Санкт-Петербургский политехнический университет Петра Великого Институт компьютерных наук и технологий

Высшая школа интеллектуальных систем и суперкомпьютерных технологий

Расчётное задание

Дисциплина: Практикум по теории вероятностей и математической статистике Тема: Применение формулы Байеса

Вариант: ball\_boxes\_arrange

Выполнил студент гр. 3530901/90002

(подпись)

Д. Е. Бакин

Принял доц. каф. КСПТ К. В. Никитин

(подпись)

“ ” 2021 г.

Санкт-Петербург 2021

Содержание

[Описание 3](#_bookmark0)

[Задание 3](#_bookmark1)

[Практическое решение, пункт 1 5](#_bookmark2)

[Практическое решение, пункт 2 7](#_bookmark3)

[Практическое решение, пункт 3 11](#_bookmark4)

[Вывод 15](#_bookmark5)

[Приложение 16](#_bookmark6)

# Описание

N пронумерованных корзин с известным распределением шаров (различным) случайным образом переставляются. Затем игрок подходит к первой корзине, последовательно вынимает d шаров, запоминает их и кладет обратно, потом по аналогии вытаскивает d шаров из 2-ой корзины, смотрит их и возвращает и т.д. в цикле. Такая процедура выполняется несколько раз. Требуется, владея исходной информацией о распределении шаров по корзинам и о том, какие шары вынимались и перекладывались из корзин, вычислить на каждом шаге:

1. Для каждой корзины – вероятность того, что она имеет номер 𝑖,𝑖∈{1,2,...,𝑁}
2. Для набора из всех корзин – вероятность того, что этот набор корзин имеет последовательные номера 𝑖1,𝑖2,… 𝑖𝑁, 𝑖𝑗∈{1,2,...,𝑁},𝑗=1,2,...,𝑁
3. Наиболее правдоподобную комбинацию номеров корзин

# Задание

1a. После каждого k опыта необходимо вычислить ряд распределения апостериорных вероятностей гипотез – о том, какая корзина имеет какой номер. Представить соответствующие результаты визуально на графике в форме изменения с течением опытов диаграмм распределений вероятностей гипотез.

Замечание: в данной задаче количество гипотез равно количеству вариантов переставить корзины, то есть N!

1b. Определять после каждого извлечения, какая гипотеза о порядке корзин имеет наибольшую вероятность. Визуализировать эволюцию изменения наиболее вероятной корзины.

1c. Построить зависимость числа превалирующих гипотез от числа проведенных опытов.

2a. Рассмотреть каждую корзину по отдельности и в качестве гипотез

выдвигать то, какой номер имеет соответствующая корзина. Всего таким образом для одной корзины получится N гипотез. Вычислить для каждой из корзин распределения вероятностей гипотез (о том, какой у нее номер) после каждого опыта. Представить результаты визуально по аналогии с п. 1a.

2b. Определить для каждой корзины наиболее вероятную гипотезу на каждом шаге и визуализировать эволюцию этой гипотезы.

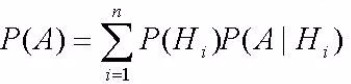
2c. Объединить результаты для всех корзин, получить наиболее вероятную перестановку корзин и сравнить ее с полученной перестановкой в п.1. Провести анализ сравнения.

3a. Определить приближенно частоту вынимания шаров каждого цвета из каждой корзины (получится N экспериментальных профилей). Рассчитать теоретические вероятности вынимания шаров каждого цвета из каждой корзины – получится N теоретических профилей для каждой корзины.

3b. Сопоставить теоретические профили с каждым из полученных экспериментальных и найти их наиболее правдоподобное соответствие. Сравнить с полученным результатом в п.1 и 2. Провести анализ сравнения.

3c. Привести графики изменения экспериментальных профилей для различного количества опытов.

Теоретические основы**:** Формула полной вероятности:



Формула Байеса:

# Практическое решение, пункт 1

Для расчёта вероятностей для разных гипотез была написана программа на языке программирования Kotlin. А для построения графиков использовалась сторонняя библиотека java XChart.

Всего у нас 5 коробок, поэтому количество гипотез равно количеству перестановок 5! = 120. Для решения задачи нахождения исходной перестановки мы циклически пересчитываем вероятности каждой гипотезы по формуле Байеса на каждом шаге, где шаг у нас это 5 извлечений различных шаров из каждой корзины. На каждом шаге мы считаем и сохраняем апостериорные вероятности различных перестановок, после чего эти вероятности используем как априорные, а также по этим сохраненным данным в результате строим графики распределения вероятности по гипотезам с ходом количеств экспериментов.

Условными вероятностями являются вероятности достать последовательно из каждой урны шары из очередного эксперимента.

|  |
| --- |
|  |
| Рис. 1 Распределение гипотез. |

Из рисунка 1 видны необходимые нам данные для пунктов 1a и 1b. Синим цветом у нас изображен график наивероятнейшей перестановки [1 4 2 3 5]. Ее

вероятность резко начинает стремиться к единице с 270 итерации.

