Algorithms M2–IF TD 2

October 6, 2021

1 A crowded theater

A theater has n seats, numbered $1, \ldots, n$. Tonight's show is sold out, so there are n spectators who want to enter. Every spectator has a ticket with the number of her seat. The spectators enter the theater in a random order. We assume that the following happens:

- The first person to enter ignores the number on his ticket and picks uniformly at random a seat to sit in. (What an idiot!)
- Every other spectator is nice, but a little shy. They all do the following: first, they go to their assigned seat and if it is free, they sit there. If not, they pick a seat uniformly at random and sit there.

Calculate the following:

1. The probability that the first spectator (the idiot) sits in someone else's seat.
2. The probability that the second spectator (the one who enters after the idiot) sits in his own seat.
3. The probability that the second spectator (the one who enters after the idiot) sits in the idiot's seat.
4. The probability that the last spectator sits in his seat. (If it helps, start with small values of n first).
5. Consider now a variation where the spectators go in in a completely random order (that is, the idiot is not necessarily the first to enter). Calculate the probability that the last spectator sits in his seat in this case.

Solution:

1. $\frac{n-1}{n}$
2. Also, $\frac{n-1}{n}$.

3. Let A_2 be the event that the idiot sits in the second spectator's seat. Let B be the event that the second spectator sits in the idiot's seat. Then

$$Pr[B] = Pr[B \mid A_2]Pr[A_2] + Pr[B \mid \neg A_2]Pr[\neg A_2]$$

But, $Pr[B \mid \neg A_2] = 0$, because if $\neg A_2$, then the second spectator seats in his own seat. Furthermore, $Pr[A_2] = \frac{1}{n}$ and $Pr[B \mid A_2] = \frac{1}{n-1}$ so we have $Pr[B] = \frac{1}{n(n-1)}$.

4. Let A_i be the event that the first spectator (the idiot) took the seat of spectator i. Let C_n be the event that the last spectator sits in the correct seat, assuming that the theater has n seats. We have

$$Pr[C_n] = \sum_{i=1}^{n} Pr[C_n \mid A_i]Pr[A_i] = \frac{1}{n}\sum_{i=1}^{n} Pr[C_n \mid A_i]$$

We now observe that $Pr[C_n \mid A_1] = 1$, because in this case everyone takes the right seat. Also $Pr[C_n \mid A_n] = 0$, because in this case the first spectator took the seat of the last spectator.

The main insight now is that $Pr[C_n \mid A_i] = Pr[C_{n-i+1}]$ for all $2 \leq i \leq n-1$. To see this, suppose that A_i is true. Then, spectators $2, \ldots, i-1$ take their assigned seats. We now have $n-i+1$ remaining spectators such that: the first among them will take a random seat, and everyone else will act as normal. This is the same experiment, but with $n-i+1$ seats remaining.

We now use induction. First, it is not hard to show that $Pr[C_2] = Pr[C_3] = \frac{1}{2}$ by analyzing all cases. Then we have

$$Pr[C_n] = \frac{1}{n} + \frac{1}{n}\sum_{i=2}^{n-1} Pr[C_n \mid A_i] = \frac{1}{n} + \frac{1}{n}\sum_{i=2}^{n-1} Pr[C_{n-i+1}] = \frac{1}{n} + \frac{n-2}{2n} = \frac{1}{2}$$

where we have assumed by the inductive hypothesis that $Pr[C_{n-i+1}] = \frac{1}{2}$, because $2 \geq n-i+1 \leq n-1$ for the values of i in the sum.

5. As shown in the previous question, the probability is still essentially $\frac{1}{2}$: if the idiot is the i-th spectator, the first $i-1$ spectators take their assigned seats, so we run the same experiment in the remaining seats. There is, however, one complication: if the idiot is last, then he will definitely take his proper seat. Hence, the probability is $\frac{n-1}{2n} + \frac{1}{n} = \frac{n+1}{2n}$.

2 Another crowded theater

This is similar to the previous exercise, except now all spectators behave like the first spectator: when they enter, they pick uniformly at random a random seat, and sit there (ignoring their assigned number).

1. What is the probability that the i-th spectator sits in her assigned place?

2. What is the expected number of spectators sitting in their assigned places?

Solution:

1. Let F_i be the event that the seat of the i-th spectator is still free when she enters the room, C_i be the event that the i-th spectator takes the right seat. We have $Pr[C_i] = Pr[C_i \mid F_i]Pr[F_i] + Pr[C_i \mid \neg F_i]Pr[\neg F_i]$. The second term is clearly 0. Since the i-th spectator picks a seat uniformly at random, we have $Pr[C_i \mid F_i] = \frac{1}{n-(i-1)}$, because $i-1$ seats are taken when she enters. So what is left is to calculate $Pr[F_i]$.

