Knowing when to quit: A rant, how e appears in decision theory and a solution to the Secretary Problem

Georgia Skiada

March 2023

1 Introduction

How fast do things grow? We often use the phrase 'exponential growth' to indicate something growing very fast but how quickly is this exactly? How is it that an irrational number (having an infinitely recurring expansion) appears in so many areas of mathematics and manages to fit in so many beautiful equations? In this essay I will attempt to take you on a journey to discover the curious nature of the constant e and its appearance in the most unexpected places even though nobody knows why. Hopefully by the end you will have knowledge from how to bargain the best deal for your savings account (perhaps a bit of a hyperbole) to (I must admit more bizarrely) how to use e to secure a job position.

2 A Tale and a Problem of Interest

To begin we must first consider what is meant by exponential growth. To do this let's ease ourselves in with a story. A long time ago there lived a wise man who discovered the game of chess. The wealthy king of the land was so impressed with the game that he offered the man any reward he wished. The man replied:

> "My wish is for you to give me 1 grain of wheat for the first square of the chessboard, 2 for the second, 4 for the third and so on, doubling the number of grains on each successive square until the board is filled."

The king was surprised to be asked for such an insignificant reward until his treasurers calculated the total number of grains to be:

$$1 + 2 + 2^2 + 2^3 + 2^4 + ... + 2^{63}$$

This works out to be 18,446,744,073,709,551,615 grains. When placed end to end, they would reach the nearest star, Alpha Centauri, and back again! Now that we have introduced exponential growth, let us see how this links to the constant e. In 1683, the Swiss mathematician Jakob Bernoulli was investigating financial problems relating to compound interest. He considered that given a sum of money how fast it will grow depends on how often we calculate the interest (given that the compound interest is fixed). For example, let's suppose we have £1 in the bank at a compound interest of 100%. Yes, this is indeed a very generous bank. Compound interest works by adding 100% of the value of the current year to find the value of each subsequent year. Therefore, after one year, the amount rises to £2 and after 2 years the amount has risen to £4 and so on. Generally, for an investment of £1 after k years, our £1 has risen to 1×2^k. Now let's consider how the profit changes when calculated at different time intervals: After 1 year, our £1 will become £2. Now offering 50% interest every 6 months, our £1 becomes £1.50 and then £2.25 (£1.50+75p) at the end of the year. We can see that this second option is more profitable. Now what if we calculated $\frac{1}{12}$ % every month for a year? After a month, the investment would be $1 \times (1 + \frac{1}{12})$. Therefore, after a year, $1 \times (1 + \frac{1}{12})^{12} = 2.61$. From this, we can conclude that the more frequent the interest is, the higher the profit at the end of the year. Now even better, let's see what happens when we take $\frac{1}{52}$ % every week for a year. This will be $1 \times (1 + \frac{1}{52})^{52} = 2.69$. Finally, at $\frac{1}{365}$ % every day for a year, our investment would be $1 \times (1 + \frac{1}{365})^{365} = 2.71$. From the above calculations, we can generalise for a given amount of times per year(n) and an initial investment of £1:

$$(1 + \frac{1}{n})^n$$

What if I could do this continuously? What if I am earning interest every nanosecond? We can picture this as n tends very very very large and more mathematically,

$$\lim_{x \to \infty} (1 + \frac{1}{n})^n$$

What is fascinating is that the above expression tends to a limiting value and unsurprisingly this is e.

Likewise, if the rate of interest was x%, the limiting value obtained after continuous interest taken would be:

$$\lim_{x \to \infty} (1 + \frac{x}{n})^n$$

and this limiting value is e^x.

At the time Bernoulli did not know the precise value of this constant but only knew it must be between 2 and 3. It was only later in 1740 that Leonhard Euler managed to work out the value of the constant to 18 decimal places: 2.718281828459045235 (a staggering number considering he only used pen and paper). Euler gave this exponential number the name 'e'. (Bernoulli had used the letter 'b' but I wholeheartedly believe that neither chose the first letter of

their name deliberately.) Euler managed to make some pioneering discoveries concerning e one being its involvement in infinite series which we will explore in the next section.

