Final Review

Computer Science Mentors 70

Dec 7 - Dec 12
Fall 2020

Prepared by: Aishani Sil, Austin Lei, Agnibho Roy, Debayan Bandyopadhyay, Abinav Routhu

Roast us here: *https://tinyurl.com/csm70-feedback20*

I Pre-Midterm

1. **Vertex Colorings**
   Show by induction that for $n \geq 3$, K_n can be vertex colored by n colors.

Solution: We proceed by induction.
Base case: $n = 3$
If $n = 3$, then we color each vertex a different color, and we are done.
Inductive step: Suppose that K_m can be vertex colored by m colors, for $m \geq 1$. We show that K_{m+1} can be vertex colored by $m + 1$ colors.
Isolate one vertex of K_{m+1} and remove the edges connected to that vertex. The remaining m vertices form K_m, so by the inductive hypothesis, the graph can be m colored. Then coloring the removed vertex with the $m + 1$th color, it does not share a color with the other vertices, so we are done.

2. **CRT, FLT, and Friends**

(a) Find $7^{17} \pmod{15}$.

Solution: Remember that by FLT, $a^{(p-1)(q-1)} \equiv 1 \pmod{pq}$.

$$7^{2 \cdot 4} \equiv 1 \pmod{(3 \cdot 5)}$$
$$7^{17} \equiv 7^8 \cdot 7^8 \cdot 7 \equiv 1 \cdot 1 \cdot 7 \pmod{15}$$

(b) Find $7^{60} \pmod{77}$.

Solution: We can solve $7^{60} \pmod{11}$ and $7^{60} \pmod{7}$ separately, and use CRT to get the answer $\pmod{77}$.

$$7^{60} \equiv 0 \pmod{7}$$
$$7^{60} \equiv (7^{10})^6 \equiv 1^6 \pmod{11}$$
$$x \equiv 1 \pmod{11}$$
$$x \equiv 0 \pmod{7}$$

By inspection, we arrive at 56, but mechanically applying CRT would also give the same answer.

3. **Breaking the Encryption**
   Austin, having found one large prime p but struggling to find another for RSA, gives up and decides to make $N = p^2$, with e and d, such that $ed \equiv 1 \pmod{p(p-1)}$. (Recall that $x^{p(p-1)} \equiv 1 \pmod{p^2}$ when $p \nmid x$.) Aishani knows that $N = p^2$ and knows the value of e, but her computer is unable to do square roots or division quickly. However, her computer is very good at addition, subtraction, and multiplication, including taking powers. How can Aishani quickly determine the value of p?

Solution: Since $N = p^2$, $x^N \equiv x^{p(p-1)+p} \equiv x^p \pmod N$. Moreover, using the Euclidean algorithm, Aishani can find the inverse of x modulo N. Thus, for any x, Aishani can compute the value of x^{p-1} modulo N. However, by Fermat's Little Theorem, $x^{p-1} \equiv 1 \pmod p$. Thus, Aishani just needs to choose a value of x, of which there are many, such that $x^{p-1} \not\equiv 1 \pmod N$. Then we have that $p \mid x^{p-1} - 1$ and $p^2 \nmid x^{p-1} - 1$, so the gcd of $x^{p-1} - 1$ and N is p, which can be computed quickly using the Euclidean Algorithm. Thus, following these quick steps, Aishani can determine the value of p.

II Counting and Introductory Probability

1. **Fun with Polya's Urn**

Consider an urn of balls, that initially contains r red balls and g green balls. The scheme is that on every turn, if we draw a ball of some color, then we put the ball back, and then put another ball of the same color back in the urn. For this question, consider $r = g = 1$.

(a) What is the probability that we have more red balls than green balls after n turns? Assume n to be odd.

Solution: Since we start with the same number of red and green balls, we would expect by symmetry that since we are drawing uniformly at random, that no ball of a certain color is more likely to have a higher count in the urn than another color. Since n is odd, $R_n > G_n$ and $R_n < G_n$ (the number of red and green balls after the nth draw) occur with equal probability. Thus, it is $\frac{1}{2}$.

