CS 70 Discrete Mathematics and Probability Theory
Spring 2016 Walrand and Rao HW 10

Due Thursday April 7 at 10PM

Before you start your homework, write down your team. Who else did you work with on this homework? List names and student ID's. (In case of hw party, you can also just describe the group.) How did you work on this homework? Working in groups of 3-5 will earn credit for your "Sundry" grade.

1. (10 points) Suppose that Allen and Alvin are flipping coins for fun. Allen flips a fair coin k times and Alvin flips $n-k$ times. In total there are n coin flips. Prove that the probability that Allen and Alvin flip the same number of heads is equal to the probability that there are a total of k heads.

   Answer: Let's first figure out the probability that there are a total of k heads. Suppose X as a random variable that denotes the total number of heads in the n flips. It can be seen clearly that X follows a binomial distribution with parameter $p = \frac{1}{2}$ and n. Then the probability that there are a total of k heads can be calculated using binomial distribution: $\mathbf{P}(X = k) = \binom{n}{k}(1/2)^n$.

   Now, let's do mathematical derivation to prove it's same as the probability that Allen and Alvin flip the same number of heads.

$$\begin{aligned}
\mathbf{P}(\text{Allen and Alvin flip the same number of heads}) &= \sum_i \mathbf{P}(A = i, B = i) \\
&= \sum_i \binom{k}{i}(\frac{1}{2})^k \binom{n-k}{i}(\frac{1}{2})^{n-k} \\
&= \sum_i \binom{k}{i}\binom{n-k}{i}(\frac{1}{2})^n \\
&= \sum_i \binom{k}{k-i}\binom{n-k}{i}(\frac{1}{2})^n \\
&= \binom{n}{k}(\frac{1}{2})^n.
\end{aligned}$$

   For the last equality, we used the identity that $\sum_i \binom{k}{k-i}\binom{n-k}{i} = \binom{n}{k}$. You can prove this identity by combinatorial proof. It's clear that on the right-hand side of the equation that we are calculating the total number of ways to choose k items from a group of n items. On the left-hand side, we first partition the whole group into two subgroup with k items and $n-k$ items. In order to choose k items in total from those two groups, we pick $k-i$ items from the first group and i items from the second group (i can be any number between 0 to k.) Therefore, the total number of ways to do so is simply a summation with all possible values of i.

2. Suppose X follows a binomial distribution with parameters n and p.

(a) (8 points) Prove that as x goes from 0 to n, $\mathbf{P}(X=x)$ first increases monotonically. After it reaches its largest value, $\mathbf{P}(X=x)$ then decreases monotonically.

(b) (5 points) What is the value of x that $\mathbf{P}(X=x)$ reaches its largest value?

Answer: In this problem, we will investigate on the ratio $\frac{\mathbf{P}(X=x)}{\mathbf{P}(X=x-1)}$. If the ratio is always greater than 1, we can tell that $\mathbf{P}(X=x)>\mathbf{P}(X=x-1)$, implying that the probability distribution increases monotonically.

$$\begin{aligned}\frac{\mathbf{P}(X=x)}{\mathbf{P}(X=x-1)} &= \frac{\binom{n}{x}p^x(1-p)^{n-x}}{\binom{n}{x-1}p^{x-1}(1-p)^{n-x+1}}\\ &= \frac{\frac{n!}{(n-x)!x!}p^x(1-p)^{n-x}}{\frac{n!}{(n-x+1)!(x-1)!}p^{x-1}(1-p)^{n-x+1}}\\ &= \frac{n-x+1}{x}\frac{p}{1-p}.\end{aligned}$$

Now, let's set $\frac{\mathbf{P}(X=x)}{\mathbf{P}(X=x-1)}\geq 1$:

$$\begin{aligned}\frac{\mathbf{P}(X=x)}{\mathbf{P}(X=x-1)}\geq 1 &\iff \frac{n-x+1}{x}\frac{p}{1-p}\geq 1\\ &\iff (n-x+1)p\geq x(1-p)\\ &\iff np-xp+p\geq x-xp\\ &\iff np+p\geq x\\ &\iff (n+1)p\geq x.\end{aligned}$$

Based on the result above, we proved that from 0 to n, $\mathbf{P}(X=x)$ first increases monotonically. After it reaches its largest value, $\mathbf{P}(X=x)$ then decreases monotonically. The peak point x is therefore:

(a) if $(n+1)p$ is an integer, x equals either $(n+1)p-1$ or $(n+1)p$

(b) if $(n+1)p$ is not an integer, x is an integer between $(n+1)p-1$ or $(n+1)p$

3. Suppose that Allen and Fan are playing a series of games. Assume all games are independent, Allen has probability of p winning each game and Fan has probability of $1-p$ winning. The winner of the series is the first one to win k games.