3   To Infinity and Beyond

As mentioned in the previous section, Euler was the first to make real progress regarding the expansion of e. He achieved this by discovering that e is also the sum of the infinite series

$$e = 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + ...$$

More generally,

$$e^x = 1 + \frac{1}{1!}x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + ...$$

Now let us consider exactly why this is the case. Before we do so we must clarify some notation and theory used. The notation n! (pronounced n factorial) simply means that we multiply all the numbers up to and including n. For example, $5! = 5 \times 4 \times 3 \times 2 \times 1$. Now, in our proof, we will use something called the Binomial Expansion. This is a formula we can use to multiply out brackets of high powers. This is a very interesting topic in itself but it will only feature in this section of the essay and so providing the formula should be sufficient without going into further detail.

$$(a + b)^n = \sum_{k=0}^{n} \frac{n!}{(n-k)!k!}$$

(This means the sum from k=0 to n of the fraction to the right. Essentially, we are substituting each value of k from 0 to n into our fraction in order to form a sequence of terms. We then add all those terms together.)

Now let us start with the binomial expansion of $(1 + a)^n$.

$$(1+a)^n = 1 + \frac{n!}{(n-1)!1!}a + \frac{n!}{(n-2)!2!}a^2 + \frac{n!}{(n-3)!3!}a^3 + \frac{n!}{(n-4)!4!}a^4 + \cdots + a^n$$

Cancelling out the smaller factorials in the denominator will give us

$$= 1 + \frac{n}{1!}a + \frac{n(n-1)}{2!}a^2 + \frac{n(n-1)(n-2)}{3!}a^3 + \frac{n(n-1)(n-2)(n-3)}{4!}a^4 + \cdots + a^n$$

Now if we substitute $a = \frac{x}{n}$,

$$(1+\frac{x}{n})^n = 1 + \frac{n}{1!}(\frac{x}{n}) + \frac{n(n-1)}{2!}(\frac{x}{n})^2 + \frac{n(n-1)(n-2)}{3!}(\frac{x}{n})^3 + \frac{n(n-1)(n-2)(n-3)}{4!}(\frac{x}{n})^4 + \cdots + (\frac{x}{n})^n$$

Cancelling out the ns and simplifying the fractions will give us:

$$(1+\frac{x}{n})^n = 1 + \frac{1}{1!}x + \frac{1}{2!}(1-\frac{1}{n})x^2 + \frac{1}{3!}(1-\frac{1}{n})(1-\frac{2}{n})x^3 + \frac{1}{4!}(1-\frac{1}{n})(1-\frac{2}{n})(1-\frac{3}{n})x^4 + \cdots$$

Now taking the limit as n tends to infinity should provide us with a recognisable expression on the left hand side. Regarding the right hand side, each bracket in the form $(1 - \frac{k}{n})$ will tend to 1. Therefore,

$$e^x = 1 + \frac{1}{1!}x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \cdots$$

as required. What is also very important about this infinite series is that it converges quickly due to the factorials in the denominators increasing very rapidly. For instance, taking only the first ten terms of the series yields the approximation of e correct to 5 decimal places.
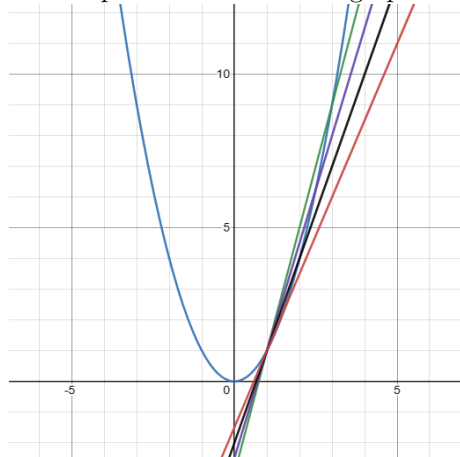
I think this is a beautiful proof not only because it enabled the start of the exploration of e but also because it connects the ideas of two of the most influential mathematicians to have ever lived. As we will discover these fascinating properties of e extend even further to the field of calculus.