(b) Still considering $r = g = 1$, show that the probability that we have the same number of red balls as green balls after n turns is $\frac{1}{n+1}$.

Solution: We will approach this question by first looking at the sequences in which we draw an equal number of red and green balls, and multiply that by the probability of observing each of those sequences. First, notice that we have $\binom{n}{\frac{n}{2}}$ ways to organize a sequence of $\frac{n}{2}$ and $\frac{n}{2}$ blue balls. Now we need to find the probability of each of those sequences. We know that if we draw a red ball $\frac{n}{2}$ times, then there are initially 1 way to draw the first red ball, 2 ways to draw the second red ball, so on and so forth till $\frac{n}{2}$ choices for drawing the $\frac{n}{2}$th red ball. The same goes for drawing green balls. The total number of balls that we can choose from to form the denominator of our probability, is simply going to be $(n + 1)!$ because we have 2 balls to choose from on our first turn, 3 balls to choose from on our second turn, so and and so forth to n balls to choose from on our nth turn. Our total probability is then:

$$\binom{n}{\frac{n}{2}} \frac{\left(\frac{n}{2}\right)! \cdot \left(\frac{n}{2}\right)!}{(n+1)!} = \frac{n!}{\left(\frac{n}{2}\right)! \cdot \left(\frac{n}{2}\right)!} \frac{\left(\frac{n}{2}\right)! \cdot \left(\frac{n}{2}\right)!}{(n+1)!} = \frac{1}{n+1}$$

(c) Still considering $r = g$, what is the probability that we have more red balls than green balls after n turns? Assume n to be even this time.

Solution: We can combine our answers from part (a) and part (b). We know that the events that we have more red then green and more green than red balls is symmetric, but if n is even, we also have the case where $R_n = G_n$, which occurs with probability $\frac{1}{n+1}$. Thus, we have:

$$P(R_n > G_n) = \left(\frac{1}{2}\right)\left(1 - \frac{1}{n+1}\right)$$

(d) Let the event R_1 denote the first time that we draw a red ball from this bag. Show that $E[R_1]$ is infinite. (*hint*: think about how you can use the tail sum definition of expectation)

Solution: Thinking about using the tail sum definition, we will likely have to look at $P(R_1 \geq n)$. This is the probability that on at least the nth turn, we draw a red ball for the first time. Equivalently, it means that we have been drawing green balls for the previous $n-1$ turns. There is only one sequence of draws that this corresponds to so looking at the probability of this sequence:

$$P(G_1 G_2 \ldots G_{n-1}) = \left(\frac{1}{2}\right)\left(\frac{2}{3}\right)\ldots\left(\frac{n-1}{n}\right) = \frac{1}{n}$$

Using the tail sum:

$$E[R_1] = \sum_{n=1}^{\infty} P(R_1 \geq n) = \sum_{n-1}^{\infty} \frac{1}{n} = +\infty$$

By the divergence of the harmonic series.

2. **Guess and Check**

Thomas did not attend any of his CSM sections, and has resorted to guessing answers on the CS70 final. The questions are structured in a mysterious way, and the number of minutes it takes him to get a question right by guessing is distributed $\sim Geom(\frac{1}{10})$ (he also gets some indication that his answer is correct). To save time, Thomas has decided that he will give up on a question if it takes him more than 5 minutes, and he will also randomly skip $\frac{1}{4}$ of the questions. What is the expected amount of time it will take Thomas to finish a 20 question exam?

