CS70–Spring 2013 — Homework 10

Felix He, SID 303308****

May 24, 2024

Collaborators: None

1. Introductions

1.

2.

3. Let X^k represents the number of students end up in groups of size k. Then $E(X_i^k)$ represents student i ends up in groups of size k.

$$\begin{aligned} E(X^k) &= E(X_1^k) + E(X_2^k) + \ldots + E(X_n^k) \\ &= \frac{1}{n} + \frac{1}{n} + \ldots + \frac{1}{n} \\ &= 1 \end{aligned}$$

If sum the results for all $1 \leq k \leq n$, the result would be n.

$$E(X) = \sum_{k=1}^{n} E(X^k)$$

$E(X)$ is the expected number of students end up in all the groups of size from 1 to n, which must be the total number of students in the class.

4. We already know the expected number of students in group of size k, then how many expected such groups should there be? We should divide by the group size k to get the result.

$$E(G_k) = \frac{\text{Expected number of students in group of size k}}{k} = \frac{1}{k}$$

5. The key observation is we can only have at most one group with size more than half

$$\begin{aligned} \text{P(a group of more than half)} &= \mathbb{E}[\mathcal{Z}] \\ \mathcal{Z} : \text{the number of groups of size} &\geq \frac{n}{2} \\ \mathcal{Z} &= \sum_{k \geq \frac{n}{2}} \mathcal{Z}_k \\ \mathcal{Z}_k : \text{number of groups of size } k \end{aligned}$$

$$
\begin{aligned}
\mathbb{E}[\mathcal{Z}] &= \sum_{k \geq \frac{n}{2}} \mathbb{E}[\mathcal{Z}_k] \\
&= \sum_{k \geq \frac{n}{2}} \mathrm{P}(\mathcal{Z}_k = 1) \\
&= \sum_{k \geq \frac{n}{2}} \frac{1}{k} \\
&= \begin{cases} H_n - H_{\frac{n}{2}}, & \text{if } n \text{ is even} \\ H_n - H_{\lfloor \frac{n}{2} \rfloor}, & \text{if } n \text{ is odd} \end{cases}
\end{aligned}
$$

6. n is even,

$$
\begin{aligned}
H_n - H_{n/2} &= \ln n + \gamma - \ln n + \ln 2 - \gamma \\
&= \ln 2
\end{aligned}
$$

   n is odd,

$$
\begin{aligned}
H_n - H_{\lfloor \frac{n}{2} \rfloor} &= \ln n + \gamma - \ln n - 1 + \ln 2 - \gamma \\
&= \ln \frac{n}{n-1} + \ln 2 \\
&= \ln 2
\end{aligned}
$$

2. Round the Clock

1. Expected waiting time is $E[W] = \frac{24}{8} = 3$.
   Suppose we arrange 3 meals in order after they are chosen based on their time from midnight to 11pm. And their time is denoted as X_1, X_2, X_3. Then the average waiting time is

$$
\begin{aligned}
E[W] &= \frac{E(X_2 - X_1) + E(X_3 - X_2) + E(24 + X_1 - X_3)}{3} \\
&= \frac{E(X_2) - E(X_1) + E(X_3) - E(X_2) + E(24) + E(X_3) - E(X_2)}{3} \\
&= \frac{24}{3} \\
&= 8
\end{aligned}
$$

2. This case it's similar to the question 1 part 1. Instead of having 3 meals, we have 4 meals and we pretend the call is one of the meal. So the question would become if I have 4 meals a day, what's the expected waiting time after I have 1 meal. Similarly we have

$$
\begin{aligned}
E[W] &= \frac{E(X_2 - X_1) + E(X_3 - X_2) + E(X_4 - X_3) + E(24 + X_4 - X_1)}{3} \\
&= \frac{E(X_2) - E(X_1) + E(X_3) - E(X_2) + E(24) + E(X_1) - E(X_4)}{4} \\
&= \frac{24}{4} \\
&= 6
\end{aligned}
$$

3. Expectation of Geometric Distribution

1.

$$S = \sum_{i=1}^{\infty} ir^{i-1} = 1 + 2r + 3r^2 + \ldots + ir^{i-1}$$
$$rS = r + 2r^2 + 3r^3 + \ldots + ir^i$$

$S - rS$, we have

$$(1-r)S = r + r^2 + r^3 + \ldots + r^{i-1} + ri^r$$
$$(1-r)S = \frac{1 \cdot (1-r^i)}{1-r} + ir^i$$

Since $-1 < r < 1$, RHS approaches to

$$(1-r)S = \frac{1}{1-r}$$
$$S = \frac{1}{(1-r)^2}$$

2.

$$E(X) = \sum_{i=1}^{\infty} iPr[X = i] = \sum_{i=1}^{\infty} i(1-p)^{i-1}p = p\sum_{i=1}^{\infty} i(1-p)^{i-1}$$

According Equation(1), let $r = 1 - p$, we know

$$E(X) = p\sum_{i=1}^{\infty} i(1-p)^{i-1} = (1-r)\frac{1}{(1-r)^2} = p\frac{1}{(1-(1-p))^2} = \frac{1}{p}$$

4. Surviving Challenge

It's worth betting.
Let $X_i = 1$ be the event that survive in round i, o.w $E(X_i) = 0$ then expected number of survival rounds in total is

$$E[X] = E[X_1] + E[X_2] + \ldots + E[X_i]$$

And

$$Pr[X_i = 1] = Pr[X_{i-1} = 1] \times \frac{(\#cardsLeft - 3 \times (round - 1))}{\#cardsLeft}$$

