The Secretary Problem

Optimal Stopping

Arnab Chakraborty

In many spheres of activity, decisions must often be made under uncertain conditions. One such problem relates to selecting a candidate from a known number when: (a) candidates arrive in a sequence; (b) the selection process has to decide on a candidate then and there; (c) the process terminates if a candidate is selected; (d) the process continues if the candidate is not selected. The question is: What is the strategy that maximises the probability of selecting the best candidate? How does one use the 'scores' of each of the candidates seen so far to decide if the present candidate must be selected or if the process be continued, in the quest for the best candidate? This is the problem of *optimal stopping*, an example of which is discussed here.

Arnab Chakraborty is a student at Indian Statistical Institute, Calcutta. He has just completed the B.Stat. (Hons.) degree course and will soon join the M.Stat. course. Besides mathematics, probability and statistics, he is interested in astronomy. He also writes limericks and translates poems from English to Bengali.

Let us suppose that the managing director (MD) of a firm wants to employ a private secretary and has to select one out of ten candidates on the basis of an interview. The candidates, however, cannot all come on the same day; they come on different days and have to be either offered the position right away or be rejected. Once a candidate has been offered the job the others are asked not to bother to come for an interview. After each interview the MD gives the candidate a score. Naturally, she wants to appoint the person with the highest score (the best person). The MD thus has a problem. How on earth can she know before hand whether some future candidate will not be better than the ones she has interviewed already? Thus she has to work under the scenario where the order in which the ten candidates come can be any one of the 10! possible permutations of 10 persons, say with equal probability; that is, the candidates arrive in a *random order*, to use a statistical term. It is obvious that under these circumstances, there is no strategy which ensures that the MD will **always** get

Consider now the probability that a particular selection strategy gets the right candidate. If we can get these probabilities for all the possible strategies (a large set, no doubt!) then possibly we may get hold of the strategy with the highest probability and declare that as the 'best' strategy.

the right person. Thus, one way of approaching this problem is to try and work out a strategy by which the *probability* of getting the best candidate is the highest for all strategies. By strategy here, we mean a procedure which tells the MD what action to take – select or reject – at the end of each interview using the scores of all candidates interviewed thus far. For instance, a strategy may say: leave the first three candidates and select the candidate after that whose score exceeds the scores of each of his/her predecessors and if this does not happen at all, select the last candidate. We shall examine strategies like these below.

This may not be a completely satisfactory formulation of the problem. After all, although the MD wants the best candidate, she may be quite happy with a candidate whose score is fairly close to that of the best. So, another way of formulating the problem, taking into account the chance elements involved, may be to ask for a strategy which will minimise the expected value (see R L Karandikar 1996, On randomness and probability, *Resonance*, Vol. 1, No. 2, pp. 55–68, for a definition of expected value) of the difference between the scores of the best and the selected one. However, in our formulation, we have concentrated on getting the best.

It appears a bit unrealistic that the candidates arrive at different times and have to be told the results right away. However, there are situations in which this may actually happen. For instance, in the traditional Indian method of finding a spouse, alliances are generally considered one by one, each case evaluated with respect to various considerations (and say, a score is given) and the alliance accepted or rejected[1]. Realistically, the number of 'alliances' that can be considered is not known.

Consider now the probability that a particular selection strategy gets the right candidate. If we can get these probabilities for all the possible strategies (a large set, no doubt!) then possibly we may get hold of the strategy with the highest probability and declare that as the 'best' strategy. In this way, probability helps us to

[1] For obvious reasons this problem is also called the *dowry problem* or the *marriage problem* in the literature.

compare strategies whose outcomes depend on chance. Statisticians call such problems of choosing strategies as *decision problems.*

In a decision problem, we take both the losses and the probabilities of incurring them into account, and act accordingly.

Such decision problems often occur in our daily lives. In the rainy season, we all have to make a decision as to whether to carry an umbrella or not when we venture out. Here you have to choose between the actions:

(a) Take umbrella.
(b) Do not take it.

If you follow (a) then either (1) it rains and you use the umbrella, or (2) it does not rain and you have carried the umbrella in vain.

If you follow (b) then either (1) it rains and you get drenched, or (2) it does not rain and you did not have to carry the umbrella as a burden.