Solution: Let us define T to be the time taken to complete the whole exam, and $T_1, T_2, \ldots, T_{20}$ to be the times taken for each question. Given that $E[T] = E[T_1] + E[T_2] + \ldots + E[T_{20}]$, we will find some $E[T_k]$.

$$E[T_k] = 0 \cdot P[\text{skips question}] + \sum_{i=1}^{5} i \cdot P[\text{takes } i \text{ minutes to guess}] + 5 \cdot P[\text{couldn't guess in 5 minutes}]$$

Note that $P[\text{takes } i \text{ minutes to guess}] = (\frac{9}{10})^{i-1}(\frac{1}{10})$. Also, if $X \sim Geom(\frac{1}{10})$, $P[X > 5] = (\frac{9}{10})^5$. Plugging back in,

$$E[T_k] = 0 + \sum_{i=1}^{5} i \cdot (\frac{9}{10})^{i-1}(\frac{1}{10}) + 5(\frac{9}{10})^5.$$

$$E[T] = \sum_{k=1}^{20} E[T_k] = \frac{\frac{1}{2} + 9 + 9^2 + 9^3 + 9^4 + 10(9^5)}{5}.$$

III Distributions

1. **Basketball**

Jordan and James are both playing basketball. Jordan makes shots with probability 0.6, and James makes shots with probability 0.7, and the probabilities of making shots are independent between different shots. Starting at time $t = 1$, Jordan and James take a shot simultaneously, and continue to do so at every integer time t.

(a) Suppose that exactly one person makes their first shot. What is the probability that it was Jordan?

Solution: The probability that Jordan was the person who made the first shot is $0.6 \cdot 0.3 = 0.18$, since James also needs to miss his shot. Similarly, the probability that James was the person who made the first shot is $0.7 \cdot 0.4 = 0.28$. Thus, by Bayes' Rule, the probability that Jordan the first shot given that exactly one person made their first shot is $\frac{0.18}{0.18+0.28} = \frac{9}{23}$.

(b) What is the expected amount of time until someone makes a shot?

Solution: We can model the process of someone making their first shot as a Geometric variable. The probability that someone makes a shot at any time is $0.6 + 0.7 - 0.6 \cdot 0.7 = 0.88$. Thus, the expected amount of time it takes for someone to make a shot is $\frac{1}{0.88}$.

(c) What is the probability that James makes a shot before Jordan? (Note that this does not count the situation where they both make their first shot at the same time.)

Solution: Suppose that James makes his first shot at time $t = k$, and Jordan has not made a shot. The probability of this occuring is $(0.4)^k(0.3)^{k-1}(0.7)$, since Jordan needs to miss all of shots, and James needs to miss all but his last shot. Thus, the probability that James makes a shot before Jordan is

$$\sum_{k=1}^{\infty}(0.4)^k(0.3)^{k-1}(0.7) = \frac{0.4 \cdot 0.7}{1 - 0.4 \cdot 0.3} = \frac{7}{22}.$$

2. **Buggy Code**

Leanne is debugging her partner's project code. She knows the number of mistakes that her partner makes is Poisson(λ) distributed, where λ is an integer chosen uniformly at random from the interval $[1, 10]$, inclusive, but she does not know the value of λ. Let X be the random variable representing the number of mistakes that Leanne's partner makes.

(a) What is the expected value of X?

Solution: Given that $\lambda = k$, the expected value of X is k, since the mean of a Poisson(λ) variable is λ. In other words, $E[X|\lambda = k] = k$. Thus,

$$\begin{aligned} E[X] &= \sum_{k=1}^{10} E[X|\lambda = k]P(\lambda = k) \\ &= \sum_{k=1}^{10} k \cdot \frac{1}{10} \\ &= \frac{\frac{10 \cdot 11}{2}}{10} \\ &= \frac{11}{2}. \end{aligned}$$

(b) Frustrated with her partner, Leanne decides to write her own code. However, the number of mistakes she makes is Poisson(λ) distributed, where λ is chosen from a Binomial$(10, 0.2)$ distribution, but the value of λ is unknown. Let Y be the random variable representing the number of mistakes that Leanne makes. Compute the expected value of Y.