(a) (8 points) If $k=4$, what's the probability that a total of 7 games are played?

Answer: In order for a total of 7 games, we need to ensure that in the first 6 games, results are 3 wins and 3 losses. Therefore it follows a binomial distribution.

$$\begin{aligned}\mathbf{P}(7\text{ games}) &= \mathbf{P}(3\text{ wins in the first 6 games})\\ &= \binom{6}{3}p^3(1-p)^3.\end{aligned}$$

(b) (5 points) What's the maximum probability from previous part? What's the value of p when probability is maximized?

Answer: Based on question 2, we know that binomial distribution can achieve its maximum point at some points. In order to find the maximum, we will take derivative of the probability above regarding to p:

$$\begin{aligned}\frac{d}{dp}\binom{6}{3}p^3(1-p)^3 &= 20[3p^2(1-p)^3 - 3p^3(1-p)^2] \\ &= 60p^2(1-p)^2(1-2p).\end{aligned}$$

In order to get the maximum, the derivative will be 0 when $p = 1/2$. You can also verify that it's the maximum by taking the second derivative. However, it's not required in this question if you show that it must be the maximum point based on question 2. You can also directly use the formula derived in the second part of question 2. It will give you the same result.

(c) (10 points) If $k = 3$, what's the expected number of games that are played?

Answer: When $k = 3$, Allen and Fan need to play 3 games minimum to decide a winner. In the meantime, they only need to play 5 games maximum. If they play the 6th game, there must be a person who has already won 3 games. Let's define X as the expected number of games that they need to play. So X can only take value 3, 4 and 5. We will calculate the probability for each event and sum those up to calculate the expectation.

When X is 3, either Allen or Fan wins the whole three games:

$$\mathbf{P}(X = 3) = p^3 + (1-p)^3$$

.

When X is 4, either Allen or Fan has 2 wins in the first three games:

$$\begin{aligned}\mathbf{P}(X = 4) &= \mathbf{P}(X = 4, \text{Allen has 2 wins in the first 3 games}) \\ &\quad + \mathbf{P}(X = 4, \text{Fan has 2 wins in the first 3 games}) \\ &= \binom{3}{2}p^2(1-p)p + \binom{3}{2}p(1-p)^2(1-p).\end{aligned}$$

When X is 5, each play must have 2 wins in the first four games:

$$\mathbf{P}(X = 5) = \binom{4}{2}p^2(1-p)^2$$

.

Now let's use the definition of expectation to get the answer:

$$\begin{aligned}\mathbf{E}(X) &= 3\mathbf{P}(X = 3) + 4\mathbf{P}(X = 4) + 5\mathbf{P}(X = 5) \\ &= 3[p^3 + (1-p)^3] + 4[\binom{3}{2}p^2(1-p)p + \binom{3}{2}p(1-p)^2(1-p)] + 5[\binom{4}{2}p^2 + (1-p)^2] \\ &= 3[p^3 + (1-p)^3] + 12p(1-p)[p^2 + (1-p)^2] + 30[p^2(1-p)^2].\end{aligned}$$

(d) (5 points) What's the maximum number from previous part? What's the value of p at that point?

Answer: Similar to part b, take the derivative of the expectation above regarding to p. Set the result to 0 shows the maximum is attained when $p = 1/2$.

4. (5 points) Two faulty machines, M_1 and M_2, are repeatedly run synchronously in parallel (i.e., both machines execute one run, then both execute a second run, and so on). On each run, M_1 fails with probability p_1 and M_2 fails with probability p_2, all failure events being independent. Let the random variables X_1, X_2 denote the number of runs until the first failure of M_1, M_2 respectively; thus X_1, X_2 have geometric distributions with parameters p_1, p_2 respectively. Let X denote the number of runs until the first failure of *either* machine. Show that X also has a geometric distribution, with parameter $p_1 + p_2 - p_1 p_2$.

