10/10

1. (Geometric distribution)

(a) **(In parallel)** Solution: From the description of the problem, we know that $X_1 \sim Geom(p1)$ and $X_2 \sim Geom(p2)$. We can write the two distributions as follows,

$$Pr[X_1 = i] = (1-p_1)^{i-1}p_1 \text{ for } i \geq 1$$
$$Pr[X_2 = i] = (1-p_2)^{i-1}p_2 \text{ for } i \geq 1$$

3/3

For $X = \min\{X_1, X_2\}$, we can consider an imaginary new machine M, for each run, it is fails when either M_1 or M_2 fails. Thus, the new machine will have a failure probability $p = 1-(1-p_1)(1-p_2)$. Thus, $X \sim Geom(1-(1-p_1)(1-p_2))$.

$$Pr[X = i] = \Big[(1-p_1)(1-p_2)\Big]^{i-1}\Big(1-(1-p_1)(1-p_2)\Big) \text{ for } i \geq 1.$$

(b) **(007 stype)** Solution: Let X_1 and X_2 denote the number of times that James Bond will choose the air-conditioning duct and the sewer pipe respectively before choosing the unlocked door, and $X = X_1 + X_2$ is the number of wrong choices before choosing the unlocked door. Then $(X+1) \sim Geom(1/3)$.

7/7

$$\begin{aligned} Pr[X = i] &= (1-1/3)^i(1/3) \\ &= (2/3)^i(1/3) \\ \mathbb{E}(X) &= 1/(1/3) - 1 = 2 \end{aligned}$$

Since the air-conditioning and the sewer pipe have an equal probability of being chosen, then $\mathbb{E}(X_1) = \mathbb{E}(X_1) = \mathbb{E}(X)/2 = 1$. Then the expected time needed before choosing the door is $2\mathbb{E}(X_1) + 5\mathbb{E}(X_2) = 7$ hours.

2. (Fish)

$\frac{10}{10}$

(a) Solution: This is a problem similar to the Coupon Collector's Problem covered in lecture. Let X_i denote the number of fish caught while we already have $i-1$ different fish caught. Then the total number of fish caught in order to get all kinds is $X = \sum_{i=1}^{n} X_i$ where X_i follows a geometric distribution, i.e, $X_i \sim Geom(p_i)$, and $p_i = 1 - \frac{i-1}{n}$.

$$\begin{aligned}
\mathbb{E}(X_i) &= \frac{1}{p_i} = \frac{n}{n-i+1} \\
\mathbb{E}(X) &= \mathbb{E}\Big[\sum_{i=1}^{n} X_i\Big] \\
&= \sum_{i=1}^{n} \mathbb{E}[X_i] \\
&= \sum_{i=1}^{n} \frac{n}{n-i+1} \\
&= n\sum_{i=1}^{n} \frac{1}{i} \\
&= nH(n) \approx n(\ln n + \Theta(1)) = n\ln n + \Theta(n)
\end{aligned}$$

$\frac{1}{1}$

(b) Solution: Let X_i be an indicator random variable denoting whether the fish kind i is not caught when $2n$ fish have been caught. Let X denote the number of kinds of fish not caught when $2n$ fish have been caught.

$$\begin{aligned}
\mathbb{E}(X_i) &= Pr[X_i = 1] \\
&= (1 - \frac{1}{n})^{2n} \\
\mathbb{E}(X) &= \mathbb{E}(\sum_{i=1}^{n} X_i) \\
&= \sum_{i=1}^{n} \mathbb{E}(X_i) \\
&= n(1 - \frac{1}{n})^{2n}
\end{aligned}$$

$\frac{2}{2}$

(c) Solution: Let Y_i be an indicator random variable denoting whether the fish kind i has only been caught exactly once after $3n$ fish being caught.

