Problem Set 11 Solutions

Problem 1. (10 points) Several statements involving probability are listed below. Assume that A and B are events, S and T are random variables, and the underlying sample space is countable. Furthermore, assume that all expectations exist. For each statement, describe restrictions on A, B, S, and T sufficient to ensure that the statement is true. Make your restrictions as minimal as you can. No proofs are necessary.

(a) $\Pr(A \cup B) = \Pr(A) + \Pr(B)$

This statement is true provided that the event $A \cap B$ has probability zero.

(b) $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$

This statement is always true.

(c) $\Pr(A \cap B) = \Pr(A) \cdot \Pr(B)$

This statement is true provided that events A and B are independent.

(d) $\Pr(A \cap B) < \Pr(A)$

This statement is true provided $\Pr(A \cap \overline{B}) > 0$.

(e) $\text{Ex}(S + T) = \text{Ex}(S) + \text{Ex}(T)$

This statement is always true.

(f) $\text{Ex}(S \cdot T) = \text{Ex}(S) \cdot \text{Ex}(T)$

This statement is true provided that random variables S and T are independent.

Problem 2. (5 points) Suppose that I roll a 4-sided die, a 6-sided die, an 8-sided die, a 10-sided die, a 12-sided die, and a 20-sided die. What is the expected number of 6's that come up? (Assume that all of the dice are fair.)

$$\begin{aligned}
\text{Ex}(\text{6's on all dice}) &= \text{Ex}(\text{6's on 4-sided die}) + \text{Ex}(\text{6's on 6-sided die}) + \ldots \\
&\quad + \text{Ex}(\text{6's on 20-sided die}) \\
&= \frac{0}{4} + \frac{1}{6} + \frac{1}{8} + \frac{1}{10} + \frac{1}{12} + \frac{1}{20} \\
&= \frac{21}{40}
\end{aligned}$$

The first step uses linearity of expectation, and the second step is simplification.

Problem 3. (10 points) Suppose that I roll n dice that are 6-sided, fair, and mutually independent. What is the expected value of the largest number that comes up?

Let the random variable M be the largest number that comes up. We will use the identity

$$\text{Ex}(M) = \sum_{i=0}^{\infty} \Pr(M > i)$$

We must compute the probability of the event $M > i$; that is, the event that some die shows a value greater than i.

$$\begin{aligned}
\text{Ex}(M > i) &= \sum_{i=0}^{5} \Pr(M > i) \\
&= \sum_{i=0}^{5} 1 - \Pr(\text{every die} \leq i) \\
&= \sum_{i=0}^{5} 1 - (\Pr(\text{one die} \leq i))^n \\
&= \sum_{i=0}^{5} 1 - \left(\frac{i}{6}\right)^n \\
&= 6 - \frac{1^n + 2^n + 3^n + 4^n + 5^n}{6^n}
\end{aligned}$$

Problem 4. (10 points) Suppose that I choose a listing of the numbers $1, 2, \ldots, n$ uniformly at random. What is the expected number of entries that are greater than all preceding entries? For example, in the listing 4, 2, 1, 5, 3, the numbers 4 and 5 are greater than all preceding entries.

Let i_k be an indicator for the event that the k-th entry is greater than the preceding $k - 1$ entries. By linearity of expectation

$$\begin{aligned}
\text{Ex}(\#\text{ new maxima}) &= \sum_{k=1}^{n} \text{Ex}(i_k) \\
&= \sum_{k=1}^{n} 1 \cdot \Pr(i_k = 1) + 0 \cdot \Pr(i_k = 0) \\
&= \sum_{k=1}^{n} \Pr(i_k = 1)
\end{aligned}$$

All that remains is to compute $\Pr(i_k = 1)$, which is the probability that the k-th entry is greater than all preceding entries. Note that each of the first k entries is equally likely to be the largest. Therefore, $\Pr(i_k = 1) = \frac{1}{k}$. Substituting into the equation above gives:

$$\begin{aligned}
\text{Ex}(\#\text{ new maxima}) &= \sum_{k=1}^{n} \Pr(i_k = 1) \\
&= \sum_{k=1}^{n} \frac{1}{k} \\
&= H_n
\end{aligned}$$

Problem 5. (10 points) Biologists are inexplicably obsessed with *genetic strings*, which consist of the characters A, C, T, and G. A biology student once observed that the sequence AAA occurs more often in a random genetic string than, say, CTG.

His reasoning is that once AAA has appeared, there is a $\frac{1}{4}$ chance that the next symbol will be an A. If this happens, then the sequence AAA has appeared a second time, offset one character to the right of the first occurrence. By contrast, once CTG has appeared, the next appearance of CTG must be offset at least three characters to the right of the first occurrence. Furthermore, the probability that a second CTG appears immediately to the right of the first is a miniscule $\frac{1}{4^3} = \frac{1}{64}$.

