Практическая работа № 2. Компьютерное моделирование выборок дискретных случайных величин, первичная обработка выборки

2.1. О содержании и задачах практической работы

В настоящей практической работе обсуждаются три задачи.

1. Компьютерное моделирование выборок дискретных случайных величин с заданным законом распределения.

2. Компьютерное моделирование выборок дискретных случайных величин путем имитации случайного эксперимента.

3. Иллюстрация некоторых фундаментальных законов теории вероятностей (статистическая «проверка» основных теорем).

О первой задаче. Как правило, моделирование случайных величин с заданным законом распределения осуществляется путем преобразования одного или нескольких независимых значений случайного числа Y, распределенного равномерно на отрезке $[0,1]$. Последовательности выборочных значений Y обычно получают с помощью теоретико-числовых алгоритмов и называют псевдослучайными числами. Один из стандартных методов моделирования дискретной случайной величины X с рядом распределения $\{x_0, x_1, x_2, \dots\}$, $P\{X = x_k\} = p_k$ $(k = 0,1,2,\dots)$ на основе псевдослучайных чисел состоит в следующем: полагают $X = x_m$, если для выбранного значения y случайной величины Y выполняется соотношение $\sum_{k=0}^{m-1} p_k \leq y < \sum_{k=0}^{m} p_k$.

В языке программирования Python широко представлены инструменты моделирования дискретных случайных величин с заданным законом распределения. Имеются генераторы выборок дискретных распределений, явно заданных рядом, или подчиняющихся базовым законам (биномиальному, Пуассона и т.п.) с конкретными параметрами. В разделе 2.2 настоящего практикума приведены некоторые из них.

О второй задаче. Для получения выборки значений случайной величины в некоторых случаях можно заменить реальный эксперимент его компьютерной моделью. При разработке алгоритма и программы, имитирующих случайный эксперимент, как правило, используют следующий подход: вначале моделируют одно случайное испытание, а затем повторяют эту процедуру n раз. Таким образом искусственно порождают (генерируют) «выборку» значений реальной случайной величины. Затем эту «выборку» обрабатывают с помощью методов математической статистики, чтобы получить представление о числовых характеристиках случайной величины и вероятностях связанных с ней событий.

О третьей задаче. Под иллюстрацией какого-либо фундаментального закона теории вероятностей будем понимать проверку согласия этого закона с опытными данными, подразумевая что эти данные получены в результате проведения компьютерного моделирования реальных экспериментов. Поясним, как провести эту проверку, на примере закона больших чисел в форме Чебышёва для одинаково распределенных слагаемых. Напомним формулировку закона.

Если $X_1, X_2, \ldots, X_n, \ldots$ – последовательность попарно независимых, одинаково распределенных случайных величин с $M[X_k] = m$, $D[X_k] = \sigma^2$, то, каково бы ни было постоянное $\varepsilon > 0$,

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{k=1}^{n} X_k - m\right| < \varepsilon\right\} = 1 .$$

Если пренебречь строгой формулировкой, то суть этого закона можно выразить следующим образом: в то время как отдельная случайная величина может довольно часто принимать значения, далекие от математического ожидания, среднее арифметическое большого числа таких величин с вероятностью, близкой к единице, принимает значения очень близкие к нему.

Для иллюстрации этого закона рассмотрим случайную величину X с известным законом распределения. Путем n -кратного проведения компьютерного эксперимента (серии) получим выборку объема n и подсчитаем среднее арифметическое элементов выборки. Сравним результат с математическим ожиданием. Многократно повторим серии экспериментов, чтобы убедиться в статистической устойчивости результата. При достаточно больших значениях n найденные в сериях средние арифметические будут близки к математическому ожиданию, их разброс относительно математических ожиданий будет тем «меньше», чем больше n.

2.2. Математические понятия и утверждения: краткая информация и ссылки на источники

1. Дискретная случайная величина. Ряд распределения [1, с. 53 – 55; 2, с. 58 – 63].

2. Закон больших чисел в форме Чебышёва и в форме Бернулли для одинаково распределенных слагаемых [1, с. 124 – 127; 2, с. 130 – 136].

3. Генеральная совокупность; случайная выборка объема n; выборка объема n (реализация случайной выборки); вариационный ряд; статистический ряд [1, с. 132 – 136; 2, с. 185 – 192].

