Dennis Kuzminer
CSCI-UA 310-001 PS6

1.

a. For any given pair, the probability of $X_i = X_j$ is 1/m. Considering all combinations, we take the union of the sets. We can see that from rewriting the union bound, the total number of $X_i = X_j$ is n(n-1)/2 (n-i combos for X_i, summing yeilds n(n-1)/2). Combining the two, we can bound the overall probability of a collision by $p_{n,m} \le \frac{n(n-1)}{2m}$.

b. 1 - p = q, where q is Pr[$X_i \ne X_j$]. We can enumerate all options and possibilities of X_i and X_j being distinct $\frac{m-0}{m} * \frac{m-1}{m} * \ldots * \frac{m-n}{m}$ (Mutual independence)→ $\prod_{i=1}^{n} \frac{m-(i-1)}{m} \to \prod_{i=1}^{n} \frac{m}{m} - \frac{(i-1)}{m} \to \prod_{i=1}^{n} 1 - \frac{(i-1)}{m} \to \prod_{i=1}^{n} (1 - \frac{(i-1)}{m}) = 1 - p.$

c. $1 + x \le e^x \to 1 - e^x \le -x \to$ We know that $p_{n,m} \le \frac{n(n-1)}{2m} \to -p_{n,m} \ge -\frac{n(n-1)}{2m} \to$ Stubstitute the first inequality with the second, $x = -p_{n,m}$. This results in $1 - e^{-\frac{n(n-1)}{2m}} \le p_{n,m}$.

d. Prove that $-\frac{n(n-1)}{2m} \le ln(.5)$, as $p_{n,m} \ge 1 - e^{ln(.5)} \to p_{n,m} \ge 1 - .5 \to p_{n,m} \ge .5$
Given that $n \ge \sqrt{2ln(2)m} + 1 \to n(n-1) \ge (n-1)^2 \ge 2ln(2)m \to$ $\frac{n(n-1)}{2m} \ge \frac{(n-1)^2}{2m} \ge ln(2) \to -\frac{n(n-1)}{2m} \le -ln(2) \to -\frac{n(n-1)}{2m} \le ln(1/2)$. Because this exponent is bounded, we can substitute and see that $p_{n,m} \ge .5$.

Dennis Kuzminer
CSCI-UA 310-001 PS6

2. $E[S^2] \rightarrow E[(X_1 + ... + X_n)^2] \rightarrow (\sum_{i=1}^{n} X_i)^2 = (\sum_{i=1}^{n} X_i)(\sum_{j=1}^{n} X_j) = \sum_{i=1}^{n} X_i^2 + \sum_{i \neq j} X_i X_j$ (Example 30 in Probability Primer) $\rightarrow \sum_{i=1}^{n} E[X_i^2] + \sum_{i \neq j} E[X_i X_j]$. This is because of linearity of expectation.

   To find $\sum_{i=1}^{n} E[X_i^2]$, we must also consider that the variables are all *independently* and *uniformly* distributed; therefore, regardless of the value of *i,* all $E[X_i]$ will have the same values and all $E[X_i^2]$ will have the same value. Meaning, $\sum_{i=1}^{n} E[X_i^2] = nE[X_1^2]$. X_1^2 is distributed on the set of the squares of the original set = $\{0, 1, 1, 4, 4, 9, 9\}$. The mean/expected value of this set is 4 ([0+1+1+4+4+9+9]/7). This shows that $\sum_{i=1}^{n} E[X_i^2] = 4n$.

   To find $\sum_{i \neq j} E[X_i X_j]$, we use linearity of expectation once more. $E[X_i X_j] = E[X_i]E[X_j] \rightarrow$ $E[X_i] = (-3 + -2 + -1 + 0 + 1 + 2 + 3)/7 = 0$.

   This means $\sum_{i \neq j} E[X_i X_j] = E[X_i X_j] = E[X_i]E[X_j] = 0$.

   Finally, we combine the terms to get $E[S^2] = 4n + 0$ or just $4n$.

Dennis Kuzminer
CSCI-UA 310-001 PS6

3. $E[k] = E[X],\ X = X_1 + ... + X_k \rightarrow$ Each X has the same distribution, so for any X, the probability that a head will appear is p = .5. So, the *expected* number of coin tosses you make to get the ith head after you have already gotten i-1 heads is $\frac{1}{p} = 2.$ $E[X] = E[X_1] + ... + E[X_k] \rightarrow$ $E[X] = \frac{1}{p} + ... + \frac{1}{p}$, k times. Meaning that $E[k] = \frac{k}{p}$. In this case, $p = .5,\ E[k] = \frac{k}{p} = 2k.$

