

9. The Secretary Problem

In this section we will study a wonderful problem known variously as the **secretary problem** or the **marriage problem**. It is simple to state and not difficult to solve, but the solution is interesting and a bit surprising. Also, the problem serves as a nice introduction to the general area of statistical decision making.

Statement of the Problem

As always, we must start with a clear statement of the problem. We have n **candidates** (perhaps applicants for a job or possible marriage partners). Here are our assumptions:

- The candidates are totally ordered from best to worst with no ties.
- The candidates arrive sequentially in random order.
- We can only determine the *relative ranks* of the candidates as they arrive. We cannot observe the *absolute ranks*.
- Our goal is choose the very best candidate; no one less will do. The second best candidate is of no more value to us than the worst candidate.
- Once a candidate is rejected, she is gone forever and cannot be recalled.
- The number of candidates n is known to us.

The assumptions, of course, are not entirely reasonable in real applications. The last assumption, for example, that n is known, is more appropriate for the secretary interpretation than for the marriage interpretation.

What is an optimal strategy? What is the probability of success with this strategy? What happens to the strategy and the probability of success as $n \rightarrow \infty$? In particular, when n is large, is there any reasonable hope of finding the best candidate?

Strategies

1. Play the **secretary game** several times with $n = 10$ candidates. See if you can find a good strategy just by trial and error.

After playing the secretary game a few times, it should be clear that the only reasonable type of strategy is to let a certain number $k - 1$ of the candidates go by, and then select the first candidate we see who is better than all of the previous candidates (if she exists). If she does not exist (that is, if no candidate better than all previous candidates appears), we will agree to accept the last candidate, even though this means failure. The parameter k must be between 1 and n ; if $k = 1$, we select the first candidate; if $k = n$, we select the last candidate; for any other value of k , the selected candidate is random, distributed on $\{k, k + 1, \dots, n\}$. We will refer to this “let $k - 1$ go by” strategy as **strategy k** .

Thus, we need to compute the probability of success $p_n(k)$ using strategy k with n candidates. Then we can

maximize the probability over k to find the optimal strategy, and then take the limit over n to study the asymptotics.

Analysis

First, let's do some basic computations.

2. For the case $n = 3$, list the 6 permutations of $\{1, 2, 3\}$. and verify the probabilities in the table below. Note that $k = 2$ is optimal.

k	1	2	3
$p_3(k)$	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{3}$

3. Play the **secretary game** several times with $n = 3$ candidates using the following strategies. See if you can detect a difference in the performance.
- Strategy 1
 - Strategy 2
 - Strategy 3

4. For the case $n = 4$, list the 24 permutations of $\{1, 2, 3, 4\}$. and verify the probabilities in the table below. Note that $k = 2$ is optimal.

k	1	2	3	4
$p_4(k)$	$\frac{6}{24}$	$\frac{11}{24}$	$\frac{10}{24}$	$\frac{6}{24}$

5. Play the **secretary game** several times with $n = 4$ candidates using the following strategies. See if you can detect a difference in the performance.
- Strategy 1
 - Strategy 2
 - Strategy 3

6. For the case $n = 5$, list the 120 permutations of $\{1, 2, 3, 4, 5\}$. and verify the probabilities in the table below. Note that $k = 3$ is optimal.

k	1	2	3	4	5
$p_5(k)$	$\frac{24}{120}$	$\frac{50}{120}$	$\frac{52}{120}$	$\frac{42}{120}$	$\frac{24}{120}$

7. Play the **secretary game** several times with $n = 5$ candidates using the following strategies. See if you can detect a difference in the performance.

- a. Strategy 2
- b. Strategy 3
- c. Strategy 4

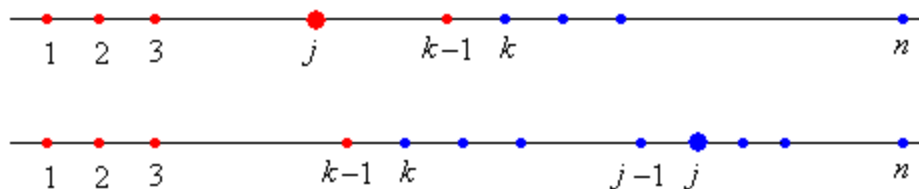
Well, clearly we don't want to keep doing *this*. Let's see if we can find a general analysis. With n candidates, let X_n denote the number (arrival order) of the best candidate, and let $S_{n,k}$ denote the event of success for strategy k (we select the best candidate).

