

# **Probability & Statistics for EECS:**

## **Homework #06**

Due on March 26, 2023 at 23:59

Name: **Wang Yunfei**  
Student ID: 2021533135

## Problem 1

### Solution

- (a) First assume the indicator  $I_j$  is the event having the  $j$ -th toy in our collection after collecting  $t$  toys,  $1 \leq j \leq n$

Then what we want is the total number of distinct toy types and assume it as  $I$ , then  $I = \sum_{j=1}^n I_j$ .

Then  $E[I] = E[\sum_{j=1}^n I_j] = \sum_{j=1}^n E[I_j]$ , because of the linearity of the expectation.

And because the events that  $I_j$  represents are i.i.d.s.

Then we can simplify the formula, that is  $E[I] = n * E[I_1]$ . Then what we need to calculate is  $E[I_1]$ , for example.

Obviously it is easy to calculate from the opposed side, which means we collect other types except the first type in  $t$  times, that is  $P = (\frac{n-1}{n})^t$ .

Then we can get  $E[I_1] = (1 - P) * 1 + P * 0 = (1 - (\frac{n-1}{n})^t)$ .

Therefore, the answer is  $E[I] = n * E[I_1] = n * (1 - (\frac{n-1}{n})^t)$ .

## Problem 2

1. First assume the indicator  $I_j$  is the event that position  $j$  is a new run, for  $1 \leq j \leq n$ .

And assume the event  $X$  is the number of new runs, and we can have  $X = \sum_{j=1}^n I_j$ .

Then we can get  $I_1 = 1$  always, because this is the first time we flip. And for  $2 \leq j \leq n$ , we can have  $I_j = 1$ , when the result of  $j$ -th toss is different from that of the  $(j-1)$ th toss.

Then we can calculate when  $j = 1$ ,  $E[I_j] = 1 * 1 + 0 * 0 = 1$ ,

and when  $2 \leq j \leq n$ ,  $P(I_j \text{ occurs}) = P(j-1\text{th toss is H})P(j\text{th toss is T}) + P(j-1\text{th toss is T})P(j\text{th toss is H}) = 2 * p(1-p)$ .

Then  $E[X] = E[\sum_{j=1}^n I_j] = \sum_{j=1}^n E[I_j] = 1 + (n-1) * 2 * p(1-p)$ , because of the linearity of the expectation and the sameness of the events which the indicator  $I_j$ s represent. Therefore the result of the expected new runs is  $2p(n-1)(1-p) + 1$ .

## Problem 3

### Solution

- (a) First from the question we can have that  $n$  elk is tagged and  $N-n$  is untagged. At the meanwhile, for the new method we need to make sure the last elk is tagged and it is one by one capturing without replacement. Therefore we can have the following formula:

$$P(X = k) = \frac{\binom{n-1}{m-1} \binom{N-n}{k-m}}{\binom{N-n}{m+k-1}} * \frac{n-(m-1)}{N-(m+k-1)}, \text{ for } 0 \leq k \leq N-n.$$

And at the same time we can get  $Y = X + m$ , then we can substitute it into  $P(X = k)$ .

Then we can have  $P(Y = i) = P(X = i - m) = \frac{\binom{n-1}{m-1} \binom{N-n}{i-m}}{\binom{N-n}{i-1}} * \frac{n-(m-1)}{N-(i-1)}$ , for  $m \leq i \leq N-n+m$ .

Therefore we get them.

- (b) First we label the untagged elk as  $1, 2, 3, \dots, N-n$ , and then assume the indicator  $I_j$  is used for representing the event we capture the  $j$ -th untagged elk being captured before any tagged elk during the new sample, for  $1 \leq j \leq N-n$ .

And then we can define  $X_i$ , for  $1 \leq i \leq m$ , which means before the  $i$ th tagged elk being captured and after the  $(i-1)$ th elk being captured, the number of capturing untagged elk.

So we have  $X = X_1 + \dots + X_m$ , and  $X_i = I_1 + \dots + I_{N-n}$ . And for  $E[I_j]$ , when it is under the condition of the  $X_1$ , it is equal to  $\frac{1}{n+1}$  obviously, because it is before every tagged elk and the method is capturing

one by one without replacement.

Then we can have  $E[X_1] = E[\sum_{j=1}^{N-n} I_j] = \sum_{j=1}^{N-n} E[I_j] = \frac{(N-n)}{n+1}$ , because of the linearity and they are same distribution.

And then we can have  $E[X] = E[\sum_{i=1}^m X_i] = \sum_{i=1}^m E[X_i] = \frac{m(N-n)}{n+1}$ , because of the linearity and they are same distribution.

And then we can get  $E[Y] = E[X+m] = E[X] + m = \frac{m(N+1)}{n+1}$ , because of the property of the expectation.