Также для пункта 1c можно составить график, количества превалирующих гипотез, вероятность которых, например больше 10%.

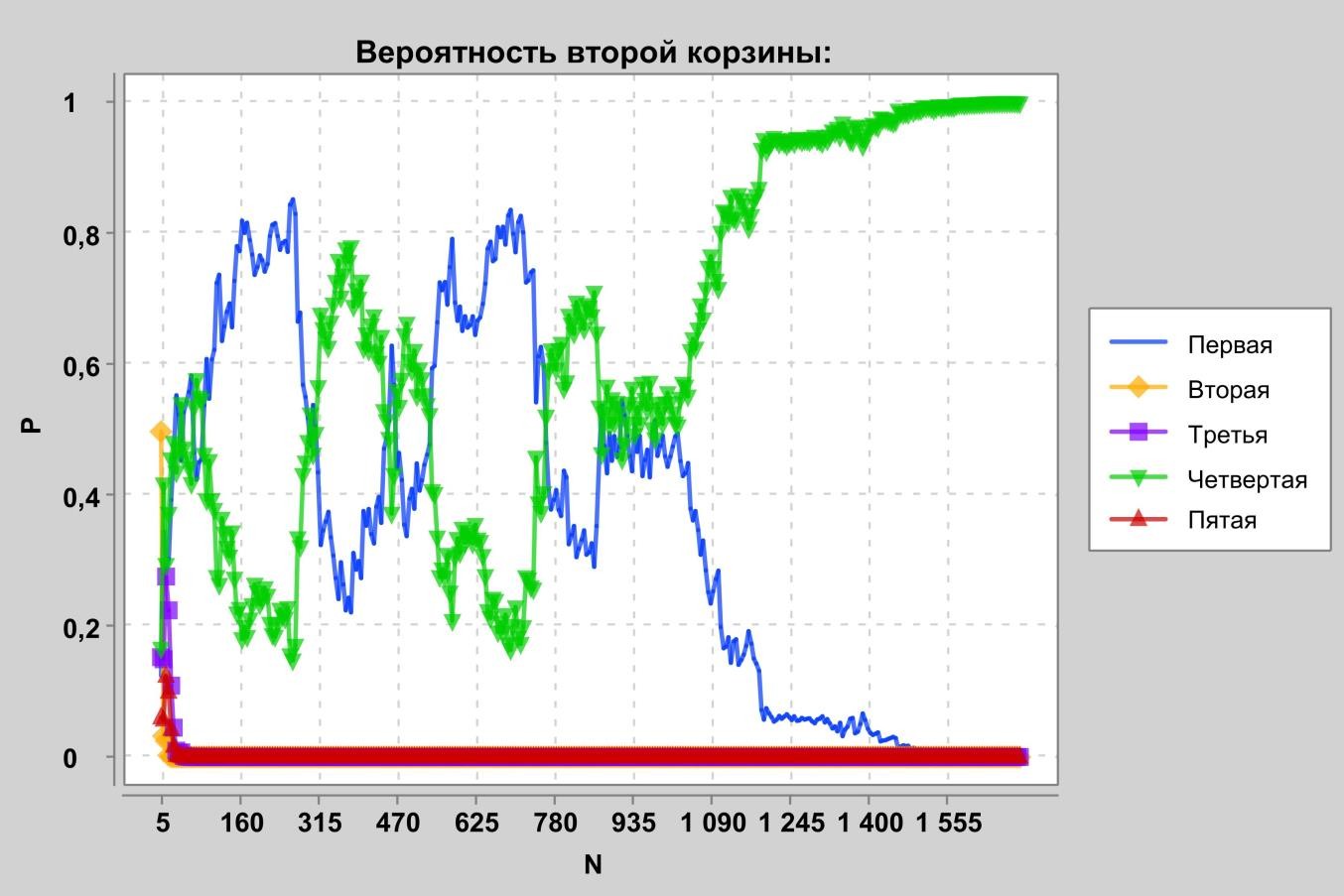
|  |
| --- |
|  |
| Рис. 2 Зависимость числа превалирующих гипотез от числа опытов.  Из рисунка видно, что в самом начале количество гипотез, вероятность которых больше 10%, было равно 4. Примерно с 100 итерации это количество уменьшилось до 2, а с 285 итерации уменьшилось до одного. Эти данные легко  сопоставить с рисунком номер 1, и убедится в их корректности. |

# Практическое решение, пункт 2

Теперь рассмотрим другой случай. Берем отдельную урну с неизвестным нам номером и рассматриваем ее вероятности быть определенной урной. Таким образом количество гипотез будет равно 5. Аналогично циклично пользуемся формулой Байеса, только теперь апостериорная вероятность пересчитывается для одной корзины, а не перестановки. Условные вероятности это вероятности последовательно достать определенные шарики из каждой корзины. Рассмотрим графики вероятностей соответствия каждой корзины исходной позиции.



Рис. 3 Первая корзина.



|  |
| --- |
| Рис. 4 Вторая корзина. |
| Рис. 5 Третья корзина. |

|  |
| --- |
|  |
| Рис. 6 Четвёртая корзина.    Рис. 7 Пятая корзина. |

Эволюция гипотез видна на графиках. Пункт 2a и 2b выполнен.

