Санкт-Петербургский политехнический университет Петра Великого

Высшая школа интеллектуальных систем и суперкомпьютерных технологий

**Отчёт по расчетной работе №1**

**Дисциплина**: Теория вероятностей и математическая статистика

**Тема:** Применение формулы Байеса для принятия решений.

Выполнил студент гр. 3530901/10001 Д. Симоновский

Преподаватель К.В.Никитин

Санкт-Петербург

2023

Оглавление

[1. Задание 3](#_Toc135485907)

[1.1. Вариант 3](#_Toc135485908)

[1.2. Описание. 3](#_Toc135485909)

[1.3. Задание 3](#_Toc135485910)

[2. Теоретическая основа 4](#_Toc135485911)

[3. Решение 4](#_Toc135485912)

[4. Анализ результатов 6](#_Toc135485913)

[5. Листинг 7](#_Toc135485914)

# Задание

## Вариант

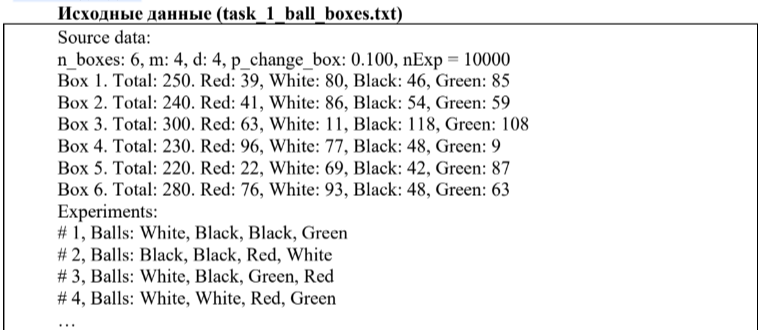
Номер варианта: 1.3.

Вытягивание шаров из случайно выбранной корзины. Угадывание корзины. Набор корзин с шарами. Требуется угадать, где какая корзина путем последовательного извлечения шаров.

## Описание.

Имеется N корзин, в каждой из которой находится известная комбинация шаров различного цвета. Игрок вначале случайно выбирает одну из корзин и затем последовательно K раз вынимает из нее по несколько d=2÷3 шаров с возвращением. Перед каждым подходом игрока к корзине ведущий с заданной вероятностью может случайно заменить ее на какую-то другую.

По истечении опытов игроку необходимо определить, к какой из N корзин он скорее всего подходил.



В примере есть 6 корзин с шарами 4-х цветов. В первой урне 250 шаров, 39 красных,80 белых, 46 черных и 85 зеленых и т.д. Извлекается в каждый опыт по d=4 шара. Вероятность замены искомой корзины равна 0.1. Всего 10000 опытов. В первом опыте вытягиваются белый, черный, черный, зеленый шары и т.д

## Задание

**1а.** После каждого k опыта необходимо вычислить ряд распределения апостериорных вероятностей гипотез – к какой из корзин подошел игрок. Представить соответствующие результаты визуально на графике в форме изменения с течением опытов диаграммы распределений вероятностей гипотез.

**1b.** Определять после каждого извлечения, какие корзины имеют наибольшую вероятность. Визуализировать эволюцию изменения наиболее вероятной корзины.

**1с.** Построить зависимость числа превалирующих гипотез от числа проведенных опытов.

**2a.** Определить приближенно частоту вынимания шаров каждого цвета из корзины (экспериментальный профиль). Рассчитать теоретические вероятности вынимания шаров каждого цвета из каждой корзины–получится N теоретических профилей для каждой корзины.

**2b.** Сравнить теоретический Профиль с каждым из полученных экспериментальными найти наиболее похожий. Сравнить с полученным результатов в п. 1

**2c.** Построить графики изменения экспериментальных профилей для различного количества опытов

# Теоретическая основа

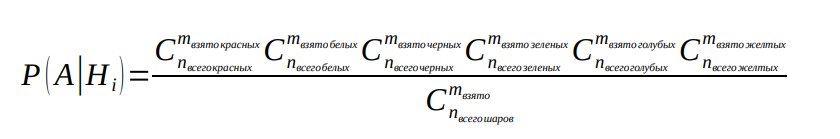
Формула Байеса имеет следующий вид:



Имеется N коробок, следовательно, гипотезы будут иметь следующий вид:

H\_i – игрок подошел к i ой коробке 1/ 𝑁.