(a) Compute the expected number of occurrences of AAA and of CTG in a genetic string of length 4 chosen uniformly at random.

We have a uniform sample space containing the $4^4 = 256$ possible length-4 genetic strings. Let the random variable X equal the number of occurrences of AAA, and let the random variable Y equal the number of occurrences of CTG.

$$\begin{aligned}
\text{Ex}(X) &= \sum_{x \in \mathbb{N}} x \cdot \Pr(X = x) \\
&= 0 \cdot \frac{249}{256} + 1 \cdot \frac{6}{256} + 2 \cdot \frac{1}{256} \\
&= \frac{1}{32}
\end{aligned}$$

The six outcomes with $X = 1$ are $CAAA$, $TAAA$, $GAAA$, $AAAC$, $AAAT$, and $AAAG$. The single outcome with $X = 2$ is $AAAA$.

$$\begin{aligned}
\mathrm{Ex}(Y) &= \sum_{y \in \mathbb{N}} y \cdot \Pr(Y = y) \\
&= 0 \cdot \frac{248}{256} + 8 \cdot \frac{1}{256} \\
&= \frac{1}{32}
\end{aligned}$$

The eight outcomes with $Y = 1$ are $ACTG$, $CCTG$, $TCTG$, $GCTG$, $CTGA$, $CTGC$, $CTGT$, and $CTGG$.

(b) Compute the expected number of occurrences of AAA and of CTG in a genetic string of length n chosen uniformly at random.

There are $n - 2$ sequences of three consecutive characters in a genetic string of length n. Each sequence is AAA with probability $\frac{1}{64}$. The expected number of occurrences of AAA in three consecutive positions is therefore $\frac{1}{64}$. By linearity of expectation, the expected number of occurrences of AAA in a genetic string of length n is $\frac{n-2}{64}$. Precisely the same calculation is valid for the sequence CTG, and so the answer is the same.

Problem 6. **(20 points)** Suppose that R is a random variable. Let E be an event. The *conditional expectation* of R given that event E occurs is denoted $\mathrm{Ex}(R \mid E)$ and is defined to be

$$\mathrm{Ex}(R \mid E) = \sum_{x \in \mathrm{range}(R)} x \cdot \Pr(R = x \mid E).$$

(a) Compute the expected value of the number rolled on a fair, 6-sided die, given that the outcome is even.

Let the random variable R be the number rolled, and let E be the event that the number rolled is even.

$$\begin{aligned}
\mathrm{Ex}(R \mid E) &= \sum_{x=1}^{6} x \cdot \Pr(R = x \mid E) \\
&= 1 \cdot 0 + 2 \cdot \frac{1}{3} + 3 \cdot 0 + 4 \cdot \frac{1}{3} + 5 \cdot 0 + 6 \cdot \frac{1}{3} \\
&= 4
\end{aligned}$$

(b) Compute the expected value of the number rolled on a fair, 6-sided die, given that the outcome is odd.

Define R and E as before.

$$\begin{aligned}
\mathrm{Ex}(R \mid \overline{E}) &= \sum_{x=1}^{6} x \cdot \Pr(R = x \mid \overline{E}) \\
&= 1 \cdot \frac{1}{3} + 2 \cdot 0 + 3 \cdot \frac{1}{3} + 4 \cdot 0 + 5 \cdot \frac{1}{3} + 6 \cdot 0 \\
&= 3
\end{aligned}$$

(c) Prove the following identity, assuming that all expectations exist.

$$\mathrm{Ex}(R) = \Pr(E) \cdot \mathrm{Ex}(R \mid E) + \Pr(\overline{E}) \cdot \mathrm{Ex}(R \mid \overline{E})$$

We transform the right side into the left side. All summations are over $x \in \mathrm{Range}(R)$.

$$\begin{aligned}
&\Pr(E) \cdot \mathrm{Ex}(R \mid E) + \Pr(\overline{E}) \cdot \mathrm{Ex}(R \mid \overline{E}) \\
&= \Pr(E) \cdot \left(\sum_x x \cdot \Pr(R = x \mid E) \right) + \Pr(\overline{E}) \cdot \left(\sum_x x \cdot \Pr(R = x \mid \overline{E}) \right) \\
&= \Pr(E) \cdot \left(\sum_x x \cdot \frac{\Pr(R = x \cap E)}{\Pr(E)} \right) + \Pr(\overline{E}) \cdot \left(\sum_x x \cdot \frac{\Pr(R = x \cap \overline{E})}{\Pr(\overline{E})} \right) \\
&= \sum_x x \cdot \left(\Pr(R = x \cap E) + \Pr(R = x \cap \overline{E}) \right) \\
&= \sum_x x \cdot \left(\Pr((R = x \cap E) \cup (R = x \cap \overline{E})) \right) \\
&= \sum_x x \cdot \Pr(R = x) \\
&= \mathrm{Ex}(R)
\end{aligned}$$