Then we can calculate $E(X) = 4.7$, which is > 4, so it's worth betting. The detailed calculation is as below

Round	**Cards Left**	**Factor**	**Probability of Survival**
1	52	52	1
2	51	48	0.941176471
3	50	44	0.828235294
4	49	40	0.676110444
5	48	36	0.507082833
6	47	32	0.345247886
7	46	28	0.210150887
8	45	24	0.112080473
9	44	20	0.05094567
10	43	16	0.018956528
11	42	12	0.005416151
12	41	8	0.00105681
13	40	4	0.000105681
14	39	0	0

Table 1: Expected Rounds of Survival Calculation

5. Manufacturing Failures

The product fails when either step A or step B fails. And since the step A and step B are independent, the probability that the product does not fail is

$$\begin{aligned} P[\text{product not fail}] &= (1-p_a)(1-p_b) = 1 - p_a - p_b + p_a p_b \\ P[\text{product fail}] &= 1 - P[\text{product not fail}] = p_a + p_b - p_a p_b \end{aligned}$$

X denote the number of products until and including the first defective one. X is a geometric distribution.

$$X \sim geom(p_a + p_b - p_a p_b)$$

6. James Bond

Since every time after James fall, he randomly chooses the 3 doors with the same probability. It means if he chooses either of the 2 wrong doors, the expected travel time is the same as before. Therefore we have

$$\begin{aligned} E[X] &= \frac{1}{3}(2 + E[X]) + \frac{1}{3}(5 + E[X]) + \frac{1}{3} \cdot 0 \\ \frac{1}{3}E[X] &= \frac{2+5}{3} \\ E[X] &= 7 \end{aligned}$$

Another way of solving this problem is to think it as geometric distribution for the number of times need to escape. Let X denote the number of tries to escape.

$$X \sim geom(p = \frac{1}{3})$$
$$E[X] = \frac{1}{p} = 3$$

Then the expected fail times is $E[F] = E[X] - 1 = 2$. Let T be the travel time if James fail once. Let W denotes travel time. Then the The expected travel time is $E[W] = E[T] \cdot E[F]$.

$$\begin{aligned} E[T] &= \frac{1}{2} \cdot 2 + \frac{1}{2} \cdot 5 \\ &= \frac{7}{2} \end{aligned}$$

$$E[W] = \frac{7}{2} \times 2 = 7$$

7. Hug Collecting

1. This problem is somewhat similar to the cards collection problem.
   Probability that picking a sad person is $P[S] = \frac{250}{450} = \frac{5}{9}$. Let X denote number of volunteer to hug all the sad people. Then we have

   $$E[X] = E[X_1] + E[X_2] + \ldots + E[X_n]$$

   X_i denote the number of volunteer to hug the i^{th} new sad person after hugging $i-1$ sad people. And $X_i \sim geom(P[i])$, $E[X_i] = \frac{1}{P[i]}$
   Let $P[i]$ denote the probability to hug the i^{th} new sad person after hugging $i-1$ sad people. Then

   $$P[i] = \frac{250 - i + 1}{450}$$

   Then

   $$\begin{aligned} E[X] &= \frac{1}{\frac{250}{450}} + \frac{1}{\frac{249}{450}} + \ldots + \frac{1}{\frac{1}{450}} \\ &= \sum_{i=1}^{250} \frac{1}{\frac{250-i+1}{450}} \\ &= \sum_{k=1}^{250} 450 \cdot \frac{1}{k} \\ &\approx 450 \cdot (\ln 250 + 0.5772) \qquad (\sum_{k=1}^{n} 450 \cdot \frac{1}{k} \approx \ln n + \gamma) \\ &\approx 2744 \end{aligned}$$

2. We can apply the union bound here.
   Let S_i be the i^{th} sad person that did not get hug after t volunteers were sent to reached them. Then the probability of S_i is

$$P[S_i] = (\frac{n-1}{n})^t$$

Let S be the event that after sending t volunteers, there are still at least one sad person did not get hug. Then

$$S = \cup_{i=1}^{m} S_i$$

m is the number of sad people.

Then

$$\begin{aligned} P[S] = P[\cup_{i=1}^{m} S_i] &\leq \sum_{i=1}^{m} P[S_i] \\ &= m(\frac{n-1}{n})^t \end{aligned}$$

And we want

$$P[S] \leq m(\frac{n-1}{n})^t < \frac{1}{3}$$

Take ln to the both sides.

$$\ln m + t \ln \frac{n-1}{n} < \ln(\frac{1}{3})$$

Put in $m = 250, n = 450$, we get

$$t > 2975.7$$

It means we should send at least 2976 volunteers.

3. The code is as below. We can see for 2976 students, the chance to cover all sad students 0.7192. To make it $\frac{2}{3}$, the actual number is slightly smaller, which is 2900.

```python
import numpy as np

m = 250
n = 450
vol = 2976
sim = 10000

hug = set()
count = 0

for _ in range(sim):
    for _ in range(vol):
        student = np.random.randint(1, 451)
        if student <= m:
```

```
                hug.add(student)
        if len(hug) == m:
            count += 1
        hug = set()

print("times of covering all sad students: " + str(count) + " out of " + str(sim) + ". Probability is {}".format(count * 1.0 / sim))
```

Listing 1: Simulating sad student selection