   Let $A_1, A_2, \ldots, A_{i-1}$ be the event that the seat of the i-th spectator is still free after the first, second, $\ldots$, $(i-1)$-th spectators take their seats. We have $Pr[F_i] = Pr[A_{i-1}]$, $Pr[A_1] = \frac{n-1}{n}$, and for all $j \in \{2, \ldots, i-1\}$ we have $Pr[A_j] = Pr[A_j \mid A_{j-1}]Pr[A_{j-1}] + Pr[A_j \mid \neg A_{j-1}]Pr[\neg A_{j-1}]$. The second term is 0. From the first term we get $Pr[A_j] = \frac{n-j}{n-j+1}Pr[A_{j-1}]$. We therefore get,

$$Pr[F_i] = Pr[A_{i-1}] = \frac{n-i+1}{n-i+2}\frac{n-i+2}{n-i+3}\cdots\frac{n-1}{n} = \frac{n-i+1}{n}$$

   We get $Pr[C_i] = \frac{1}{n}$ for all i.

2. Let X_i be a random variable that is 1 if C_i is true and 0 otherwise. Then the number of spectators who sit in their seats is $\sum_{i=1}^{n} X_i$. The expectation of this variable is

$$E[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} Pr[C_i] = 1$$

   So, we expect 1 person on average to sit in the right place (independent of the size of the theater!)

3 Dice and Expectations

[MU Ex 2.1] Suppose we roll a fair k-sided die with the numbers $1, \ldots, k$ on its faces. Let X be the number that appears. What is $E[X]$?

[MU Ex 2.9] Suppose we roll a fair k-sided die **twice**. Let X_1, X_2 be the two values we obtained. What is $E[\min\{X_1, X_2\}]$? What is $E[\max\{X_1, X_2\}]$. Calculate also $E[\min\{X_1, X_2\}] + E[\max\{X_1, X_2\}]$.

Solution:

For the first question we have $E[X] = \sum_{i=1}^{k} iPr[X = i] = \frac{1}{k}\sum_{i=1}^{k} i = \frac{k(k+1)}{2k} = \frac{k+1}{2}$.

For the second question, $E[\min\{X_1, X_2\}] = \sum_{i=1}^{k} iPr[\min\{X_1, X_2\} = i]$. Let us therefore calculate $Pr[\min\{X_1, X_2\} = i]$. We have

$$\begin{aligned} Pr[\min\{X_1, X_2\} = i] &= Pr[(X_1 = i) \wedge (X_2 > i)] + Pr[(X_2 = i) \wedge (X_1 > i)] + Pr[(X_1 = i) \wedge (X_2 = i)] = \\ &= Pr[X_1 = i]Pr[X_2 > i] + Pr[X_2 = i]Pr[X_1 > i] + Pr[X_1 = i]Pr[X_2 = i] = \\ &= \frac{k-i}{k^2} + \frac{k-i}{k^2} + \frac{1}{k^2} = \frac{2k-2i+1}{k^2} \end{aligned}$$

We now have

$$\begin{aligned} E[\min\{X_1, X_2\}] &= \sum_{i=1}^{k} iPr[\min\{X_1, X_2\} = i] = \\ &= \sum_{i=1}^{k} \frac{i(2k-2i+1)}{k^2} = \\ &= (\frac{2}{k}\sum_{i=1}^{k} i) - (\frac{2}{k^2}\sum_{i=1}^{k} i^2) + \frac{1}{k^2}\sum_{i=1}^{k} i \end{aligned}$$

Using the fact that $\sum_{i=1}^{k} i = \frac{k(k+1)}{2}$ and $\sum_{i=1}^{k} i^2 = \frac{k(k+1)(2k+1)}{6}$ we get $E[\min\{X_1, X_2\}] = \frac{(k+1)(2k+1)}{6k}$.

Instead of calculating $E[\max\{X_1, X_2\}]$ in the same way, we observe that for all X_1, X_2 we have $\min\{X_1, X_2\} + \max\{X_1, X_2\} = X_1 + X_2$. Therefore, $E[\max\{X_1, X_2\}] = E[X_1] + E[X_2] - E[\min\{X_1, X_2\}]$.

4 Fair coins again

We flip a fair coin $2n$ times. Let X_1 be the number of times the result was heads, and X_2 the number of times the result was tails. Prove that for any $\epsilon > 0$ there exists a c such that we have $Pr[X_1 - X_2 > c\sqrt{n}] < \epsilon$.

Solution:

Let $Y = X_1 - X_2$. Then, $E[Y] = 0$, because $E[X_1] = E[X_2]$ (the coin is fair). $Var[Y] = E[Y^2] - E[Y]^2 = E[Y^2] = E[X_1^2] + E[X_2^2] - 2E[X_1X_2]$.