Пункт 2c:

Для всех корзин, кроме второй, вероятность определенного места выявляется уже с 300 итерации. Однако у второй корзины колебания между первым и четвертым местом были вплоть до 1100 итерации. Но это никак не отразилось на графиках вероятностей гипотез перестановок, т. к. первое место было занято первой корзиной уже с 300 итерацией, вместе со всеми остальными местами, кроме четвертого. Поэтому, смотря на рисунок 1 из первого пункта мы видим, как вероятность определенной последовательности [1 4 2 3 5] стремится к единице уже с 300 итерации. Если рассматривать гипотезу, как соответствие корзине определенного места, то мы аналогично получаем сходимость к последовательности [1 4 2 3 5], но только гораздо позже, примерно начиная с 1200 итерации. Это связано с тем, что корзины ничего не знают, о том какие места уже заняты другими корзинами.

# Практическое решение, пункт 3

Определим теоретические профиля для данных нам корзин.

Табл. 1 - Теоретические профиля.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Коробка номер | Частота доставания Красного  шара | Частота доставая Белого шара | Частота доставания Чёрного  шара | Частота доставания Зелёного  шара | Частота доставания Синего шара |
| 1 | 0.0(36) | 0.03(18) | 0.3(09) | 0.3(27) | 0.3 |
| 2 | 0.256 | 0.244 | 0.22 | 0.06 | 0.22 |
| 3 | 0.2041(6) | 0.091(6) | 0.141(6) | 0.28(3) | 0.2791(6) |
| 4 | 0.035714… | 0.0464285… | 0.264285… | 0.3321428… | 0.321428… |
| 5 | 0.265217… | 0.043478… | 0.121739… | 0.3 | 0.269565… |

После определим частоту вынимания шаров каждого цвета из корзины.

Табл. 2 - Экспериментальные профиля.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Положение Корзины | Частота доставания Красного  шара | Частота доставая Белого  шара | Частота доставания Чёрного  шара | Частота доставания Зелёного  шара | Частота доставания Синего  шара |
| 1..6...11... | 0.0355 | 035(3) | 0.3038(3) | 0.32(3) | 0.302 |
| 2..7...12... | 0.0391(6) | 0.0441(6) | 0.2658(3) | 0.3335 | 0.317(3) |
| 3..8…13... | 0.2485 | 0.245 | 0.2291(6) | 0.0545 | 0.2228(3) |
| 4..9...14... | 0.2138(3) | 0.0958(3) | 0.134(6) | 0.2831(6) | 0.2725 |
| 5..10...15... | 0.2628(3) | 0.044 | 0.119(3) | 0.303(6) | 0.2701(6) |

Напишем функцию, которая найдёт лучшее совпадение профилей (Найдет наименьшую погрешность перестановки).

Мы также получили расстановку [1 4 2 3 5], что опять совпадает со всеми предыдущими пунктами. Пункт 3a и 3b выполнен.

Графики изменения экспериментальных профилей для 10 000 опытов, для пункта 3c представлены ниже:

|  |
| --- |
|  |
| Рис. 8 Изменение профиля для первой корзины. |
|  |
| Рис. 9 Изменение профиля для второй корзины. |

|  |
| --- |
|  |
| Рис. 10 Изменение профиля для третьей корзины. |
|  |
| Рис. 11 Изменение профиля для четвертой корзины. |

