Diploma Programme

Cases of Academic Misconduct

May 2015
Examination Session

Subject Group 5

Mathematics HL
Exploration
Infringement: Plagiarism

International Baccalaureate®
Baccalauréat International
Bachillerato Internacional

Original Script

IB Math HL Period 2
1/20/15

Math IA: "The Odds and Relationships of Dating"

The concept of dating can be superficial for some but pretty hard for others. For candidates like myself, I have always wondered when or how can I find my ideal soul mate. The basis of the investigation is grounded on the chance (or probability) that my ideal mate will come to me after X amount of women I reject or pass through relations with. The entire concept is based the idea of the secretary problem, a problem that for years stumped even the best of mathematicians. The math later in the IA is all based of piecing together solutions and concepts from multiple sites, however, terminology that is needed to speak in the mathematical language properly was repeated at some points. In general-the Secretary problem was a problem designed by CEO's to determine which candidate (for secretary) was most ideal for the job. At the time however, the problem had several functions in society aside from helping a businessman find a valuable employee. The main use of this problem had appeared in the dating market. People believed the problem could determine their own chance of finding their true love and hone in the exact person (i.e the first person, the fifth, etc.). The problem, while solid in nature and generally accepted by mathematicians worldwide, does not address the other factors that can change the ultimate solution. It lacks personal connection, and is just a solution in general-(meaning my chances are different than yours or the one in the problem). Love, after all, is something not calculated, but left completely to chance, deriving itself from the personal and emotional levels.

Since the problem can be confusing, the process is defined throughout the investigation and can be simplified into 4 simple steps which are referenced throughout the problem. Not only does the secretary problem allow one to determine an ideal candidate in a set, but ultimately determine the general probability of finding the ideal partner. The problem begins with investigating specific probabilities of n candidates, but then transitions to its general form.

1. Use specific numbers (the solution uses 1-5) and take the sum of their Permutations (${}_nP_r$ where $R \leq 0 \leq N$)

Key strategy: 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 does not exist, we will

Copied Source

Comments

Although it is listed in the bibliography, the copied text is not referenced in the body of the text. It is essential that we indicate the author of the words we are using at all times, in a referencing style of the candidate's choosing. The candidate has not used any referencing style here.

http://www.math.uah.edu/stat/urn/Secretary.html

9. The Secretary Problem

... *secretary problem* ... *marriage problem* ...

Strategies

... 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 ...

International Baccalaureate
Baccalauréat International
Bachillerato Internacional

Original Script

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,…,n} . We will refer to this let k go by strategy as *strategy* k .

Process Required to Determine Solution: 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 asymptotic behavior.

Analysis

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)	2/6	3/6	1/6

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)	4/24	9/24	6/24	4/24

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)	15/120	30/120	32/120	28/120	15/120

Copied Source

s, f o ca d da e e e a a p ev o s ca d da es appea s), we w ag ee o accep e as ca d da e, eve o g s ea s fa e T e pa a e e *k* s e e wee 1 a d *n* f *k* 1, we se ec e f s ca d da e f *k* *n*, we se ec e as ca d da e f a y o e va e of *k*, e se ec ed ca d da e s a do , d s ed o *k*, *k*+1, …, *n*} We w efe o s " e *k* go y" s a egy as *strategy k*

T s, we eed o co p e e p o a y of s ccess $p_n(k)$ s g s a egy *k* w *n* ca d da es T e we ca ax ze e p o a y ove *k* o f e op a s a egy, a d e ake e ove *n* o s dy e asy p o c e av o

Analysis

F s , e 's do so e as c co p a o s

3 Fo e case *n* 3, s e 6 pe a o s of 1, 2, 3} a d ve fy e p o a es e a e e w No e a *k* 2 so a

k	1	2	3
$p_3(k)$	$\frac{2}{6}$	$\frac{3}{6}$	$\frac{2}{6}$

4 I e sec e a y expe e , se e e of ca d da es o *n* 3 R e expe e 1000 es w eac s a egy $k \in$ 1, 2, 3}

5 Fo e case *n* 4, s e 24 pe a o s of 1, 2, 3, 4} a d ve fy e p o a es e a e e ow No e a *k* 2 s op a

k	1	2	3	4
p (k)	$\frac{}{2}$	$\frac{11}{2}$	$\frac{1}{2}$	$\frac{6}{2}$

6 I e sec e a y expe e , se e e of ca d da es o *n* 4 R e expe e 1000 es w eac s a egy $k \in$ 1, 2, 3, 4}

7 Fo e case *n* 5, s e 120 pe a o s of ,2, 3, 4, 5} a d ve fy e p o a es e a e e ow No e a *k* 3 s op a

k		2	3	4	5
$p_5(k)$	$\frac{2}{120}$	$\frac{5}{120}$	$\frac{52}{120}$	$\frac{2}{120}$	$\frac{2}{120}$

8 I e sec e a y expe e , se e e of ca d da es *n* 5 R e expe e 000 es w eac s a egy $k \in$ 1, 2, 3, 4, 5}

Original Script

General Note: The intention of the probabilities notated in the tables above are to display a parabolic trend as shown later in the solution, for the number of outcomes cannot exceed the amount of possible outcomes via the factorial of n. The numbers were modified based on the actual permutation and not based on a mixed result.