Answer: We have that $X_1 \sim \text{Geom}(p_1)$ and $X_2 \sim \text{Geom}(p_2)$. Also, X_1, X_2 are independent r.v.'s. We also use the following definition of the minimum:

$$\min(x,y) = \begin{cases} x & \text{if } x \le y; \\ y & \text{if } x > y. \end{cases}$$

Now, for all $k \in \{1,2,\dots\}$, $\min(X_1,X_2) = k$ is equivalent to $(X_1 = k) \cap (X_2 \ge k)$ or $(X_2 = k) \cap (X_1 > k)$. Hence,

$$\begin{aligned} \mathbf{P}(X = k) &= \mathbf{P}(\min(X_1,X_2) = k) \\ &= \mathbf{P}((X_1 = k) \cap (X_2 \ge k)) + \mathbf{P}((X_2 = k) \cap (X_1 > k)) \\ &= \mathbf{P}(X_1 = k) \cdot \mathbf{P}(X_2 \ge k) + \mathbf{P}(X_2 = k) \cdot \mathbf{P}(X_1 > k) && \text{since } X_1, X_2 \text{ are independent} \\ &= ((1-p_1)^{k-1} p_1)(1-p_2)^{k-1} + ((1-p_2)^{k-1} p_2)(1-p_1)^k && \text{since } X_1, X_2 \text{ are geometric} \\ &= ((1-p_1)(1-p_2))^{k-1} (p_1 + p_2(1-p_1)) \\ &= (1 - p_1 - p_2 + p_1 p_2)^{k-1} (p_1 + p_2 - p_1 p_2). \end{aligned}$$

But this final expression is precisely the probability that a geometric r.v. with parameter $p_1 + p_2 - p_1 p_2$ takes the value k. Hence $X \sim \text{Geom}(p_1 + p_2 - p_1 p_2)$, and $\mathbb{E}(X) = \frac{1}{p_1 + p_2 - p_1 p_2}$.

An alternative, slightly cleaner approach is to work with the *tail probabilities* of the geometric distribution, rather than with the usual point probabilities as above. In other words, we can work with $\mathbf{P}(X \ge k)$ rather than with $\mathbf{P}(X = k)$; clearly the values $\mathbf{P}(X \ge k)$ specify the values $\mathbf{P}(X = k)$ since $\mathbf{P}(X = k) = \mathbf{P}(X \ge k) - \mathbf{P}(X \ge (k+1))$, so it suffices to calculate them instead. We then get the following argument:

$$\begin{aligned} \mathbf{P}(X \ge k) &= \mathbf{P}(\min(X_1,X_2) \ge k) \\ &= \mathbf{P}((X_1 \ge k) \cap (X_2 \ge k)) \\ &= \mathbf{P}(X_1 \ge k) \cdot \mathbf{P}(X_2 \ge k) && \text{since } X_1, X_2 \text{ are independent} \\ &= (1-p_1)^{k-1} (1-p_2)^{k-1} && \text{since } X_1, X_2 \text{ are geometric} \\ &= ((1-p_1)(1-p_2))^{k-1} \\ &= (1 - p_1 - p_2 + p_1 p_2)^{k-1}. \end{aligned}$$

This is the tail probability of a geometric distribution with parameter $p_1 + p_2 - p_1 p_2$, so we are done.

5. In this question, we will use another approach to calculate the expectation of the geometric distribution.

   (a) (8 points) If $S = \sum_{i=1}^{\infty} ir^{i-1} = 1 + 2r + 3r^2 + \ldots$ where $-1 < r < 1$, prove

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

   (Hint: what is $S - rS$?)

   Answer:

   $$\begin{aligned} S - rS &= \sum_{i=1}^{\infty} ir^{i-1} - \sum_{i=1}^{\infty} ir^i \\ &= (1 + 2r + 3r^2 + \ldots) - (r + 2r^2 + 3r^3 + \ldots) \\ &= 1 + r + r^2 + \ldots \\ &= \frac{1}{1-r} \\ \Longrightarrow S &= \frac{1}{(1-r)^2}. \end{aligned}$$

   (b) (8 points) Given a random variable X having the geometric distribution with parameter p where $0 < p \leq 1$, *i.e.*,

   $$\mathbf{P}(X = i) = (1-p)^{i-1}p \qquad \text{for } i = 1, 2, 3, \ldots,$$

   use Equation (1) to prove $\mathbf{E}(X) = \frac{1}{p}$.