Now we observe that $E[X_1^2] = E[X_2^2]$ (because the coin is fair). Also $E[X_1] = n$ and $X_1 + X_2 = 2n$. We therefore have

$$\begin{aligned} Var[Y] &= 2E[X_1^2] - 2E[X_1(2n - X_1)] = \\ &= 4E[X_1^2] - 4nE[X_1] = \\ &= 4E[X_1^2] - 4(E[X_1])^2 = 4Var[X_1] \end{aligned}$$

Let us now calculate $Var[X_1]$. Let H_i, $i \in \{1, \ldots, 2n\}$ be a random variable that takes value 1 if the i-th flip was heads and 0 otherwise. Then $X_1 = \sum_{i=1}^{2n} H_i$. Also the variables H_i are independent, so $Var[X_1] = \sum_{i=1}^{2n} Var[H_i]$. But $Var[H_i] = E[H_i^2] - E[H_i]^2 = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}$. So, $Var[X_1] = \frac{n}{2}$, which gives $Var[Y] = 2n$.

We now use Chebyshev's inequality:

$$Pr[X_1 - X_2 > c\sqrt{n}] = Pr[Y - E[Y] > c\sqrt{n}] \leq Pr[|Y - E[Y]| > c\sqrt{n}] \leq \frac{Var[Y]}{c^2 n} = \frac{2}{c^2}$$

Therefore, setting $c = \sqrt{2/\epsilon}$ gives the result.

5 Secretary Problem

We are interviewing candidates for a job. Suppose there are n candidates overall. If we had perfect information, we could assign each candidate a score from $\{1, \ldots, n\}$ such that all candidates have distinct scores and the best candidate has score n. (In other words, if we could interview everyone before deciding, we could produce a ranking of the candidates).

The problem is that we see candidates one by one, in a random order, and after seeing a candidate we have to immediately decide if this candidate is hired. If not, this candidate leaves (and gets a job somewhere else).

We want to adopt a strategy that maximizes the probability of hiring the best candidate. Suppose we adopt the following strategy: we interview m candidates just to get a feeling of their level, but we don't hire anyone among them; then we hire the first candidate who is better than *all* the m candidates in our initial sample.

- Let E be the event that we hire the best candidate. Show that $Pr[E] = \frac{m}{n} \sum_{j=m+1}^{n} \frac{1}{j-1}$.
- Using the approximation $\sum_{i=1}^{n} \frac{1}{n} \approx \ln n$ show that

$$\frac{m}{n}(\ln n - \ln m) \leq Pr[E] \leq \frac{m}{n}(\ln(n-1) - \ln(m-1))$$

- Show that $Pr[E]$ is maximized when $m = n/e$. How much is $Pr[E]$ for this value?

Solution:

For the first question, let E_i be the event that the i-th candidate to be interviewed is in fact the best (this candidate may or may not be interviewed in the end, here we are talking about her order in the interview list). Then $Pr[E_i] = \frac{1}{n}$, since we consider all candidates with random order.

Now we have, $Pr[E] = \sum_{i=1}^{n} Pr[E \mid E_i] \cdot Pr[E_i] = \frac{1}{n} \sum_{i=1}^{n} Pr[E \mid E_i]$.

Observe that if $i \leq m$ then $Pr[E \mid E_i] = 0$, because in this case we surely reject the best candidate. We therefore have $Pr[E] = \frac{1}{n} \sum_{i=m+1}^{n} Pr[E \mid E_i]$.

If $i > m$, then (assuming E_i), the candidate in position i will be hired if and only if we do not hire anyone among the $(m+1)$-th and the $(i-1)$-th candidates, because if we reach candidate i, she is the best (therefore she is better than the first m), so she is hired. Therefore, it must be the case that among the first $i-1$ candidates, the best is one the first m candidates of the sample. This happens with probability $\frac{m}{i-1}$. We therefore have: $Pr[E] = \frac{m}{n} \sum_{i=1}^{n} \frac{1}{i-1}$.

For the second question we have $\sum_{i=m+1}^{n} \frac{1}{i-1} = \sum_{i=m}^{n-1} \frac{1}{i} = \sum_{i=1}^{n-1} \frac{1}{i} - \sum_{i=1}^{m-1} \frac{1}{i}$

For the last question we approximate $Pr[E] \approx \frac{m}{n}(\ln n - \ln m)$. Suppose $m = \alpha n$. We would like to find the value of α for which $Pr[E]$ is maximized. We have $Pr[E] = -\alpha \ln \alpha$. By taking the derivative (which is $-\ln \alpha - 1$) we see that this function has a maximum at $\alpha = \frac{1}{e}$.