

# Probability & Statistics for EECS: Homework #01

Due on Feb 29, 2024 at 23:59

Name: **Fei Pang**  
Student ID: 2022533153

## Problem 1

- (1) Consider we are forming  $k$  groups of  $n + 1$  people and I am among the  $n + 1$  people. There are two possibilities: the other  $n$  people form  $k - 1$  groups and I am in a group by myself, or the other  $n$  people form  $k$  groups already, so I only have to join one of the  $k$  groups.
- (2) Consider we are forming  $k + 1$  groups of  $n + 1$  people and I am among the  $n + 1$  people. There are possibilities that 1, 2, ...,  $k$  people are not in my group. So we can divide it into  $k$  circumstances and the  $j$  people (not in my group) form  $k$  groups and I lead a group myself.

## Problem 2

There are  $P_{26}^{26}$  *norepeatwords* with 26 alphabets, and  $P_{26}^1 + P_{26}^2 + \dots$  *norepeatwords*.

$$P = \frac{P_{26}^{26}}{P_{26}^1 + P_{26}^2 + \dots + P_{26}^{26}}$$

$$\frac{1}{P} = 1 + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \dots + \frac{1}{25!} = \sum_{n=1}^{25} \frac{1}{n!} \approx \sum_{n=1}^{\infty} \frac{1}{n!} = e$$

$$\text{So } P \approx \frac{1}{e}.$$

## Problem 3

- (a) Apparently, we have  $n^n$  possible bootstrap samples.
- (b) Assume  $(a_1, a_2, \dots, a_n)$  and  $(b_1, b_2, b_3, \dots, b_n)$ .  
Each  $b_n$  presents the number of times that each  $a_n$  is chosen. So  $\sum_{i=1}^n b_i = n$ .  
Then it turns out to be a partition problem: We should put  $n - 1$  partitions to separate the numbers into  $k$  different boxes. So, there are  $2n - 1$  places to be inserted, which is a  $\binom{2n-1}{n-1}$ . So there are  $\frac{(2n-1)!}{n!(n-1)!}$  kinds.
- (c) Assume the bootstrap samples have orders so there are  $n^n$  kinds.  
 $b1$ : only one kind:  $a_1, a_1, a_1, \dots, a_1$ .  
 $b2$ : each number appears once:  $a_1, a_2, a_3, \dots, a_n$  is  $n!$ .  
So  $p_1 = \frac{n!}{n^n}$ ,  $p_2 = \frac{1}{n^n}$ .  $\frac{p_1}{p_2} = n!$   
Probability of getting  $p_1$ :  $\frac{n!}{n^n}$ .  
probability of getting  $p_2$ :  $n \times \frac{1}{n^n}$ .  
The ratio is  $(n - 1)!$

## Problem 4

Assume the length of the stick is  $k$  and the three pieces are  $a$ ,  $b$ , and  $k - a - b$ . To form a triangle, we have to satisfy:

$$\begin{aligned} a + b &> k - a - b \\ a + k - a - b &> b \\ b + k - a - b &> a \end{aligned}$$

The fig below gives a way to calculate the possibility, which is the proportion of the area of the shadow.

$$\text{So } P = \frac{1}{4}.$$

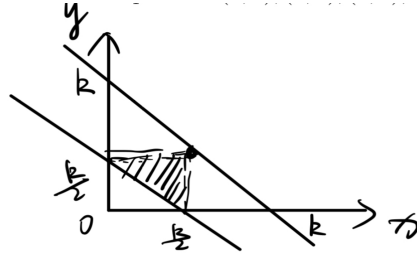


Figure 1: LP

## Problem 5

- (a) There are  $k$  people, then the probability of at least one holiday match is

$$1 - k!e_k(p_1, p_2, p_3, \dots, p_{365}) = 1 - k!e_k(\vec{p}).$$

- (b) Consider  $k = 2$ ,  $P = 1 - 2e_2(\vec{p}) = \sum_{j=1}^{365} p_j^2 \geq \frac{(p_1 + p_2 + \dots + p_{365})^2}{365}$   
We know that it is only equal when  $p_1 = p_2 = \dots = p_{365} = \frac{1}{365}$ .

- (c) 1: A's birthday is on 1 day, B's birthday is on another day, others' don't match and are not on day1 or day2  $\Rightarrow x_1 x_2 e_{k-2}(x_3, \dots, x_n)$

2: A's birthday and B's birthday are on one day, others' on another day  $\Rightarrow (x_1 + x_2)e_{k-1}(x_3, \dots, x_n)$

3: All people's birthday don't match.

