How to Find a Spouse
A Problem in Discrete Mathematics
With an Assist From Calculus

You are seeking a spouse and, obviously, want to find the best match possible. As you meet and date "candidates", you have the opportunity to determine how well matched you are as a couple. There are several rules to this dating game:

- It is generally considered bad form to date seriously two different people simultaneously, so you consider each person one at a time.

- You can date someone for any length of time, but eventually, you must either "select" them or say "no", and move on to another candidate.

- Once someone has been passed over, you cannot go back. No is forever.

- If there are N candidates, how can you maximize the probability that you select your best match?

It is essential that you know when a candidate is a good one and when they are not so good. The only way to gain some understanding of what is "good" is to "play the field". Date several people without serious intent to determine what attributes are important to you. This is similar to the baseball strategy of "taking a strike" before hitting. Taking a strike gives the hitter the opportunity to better judge what is a good pitch from this pitcher. In this model, we will employ the "play the field" or "take a strike" strategy.

Strategy for Finding a Spouse: Date k people without making a selection. Then, select the first person judged to be better than any of the first k.

What is the relationship between N and k that maximizes our probability of selecting the very best spouse from N choices. If k is small, we have little information. Without sufficient information about the quality of the choices, we can make a hasty and unwise uninformed choice. If k is large, then the very best choice has a greater probability of being among the first k, which guarantees that our selection will not be optimal. This, then, is the max-min dynamic. As k increases, we can make a better and better choice. But as k increases, we face the likelihood that our best choice has already passed us by before we begin the selection process.

A Mathematical Model

We want to find the value of k (relative to N) that gives us the greatest probability of selecting the best spouse for among the N potential choices. We will develop a function $P(k)$ that will compute the probability of success as a function of k. Remember, k is an integer, so the domain of this function will be $k = 0,1,2,\cdots,N-1$. If $k = 0$, this is equivalent to selecting the first person and if $k = N-1$, we select the last person.

To define $P(k)$, we consider the possible location of the best choice. They could be anywhere from 1 to N. We will be successful if we select the best person, otherwise we are unsuccessful. If we let k go by and then select the first person better than any of the first k, the probability of success can be computed using the diagram below:

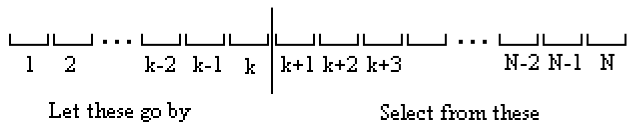

Figure 1: The Positions of the Candidates

The best person could be in the first or the second or the third, or, in fact, any position in the list. So the probability is the sum of the individual probabilities of being in a position, and being selected.
So

$$P(k)=\sum_{n=1}^{N}(\text{probofbeinginPosition } n)(\text{probof being selected giventheyareinPosition } n).$$

We need to determine each of these probabilities and add them all up.

Calculating the Probabilities

What is the probability that the best person is in Position 1? The best person is equally likely to be in Position 1 as any other position. All positions are equally likely, so each has probability $p=\frac{1}{N}$. Now, if the best person is in Position 1, what is the probability that they will be chosen? Since we will not choose any of the first k, this probability is zero. So the combined probability of being in the first position and being chosen is $P=\left(\frac{1}{N}\right)\cdot 0=0$.

In fact, for the first k positions, the value of P is $P=\left(\frac{1}{N}\right)\cdot 0=0$. So, we have

$$P(k)=\left(\frac{1}{N}\right)\cdot 0+\left(\frac{1}{N}\right)\cdot 0+\left(\frac{1}{N}\right)\cdot 0+\cdots+\left(\frac{1}{N}\right)\cdot 0+\cdots.$$