You can never be sure about the rain. (Some people say that it will surely not rain if and only if the weather bureau predicts rain; if it is so, then of course there is no problem!) So your decision has to be based on assessments of chances, heuristically or scientifically. Also, you compare the penalties to be paid for wrong decisions. Most of us consider getting drenched a more serious loss or penalty than the burden of an umbrella. That is why even a few clouds in the sky finds us carrying umbrellas. So, in a decision problem, we take both the losses and the probabilities of incurring them into account, and act accordingly.

But the situation that the MD faces is slightly different. To understand this let us make a game out of the MD's problem. It will take two to play it. First you write 10 distinct numbers (unknown to your opponent) on 10 pieces of identical-looking paper. Fold the pieces and mix them well. Your opponent plays the role of the MD. She randomly draws one paper, unfolds it and sees the number on it. If she signals 'acceptance' the game terminates. Otherwise she throws away that paper and selects another paper randomly from the rest, and the game continues. If she comes to the last paper, she has to accept it and, of course, the game terminates there. When the game terminates, the MD wins the game if what she accepts is the maximum number. Otherwise, she loses the game. You can see that this game imitates the secretary selection procedure. Repeat this game a large number of times (with different sets of numbers, of course) and find out how many times the opponent (MD) wins.

The object here is to find a strategy with the highest probability of choosing the best person.

With this game in hand we are ready to investigate the problem of the MD more deeply. Firstly, notice the difference between the umbrella problem and the MD's problem. In the umbrella problem, the decision was to be made just once. In the MD's problem, a decision is to be made after each interview in a sequential manner. In the umbrella problem, if some expert tells us that the probability of rain that day is 0.7 and if we quantify the loss or penalty for getting drenched and for carrying the umbrella, the problem is solved immediately. Let the loss in getting drenched be Rs 10 (the amount required to wash your clothes and for the medicine to treat the consequent cold) and the loss in carrying the umbrella is Rs 2 (let us say that you have to drink a cup of tea to relax your arm muscles), then the computation goes as follows:

Expected loss in *not* carrying umbrella: Rs $10 \times 0.7 = 7.0$
Expected loss in carrying umbrella: Rs $2 \times 0.3 = 0.6$

Evidently, it is better to carry an umbrella. This computation is explained as follows: If you do not carry an umbrella, there is loss only if it rains; the expected loss is the amount of loss (Rs 10) multiplied by the probability of rain (0.7). Similarly, if you carry an umbrella there is loss only if it does not rain and the expected loss is the amount of loss (Rs 2) multiplied by the probability of no rain (0.3). Similarly, if the probability of rain is 0.2, then the expected losses are 2 and 1.6 respectively and it is still better to carry the umbrella. If the probability of rain is 0.1, then these numbers are respectively 1 and 1.8 and it is better not to carry the umbrella.

But in the MD's problem, there is only one correct decision and thus there is no question of loss or expected loss. The object here is to find a strategy with the highest probability of choosing the best person. To outline the solution is not difficult and we do it now. We do it in terms of the game.

First, it is easy to see that what matters is the order or rank of the numbers and not the actual numbers themselves. Suppose you have drawn (and rejected) 5 papers bearing the numbers 1, –2, 59, 2.3, 7.999. The 6^{th} paper that you draw shows the number 10. Will you accept it? Surely not, since it has already been exceeded by 59 and so this cannot be the maximum number. Hence:

First, it is easy to see that what matters is the order or rank of the numbers and not the actual numbers themselves.

RULE 1: Never accept a paper bearing a number less than any on the previously drawn papers.

So we confine our attention only to those papers that show numbers higher than their predecessors. These papers, including the first one drawn, will henceforth be called 'hopefuls'. (In the literature these papers are called 'candidates', but to avoid confusion with the candidates for the secretary's post, we are using a different terminology.)

Consider once more the first draw. There is only one paper that you have seen and all the other nine are lying folded on the table. The probability that the maximum number is among those nine is obviously quite high. Suppose you reject that paper and the paper drawn next is a hopeful (otherwise reject that also by Rule 1). Now the probability that the maximum is among the remaining papers is less than that at the first step. Thus the probability that the maximum is yet to come gradually decreases as you go along and the probability of winning increases as you keep going. This, at first, seems to mean that we should go on and accept the tenth candidate. This argument, unfortunately, is incorrect, since the tenth paper may not be a hopeful; if it is, then of course we are sure to win.

