Homework 7

Due: Monday, August 11 at 11:59pm

CS 70: Discrete Mathematics and Probability Theory, Summer 2014

1. **Expectation of Solace** [8 points] James Bond is imprisoned in a cell from which there are three possible ways to escape: an air-conditioning duct, a sewer pipe, and the door (which is unlocked). The air-conditioning duct leads him on a two-hour trip whereupon he falls through a trap door onto his head, much to the amusement of his captors. The sewer pipe is similar but takes five hours to traverse. Each fall produces temporary amnesia and he is returned to the cell immediately after each fall. Assume that he always immediately chooses one of the three exits from the cell with equal probabilities. On average, how long does it take before he realizes that the door is unlocked and escapes?

   Answer: Let the random variable T denote the time to escape, and let A, S, D be the events that Bond chose (as his first attempt) to go through the AC-duct, the sewer, and the door, respectively. $\mathrm{E}(T \mid A)$ means the expected time to escape given that Bond went through the AC-duct, etc.

   Note that:

$$\begin{aligned} E[T] &= E[T \mid A] \Pr[A] + E[T \mid S] \Pr[S] + E[T \mid D] \Pr[D] \\ &= \frac{1}{3}(E[T \mid A] + E[T \mid S] + E[T \mid D]) \end{aligned}$$

   Also, because of the memorylessness of the situation, we have the following:

$$\begin{aligned} E[T \mid A] &= E[T] + 2 \\ E[T \mid S] &= E[T] + 5 \end{aligned}$$

   Lastly, Mr. Bond escapes immediately if he takes the door, so:

$$E[T \mid D] = 0$$

   So we have that $E[T] = \frac{1}{3}(E[T] + 2 + E[T] + 5 + 0)$. Solving this for $E[T]$, we get that $E[T] = \boxed{\mathbf{007}}$.

2. **Casino Pascal** [12 points] Use the appropriate distributions to answer these questions.

   2a. [4 points] James Bond rolls twenty (fair, six-sided) dice. What's the probability that exactly three of them land on a 6?

   Answer: This is a binomial distribution with $n = 20$ and $p = \frac{1}{6}$. (the probability that a given die lands on a 6) Then,

$$\Pr[X = 3] = \binom{20}{3}(\frac{1}{6})^3(\frac{5}{6})^{17}$$

2b. [4 points] Again, 007 rolls twenty (fair, six-sided) dice. What's the probability that three or fewer land on a 6?

Answer: This is an extension of the previous part. The probability we're interested in is:

$$\Pr[X = 0 \cup X = 1 \cup X = 2 \cup X = 3]$$

Since each of these are disjoint, we don't need to use inclusion-exclusion, and can just write:

$$\Pr[X = 0] + \Pr[X = 1] + \Pr[X = 2] + \Pr[X = 3]$$
$$= \binom{20}{0}(\frac{1}{6})^0(\frac{5}{6})^{20} + \binom{20}{1}(\frac{1}{6})^1(\frac{5}{6})^{19} + \binom{20}{2}(\frac{1}{6})^2(\frac{5}{6})^{18} + \binom{20}{3}(\frac{1}{6})^3(\frac{5}{6})^{17}$$

2c. [4 points] James now decides to throw dice until he gets a 6. What's the probability that he throws at least four dice?

Answer: This is a geometric distribution with $p = \frac{1}{6}$. One approach is to calculate $1 - \Pr[X = 1 \cup X = 2 \cup X = 3]$, which is:

$$1 - \Pr[X = 1 \cup X = 2 \cup X = 3]$$
$$1 - (\Pr[X = 1] + \Pr[X = 2] + \Pr[X = 3])$$
$$1 - (\frac{1}{6} + \frac{5}{6} \cdot \frac{1}{6} + \frac{5}{6} \cdot \frac{5}{6} \cdot \frac{1}{6}) = \frac{125}{216}$$

An easier way: we can just say that in order to throw at least four dice, James must throw a non-6 three times, which is $(\frac{5}{6})^3 = \frac{125}{216}$.

3. **Licence To Thrill** [12 points] Use the Poisson distribution to answer these questions.