Dennis Kuzminer
CSCI-UA 310-001 PS6

4. Bob should choose the guessed number (or a number out of a set of numbers) by enumerating all the possibilities of 2 dice summing to ℓ and picking the number appearing in the most possibilities. For example, $\ell = 4$, possibilities: { {1, 3} {2, 2} {3, 1} }. The optimal choice would be 1 or 3 because they appear in 2 of three possibilities while 2 only appears in 1. Pr[Z = ℓ] = number of possibilities where sum = ℓ / total possibilities = 36
Case where $\ell = 2$: Bob should choose 1, and the probability that he win, p = 1. E[W | Z = 2] = 2 because he should double down, as p > q, where q = the probability he will lose. We find E[W | Z = l] by computing (1)p + (-1)q to find expected earnings. If this result is positive, this will indicated that Bob is expected to earn money, so he should double down to double his expected earnings. Pr[Z = 2] = 1/36
Case $\ell = 3$: Bob should choose 1 or 2, and p = 1. E[W | Z = 3] = 2 because he should double down, as p > q. Pr[Z = 3] = 2/36
Case $\ell = 4$: Bob should choose 1 or 3, and p = 2/3. E[W | Z = 4] = 1/3 → 2/3 because he should double down, as p > q. Pr[Z = 4] = 3/36
Case $\ell = 5$: Bob should choose 1, 2, 3, or 4, and p = 2/4 = 1/2. E[W | Z = 5] = 0. It does not matter if Bob doubles down because p = q. Pr[Z = 5] = 4/36
Case $\ell = 6$: Bob should choose 1, 2, 4, or 5, and p = 2/5. E[W | Z = 6] = -1/5. Bob should not double down, as he is expected to lose because p < q. Pr[Z = 6] = 5/36
Case $\ell = 7$: Bob should choose 1, 2, 3, 4, 5, or 6, and p = 2/6 = 1/3. E[W | Z = 7] = -1/3. Bob should not double down, as he is expected to lose because p < q. Pr[Z = 7] = 6/36
Case $\ell = 8$: Bob should choose 2, 3, 5, or 6, and p = 2/5. E[W | Z = 8] = -1/5. Bob should not double down, as he is expected to lose because p < q. Pr[Z = 8] = 5/36
Case $\ell = 9$: Bob should choose 3, 4, 5, or 6, and p = 2/4 = 1/2. E[W | Z = 9] = 0. It does not matter if Bob doubles down because p = q. Pr[Z = 9] = 4/36
Case $\ell = 10$: Bob should choose 4 or 6, and p = 2/3. E[W | Z = 10] = 1/3 → 2/3 because he should double down, as p > q. Pr[Z = 10] = 3/36
Case $\ell = 11$: Bob should choose 5 or 6, and p = 1. E[W | Z = 11] = 2 because he should double down, as p > q. Pr[Z = 11] = 2/36
Case $\ell = 12$: Bob should choose 6, and p = 1. E[W | Z = 12] = 2 because he should double down, as p > q. Pr[Z = 12] = 1/36
By the law of total expectation

$$\mathrm{E}[W] = \sum_{\ell=2}^{12} \mathrm{E}[W \mid Z = \ell] \Pr[Z = \ell].$$

Summing, we get
2(1/36)+2(2/36)+(2/3)(3/36)+0(4/36)+(-1/5)(5/36)+(-1/3)(6/36)+(-1/5)(5/36)+0(4/36)+(2/3)(3/36)+2(2/36)+2(1/36) = 1/3

Dennis Kuzminer
CSCI-UA 310-001 PS6

5.

a. $E[X+X] = E[X] + E[X] = 2E[X] \rightarrow E[X] = \frac{1}{n}\sum_{i=1}^{n} i = \frac{1}{n} * \frac{n(n+1)}{2} \rightarrow 2 * \frac{1}{n} * \frac{n(n+1)}{2} \rightarrow$
$E[X+X] = n + 1 \rightarrow$ When n = 10, $E[X+X] = 11$

b. $E[XX] = E[X]E[X] \rightarrow E[X] = \frac{1}{n}\sum_{i=1}^{n} i = \frac{1}{n} * \frac{n(n+1)}{2} \rightarrow E[X]E[X] = (.5(n+1))^2 \rightarrow$
When n = 10, $E[XX] = 30.25$

Dennis Kuzminer
CSCI-UA 310-001 PS6

6. $E[T] = \sum_{i=1}^{n} E[T \mid X = i]\, Pr[X = i] \rightarrow$ The probability of X = i is uniform and will always be = 1/10 in this case $\rightarrow$ $E[T] = \frac{1}{n}\sum_{i=1}^{n} E[T \mid X = i] \rightarrow$ If i = 1, we can say that we will expect 10 tries before we reach at most 1, as the distribution is uniform. Similarly, if we spin a 2, we expect it to take 5 tries to get to a number of at most 2 after (uniform and independent spins). Therefore, we can see that $\sum_{i=1}^{n} E[T \mid X = i] = \sum_{i=1}^{n} \frac{n}{i} \rightarrow$ Substituting back $\rightarrow$ $E[T] = \frac{1}{n}\sum_{i=1}^{n} \frac{n}{i} \rightarrow \sum_{i=1}^{n} \frac{1}{i} \rightarrow$ We know that this sum is bounded by the integral of 1/x. More specifically, $\sum_{i=1}^{n} \frac{1}{i} \leq 1 + ln(n)$. Generally, because $E[T] = \sum_{i=1}^{n} \frac{1}{i}$, $|E[T] - ln(n)| \leq c$ for some positive constant c and all sufficiently large n. Therefore, $E[T] = ln(n) + O(1)$.