Thus the best strategy is now clear. At each draw, apply Rule 1 to reject any non-hopefuls. If the draw is a hopeful, then compare

(a) the probability of winning by accepting it, and
(b) the probability of winning by rejecting it.

If (a) is higher, accept the draw; otherwise, reject the draw and proceed to the next draw. At the last draw, accept.

So all that remains to be done is to compute these probabilities. As pointed out earlier, these only depend on n, the number of candidates and not on the numbers on the papers, so long as they are distinct. We know that the probability (a) increases with draws and (b) decreases with draws. Thus initially (b) is high and hence you should go on rejecting the first few initial draws. Then after a certain stage, say, after the x^{th} draw (b) is less than (a). Therefore from the $(x+1)^{\text{th}}$ onwards you should accept at the very first draw which is a hopeful. This is the best strategy. Computation of these probabilities is somewhat technical and is explained in the sequel; however, we give here a table of the probabilities for $n=10$. From this it is evident that the right choice of x is 3 (when n is 10); that is, leave the first three draws and from the fourth draw accept the first one for which the number is greater than all the previously seen numbers. Of course, if you are at the tenth draw, just accept it. The probability of winning at each draw is shown in the table. In terms of the MD's problem, the MD should reject the first three candidates and from the fourth accept the one whose score is higher than all the earlier ones; if however it is the tenth candidate, accept. Of course, as mentioned earlier, once a candidate is selected no more candidates will be interviewed. It goes without saying that if n is different from 10, these probabilities and hence the value

Thus the best strategy is now clear. At each draw, apply Rule 1 to reject any non-hopefuls.

Table 1. Probabilities of Winning by Accepting the next 'Hopeful' after the $(x+1)^{th}$ Draw Onwards (n=10)

x	Probability	x	Probability
0	0.100	5	0.373
1	0.283	6	0.327
2	0.366	7	0.265
3	0.399	8	0.189
4	0.398	9	0.100

of x will change. As you can see, the value of x will monotonically increase with the value of n.

Let us now explain how a table like this is obtained. This requires a bit of analysis and some concepts like *conditional probabilities*. Let us call the strategy which skips the initial s candidates ($0 \le s \le n-1$) and selects the first hopeful after that as STRAT(s). We want to compute

$$P(\text{Win by STRAT}(s)).$$

Now, STRAT(s) can possibly win only when the maximum occurs at a draw after the s^{th} draw. Suppose the maximum occurs at the k^{th} draw ($s+1 \le k \le n$). Then

$$P(\text{Win by STRAT}(s))$$

$$= \sum_{k=s+1}^{n} P(\text{Win by STRAT}(s) \text{ and maximum is at the } k^{\text{th}} \text{ draw})$$

$$= \sum_{k=s+1}^{n} P(\text{Win by STRAT}(s) \mid \text{maximum is at the } k^{\text{th}} \text{ draw}) \times P(\text{maximum is at the } k^{\text{th}} \text{ draw}). \quad (1)$$

But the order being random, the maximum may occur at any of the n draws $1, 2, \ldots, n$, with equal probability $1/n$. Hence $P(\text{maximum is at the } k^{\text{th}} \text{ draw}) = 1/n$.

Now we focus our attention on the term

$P(\text{Win by STRAT}(s) \mid \text{maximum is at the } k^{\text{th}} \text{ draw}), (s+1 \le k \le n)$.

Since the draw showing up the maximum is always a hopeful, all that needs to be ensured for a win is that it is the first hopeful after the s^{th} draw. This happens only if the maximum of the initial $(k-1)$ draws lies within the first s draws. But, due to randomness of the order, this maximum may be at any of the initial $(k-1)$ draws with equal probability $1/(k-1)$. Hence the chance of its being among the first s draws is $s/(k-1)$. Thus

$$P(\text{Win by STRAT}(s) \mid \text{maximum is at } k^{\text{th}} \text{ draw})$$

$$= \frac{s}{k-1}, \quad (s+1 \le k \le n).$$

Since the draw showing up the maximum is always a hopeful, all that needs to be ensured for a win is that it is the first hopeful after the s^{th} draw.