3a. [4 points] Suppose that on average, 20 people ride your roller coaster per day. What is the probability that exactly 7 people ride it tomorrow?

Answer: We can model this as a poisson distribution $X \sim \text{Poiss}(\lambda = 20)$. Thus,

$$\Pr[X = 7] = \frac{20^7}{7!}e^{-20} \approx 5.23 \times 10^{-4}$$

3b. [4 points] Suppose that on average, you go to Six Flags twice a year. What is the probability that you will go *at most* once in 2014?

Answer: We can model this as a poisson distribution $X \sim \text{Poiss}(\lambda = 2)$. Thus,

$$\Pr[X \leq 1] = \Pr[X = 0 \cup X = 1] = \Pr[X = 0] + \Pr[X = 1] = \frac{2^0}{0!}e^{-2} + \frac{2^1}{1!}e^{-2} \approx 0.41$$

Note: We do not need to use inclusion-exclusion here because the events are disjoint (you cannot go zero times *and* go one time in the same year).

3c. [4 points] Suppose that on average, there are 5.7 accidents per day on California roller coasters. (I hope this is not true.) What is the probability there will be *at least* 3 accidents throughout the *next two days* on California roller coasters?

Answer: Let Y be the number of accidents that occur in the next two days. We can approximate Y as a poisson distribution $Y \sim \text{Poiss}(\lambda = 11.4)$, where λ is the average number of accidents over two days. Now, we compute

$$\begin{aligned}
\Pr[Y \geq 3] &= 1 - \Pr[Y < 3] \\
&= 1 - \Pr[Y = 0 \cup Y = 1 \cup Y = 2] \\
&= 1 - (\Pr[Y = 0] + \Pr[Y = 1] + \Pr[Y = 2]) \\
&= 1 - (\frac{11.4^0}{0!} e^{-11.4} + \frac{11.4^1}{1!} e^{-11.4} + \frac{11.4^2}{2!} e^{-11.4}) \\
&\approx 0.999
\end{aligned}$$

Note: We can show what we did above formally with the following claim: the sum of two independent poisson random variables is poisson. We won't prove this, but from the above, you should intuitively know why this is true. Now, we can model accidents on day i as a poisson distribution $X_i \sim \text{Poiss}(\lambda = 5.7)$. Now, Let X_1 be the number of accidents that happen on the next day, and X_2 be the number of accidents that happen on the day after next. We are interested in $Y = X_1 + X_2$. Thus, we know $Y \sim \text{Poiss}(\lambda = 5.7 + 5.7 = 11.4)$.

4. **A Variance To Kill** [16 points] This problem will give you practice using the "standard method" to compute the variance of a sum of random variables that are not pairwise independent (so you cannot use "linearity" of variance).

4a. [8 points] A building has n floors numbered $1, 2, \ldots, n$, plus a ground floor G. At the ground floor, m people get on the elevator together, and each gets off at a uniformly random one of the n floors (independently of everybody else). What is the *variance* of the number of floors the elevator *does not* stop at? (In fact, the variance of the number of floors the elevator *does* stop at must be the same (do you see why?) but the former is a little easier to compute.)

Answer: Let X be the number of floors the elevator does not stop at. As in the previous homework, we can represent X as the sum of the indicator variables $X_1, \ldots, X_n$, where $X_i = 1$ if no one gets off on floor i. Thus, we have

$$\mathbb{E}(X_i) = \Pr[X_i = 1] = \left(\frac{n-1}{n}\right)^m,$$

and from linearity of expectation,

$$\mathbb{E}(X) = \sum_i^n \mathbb{E}(X_i) = n\left(\frac{n-1}{n}\right)^m$$

To find the variance, we cannot simply sum the variance of our indicator variables. However, we can still compute $\text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$ directly using linearity of expectation.