|  |
| --- |
|  |
| Рис. 12 Изменение профиля для пятой корзины. |

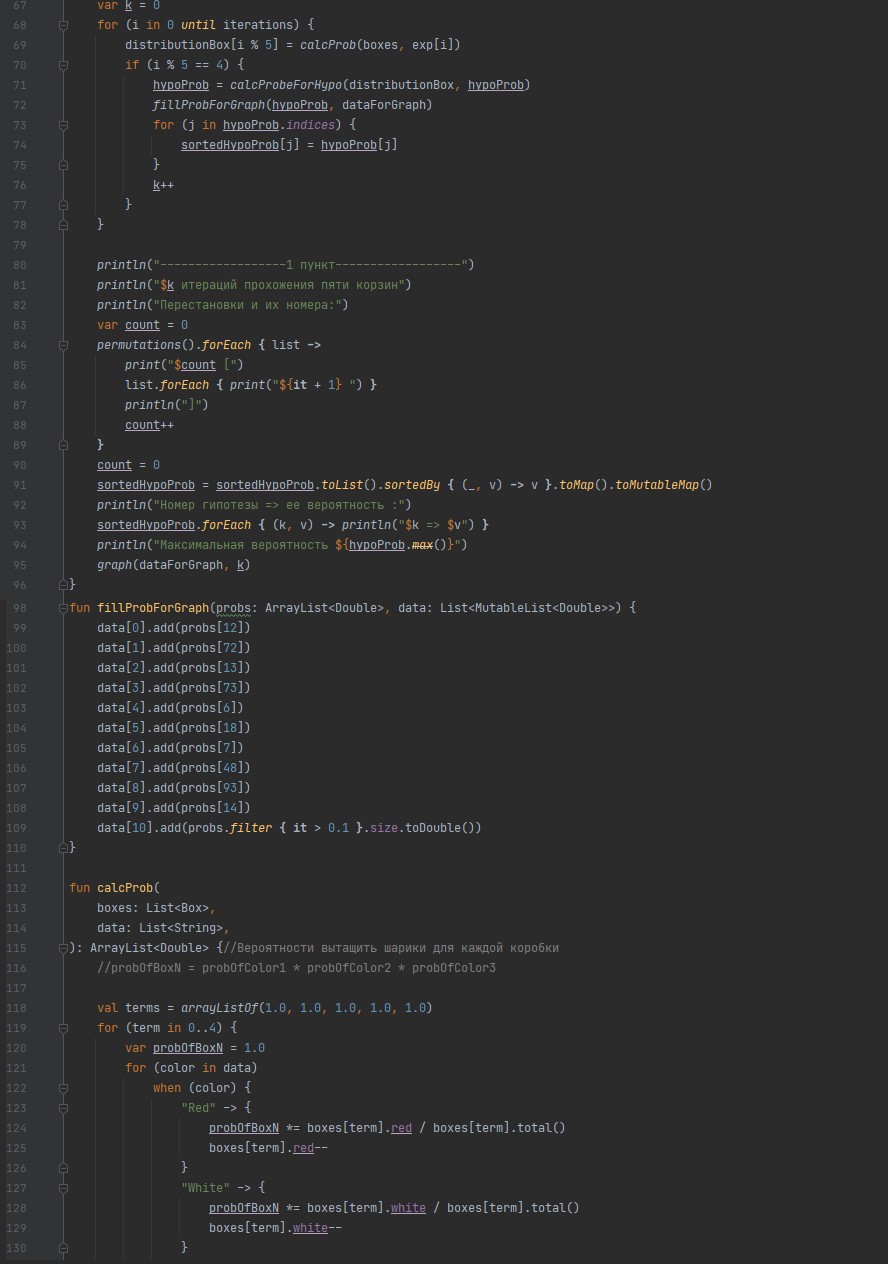
# Вывод

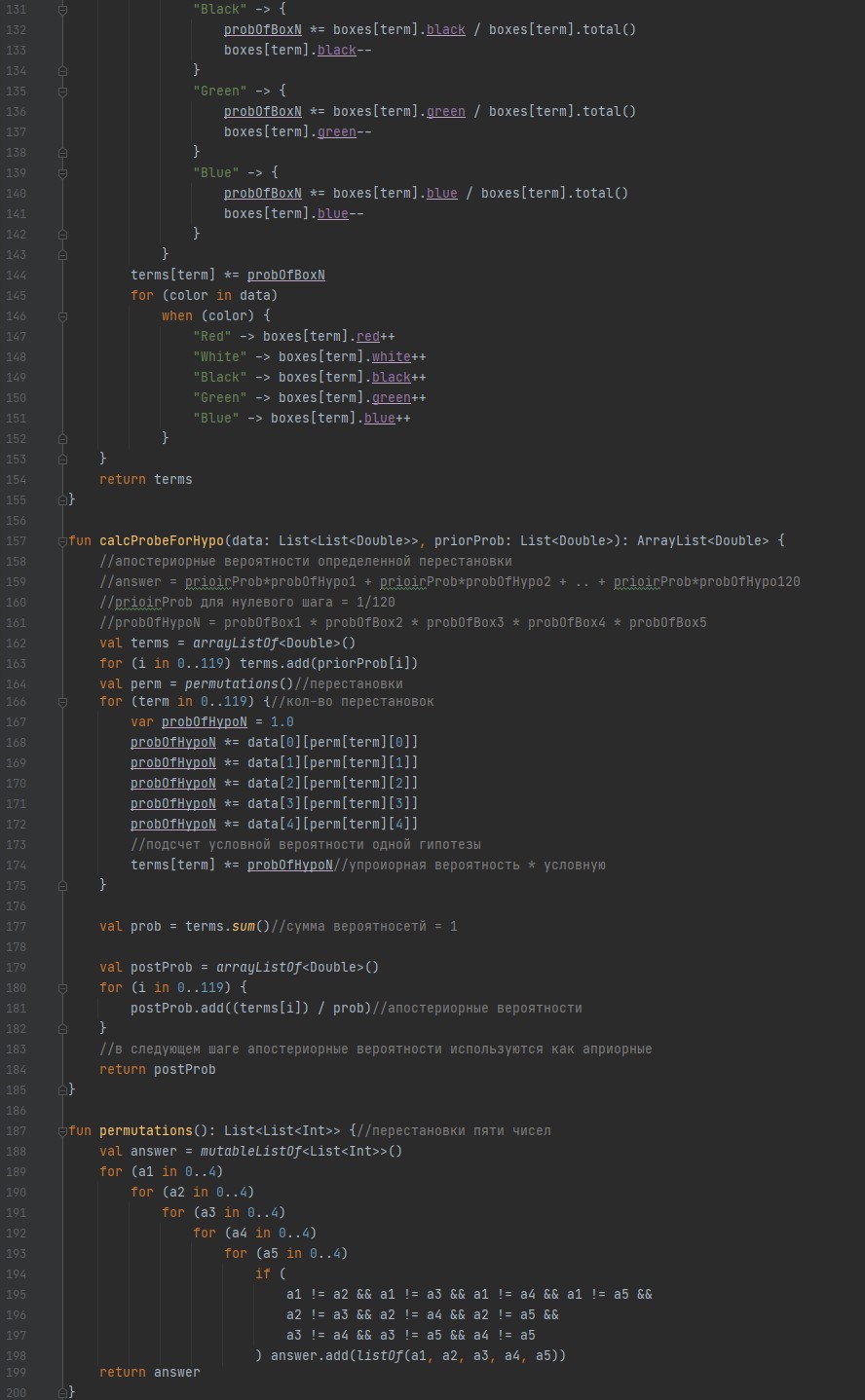
В ходе работы мы активно пользовались формулой Байеса для расчета апостериорных вероятностей. Варианты рассмотрения одной и той же задачи с разных углов показали один и тот же результат. Найденная наивероятнейшая расстановка фигурирует во всех пунктах, что подтверждает правильность решения задачи.

