Chris Piech
CS 109

Oct 30, 2017

Practice Midterm Solutions

With solutions by Mehran + Will

1. a. The answer to this question is simply a multinomial coefficient, which can be written/computed in numerous ways:

$$\binom{12}{5,4,3} = \frac{12!}{5!4!3!} = \binom{12}{5}\binom{7}{4}\binom{3}{3} = \binom{12}{5}\binom{7}{4}$$

b. $\binom{10}{3,4,3} + \binom{10}{5,2,3} + \binom{10}{5,4,1}$

We select (remove) two fruits of the same type to give to Larry and Sergey (there is only 1 way to do this for each type of fruit). The remaining 10 fruits are then distributed to the remaining 10 students. The three terms above correspond respectively to apples, mandarins, and persimmons being given to Larry and Sergey.

Note that each of the multinomial coefficients could have been written in different ways (analogously to what was shown in part (a)).

c. $\binom{10}{3,4,3} + \binom{10}{5,2,3} + \binom{10}{5,4,1} + \binom{10}{4,3,3} + \binom{10}{4,4,2} + \binom{10}{5,3,2}$

We select two fruits to remain in the bag and the remaining 10 fruits are then distributed to the 10 students. The six terms above correspond respectively to the cases where the two fruits left in the cooler are: (a) 2 apples, (b) 2 mandarins, (c) 2 persimmons, (d) 1 apple and 1 mandarin, (e) 1 apple and 1 persimmon, and (f) 1 mandarin and 1 persimmon.

Note that each of the multinomial coefficients could have been written in different ways (analogously to what was shown in part (a)).

2. Let W = amount the team wins in dollars, and let D be the distance of the throw. $D \sim \mathrm{N}(7, 4)$.

$$\begin{aligned}
E[W] &= \sum_{w:p_W(w)>0} w \cdot p_W(w) \\
&= 0 \cdot p_W(0) + 5 \cdot p_W(5) + 20 \cdot p_W(20) \\
&= 5 \cdot p_W(5) + 20 \cdot p_W(20) \\
&= 5 \cdot P(4 < D \leq 7) + 20 \cdot P(7 < D \leq 10) \\
&= 5 \cdot P\left(\frac{4-7}{2} < \frac{D-7}{2} \leq \frac{7-7}{2}\right) + 20 \cdot P\left(\frac{7-7}{2} < \frac{D-7}{2} \leq \frac{10-7}{2}\right) \\
&= 5 \cdot P(-1.5 < Z \leq 0) + 20 \cdot P(0 < Z \leq 1.5) \\
&= 5 \cdot (\Phi(0) - \Phi(-1.5)) + 20 \cdot (\Phi(1.5) - \Phi(0)) \\
&= 5 \cdot (\Phi(0) - (1 - \Phi(1.5))) + 20 \cdot (\Phi(1.5) - \Phi(0)) \\
&= \boxed{5 \cdot (0.5 - (1 - 0.9332)) + 20 \cdot (0.9332 - 0.5)} \\
&= 5 \cdot (0.9332 - 0.5) + 20 \cdot (0.9332 - 0.5) \\
&= 25 \cdot 0.4332 = 10.83
\end{aligned}$$

(The boxed answer is sufficient; further simplification is not required.)

3. a. We are given the PMF for the random variable X, which is the popularity rank of the song for a random play. So we can plug in $i = 10$:

$$P(X = 10) = \frac{\frac{1}{10}}{\sum_{n=1}^{3\cdot 10^7} \frac{1}{n}}$$

b. Let Y be a random variable equal to the number of times the most popular song is listened to over the course of the day. If we consider each play to be a trial which succeeds if the song is the most popular, then $Y \sim \text{Bin}(n, p)$, where n is the number of plays (1 billion = 10^9) and p is the probability that the song is the most popular. From the PMF, the probability that the song is the most popular is

$$p = P(X = 1) = \frac{\frac{1}{1}}{\sum_{n=1}^{3\cdot 10^7} \frac{1}{n}}$$

Here, n is very large, and p is fairly small (using the fact that $\sum_{n=1}^{3\cdot 10^7} \frac{1}{n} \approx 17.8$, we can figure out that $p = \frac{1}{17.8} \approx 0.056$). So a Poisson approximation is a good choice here. We can approximate $Y \approx W \sim \text{Poi}(\lambda = np)$.