But now how can we find $\mathbb{E}\left(X^2\right)$? Recall that

$$\begin{aligned}
\mathbb{E}\left(X^2\right) &= \mathbb{E}\left([X_1 + \cdots + X_n]^2\right) \\
&= \mathbb{E}\left(\sum_{i,j} X_i X_j\right) \\
&= \sum_{i,j} \mathbb{E}\left(X_i X_j\right) \\
&= \sum_{i}^{n} \mathbb{E}\left([X_i]^2\right) + \sum_{i \neq j} \mathbb{E}\left(X_i X_j\right)
\end{aligned}$$

The first term is simple to calculate: $\mathbb{E}\left([X_i]^2\right) = 1^2 \Pr[X_i = 1] = \left(\frac{n-1}{n}\right)^m$, meaning that

$$\sum_{i}^{n} \mathbb{E}\left([X_i]^2\right) = n\left(\frac{n-1}{n}\right)^m$$

$X_i X_j = 1$ when both X_i and X_j are 1, which means no one gets off of the elevator on floor i and floor j. This happens with probability

$$\begin{aligned}
\Pr[X_i = X_j = 1] &= \Pr[X_i = 1 \cap X_j = 1] \\
&= \left(\frac{n-2}{n}\right)^m
\end{aligned}$$

Thus, we can now compute

$$\sum_{i \neq j} \mathbb{E}\left(X_i X_j\right) = n(n-1)\left(\frac{n-2}{n}\right)^m$$

Finally, we plug in to see that

$$\begin{aligned}
\operatorname{Var}(X) &= \mathbb{E}\left(X^2\right) - \mathbb{E}\left(X\right)^2 \\
&= n\left(\frac{n-1}{n}\right)^m + n(n-1)\left(\frac{n-2}{n}\right)^m - \left(n\left(\frac{n-1}{n}\right)^m\right)^2
\end{aligned}$$

4b. [8 points] A group of three friends has n books they would all like to read. Each friend (independently of the other two) picks a random permutation of the books and reads them in that order, one book per week (for n consecutive weeks). Let X be the number of weeks in which all three friends are reading the same book. Compute $\operatorname{Var}(X)$.

Answer: Let $X_1, \ldots, X_n$ be indicator variables such that $X_i = 1$ if all three friends are reading the same book on week i. Thus, we have

$$\mathbb{E}\left(X_i\right) = \Pr[X_i = 1] = \left(\frac{1}{n}\right)^2,$$

and from linearity of expectation,

$$\mathbb{E}\left(X\right) = \sum_{i}^{n} \mathbb{E}\left(X_i\right) = n\left(\frac{1}{n}\right)^2 = \frac{1}{n}$$

As before, we know that

$$\mathbb{E}\left(X^2\right) = \sum_{i}^{n} \mathbb{E}\left([X_i]^2\right) + \sum_{i \neq j} \mathbb{E}\left(X_i X_j\right)$$

Furthermore, because X_i is an indicator variable, $\mathbb{E}\left([X_i]^2\right) = 1^2 \Pr[X_i = 1] = \left(\frac{1}{n}\right)^2$, and

$$\sum_{i}^{n} \mathbb{E}\left([X_i]^2\right) = n\left(\frac{1}{n}\right)^2 = \frac{1}{n}$$

Again, because X_i and X_j are indicator variables, we are interested in

$$\Pr[X_i = X_j = 1] = \Pr[X_i = 1 \cap X_j = 1] = \frac{1}{[n(n-1)]^2},$$

the probability that all three friends pick the same book on week i and week j. Thus,

$$\sum_{i \neq j} \mathbb{E}\left(X_i X_j\right) = n(n-1) \cdot \frac{1}{[n(n-1)]^2} = \frac{1}{n(n-1)}$$

Finally, we compute

$$\begin{aligned} \mathrm{Var}\left(X\right) &= \mathbb{E}\left(X^2\right) - \mathbb{E}\left(X\right)^2 \\ &= \frac{1}{n} + \frac{1}{n(n-1)} - \left(\frac{1}{n}\right)^2 \end{aligned}$$

5. **Octoproof** [8 points] Consider any random variable X with a finite sample space. Prove that $\mathrm{E}(X^2) \geq \mathrm{E}(X)^2$, and that $\mathrm{E}(X^2) = \mathrm{E}(X)^2$ iff X is a constant random variable. (Hint: Think in terms of variance.)

