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. 5130901/20003 А.А Вагнер

(подпись)

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

(подпись)

“\_\_\_” \_\_\_\_\_\_\_\_\_\_\_\_\_ 2024 г.

Санкт-Петербург

2024

**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:

n\_boxes: 7, m: 5, d: 5, p\_change\_box: 0.100, nExp = 10000

Box 1. Total: 210. Red: 21, White: 27, Black: 25, Green: 69, Blue: 68

Box 2. Total: 240. Red: 74, White: 19, Black: 56, Green: 68, Blue: 23

Box 3. Total: 210. Red: 67, White: 47, Black: 27, Green: 58, Blue: 11

Box 4. Total: 250. Red: 52, White: 79, Black: 82, Green: 20, Blue: 17

Box 5. Total: 280. Red: 42, White: 84, Black: 40, Green: 36, Blue: 78

Box 6. Total: 280. Red: 53, White: 5, Black: 84, Green: 50, Blue: 88

Box 7. Total: 250. Red: 4, White: 77, Black: 31, Green: 62, Blue: 76

1. **Аналитика**

Для решения этой задачи потребуется применить формулу Байеса.

Гипотезой в данной задаче выберем факт выбора коробки под номером i. Тогда вероятность для каждой из гипотез , при условии, что у игрока нет предрасположенности к определённой коробке, будет равна . Определим вероятность того, что шар вынут из данной корзины i. Введём обозначения: вс – всего, вз – взято, к – красных, б – белых, ч – чёрных, з – зелёных, с – синих.

Для решения пунктов 1a, 1b требуется при каждом прочтении искать по вышенаписанной формуле корзину, к которой с наибольшей вероятностью относится текущий набор шаров и записывать в соответствующую ячейку результат.

1. **Практика**

Алгоритм решения этой задачи был реализован на языке Matlab, а результаты представлены в виде графиков

Изображение выглядит как текст, диаграмма, линия, снимок экрана

Автоматически созданное описание

Рис.1 Графики превалирующих гипотез

Из первого графика видно, что искомая корзина – под номером 7

Также сравним экспериментальные профили каждой из корзин с профилем выемки шаром и найдём наиболее схожие

Изображение выглядит как текст, диаграмма, снимок экрана, Шрифт

Автоматически созданное описание

Рис.2 Профили корзин

Очевидно, профиль 7 корзины предельно схож с экспериментальным.

Также следует оценить, как экспериментальные профили менялись по ходу эксперимента. Построим график соответствующей зависимости профиля от количества произведённых опытов.

Изображение выглядит как текст, линия, График, снимок экрана

Автоматически созданное описание

Рис.3 График изменения профилей

Из графика следует, что значительные колебания профилей прекратились после 2500 эксперимента

**Вывод**

Исходя из полученных графиков и гистограмм, можно точно определить, что выемка происходила из седьмой корзины, для определения этого результата применялась формула Байеса.

**Приложение**

threshold = 0.05;

PAHi = ones(1, 7);

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

i = 1;

b = 1;

balls = zeros(2, 5);

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

while ~feof(fid)

% парсинг

balls(1, :) = zeros(1, 5);

tline = fgetl(fid);

% n\_boxes: 7, m: 5, d: 5, p\_change\_box: 0.100, nExp = 10000

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)));

% Box 1. Total: 210. Red: 21, White: 27, Black: 25, Green: 69, Blue: 68

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 = 1:5

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;

end

end

% экспериментальные профили к 2с

for j = 1:5

expProfiles(i, j) = balls(2, j) / (i \* 5);

end

for j = 1:7

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

for k = 1:7

if (k ~= j)

m = m + getPAHi(boxes(k, :), balls(1, :)) \* 0.1 / 7;

end

end

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

end

prevalCount = 0;

for j = 1:7

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

if results(i, j) > threshold

prevalCount = prevalCount + 1;

end

end

if ~isnan(results(i, 1))

results(i, 8) = prevalCount;

end

i = i + 1;

else

end

end

nexttile

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

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

title('1a')

nexttile

%1b

[~,c]=find(results(:, 1:7) == max(max(results(:, 1:7))));

plot(results(:, c));

title('1b')

nexttile;

%1c

plot(results(1:find(isnan(results)), 8));

title('1c')

%pause;

clf

%2a

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

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)

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(sum(boxes), 5);

end