So, we have from (1)

$$P(\text{Win by STRAT}(s)) = \frac{s}{n} \sum_{k=s+1}^{n} \frac{1}{k-1}$$

$$\frac{s}{n}\left[\frac{1}{s} + \frac{1}{s+1} + \cdots + \frac{1}{n-1}\right] = \pi(s, n), \text{ say.} \qquad (2)$$

The interested reader may check that the probabilities tabulated above are indeed the values $\pi(s, 10)$ as s varies from 0 to 9.

Let $s*$ (which may depend on n) maximise $\pi(s, n)$ for a fixed n. That is, let

$$\pi(s*, n) \ge \pi(s, n) \text{ for all } s.$$

It is easy to check that this is the same as

$$\frac{1}{s*} + \frac{1}{s*+1} + \cdots + \frac{1}{n-1} \ge 1 \ge \frac{1}{s*+1} + \cdots + \frac{1}{n-1}. \qquad (3)$$

From this we can see that $s^* \to \infty$ as $n \to \infty$. Now recall from analysis the following result:

Result: $\sum_{i=1}^{n} \frac{1}{i} \approx \ln(n) + C$, where C is the Euler's constant.

Using this result we have from (3)

$$\ln\left(\frac{n-1}{s^*}\right) \approx 1 \text{ for large } n \text{ (that is, for large } s^* \text{ also),}$$

$$\text{that is } \ln(n/s^*) \approx 1 \text{ for large } n$$

$$\Rightarrow s^* \approx n/e \text{ for large } n. \qquad (4)$$

This means that if n is large, you need not do the tedious maximisation of $\pi(s, n)$. Rather, just divide n by e and the integer nearest to the answer gives you the value of s^*. What is the probability of winning now? To find out, apply the above result to (2) to get

This means that if n is large, you need not do the tedious maximisation of $\pi(s, n)$. Rather, just divide n by e and the integer nearest to the answer gives you the value of s^*.

$$\pi(s, n) \approx \frac{s}{n} \ln\left(\frac{n-1}{s-1}\right) \text{ for large } n \text{ and } s$$

$$\approx \frac{s}{n} \ln\left(\frac{n}{s}\right) \text{ for large } n \text{ and } s.$$

Therefore

$$\pi(s^*, n) \approx \frac{s^*}{n} \ln\left(\frac{n}{s^*}\right)$$

$$\approx \frac{1}{e} \text{ for large } n, \text{ by (4).}$$

Here is an example: Suppose you are the MD and 1000 candidates appear at the interviews. Then, this is what you should do: Compute $(1000/e) = 367.9 \approx 368$. Having done that, you only score the first 368 candidates and not offer them the job. Then on, you select the first candidate whose score exceeds all the previous ones. By this process, you have a 36.79% chance of securing the best candidate. Do not feel disheartened by such a low probability. Remember that this is the maximum possible probability.

As pointed out earlier, if you formulate the problem differently, the solution will also be different. In any case, problems of this sort are often successfully tackled by a probabilistic or statistical formulation. For this, as you have seen here, probability calculations are required. Some of these probability calculations pose challenging and interesting mathematical problems and it is no wonder then that the theory of probability is such a fascinating subject!

Suggested Reading

- J H Fox, L G Marnie. Martin Gardner's column *Mathematical Games*. *Sci. Am.* 202:150-153, 1960.
- J P Gilbert, F Mosteller. Recognizing the maximum of a sequence. *J. Am. Stat. Assoc.* 61:35-79, 1966.
- J D Petruccelli. Secretary problem. *Encyclopedia of Statistical Sciences*. (Eds) S Kotz, N L Johnson and C B Read. Wiley New York. 8:326-329, 1988.

Address for Correspondence
Arnab Chakraborty
M. Stat Ist Year Student
c/o Dean of Studies
Indian Statistical Institute
Calcutta 700 035, India

In *Scientific American,* 1961... Cosmonaut Yuri A Gagarin has become the first person to cross "the border between the earth and interplanetary space" in his spaceship *Vostok.*
A report on the design and construction of satellites engineered to transmit telephone and television signals predicts that the first of these systems will be operating within five years. Progress is faster than expected, and less than a year later *Scientific American* tells of the successful launch of Bell Telephone Laboratories's *Telstar.* (From *Scientific American,* September 1995)