$$
\begin{aligned}
\mathbb{E}(Y_i) &= Pr[Y_i = 1] \\
&= \binom{3n}{1}(1-\frac{1}{n})^{3n-1}(\frac{1}{n}) \\
\mathbb{E}(Y) &= \mathbb{E}(\sum_i^n Y_i) \\
&= \sum_i^n \mathbb{E}(Y_i) \\
&= n\binom{3n}{1}(1-\frac{1}{n})^{3n-1}(\frac{1}{n}) = 3n(\frac{n-1}{n})^{3n-1}
\end{aligned}
$$

$\frac{2}{2}$

(d) Solution: We can use similar method in (a). Let X_i denote the number of fish caught while we already have $i-1$ different fish caught. Then the total number of fish we need to get $n/2$ kinds of fish is $X = \sum_{i=1}^{n/2} X_i$ where $X_i \sim Geom(1 - \frac{i-1}{n})$.

$$
\begin{aligned}
\mathbb{E}(X) &= \mathbb{E}(\sum_i^{n/2} X_i) \\
&= \sum_{i=1}^{n/2} \frac{n}{n-i+1} \\
&= n \sum_{i=n/2+1}^{n} \frac{1}{i} \\
&= n(H(n) - H(n/2)) \approx n(\ln n - \ln n/2) + \Theta(n) = n\ln 2 + \Theta(n)
\end{aligned}
$$

$\frac{2}{2}$.

(e) Solution: Let F_i be the total number of fish for generation i and F be the number of fish in all generations.

$$
\begin{aligned}
&\mathbb{E}(F_1) = 1p_1 + 2p_2 \\
&\mathbb{E}(F_2) = \mathbb{E}(F_1)p_1 + 2\mathbb{E}(F_1)p_2 = \mathbb{E}(F_1)(p_1 + 2p_2) = (p_1 + 2p_2)^2 \\
&\mathbb{E}(F_3) = \mathbb{E}(F_2)p_1 + 2\mathbb{E}(F_2)p_2 = \mathbb{E}(F_2)(p_1 + 2p_2) = (p_1 + 2p_2)^3 \\
&\quad\vdots \\
&\mathbb{E}(F_i) = (p_1 + 2p_2)^i
\end{aligned}
$$

$\frac{3}{3}$

Thus, the expected number of fish in the talk will be

$$
\begin{aligned}
\mathbb{E}(F) &= \mathbb{E}(\sum_{i=1}^{\infty} F_i) \\
&= \sum_{i=1}^{\infty}(p_1 + 2p_2)^i = \frac{1}{1-(p_1+2p_2)}, \text{ given } p_1 + 2p_2 < 1
\end{aligned}
$$

The above sum is bounded only when $p_1 + 2p_2 < 1$.

$\frac{8.5}{10}$

Peilun Dai

3. (Consecutive ones)

(+3)

(a) Solution: Let R.V. R_2 denote the number of rolls until getting a pair of ones, and X_i be the random variable that denote the result at the i^{th} roll. Then by law of total probability and linearity of expectation, we have

$$\begin{aligned}\mathbf{E}[R_2] =&(1 - Pr[X_1])(\mathbf{E}[R_2] + 1)\\ &+ (Pr[X_1 = 1](1 - Pr[X_2 = 1]))(\mathbf{E}[R_2] + 2)\\ &+ (Pr[X_1 = 1]Pr[X_2 = 1])2\\ =&\frac{k-1}{k}(\mathbf{E}[R_2]) + \frac{1}{k}\frac{k-1}{k}(\mathbf{E}[R_2] + 2) + \frac{1}{k^2}2\end{aligned} \tag{1}$$

R_1 and R_2 undefined

If we solve (1), we can obtain $\mathbf{E}[R_2] = k^2 + k$.

(b) Solution: Similar to the last question, let R.V. R_3 denote the number of rolls until getting a triple of consecutive ones.

$$\begin{aligned}\mathbf{E}[R_3] =&(1 - Pr[X_1 = 1])(\mathbf{E}[R_3] + 1)\\ &+ Pr[X_1 = 1](1 - Pr[X_2 = 1])(\mathbf{E}(R_3) + 2)\\ &+ Pr[X_1 = 1]Pr[X_2 = 1](1 - Pr[X_3 = 1])(\mathbf{E}(R_3) + 3)\\ &+ Pr[X_1 = 1]Pr[X_2 = 1]Pr[X_3 = 1]3\end{aligned} \tag{2}$$

We solve (2) for $\mathbf{E}[R_3] = k^3 + k^2 + k - 3$.

