CSM Fall 2018 Interview Materials (For interviewers)

Nick Titterton, Max Ovsiankin, Anindit Go, Avni Singhal

Format

The format of the interview will be as follows:

- **Part 1 (4 min):** Candidates will all be provided with 5 relatively difficult questions beforehand. They will also have access to very bare bones solutions to these questions. They will pick one question and present their solution to it (does not have to follow the same solution that we provide).
- **Part 2 (4 min):** We will then present a problem and a made-up student's attempt at solving the problem. The made up student will have made a common mistake or misconception while solving the problem, and the candidate will be expected to find the mistake and explain the correct solution to the question. These questions will not be difficult and will be randomly selected from a set of 3. Candidates will not be aware of these questions beforehand.
- **Part 3: (Remaining time):** With whatever remaining time there is, we will ask some soft questions to the candidates. We have to try to design these questions such that all candidates do not provide the same stock answers.

Prepared Question/Solutions:

Use a combinatorial proof to argue

$$\sum_{i=0}^{n} i^2 = \binom{n+1}{2} + 2\binom{n+1}{3}$$

Solution 1:

This is the number of ways to choose triplets (a, b, c) such that $c \geq a, b$, for every c from 0 to n. $\binom{n+1}{2}$ covers the case when $a = b$, and $2\binom{n+1}{3}$ covers the case when $a \neq b$.

Good:

- Explains why it's necessary to split the right hand side up into 2 cases
- Tries to provide some intuition for how to approach similar problems

Bad:

- Does algebraic proof (pulls out factorial)
- Logical gaps in their explanation. Push them to fill in the gaps.

Question 2:

You draw randomly from a hat that has 100 slips of paper in it, labeled 1-100. You can either chose to keep the dollar amount equal to the number you drew, or you can pay 1 dollar, place the number back in the hat, and draw again. Determine the optimal strategy to maximize your expected value from playing the game. Hint: Think about how we calculated expected value for a geometric series.

Solution 2:

Note: In your presented solutions, you have to show how you would go about maximizing something, but you don't have to actually calculate derivatives/solve for roots.

Let P be your profit from the game. The optimal strategy is to have some threshold X, such that when the spin is greater than or equal to X you take the money and when it is less X chose to re-draw. We see that:

$$\mathbf{E}[P] = \frac{X-1}{100}(\mathbf{E}[P]-1) + (1 - \frac{X-1}{100}) \cdot \frac{X+100}{2}$$

$\frac{X-1}{100}$ is the probability that we decide to play the game again. In this case, we expect to get the same expected value, minus 1 for the dollar we pay to play again. If we do not play the game, then the expected value of the money we get is the average of X and 100. Maximizing the above with respect to X, we find the optimal value for $X = 86.85$. The game only uses discrete values, so using $X = 87$ gives us an expected value of 86.37. Alternatively,

$$\mathbf{E}[P] = \frac{100+X}{2} - \frac{100}{101-X}.$$

Maximizing this with respect to X gives us the same results as the above.

Good:

- Explain why the optimal strategy stays the same from round to round
- Intuition about using recursive definition of Expectation from the fact that game repeats itself after each round.
- Explains the logic behind one of the 2 equations provided in the solution (or some other valid equation)
- Explains how to maximize the expression/with respect to what variable

Bad:

- Logical gaps in their explanation. Push them to fill in the gaps.

Follow-Up Questions (Optional):

- If student did not address any of the items in the Good category, ask them to explain.

Question 3:

You are practicing your max range shot put. Each day for 30 consecutive days, you throw the shot as far as you can. The shotput lands distance D away from you, where $D \sim Exp(\lambda)$ yards. Define a "Personal Record" as a throw that is farther than all the throws on the days before it. Let P be the number of days where you achieve a personal record. Calculate $\mathbf{E}[P]$.

Solution 3:

Let $P_i = 1$ if you set a PR on day i, and 0 otherwise.

$$\begin{aligned}\mathbf{E}[P] &= \sum_{i=1}^{n} \mathbf{E}[P_i] \\ &= \sum_{i=1}^{n} \frac{1}{i} \\ &\approx \ln 30 + \gamma\end{aligned}$$

Good:

- Explain why it makes sense to use indicator variables.
- Explains why the expectation of the indicator variable is $\frac{1}{i}$. Mentions the fact that since all independent and drawn from the same distribution, all orderings are equally likely.
- Explains the approximation of the sum as a logarithm plus a constant

Bad:

- Logical gaps in their explanation. Push them to fill in the gaps.