2.3. Библиотечные инструменты языка программирования Python

Загрузка основных модулей

```
import numpy as np
import scipy.stats as sts
import scipy.special as sc
import matplotlib.pyplot as plt
import random
%matplotlib inline
```

Генераторы выборок дискретных распределений в библиотеке numpy

Функция `np.random.choice(a, size, replace=True, p)`
Возвращает выборку заданного объема `size` (если `size` – число) из массива `a.` По умолчанию повторную, если `replace=False`, то бесповторную. Выбор элемента из массива `a` осуществляется с соответствующей вероятностью массива `p`. Если `size` – кортеж, то генерируется массив заданной формы.
Другие варианты: если `a` – целое число, то генерируется случайное число из массива `arange(a)`; если параметр `p` не задан, то элементы из `a` равновозможны.

Функция `np.random.binomial(n, p, size)`
Возвращает выборку заданного объема `size` (если `size` – число) биномиального распределения с параметрами `n` и вероятностью `p`. Если `size` – кортеж, то генерируется массив заданной формы.

Функция `np.random.poisson (lm, size)`
Формирует выборку заданного объема `size` (если `size` – число) распределения Пуассона с параметром `lm`. Если `size` – кортеж, то генерируется массив заданной формы.

В модуле numpy.random также имеются генераторы выборок следующих распределений: геометрического, гипергеометрического, отрицательного биномиального, распределения logser и др.

Средства формирования вариационного и статистического рядов в библиотеке numpy

Функция `np.bincount(x)`
Вычисляет количество появлений в массиве каждого целого числа от `0` до `max(x)`.

Функция `np.sort(list(set(x)))`
Формирует набор уникальных элементов (вариант) выборки.

Случайное перемешивание

Функция `np.random.shuffle(x)` случайным образом перемешивает массив `x`.

2.4. Примеры для совместного обсуждения

Пример 1. Пусть случайная величина распределена по биномиальному закону с параметрами $n = 4$, $p = 1/3$, т.е. $X \sim B(4,\ 1/3)$. Сгенерируем выборку значений X объема 30 и визуализируем статистический ряд.

Вариант решения аналогичной задачи представлен в приложении к практической работе № 2 (пример 1).

Пример 2. Рассмотрим следующий случайный эксперимент. На карточках написаны цифры от 1 до 6. Опыт состоит в случайном выборе трех карточек и раскладывании их в порядке поступления слева направо. «Успехом» считается появление числа, не содержащего цифры 2 и 3. Наблюдаемая случайная величина X – индикатор «успеха» (она принимает значение 1, если запись числа не содержит цифры 2 и 3, и 0 – в противном случае). Поставим задачу: с помощью компьютерного моделирования случайного эксперимента получить выборку объема 10 000 значений случайной величины.

При разработке компьютерной модели будем исходить непосредственно из описания эксперимента (а не генерировать выборку по предварительно составленному теоретическому закону распределения). Результат представим в виде статистического ряда.

Сопоставим полученный статистический ряд с теоретическими вероятностями соответствующих значений случайной величины. Увидим, что результаты моделирования довольно близки к теоретическим вероятностям.

Вариант решения аналогичной задачи (карточки с цифрами 1, 2, 3, 4, 5, случайная величина – индикатор «не выпадения» двойки) представлен в приложении к практической работе № 2 (пример 2).

Пример 3. На практике при построении вероятностного пространства неизвестные вероятности часто определяют из опыта, используя в качестве оценки относительную частоту наступления элементарного события в n испытаниях. При этом в теоретическом отношении опираются на закон больших чисел в форме Бернулли, который гласит следующее.

Пусть μ_n *– число наступлений события* A *в* n *независимых испытаниях и* p *есть вероятность наступления события* A *в каждом из испытаний. Тогда, каково бы ни было постоянное* $\varepsilon > 0$, $\lim\limits_{n\to\infty} P\left\{\left|\frac{\mu_n}{n} - p\right| < \varepsilon\right\} = 1$.

Проверка согласия закона с опытом проводилась исследователями многократно. При этом рассматривались события, удовлетворяющие двум условиям: с одной стороны, вероятности этих событий можно было считать по тем или иным соображениям известными, а с другой – для наблюдения за ними было легко проводить независимые испытания.

