Due ~~Monday July 27~~ Tuesday July 28 at Noon

1. **Best-of-Five Series — Go Bears!** (60 points, 3/3/10/5/10/5/3/3/3/3/3/3/3/3 points for each part)

   A best-of-five series is the competition between two teams where a team wins the series by winning three games in the series. After a team wins three games, the following games will not be played. An outcome of a best-of-five series is defined by the result of each game, *e.g.*, WWW, WWLW, WLWW, and WLWLW are all different outcomes for a team. Now, two basketball teams, California Golden Bears and Stanford Cardinal, are going to play a best-of-five series.

   (a) How many possible outcomes for Golden Bears to win the series with 3 wins and 2 losses?

   (b) How many possible outcomes for Golden Bears to win the series?

   (c) Prove $\sum_{i=0}^{n} \binom{n+i}{i} = \binom{2n+1}{n+1}$ with explicit calculations ($n \geq 0$).

   (d) If it is a best-of-$(2n+1)$ series, use the above equality to prove that there are $\binom{2n+1}{n+1}$ outcomes for Golden Bears to win the series.

   (e) If it is a best-of-$(2n+1)$ series, explain "there are $\binom{2n+1}{n+1}$ outcomes for Golden Bears to win the series" without any explicit calculation.

   (f) If it is a best-of-(n_1+n_2-1) series where Golden Bears win the series by winning n_1 games in the series, and Cardinal wins the series by winning n_2 games in the series, explain "there are $\binom{n_1+n_2-1}{n_1}$ outcomes for Golden Bears to win the series" without any explicit calculation.

   It is believed that Golden Bears have a 0.7 probability to win each game, if the game is played.

   (g) What is the probability for Golden Bears to win the series with 3 wins and 2 losses?

   (h) What is the probability for Golden Bears to win the series?

   (i) The captain of Cardinal proposes to change the series from best-of-five to best-of-three. Is this change advantageous to Golden Bears or Cardinal? Prove your answer.

   Now, let's consider some conditional probabilities. We still assume that Golden Bears have a 0.7 probability to win each game, if the game is played.

   (j) Pr[(Golden Bears win the series) | (Golden Bears win the first game)] =?

   (k) Pr[(Golden Bears win the first game) | (Golden Bears win the series)] =?

   (l) Pr[Golden Bears win the fifth game] =? (We are considering this before the series starts.)

   (m) Pr[(Golden Bears win the series) | (Golden Bears win the fifth game)] =?

   (n) Pr[(Golden Bears win the fifth game) | (Golden Bears win the series)] =?

   Answer:

(a) Given that Golden Bears win the series with 3 wins and 2 losses, Golden Bears must win the fifth game and two of the first four games, so the number of possible outcomes is $\binom{4}{2} = 6$. We can also list all possible outcomes: WWLLW, WLWLW, WLLWW, LWWLW, LWLWW, and LLWWW.

(b) The number of all possible outcomes for Golden Bears to win the series is the summation of the numbers of the possible outcomes with 3-to-0, 3-to-1, and 3-to-2, so it is $\binom{2}{0} + \binom{3}{1} + \binom{4}{2} = 10$.

(c) We can use induction on m ($0 \le m \le n$) to prove $\sum_{i=0}^{m} \binom{n+i}{i} = \binom{n+m+1}{m}$, *e.g.*, the summation of the first m terms is $\binom{n+m+1}{m}$.

- Base case: when $m = 0$, $\sum_{i=0}^{0} \binom{n+i}{i} = \binom{n+0}{0} = 1 = \binom{n+0+1}{0} = \binom{n+m+1}{m}$.
- Inductive hypothesis: assume when $m = k$, $\sum_{i=0}^{k} \binom{n+i}{i} = \binom{n+k+1}{k}$.
- Inductive step: when $m = k+1$,

$$\begin{aligned}
\sum_{i=0}^{k+1} \binom{n+i}{i} &= \left(\sum_{i=0}^{k} \binom{n+i}{i}\right) + \binom{n+k+1}{k+1} \\
&= \binom{n+k+1}{k} + \binom{n+k+1}{k+1} \\
&= \frac{(n+k+1)!}{k!(n+1)!} + \frac{(n+k+1)!}{(k+1)!n!} \\
&= \frac{(n+k+1)!}{k!n!}\left(\frac{1}{n+1} + \frac{1}{k+1}\right) \\
&= \frac{(n+k+1)!}{k!n!}\left(\frac{n+k+2}{(n+1)(k+1)}\right) \\
&= \frac{(n+k+2)!}{(k+1)!(n+1)!} \\
&= \binom{n+k+2}{k+1}.
\end{aligned}$$

By the induction, we have proved that the summation of the first k terms is $\binom{n+k+1}{k}$, so the summation of the total n terms is $\binom{n+n+1}{n} = \binom{2n+1}{n} = \binom{2n+1}{n+1}$.