4   Unique Graphical Properties

The 17th century saw the development of calculus which is the study of continuous change. It is made up of two unique strands: Differentiation (concerned with instantaneous rates of change and gradients of curves) and Integration (concerned with the area under or between curves). When we are differentiating, our main purpose is to find the gradient of a curve at a given point. In order to do this, let's pick two points on a curve and consider the gradient of the chord between those two points. We can do this using the standard formula of the difference in y values divided by the difference in x values. What will happen when the difference between the x values becomes very very very small? This is called Differentiation from First Principles and can be expressed as follows:

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

Let's explore how this works graphically.

In the above example, we are aiming to find the gradient of the curve $y = x^2$ at x=1. Notice how as we reduce the difference between our x coordinates, the gradient of the straight lines approaches the gradient we are trying to find. Now we will use our equation above to find the gradient by using the values x=x and $y = x^2$ and x=(x+h) and $y = (x + h)^2$.

$$f'(x) = \lim_{h \to 0} \frac{f(x + h) - f(x)}{h}$$

$$= \lim_{h \to 0} \frac{(x + h)^2 - x^2}{h}$$

$$= \lim_{h \to 0} \frac{2xh + h^2}{h}$$

$$= \lim_{h \to 0} 2x + h$$

$$= 2x$$

Substituting in the value of x=1 will give us the gradient of $y = x^2$ at the point (1,1). From the above, we can generate the following rule: $\frac{dy}{dx} = nx^{n-1}$ for a curve $y = x^n$.

Other rules concerning differentiation: The power rule,

$$\text{Given a function } f(x) = g(x)h(x), f'(x) = g'(x)h(x) + g(x)h'(x))$$

The quotient rule,

$$\text{Given } f(x) = \frac{g(x)}{h(x)}, f'(x) = \frac{g'(x)h(x) - g(x)h'(x)}{h^2(x)}$$

.

The chain rule is also very important. This states that:

$$\frac{dy}{dx} f(g(x)) = f'(g(x)) \times g'(x)$$

.

Essentially we are working our way through a composite function and multiplying each individual derivative of each function together. But what I really want to discuss (and is only proper that I do so given the subject of this essay) is the derivative of exponentials.

I would like to start by defining the function $f(x) = a^x$ for some constant a. We should start by using the formula to differentiate from First Principles.

$$f'(x) = \lim_{h \to 0} \frac{a^{x+h} - a^x}{h}$$

$$= \lim_{h \to 0} \frac{a^x(a^h - 1)}{h}$$

$$= a^x \lim_{h \to 0} \frac{a^h - 1}{h}$$

Notice how the $\lim_{h \to 0} \frac{a^h - 1}{h}$ has no dependency on x; it is a constant. This shows that the derivative of an exponential is always proportional to itself, with the constant of proportionality being $\lim_{h \to 0} \frac{a^h - 1}{h}$. Now the bigger question we must ask ourselves is for what value of 'a' will this constant equal 1? In other words, the derivative of which exponential will equal to itself? Let $\lim_{h \to 0} \frac{a^h - 1}{h}$ = L.