Пополним ряды экспериментаторов, используя возможности компьютерного моделирования. Будем многократно «подбрасывать» монету и подсчитывать относительную частоту выпадений «герба» при n-кратном проведении опыта Для наглядного представления изменения относительной частоты $\frac{\mu_n}{n}$ в зависимости от числа испытаний n построим график (см. приложение к практикуму № 2, пример 3). На графике видно, что пока число опытов невелико, ломаная линия иногда значительно отклоняется от прямой $y = p = 0,5$. Но с увеличением числа опытов ломаная практически все ближе и ближе подходит к этой прямой.

Таким образом, при увеличении объема моделирования точность оценки возрастает.

Теперь рассмотрим эксперимент, состоящий в подбрасывании двух игральных костей. При моделировании этого эксперимента с помощью классической схемы множество элементарных исходов строят как множество упорядоченных пар чисел, поскольку если порядок не учитывать, то исходы не равновозможны. Убедимся в этом путем компьютерного моделирования: проверим, например, утверждение, что при многократном подбрасывании двух игральных костей частота выпадения двух единиц «в два раза» меньше частоты выпадения единицы и двойки.

2.5. Задания для самостоятельного выполнения

Задание 1_2025. «Запишите» на карточках поочередно буквы, входящие в написание ваших имени, отчества и фамилии. На оборотной стороне карточек запишите цифры, руководствуясь следующим правилом: запишите цифру 1, если это карточка с одной из букв А, Б, В, Г, Д, Е, цифру 2, если карточка с одной из букв Ё, Ж, З, И, Й, К, цифру 3, если карточка с одной из букв Л, М, Н, О, П, Р, цифру 4, если карточка с одной из букв С, Т, У, Ф, Х, Ц, цифру 5, если карточка с одной из букв Ч, Ш, Щ, Ъ, Ы, Ь, цифру 6, если карточка с одной из букв Э, Ю, Я. Рассмотрите эксперимент, состоящий в вытягивании наугад одной из карточек. Пусть случайная величина X – цифра, написанная на вытянутой карточке.

1) Составьте аналитически ряд распределения случайной величины X, найдите математическое ожидание и дисперсию. Сгенерируйте выборку значений X объема 50, составьте и визуализируйте статистический ряд. Сопоставьте полученный статистический ряд и с теоретическим законом распределения случайной величины X.

2) Проиллюстрируйте теорему Чебышёва для одинаково распределенных слагаемых.

Задание 2_2025. Задание выполняется по вариантам.

Вариант 1. Проводится эксперимент, состоящий в восьми подбрасываниях игральной кости и наблюдении цифры, выпавшей на верхней грани.

1) Выполните компьютерное моделирование эксперимента, исходя непосредственно из его описания.

2) Проиллюстрируйте теорему Бернулли, построив график зависимости относительной частоты наблюдения события A = «тройка выпала не менее четырех раз» от длины серии экспериментов.

Вариант 2. Проводится эксперимент, состоящий в подбрасывании семи игральных костей и наблюдении цифр, выпавших на верхних гранях.

1) Проведите компьютерное моделирование эксперимента, непосредственно исходя из его описания.

2) Проиллюстрируйте теорему Бернулли, построив график зависимости относительной частоты наблюдения события A = «выпадут две и только две пары одинаковых цифр» от длины серии экспериментов.

Вариант 3. Проводится эксперимент, состоящий в подбрасывании семи игральных костей и наблюдении цифр, выпавших на верхних гранях.

1) Проведите компьютерное моделирование эксперимента, непосредственно исходя из его описания.

2) Проиллюстрируйте теорему Бернулли, построив график зависимости относительной частоты наблюдения события A = «выпадут хотя бы три четные цифры» от длины серии экспериментов.

Вариант 4. Проводится эксперимент, состоящий в подбрасывании пяти игральных костей и наблюдении цифр, выпавших на верхних гранях.

1) Проведите компьютерное моделирование эксперимента, непосредственно исходя из его описания.

2) Проиллюстрируйте теорему Бернулли, построив график зависимости относительной частоты наблюдения события A = «в сумме выпадет четное число» от длины серии экспериментов.

Вариант 5. Из урны, содержащей 3 белых и 7 черных шаров, случайным образом и без возвращения извлекаются 5 шаров и цвета извлеченных шаров фиксируются.

1) Проведите компьютерное моделирование эксперимента, непосредственно исходя из его описания.