(d) The number of all possible outcomes for Golden Bears to win the series is the summation of the numbers of possible outcomes with $(n+1)$-to-0, $(n+1)$-to-1, ..., $(n+1)$-to-n. Do not forget that Golden Bears must win the last game, so it is $\binom{n}{0} + \binom{n+1}{1} + \ldots + \binom{2n}{n} = \sum_{i=0}^{n} \binom{n+i}{i}$. By the above equality, it is $\binom{2n+1}{n+1}$.

(e) We can regard $\binom{2n+1}{n+1}$ as the number of all possible outcomes of a new scenario: **all** $(2n+1)$ **games are played and Golden Bears win exactly** $n+1$ **games**.

- In the original scenario, a possible outcome for Golden Bears to win the series can be mapped to exactly one outcome in the new scenario because, after the $(n+1)$-th win (the last win), we can assume that the following games are still played, and Golden Bears lose all of the following games (*e.g.*, in a best-of-five series, WWW can be mapped to WWWLL; WWLW can be mapped to WWLWL). Besides, different possible outcomes for Golden Bears to win the series are mapped to different outcomes in the new scenario because their W/L sequences have been different before the $(n+1)$-th win (the last win).
- In the new scenario, a possible outcome can be mapped to exactly one outcome in the original scenario because, after the $(n+1)$-th win (the last win), we can assume that the following games are not played (*e.g.*, in a best-of-five series, WWWLL can be mapped to

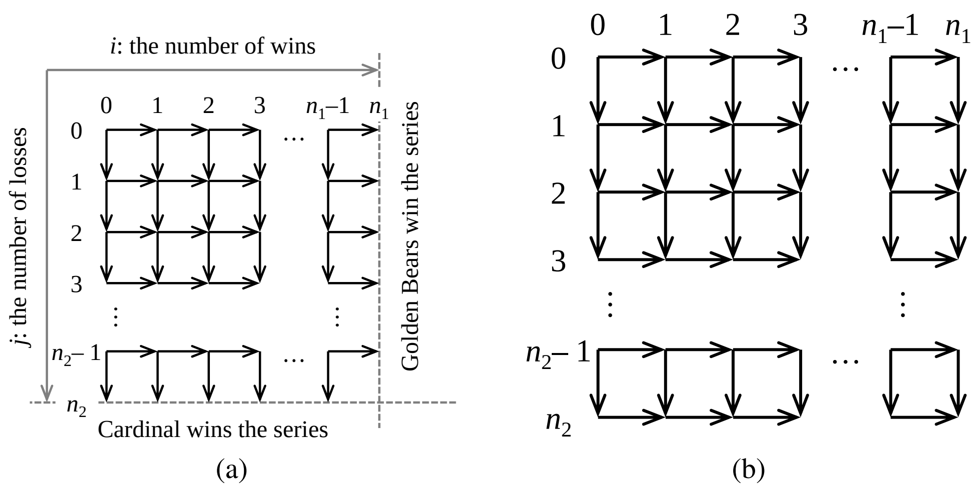

Figure 1: The computation for the number of paths corresponding to n_1 wins and j losses where $j < n_2$.

WWW; WWLWL can be mapped to WWLW). Besides, different possible outcomes are mapped to different outcomes in the original scenario because, again, their W/L sequences have been different before the $(n+1)$-th win (the last win).

As a result, there is a one-to-one mapping between the original scenario and the new scenario, so there are $\binom{2n+1}{n+1}$ outcomes for Golden Bears to win the series.

(f) Similar to the above one, we can regard $\binom{n_1+n_2-1}{n_1}$ as the number of all possible outcomes of a new scenario: **all (n_1+n_2-1) games are played and Golden Bears win exactly n_1 games**.