# Приложение

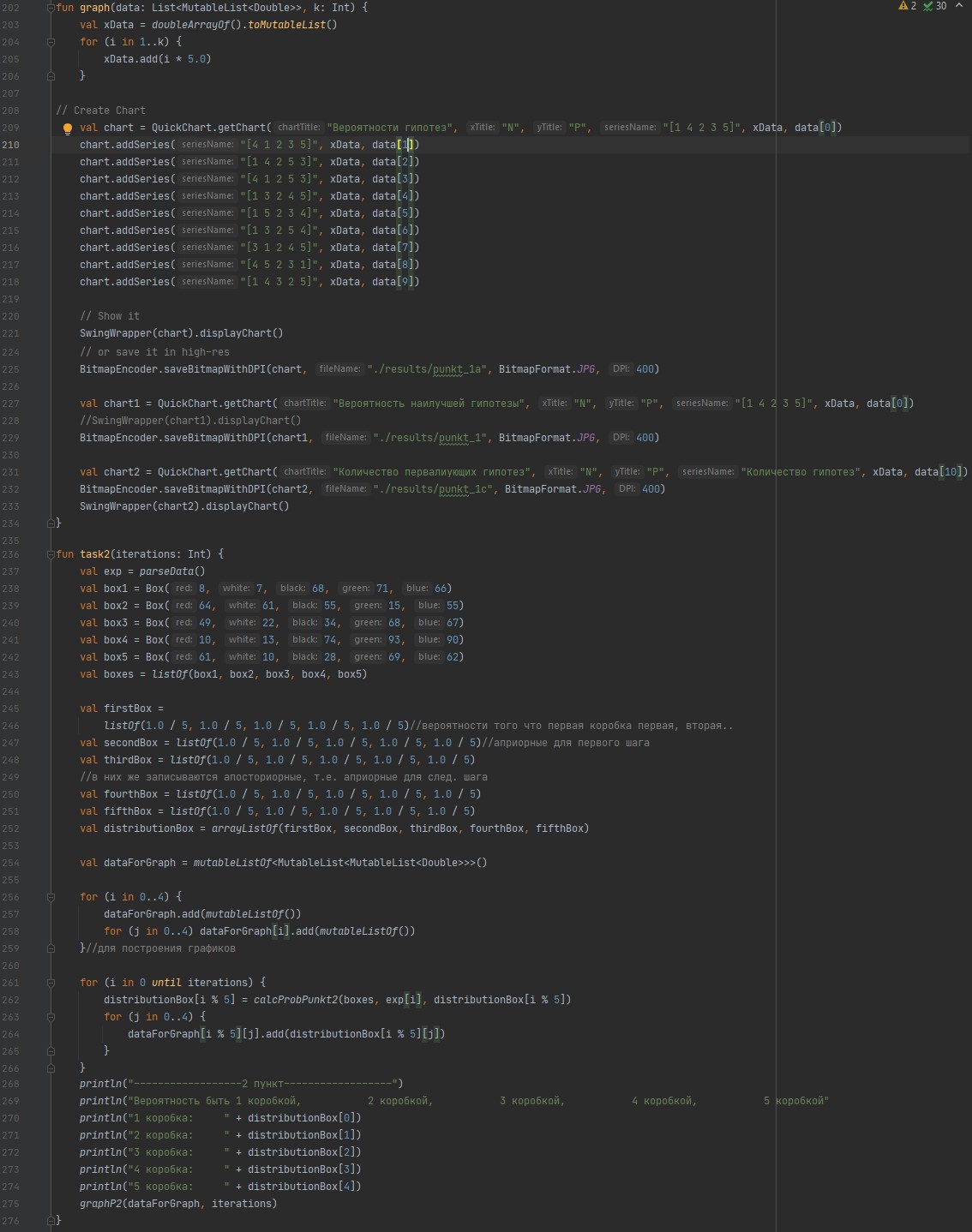
Листинг программы на ЯП Kotlin прилагается.