- (c) From the definition of the capture-recapture, and assume the number of the tagged elk is  $M$ . Then we can have a formula  $\frac{n}{N} = \frac{E[M]}{E[Y]}$ .

Then  $E[M] = \frac{n}{N} * \frac{m(N+1)}{n+1}$ . Because  $n < N$ , then we can get  $E[M] < m$ .

## Problem 4

### Solution

- (a) First we can assume that  $p$  is the chance of a birthday match for 22 people, and  $0 < p \leq 0.5$  from the question.

And for  $X = 23$ ,  $P(X \leq 23) = 0.507 > 0.5$ , at the same time  $P(X \geq 23) = 1 - P(X \leq 22) = 1 - p > 0.5$ .

Therefore, the median of  $X$  is 23.

For  $m > 23$ ,  $P(X \geq m) = 1 - P(X < m) \leq 1 - P(X \leq 23) < 0.5$

For  $m < 23$ ,  $P(X \leq m) \leq P(X \leq 22) = p < 0.5$

Therefore, 23 is the unique median of  $X$ .

- (b) Because  $X = I_1 + I_2 + \dots + I_{366}$  and the linearity of the expectation,

then  $E[X] = E[\sum_{j=1}^{366} I_j] = \sum_{j=1}^{366} E[I_j] = \sum_{j=1}^{366} P(X \geq j)$ .

And obviously we have  $P(X \geq 1) = 1 = p_1$ ,  $P(X \geq 2) = 1 = p_2$ , and  $P(X \geq j) = P(\text{the first } j-1 \text{ people do not have the match}) = \frac{365 * \dots * (365-j+2)}{365^{j-1}}$ .

After simplifying, it is equal to  $p_j$ , for  $3 \leq j \leq 366$ .

Finally,  $E[X] = \sum_{j=1}^{366} p_j$ .

- (c) After calculating,  $E[X] \approx 24.6166$ .

- (d) Because  $Var[X] = E[X^2] - E[X]^2$ , then we can have follows for  $i < j$ .

$$X^2 = I_1^2 + I_2^2 + \dots + I_{366}^2 + 2 \sum_{j=2}^{366} \sum_{i=1}^{j-1} I_i I_j$$

Because the property of the indicator

$$\begin{aligned} &= I_1 + I_2 + \dots + I_{366} + 2 \sum_{j=2}^{366} \sum_{i=1}^{j-1} I_j \\ &= I_1 + I_2 + \dots + I_{366} + 2 \sum_{j=2}^{366} (j-1) I_j \end{aligned}$$

$$\begin{aligned} E[X^2] &= \sum_{j=1}^{366} E[I_j] + 2 * \sum_{j=2}^{366} (j-1) E[I_j] \\ &= \sum_{j=1}^{366} p_j + 2 * \sum_{j=2}^{366} (j-1) p_j \end{aligned}$$

Therefore,  $Var[X] = E[X^2] - E[X]^2 = \sum_{j=1}^{366} p_j + 2 * \sum_{j=2}^{366} (j-1)p_j - (\sum_{j=1}^{366} p_j)^2 \approx 148.640$

## Problem 5

### Solution

- (a) First we can calculate the total number through the method of unordered sampling with replacement, that is  $\binom{14+5-1}{5-1} = 3060$ .

Second, we can eliminate out-of-bounds cases as follows.

(1) Randomly select one box and put seven balls in it and then assign seven balls to remaining four boxes, that is  $5 * \binom{7+4-1}{4-1} - \binom{5}{2} = 590$ , because it is still possible to assign seven balls all in one box.

(2) Randomly select one box and put eight balls in it and then assign six balls to remaining four boxes, that is  $5 * \binom{6+4-1}{4-1} = 420$ .

(3) Randomly select one box and put nine balls in it and then assign five balls to remaining four boxes, that is  $5 * \binom{5+4-1}{4-1} = 280$ .

(4) Randomly select one box and put ten balls in it and then assign four balls to remaining four boxes, that is  $5 * \binom{4+4-1}{4-1} = 175$ .

(5) Randomly select one box and put eleven balls in it and then assign three balls to remaining four boxes, that is  $5 * \binom{3+4-1}{4-1} = 100$ .

(6) Randomly select one box and put twelve balls in it and then assign two balls to remaining four boxes, that is  $5 * \binom{2+4-1}{4-1} = 50$ .

(7) Randomly select one box and put thirteenth balls in it and then assign one balls to remaining four boxes, that is  $5 * \binom{1+4-1}{4-1} = 20$ .

(8) Randomly select one box and put 14 balls in it and then assign 0 balls to remaining four boxes, that is  $5 * \binom{0+4-1}{4-1} = 5$ .

Then the result is  $3060 - (590 + 420 + 280 + 175 + 100 + 50 + 20 + 5) = 1420$ .