So we have  $e_k(x_1, \dots, x_n) = x_1 x_2 e_{k-2}(x_3, \dots, x_n) + (x_1 + x_2)e_{k-1}(x_3, \dots, x_n) + e_k(x_3, \dots, x_n)$ .

$$e_k(\vec{p}) = p_1 p_2 e_{k-2}(p_3, \dots, p_n) + (p_1 + p_2)e_{k-1}(p_3, \dots, p_n) + e_k(p_3, \dots, p_n)$$

$$e_k(\vec{r}) = \frac{(p_1 + p_2)^2}{4} e_{k-2}(p_3, \dots, p_n) + (p_1 + p_2)e_{k-1}(p_3, \dots, p_n) + e_k(p_3, \dots, p_n)$$

refer to  $\frac{x+y}{2} \geq \sqrt{xy}$  (only equal when  $x=y$ )

$$\text{so } e_k(\vec{p}) \leq e_k(\vec{r}), \text{ only equal when } p_1 = p_2$$

So  $P(\text{at least one birthday match} | p) \geq P(\text{at least one birthday match} | r)$ .

then we define a vector  $\vec{r} = (r_1, r_2, p_3, p_4, p_5, \dots)$ ,  $\vec{m} = (r_1, m_2, m_3, p_4, p_5, \dots)$ ,  $m_2 = m_3 = \frac{r_2 + p_3}{2}$

$$\text{so } e_k(\vec{r}) \leq e_k(\vec{m}), \text{ only equal when } m_2 = m_3 = \frac{p_3 + r_2}{2} = \frac{p_3 + \frac{p_1 + p_2}{2}}{2}$$

then  $e_k(\vec{p}) \leq e_k(\vec{r}) \leq e_k(\vec{m}) \leq \dots \leq e_k(\vec{P})$ .

So only equal when  $P_{n+1} = \frac{P_n + P_{n-1}}{2}$ , and  $P_1 = \frac{P_{365} + P_{364}}{2}$

$$\sum_{k=1}^{365} P_j = 1 \Rightarrow P_j = \frac{1}{365}$$

So  $e_k(\vec{P})$  is max value and the value of  $p$  that minimizes the probability of at least one birthday match is given by  $p_j = \frac{1}{365}$  for all  $j$ .

## Problem 5

$$P = \frac{108! \binom{n}{108}}{108^n} = \frac{\sum_{k=0}^{108} (-1)^k C_{108}^k (108-k)^n}{108^n} = \sum_{k=0}^{108} (-1)^k C_{108}^k \left(\frac{108-k}{108}\right)^n$$

Then we use python to plot an image:

```
import numpy as np
import matplotlib.pyplot as plt
from scipy.special import comb

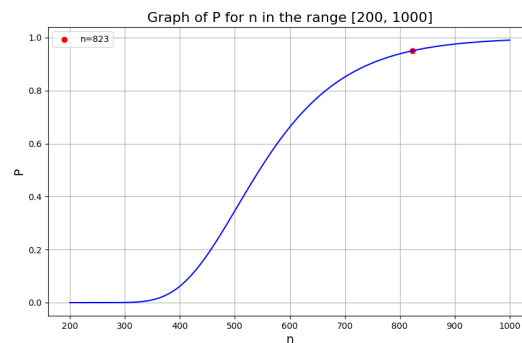
def P_function(n):
    result = 0
    for k in range(109):
        result += (-1)**k * comb(108, k) * ((108 - k) / 108)**n
    return result

n_values = np.arange(200, 1001)
P_values = [P_function(n) for n in n_values]

first_index_above_095 = next((i for i, p in enumerate(P_values) if p > 0.95), None)
n_value_above_095 = n_values[first_index_above_095]

plt.figure(figsize=(10, 6))
plt.plot(n_values, P_values, color='blue', linestyle='-')
plt.scatter(n_value_above_095, P_values[first_index_above_095], color='red', label=f'n={n_value_above_095}')

plt.xlabel('n', fontsize=14)
plt.ylabel('P', fontsize=14)
plt.title('Graph of P for n in the range [200, 1000]', fontsize=16)
plt.grid(True)
plt.legend()
plt.show()
```



We can tell from the fig that the minimum number of  $n$  is 823 when such probability is no less than 95%.