Probability & Statistics for EECS: Homework #05

Due on March 19, 2023 at 23:59

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

Problem 1

Solution

a For the enumeration strategy,
the probability of the first time Kratos geting the right answer is $\frac{1}{9}$
the Probability of the second time Kratos geting the right answer is $\frac{8}{9} * \frac{1}{8} = \frac{1}{9}$
...
the Probability of the eighth time Kratos geting the right answer is $\frac{8}{9} * \cdots * \frac{2}{3}$. Only this item is different from others, that is because if yes, we can get the right answer and if not, another one is the right answer.
Therefore, the expectation of this strategy is

$$E(enumeration\ strategy) = 1 * \frac{1}{9} + 2 * \frac{1}{9} + \cdots + 7 * \frac{1}{9} + 8 * \frac{2}{9} = \frac{44}{9}.$$

b For the bisection strategy, we can get the result by enumeration.
When the result is 1 or 2, we need four steps to get the final answer: <= 5, <= 3, <= 2 and <= 1.
When the result is one of remaining numbers, we need three steps to get the final answer.
Take 3 as an example, whether <= 5, yes, whether <= 3, yes, whether <= 2, no, then we can get 3 only through three steps.
Take 9 as an example, whether <= 5, no, whether <= 7, no, whether <= 8, no, then we can get 9.
The only point to note in this process is that at the even length edge we need to make sure that the indicators for the different number enumerations are the same, so that we get the correct answer.
Therefore, $E(bisection\ strategy) = \frac{3*7+2*4}{9} = \frac{29}{9}$.

Problem 2

Solution

(a) This is obviously a Coupon collector problem. And we have five different types to collect.
So assume N_j number of days to get the j-th new types, $1 <= j <= 5$.
Assume N is the total number of getting all types of steak, and because the process of collecting different j-th new types is independent.
Then we have $N = N_1 + N_2 + N_3 + N_4 + N_5$.
Obviously,

$$\begin{aligned} &N_1 = 1 \\ &N_2 \sim Fs(1 - \frac{1}{5}) \\ &N_3 \sim Fs(1 - \frac{2}{5}) \\ &N_4 \sim Fs(1 - \frac{3}{5}) \\ &N_5 \sim Fs(1 - \frac{4}{5}) \end{aligned}$$

From the linearity of expectation, we have $E[N] = E[N_1] + E[N_2] + E[N_3] + E[N_4] + E[N_5]$.
Because the property of the first success distribution, take $X \sim Fs(p)$ for example , $E[X] = \frac{1}{p}$.
Then we can get $E[N] = 1 + \frac{5}{4} + \frac{5}{3} + \frac{5}{2} + \frac{5}{1} = \frac{137}{12} \approx 11.41$.
Therefore I expect to spend 11.41 days, and then at least 12 days.

Problem 3

Solution

(a) From the question, we can get this is a Fs distribution.
Because of $X_i \sim Bern(p_1)$ and $Y_i \sim Bern(p_2)$, we can assume Z_i, which means i-th time they are simultaneously successful.
Obviously this is also a Bernoulli distribution with parameter p_1p_2.
Then we can get $Z \sim Fs(p_1p_2)$, that is $P(Z = k) = (1 - p_1p_2)^{k-1}p_1p_2$. And from what we have learned about Fs distribution, $E[Z] = \frac{1}{p_1p_2}$.

(b) From the question, we can get this is a Fs distribution.
Because of $X_i \sim Bern(p_1)$ and $Y_i \sim Bern(p_2)$, we can assume Z_i, which means i-th time they are simultaneously successful.
Obviously this is also a Bernoulli distribution with parameter $1 - (1 - p_1)(1 - p_2) = p_1 + p_2 - p_1p_2$.
Then we can get $Z \sim Fs(p_1 + p_2 - p_1p_2)$, that is $P(Z = k) = [1 - (p_1 + p_2 - p_1p_2)]^{k-1}(p_1 + p_2 - p_1p_2)$.
And from what we have learned about Fs distribution, $E[Z] = \frac{1}{p_1+p_2-p_1p_2}$.