$$L(2) = 0.69317 \cdots$$

$$L(3) = 1.09867 \cdots$$

Therefore, L(a) = 1 for some $2 < a < 3$.
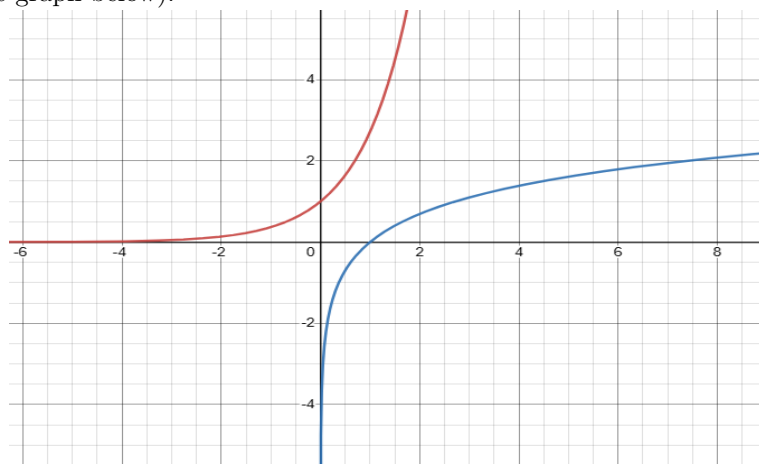
$$L(2.7) = 0.99330 \cdots$$

$$L(2.8) = 1.02967 \cdots$$

I suppose most of you would have suspected this. The value of e which gives us a constant of proportionality of 1 is e! Therefore,

$$\frac{dy}{dx} e^x = e^x$$

The derivatives of other exponentials can be understood in terms of the chain rule. For example, $\frac{dy}{dx} e^{5x} = 5e^{5x}$. This is because as we discussed, you take the derivative of the outermost function (which due to the nature of e is just itself) and multiply by the derivative of the innermost function (the derivative of 5x which is simply 5).

On the topic of calculus, the inverse function of e^x has some very interesting properties to. This is ln(x) (natural logarithm of x which is the blue curve on the graph below).

In order to understand what this means, we must ask ourselves: e to the power of what will give us x? Keeping this in mind, we can use the natural log to express any exponential with any base to one of base e. For instance, $2^x = e^{x ln(2)}$ and by extension, using this definition, we can find the derivative of any exponential with any base:

$$2^x = e^{x ln(2)}$$

$$\frac{dy}{dx} = ln(2)e^{x ln(2)}$$

$$= ln(2)2^x$$

This shows us that the constant of proportionality that we defined before of any exponential is just the natural log of its base.

As you might suspect, we can use differentiation from First Principles to also calculate the derivative of y=ln(x). In addition we will use the laws of logarithms. I will not prove these here as I predict this essay will already be too long.

$$ln(ab) = ln(a) + ln(b)$$

$$ln\frac{a}{b} = ln(a) - ln(b)$$

$$ln(a)^b = (b)ln(a)$$

Now let's try it out!

$$\frac{dy}{dx} = \lim_{h \to 0} \frac{ln(x+h) - ln(x)}{h}$$

$$= \lim_{h \to 0} \frac{ln(\frac{x+h}{x})}{h}$$

$$= \lim_{h \to 0} \frac{1}{h}ln(\frac{x+h}{x})$$

$$= \lim_{h \to 0} ln(\frac{x+h}{x})^{\frac{1}{h}}$$

$$= \lim_{h \to 0} ln(1 + \frac{h}{x})^{\frac{1}{h}}$$

Now let $n = \frac{h}{x}$ so $h = nx$. When h tends to 0, n tends to 0 too.