- In the original scenario, a possible outcome for Golden Bears to win the series can be mapped to exactly one outcome in the new scenario. Besides, different possible outcomes for Golden Bears to win the series are mapped to different outcomes in the new scenario. (The reasons are all the same as the above parts.)
- In the new scenario, a possible outcome can be mapped to exactly one outcome in the original scenario. Besides, different possible outcomes are mapped to different outcomes in the original scenario. (The reasons are all the same as the above parts.)

As a result, there is a one-to-one mapping between the original scenario and the new scenario, so there are $\binom{n_1+n_2-1}{n_1}$ outcomes for Golden Bears to win the series.

Comments: As shown in Figure 1, we can also solve the question by considering the number of paths corresponding to n_1 wins and j losses where $j < n_2$. In Figure 1 (a), it is the summation of the numbers of paths from $(0,0)$ to $(n_1,0)$, from $(0,0)$ to $(n_1,1)$, ..., from $(0,0)$ to (n_1,n_2-1). In Figure 1 (b), it is the number of paths from $(0,0)$ to (n_1,n_2-1).

(g) There are $\binom{4}{2} = 6$ possible outcomes, and the probability of each outcome is $0.3^2 0.7^3$, so the total probability is $\binom{4}{2}0.3^2 0.7^3 \approx 0.185$.

(h) The probability for Golden Bears to win the series is the summation of the probabilities with 3-to-0, 3-to-1, and 3-to-2, so it is $\binom{2}{0}0.7^3 + \binom{3}{1}0.3^1 0.7^3 + \binom{4}{2}0.3^2 0.7^3 \approx 0.837$.

(i) In the best-of-three series, the probability for Golden Bears to win the series is the summation of the probabilities with 2-to-0 and 2-to-1, so it is $\binom{1}{0}0.7^2 + \binom{2}{1}0.3^1 0.7^2 = 0.784$. This change is advantageous to Cardinal. (However, Golden Bears still have a 0.784 probability to win the series.)

(j) Given two events, $\Pr[A|B] = \frac{\Pr[A\cap B]}{\Pr[B]}$. Now, A is the event that Golden Bears win the series, and B is the event that Golden Bears win the first game.

$$\begin{aligned}
& \Pr[\text{ (Golden Bears win the series) } | \text{ (Golden Bears win the first game) }] \\
= & \frac{\Pr[\text{ Golden Bears win the first game and win the series }]}{\Pr[\text{ Golden Bears win the first game }]} \\
= & \frac{\binom{1}{0}0.7^3 + \binom{2}{1}0.3^1 0.7^3 + \binom{3}{2}0.3^2 0.7^3}{0.7} \\
= & \frac{0.64141}{0.7} \\
\approx & \ 0.916.
\end{aligned}$$

(k) A is the event that Golden Bears win the first game, and B is the event that Golden Bears win the series.

$$\begin{aligned}
& \Pr[\text{(Golden Bears win the first game) } | \text{ (Golden Bears win the series) }] \\
= & \frac{\Pr[\text{ Golden Bears win the first game and win the series }]}{\Pr[\text{ Golden Bears win the series }]} \\
= & \frac{\binom{1}{0}0.7^3 + \binom{2}{1}0.3^1 0.7^3 + \binom{3}{2}0.3^2 0.7^3}{\binom{2}{0}0.7^3 + \binom{3}{1}0.3^1 0.7^3 + \binom{4}{2}0.3^2 0.7^3} \\
= & \frac{\binom{1}{0} + \binom{2}{1}0.3^1 + \binom{3}{2}0.3^2}{\binom{2}{0} + \binom{3}{1}0.3^1 + \binom{4}{2}0.3^2} \\
= & \frac{1.87}{2.44} \approx 0.766.
\end{aligned}$$

(l) It must be 2-to-2 after the first four games, and Golden Bears win the fifth game. The probability is $\binom{4}{2}0.3^2 0.7^3 \approx 0.185$. Note that it is possible that the fifth game is not played, so it is not 0.7.