Solution: Given that $\lambda = k$, the expected value of X is k, since the mean of a Poisson(λ) variable is λ. In other words, $E[X|\lambda = k] = k$. Thus,

$$\begin{aligned} E[Y] &= \sum_{k=0}^{10} E[Y|\lambda = k]P(\lambda = k) \\ &= \sum_{k=0}^{10} kP(\lambda = k). \end{aligned}$$

However, note that $\sum_{k=0}^{10} kP(\lambda = k)$ is exactly the expression for the expected value of the Binomial$(10, 0.2)$ distribution. Thus, $E[Y] = 0.2 \cdot 10 = 2$.

3. **Balls in Bins™**

You are playing a game where you are randomly throwing balls into 15 lined-up bins, and each bin has an Oski sticker that you get to keep if at least one ball lands in that bin.

(a) If you're given 10 balls to throw, what is the expected number of stickers you win?

Solution: For each sticker, we need at least one ball to hit the corresponding bin. Since this value is harder to calculate as it involves a summation, we will instead try to find the number of stickers we cannot win (so the number of bins that remain empty, which we'll define as X). Let's now define X_i to be the probability that no ball hits bin i, which is $(\frac{14}{15})^{15}$. Using linearity of expectation,

$$E[X] = E[X_1] + E[X_2] + ... + E[X_{15}] = 15(\frac{14}{15})^{15}.$$

Remember that we are actually looking for the number of stickers we do win, so that is $15 - E[X] = 15 - 15(\frac{14}{15})^{15}$.

(b) What is the variance of the above?

Solution: We can find $E[X]^2$ trivially from the last part, so let us now find $E[X^2]$.

$$E[X^2] = \sum_{ij} E[X_i X_j] = \sum_{i} E[X_i^2] + \sum_{i \neq j} E[X_i X_j] = E[X] + (15 \cdot 14)(\frac{13}{15})^{15}.$$

Note that $E[X_i X_j]$ is the probability that no balls go into any given two bins.
Now, remember that $\text{Var}(X) = E[X^2] - E[X]^2 = (\frac{14}{15})^{14} + 14(\frac{13}{15})^{14} - ((\frac{14}{15})^{15})^2$. Further simplification is left as an exercise to the reader.

IV Continuous RVs and Applications

1. **Lognormal Distribution**

In this question, we will explore an interesting distribution called the lognormal distribution. The distribution is described as $X = e^{\mu + \sigma Z}$, where $Z \sim \mathcal{N}(0, 1)$.

(a) First, derive the CDF of X. You may represent your answer in terms of ϕ, the CDF of the standard normal distribution.

Solution:

$$\begin{aligned} P(X \leq x) &= P\left(e^{\mu + \sigma Z} \leq x\right) \\ &= P(\mu + \sigma Z \leq \ln x) \\ &= P\left(Z \leq \frac{\ln x - \mu}{\sigma}\right) \\ &= \phi\left(\frac{\ln x - \mu}{\sigma}\right) \end{aligned}$$

(b) Now, find the PDF of the lognormal distribution (remember the chain rule here when taking the derivative).

Solution: We derive the CDF that we found in part (a). Recall the PDF of the standard normal, Z. Using the chain rule,

$$\begin{aligned}\frac{d}{dx}\phi\left(\frac{\ln x-\mu}{\sigma}\right) &= \frac{d}{dx}\left(\frac{\ln x}{\sigma}\right)f_Z\left(\frac{\ln x-\mu}{\sigma}\right)\\ &= \frac{1}{\sigma x}\left(\frac{1}{\sqrt{2\pi}}e^{\frac{((\ln x-\mu)/\sigma)^2}{2}}\right)\\ &= \frac{1}{\sigma x\sqrt{2\pi}}e^{\frac{(\ln x-\mu)^2}{2\sigma^2}}\end{aligned}$$

(c) How would you find the expectation of the lognormal distribution? Leave your answer as an integral in terms of the the PDF of the normal distribution, $f_Z(z)$.

