Санкт-Петербургский политехнический университет Петра Великого

Высшая школа интеллектуальных систем и суперкомпьютерных технологий

**Отчёт по расчётному заданию № 1**

**Дисциплина**: Теория вероятностей и математическая статистика

**Тема**: Применение формулы Байеса

Выполнили студенты гр. 3530901/90005 И.Д.Галактионов

(подпись)

Преподаватель \_\_\_\_\_\_\_\_\_\_ К.В. Никитин

(подпись)

“\_\_\_” \_\_\_\_\_\_\_\_\_\_\_\_\_ 2021 г.

Санкт-Петербург

2021

**Содержание решенных заданий**

1. *Ball Boxes*

Updates: добавил в конце листинг программы

**Ball Boxes**

1. **Задание**

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 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 корзин он скорее всего подходил.

Практические задания

1.a. 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 диаграмм распределений вероятностей гипотез. В данной задаче всего N гипотез.

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

1.с. 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.

2.b. Сравнить теоретический профиль с каждым из полученных экспериментальным и найти наиболее похожий. Сравнить с полученным результатов в п. 1

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

Вариант

Source data:  
n\_boxes: 6, m: 4, d: 4, p\_change\_box: 0.100, nExp = 10000  
Box 1. Total: 250. Red: 68, White: 41, Black: 106, Green: 35  
Box 2. Total: 300. Red: 97, White: 69, Black: 30, Green: 104  
Box 3. Total: 200. Red: 85, White: 58, Black: 5, Green: 52  
Box 4. Total: 270. Red: 76, White: 43, Black: 56, Green: 95  
Box 5. Total: 200. Red: 12, White: 83, Black: 83, Green: 22  
Box 6. Total: 200. Red: 23, White: 15, Black: 80, Green: 82

1. **Теоретическая основа**Формула Байеса:

.

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

Hi – игрок подошел к iой коробке

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

В итоге на каждой итерации мы находим корзину с наиболее высокой условной вероятностью. Когда итерации закончатся, мы можем узнать наиболее вероятность корзину, к которой мог подходить игрок, она будет с наибольшим счетчиком.

1. **Практическая реализация**

По вышеприведённым расчётным формулам нашли вероятности гипотез, возьмём в качестве отсечки превалирующих формул отметку 0.05. Код программы на языке MATLAB, выполняющий расчёты по приведённым выше формулам для выполнения заданий 1а-1с представлен ниже