(m) A is the event that Golden Bears win the series, and B is the event that Golden Bears win the fifth game. It is trivial to be 1. It can also be calculated as

$$\begin{aligned}
& \Pr[\text{ (Golden Bears win the series) } | \text{ (Golden Bears win the fifth game) }] \\
= & \frac{\Pr[\text{ Golden Bears win the fifth game and win the series }]}{\Pr[\text{ Golden Bears win the fifth game }]} \\
= & \frac{\binom{4}{2}0.3^2 0.7^3}{\binom{4}{2}0.3^2 0.7^3} \\
= & \ 1.
\end{aligned}$$

(n) A is the event that Golden Bears win the fifth game, and B is the event that Golden Bears win the series.

$$\begin{aligned}
& \Pr[\text{ (Golden Bears win the fifth game) } | \text{ (Golden Bears win the series) }] \\
= & \frac{\Pr[\text{ Golden Bears win the fifth game and win the series }]}{\Pr[\text{ Golden Bears win the series }]} \\
= & \frac{\binom{4}{2}0.3^2 0.7^3}{\binom{2}{0}0.7^3 + \binom{3}{1}0.3^1 0.7^3 + \binom{4}{2}0.3^2 0.7^3} \\
= & \frac{\binom{4}{2}0.3^2}{\binom{2}{0} + \binom{3}{1}0.3^1 + \binom{4}{2}0.3^2} \\
= & \frac{0.54}{2.44} \approx 0.221.
\end{aligned}$$

2. **Bayesian Inference and Pancakes** (15 points, 3 points for each part)

Kunal is making golden-brown pancakes and you are hungry!

(a) Kunal serves up a stack of 3 pancakes, but he forgot to butter the pan! Pancake A is perfect (golden-brown on both sides), pancake B is burnt on one side, and pancake C is burnt on both sides. The top of the stack is burnt. What's the probability that the other side of the top pancake is also burnt? Justify your answer.

(b) Kunal agrees that a burnt pancake on top of the stack looks un-appetizing, and suggests flipping the stack over. In the same situation as before, what's the probability that the pancake side touching the plate is burnt?

(c) Suppose Kunal makes a stack of n pancakes such that x pancakes are burnt on both sides and y pancakes are burnt on one side. If the top of the stack is burnt, what's the probability that the other side of the top pancake is also burnt? What if the top of the stack is golden-brown? Justify your answer.

(d) You asked for chocolate chips, so Kunal adds lots of chocolate chips to the batter. He makes a stack of m pancakes next to the stack of n pancakes from before. However, the k-th pancake ($1 \leq k \leq m$) in the new stack only has a k/m chance of having chocolate chips (independent from the rest of the pancakes). If you choose a pancake randomly from either stack, what's the probability that you get chocolate chips?

(e) Kunal realizes that the top few pancakes in the new stack don't really have chocolate chips in them. He shifts the top 10 pancakes from that stack (those with the smallest chance of chocolate chips) to the old stack. Given you randomly choose a pancake and it has chocolate chips, what's the probability it came from the new stack?

Answer:

Event A: A golden-brown pancake is on top of the stack.

Event B: A half-burnt pancake is on top of the stack.

Event C: A fully-burnt pancake is on top of the stack.

Event D: The top side of the stack is burnt.

Event E: The bottom side of the stack is burnt.

(a) $\Pr[A] = \Pr[B] = \Pr[C] = \frac{1}{3}$; and $\{A, B, C\}$ partition the sample space.

$\Pr[D|A] = 0, \Pr[D|B] = 0.5, \Pr[D|C] = 1$;

$\Pr[C|D] = \frac{\Pr[D|C]\Pr[C]}{\Pr[D]} = \frac{1\left(\frac{1}{3}\right)}{0\left(\frac{1}{3}\right) + 0.5\left(\frac{1}{3}\right) + 1\left(\frac{1}{3}\right)} = \frac{2}{3}$.

Alternate: 6 sides, 3 are burnt, 2 of which correspond to Pancake C.

(b)

$$\begin{aligned}
\Pr[E|D] &= \Pr[E \cap A|D] + \Pr[E \cap B|D] + \Pr[E \cap C|D] \\
&= \frac{3}{4} \times Pr[A|D] + \frac{2}{4} \times \Pr[B|D] + \frac{1}{4} \times \Pr[C|D] \\
&= \frac{3}{4}(0) + \frac{2}{4}\left(\frac{1}{3}\right) + \frac{1}{4}\left(\frac{2}{3}\right) \\
&= \frac{1}{3}.
\end{aligned}$$

(c) $\Pr[C|D] = \dfrac{\Pr[D|C]\Pr[C]}{\Pr[D]} = \dfrac{1\left(\frac{x}{n}\right)}{0\left(\frac{n-x-y}{n}\right) + 0.5\left(\frac{y}{n}\right) + 1\left(\frac{x}{n}\right)} = \dfrac{2x}{2x+y}$;

$\Pr[B|\overline{D}] = \dfrac{\Pr[\overline{D}|B]\Pr[B]}{\Pr[\overline{D}]} = \dfrac{0.5\left(\frac{y}{n}\right)}{1\left(\frac{n-x-y}{n}\right) + 0.5\left(\frac{y}{n}\right) + 0\left(\frac{x}{n}\right)} = \dfrac{y}{2n-2x-y}$.