So, we have

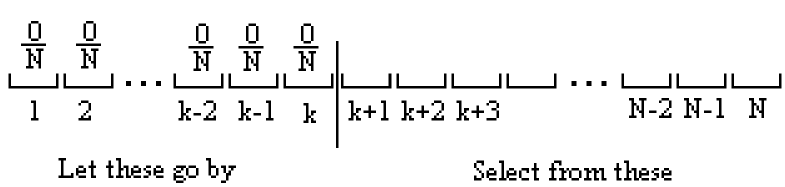

Figure 2: Probability of Success if Best Candidate is in Positions 1 to k

The first non-zero term in P comes from Position $(k+1)$. The probability that the best choice is at this position is again $p = \frac{1}{N}$. If the best choice is at this position, the probability that it will be selected using this procedure is $p = 1$. At this point,

$$P(k) = \left(\frac{1}{N}\right)\cdot 0 + \left(\frac{1}{N}\right)\cdot 0 + \left(\frac{1}{N}\right)\cdot 0 + \cdots + \left(\frac{1}{N}\right)\cdot 0 + \left(\frac{1}{N}\right)\cdot 1 + \cdots$$

Now, here come the difficult term in the summation. If we can figure out the probability for Position $k+2$, we can do all the others. Let's think about Position $k+2$.

One way to think about this is to consider how it would not be chosen. In this situation, we are given that the best choice is at Position $k+2$, so none of the first k are better than it. However, if the choice in Position $k+1$ is better than any in the first k, then it would be selected before you arrived at Position $k+2$. We are assuming there are no ties, so there will be a best in the first $k+1$ positions. If this best choice in the first $k+1$ positions falls among the first k, then the choice in Position $k+1$ cannot be better than any of the previous and so we will arrive at Position $k+2$ and select this choice. If the best choice is in Position $k+2$, we will successfully select it with probability $p = \frac{k}{k+1}$. (This is the most important point in the mathematical development. Make sure you understand why this probability is correct.)

So, at this point we have

$$P(k) = \left(\frac{1}{N}\right)\cdot 0 + \left(\frac{1}{N}\right)\cdot 0 + \left(\frac{1}{N}\right)\cdot 0 + \cdots + \left(\frac{1}{N}\right)\cdot 0 + \left(\frac{1}{N}\right)\cdot 1 + \left(\frac{1}{N}\right)\cdot\left(\frac{k}{k+1}\right) + \cdots$$

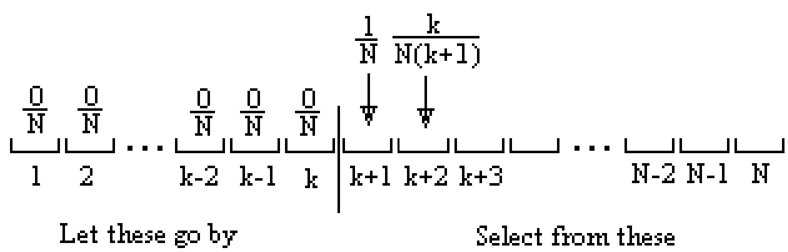

Figure 3: Probability for Positions $k+1$ and $k+2$

Continuing, we find that the probability that the best choice is at Position $k+3$ is still $p = \frac{1}{N}$. We will select it if the best choice in the first $k+2$ positions lie among the first k. This happens with probability $p = \frac{k}{k+2}$. So,

$$P(k)=\left(\frac{1}{N}\right)\cdot 0+\left(\frac{1}{N}\right)\cdot 0+\cdots+\left(\frac{1}{N}\right)\cdot 0+\left(\frac{1}{N}\right)\cdot 1+\left(\frac{1}{N}\right)\cdot\left(\frac{k}{k+1}\right)+\left(\frac{1}{N}\right)\cdot\left(\frac{k}{k+2}\right)+\cdots$$

All of the other probabilities are determined in the same way. So, we have

$$P(k)=\left(\frac{1}{N}\right)\cdot 0+\cdots+\left(\frac{1}{N}\right)\cdot 0+\left(\frac{1}{N}\right)\cdot 1+\left(\frac{1}{N}\right)\cdot\left(\frac{k}{k+1}\right)+\left(\frac{1}{N}\right)\cdot\left(\frac{k}{k+2}\right)+\cdots+\left(\frac{1}{N}\right)\cdot\left(\frac{k}{N-1}\right)$$

which simplifies to

$$P(k)=\left(\frac{1}{N}\right)\left(1+\frac{k}{k+1}+\frac{k}{k+2}+\frac{k}{k+3}+\cdots+\frac{k}{N-1}\right),\text{ for } k=0,1,2\ldots N-1$$

or

$$P(k)=\left(\frac{k}{N}\right)\left(\frac{1}{k}+\frac{1}{k+1}+\frac{1}{k+2}+\frac{1}{k+3}+\cdots+\frac{1}{N-1}\right),\text{ for } k=1,2,\cdots N-1.$$