По результатам выполнения получили графики, представленные рис. 1

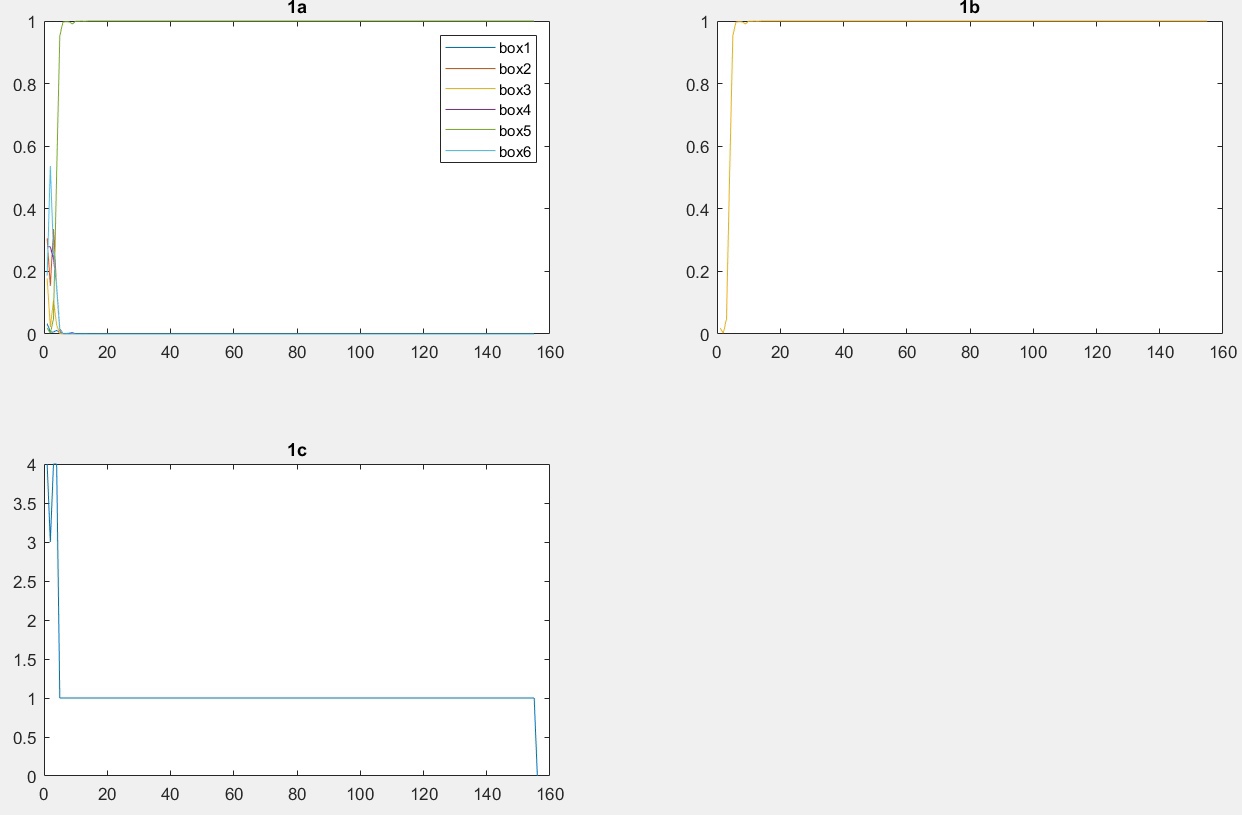


Рис. 1 Графики гипотез, превалирующей гипотезы и числа превалирующих гипотез

Из полученных результатов следует, что выемка шаров производилась из пятой корзины.

# Часть 2 (2а-2с)

## Теоретическая часть

В этой части в основном применяется формула вычисления вероятности выемки шара определённого цвета:

## Практическая часть

Программа на MATLAB выполняет построение гистограмм экспериментальных профилей для каждой из корзин.

Полученные в итоге гистограммы представлены рис. 2

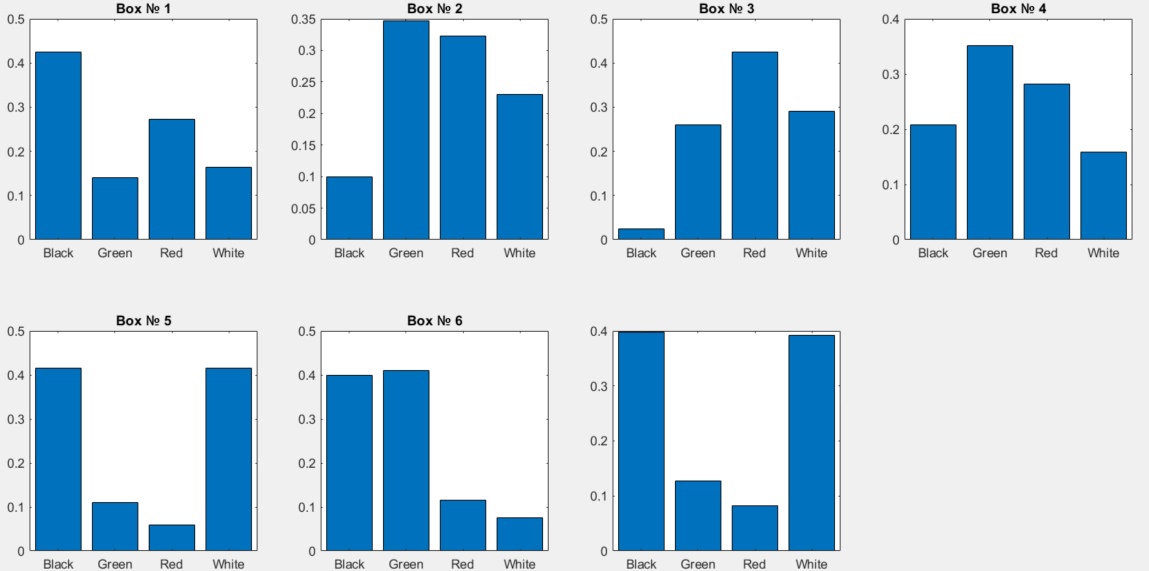


Рис. 2 Экспериментальные профили

Исходя из гистограмм экспериментальных профилей, наиболее схожий экспериментальны профиль с профилем 5 корзины.