   Answer: First, we will show $\mathbb{E}\left(X^2\right) \geq \mathbb{E}\left(X\right)^2$. First note that

$$\mathrm{Var}(X) = \mathbb{E}\left([X - \mathbb{E}\left(X\right)]^2\right) = \mathbb{E}\left(X^2\right) - \mathbb{E}\left(X\right)^2$$

   However, $\mathbb{E}\left([X - \mathbb{E}\left(X\right)]^2\right)$ can never be negative, meaning that

$$\mathrm{Var}(X) = \mathbb{E}\left(X^2\right) - \mathbb{E}\left(X\right)^2 \geq 0$$

   Therefore, we can conclude $\mathbb{E}\left(X^2\right) \geq \mathbb{E}\left(X\right)^2$.$\square$

   Claim: $\mathbb{E}\left(X^2\right) = \mathbb{E}\left(X\right)^2 \iff X$ is a constant random variable.

Proof: $\mathbb{E}\left(X^2\right) = \mathbb{E}\left(X\right)^2 \implies X$ is a constant random variable.
Let $\mu = \mathbb{E}\left(X\right)$. Recall that

$$\text{Var}(X) = \mathbb{E}\left([X-\mu]^2\right) = \mathbb{E}\left(X^2\right) - \mu^2$$

$\mathbb{E}\left(X^2\right) = \mathbb{E}\left(X\right)^2$ means that $\mathbb{E}\left(X^2\right) - \mathbb{E}\left(X\right)^2 = 0 = \text{Var}(X)$. Or, in other words,

$$\mathbb{E}\left([X-\mu]^2\right) = \sum_i \Pr[X=i](i-\mu)^2 = 0$$

No term in the summation can be negative, which means that in order for the sum to be 0, $i = \mu$ for all i such that $\Pr[X=i] \neq 0$. Therefore, X must be a constant random variable.$\square$

Proof: X is a constant random variable $\implies \mathbb{E}\left(X^2\right) = \mathbb{E}\left(X\right)^2$.
If X is a constant random variable, then $i = c$ for all i such that $\Pr[X=i] \neq 0$. This means

$$\mu = \mathbb{E}\left(X\right) = \sum_i i \times \Pr[X=i] = \sum_i c \times \Pr[X=i] = c \sum_i \Pr[X=i] = c$$

Furthermore,

$$\mathbb{E}\left(X^2\right) = \sum_i i^2 \times \Pr[X=i] = \sum_i c^2 \times \Pr[X=i] = c^2 \sum_i \times \Pr[X=i] = c^2$$

Lastly, we observe that $\mathbb{E}\left(X^2\right) = \mathbb{E}\left(X\right)^2$, and we are done. $\square$

We have shown that both sides of the if and only if statement hold true, and therefore have shown our claim to be true. $\square$

6. **Course Schedules Are Forever** [20 points] A friend tells you about a course called "Laziness in Modern Society" that requires almost no work. You hope to take this course so that you can devote all of your time to CS70. At the first lecture, the professor announces that grades will depend on only a midterm and a final. The midterm will consist of three questions, each worth 10 points, and the final will consist of four questions, also each worth 10 points. He will give an A to any student who gets at least 60 of the possible 70 points.

   However, speaking with the professor in office hours you hear some very disturbing news. He tells you that to save time he will be grading as follows. For each student's midterm, he'll choose a number randomly from some distribution with mean $\mu = 5$ and variance $\sigma^2 = 1$. He'll mark each of the three questions with that score. To grade the final, he'll again choose a random number from the same distribution, independent of the first number, and will mark all four questions with that score.

   6a. [5 points] What will the mean (i.e., expectation) of your total score be?
   Answer: Let Q_M be the score you get on each question of the midterm, and let Q_F be the score you get on each question of the final. Let M be your midterm score, F be your final score, and T be your total score. We see that

$$\begin{aligned} T &= M + F \\ M &= 3Q_M \\ F &= 4Q_F \end{aligned}$$

Thus, we see that

$$\begin{aligned}\mathbb{E}(T) &= \mathbb{E}(3Q_M) + \mathbb{E}(4Q_F) \\ &= 3\mathbb{E}(Q_M) + 4\mathbb{E}(Q_F) \\ &= 3 \times 5 + 4 \times 5 \\ &= 35\end{aligned}$$

6b. [5 points] Use Markov's Inequality to prove an upper bound on the probability of getting an A.