$$= \lim_{h \to 0} ln(1 + n)^{\frac{1}{nx}}$$
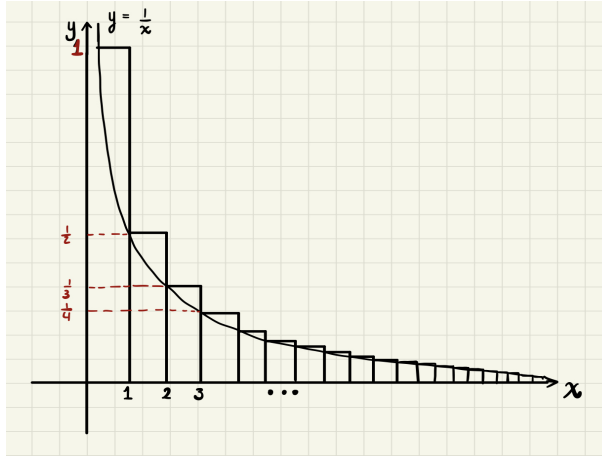
$$= \lim_{h \to 0} ln((1 + n)^{\frac{1}{n}})^{\frac{1}{x}}$$

$$= \frac{1}{x} \lim_{h \to 0} ln(1 + n)^{\frac{1}{n}}$$

$$= \frac{1}{x} ln\big(\lim_{h\to 0}(1+n)^{\frac{1}{n}}\big)$$

The expression $\lim_{h\to 0}(1+n)^{\frac{1}{n}}$ seems familiar. This is very similar to Bernoulli's expression in section 1 : $\lim_{x\to\infty}(1+\frac{1}{n})^{n})$. In fact, both of these expressions are identical and so they are both equal to e. Therefore,

$$\frac{dy}{dx} ln(x) = \frac{1}{x} ln(e) = \frac{1}{x}$$

Just as we have the differentiation from First Principles, we can use a similar logic for integration. Consider the following graph.

As we can see, the infinite sum of

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

can be used to approximate the area under the curve of $y = 1/x$. Just as we did with differentiation, we can use a formula to calculate the area precisely and avoid a rather tiresome process.

$$\int x^n dx = \frac{x^{n+1}}{n+1} + c$$

where c is any constant.

5   The secretary Problem

The Secretary Problem refers to a family of problems which display a scenario involving optimal stopping theory. This means that it is concerned with choosing a time to perform a particular action in order to achieve the best (maximum) possible outcome. All of such problems have a set structure which is as follows. Imagine an interviewer who wants to hire the best secretary out of n (known

number) applicants for a position. The applicants are interviewed one by one in random order. A decision about each particular applicant is to be made immediately after the interview. Once rejected, an applicant cannot be recalled. During the interview, the administrator gains information sufficient to rank the current applicant among all applicants interviewed so far, but is of course unaware of the quality of applicants who have not been interviewed yet. What should be the the optimal strategy (stopping rule) of the interviewer in order to maximize the probability of selecting the best applicant? The origins of the problem are uncertain. Some claim it was first introduced in February 1960 by Martin Gardner in his Mathematical Games column in the newspaper 'Scientific American'. Others attribute it to mathematician Merrill Meeks Flood who used it in one of his lectures in 1949 and then again several times in the 1950s.

So far this problem may seem detached from what we have covered so far but bear with it and I promise the pieces will fit together in the end.

Let us start by defining some key vocabulary. A permutation is a rearrangement of a set of objects. For example, 'sniitser', 'nitisers' and 'ternissi' are all distinct permutations of the word 'sinister'. We are only concerned about distinct permutations (all the letters of the word are included) but it is important to note that generally, the term is also used for any arrangement with any number of letters ('iins', 'ters' and 'site' being 4 letter permutations of the word sinister). A derangement is a permutation of a list of objects where after the rearrangement, none of the objects are still in their original place.

Now let us begin with analysing the problem. I must admit that in face of such a scenario, a tempting strategy would be to abandon yourself to luck. You may decide to make an arbitrary decision such as simply choosing the first applicant. Unsurprisingly, this strategy performs poorly. In fact, the same is true when always choosing the nth applicant. The probability of choosing the best applicant will always be $1/n$. As you can imagine, the larger the pool of applicants n, the smaller your probability of picking the best one.