8. Argue that since the candidates arrive in random order, X_n is uniformly distributed on $\{1, 2, \dots, n\}$.

9. Argue that

- a. $\mathbb{P}(S_{n,k}|X_n = j) = 0$ if $j \in \{1, 2, \dots, k-1\}$. If the arrival number of the best candidate is $j < k$, then strategy k will certainly fail.
- b. $\mathbb{P}(S_{n,k}|X_n = j) = \frac{k-1}{j-1}$ if $j \in \{k, k+1, \dots, n\}$. If the arrival number of the best candidate is $j \geq k$, then strategy k will succeed if and only if one of the first $k-1$ candidates (the ones that are automatically rejected) is the best among the first $j-1$ candidates.

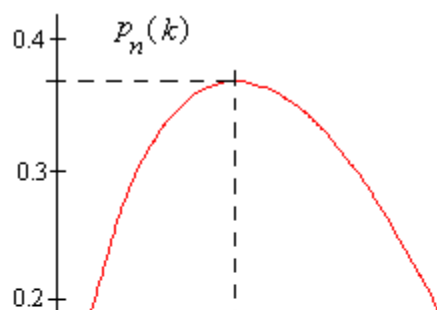
The two cases are illustrated below. The large dot indicates the best candidate. Red dots indicate candidates that are rejected out of hand, while blue dots indicate candidates that are considered.

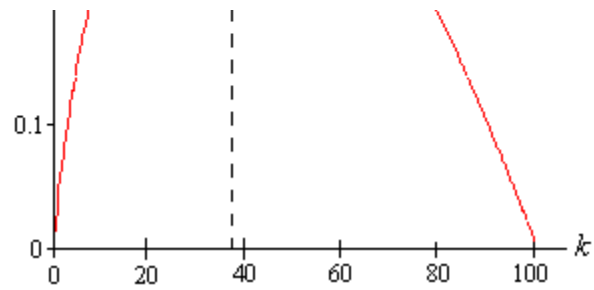


10. Use the results of the previous exercise and a conditioning argument to show that

$$p_n(k) = \mathbb{P}(S_{n,k}) = \sum_{j=k}^n \frac{1}{n} \frac{k-1}{j-1} = \frac{k-1}{n} \sum_{j=k}^n \frac{1}{j-1}$$

Values of the function p_n can be computed by hand for small n and by a computer algebra system for moderate n . The graph of p_{100} is shown below. Note the concave downward shape of the graph and the optimal value of k , which turns out to be 38. The optimal probability is about 0.37104.