$$P(Y > 10^8) \approx P(W > 10^8) = 1 - \sum_{i=0}^{10^8} e^{-np} \frac{(np)^i}{i!} = 1 - \sum_{i=0}^{10^8} e^{-\frac{10^9}{17.8}} \frac{\left(\frac{10^9}{17.8}\right)^i}{i!}$$

(The last step, plugging in the values we already defined for n and p, is not necessary for full credit.)

4. a. Let X_i = the value rolled on die i, where $1 \le i \le 4$. $P(X \ge k) = P(X_1 \ge k, X_2 \ge k, X_3 \ge k, X_4 \ge k) = \left(\frac{6-k+1}{6}\right)^4$, since all four rolls must be greater than or equal to k.

b. Using the definition of expectation:

$$\begin{aligned} E[X] &= \sum_{x=1}^{6} x \cdot P(X = x) = \sum_{x=1}^{6} x \cdot [P(X \ge x) - P(X \ge x + 1)] \\ &= \sum_{x=1}^{6} x \cdot \left[\left(\frac{6 - x + 1}{6}\right)^4 - \left(\frac{6 - x}{6}\right)^4\right] \end{aligned}$$

Alternatively, one can use a property covered in Lecture 12 (and therefore not required knowledge for the midterm), which is that if X is non-negative, then:

$$E[X] = \sum_{x=1}^{6} P(X \ge x) = \sum_{x=1}^{6} \left(\frac{6 - x + 1}{6}\right)^4 = \left(\frac{6}{6}\right)^4 + \left(\frac{5}{6}\right)^4 + \left(\frac{4}{6}\right)^4 + \left(\frac{3}{6}\right)^4 + \left(\frac{2}{6}\right)^4 + \left(\frac{1}{6}\right)^4$$

The two expressions to compute $E[X]$ above are, indeed, equivalent.

c. $E[S] = E[T - X] = E[T] - E[X]$
Let X_i = the value rolled on die i, where $1 \le i \le 4$. As computed in class, we know that $E[X_i] = 3.5$ for all $1 \le i \le 4$.
$E[T] = E[X_1 + X_2 + X_3 + X_4] = E[X_1] + E[X_2] + E[X_3] + E[X_4] = 4(3.5) = 14$
So, $E[S] = 14 - E[X]$, where $E[X]$ is as computed in part (b).

5. Let X = lifetime of screen in our laptop.

Let event A = manufacturer A produced the screen.

Let event B = manufacturer B produced the screen.

a. We want to compute $P(A \mid X > 18)$. Using Bayes Theorem, we have:

$$P(A \mid X > 18) = \frac{P(X > 18 \mid A)P(A)}{P(X > 18)} = \frac{(1 - P(X \le 18 \mid A)) \cdot 0.5}{P(X > 18)}$$

Noting that $(X \mid A) \sim \mathrm{N}(20, 4)$, we have:

$$\begin{aligned} P(A \mid X > 18) &= \frac{(0.5)\left(1 - P(\frac{X-20}{2} \le \frac{18-20}{2})\right)}{P(X > 18)} \\ &= \frac{(0.5)\Phi(1)}{P(X > 18)} = \frac{(0.5)(0.8413)}{P(X > 18)} \end{aligned}$$