Instead, it seems that we require a strategy that can be applied to any number of applicants. At this point, it is important to highlight that the only control variable we have on the entire interview process is the number of applicants we reject. Therefore, our strategy must revolve around how many people we want to reject before properly starting to decide. In other words, we should wait long enough to get a good reference point on the quality of all applicants and then choose the next applicant that beats our reference. But how long is this exactly?

Let us define P(S). This is the probability of choosing the best applicant. S is the number of applicants we stop at and when I am going to pick the best applicant after that stopping point.

$$P(S) = \sum_{k=1}^{n} \text{P(interviewing applicant n)} \times \text{P(selecting applicant n)}$$

(We have encountered the summation symbol in the second section.) This

is necessary as interviewing someone clearly does not guarantee their selection.

For all the applicants up to S, P(S) is equal to 0. This is because P(selecting applicant n)=0 as we have decided to reject all applicants until this cut-off point.

Now let us consider all applicants after S (in position S+1, S+2 etc.). Remember we are looking for the first applicant after S that beats all those up to S. Now is a good time to clarify that if none arise (no applicant after S beat the best applicant up to S), we must pick the last one. The applicant which is in position S+1 will have a $1/n$ probability of being in that position since there are still n applicants. If that candidate is better than all those before him, the probability that he is chosen is 1 (certain). Therefore if the (S+1)th applicant was indeed the best one, we have a 100% chance of picking him! However, what if the best applicant was not S+1? This means that we missed out on the opportunity of picking him. Therefore, the probability of choosing the best applicant will no longer be 100%. So instead of considering the probability of selecting applicant n, let us calculate the probability of not selecting him.

$$\text{P(not selecting applicant S+2)} \ = \frac{1}{S+1}$$

This is because there were S+1 applicants before him and any one could have been selected. Therefore,

$$\text{P(selecting applicant S+2)} \ = 1 - \frac{1}{S+1} = \frac{S}{S+1}$$

Continuing this pattern,

$$\text{P(selecting applicant S+3)} \ = \frac{S}{S+2}$$
$$\text{P(selecting applicant S+4)} \ = \frac{S}{S+3}$$
$$\text{P(selecting applicant S+5)} \ = \frac{S}{S+4}$$
$$\dots$$

Now all we need to do is find the value for S. We can start by returning to our original expression:

$$P(S) = \sum_{k=1}^{n} \text{P(interviewing applicant n)} \ \times \ \text{P(selecting applicant n)}$$

Now we have all the information we need to calculate the above summation.

Recall that all the probabilities up to and including S are 0 so we are only concerned about the probabilities after that point.

$$P(S) = \frac{1}{n} \times \frac{S}{S} + \frac{1}{n} \times \frac{S}{S+1} + \frac{1}{n} \times \frac{S}{S+2} + \dots + \frac{1}{n} \times \frac{S}{S-1}$$

$$= \frac{S}{n}(\frac{1}{S} + \frac{1}{S+1} + \frac{1}{S+2} + \cdots + \frac{1}{S-1})$$

Now let us link some of our calculus knowledge from the previous section! Just as we can use the formula of integration to work out the precise area under a curve, we can approximate it using infinite sums. Just as we showed in the previous section, this sum is the approximation of the area under the graph of $y = \frac{1}{x}$!

Now we can work out the precise area under the curve using our knowledge of integration. Just as a reminder, in the previous section, we proved why the derivative of ln(x) is $\frac{1}{x}$ and by extension, why the integral of $\frac{1}{x}$ is ln(x) (as they are inverse operations). Therefore we will integrate $\frac{1}{x}$ between S (the first permissible choice of applicant) and n (the last applicant).