   Answer: Assume $r = 1 - p$,

   $$\begin{aligned} \sum_{i=1}^{\infty} i(1-p)^{i-1}p &= p\sum_{i=1}^{\infty} i(1-p)^{i-1} \\ &= p\sum_{i=1}^{\infty} ir^{i-1} \\ &= p\left(\frac{1}{(1-r)^2}\right) \\ &= \frac{p}{(1-(1-p))^2} \\ &= \frac{1}{p}. \end{aligned}$$

6. Suppose Professor Walrand now offers you a game: you bet any amount of homework points in this homework, then you flip a coin and if it comes up heads you win that amount, and if it comes up tails you lose that amount. Suppose you follow this strategy: you start with a bet of 1 homework point, and if you lose you increase your bet to 2 homework points, and again if you lose you double your bet to 4 homework points, and so on. As soon as you win, you take your winnings and you go out (i.e. you bet 0 homework points for the next rounds). So in round n you bet 2^{n-1} homework points if you have lost all the previous rounds, and you bet 0 homework points if you have won any of the previous rounds.

(a) (5 points) What is your net winnings (i.e. subtract your losses from your wins) if you win after the n-th coin flip.

Answer: You win $2^{n-1} - 2^{n-2} - \ldots - 2^0 = 1$ homework point.

(b) (5 points) What is the expected value of your winnings on round i?

Answer: On round i, you bet 2^{i-1} with some probability p and you bet 0 homework points with probability $1-p$. So your expected return is

$$2^{i-1} \times \frac{p}{2} - 2^{i-1} \times \frac{p}{2} + 0 \times (1-p) = 0$$

(c) (5 points) What is you expected net winnings after round n?

Answer: If you sum all of the winnings for every round up to round n, you get your expected winnings which is 0, by linearity of expectation.

7. (10 points) Suppose we put a total of r balls, independently one at a time, in k boxes. Probability that each ball goes into box i is p_i, such that $\sum_{i=1} p_i = 1$. Collision occurs when we put a ball in a nonempty box. There are 9 collisions if a bin is filled with 10 balls. What's the expected number of collisions.

Answer: Let's define N_i be the number of balls in box i ($i = 1, \ldots, k$), X be the number of collisions we have.

So in each box, we have N_i balls. N_i follows a binomial distribution with parameter $n = r$ and $p = p_i$. So $\mathbf{E}(N_i) = rp_i$.If $N_i > 0$, then the number of collisions in the box is equal to $N_i - 1$. If $N_i = 0$, the number of collisions is 0. So we can define an indicator variable I_i where $I_i = 1$ if $N_i = 0$, and $I_i = 0$ if $N_i > 0$. So $\mathbf{E}(I_i) = \mathbf{P}(I_i = 1) = \mathbf{P}(N_i = 0) = (1-p_i)^r$. Therefore we can write the number of collisions in a box is equal to $N_i - 1 + I_i$.

Therefore, we can use linearity of expectation to calculate X.

$$\begin{aligned}
\mathbf{E}(X) &= \mathbf{E}(\sum_{i=1}^{k} N_i - 1 + I_i) \\
&= \sum_{i=1}^{k} (\mathbf{E}(N_i) - 1 + \mathbf{E}(I_i)) \\
&= \sum_{i=1}^{k} (rp_i - 1 + (1-p_i)^r) \\
&= \sum_{i=1}^{k} rp_i - \sum_{i=1}^{k} 1 + \sum_{i=1}^{k} (1-p_i)^r \\
&= r(\sum_{i=1}^{k} p_i) - k + \sum_{i=1}^{k} (1-p_i)^r \\
&= r - k + \sum_{i=1}^{k} (1-p_i)^r
\end{aligned}$$

We use the fact that all the probability p_i sums up to 1 in the second to last equality.

8. Suppose you roll a fair die n times. What is the expectations of each of the following random variables?

   (a) (5 points) A is the random variable that denotes the sum of the numbers in those rolls.

   Answer: Let X_i be the value of the i-th roll. Since the die is rolled independently. All X_i follows an identical and independent distributed Uniform distribution, in the range of (1,2,...,6).

$$\mathbf{E}(A) = \sum_{i=1}^{n} \mathbf{E}(X_i) = \sum_{i=1}^{n} \frac{1+2+3+4+5+6}{6} = 3.5 \cdot n.$$

   (b) (5 points) B is the random variable that denotes the maximum number in the those rolls.

