

The Tinder Problem

This lesson discusses a sample problem and works out the complexity using probabilistic analysis.



Meet Neha. She's a millennial and an avid Tinder user. She swipes Tinder daily and dumps her latest boyfriend if she finds a better match than her current boyfriend (don't judge). In her world, there's never a tie between two potential boyfriends (i.e. she can always decide that one is better than the other). However, dating isn't easy or cheap. When Neha finds a mutual match, she'll go out on a date with the new guy and pay for it herself - which usually costs her about \$30. If she finds she prefers the new match she'll immediately break-up with her current boyfriend over text while also sending him a \$100 gift card at the same time to smooth out any grudges.

There are two costs in Neha's approach: one when she goes out on a date, and one when she breaks-up with someone. Let's see what the costs look like in the worst and best cases over ten consecutive mutual matches.

In the best case, she finds the best boyfriend on the first match and only incurs the cost of dating the remaining 9 matches without breaking up with

incurs the cost of dating the remaining 9 matches without breaking up with her first boyfriend. This brings her total to $\$30 * 10 = \300 .

In the worst case, each new match that she dates turns out to be better than her current boyfriend. Over the course of 10 matches, she will have broken up with 9 guys and dated all 10 of them. Her cost, in the worst case, comes out to be $\$300 + \$900 = \$1200$! This is much worse than her best case cost, or four-fold of that former amount.

It is unlikely that she'll hit the best or the worst case, although she can. The question is, what would be her average or expected cost since the matches she receives are guaranteed to be in a random order?



Solution

We can rank each of the 10 matches using an integer from 1 to 10. A match represented by a higher integer is considered to be better than the matches represented by the lower integers. We can simply use an array of 10 integers, representing the 10 matches for Neha, as input to the algorithm that outputs her final cost. Note that the array can have $10!$ permutations, and each permutation can represent a particular order in which she encounters her 10 matches and dates them. Below is the algorithm followed by the expected cost analysis.

Let's start with calculating the probability that she prefers the 5th match more than the previous four? It's simply $\frac{1}{5}$. Generalizing, when she dates the k th

than the previous four! It's simply $\frac{1}{5}$. Generalizing, when she dates the k th match, the probability is $\frac{1}{k}$ that she'll prefer the k th match more than the previous $k-1$ matches and breaks-up to be with the k th match. We can now model the possibility that Neha prefers the k th match more than the previous $k-1$ matches as an indicator random variable:

$I_k = 1$ if k th match is best so far or,
= 0 if a better match existed in previous $k-1$ dates

If we want to know how many times is Neha *expected* to break-up, It's simply a summation of all the indicator variables for each of k matches.

$$X = X_1 + X_2 + X_3 \cdots X_k$$

The above equation is essentially counting the number of times Neha starts a relationship when she goes through k tinder matches. We can now apply expectation on both sides of the equation:

$$\begin{aligned} E[X] &= E[X_1 + X_2 + X_3 \cdots X_k] \\ E[X] &= E[X_1] + E[X_2] + E[X_3] \cdots E[X_k] \end{aligned}$$

Expected value of a random indicator variable is just the probability of the event it represents happening.

$$E[X] = 1 + \frac{1}{2} + \frac{1}{3} \cdots \frac{1}{k}$$

Most people would not know the summation of the above series off the top of their heads, but if one can solve to this point, one has learned complexity analysis well. The above series mathematically sums to:

$$E[X] = \ln k$$

Hence we can say that Neha is expected to break-up with ***lnk*** number of guys, so her expected cost over dating ***k*** tinder matches can be represented as:

$$T(k) = 30 * k + 100lnk$$

That is roughly 530, much better than our expected worst case cost of 1200!