



上海科技大学
ShanghaiTech University

SI140 Probability & Mathematical Statistics

Homework 2

陈昱聪

Chen Yucong ><E<>N

Student ID: 2019533079

Email: chenyc@shanghaitech.edu.cn

SCHOOL OF INFORMATION SCIENCE AND TECHNOLOGY

© Group#2 (TA: 曾理)

Problem 1.9

Solution:

- (a) From point $(0, 0)$ to point $(110, 111)$, we have to go through 110 steps of going one unit to the right and 111 steps of going one unit up, so the number of paths there is

$$\frac{(110 + 111)!}{110! \cdot 111!} = \binom{221}{110}$$

- (b) Break this process into two parts. First we aim to get to point $(110, 111)$, then we go from $(110, 111)$ to $(210, 211)$, then we have the number of paths

$$\binom{221}{110} \cdot \binom{210 + 211 - 110 - 111}{210 - 110} = \binom{221}{110} \cdot \binom{200}{100}$$

Problem 1.25

Solution:

There are k^n choices to store these phone number. And there are $\binom{n}{k} \cdot k!$ choices that no two of the numbers are stored in the same location. So the probability that at least one location has more than one phone number stored there is

$$\frac{n^k - \binom{n}{k} \cdot k!}{n^k}$$

Problem 1.40

Proof:

The total number of *norepeatwords* is

$$\binom{26}{1} \cdot 1! + \binom{26}{2} \cdot 2! + \binom{26}{3} \cdot 3! + \cdots + \binom{26}{26} \cdot 26! = \sum_{k=1}^{26} \binom{26}{k} \cdot k!$$

The number of *norepeatwords* using 26 letters is

$$\binom{26}{26} \cdot 26!$$

So the probability is

$$\begin{aligned} \frac{\binom{26}{26} \cdot 26!}{\sum_{k=1}^{26} \binom{26}{k} \cdot k!} &= \frac{26!}{\sum_{k=1}^{26} \frac{26!}{(26-k)!}} \\ &= \frac{1}{\sum_{k=1}^{26} \frac{1}{(26-k)!}} \\ &= \frac{1}{\sum_{k=0}^{25} \frac{1}{k!}} \end{aligned}$$

From the inverse of Taylor's formula, we have $1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots = e^x$, when $x = 1$, we know that $1 + 1 + \frac{1}{2!} + \frac{1}{3!} + \cdots = e$ so we know that $\frac{1}{\sum_{k=0}^{25} \frac{1}{k!}} \approx \frac{1}{e}$.

□

Problem 1.58

(a) **Solution:**

There is n numbers for n positions which allowed for repetitions, so we have n^n samples.

(b) **Solution:**

Using Bose-Einstein Counting, We have $\binom{n+n-1}{n-1} = \binom{2n-1}{n}$ samples.

(c)

• **Proof:**

The probability of getting a certain unordered bootstrap sample is the number of corresponding ordered samples divided by the total number of ordered samples. Since the numbers of different types of ordered samples are different (such as there is just one sample that all the elements of it are a_1 but there are $n!$ samples that each of them contains all the a_i s.) \square

• **Solution:**

The most unlikely sample is that it just contains one specific type of elements because the corresponding ordered samples are the fewest. Since there are n elements, we get that $p_2 = \frac{n}{n^n} = \frac{1}{n^{n-1}}$.

The most likely sample is that it contains all types of elements (a_1, a_2, \dots, a_n) because the number of corresponding ordered samples are the largest. Since there are n elements, we get that $p_1 = \frac{n!}{n^n}$.

\Rightarrow So $\frac{p_1}{p_2} = n!$.

• **Solution:**

There is just one unordered sample that it contains all the elements, but n samples that only contains one type of elements since there is n a_i s. So the ratio is $\frac{1}{n}$.

Problem 1.60

(a) **Solution:**

First we find the probability of no birthday matches. So we need to take k birthday without repetition. Since the order matters (because people are different from each other), so the probability is $k! \cdot e_k(\mathbf{p})$. Now we can take the complement, so the probability is

$$1 - k! \cdot e_k(\mathbf{p})$$

(b) **Solution:**

An extreme case is that we just have one day that the probability is 1, and the rest of it are all 0. That is $\mathbf{p} = (0, 0, 0, \dots, 1, 0, \dots, 0)$. So all people will have the same birthday and so the P will be maxed to 1. So we can easily think that if the probability distribution is more evenly distributed, P is going to go down. So that, when the probability of each date becomes the same (i.e. $p_i = \frac{1}{365}$), P gets the minimum.

(c) **Proof:**

$$\begin{aligned}
 e_k(x_1, x_2, \dots, x_n) &= \sum_{1 \leq j_1 < j_2 < \dots < j_k \leq n} x_{j_1} \dots x_{j_k} \\
 &= \sum_{3 \leq j_1 < j_2 < \dots < j_{k-2} \leq n} x_1 x_2 x_{j_1} \dots x_{j_{k-2}} \quad (\text{Which contains both } x_1 \text{ and } x_2) \\
 &+ \sum_{3 \leq j_1 < j_2 < \dots < j_{k-1} \leq n} x_1 x_{j_1} \dots x_{j_{k-1}} \quad (\text{Which contains } x_1 \text{ but not } x_2) \\
 &+ \sum_{3 \leq j_1 < j_2 < \dots < j_{k-1} \leq n} x_2 x_{j_1} \dots x_{j_{k-1}} \quad (\text{Which contains } x_2 \text{ but not } x_1) \\
 &+ \sum_{3 \leq j_1 < j_2 < \dots < j_k \leq n} x_{j_1} x_{j_2} \dots x_{j_k} \quad (\text{Which doesn't contain } x_1 \text{ nor } x_2) \\
 &= x_1 x_2 e_{k-2}(x_3, x_4, \dots, x_n) + (x_1 + x_2) e_{k-1}(x_3, x_4, \dots, x_n) + e_k(x_3, \dots, x_n)
 \end{aligned}$$

So we have

$$\begin{aligned}
 e_k(\mathbf{p}) &= p_1 p_2 e_{k-2}(p_3, p_4, \dots, p_n) + (p_1 + p_2) e_{k-1}(p_3, p_4, \dots, p_n) + e_k(p_3, \dots, p_n) \\
 &\leq \frac{(p_1 + p_2)^2}{4} e_{k-2}(p_3, p_4, \dots, p_n) + (p_1 + p_2) e_{k-1}(p_3, p_4, \dots, p_n) + e_k(p_3, \dots, p_n) \\
 &= r_1 r_2 e_{k-2}(r_3, r_4, \dots, r_n) + (r_1 + r_2) e_{k-1}(r_3, r_4, \dots, r_n) + e_k(r_3, \dots, r_n) \\
 &= e_k(\mathbf{r})
 \end{aligned}$$

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

So for the vector \mathbf{p} , we can always choose two elements of it and let them become p_1 and p_2 repeatedly, and then take the average so that \mathbf{p} becomes a new \mathbf{r} over and over again. Every time we do this work, the P becomes smaller. Until each elements in \mathbf{p} becomes exactly the same so that it'll never change. Then P gets the minimum. That is, at this time, $p_i = \frac{1}{365}$.