Тогда, чтобы рассчитать условные вероятности, что шар вынут из этой корзины будут иметь следующий вид:



В цикле мы находим корзину с наиболее высокой условной вероятностью, а по окончании итераций можно определить наиболее вероятную корзину, к которой мог подойти игрок.

# Решение

Для решения задачи был использован код на языке Matlab с отсечкой превалирующих гипотез на уровне 0.05. Результаты анализа требуют дальнейшего изучения.

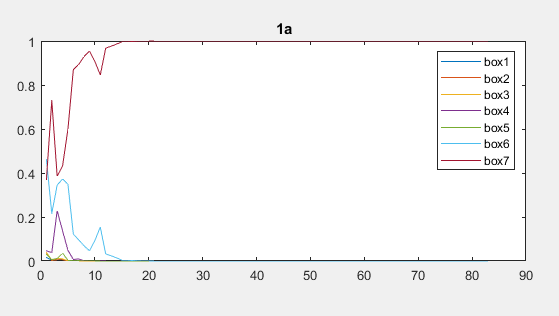


Рис. 1. Ряд распределения апостериорных вероятностей гипотез с течением опытов

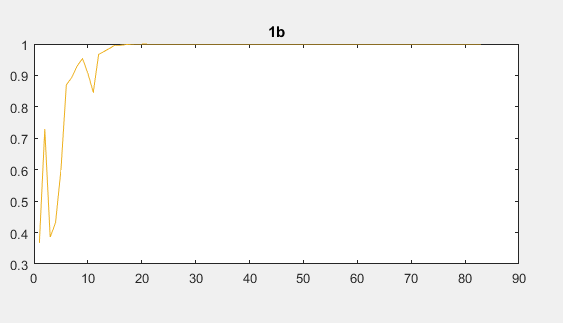


Рис. 2. Эволюция изменения наиболее вероятной корзины.

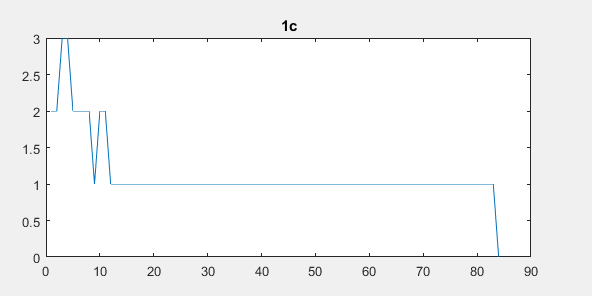


Рис. 3. Зависимость числа превалирующих гипотез от числа проведенных опытов.