It is important to note that the two equations above are not equivalent. In the first, the domain is $k=0,1,2\ldots N-1$ while in the second, it is restricted to $k=1,2,\cdots N-1$. We should consider whether the loss of an element in the domain will create significant problems for us. When we consider the case of $k=0$, this represents selecting the first choice. This choice will be the best with probability $p=\frac{1}{N}$. We lose this value from our function, but we are losing something whose value we already know. We can proceed to use the function

$$P(k)=\left(\frac{k}{N}\right)\left(\frac{1}{k}+\frac{1}{k+1}+\frac{1}{k+2}+\frac{1}{k+3}+\cdots+\frac{1}{N-1}\right),\text{ for } k=1,2,\cdots N-1$$

as our model.

Let's look at an example. Suppose $N=5$, then for $k=1,2,3,$ and 4, we have

$$P(0)=0.2$$
$$P(1)=\left(\frac{1}{5}\right)\left(1+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}\right)=0.4167$$
$$P(2)=\left(\frac{2}{5}\right)\left(\frac{1}{2}+\frac{1}{3}+\frac{1}{4}\right)=0.4333$$
$$P(3)=\left(\frac{3}{5}\right)\left(\frac{1}{3}+\frac{1}{4}\right)=0.3500$$
$$P(4)=\left(\frac{4}{5}\right)\left(\frac{1}{4}\right)=0.2000$$

In this example, we see that $k=2$ is the optimum value, giving us a probability of success of 0.4333. The table below presents the results for other values of N.

k	$N = 3$	$N = 4$	$N = 5$	$N = 6$	$N = 7$	$N = 8$	$N = 9$	$N = 10$	$N = 11$
0	0.333	0.25	0.200	0.167	0.143	0.125	0.111	0.100	0.091
1	0.500	0.458	0.417	0.381	0.350	0.324	0.302	0.283	0.266
2	0.333	0.457	0.433	0.428	0.414	0.398	0.382	0.366	0.351
3		0.25	0.350	0.392	0.407	0.410	0.406	0.399	0.390
4			0.200	0.300	0.352	0.380	0.393	0.398	0.398
5				0.167	0.262	0.318	0.353	0.373	0.384
6					0.143	0.232	0.290	0.327	0.352
7						0.125	0.208	0.265	0.305
8							0.111	0.189	0.244
9								0.100	0.173
10									0.091
k/N	0.333	0.500	0.400	0.333	0.286	0.375	0.333	0.300	0.364

Table1 : Best Positions for N = 3 to 11

We see the maximal probability of success is steadily decreasing. This is to be expected. We should have a better shot at picking the best from 3 than from 11 or from 111. Will the probability eventually approach the value $\frac{1}{N}$ as N increases? How can we determine the optimal value of k without listing all the values and looking to see which is the largest? Can we devise an algorithm that will more easily find the correct value?

Can we use calculus to help? Why don't we differentiate function P with respect to k, set the derivative equal to zero, and solve? It is important to remember that $P(k)$ is not a continuous function. It is only defined for $k=1,2,3\cdots N$ and so the standard calculus techniques are not an option for us. We will, however, think more about calculus later.