Answer: We are interested in $\Pr[T \geq 60]$. Using Markov's Inequality, we have

$$\Pr[T \geq 60] \leq \frac{\mathbb{E}(T)}{60} = \frac{35}{60} = \frac{7}{12}$$

6c. [5 points] What will the variance of your total score be?

Answer:

$$\begin{aligned}\mathrm{Var}(T) &= \mathrm{Var}(3Q_M) + \mathrm{Var}(4Q_F) \\ &= 9 \times \mathrm{Var}(Q_M) + 16 \times \mathrm{Var}(Q_F) \\ &= 9 \times 1 + 16 \times 1 \\ &= 25\end{aligned}$$

6d. [5 points] Use Chebyshev's Inequality to prove an upper bound on the probability of getting an A.

Answer: Again, we are interested in $\Pr[T \geq 60]$, but we need to do some manipulation before we can apply Chebyshev's Inequality.

$$\begin{aligned}\Pr[T \geq 60] &= \Pr[T - 35 \geq 25] \\ &\leq \Pr[|T - 35| \geq 25]\end{aligned}$$

The last step here is a crucial one! Convince yourself this is true before moving on.

Now we have our probability in a form we can apply Chebyshev's Inequality to, which yields

$$\Pr[T \geq 60] \leq \Pr[|T - 35| \geq 25] \leq \frac{\mathrm{Var}(T)}{25^2} = \frac{25}{25^2} = \frac{1}{25}$$

This is a much tighter upper bound than what we found in part (b). If I were you, I would seriously reconsider taking this class...

7. **Golden-I.I.D.** [6 points] You have a random variable X whose expectation you'd like to estimate. You have the ability to draw samples from X's distribution, and you know nothing about $\mathrm{E}(X)$, and you only know that $\mathrm{Var}(X) \leq 10$. (You may assume $\mathrm{Var}(X) = 10$, since this is the worst case.) To estimate $\mathrm{E}(X)$, you take n i.i.d. samples of X and average them. Using the bound based on Chebyshev's Inequality (equation (3) in lecture note 17):

7a. [2 points] Suppose you take 1000 samples. How confident are you that your estimate is within an absolute error of 0.5?

Answer: We have $n = 1000$ samples, and $\varepsilon = 0.5$ error bound. We have $n \geq \frac{\sigma^2}{\varepsilon^2 \delta}$, and wish to find a bound on δ, our confidence. Since $n > 0$ and $\delta > 0$, we manipulate this equation and see that $\delta \geq \frac{\sigma^2}{\varepsilon^2 n} = \frac{\sqrt{10}^2}{0.5^2 \cdot 1000} = 0.04$. This means we are 96% sure in our answer.

7b. [2 points] Suppose instead that you want an absolute error of at most 2 and a confidence parameter of 0.02 (you want to be "98% confident"). How many samples do you need?

Answer: We have $\varepsilon = 2$ error bound, and $\delta = 0.02$ confidence. We have $n \geq \frac{\sigma^2}{\varepsilon^2 \delta}$ and wish to find a bound on n, our number of samples. Plugging in our numbers, $n \geq \frac{\sqrt{10}^2}{2^2 \cdot 0.02} = 125$, so we should take at least 125 samples to achieve our desired error and confidence.

7c. [2 points] Suppose instead that you take 2500 samples and you want a confidence parameter of 0.1 ("90% confident"). What absolute error bound will you get with this confidence?

Answer: We have $n = 2500$ samples and want $\delta = 0.1$ confidence. We have $n \geq \frac{\sigma^2}{\varepsilon^2 \delta}$, and wish to find a bound on ε, our error bound. Since $n > 0$ and $\varepsilon > 0$, we manipulate this equation and see that $\varepsilon \geq \sqrt{\frac{\sigma^2}{\delta \cdot n}} = \sqrt{\frac{\sqrt{10}^2}{0.1 \cdot 2500}} = 0.2$, so our error bound is 0.2.