2) Проиллюстрируйте теорему Бернулли, построив график зависимости относительной частоты наблюдения события A = «вынут один белый и четыре черных шара» от длины серии экспериментов.

Вариант 6. Из урны, содержащей 6 белых и 5 черных шаров, случайным образом и без возвращения извлекаются 5 шаров и цвета извлеченных шаров фиксируются.

1) Проведите компьютерное моделирование эксперимента, непосредственно исходя из его описания.

2) Проиллюстрируйте теорему Бернулли, построив график зависимости относительной частоты наблюдения события $A =$ «вынуто не менее трех черных шаров» от длины серии экспериментов.

Вариант 7. Для сборки требуется 6 однотипных деталей. Всего имеется 11 деталей, из которых только 7 не имеют брака. Наудачу отбираются 7 деталей (одна деталь «про запас»).

1) Проведите компьютерное моделирование эксперимента, непосредственно исходя из его описания.

2) Проиллюстрируйте теорему Бернулли, построив график зависимости относительной частоты наблюдения события $A =$ «можно будет произвести сборку прибора» от длины серии экспериментов.

Вариант 8. Девять вариантов контрольной работы, написанных каждый на отдельной карточке, перемешиваются и распределяются случайным образом среди семи студентов, сидящих в одном ряду, причем каждый получает по одному варианту (два варианта остаются не использованными).

1) Проведите компьютерное моделирование эксперимента, непосредственно исходя из его описания.

2) Проиллюстрируйте теорему Бернулли, построив график зависимости относительной частоты наблюдения события $A =$ «варианты 1 и 2 окажутся использованными, причем вариант 1 достанется студенту, сидящему левее».

Вариант 9. Имеется семь отрезков длины которых равны соответственно 1, 2, 3, 5, 6, 8 и 10 см. Наудачу выбираются три из них.

1) Проведите компьютерное моделирование эксперимента, непосредственно исходя из его описания.

2) Проиллюстрируйте теорему Бернулли, построив график зависимости относительной частоты наблюдения события $A =$ «из выбранных отрезков можно составить треугольник» от длины серии экспериментов.

Вариант 10. Имеется шесть билетов стоимостью по 1000 рублей, четыре билета по 3000 рублей и два билета по 5000 рублей. Наугад берут три билета.

1) Проведите компьютерное моделирование эксперимента, непосредственно исходя из его описания.

2) Проиллюстрируйте теорему Бернулли, построив график зависимости относительной частоты наблюдения события A = «суммарная стоимость билета более 4000 рублей» от длины серии экспериментов.

Вариант 11. Из колоды в 32 карты берется наугад 9 карт.

1) Проведите компьютерное моделирование эксперимента, непосредственно исходя из его описания.

2) Проиллюстрируйте теорему Бернулли, построив график зависимости относительной частоты наблюдения события A = «среди карт будет не менее 6 одномастных» от длины серии экспериментов.

Вариант 12. 7 яблок, 3 апельсина и 5 лимонов раскладываются случайным образом в три ненумерованных пакета, но так чтобы в каждом было одинаковое количество фруктов.

1) Проведите компьютерное моделирование эксперимента, непосредственно исходя из его описания.

2) Проиллюстрируйте теорему Бернулли, построив график зависимости относительной частоты наблюдения события A = «в каждом пакете оказалось не менее двух яблок» от длины серии экспериментов.

Вариант 13. Проводится эксперимент, состоящий в подбрасывании трех игральных костей и наблюдении цифр, выпавших на верхних гранях.

1) Проведите компьютерное моделирование эксперимента, непосредственно исходя из его описания.

2) Проиллюстрируйте теорему Бернулли, построив график зависимости относительной частоты наблюдения события A = «в сумме на костях выпадет не менее 10 очков» от длины серии экспериментов.

Вариант 14. Имеется 10 карточек, на которых написаны числа 3,3,3,4,4,5,5,6,6,6. Две из этих карточек вынимается одна за другой. Число, написанное на первой карточке, берется за числитель, на второй – знаменатель.

1) Проведите компьютерное моделирование эксперимента, непосредственно исходя из его описания.

2) Проиллюстрируйте теорему Бернулли, построив график зависимости относительной частоты наблюдения события A = «получится правильная дробь» от длины серии экспериментов.