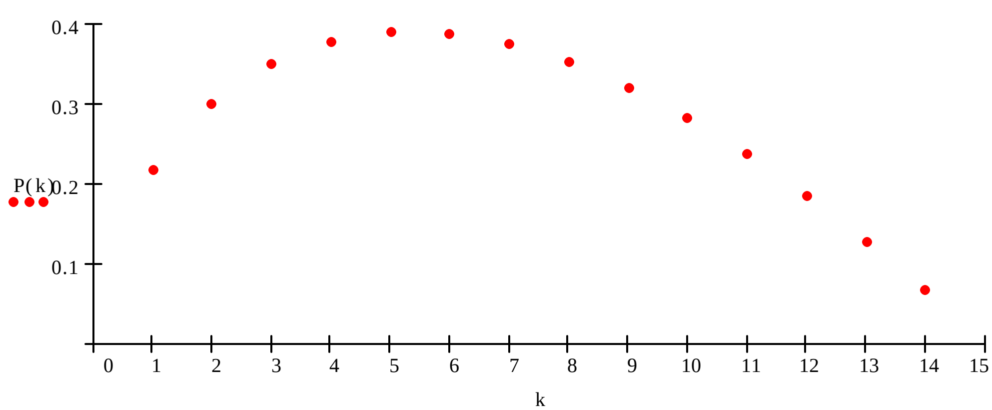

Figure 4: Plot of $P(k)$ against k for $N=15$

From the table and the plot above, we notice that the successive probabilities begin at $\frac{1}{N}$, increase to a maximum value, and then decrease back to $\frac{1}{N}$. How can we tell we are at the highest point on the graph? One way to find the maximum value is to consider the sequence of successive values of $P(k)$. The value of k we want is first value of k for which the next probability is smaller than the preceding probability, that is, for which $P(k+1)-P(k)<0$. This process is similar to using the first derivative test in calculus.

So,

$$P(k+1)-P(k)=\left(\frac{k+1}{N}\right)\left(\frac{1}{k+1}+\frac{1}{k+2}+\cdots+\frac{1}{N-1}\right)-\left(\frac{k}{N}\right)\left(\frac{1}{k}+\frac{1}{k+2}+\cdots+\frac{1}{N-1}\right)$$

$$=\left(\frac{k+1}{N}\right)\left(\frac{1}{k+1}+\frac{1}{k+2}+\cdots+\frac{1}{N-1}\right)-\left(\frac{1}{N}\right)-\left(\frac{k}{N}\right)\left(\frac{1}{k+1}+\frac{1}{k+2}+\cdots+\frac{1}{N-1}\right)$$

$$=\left(\frac{k+1}{N}-\frac{k}{N}\right)\left(\frac{1}{k+1}+\frac{1}{k+2}+\cdots+\frac{1}{N-1}\right)-\left(\frac{1}{N}\right)$$

$$=\left(\frac{1}{N}\right)\left(\frac{1}{k+1}+\frac{1}{k+2}+\cdots+\frac{1}{N-1}-1\right).$$

So, if $P(k+1)-P(k)<0$, then $\left(\frac{1}{k+1}+\frac{1}{k+2}+\cdots+\frac{1}{N-1}-1\right)<0$.

If we can compute the cumulative sum $S_k=\left[\sum_{n=k+1}^{N-1}\frac{1}{n}\right]-1$ for different values of k, the first sum that is negative will give us the solution. The table below illustrates this process for $N=15$.

k	S_k	$P(k)$	$P(k+1)-P(k)$
1	1.252	0.217	0.083
2	0.752	0.300	0.050
3	0.418	0.350	0.028
4	0.168	0.378	0.011
5	-0.032	0.389	-0.002
6	-0.198	0.387	-0.013
7	-0.341	0.374	-0.023

Table 2: Summation for N = 5

In optimization problems in discrete mathematics, solutions are often algorithms that will generate the solution, rather than a specific, clean solution. Often, students are unsatisfied with this kind of "solution". But, writing a program that terminates when $S_k = \left[\sum_{n=k+1}^{N-1} \frac{1}{n} \right] - 1 < 0$ is quite simple, so we can easily find a clean, specific solution for any given *N*. A short calculator program to find *k* is given below:

```
PROGRAM: SPOUSE
:Disp "INPUT N"
:Input N
:0→S
:N−1→X
:While S<1
:1/X+S→S
:X−1→X
:End
:Disp X+1
:X+1→K
:(K/N)(sum(seq(1/T,T,K,N−1)) )→P
:Disp P
```

Students may want to use this program for different values of N to determine empirically the relationship between *k* and *N* as well as the ultimate probability of success as *N* increases.