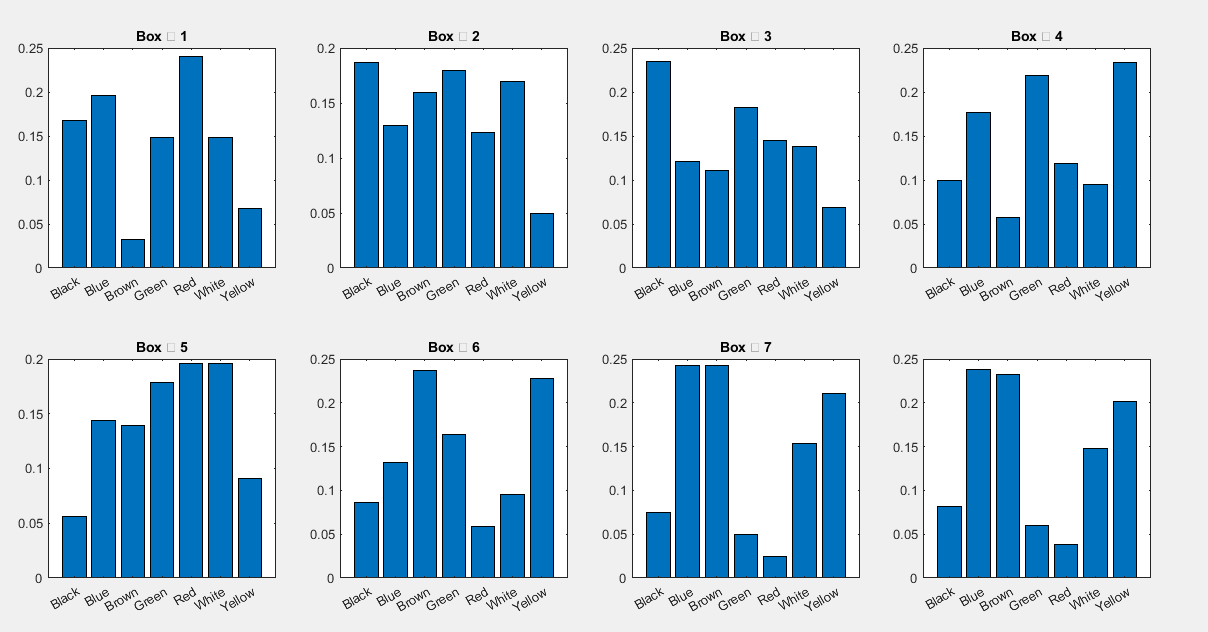


Рис. 4. Экспериментальные профили.

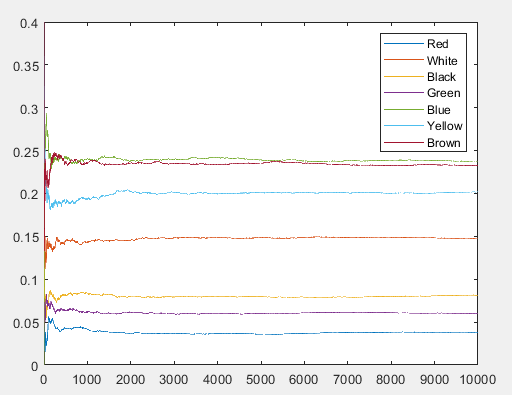


Рис. 5. Изменение экспериментальных профилей.

# Анализ результатов

По рисунку 1 видно, что наиболее вероятной оказалась корзина под номером 7.

Рисунок 2 представляет ту же информацию, что и рисунок 1, но с изменением выбранной корзины, к которой с наибольшей вероятностью подошел человек. По результатам более чем 30 экспериментов можно с уверенностью близкой к 100% утверждать, что искомая корзина имеет номер 7.

Из рисунка 3 видно, что количество превалирующих гипотез уменьшалось с увеличением числа проведенных опытов.

Последний график на рис. 4 показывает профиль не конкретной корзины, а изменение вероятности выбора каждой корзины по ходу экспериментов. Видно, что профиль наиболее совпадает с корзиной №7.

Рисунок 5 показывает, что после 2000 экспериментов экспериментальный профиль уже окончательно сформирован. Также видно, что практические вероятности вынимания шаров различных цветов совпадают с теоретическими.

В результате проведения всех опытов мы установили, что наиболее вероятной корзиной является 7-ая, что также подтверждается эволюцией вероятностной корзины. Мы использовали формулу Байеса для определения корзины, к которой подошел человек, и практические вычисления подтверждают ее корректность.

# Листинг

threshold = 0.05;

number\_boxes = 7;

m\_default = 7;

d = 5;

change\_probability = 0.1;

not\_change\_probability = 1-change\_probability;