-Now that certain values are examined, a proof of the general problem can be established- recalling that the end result of the probabilities results in the middle candidate being the optimal, which graphical is noted as a optimal point-or better known as the maximum of a parabolic function.

General form of the problem: 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)

X n is uniformly distributed on {1,2,…,n} .

This follows since the candidates arrive in random order.

For n∈N + and k∈{2,3,…,n} ,

P(S n,k |X n =j)={0 j∈{1,2,…,k 1}, k 1/j−1 j∈{k,k+1,…,n}

Copied Source

We , c ea y we do ' wa o keep do g *this* Le 's see f we ca f d a ge e a a a ys s W n ca d da es, e X_n de o e e e (a va o de) of e es ca d da e, a d e $S_{n,k}$ de o e e eve of s ccess f s a egy k (we se ec e es ca d da e)

9 X_n s f y d s ed o 1, 2, …, n}

▶ *Proof:*

Nex we w co p e e co d o a p o a y of s ccess g ve e a va o de of e es ca d da e

10 Fo $n \in$ N a d $k \in$ 2, 3, …, n},

$$P(S_{n,k} \quad X_n \quad j) \begin{cases} 0, & j \in \quad ,2, \ldots, k-1\} \\ \frac{k \quad 1}{j \quad 1}, & j \in \; k, k+1, \ldots, n\} \end{cases}$$

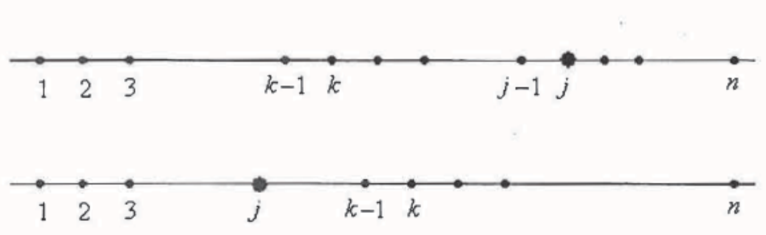
▶ *Proof:*

Original Script

Proof:

For the first case, note that if the arrival number of the best candidate is j<k , then strategy k will certainly fail. For the second cases, note that if the arrival order of the best candidate is j≥k , then strategy k will succeed if and only if one of the first k 1 candidates (the ones that areautomatically rejected) is the best among the first j 1.

-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.



Now we can compute the probability of success with strategy k .

For n∈N + p n (k)=P(S n,k)={1n ,k 1n ∑ n j=k 1j 1 , k=1k∈{2,3,…,n}

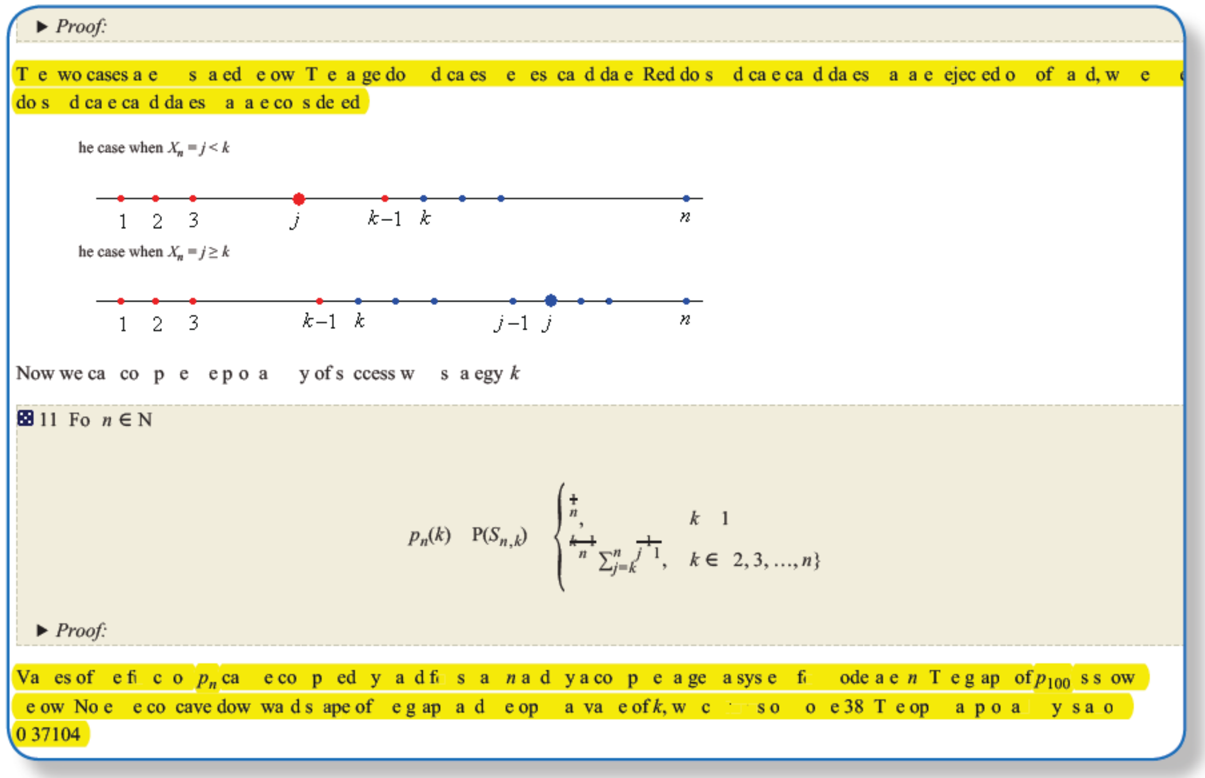
Proof: When k=1 we simply select the first candidate. This candidate will be the best one with probability 1/n . The result for k∈{2,3,…,n}