$$\int_S^n \frac{1}{x}dx = [ln(x)]_S^n$$

$$= ln(n) - ln(S)$$

$$= ln(\frac{n}{S})$$

Now our long sum can simply be replaced with $ln(\frac{n}{S})$ Therefore,

$$P(S) = \frac{S}{n} \times ln(\frac{n}{S})$$
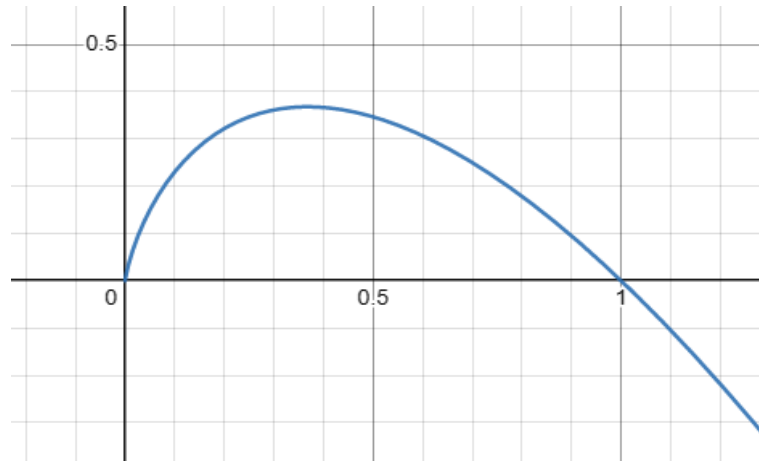
To simplify this, substitute $x = \frac{S}{n}$.

$$P(S) = x \times ln(\frac{1}{x})$$

Using our laws of logarithms,

$$P(S) = -x \times ln(x)$$

We can understand that the probability that the first applicant is the best is quite small. As we work our way through the applicants, the probability gradually increases until we reach a maximum and then fall back down again. It is the applicant with this maximum probability that we are concerned about. Now in order to locate this maximum we must employ our skills of differentiation from the previous section. Remember once we find the derivative, we must set it equal to 0. I have graphed P(S) below.

Using the product rule,

$$\frac{dP(x)}{dx} = -ln(x) - x(\frac{1}{x}) = 0$$

$$= -ln(x) - 1 = 0$$

$$ln(x) = -1$$

$$x = e^{-1} \approx 0.368$$

I know what you are thinking. Just wow! A beautiful proof with an even more beautiful ending! This tells us that we must reject 37% of the applicants and move on to the best one after that. And not only should you stop readily rejecting after 37%, but after doing so, you have a 37% chance of hiring the best applicant. What is even more fascinating is that this result applies whether you are interviewing 10 applicants or 1000. Your strategy should be the same. This is an absolutely gorgeous problem that intrigued me from the moment I encountered it. It links two seemingly unrelated strands of mathematics in the most elegant way.

In each section of this essay my main objective was to interlink different areas of mathematics gradually until a proof which would have otherwise been demanding became effortless and one that a wide range of readers could enjoy. I hope I have managed to convey to you this true beauty that I see in mathematics or even better, given you some ideas regarding how to choose the best possible date on Tinder.

Bibliography

Robin Wilson (2017) Euler's Pioneering Equation. Oxford University Press

https://en.wikipedia.org/wiki/Secretaryproblem

T. W. Kornerc (2014) Calculus for the Ambitious. Cambridge University Press

Numberphile. "e(Euler's Number)" YouTube, uploaded by Reddit, 19 Dec. 2016, hhttps://www.youtube.com/redirect?

3Blue1Brown. "What's so special about Euler's Number e?" Youtube, uploaded by 3Blue1Brown, 2 May 2017, https://www.youtube.com/watch?v=m2MIpDrF7Es

Xinlin Wu Haixiong Li (2022) A satisficing policy of the secretary problem: theory and simulation, Communications in Statistics - Theory and Methods, 51:18, 6151-6165, DOI: 10.1080/03610926.2020.1856875

Corbin, R.M. (1980). "The Secretary Problem as a Model of Choice," Journal of Mathematical Psychology 21, 1-29.

Lorenzen, T.J. (1981). "Optimal Stopping with Sampling Costs: The Secretary Problem," Annals of Probability 9, 167-172.