$= k^3 + k^2 + k$

+5.5

9.5/10

4. (Selling a plane)

(a) Solution: Since the distribution of the sequence is uniform, by symmetry, we know that each of the first $i-1$ offers has an equal probability of being the highest among the first $i-1$ offers. Let $F_{m,i-1}$ denote the event that the best among the first $i-1$ offers is in the first m offers. Then

$$Pr[F_{m,i-1}] = m/(i-1)$$

(b) Solution: In order for B_i to happen, the largest among the first $i-1$ numbers must be in the first m numbers, otherwise, another number between $m+1$ and $i-1$ will be chosen instead of i. The second condition is that the i^{th} number must be the largest number in order for B_i to happen. The second event is denoted by L_i.

what if $i \le m$?

Then $P[B_i] = 0$

-0.5

$$\begin{aligned} Pr[B_i] &= Pr[F_{m,i-1}]Pr[L_i] \\ &= \frac{m}{i-1}\frac{1}{n} = \frac{m}{n}\frac{1}{i-1} \end{aligned}$$

(c) Solution: Based on the strategy we are employing, we know that $Pr[B_i] = 0$ for $i \le m$. And also note that the events B_i are mutually exclusive.

$$\begin{aligned} Pr[B] &= \sum_{i=1}^{n} Pr[B_i] \\ &= \sum_{i=1}^{m} Pr[B_i] + \sum_{j=m+1}^{n} Pr[B_j] \\ &= \sum_{i=m=1}^{n} Pr[B_i] = \frac{m}{n} \sum_{i=m+1}^{n} \frac{1}{i-1} \end{aligned}$$

(d) Solution: We use the approximation $H(n) = \sum_{i=1}^{n} \frac{1}{i} = \ln n + \Theta(1)$ and also that $H(x) - \ln x \ge H(y) - \ln y \ge 0$ for $x > y$. Note that $\lambda = \lim_{x\to\infty}(H(x) - \ln x) > 0$ is called Euler–Mascheroni constant. These are properties from the approximation of Harmonic Numbers the derivation of which will be omitted here (It involves using integration to get the approximation as covered in Discussion).

$$\begin{aligned} Pr[B] &= \frac{m}{n} \sum_{i=m+1}^{n} \frac{1}{i-1} \\ &= \frac{m}{n} \sum_{j=m}^{n-1} \frac{1}{j} = \frac{m}{n}(H(n-1) - H(m-1)) \ge \frac{m}{n}\big(\ln(n-1) - \ln(m-1)\big) = \frac{m}{n} \ln \frac{n-1}{m-1} \qquad (1) \\ Pr[B] &\ge \frac{m}{n} \sum_{j=m+1}^{n} \frac{1}{j} = \frac{m}{n}(H(n) - H(m)) \ge \frac{m}{n}\big(\ln n - \ln m\big) = \frac{m}{n} \ln \frac{n}{m} \qquad (2) \end{aligned}$$

Thus from (1) and (2), we have shown that

$$\frac{m}{n} \ln \frac{n}{m} \le Pr[B] \le \frac{m}{n} \ln \frac{n-1}{m-1}.$$

The equality is taken when $m = n$ which should not happen, and when $m, n \to \infty$.

(e) Solution: Let $x = m/n > 0$. So we can take the derivative of the function $f(x) = x \ln \frac{1}{x}$ and set it to 0.

$$f'(x) = -(\ln x + 1) = 0$$
$$\frac{m}{n} = x = \frac{1}{e}$$

When $\frac{m}{n} = \frac{1}{e}$ i.e. $m = \frac{n}{e}$, $\frac{m}{n} \ln \frac{n}{m}$ has a maximum value of $\frac{1}{e}$. A lower bound of $\frac{1}{e}$ is obtained for $Pr[B]$ when $m = \frac{n}{e}$.