The first step uses the definition of conditional expectation. The second step uses the definition of conditional probability. The third step is simplification. The fourth step uses the fact that the probability of a disjoint union of events is equal to the sum of the event probabilities. The fifth step is simplification, and the final step uses the definition of expected value.

(d) Compute the expected value of the product of the numbers rolled on two dice that are 6-sided, fair, and independent, given that the numbers rolled are distinct.

Let the random variable R be the product of the two dice, and let E be the event that the two numbers rolled are *equal*. The following three quantities are easy to compute:

$$\begin{aligned}
\text{Ex}(R) &= \text{Ex}(\text{Die } 1) \cdot \text{Ex}(\text{Die } 2) \\
&= \frac{7}{2} \cdot \frac{7}{2} \\
&= \frac{49}{4} \\
\Pr(E) &= \frac{6}{36} \\
&= \frac{1}{6} \\
\text{Ex}(R \mid E) &= \sum_{x \in \mathbb{N}} x \cdot \Pr(R = x \mid E) \\
&= 1^2 \cdot \frac{1}{6} + 2^2 \cdot \frac{1}{6} + 3^2 \cdot \frac{1}{6} + 4^2 \cdot \frac{1}{6} + 5^2 \cdot \frac{1}{6} + 6^2 \cdot \frac{1}{6} \\
&= \frac{91}{6}
\end{aligned}$$

We want to compute $\text{Ex}(R \mid \overline{E})$, the expected value of the product of the two rolls, given that they are distinct. From the identity in part(c), we have:

$$\begin{aligned}
\text{Ex}(R \mid \overline{E}) &= \frac{\text{Ex}(R) - \Pr(E) \cdot \text{Ex}(R \mid E)}{\Pr(\overline{E})} \\
&= \frac{\frac{49}{4} - \frac{1}{6} \cdot \frac{91}{6}}{\frac{5}{6}} \\
&= \frac{35}{3}
\end{aligned}$$

Problem 7. (10 points) Suppose that I start throwing balls uniformly and mutually independently at random into n bins.

(a) Suppose that at some stage, exactly k of the bins contain at least one ball. Now I throw additional balls until exactly $k + 1$ of the bins contain at least one ball. What is the expected number of additional balls that I throw?

Let the random variable A_k equal the number of additional balls that I throw. The expected value of A_k can be computed as follows.

$$\begin{aligned}
\operatorname{Ex}(A_k) &= \sum_{i=0}^{\infty} \Pr(A_k > i) \\
&= \sum_{i=0}^{\infty} \left(\frac{k}{n}\right)^i \\
&= \frac{1}{1 - \frac{k}{n}} \\
&= \frac{n}{n-k}
\end{aligned}$$

The first equation holds because B is a random variable taking on non-negative integer values. In the second step, more than i balls are required if the first i balls all fall into the k non-empty bins; this happens with probabilty $(\frac{k}{n})^i$. The formula for the sum of an infinite geometric series is used in the third step. The final step is simplification.

(b) Suppose that the bins start out empty, and I throw balls until every bin has at least one ball. What is the expected number of balls that I throw?

Let the random variable B equal the number of balls that I must throw to put at least one ball in every bin. The number of balls required is the number of balls required to fill one bin plus the number of balls required to fill a second bin and so forth. That is, $B = A_0 + A_1 + \ldots + A_{n-1}$. We can compute the expected value of B as follows.

$$\begin{aligned}
\operatorname{Ex}(B) &= \operatorname{Ex}(A_0 + A_1 + \ldots + A_{n-1}) \\
&= \operatorname{Ex}(A_0) + \operatorname{Ex}(A_1) + \ldots + \operatorname{Ex}(A_{n-1}) \\
&= \frac{n}{n-0} + \frac{n}{n-1} + \ldots + \frac{n}{1} \\
&= n \cdot \left(\frac{1}{n} + \frac{1}{n-1} + \ldots + \frac{1}{1}\right) \\
&= nH_n
\end{aligned}$$

The first step follows by taking expectations on both sides of the equation $B = A_0 + A_1 + \ldots + A_{n-1}$. The second step uses linearity of expectation, and the third uses the result of part (a). The remaining steps are simplifications.