   Answer: Since B denotes the maximum number: $B = \max(X_1, X_2, ..., X_n)$.
   Since B is the maximum among variable Xs. So we can get the following:

$$\mathbf{P}(B \le b) = \mathbf{P}(X_1 \le b) \cdot \mathbf{P}(X_2 \le b) ... \mathbf{P}(X_n \le b) = (\frac{b}{6})^n.$$

   Therefore,

$$\mathbf{P}(B = b) = \mathbf{P}(B \le b) - \mathbf{P}(B \le b-1) = (\frac{b}{6})^n - (\frac{b-1}{6})^n.$$

   Expected value can be calculated using its definition:

$$\mathbf{E}(B) = \sum_{i=1}^{6} i \cdot P(B = i) = \sum_{i=1}^{6} i((\frac{i}{6})^n - (\frac{i-1}{6})^n).$$

   Another way to solve this is to use tail sum.

$$\begin{aligned}\mathbf{E}(B) &= \sum_{i=1}^{6} i \cdot P(B = i) = \sum_{i=1}^{6} P(B \ge i) \\ &= \sum_{i=1}^{6} 1 - P(B \le i) = \sum_{i=1}^{6} 1 - (\frac{i}{6})^n \\ &= 6 - \sum_{i=1}^{6} (\frac{i}{6})^n.\end{aligned}$$

   Those two answers should give you the same results.

   (c) (8 points) C is the random variable that denotes the sum of the largest two numbers in the first three rolls.

   Answer: Let X_1, X_2, X_3 be the value of first 3 rolls, and X_{min} be the minimum of the first 3 rolls. We can write C in terms of those four variables:

$$\mathbf{E}(C) = \mathbf{E}(X_1 + X_2 + X_3 - X_{min}) = \mathbf{E}(X_1 + X_2 + X_3) - \mathbf{E}(X_{min})$$

   Now, we will use the same method above to find out the expectation of the minimum.

$$\mathbf{P}(X_{min} \ge m) = (\frac{6-m}{6})^3.$$

So that

$$\mathbf{P}(X_{min}=m)=\mathbf{P}(X_{min}\geq m-1)-\mathbf{P}(X_{min}\geq m)=(\frac{7-m}{6})^3-(\frac{6-m}{6})^3.$$

$$\mathbf{E}(X_{min})=\sum_{i=1}^{6}i((\frac{7-i}{6})^3-(\frac{6-i}{6})^3).$$

After you calculate the expectation of the minimum, you can get the sum of the largest two easily. You can also use tail sum to calculate it. The calculation is very similar to part b so it's omitted.

(d) (8 points) D is the random variable that denotes the number of multiples of three in those rolls.

Answer: It can be seen that D follows a binomial distribution with n, $p=2/6$. So $\mathbf{E}(D)=np=n/3$.

(e) (8 points) E is the random variable that denotes the number of faces which fail to appear in those rolls.

Answer: Let I_i be an indicator that takes value 1 if face i does not appear.
So that $\mathbf{E}(I_i)=\mathbf{P}(I_i=1)=(5/6)^n$.

$$\mathbf{E}(E)=\sum_{i=1}^{6}\mathbf{E}(I_i)=\sum_{i=1}^{6}(5/6)^n=6\cdot(5/6)^n.$$

(f) (5 points) F is the random variable that denotes the number of distinct faces that appear in those rolls.

Answer: Very similar to part (e), Let I_i be an indicator that takes value 1 if face i appears.
So that $\mathbf{E}(I_i)=\mathbf{P}(I_i=1)=1-(5/6)^n$.

$$\mathbf{E}(F)=\sum_{i=1}^{6}\mathbf{E}(I_i)=\sum_{i=1}^{6}[1-(5/6)^n]=6\cdot[1-(5/6)^n].$$

9. (10 points) Suppose you flip a biased coin until two of the most recent three flips are heads. The probability that you see a head is p. X is the random variable that denotes the number of total flips. If the first two flips are heads, then $X=2$. Find the expected value of X.

Answer: Let's define $\mathbf{E}(X|H)$ as the expected remaining tosses given that the first toss is a head. In both cases, you need to toss the coin again. So,

$$\mathbf{E}(X)=p(1+\mathbf{E}(X|H))+(1-p)(1+\mathbf{E}(X|T)).$$

However, you can replace $\mathbf{E}(X|T)$ with $E(X)$. Why? If you treat two heads in the recent three flips as a series. If the first toss is tail, you need to start a new series again. So, we can simplify that into

$$\mathbf{E}(X)=p(1+\mathbf{E}(X|H))+(1-p)(1+\mathbf{E}(X))=1+p\mathbf{E}(X|H)+(1-p)\mathbf{E}(X).$$