Solution:

$$\int_0^\infty e^{\mu+\sigma x}f_Z(x)dx$$

2. **Bounding the Irwin-Hall Distribution**

There is a very interesting distribution, which is known as the Irwin-Hall Distribution. It is defined as the sum of i.i.d continuous uniform distributions $U_i \sim U[a,b]$. Thus, I can be defined as:

$$I = \sum_{i=1}^{n} U_i$$

The PDF of this distribution is quite ugly, so we will attempt to bound some probabilities using the inequalities learned from this class. For all parts, assume $a \geq 0$. Let $c = \frac{3n(b-a)}{4}$

(a) Using Markov's inequality, bound the probability that the Irwin-Hall distribution exceeds c.

Solution: We find the expectation:

$$E[I] = E\left[\sum_{i=1}^{n} U_i\right] = \sum_{i=1}^{n} E[U_i] = \frac{n(b-a)}{2}$$

Now we can use Markov's inequality:

$$P(I \geq c) = P\left(I \geq \frac{3n(b-a)}{4}\right) \leq \left(\frac{n(b-a)}{2}\right)\left(\frac{4}{3n(b-a)}\right) = \frac{2}{3}$$

(b) Using Chebyshev's inequality, bound the probability that the Irwin-Hall distribution exceeds c.

Solution: We find the variance, and keep in mind U_i are all independent. Recall that $\text{Var}(U_i) = \frac{1}{12}(b-a)^2$

$$\text{Var}(I) = \text{Var}\left(\sum_{i=1}^{n} U_i\right) = \sum_{i=1}^{n}\text{Var}(U_i) = \frac{n(b-a)^2}{12}$$

Now using Chebyshev's inequality:

$$\begin{aligned}
P(|I - E[I]| \geq a) &= P\left(\left|I - \frac{n(b-a)}{2}\right| \geq \frac{n(b-a)}{4}\right) \\
&\leq \frac{\mathrm{Var}(I)}{a^2} \\
&= \left(\frac{n(b-a)^2}{12}\right)\left(\frac{16}{n^2(b-a)^2}\right) \\
&= \frac{4}{3n}
\end{aligned}$$

3. **Tired of Balls and Urns Yet?**
An urn contains six balls, of which three are red and three are green. In each step, two balls are selected at random. If one of them is red, and the other is green, then we discard them and replace them by two blue balls, and if both of the balls are blue, then we replace those blue balls with an equal amount of red and green balls. Otherwise, we do not do anything. Find the probability that if we start with an equal number of balls of every color, what is the probability that we reach 6 blue balls before 0 blue balls in the bag?

Solution: We notice that this is a Markov chain, and we have a total of 4 states, which we can uniquely represent by the number of red balls, since we know we have the same number of green balls at all times and the remaining balls are blue. The states are state 0 (0R 0G 6B), state 1 (1R 1G 4B), state 2 (2R 2G 4B), and state 3 (3R 3G 0B). If we have i each of red and green balls, then

$$P(1R1G) = \frac{i \cdot i}{\binom{6}{2}}$$

which is the transition from state i to $i-1$. Furthermore, if we have j blue balls, then

$$P(2B) = \frac{\binom{j}{2}}{\binom{6}{2}}$$

which is the transition from state i to $i+1$. The remaining probability are self loops. The transition matrix is then:

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ \frac{1}{15} & \frac{8}{15} & \frac{6}{15} & 0 \\ 0 & \frac{4}{15} & \frac{10}{15} & \frac{1}{15} \\ 0 & 0 & \frac{9}{15} & \frac{6}{15} \end{bmatrix}$$

The probability desired can be modeled recursively by considering the hitting probability of state 0 before state 3, given you are at state i, which we denote as $\alpha(i)$. Naturally, $\alpha(0) = 1$ and $\alpha(3) = 0$. Then our equations are:

$$\begin{aligned}
\alpha(1) &= \frac{1}{15}\alpha(0) + \frac{8}{15}\alpha(1) + \frac{6}{15}\alpha(2) \\
\alpha(2) &= \frac{4}{15}\alpha(1) + \frac{10}{15}\alpha(2) + \frac{1}{15}\alpha(3)
\end{aligned}$$

Solving the equations above we get $\alpha(2) = \frac{4}{11}$