(c) Assume X is the number of times of Mario flipping untill the first success, and it is distributed as $X \sim Fs(p_1)$.
Assume Y is the number of times of Zelda flipping untill the first success, and it is distributed as $Y \sim Fs(p_2)$.
So what we want to get is

$$\begin{aligned}
P(X = Y) &= \sum_{k=1}^{\infty} P(X = Y|Y = k)P(Y = k) \\
&= \sum_{k=1}^{\infty} P(X = k|Y = k)P(Y = k) \\
&= \sum_{k=1}^{\infty} \frac{P(X = k, Y = k)}{P(Y = k)} P(Y = k) \\
&= \sum_{k=1}^{\infty} P(X = k, Y = k) \\
\textit{And because of the independence of } X \textit{ and } Y \\
&= \sum_{k=1}^{\infty} P(X = k)P(Y = k) \\
&= p_1p_2 \sum_{k=1}^{\infty} [(1 - p_1)(1 - p_2)]^{k-1} \\
\textit{Because of } p_1 = p_2 \\
&= p_1^2 \sum_{k=1}^{\infty} [(1 - p_1)^2]^{k-1} \\
&= p_1^2 \frac{1}{1 - (1 - p_1)^2} \\
&= \frac{p_1}{2 - p_1}
\end{aligned}$$

There are only three cases, either their first successes are simultaneous, or not.
So $P(X \neq Y) = 1 - \frac{p_1}{2-p_1} = \frac{2-2p_1}{2-p_1}$.

Problem 3 continued on next page. . .

And because $p_1 = p_2$, either Mario precedes Zelda, or Zelda precedes Mario, and the probabilities are equal.
So $P(Mario's\ first\ success\ precedes\ Zelda's) = \frac{1}{2}\frac{2-2p_1}{2-p_1} = \frac{1-p_1}{2-p_1}$.

Problem 4

Solution

(a) From the question, we have $X \sim Geom(p)$ and $Y \sim Geom(q)$, and they are independent.

$$\begin{aligned}
P(X=Y) &= \sum_{k=0}^{\infty} P(X=Y|Y=k)P(Y=k) \\
&= \sum_{k=0}^{\infty} P(X=k|Y=k)P(Y=k) \\
&= \sum_{k=0}^{\infty} \frac{P(X=k,Y=k)}{P(Y=k)}P(Y=k) \\
Because\ of\ they\ are\ independent &= \sum_{k=0}^{\infty} P(X=k)P(Y=k) \\
&= pq\sum_{k=0}^{\infty}[(1-p)(1-q)]^k
\end{aligned}$$

And then get the answer from the geometric progression, $P(X=Y) = \frac{pq}{p+q-pq}$.

(b) From LOTP and the properties of the geometric distribution and the definition of expectation, then we can have

$$\begin{aligned}
E[max(X,Y)] &= \sum_{k=0}^{\infty} kP(max(X,Y)=k) \\
&= \sum_{k=0}^{\infty} k[P(X=k,Y<=k)+P(X<k,Y=k)] \\
Because\ of\ independence &= \sum_{k=0}^{\infty} k[P(X=k)P(Y<=k)+P(X<k)P(Y=k)] \\
&= \sum_{k=0}^{\infty} k\{p(1-p)^k[1-(1-q)^{k+1}]+q(1-q)^k[1-(1-p)^k]\} \\
&= \sum_{k=0}^{\infty} k\{p(1-p)^k+q(1-q)^k-[(1-p)(1-q)]^k(p+q-pq)\} \\
&= \sum_{k=0}^{\infty} kp(1-p)^k+\sum_{k=0}^{\infty} kq(1-q)^k-\sum_{k=0}^{\infty} k[(1-p)(1-q)]^k(p+q-pq) \\
&= \frac{1-p}{p}+\frac{1-q}{q}-\frac{1-p-q+pq}{p+q-pq}
\end{aligned}$$

(c) From LOTP, then we have

$$\begin{aligned}
P(min(X,Y)=k) &= P(X=k, Y>=k) + P(X>k, Y=k) \\
Because\ of\ they\ are\ independent \\
&= P(X=k)P(Y>=k) + P(X>k)P(Y=k) \\
From\ the\ property\ of\ Geom,\ we\ have\ P(Y>=k) = (1-q)^k \cdots \\
&= (1-p)^k p(1-q)^k + (1-p)^k(1-p)(1-q)^k * q \\
&= [1-p-q+pq]^k(p+q-pq)
\end{aligned}$$

(d) First, for $E[X|X<=Y]$

$$\begin{aligned}
E[X|X<=Y] &= \sum_{x=0}^{\infty} xP(X|X<=Y) \\
&= \sum_{x=0}^{\infty} x\frac{P(X, X<=Y)}{P(X<=Y)} \\
&= \sum_{x=0}^{\infty} x\frac{P(X=x, x<=Y)}{P(X<=Y)}
\end{aligned}$$

Second, for$P(X<=Y)$

$$\begin{aligned}
P(X<=Y) &= \sum_{x=0}^{\infty} P(X=x, x<=Y) \\
Because\ of\ they\ are\ independent \\
&= \sum_{x=0}^{\infty} P(X=x)P(x<=Y) \\
&= \sum_{x=0}^{\infty} (1-p)^x p(1-q)^x \\
&= p\sum_{x=0}^{\infty} [(1-p)(1-q)]^x \\
&= p\frac{1}{p+q-pq} \\
&= \frac{p}{p+q-pq}
\end{aligned}$$