Now, let's consider the case $\mathbf{E}(X|H)$. We can still use the first formula to include the outcome of next toss:

$$\mathbf{E}(X|H)=p(1+\mathbf{E}(X|HH))+(1-p)(1+\mathbf{E}(X|HT)).$$

At this moment, we are done with $\mathbf{E}(X|HH)$, since we got two heads! So $\mathbf{E}(X|HH) = 0$. Remember we define the conditional expectation as the remaining tosses. Let's simplify $\mathbf{E}(X|H)$ before we move on:

$$\mathbf{E}(X|H) = 1 + (1-p)\mathbf{E}(X|HT).$$

Finally, we will consider the case $\mathbf{E}(X|HT)$. Similarly we will break it down into two cases:

$$\mathbf{E}(X|HT) = p(1 + \mathbf{E}(X|HTH)) + (1-p)(1 + \mathbf{E}(X|HTT)).$$

Let's understand the equation above. First, we are also done with $\mathbf{E}(X|HTH)$ since there are two heads in the recent three tosses. $\mathbf{E}(X|HTH) = 0$. Next, $\mathbf{E}(X|HTT)$ implies that we need to start a new series again from the next toss so $\mathbf{E}(X|HTT) = E(X)$. As a result, $\mathbf{E}(X|HT)$ is transformed into:

$$\mathbf{E}(X|HT) = p + (1-p)(1 + \mathbf{E}(X)) = 1 + (1-p)\mathbf{E}(X).$$

Now, we go back to $\mathbf{E}(X)$, so that we can put back everything together and solve $\mathbf{E}(X)$.

$$\begin{aligned}
\mathbf{E}(X) &= 1 + p\mathbf{E}(X|H) + (1-p)\mathbf{E}(X) \\
&= 1 + p[1 + (1-p)\mathbf{E}(X|HT)] + (1-p)\mathbf{E}(X) \\
&= 1 + p\{1 + (1-p)[1 + (1-p)\mathbf{E}(X)]\} + (1-p)\mathbf{E}(X) \\
&= 1 + p + p(1-p)[1 + (1-p)\mathbf{E}(X)] + (1-p)\mathbf{E}(X) \\
&= 1 + p + p(1-p) + p(1-p)^2\mathbf{E}(X) + (1-p)\mathbf{E}(X) \\
(1 - p(1-p)^2 - (1-p))\mathbf{E}(X) &= 1 + p + p(1-p) \\
\mathbf{E}(X) &= \frac{1 + 2p - p^2}{2p^2 - p^3}
\end{aligned}$$

10. Suppose you flip a biased coin n times ($n > 4$). The probability that you see a head is p. Let's define the concept of a *run* of three heads. It can be the first four flips in the pattern $HHHT$, the last four flips in the pattern $THHH$, or elsewhere in the flips in the pattern $THHHT$. Let $R(3,n)$ denotes the number of runs of three heads in the n trials.

    (a) (8 points) Find the expectation of $R(3,n)$.

    Answer: Let's define I_i as an indicator variable that denotes a head run of exactly $m = 3$ heads starting on the i-th trial.

$$\mathbf{P}(I_i = 1) = \begin{cases} p^3(1-p) & \text{if } i = 1, \text{ or } i = n-2 \\ p^3(1-p)^2 & \text{if } 2 \le i \le n-3 \\ 0 & \text{if } n-1 \le i \le n \end{cases}$$

    Therefore, we can simply write $R(3,n)$ as a sum of n indicator variables.

$$R(3,n) = I_1 + I_2 + \ldots + I_{n-3} + I_{n-2} + I_{n-1} + I_n$$

Please note that according to our piece-wise function above, we can partition all n indicator variables into three groups. I_1 and I_{n-2} belong to the first category with probability p^3q. I_{n-1} and I_{n-2} belong to the third category with probability 0. All the rest items, from I_2 to I_{n-3}, all belong to the second category.

$$\begin{aligned}
\mathbf{E}(R(3,n)) &= \mathbf{E}(I_1 + I_2 + \ldots + I_{n-3} + I_{n-2} + I_{n-1} + I_n) \\
&= \mathbf{E}(I_1) + \mathbf{E}(I_2) + \ldots + \mathbf{E}(I_{n-3}) + \mathbf{E}(I_{n-2}) + \mathbf{E}(I_{n-1}) + \mathbf{E}(I_n) \\
&= \mathbf{P}(I_1) + \mathbf{P}(I_2) + \ldots + \mathbf{P}(I_{n-3}) + \mathbf{P}(I_{n-2}) + \mathbf{P}(I_{n-1}) + \mathbf{P}(I_n) \\
&= 2 \cdot (p^3(1-p)) + 2 \cdot 0 + (n-4) \cdot p^3(1-p)^2 \\
&= 2p^3(1-p) + (n-4)p^3(1-p)^2.
\end{aligned}$$

The third equality uses the fact that the expectation of a Bernoulli variable I is the same as $\mathbf{P}(I = 1)$.