(d) Event F: You picked from the new stack.

Event G: You got chocolate chips on your pancake.

$\Pr[F \cap G] = \sum_{k \in F} \Pr[\{k\} \cap G] = \sum_{k=1}^{m} \dfrac{1}{m+n} \cdot \dfrac{k}{m} = \dfrac{1}{m+n} \cdot \dfrac{m+1}{2}$;

$\Pr[\overline{F} \cap G] = 0$;

$\Pr[G] = \Pr[F \cap G] + \Pr[\overline{F} \cap G] = \dfrac{m+1}{2(m+n)}$.

Alternate:

$\Pr[F] = \dfrac{m}{m+n}$;

$\Pr[G|F] = \sum_{k=1}^{m} \dfrac{1}{m} \cdot \dfrac{k}{m} = \dfrac{1}{m+n} \cdot \dfrac{m+1}{2}$;

$\Pr[G|\overline{F}] = 0$;

$\Pr[G] = \Pr[G|F]\Pr[F] + \Pr[G|\overline{F}]\Pr[\overline{F}] = \dfrac{m+1}{2(m+n)}$.

(e) $\Pr[F \cap G] = \sum_{k \in F} \Pr[\{k\} \cap G] = \sum_{k=11}^{m} \dfrac{1}{m+n} \cdot \dfrac{k}{m} = \dfrac{1}{m+n} \cdot \dfrac{(m+11)(m-10)}{2m}$;

$\Pr[F|G] = \dfrac{\Pr[F \cap G]}{\Pr[G]} = \dfrac{(m+11)(m-10)}{m(m+1)} = \dfrac{m^2+m-110}{m^2+m}$.

(Below algebra not required, since $\Pr[G]$ does not change):

$\Pr[F \cap G] = \sum_{k \in F} \Pr[\{k\} \cap G] = \sum_{k=11}^{m} \dfrac{1}{m+n} \cdot \dfrac{k}{m}$;

$\Pr[\overline{F} \cap G] = \sum_{k \in \overline{F}} \Pr[\{k\} \cap G] = \sum_{k=1}^{10} \dfrac{1}{m+n} \cdot \dfrac{k}{m}$;

$\Pr[G] = \Pr[F \cap G] + \Pr[\overline{F} \cap G] = \sum_{k=1}^{m} \dfrac{1}{m+n} \cdot \dfrac{k}{m} = \dfrac{m+1}{2(m+n)}$.

Alternate:

$\Pr[F] = \dfrac{m-10}{m+n}$;

$\Pr[G|F] = \sum_{k=11}^{m} \dfrac{1}{m-10} \cdot \dfrac{k}{m} = \dfrac{m+11}{2m}$;

$\Pr[\overline{F}] = \dfrac{n+10}{m+n}$;

$\Pr[G|\overline{F}] = \sum_{k=1}^{10} \dfrac{1}{n+10} \cdot \dfrac{k}{m} = \dfrac{55}{m(n+10)}$;

$\Pr[F|G] = \frac{\Pr[F \cap G]}{\Pr[F \cap G] + \Pr[\overline{F} \cap G]} = \frac{\Pr[F]\Pr[G|F]}{\Pr[F]\Pr[G|F] + \Pr[\overline{F}]\Pr[G|\overline{F}]} = \dfrac{(m+11)(m-10)}{(m+11)(m-10)+110} = \dfrac{m^2+m-110}{m^2+m}$.

3. **Error Correction Codes with Probability** (17 points, 3/3/3/3/5 points for each part)

Alice wishes to send $n = 5$ packets to Bob. However, due to channel noise, there is a 10% chance ($p = 0.1$) for each packet to be corrupted (general error) during transmission. Therefore, Alice wants to transmit m additional packets so Bob can correct potential errors. Bob uses the Berlekamp-Welch Algorithm in the lecture note to fix the errors.

(a) If $m = 2$, what is the probability that Bob gets Alice's message successfully?

(b) If $m = 3$, what is the probability that Bob gets Alice's message successfully?

(c) By the above results, is adding more packets always helpful in this case?