Then we can go back to step 1

$$Because\ of\ they\ are\ independent$$

$$\begin{aligned} E[X|X <= Y] &= \sum_{x=0}^{\infty} x \frac{P(X = x)P(x <= Y)}{P(X <= Y)} \\ &= \frac{p+q-pq}{p} \sum_{x=0}^{\infty} xP(X = x)P(x <= Y) \\ &= \frac{p+q-pq}{p} \sum_{x=0}^{\infty} xp[(1-p)(1-q)]^x \\ &= (p+q-pq) \sum_{x=0}^{\infty} xp[(1-p)(1-q)]^x \end{aligned}$$

$$This\ can\ be\ seemed\ asanother\ Geom's\ Expectation$$

Therefore, $E[X|X <= Y] = \frac{1-p-q+pq}{p+q-pq}$.

Problem 5

Solution

(a) At the beginning, we need k days to do the exploration phase, and m-k days to do the exploitation phase.
Assume T is the total number of the rank,
and X_i is the rank of the i-th exploration time, $1 <= i <= k$. There is no doubt that they are independent.
So we have $T = \sum_{i=1}^{k} X_i + (m-k)X$, and from the linearity of expectation,
we actually have $E[T] = \sum_{i=1}^{k} E[X_i] + (m-k)E[X]$
For X_i, when $i = 1$, $E[X_1]$ is obviously equal to $\frac{1}{n}\sum_{j=1}^{n} j = \frac{n+1}{2}$.
When $i = 2$, because our strategy, we wonnot try the dish we have tried.
If we had tried the dish 1, then we will try one of the remaining n-1 types equally likely.
...
If we had tried the dish n, then we will try one of the remaining n-1 types equally likely.
Then

$$\begin{aligned} E[X_2] &= \frac{1}{n}\frac{1}{n-1}(\sum_{j=2}^{n} j) + \frac{1}{n}\frac{1}{n-1}(\sum_{j=3}^{n} j + 1) + \cdots + \frac{1}{n}\frac{1}{n-1}(\sum_{j=1}^{n-1} j) \\ &= \frac{1}{n}\frac{1}{n-1}(n-1)\frac{n(n+1)}{2} \\ &= \frac{n+1}{2} \end{aligned}$$

Similarly we can get the same result for the remaining X_i.
Then $E[T] = \frac{k(n+1)}{2} + (m-k)E[X]$.

(b) From the strategy, we will choose k length of the exploration day, then the support of X can be k to n.
If $X = k$, then the dish set we choose is only $\binom{k-1}{k-1} = 1$ case, that is$\{1, 2, \cdots k\}$.
If $X = k+1$, then the dish set we choose has $\binom{k}{k-1} = \binom{k}{1} = k$ cases, that is we need to choose k-1 items from the range 1 to k.
...

If $X = n$, then the dish set we choose has $\binom{n-1}{k-1}$ cases, that is we need to choose k-1 items from the range 1 to n-1.
Therefore the PMF of X is $P(X = m) = \frac{\binom{m-1}{k-1}}{\binom{n}{k}}$, $k <= m <= n$ and otherwise is equal to 0.

(c) From (b), we have had the PMF of X, and $E[X] = \sum_{x=0}^{\infty} xP(X = x)$.
Then we can have

$$\begin{aligned} E[X] &= \sum_{x=k}^{n} x\frac{\binom{x-1}{k-1}}{\binom{n}{k}} \\ &= \frac{k}{\binom{n}{k}}\sum_{x=k}^{n}\binom{x}{k} \\ &= \frac{k\binom{n+1}{k+1}}{\binom{n}{k}} \\ &= \frac{k(n+1)}{k+1} \end{aligned}$$

Therefore we get it.

(d) From (c) and (a), then we have
$E[T] = \frac{k(n+1)}{2} + (m-k)\frac{k(n+1)}{k+1} = (n+1)(\frac{k}{2} + \frac{m-k}{k+1})$.
Then we can obtain the derivative of E[T], that is $E'[T] = (n+1)[\frac{m+1}{(k+1)^2} - \frac{1}{2}]$.
Obviously when $\frac{m+1}{(k+1)^2} = \frac{1}{2}$, $E'[T] = 0$. Then we have $k = \sqrt[2]{2(m+1)} - 1$.
And when $0 <= k <= \sqrt[2]{2(m+1)} - 1$, $E'[T] >= 0$, so $E[T]$ monotonically increasing.
And when $\sqrt[2]{2(m+1)} - 1 <= k <= n$, $E'[T] <= 0$, so $E[T]$ monotonically decreasing.
Therefore when $k = \sqrt[2]{2(m+1)} - 1$ or this rounded up or down to an integer if needed, $E[T]$ is maximum.