{10}$, $R0^{**} = 10100001_2 = 161_{10}$, тогда $R0^* + R0^{**} = 100110010_2 = 306_{10}$. Как видим, число 306 занимает 9 разрядов (в двоичной системе счисления), а ячейка $R1$ (как и $R0$) может вместить в себя максимум восемь разрядов. Поэтому перед занесением значения в $R1$ необходимо убрать один

«лишний», крайний левый бит из числа 306, в результате чего в $R1$ пойдет уже не 306, а $00110010_2 = 50_{10}$. Также заметим, что в таких языках, как Паскаль, «урезание» лишних битов при переполнении ячейки производится автоматически в соответствии с заданным типом переменной.

ЛАБОРАТОРНАЯ РАБОТА № 5

Тема. Многомерные распределения.

Цель работы. Исследование двумерного нормального распределения.

Теоретическое введение

Многомерное нормальное распределение

Многомерное нормальное распределение (или многомерное гауссовское распределение) в теории вероятностей – это обобщение одномерного нормального распределения.

Случайный вектор, имеющий многомерное нормальное распределение, называется гауссовским вектором.

Двумерное нормальное распределение

Частным случаем многомерного нормального распределения является двумерное нормальное распределение. В этом случае имеем две случайные величины X_1 и X_2 с математическими ожиданиями μ_1 и μ_2 , дисперсиями σ_1^2 , σ_2^2 и ковариацией σ_{12} .

В этом случае ковариационная матрица имеет размер 2, ее определитель рассчитывается по формуле 1.

$$\det \Sigma = \sigma_1^2 \sigma_2^2 - \sigma_{12}^2 = \sigma_1^2 \sigma_2^2 (1 - p^2) \quad (1)$$

где $p = \sigma_{12} / (\sigma_1 * \sigma_2)$ – коэффициент корреляции случайных величин.

Тогда плотность двумерного невырожденного (коэффициент корреляции по модулю не равен единице) нормального распределения можно записать в виде формулы 2.

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-p^2}} \exp\left\{-\frac{1}{2(1-p^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} - p\frac{2(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]\right\} \quad (2)$$

В том случае, если $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$, $\text{cov}(X, Y) = \sigma_{12}$ (то есть X , Y являются зависимыми), их сумма все еще распределена нормально, но в дисперсии появляется дополнительное слагаемое $2\sigma_{12}$ (формула 3).

$$X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2 + 2\sigma_{12}) \quad (3)$$

Ковариация

Ковариация – мера линейной зависимости двух величин. Она несет тот же смысл, что и коэффициент корреляции – она показывает, есть ли линейная взаимосвязь между двумя случайными величинами, и может рассматриваться как «двумерная дисперсия». Однако, в отличие

от коэффициента корреляции, который меняется от -1 до 1 , ковариация не инвариантна относительно масштаба, т. е. зависит от единицы измерения и масштаба случайных величин.

Знак ковариации указывает на вид линейной связи между рассматриваемыми величинами: если она > 0 – это означает прямую связь (при росте одной величины растет и другая), ковариация < 0 указывает на обратную связь. При ковариации $= 0$ линейная связь между переменными отсутствует (формула 4).

$$\text{cov}(X, Y) = M[(X - M[X]) * (Y - M[Y])] \quad (4)$$

Свойства ковариации:

1 Если X и Y независимы, то $\text{cov}(X, Y) = 0$.

2 Симметричность: $\text{cov}(X, Y) = \text{cov}(Y, X)$.

3 $\text{cov}(X, X) = DX$.

4 Если $\alpha, \beta = \text{const}$, то $\text{cov}(X) + \alpha, Y + \beta = \text{cov}(X, Y)$

Пример решения задачи по нахождению ковариации показан на рисунке 17.

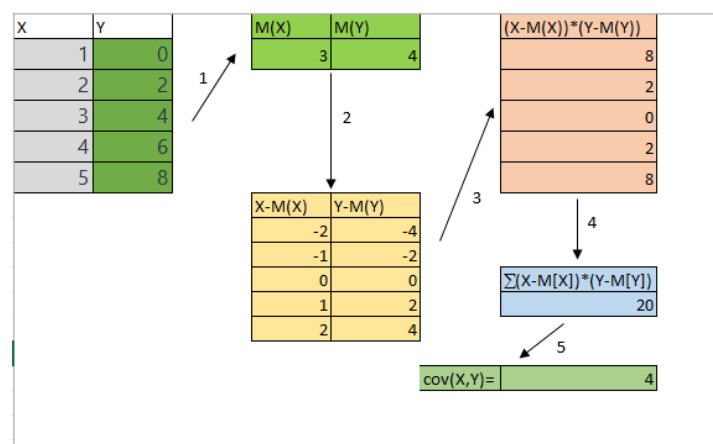


Рисунок 17 – Алгоритм нахождения ковариации двух случайных величин X и Y

Свойства многомерного нормального распределения

- Если вектор $X = (X_1, \dots, X_n)$ имеет многомерное нормальное распределение, то его компоненты $X_i, i = 1, \dots, n$ имеют одномерное нормальное распределение. Обратное верно при независимости компонентов.

- Если случайные величины $X_1, \dots, X_n$ имеют одномерное нормальное распределение и совместно независимы, то случайный вектор $X = (X_1, \dots, X_n)$ имеет многомерное нормальное распределение. Матрица ковариаций Σ такого вектора диагональна.