PAHi = ones(1, number\_boxes);

fid = fopen('Simonovsky1Balls.txt');

i = 1;

b = 1;

balls = zeros(2, number\_boxes);

while ~feof(fid)

balls(1, :) = zeros(1, number\_boxes);

tline = fgetl(fid);

if regexp(string(tline), '(\w+(: | = )\d+(.\d+)?(, )?)+', 'match') == string(tline)

fblocks = regexp(string(tline), '[A-Za-z\_: ,=]+', 'split');

param = fblocks(1, 2:6);

expProfiles = zeros(str2double(param(1, 5)), str2double(param(1, 2)));

elseif regexp(string(tline), '\w+ \d+. \w+: \d+.([\w+ :,]+)', 'match') == string(tline)

temp(1, :) = string(regexp(tline, '[A-Za-z\_: ,=.]+', 'split'));

for c = 1:str2double(param(1, 2))

boxes(b, c) = str2double(temp(1, 3 + c));

end

b = b + 1;

elseif regexp(string(tline), '#', 'start') == 1

str(i, 1) = string(tline);

fblocks = regexp(string(regexp(tline, '([a-zA-Z, ]+)$', 'match')),',','split');

for j = 1:d

switch strtrim(fblocks(1, j))

case "Red"

balls(2, 1) = balls(2, 1) + 1;

balls(1, 1) = balls(1, 1) + 1;

case "White"

balls(2, 2) = balls(2, 2) + 1;

balls(1, 2) = balls(1, 2) + 1;

case "Black"

balls(2, 3) = balls(2, 3) + 1;

balls(1, 3) = balls(1, 3) + 1;

case "Green"

balls(2, 4) = balls(2, 4) + 1;

balls(1, 4) = balls(1, 4) + 1;

case "Blue"

balls(2, 5) = balls(2, 5) + 1;

balls(1, 5) = balls(1, 5) + 1;

case "Yellow"

balls(2, 6) = balls(2, 6) + 1;

balls(1, 6) = balls(1, 6) + 1;

case "Brown"

balls(2, 7) = balls(2, 7) + 1;

balls(1, 7) = balls(1, 7) + 1;

end

end

for j = 1:m\_default

expProfiles(i, j) = balls(2, j) / (i \* d);

end

for j = 1:number\_boxes

m = getPAHi(boxes(j, :), balls(1, :), d ) \* not\_change\_probability;

for k = 1:number\_boxes

if (k ~= j)

m = m + getPAHi(boxes(k, :), balls(1, :), d) \* change\_probability / number\_boxes;

end

end

PAHi(j) = PAHi(j) \* m;

end

prevalCount = 0;

for j = 1:number\_boxes

results(i, j) = PAHi(j) / sum(PAHi);

if results(i, j) > threshold

prevalCount = prevalCount + 1;

end

end

if ~isnan(results(i, 1))

results(i, number\_boxes + 1) = prevalCount;

end

i = i + 1;

else

end

end

% 1a

nexttile

plot(results(:, 1:number\_boxes));

legend("box1","box2","box3","box4","box5", "box6", "box7");

title('1a')

nexttile;

% 1b

[~,c]=find(results(:, 1:number\_boxes) == max(max(results(:, 1:number\_boxes))));

plot(results(:, c));

title('1b')

nexttile;

% 1c

plot(results(1:find(isnan(results)), number\_boxes + 1));

title('1c')

pause;

clf

% 2a

colorsnames = categorical({'Red' 'White' 'Black' 'Green' 'Blue', 'Yellow', 'Brown'});

formatTitle = 'Box № %d';

for box = 1:str2double(param(1, 1))

nexttile;

for color = 1:str2double(param(1, 2))

profile(box, color) = boxes(box, color) / sum(boxes(box, :));

end

bar(colorsnames, profile(box, :));

title (sprintf(formatTitle, box));

end

nexttile;

bar(colorsnames, expProfiles(str2double(param(1, 5)), :));

pause;

clf

% 2c

plot(expProfiles);

legend(colorsnames);

function res = getPAHi(boxes, balls, d)

res = power(boxes(1, 1), balls(1, 1)) \* power(boxes(1, 2), balls(1, 2)) \* power(boxes(1, 3), balls(1, 3)) \* power(boxes(1, 4), balls(1, 4)) \* power(boxes(1, 5), balls(1, 5)) \* power(boxes(1,6),balls(1,6)) \* power(boxes(1,7),balls(1,7))/ power(sum(boxes), d);

end