Problem 8. **(15 points)** Let R be a random variable taking on non-negative integer values. Let r be the probability generating function for R; that is, define

$$r(x) = \Pr(R = 0) + \Pr(R = 1) \cdot x + \Pr(R = 2) \cdot x^2 + \Pr(R = 3) \cdot x^3 + \ldots .$$

(a) Write an expression involving r that is equal to the expected value of R.

$$\begin{aligned} r'(x) &= 1 \cdot \Pr(R = 1) + 2 \cdot \Pr(R = 2) \cdot x + 3 \cdot \Pr(R = 3) \cdot x^2 + \ldots \\ r'(1) &= 1 \cdot \Pr(R = 1) + 2 \cdot \Pr(R = 2) + 3 \cdot \Pr(R = 3) + \ldots \\ &= \mathrm{Ex}(R) \end{aligned}$$

(b) Suppose that $\Pr(R = k) = \frac{2}{3^{k+1}}$ for all $k \geq 0$. Find a closed-form expression for r, the probability generating function for R.

$$\begin{aligned} r(x) &= \frac{2}{3^1} + \frac{2}{3^2} \cdot x + \frac{2}{3^3} \cdot x^2 + \frac{2}{3^3} \cdot x^3 + \ldots \\ &= \frac{\frac{2}{3}}{1 - \frac{x}{3}} \\ &= \frac{2}{3 - x} \end{aligned}$$

The first equation follows from the definition of the probability generating function r. The second equation uses the formula for the sum of an infinite geometric series, and the third equation is obtained by simplification.

(c) Using your solution to part (a), determine the expected value of the random variable R as defined in part (b).

$$\begin{aligned} r'(x) &= \frac{2}{(3 - x)^2} \\ r'(1) &= \frac{2}{(3 - 1)^2} \\ &= \frac{1}{2} \end{aligned}$$

Problem 9. (10 points) 6.042 TA Eric Lehman has offered to play a "fun game" with Theory Pig, who rarely comes out of his drawer anymore after all the abuse that he has suffered this term.

Here are the rules. Eric flips a coin until he gets a heads. Let n be the number of tails preceding the first heads. For example, if he flips $TTTH$, then $n = 3$, and if he flips just H, then $n = 0$. Now Eric offers to pat Theory Pig on each of the next 3^n days. If Theory Pig declines the offer, then Eric repeats the coin-tossing experiment and instead pats Theory Pig on each of next 3^m days where m is the number of tails before the first heads in the second trial.

As usual, this game sounds great, but Theory Pig is understandably nervous. He wants to determine his best strategy.

(a) As a warmup, consider the following simpler variant of the game. Suppose that Eric rolls a fair, 6-sided die. He offers to pat Theory Pig for the number of days indicated on the die. If Theory Pig rejects the offer, then Eric rolls again and pats Theory Pig for the number of days shown on the second roll. What is expected number of patting-days, if Theory Pig uses an optimal strategy?

If Theory Pig always declines the initial offer, then his expected payoff is $\frac{7}{2}$, the expected roll on a fair, 6-sided die. Let p_k be the probability that Theory Pig accepts the initial offer, given that k is rolled. Then Theory Pig's expected payoff is

$$\text{Ex(patting days)} = \sum_{i=1}^{6} \frac{1}{6}(p_i \cdot i + (1 - p_i) \cdot \frac{7}{2}).$$

This sum is maximized by maximizing each term. Consequently, we should choose $p_1 = p_2 = p3 = 0$ and $p_4 = p_5 = p_6 = 1$. That is, Theory Pig should decline an offer of 3 or less and accept an offer of 4 or more. In this case, his expected payoff is $\frac{17}{4}$.

(b) Now return to the original coin-flipping game. Describe a strategy that maximizes the expected number of patting days.

WARNING: If less than 80% of the class solves this problem part correctly, then Eric will withdraw his offer to play the game, and THE PIG GETS NOTHING!

If Theory Pig declines the initial offer, then his expected payoff is

$$\text{Ex(patting days if decline)} = \sum_{i=0}^{\infty} \frac{1}{2^{i+1}} \cdot 3^i.$$

This summation is infinite. Therefore, if there is a positive probability that Theory Pig decines the initial offer, his expected payoff is infinite. On the other hand, suppose that there is zero probability that Theory Pig declines the initial offer; that is, he always accepts. Then his expected payoff is

$$\text{Ex(patting days if accept)} = \sum_{i=0}^{\infty} \frac{1}{2^{i+1}} \cdot 3^i.$$

This is the same infinite sum. Therefore, every strategy is optimal and gives an infinite expected number of patting days.

We are still tallying up the precentage of students that solved this problem correctly.