An Assist From Calculus

While we cannot solve the problem directly using calculus, we can generate an approximation using calculus. Students in calculus are familiar with the principle of using discrete models and methods to approximate continuous models. They see this when using Euler's method to generate approximate solutions to differential equations and when they use Reimann sums, or the Trapezoid Rule to approximate a definite integral. In this problem, we will do the reverse, We have a discrete function and we will approximate it with a continuous function. By using the more powerful techniques of calculus on the continuous approximation, we can learn something about our discrete model.

Consider the probability function

$$P(k) = \left(\frac{k}{N} \right) \left(\frac{1}{k} + \frac{1}{k+1} + \frac{1}{k+2} + \frac{1}{k+3} + \cdots + \frac{1}{N-1} \right).$$

Students in calculus may recognize the summation as an approximation for the area under the curve $y = \frac{1}{x}$ from $x = k$ to $x = N$.

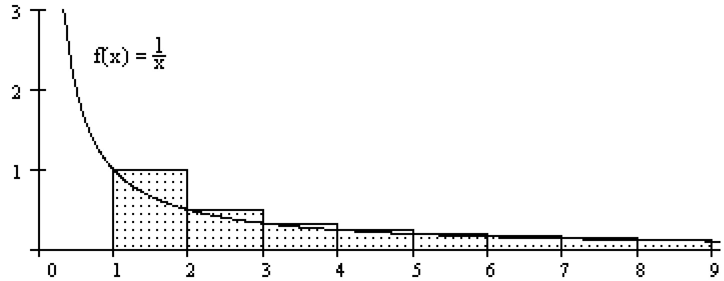

Figure 5: The area under $f(x) = \frac{1}{x}$

Calculus students recognize $\left(\frac{1}{k} + \frac{1}{k+1} + \frac{1}{k+2} + \frac{1}{k+3} + \cdots + \frac{1}{N-1}\right)$ as an approximation for $\ln(N) - \ln(k) = \ln\left(\frac{N}{k}\right)$. So, we can approximate the value of the function $P(x)$ with the function $\tilde{P}(k) = \left(\frac{k}{N}\right)\ln\left(\frac{N}{k}\right)$. From the graph above, we see that the approximation is poor when N and k are small. But as N increases in size, k also increases. Once we get beyond $N = 15$ and $k = 5$, we see the area under the curve does a good job of representing the area of the rectangles. Now, if we let the variable of interest be the ratio $\frac{k}{N} = x$, we have the approximate continuous function $\tilde{P}(x) = x\ln\left(\frac{1}{x}\right) = -x\ln(x)$. Now, calculus can help.

We have

$$\tilde{P}(x) = -x\ln(x),$$

so

$$\tilde{P}'(x) = -x\left(\frac{1}{x}\right) - \ln(x) = -1 - \ln(x).$$

Setting $\tilde{P}'(x) = 0$, we find that

$$\ln(x) = -1 \text{ and } x = \frac{1}{e}.$$

Students are always excited to see *e* show up in a problem. This says that we should let $\frac{k}{N} = \frac{1}{e} \approx 0.368$ of the positions go by before beginning our selection process. They should check their computed values from their program to see if this is a reasonable approximation.

What is the probability of success? Does it decrease to $\frac{1}{N}$? We can now approximate the probability of success using

$$\tilde{P}\left(e^{-1}\right) = -e^{-1}\ln\left(e^{-1}\right) = \frac{1}{e}.$$

The probability of success settles down as k increases to approximately 0.368 as well. Using this process, we find that we can be successful in selecting the best from a group of *N* by letting approximately 37% of the available positions go by, then selecting the first choice better than any seen before about 37% of the time. And this is true no matter how large *N* is! This is a strikingly high probability. Using this process, you can select the best out of 5000 almost 37% of the time by letting the first 1839 go by and then selecting the first choice better than any of those 1839.

It also suggests to students that marrying your high school sweetheart is not a particularly good strategy. Don't get too serious too soon. Go out with a number of people to see who you like and who likes you. Then make your choice.

References:

Shultz, Harris, and Bill Leonard, "Strategy for Hiring a Secretary", *The UMAP Journal*, Volume 8 Number 4, Winter, 1987.

Paulos, John Allen, *Innumeracy*, Hill and Wang, New York, New York, 1988.

Dan Teague
NCSSM
1219 Broad Street
Durham, NC 27705
teague@ncssm.edu