(d) To guarantee that the probability that Bob gets Alice's message successfully is at least 0.9, how many additional packets does Alice need to transmit?

Now, we are considering a general case where $n \geq 0$, $m \geq 0$, and $0 \leq p \leq 1$, where p is the probability for each packet to be corrupted.

(e) Prove it or provide a counterexample: For any n, m, p, transmitting $m+2$ additional packets is always at least as good as (never worse than) transmitting m additional packets, in terms of the probability that Bob gets Alice's message successfully.

Answer:

(a) When $n = 5$ and $m = 2$, Bob can correct the message if there is at most $\lfloor \frac{2}{2} \rfloor = 1$ corrupted packet. We have

$$\begin{aligned} \Pr[\text{no packet is corrupted}] &= \binom{7}{0} 0.9^7, \\ \Pr[\text{one packet is corrupted}] &= \binom{7}{1} 0.1^1 0.9^6, \end{aligned}$$

so the probability that Bob gets Alice's message successfully is $\binom{7}{0}0.9^7 + \binom{7}{1}0.1^1 0.9^6 \approx 0.850$.

(b) When $n = 5$ and $m = 3$, Bob can correct the message if there is at most $\lfloor \frac{3}{2} \rfloor = 1$ corrupted packet. We have

$$\begin{aligned} \Pr[\text{no packet is corrupted}] &= \binom{8}{0} 0.9^8, \\ \Pr[\text{one packet is corrupted}] &= \binom{8}{1} 0.1^1 0.9^7, \end{aligned}$$

so the probability that Bob gets Alice's message successfully is $\binom{8}{0}0.9^8 + \binom{8}{1}0.1^1 0.9^7 \approx 0.813$.

(c) No. We can see $0.813 < 0.850$, meaning that adding 3 packets is worse than adding 2 packets.

(d) We can try $m = 4$ first. When $n = 5$ and $m = 4$, Bob can correct the message if there is at most $\lfloor \frac{4}{2} \rfloor = 2$ corrupted packet. We have

$$\begin{aligned} \Pr[\text{no packet is corrupted}] &= \binom{9}{0} 0.9^9, \\ \Pr[\text{one packet is corrupted}] &= \binom{9}{1} 0.1^1 0.9^8, \\ \Pr[\text{two packets are corrupted}] &= \binom{9}{2} 0.1^2 0.9^7, \end{aligned}$$

so the probability that Bob gets Alice's message successfully is $\binom{9}{0}0.9^9 + \binom{9}{1}0.1^1 0.9^8 + \binom{9}{2}0.1^2 0.9^7 \approx 0.947 > 0.9$. Therefore, Alice needs to transmit at least 4 additional packets.

(e) False. Suppose $n = 1$, $m = 0$, and $p = 0.9$. The probability that Bob gets Alice's message successfully with 2 additional packets is $\binom{3}{0}0.1^3 + \binom{3}{1}0.9^1 0.1^2 = 0.028$; the probability that Bob gets Alice's message successfully with 0 additional packets is $\binom{1}{0}0.1 = 0.1$. We can see that $0.028 < 0.1$, *i.e.*, transmitting 2 additional packets is worse than transmitting 0 additional packet.

4. **Phase Two** (12 points, 3 points for each part)

   Your Tele-BEARS appointment is tomorrow, but unfortunately one of the classes you need and all its sections and labs are full. The first class has four sections. To enroll in the first class, you must enroll in exactly one section.

   (a) For the first class, Tele-BEARS (miraculously) lets you waitlist all four sections simultaneously (if you please). Assume the probability that you will be able to get off the waitlist for that given section is $1/(n+2)$, where n is the number of people ahead of you on the waitlist. During your appointment, you notice Section 1 has a waitlist of 3 people, Section 2's waitlist has 7 people, Section 3's has 8 people, and Section 4's has 5 people. What is the expected number of sections you will get into if you waitlist all four?

   (b) Suppose Section 1 is at 8am, so you decide not to waitlist it. What is the new expected number of sections you will get into if you waitlist all but Section 1?

   (c) Following Part (a), the professor surprises everyone by deciding to add labs to the course which immediately fill up before your Tele-BEARS appointment. He adds two three hour labs — Lab 1 and Lab 2. Lab 1 has a time conflict with Section 1, and Lab 2 has a time conflict with Section 3. The probability of getting into a lab with n people ahead of you on the waistlist is $1/(n+1)$ and is independent of the probability of getting into any section. There are 10 people on the waitlist for Lab 1 and 12 people on the waitlist for Lab 2 at the time of your Tele-BEARS appointment. Unfortunately, you can only waitlist one of the labs at a time, and to enroll or waitlist any section you MUST be enrolled in or on the waitlist of a lab first, and there must not be any time conflicts. What lab should you waitlist to maximize the expected number of sections you will get into? You MUST get in!!! Justify your answer.