Now, we need to compute $P(X > 18)$:

$$\begin{aligned} P(X > 18) &= P(X > 18 \mid A)P(A) + P(X > 18 \mid B)P(B) \\ &= P(X > 18 \mid A)(0.5) + P(X > 18 \mid B)(0.5) \\ &= 0.5 \cdot \left(1 - P\left(\frac{X - 20}{2} \le \frac{18 - 20}{2}\right)\right) + 0.5\left[1 - \left(1 - e^{-\frac{18}{20}}\right)\right] \\ &= 0.5 \cdot (1 - P(Z \le -1)) + 0.5e^{-\frac{9}{10}} \\ &= 0.5 \cdot (1 - (1 - P(Z \le 1))) + 0.5e^{-\frac{9}{10}} \\ &= 0.5\Phi(1) + 0.5e^{-\frac{9}{10}} \\ &= 0.5 \cdot 0.8413 + 0.5e^{-\frac{9}{10}} \end{aligned}$$

Substituting $P(X > 18)$ into the expression for $P(A \mid X > 18)$, yields the answer:

$$P(A \mid X > 18) = \frac{(0.5)(0.8413)}{P(X > 18)} = \frac{0.8413}{0.8413 + e^{-\frac{9}{10}}}$$

b. Here, we want to compute $P(B \mid X > 18)$. Using Bayes Theorem, we have:

$$P(B \mid X > 18) = \frac{P(X > 18 \mid B)P(B)}{P(X > 18)} = \frac{(1 - P(X \le 18 \mid B)) \cdot 0.5}{P(X > 18)}$$

Noting that $(X \mid B) \sim \text{Exp}(1/20)$, we have:

$$P(B \mid X > 18) = \frac{0.5\left(1 - \left(1 - e^{-\frac{18}{20}}\right)\right)}{P(X > 18)} = \frac{0.5e^{-\frac{9}{10}}}{P(X > 18)}$$

Substituting the previously computed value for $P(X > 18)$ into the expression for $P(B \mid X > 18)$, yields the final answer:

$$P(B \mid X > 18) = \frac{0.5e^{-\frac{9}{10}}}{P(X > 18)} = \frac{e^{-\frac{9}{10}}}{0.8413 + e^{-\frac{9}{10}}}$$

6. As in the normal birthday problem, we start with $P(\text{at least one shared birthday}) = 1 - P(\text{no shared birthdays})$.

We need to treat the weekday and weekend babies differently. To do this, we can model the number that get born on weekdays as a binomial distribution. Let W be the number of people born on a weekday. $W \sim \text{Bin}(n, 260 \cdot p_a)$.

We then sum over this number (using the general law of total probability) to get the overall probability that there are no collisions:

$$P(\text{no shared birthdays}) = \sum_{i=0}^{n} P(\text{no shared birthdays} \mid W = i) P(W = i)$$

We can take advantage of the fact that all weekdays are equally likely (and same for weekends) to compute the conditional probability by counting:

$$P(\text{no shared birthdays} \mid W = i) = \frac{\binom{260}{i} i! \cdot \binom{105}{n-i} (n-i)!}{260^i 105^{n-i}}$$

And $P(W = i)$ is just the PMF of a binomial:

$$P(W = i) = \binom{n}{i} (260 \cdot p_a)^i (105 \cdot p_b)^{n-i}$$

So the final answer is

$$\boxed{1 - \sum_{i=0}^{n} \frac{\binom{260}{i} i! \cdot \binom{105}{n-i} (n-i)!}{260^i 105^{n-i}} \binom{n}{i} (260 \cdot p_a)^i (105 \cdot p_b)^{n-i}}$$

$$= 1 - \sum_{i=0}^{n} \binom{260}{i} \binom{105}{n-i} n! \cdot {p_a}^i {p_b}^{n-i}$$

The simplification at the end suggests another way of arriving at the right answer: sum up $P(\text{no shared birthdays } \textbf{and } W = i)$ over values of i. The value of that probability, for a given i, can be computed by choosing the i weekdays and $n - i$ weekend days, then arranging the people in the room in all possible orders, and computing the probability that each person has exactly the assigned birthday.

We would expect that any unevenness will lead to a $\boxed{\text{larger}}$ probability of having two people with the same birthday. Intuitively, this is because the more popular birthdays are going to be shared by more people. As an extreme case, consider if no one were born on weekends. Then it would be as if we had 105 fewer days in the year, so there's less room for people to be born on different days.