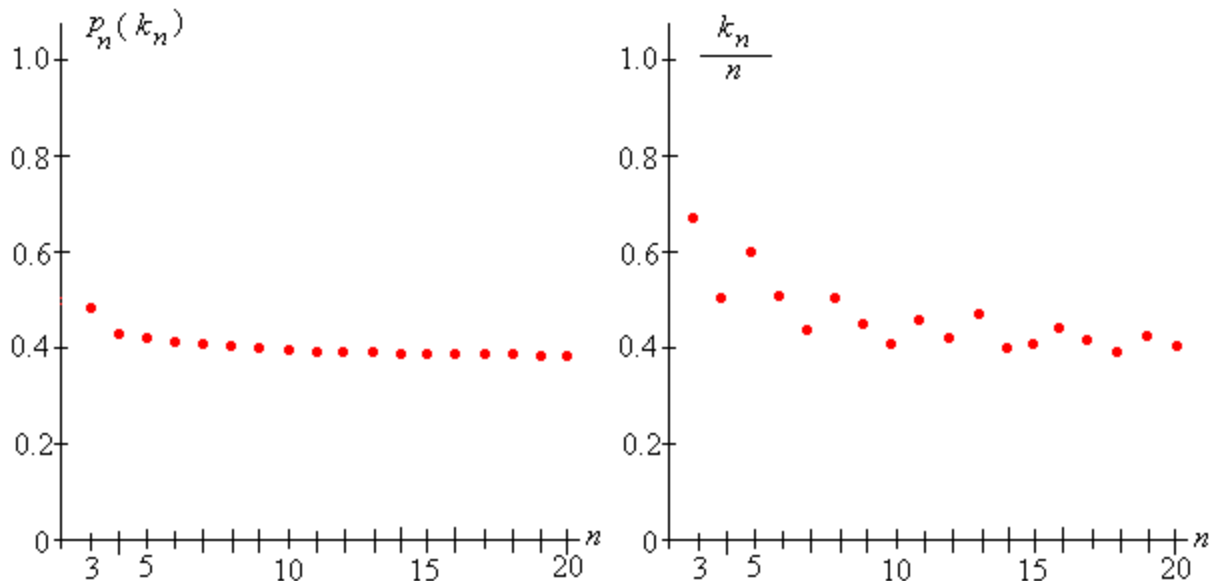




The optimal strategy k_n , the ratio $\frac{k_n}{n}$, and the optimal probability $p_n(k_n)$ of finding the best candidate, as functions of $n \in \{3, 4, \dots, 20\}$ are given in the following table:

Candidates n	Optimal strategy k_n	Ratio $\frac{k_n}{n}$	Optimal probability $p_n(k_n)$
3	2	0.6667	0.5000
4	2	0.5000	0.4853
5	3	0.6000	0.4333
6	3	0.5000	0.4278
7	3	0.4286	0.4143
8	4	0.5000	0.4098
9	4	0.4444	0.4060
10	4	0.4000	0.3987
11	5	0.4545	0.3984
12	5	0.4167	0.3955
13	6	0.4615	0.3923
14	6	0.4286	0.3917
15	6	0.4000	0.3894
16	7	0.4375	0.3881
17	7	0.4118	0.3873
18	7	0.3889	0.3854
19	8	0.4211	0.3850
20	8	0.4000	0.3842

Apparently, as we might expect, the optimal strategy k_n increases and the optimal probability $p_n(k_n)$ decreases as n increases. On the other hand, it's encouraging, and a bit surprising, that the optimal probability does not appear to be decreasing to 0. It's perhaps least clear what's going on with the ratio. Graphical displays of some of the information in the table may help:



Could it be that the ratio $\frac{k_n}{n}$ and the probability $p_n(k_n)$ are both converging, and moreover, are converging to the same number? First let's try to establish rigorously some of the trends observed in the table.

11. Show that

- $p_n(k-1) < p_n(k)$ if and only if $\sum_{j=k}^n \frac{1}{j-1} > 1$
- For each n , the function p_n at first increases and then decreases.
- The maximum occurs at the largest k with $\sum_{j=k}^n \frac{1}{j-1} > 1$. This is the optimal strategy k_n .
- As n increases, k_n increases

12. Give a probabilistic argument that the optimal probability $p_n(k_n)$ decreases as n increases.

Asymptotic Analysis

We are naturally interested in the asymptotic behavior of the function p_n , and the optimal strategy as $n \rightarrow \infty$. The key is recognizing p_n as a **Riemann sum** for a simple integral. (Riemann sums, of course, are named for **Georg Riemann**.)