Follow-Up Questions (Optional):

- If student did not address any of the items in the Good category, ask them to explain.
- If the distribution was another continuous distribution (not exponential), does the answer change?

Question 4:

You and your friend each spin a spinner which returns a uniform, continuous random value between 0 and 1. Your team score S is calculated by taking the max of you and your friend's results and squaring it. For example, if you spun 0.5 and your friend spun 0.1, then your score S would equal $0.5^2 = 0.25$. Calculate the pdf of S.

Solution 4: Let your score be denoted by $X \sim U[0,1]$ and your friend's denoted by $Y \sim U[0,1]$. Let $T = \max(X, Y)$. $S = T^2$

$$P(T < t) = P(X < x \cap Y < y) = t^2 \text{ so } f_t(t) = 2t$$
$$\text{Then } P(S < s) = P(T < \sqrt{s}) = s, \text{ so}$$
$$f_s(s) = 1.$$

Good:

- Explains the relationship between pdf and cdf (that pdf is derivative of cdf).
- Calculates the pdf or cdf of the max of 2 uniform random variables, using the fact that X and Y are independent.
- Calculates the pdf or cdf (directly or indirectly) of the square of the max.
- EXTREME BONUS POINTS: (please let us know if they do): provides intuition for the result being a uniform distribution.

Bad:

- Logical gaps in their explanation. Push them to fill in the gaps.

Common Misconceptions

Misconception 1:

"For all graphs with $n \geq 2$ vertices, if each vertex has positive degree then it is connected."

Proof:
Base case: For a graph with only 2 vertices, if they each have degree greater than then they must be connected.
Inductive Hypothesis: Assume all graphs with n vertices that each have positive degree are connected.
Inductive Step: Consider a graph with n vertices. It must be connected by the inductive hypothesis. Add a vertex to this graph. If this new vertex has a positive degree, it must be connected to some other vertex in the original n vertex graph, which means that the entire graph must be connected. If it doesn't have positive degree, then it isn't connected to anything, so the graph is disconnected.

Correcting 1:

This is due to build up error: You cannot build all possible $n+1$ vertex graphs with all verticies of degree greater than 1, by simply adding a vertex to an n degree graph with the same property. What you have to do is start with the $n+1$ degree graph, remove a vertex, check that the property still holds, then add back a vertex. In this case, say we start with a $n+1$ degree graph with all vertices with degree greater than 0. After we remove a vertex, we cannot guarantee that the resulting n degree graph has positive degree on each of its vertices, so we cannot apply our inductive hypothesis. This is known as build up error.

Misconception 2:

Question: A random variable Z takes on values between $[-10, 0]$ and has expected value -5. Use Markov's bound to bound the probability that Z is greater than or equal to -1.
Proposed Solution:

$$P(Z \geq -1) \leq \frac{\mathbf{E}[Z]}{-1} = \frac{-5}{-1} = 5$$

Correcting 2:

Markov's Bound can only be used on variables that are strictly non-negative, which Z is not. If we wanted to use Markov's bound in this case, we would have to transform Z to be non negative. Let $Z' = Z + 10$. Then

$$P(Z \geq -1) = P(Z' \geq 9) \leq \frac{\mathbf{E}[Z']}{9} = \frac{5}{9}$$

Misconception 3:

Andy spins a spinner that returns a uniform, continuous random value between -1 and 1. He then squares his result to calculate his score. Ben spins 2 spinners that return uniform, continuous random values between -1 and 1. He multiplies the 2 results together to calculate his score. Carol determines that both players are multiplying 2 random, uniform values on the range -1 to 1 together to calculate the score, so she determines that it is a fair game and that neither player should be given a handicap.

Correcting 3:

Let $X \sim U[-1, 1]$ represent Andy's spin, and X^2 represent his score. Let $Y, Z \sim U[-1, 1]$ represent Ben's spins, and let YZ represent his score. Since Y and Z are independent, $E[YZ] = E[Y]E[Z] = 0$. However, $E[X^2] \neq E[X]E[X]$, because X is not independent from X. Instead we calculate $E[X^2]$ by doing

$$\int_{-1}^{1} \frac{1}{2}\frac{1}{3}x^2 dx = \frac{x^3}{6}\Big|_{-1}^{1} = \frac{1}{3}$$

which is quite different than Ben's expected value.