   (d) For your second class, suppose there are k different sections and students are only allowed to waitlist one of them. All k sections are full, and there are m students who want to waitlist the course. If students waitlist for a random section, what is the expected number of sections which will have students on their waitlists?

   Answer:

   (a) Let random variable $S_i = 1$ if you get into section i, else 0 and random variable T bet the total number of sections you get into.
   $T = S_1 + S_2 + S_3 + S_4$;
   $\mathbb{E}(S_i) = 1 \cdot \Pr[S_i = 1]$;
   $\mathbb{E}(S_1) = 1/(3+2)$;
   $\mathbb{E}(S_2) = 1/(7+2)$;
   $\mathbb{E}(S_3) = 1/(8+2)$;
   $\mathbb{E}(S_4) = 1/(5+2)$;
   $\mathbb{E}(T) = \mathbb{E}(S_1) + \mathbb{E}(S_2) + \mathbb{E}(S_3) + \mathbb{E}(S_4) = 1/5 + 1/9 + 1/10 + 1/7$.

   (b) $T = S_2 + S_3 + S_4$;
   $\mathbb{E}(T) = \mathbb{E}(S_2) + \mathbb{E}(S_3) + \mathbb{E}(S_4) = 1/9 + 1/10 + 1/7$.

   (c) Let random variable $X_{i,j} = 1$ if you get into lab i and section j, else 0. Let X_{labi} be the number of sections you get into after choosing Lab i.

$$
\begin{aligned}
\mathbb{E}(X_{\text{labi}}) &= \sum_{j\neq i}\mathbb{E}(X_{i,j});\\
\mathbb{E}(X_{i,j}) &= 1\cdot\Pr[X_{i,j}=1];\\
\Pr[X_{1,2}=1] &= (1/(7+2))(1/(10+1));\\
\Pr[X_{1,3}=1] &= (1/(8+2))(1/(10+1));\\
\Pr[X_{1,4}=1] &= (1/(5+2))(1/(10+1));\\
\mathbb{E}(X_{\text{lab1}}) &= (1/(7+2))(1/(10+1))+(1/(8+2))(1/(10+1))+(1/(5+2))(1/(10+1))\\
&\approx 0.032;\\
\Pr[X_{2,1}=1] &= (1/(3+2))(1/(12+1));\\
\Pr[X_{2,2}=1] &= (1/(7+2))(1/(12+1));\\
\Pr[X_{2,4}=1] &= (1/(5+2))(1/(12+1));\\
\mathbb{E}(X_{\text{lab2}}) &= (1/(3+2))(1/(12+1))+(1/(7+2))(1/(12+1))+(1/(5+2))(1/(12+1))\\
&\approx 0.035.
\end{aligned}
$$

$\mathbb{E}(X_{\text{lab2}})$ is slightly larger.

(d) This is kind of like the balls and bins from lecture notes. Let random variable Y_i be 1 if a section has a waitlist, else 0 and Y be the total number of sections with waitlists.
$Y=\sum_{i=0}^{n} Y_i$;
$\mathbb{E}(Y)=\sum_{i=0}^{n}\mathbb{E}(Y_i)$;
$\mathbb{E}(Y_i)=\Pr[Y_i=1]=1-\Pr[\text{section } i \text{ is empty}]=1-(\frac{k-1}{k})^m$;
$\mathbb{E}(Y)=k\cdot\left(1-(\frac{k-1}{k})^m\right)$.

5. **Random Variable and Expectation** (16 points, 3/3/10 points for each part)

(a) Given $\Pr[X=1]=\dfrac{1}{5},\Pr[X=2]=\dfrac{4}{5}$, what is $\mathbb{E}(X)$?

(b) Given $\Pr[X=1]=\dfrac{1}{14},\Pr[X=2]=\dfrac{4}{14},\Pr[X=3]=\dfrac{9}{14}$, what is $\mathbb{E}(X)$?