8. **On Her Majesty's Secret Estimation Team** [10 points] Suppose you have a random variable X with distribution $\mathrm{Geom}(p)$, and you'd like to estimate p.

8a. [5 points] Since $\mathrm{E}(X) = \frac{1}{p}$, a natural approach to estimate p is to get a sample from X and then take the reciprocal $\frac{1}{X}$. Show that this is *not* an unbiased estimator for p.

Answer: If $Z = \frac{1}{X}$ were an unbiased estimator of p, then $\mathrm{E}(Z) = p$ must hold. We know that $\mathrm{E}(Z) = \sum_{i=1}^{\infty} \frac{1}{i} \cdot \Pr[X = i] = \sum_{i=1}^{\infty} \frac{\Pr[X=i]}{i} = \sum_{i=1}^{\infty} \frac{(1-p)^{i-1}p}{i}$. Intuitively, this number is larger than p, since it's a sum of p (when $i = 1$) and additional positive numbers. Since $\mathrm{E}(Z) > p$, Z is not an unbiased estimator of p.

8b. [5 points] Suppose you take n i.i.d. samples $X_1, \ldots, X_n$, and you let Y be the fraction of these samples that equal 1. Show that Y *is* an unbiased estimator for p.

Answer: Y is the fraction of samples that equal 1. Define indicator random variable Y_i that is 1 iff X_i is 1, and 0 otherwise. Thus, $Y = \frac{1}{n} \sum_{i=1}^{n} Y_i$. By linearity of expectation, $\mathrm{E}(Y) = \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}(Y_i)$, and since $Y_i = 1$ iff $X_i = 1$, we know that $\mathrm{E}(Y_i) = \Pr[Y_i = 1] = \Pr[X_i = 1] = p$. Thus, $\mathrm{E}(Y) = \frac{1}{n} \sum_{i=1}^{n} p = p$. Since $\mathrm{E}(Y) = p$, we have an unbiased estimator.

9. **From Russia With 30–Love** [8 points] I am playing in a tennis tournament, and I am up against a player I have watched but never played before. Based on what I have seen, I consider three possible models for our relative strengths:

- Model A: We are evenly matched, so that each of us is equally likely to win each game.
- Model B: I am slightly better, so that I win each game independently with probability 0.6.

- Model C: My opponent is slightly better and wins each game independently with probability 0.6.

Before we play, I consider each of these possibilities to be equally likely. In our match, we play until one player wins three games. I win the second game, but my opponent wins the first, third, and fourth games. After the match, what is the posterior probability of model C (i.e., that my opponent is slightly better than me)?

Answer: Define events A, B, and C representing models A, B, and C being correct, respectively. Define event G_i being me winning game i, and G_{obs} as the event $\overline{G_1} \cap G_2 \cap \overline{G_3} \cap \overline{G_4}$. We must then find $\Pr[C|G_{obs}]$.

$$
\begin{aligned}
\Pr[C|G_{obs}] &= \frac{\Pr[G_{obs}|C]\Pr[C]}{\Pr[G_{obs}]} \\
&= \frac{\Pr[G_{obs}|C]\Pr[C]}{\Pr[G_{obs}|A]\Pr[A] + \Pr[G_{obs}|B]\Pr[B] + \Pr[G_{obs}|C]\Pr[C]} \\
&= \frac{(0.6 \cdot 0.4 \cdot 0.6 \cdot 0.6) \cdot \frac{1}{3}}{(0.5 \cdot 0.5 \cdot 0.5 \cdot 0.5) \cdot \frac{1}{3} + (0.4 \cdot 0.6 \cdot 0.4 \cdot 0.4) \cdot \frac{1}{3} + (0.6 \cdot 0.4 \cdot 0.6 \cdot 0.6) \cdot \frac{1}{3}} \\
&= \frac{864}{1873} \\
&\approx 0.4613
\end{aligned}
$$