(b) (8 points) Let's now define $R(m,n)$ as the number of heads of length m in n flips, similarly for $1 \le m \le n$. Find the expectation of $R(m,n)$.

Answer: We will generalize our answer to part a. Let's define I_i as an indicator variable that denotes a head run of exactly m heads starting on the i-th trial.

$$\mathbf{P}(I_i = 1) = \begin{cases} p^m(1-p) & \text{if } i = 1, \text{ or } i = n-m+1 \\ p^m(1-p)^2 & \text{if } 2 \le i \le n-m \\ 0 & \text{if } n-m+2 \le i \le n \end{cases}$$

Again we will write $R(m,n)$ as a sum of n indicator variables.

$$R(m,n) = I_1 + I_2 + \ldots + I_{n-m} + I_{n-m+1} + I_{n-m+2} + \ldots + I_n$$

Similarly, we can partition all n indicator variables into three groups. I_1 and I_{n-m+1} belong to the first category with probability $p^m q$. Indicators from I_{n-m+2} to I_n belong to the third category with probability 0 since none of them will make a run. All the rest items, from I_2 to I_{n-m}, all belong to the second category.

$$\begin{aligned}
\mathbf{E}(R(m,n)) &= \mathbf{E}(I_1 + I_2 + \ldots + I_{n-m} + I_{n-m+1} + I_{n-m+2} + \ldots + I_n) \\
&= \mathbf{E}(I_1) + \mathbf{E}(I_2) + \ldots + \mathbf{E}(I_{n-m}) + \mathbf{E}(I_{n-m+1}) + \mathbf{E}(I_{n-m+2}) + \ldots + \mathbf{E}(I_n) \\
&= \mathbf{P}(I_1) + \mathbf{P}(I_2) + \ldots + \mathbf{P}(I_{n-m}) + \mathbf{P}(I_{n-m+1}) + \mathbf{P}(I_{n-m+2}) + \ldots + \mathbf{P}(I_n) \\
&= 2 \cdot (p^m(1-p)) + (m-1) \cdot 0 + (n-m-1) \cdot p^m(1-p)^2 \\
&= 2p^m(1-p) + (n-m-1)p^m(1-p)^2.
\end{aligned}$$

However the answer above is only true when $m < n$. We also need to consider the case for $m = n$. In this case, we need to make sure all n flips are heads in order to make it a run of m. So $\mathbf{E}(R(m,n)) = p^n$.

(c) (8 points) Let $R(n)$ be the total number of non-overlapping head runs in n trials, counting runs of any length between 1 and n. Find the expectation of $R(n)$ by using the result of part b.

Answer: We will find $\mathbf{E}(R(n))$ by taking the expectation of the sum of $R(m,n)$ over all m.