(c) Assume n is a positive integer. A random variable X has n possible values $\{1,2,\dots,n\}$, and

$$\frac{\Pr[X=i]}{\Pr[X=1]}=i^2$$

for any $i\in\{1,2,\dots,n\}$. Prove that $\mathbb{E}(X)=\dfrac{3n(n+1)}{2(2n+1)}$.

Answer:

(a) $\mathbb{E}(X)=1\cdot\Pr[X=1]+2\cdot\Pr[X=2]=1\cdot\frac{1}{5}+2\cdot\frac{4}{5}=\frac{9}{5}$.

(b) $\mathbb{E}(X)=1\cdot\Pr[X=1]+2\cdot\Pr[X=2]+3\cdot\Pr[X=3]=1\cdot\frac{1}{14}+2\cdot\frac{4}{14}+3\cdot\frac{9}{14}=\frac{36}{14}=\frac{18}{7}$.

(c) We want to show that

$$\mathbb{E}(x)=\sum_{i=1}^{n} i\cdot\Pr[X=i]=\frac{3n(n+1)}{2(2n+1)}.$$

We know that

$$\Pr[X=i] = i^2 \Pr[X=1],$$

so we can write it as

$$\mathbb{E}(x) = \sum_{i=1}^{n} i^3 \Pr[X=1].$$

Another piece of information that we need to prove this is that the summation of the probabilities is equal to one.

$$\sum_{i=1}^{n} \Pr[X=i] = 1;$$

$$\sum_{i=1}^{n} i^2 \Pr[X=1] = 1;$$

$$\Pr[X=1] \sum_{i=1}^{n} i^2 = 1;$$

$$\Pr[X=1] = \frac{1}{\sum_{i=1}^{n} i^2}.$$

Plug it in to the expectation equation to get

$$\mathbb{E}(X) = \frac{\sum_{i=1}^{n} i^3}{\sum_{i=1}^{n} i^2}.$$

Showing the following equalities are true is a simple matter of induction (check the theorems below):

$$\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6};$$

$$\sum_{i=1}^{n} i^3 = \frac{n^2(n+1)^2}{4}.$$

Plug them in and you get

$$\mathbb{E}(X) = \frac{6n^2(n+1)^2}{4n(n+1)(2n+1)} = \frac{3n(n+1)}{2(2n+1)}.$$

Theorem 1: $\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}$.
Proof:

- **Base Case** $(n=0)$: $\sum_{i=0}^{0} i^2 = 0 = \frac{0(0+1)(2\cdot 0+1)}{6}$.
- **Inductive Hypothesis**: assume, when $n=k$, $\sum_{i=0}^{k} i^2 = \frac{k(k+1)(2k+1)}{6}$.
- **Inductive Step**: need to prove that $\sum_{i=0}^{k+1} i^2 = \frac{(k+1)(k+2)(2k+3)}{6}$:

$$\begin{aligned}
\sum_{i=0}^{k+1} i^2 &= \left(\sum_{i=0}^{k} i^2\right) + (k+1)^2 \\
&= \frac{k(k+1)(2k+1)}{6} + (k+1)^2 \\
&= \frac{(k+1)(2k^2+k) + 6(k+1)^2}{6} \\
&= \frac{(k+1)(2k^2+7k+6)}{6} \\
&= \frac{(k+1)(k+2)(2k+3)}{6}.
\end{aligned}$$

By the principle of induction, the claim follows. $\square$

Theorem 2: $\sum_{i=1}^{n} i^3 = \frac{n^2(n+1)^2}{4}$.

Proof:

- **Base Case** $(n = 0)$: $\sum_{i=0}^{0} i^3 = 0 = \frac{0^2(0+1)^2}{4}$.
- **Inductive Hypothesis**: assume, when $n = k$, $\sum_{i=0}^{k} i^3 = \frac{k^2(k+1)^2}{4}$.
- **Inductive Step**: need to prove that $\sum_{i=0}^{k+1} i^3 = \frac{(k+1)^2(k+2)^2}{4}$:

$$\begin{aligned}
\sum_{i=0}^{k+1} i^3 &= \left(\sum_{i=0}^{k} i^3\right) + (k+1)^3 \\
&= \frac{k^2(k+1)^2}{4} + (k+1)^3 \\
&= \frac{k^2(k+1)^2 + 4(k+1)^3}{4} \\
&= \frac{(k+1)^2(k^2+4k+4)}{4} \\
&= \frac{(k+1)^2(k+2)^2}{4}.
\end{aligned}$$

By the principle of induction, the claim follows. $\square$