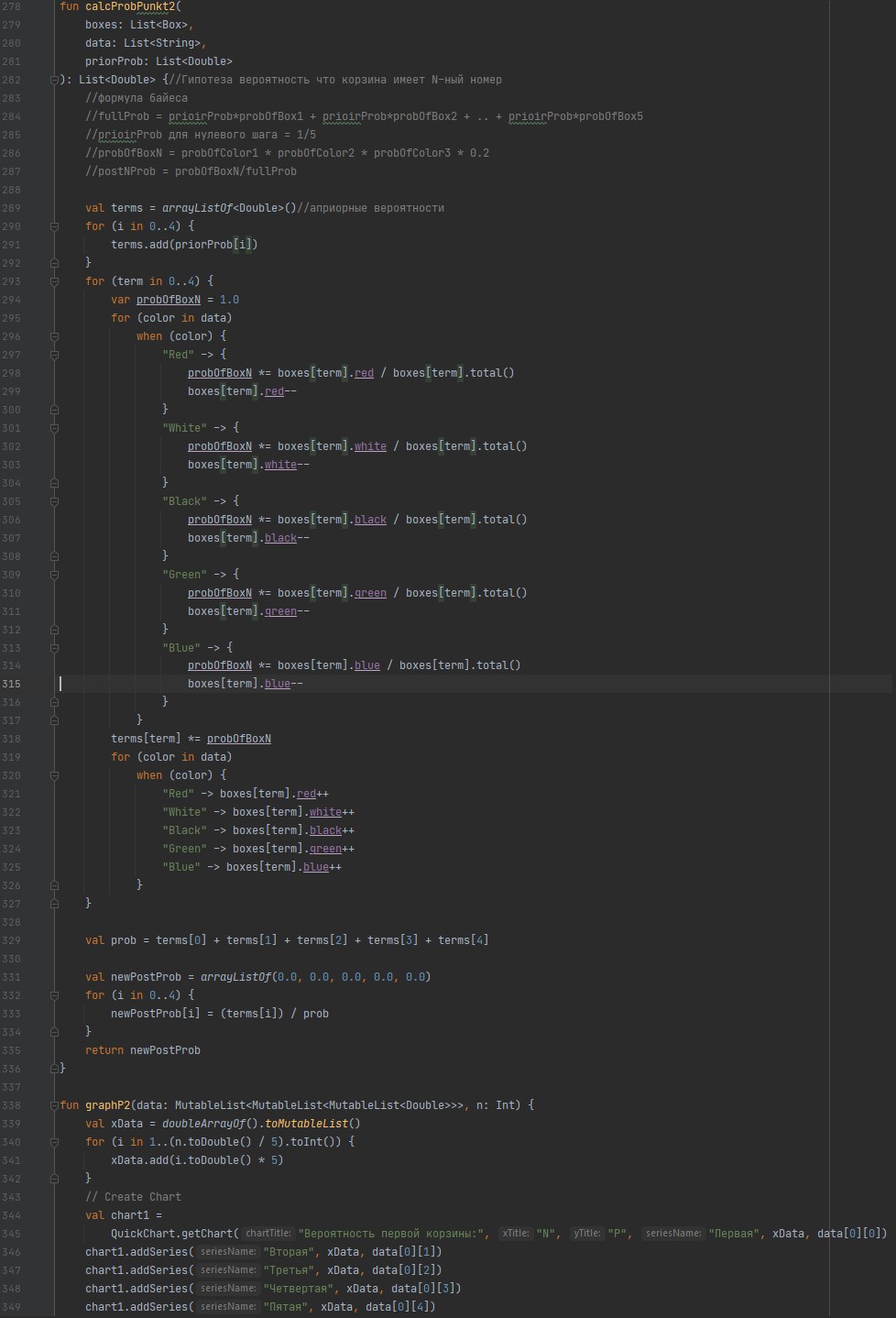