$$\begin{aligned}
\mathbf{E}(R(n)) &= \mathbf{E}(R(1,n)+R(2,n)+...+R(n,n)) = \mathbf{E}(\sum_{m=1}^{n} R(m,n)) \\
&= \mathbf{E}(\sum_{m=1}^{n-1} R(m,n)) + \mathbf{E}(R(n,n)) = \sum_{m=1}^{n-1} \mathbf{E}(R(m,n)) + \mathbf{E}(R(n,n)) \\
&= \sum_{m=1}^{n-1} 2p^m(1-p) + (n-1)p^m(1-p)^2 - mp^m(1-p)^2 + p^n \\
&= [2(1-p)+(n-1)(1-p)^2] \sum_{m=1}^{n-1} p^m - (1-p)^2 \sum_{m=1}^{n-1} mp^m + p^n \\
&= [2(1-p)+(n-1)(1-p)^2] \frac{p-p^n}{1-p} - (1-p)^2 [\frac{p-p^{n+1}-np^{n+1}+np^{n+2}}{(1-p)^2} - np^n] + p^n \\
&= [2+(n-1)(1-p)](p-p^n) - [p-p^{n+1}-np^{n+1}+np^{n+2}-(1-p)^2 np^n] + p^n \\
&= [2+(n-1)(1-p)](p-p^n) - [p-(n+1)p^{n+1}+np^{n+2}-(1-2p+p^2)np^n] + p^n \\
&= (2+n-np-1+p)(p-p^n) - [p-(n+1)p^{n+1}+np^{n+2}-(1-2p+p^2)np^n] + p^n \\
&= [(n+1)-(n-1)p](p-p^n) - [p+(n-1)p^{n+1}-np^n] + p^n \\
&= [(n+1)-(n-1)p](p-p^n) - p-(n-1)p^{n+1}+np^n+p^n \\
&= [(n+1)-(n-1)p](p-p^n) - p-(n-1)p^{n+1}+(n+1)p^n \\
&= (n+1)p-(n-1)p^2-(n+1)p^n+(n-1)p^{n+1}-p-(n-1)p^{n+1}+(n+1)p^n \\
&= np-(n-1)p^2 \\
&= p+(n-1)p(1-p). \text{You can derive it by yourself but those two are equivalent}
\end{aligned}$$

Please note that we replace the summation using the equation below.

$$\sum_{i=1}^{n} t^i = \frac{t-t^{n+1}}{1-t}$$

We could use it to solve $\sum_{i=1}^{n} it^i$. Let $S = \sum_{i=1}^{n} it^i$.

$S = 1t^1 + 2t^2 + 3t^3 + ... + nt^n$

$tS = \quad 1t^2 + 2t^3 + ... + (n-1)t^n + nt^{n+1}$

So,

$$S(1-t) = S - tS = t^1 + t^2 + t^3 + ... + t^n - nt^{n+1}$$

Therefore,

$$S = \sum_{i=1}^{n} it^i = \frac{\sum_{i=1}^{n} t^i + nt^{n+1}}{1-t} = \frac{t-t^{n+1}}{(1-t)^2} - \frac{nt^{n+1}}{1-t} = \frac{t-t^{n+1}-nt^{n+1}(1-t)}{(1-t)^2} = \frac{t-t^{n+1}-nt^{n+1}+nt^{n+2}}{(1-t)^2}.$$

Here is another to solve the sum $S = \sum_{i=1}^{n} it^i$. Consider

$$\sum_{i=0}^{n} e^{ik} t^i = \sum_{i=0}^{n} (te^k)^i = \frac{1-(te^k)^{n+1}}{1-te^k}$$

Differentiate with respect to k:

$$\sum_{i=0}^{n} ie^{ik}t^i = \frac{-(n+1)(te^k)^n(te^k)(1-te^k)-(1-(te^k)^{n+1})(-te^k)}{(1-te^k)^2}$$
$$= \frac{(1-(te^k)^{n+1})(te^k)-(n+1)(te^k)^{n+1}(1-te^k)}{(1-te^k)^2}$$

Set $k=0$:

$$\sum_{i=0}^{n} it^i = \frac{(1-t^{n+1})t-(n+1)t^{n+1}(1-t)}{(1-t)^2}$$
$$= \frac{t-t^{n+2}-(n+1)t^{n+1}+(n+1)t^{n+2}}{(1-t)^2} = \frac{t-(n+1)t^{n+1}+nt^{n+2}}{(1-t)^2}$$

It makes no difference if the sum starts from $i=1$ since the $i=0$ term is 0 anyway.

(d) (8 points) Find the expectation of $R(n)$ another way by considering for each $1 \le j \le n$ the number of runs that start on the jth trial. Check that the two methods give the same answer.

Answer: Let's define I_i as an indicator variable that denotes a success run of any length starting on the i-th trial.

$$\mathbf{P}(I_i = 1) = \begin{cases} p & \text{if } i=1 \\ p(1-p) & \text{if } 2 \le i \le n \end{cases}$$

We can write $R(n)$ as a sum of n indicator variables.

$$R(n) = I_1 + I_2 + \ldots + I_n$$

$$\begin{aligned}\mathbf{E}(R(n)) &= \mathbf{E}(I_1 + I_2 + \ldots + I_n) \\ &= \mathbf{E}(I_1) + \sum_{i=2}^{n} \mathbf{E}(I_i) \\ &= p + (n-1)p(1-p).\end{aligned}$$

You can verify that the answer here is the same as the answer we get in part 3. However, it's MUCH MUCH MUCH easier to do it this way.