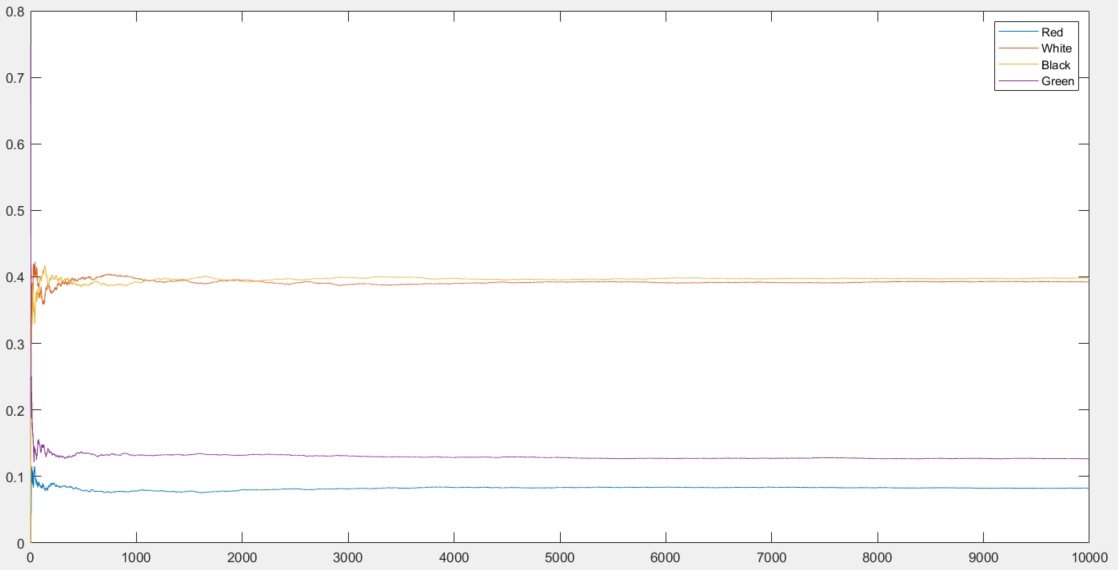


Рис. 3 График изменения экспериментальных профилей

Из графика следует, что уже после 1000 экспериментов окончательно сформирован экспериментальный профиль.

# Вывод

Исходя из результатов моделирования, выполненных как в первой части работы по формуле Байеса так и во второй части исследования по теоретическим и экспериментальным профилям, можно сделать вывод, что корзина, из которой проводилась выемка имеет номер 4.

ЛИСТИНГ

%порог превалирующих значений

threshold = 0.05;

PAHi = ones(1, 6);

fid = fopen('task\_1\_ball\_boxes.txt');

i = 1;

b = 1;

balls = zeros(2, 4);

%"Red" "White" "Black" "Green"

while ~feof(fid)

% парсинг

balls(1, :) = zeros(1, 4);

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)));

% balls = zeros(2, 4);

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 = 1:4

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;

end

end

% экспериментальные профили к 2с

for j = 1:4

expProfiles(i, j) = balls(2, j) / (i \* 4);

end

for j = 1:6

m = getPAHi(boxes(j, :), balls(1, :)) \* 0.9;

for k = 1:6

if (k ~= j)

m = m + getPAHi(boxes(k, :), balls(1, :)) \* 0.1 / 6;

end

end

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

end

prevalCount = 0;

for j = 1:6

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

if results(i, j) > threshold

prevalCount = prevalCount + 1;

end

end

if ~isnan(results(i, 1))

results(i, 7) = prevalCount;

end

i = i + 1;

else

end

end

nexttile

plot(results(:, 1:6));

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

title('1a')

nexttile

%1b

[~,c]=find(results(:, 1:6) == max(max(results(:, 1:6))));

plot(results(:, c));

title('1b')

nexttile;

%1c

plot(results(1:find(isnan(results)), 7));

title('1c')

pause;

clf

%2a

colorsnames = categorical({'Red' 'White' 'Black' 'Green'});

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);

%"Red" "White" "Black" "Green"

function res = getPAHi(boxes, balls)

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(sum(boxes), 4);

end