• Если $X = (X_1, \dots, X_n)$ имеет многомерное нормальное распределение, и его компоненты попарно некоррелированы, то они независимы. Однако, если некоторые случайные величины $X_i, i = 1, \dots, n$ имеют одномерные нормальные распределения и попарно не коррелируют, то отсюда не следует, что они независимы и имеют многомерное нормальное распределение.

Пример. Пусть $X \sim N(0, 1)$, а $\alpha = \pm 1$ с равными вероятностями и независима от указанной нормальной величины. Тогда если $Y = \alpha X \sim N(0, 1)$, то корреляция X и Y равна нулю. Однако эти случайные величины зависимы и в силу первого утверждения абзаца не имеют многомерного нормального распределения.

Многомерное нормальное распределение устойчиво относительно линейных преобразований. Если $X \sim N(\mu, \Sigma)$, а A – произвольная матрица размерности $m \times n$, то $AX \sim N(A\mu, A\Sigma A^T)$.

Таким преобразованием и сдвигом любое невырожденное нормальное распределение можно привести к вектору независимых стандартных нормальных величин.

Условное распределение

Пусть случайные векторы X и Y имеют совместное нормальное распределение с математическими ожиданиями μ_X, μ_Y , ковариационными матрицами V_X, V_Y и матрицей ковариаций C_{XY} . Это означает, что объединенный случайный вектор $z = \begin{bmatrix} X \\ Y \end{bmatrix}$ подчиняется многомерному нормальному распределению с вектором математического ожидания $\mu_z = \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}$ и ковариационной матрицей, которую можно представить в виде следующей блочной матрицы (формула 5).

$$V_z = \begin{bmatrix} V_X & C_{XY} \\ C_{YX} & V_Y \end{bmatrix}, \text{ где } C_{YX} = C_{XY}^T. \quad (5).$$

Тогда случайный вектор Y при заданном значении случайного вектора X имеет (многомерное) нормальное условное распределение со следующим условным математическим ожиданием и условной ковариационной матрицей (формула 6).

$$E(Y|X = x) = \mu_Y + C_{YX}V_X^{-1}(x - \mu_X), \quad V(Y|X = x) = V_Y + C_{YX}V_X^{-1}C_{YX} \quad (6).$$

Первое равенство определяет функцию линейной регрессии (зависимости условного математического ожидания вектора Y от заданного значения x случайного вектора X), причем матрица $C_{YX}V_X^{-1}$ – матрица коэффициентов регрессии.

Условная ковариационная матрица представляет собой матрицу ковариаций случайных ошибок линейных регрессий компонентов вектора Y на вектор X . В случае если Y – обычная случайная величина (однокомпонентный вектор), условная ковариационная матрица –

это условная дисперсия (по существу дисперсия случайной ошибки регрессии Y на вектор X).

Выполнение работы

Задания:

- 1) Ознакомиться с теоретическим введением по многомерным распределениям.
- 2) Разделиться на группы от 10 человек.
- 3) Составить две выборки (например, X – рост, Y – вес) участников группы.
- 4) Произвести расчет необходимых параметров для построения графиков одномерного нормального распределения случайной величины.
- 5) Построить графики распределения случайных величин X и Y .
- 6) Сделать вывод о зависимости случайных величин X и Y и по работе.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

- 1 Болотюк В. А. Теория вероятностей. Практикум и индивидуальные задания по комбинаторике (типовые расчеты) : учеб. пособие / В. А. Болотюк, Л. А. Болотюк. – Санкт-Петербург : Издательство «Лань», 2018. – 72 с.
- 2 Гмурман В. Е. Руководство к решению задач по теории вероятностей и математической статистике. / В. Е. Гмурман. – Москва : «Высшая школа», 1999. – 400 с.
- 3 Гмурман В. Е. Теория вероятностей и математическая статистика / В. Е. Гмурман. – Москва : В. Ш., 2003. – 480с.
- 4 Иванов Б. Н. Теория вероятностей и математическая статистика : учеб. пособие / Б. Н. Иванов. – 2-е изд., испр. и доп. – Санкт-Петербург : Издательство «Лань», 2019. – 224 с.
- 5 Ивченко Г. И. Математическая статистика. / Г. И. Ивченко, Ю. И. Медведев. – Москва : Книжный дом «ЛИБРОКОМ», 2014. – 352 с.
- 6 Мастицкий С. Э. Статистический анализ и визуализация данных с помощью R / С. Э. Мастицкий, В. К. Шитиков. – Хайдельберг – Лондон – Тольятти, 2014.
- 7 Фролов А. Н. Краткий курс теории вероятностей и математической статистики : учеб. пособие / А. Н. Фролов. – Санкт-Петербург : Издательство «Лань», 2017. – 304 с.

СОДЕРЖАНИЕ

Введение	3
Лабораторная работа № 1	4
Лабораторная работа № 2	6
Лабораторная работа № 3	9
Лабораторная работа № 4	12
Лабораторная работа № 5	20
Библиографический список	25

Человечкова Анна Владимировна
Иванов Даниил Сергеевич

**ТЕОРИЯ ВЕРОЯТНОСТИ
И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА**

Методические рекомендации
к выполнению лабораторных работ
по дисциплине «Теория вероятности и математическая статистика»
для студентов очной формы обучения
10.05.03 «Информационная безопасность автоматизированных систем»

Редактор В. А. Лисина

БИЦ Курганского государственного университета.
640020, г. Курган, ул. Советская, 63/4.
Курганский государственный университет.