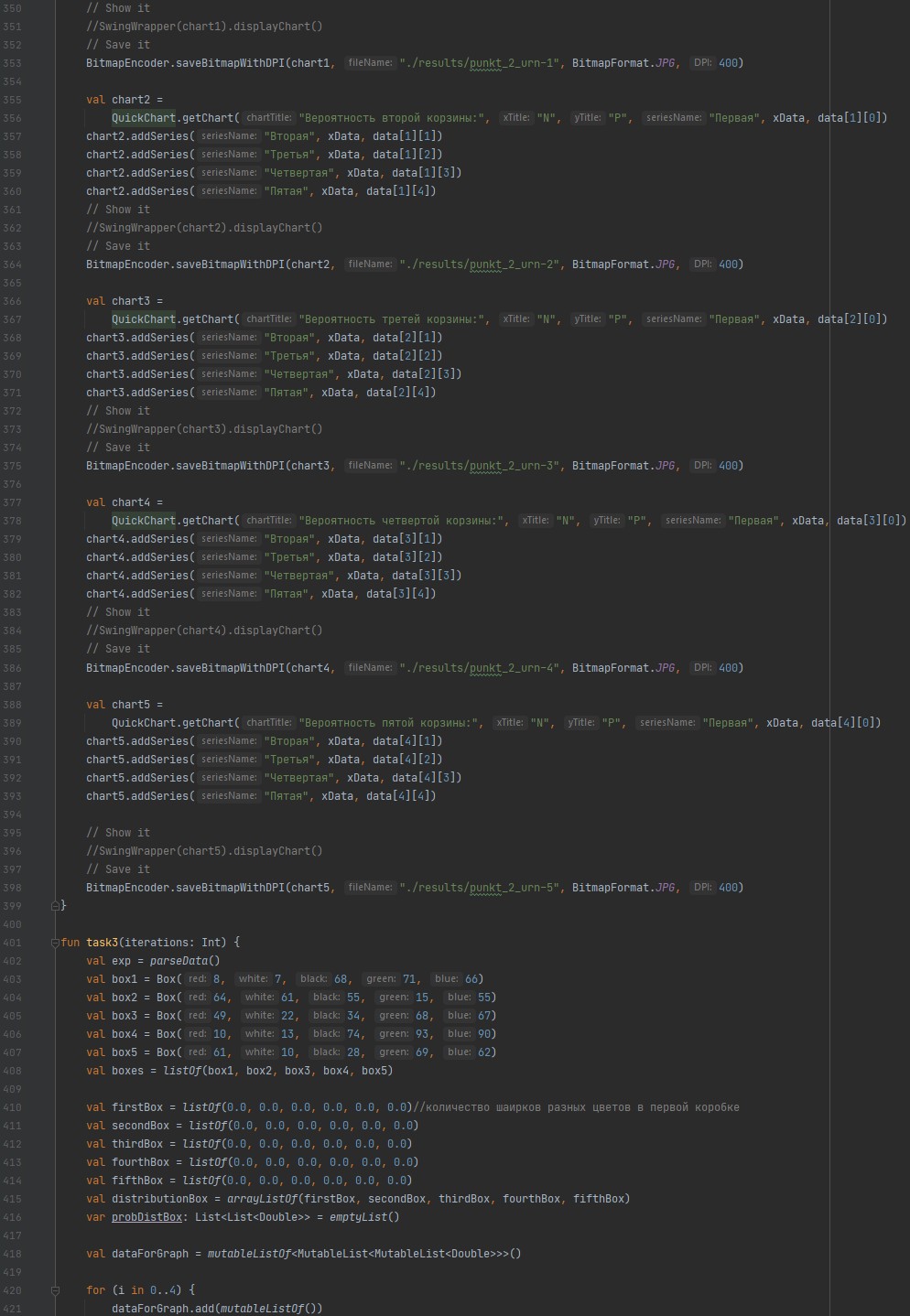


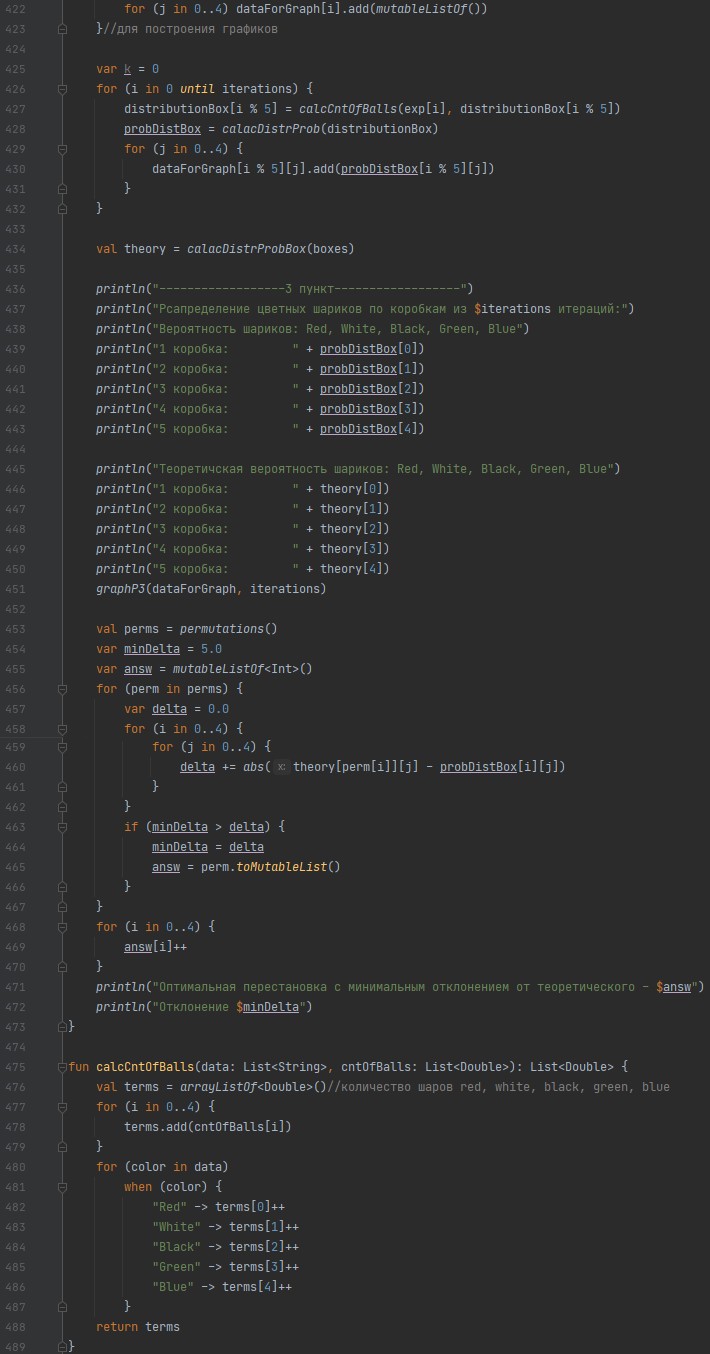