13. Show that

$$p_n(k) = \frac{k-1}{n} \sum_{j=k}^n \frac{1}{n} \frac{n}{j-1}$$

14. Recognize the sum in the previous exercise as the left Riemann sum for the the function $f(t) = \frac{1}{t}$ corresponding to the partition of the interval $[\frac{k-1}{n}, 1]$ into $(n-k) + 1$ subintervals of length $\frac{1}{n}$ each:
 $(\frac{k-1}{n}, \frac{k}{n}, \frac{k+1}{n}, \dots, \frac{n-1}{n}, 1)$

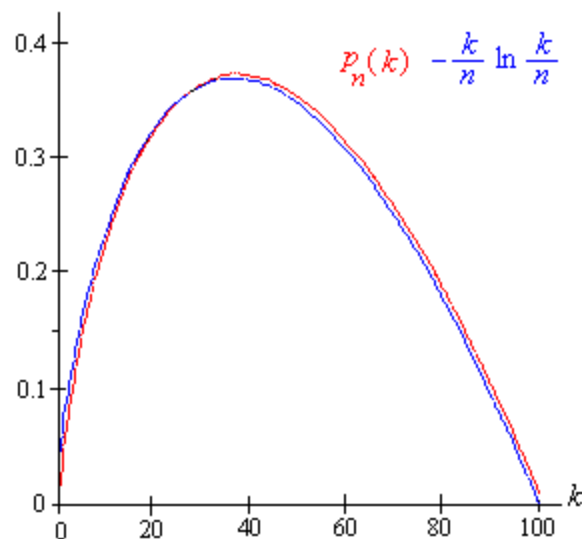
From the previous two exercises, it follows that

$$p_n(k) \approx -\frac{k-1}{n} \ln\left(\frac{k-1}{n}\right) \approx -\frac{k}{n} \ln\left(\frac{k}{n}\right)$$

More rigorously, suppose that k_n depends on n and that $\frac{k_n}{n} \rightarrow x$ as $n \rightarrow \infty$. Thus, $x \in [0, 1]$ is the limiting proportion of the candidates that we reject out of hand.

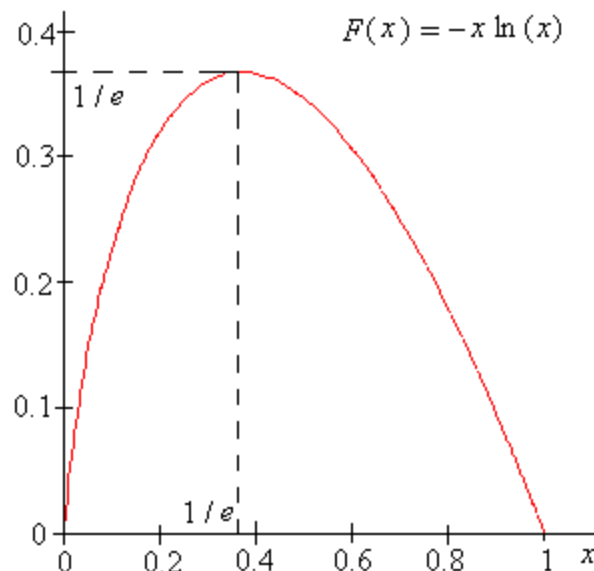
15. Show that $p_n(k_n) \rightarrow -x \ln(x)$ as $n \rightarrow \infty$.

The graph below shows the true probabilities $p_n(k)$ and the limiting values $-\frac{k}{n} \ln\left(\frac{k}{n}\right)$ as a function of k with $n = 100$.



It's easy to maximize the limiting function.

16. Show that the maximum value of $-x \ln(x)$ occurs at $x = \frac{1}{e}$ and the maximum value is also $\frac{1}{e}$.



Thus, the magic number $\frac{1}{e} \approx 0.36788$ occurs twice in the problem. For large n :

- Our approximate optimal strategy is to reject out of hand the first 37% of the candidates and then select the first candidate (if she appears) that is better than all of the previous candidates.
- Our probability of finding the best candidate is about 0.37.

The article "[Who Solved the Secretary Problem?](#)" by Tom Ferguson has an interesting historical discussion of the problem, including speculation that Johannes Kepler may have used the optimal strategy to choose his second wife. The article also discusses several generalizations of the problem.

[Virtual Laboratories](#) > [12. Finite Sampling Models](#) > [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#)

[Contents](#) | [Applets](#) | [Data Sets](#) | [Biographies](#) | [External Resources](#) | [Keywords](#) | [Feedback](#) | [©](#)