Dennis Kuzminer
CSCI-UA 310-001 PS6

7. $E[X],\ X = X_1 + ... + X_n \rightarrow$ where X_i is the number of spins you make to get the ith distinct number after you have already gotten i - 1 distinct numbers.
$E[X_1 + ... + X_n] = E[X_1] + ... + E[X_n] \rightarrow$
Let's compare the values of the individual expectations for n = 10 and derive a trend.
Clearly, $E[X_1] = 1$
If 1 number is achieved, and each number has the same probability of being chosen, p of choosing a new number = 9/10; therefore, $E[X_2] = \frac{10}{9}$
If 2 number is achieved, and each number has the same probability of being chosen, p of choosing a new number = 8/10; therefore, $E[X_3] = \frac{10}{8}$
Generalizing, we can see that $E[X_i] = \frac{10}{10-i},\ i = [0,\ ...,\ 10) \rightarrow E[X_i] = \frac{n}{n-i},\ i = [0,\ ...,\ n) \rightarrow$
$E[X] = \sum_{i=0}^{n-1} \frac{n}{n-i} \rightarrow n \sum_{i=0}^{n-1} \frac{1}{n-i} \rightarrow$ Writing out the terms, we can rearrange this sum to $n \sum_{i=1}^{n} \frac{1}{i} \rightarrow$
We can estimate this with an integral, $\int_1^n \frac{1}{x}\ dx + M,\ M = max(1/1,\ 1/n) = 1 \rightarrow$
$ln(n) - ln(1) + 1 + c_1 \rightarrow ln(n) + c_2,\ c_2 = 1 + c_1 \rightarrow$ Putting back into the equation and generalizing c, we get $E[X] = n(ln(n) + c) \rightarrow nln(n) + nc$
$|E[X] - nln(n)|$ is bounded by nc. Additionally, we see that nc grows at most the rate of n.
Meaning, we can generalize be saying $E[X] = nln(n) + O(n)$

Dennis Kuzminer
CSCI-UA 310-001 PS6

8.

a. Consider the case where we only spin once, t = 1, and we will generalize the probability of $Pr[M \geq j]$. If j = 1, clearly $Pr[M \geq j] = 1$. If j = 2, $Pr[M \geq j] = 9/10$. We can see that this probability can be generalized to (n - j + 1) / n, for any j when t = 1. When t = 2, we consider the case when *both* spins have M as greater or equal to than j $\rightarrow$ $Pr[M \geq j] \cap Pr[M \geq j] = Pr[M \geq j] * Pr[M \geq j] \rightarrow \frac{(n-j+1)}{n} * \frac{(n-j+1)}{n}$, given that each spin is independent. More generally, for any t and j = 1, …, n, $Pr[M \geq j] = \frac{(n-j+1)^t}{n^t}$.

b. Given the tail-sum formula, $E[M] = \sum_{j=1}^{n} Pr[M \geq j] \rightarrow \sum_{j=1}^{n} \frac{(n-j+1)^t}{n^t} \rightarrow \frac{1}{n^t} \sum_{j=1}^{n} (n-j+1)^t$ $\rightarrow$ To simplify the sum, we can compare the values of j and n - j + 1. When j = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, n - j + 1 = 10, 9, 8, 7, 6, 5, 4, 3, 2, 1. Therefore, we can rearrange the sum to be $\sum_{i=1}^{n} i^t$, as the set of numbers in j perfectly spans/covers the set of numbers represented by n - j + 1. Therefore, $E[M] = \frac{1}{n^t} \sum_{i=1}^{n} i^t$.

c. $E[M] = \frac{1}{n^t} \sum_{i=1}^{n} i^t \rightarrow \frac{1}{n^t} \sum_{i=1}^{n} i^t \rightarrow$ Approximating as an integral $\rightarrow \frac{1}{n^t} \int_1^n x^t \, dx + M \rightarrow$

$$\frac{1}{n^t}\left(\left[\frac{x^{t+1}}{t+1}\right]_1^n + max(1^t, n^t)\right) \rightarrow$$

$$\frac{1}{n^t}\left(\frac{n^{t+1}}{t+1} - \frac{1}{t+1} + \frac{n^t}{n^t}\right) \rightarrow \frac{n^{t+1}-1}{n^t(t+1)} + \frac{1}{1} \rightarrow \frac{n * n^t}{n^t(t+1)} - \frac{1}{n^t(t+1)} + \frac{1}{1} \rightarrow$$

$\frac{n}{(t+1)} - \frac{1}{n^t(t+1)}$, we can see that for all sufficiently large n, the term c, where $c = \frac{1}{n^t(t+1)} + 1$, bounds the sum such that $\left|E[M] - \frac{n}{(t+1)}\right| \leq c$. Additionally, we can say that the constant c is significantly less than $\frac{n}{(t+1)}$ (especially if n will never be smaller than 1), so we can generalize the formula to be $E[M] = \frac{n}{(t+1)} + O(1)$.