18

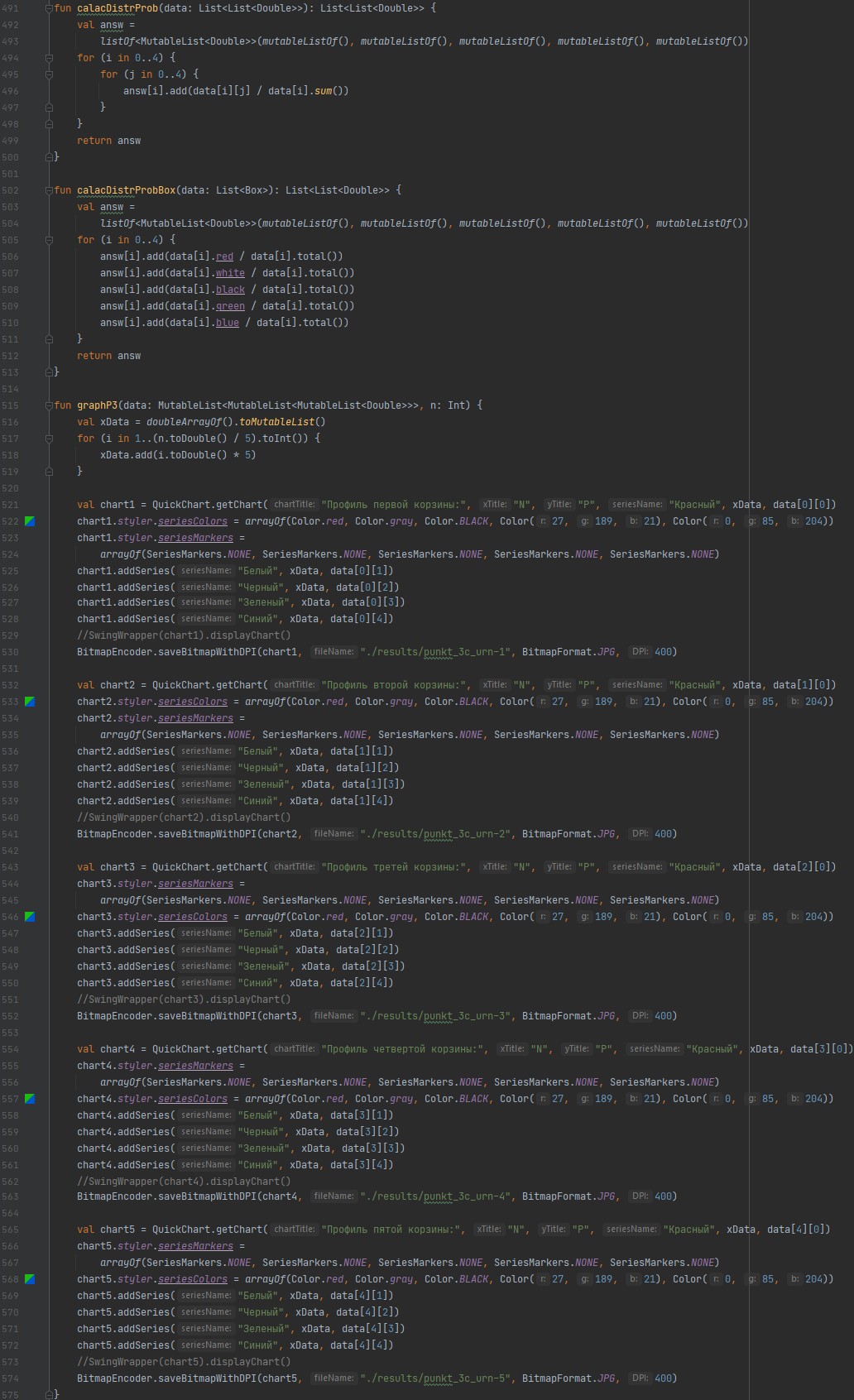








22



23