P(S n,k)=∑ j=1 n P(X n =j)P(S n,k |X n =j)=∑ j=k n 1n k 1j 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.

Copied Source



Original Script

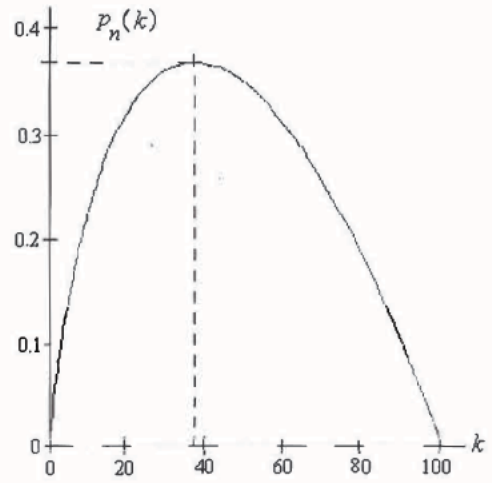
Candidates n	Optimal strategy k n	Ratio k n /n	Optimal probability p n (k n)
3	2	0.6667	0.5000
4	2	0.5000	0.4583
5	3	0.6000	0.4333
6	3	0.5000	0.4278
7	3	0.4286	0.4143
8	4	0.5000	0.4098
9	4	0.4444	0.4060
10	4	0.4000	0.3987
11	5	0.4545	0.3984
12	5	0.4167	0.3955
13	6	0.4615	0.3923
14	6	0.4286	0.3917
15	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

2. Graph these numbers using the notation P(n) where n is the number of candidates I meet and P(n) is the sum of the permutations.

Copied Source

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

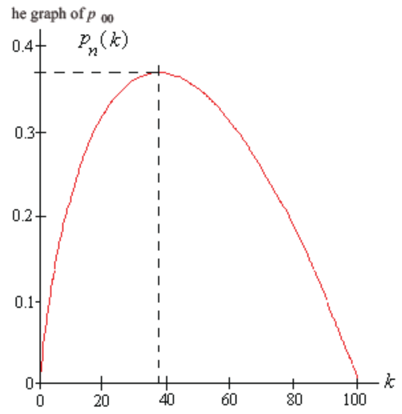
Original Script



The optimal strategy k n that maximizes k↦p n (k) , the ratio k n /n , and the optimal probability p n (k n) of finding the best candidate, as functions of n∈{3,4,…,20} are given.

Apparently, as we might expect, the optimal strategy k n increases and the optimal probability p n (k n) decreases as n ↦∞ . 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:

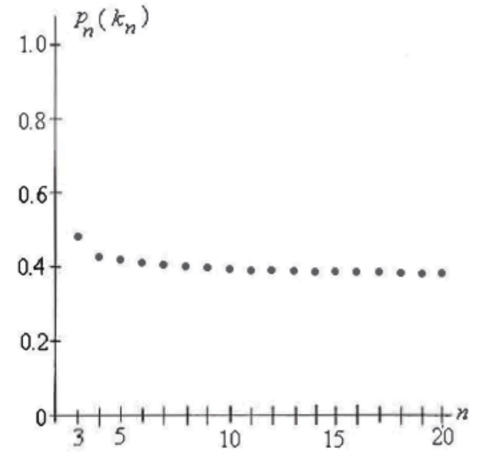
Copied Source

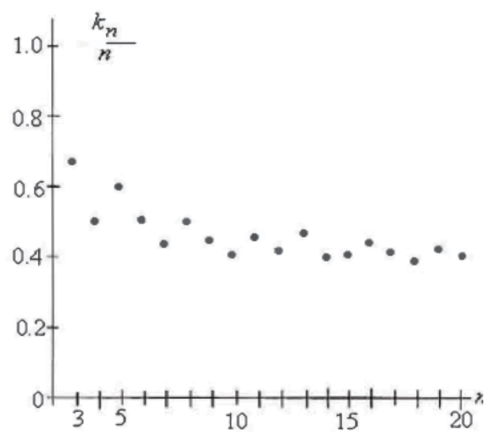


T e op a s a egy k_n a ax zes $k \mapsto p_n(k)$, e a o k_n/n, a d e op a p o a y $p_n(k_n)$ of f d g e es ca d da e, as fi c o s c $n \in$ 3, 4, …, 20} a e g ve e fi ow g a e

Appa e y, as we g expec , e op a s a egy k_n c eases a d e op a p o a y $p_n(k_n)$ dec eases as $n \to \infty$ O e o e a d, 's e co ag g, a d a s p s g, a e op a p o a y does o appea o e dec eas g o 0 I 's pe aps eas c ea w a 's go g o w e a o G ap ca d sp ays of so e of e fi a o e a e ay e p

Original Script





Could it be that the ratio 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.

The success probability p (n) satisfies

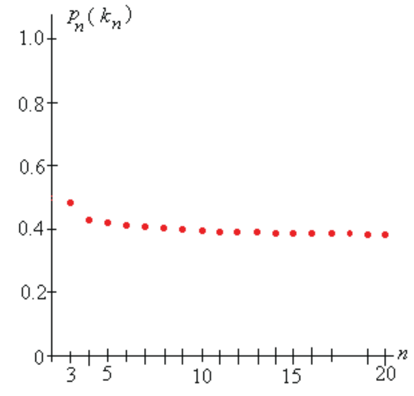
p n (k 1)<p n (k) if and only if ∑ j=k n 1j 1 >1

It follows that for each n∈N + , the function p n at first increases and then decreases. The maximum value of p n occurs at the largest k with ∑ n j=k 1j 1 >1 .

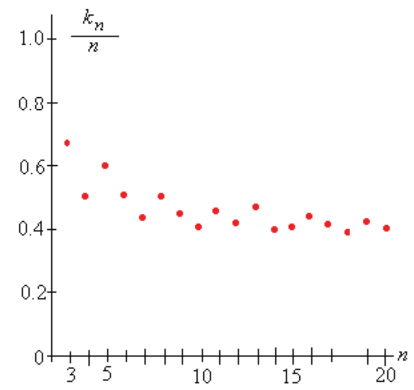
3. Find a general pattern to all of the graphs, and prove a general case using mathematical induction and setting the number of candidates to a number K value(the solution goes up to K=100)

Asymptotes: We are naturally interested in the asymptotic behavior of the function p n , and the optimal strategy as n→∞ . The key is recognizing p (n) as a *Riemann sum* for a simple integral.

Copied Source



he optima ratio k_n n



Co d e a e a o k_n/n a d e p o a y $p_n(k_n)$ a e o co ve g g, a d o eove , a e co ve g g o e sa e e ? F s e 's y o es a s go o sy so e of e e ds o se ved e a e

12 T e s ccess p o a y p_n sa sf es

$$p_n(k-1) < p_n(k) \text{ f a d o y f } \sum_{j=k}^{n} \frac{1}{j-1} > 1$$

I f ows a f eac $n \in$ N e f c o p_n a f s c eases a d e dec eases T e ax va e of p_n occ s a e a ges k w

Asymptotic Analysis

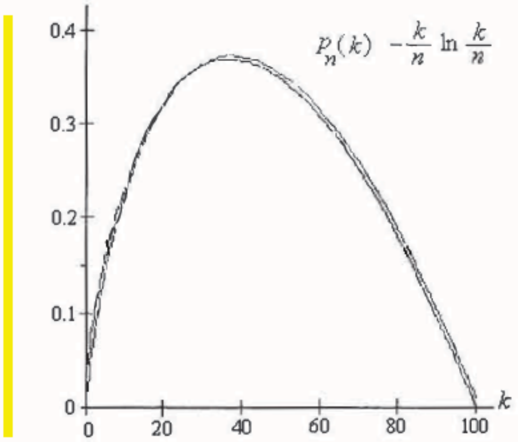
We a e a a y e es ed e asy p c e av o of e f c o p_n, a d e op a s a egy as $n \to \infty$ T e key s ecog z g p_n as a *Riemann sum* f a s p e eg a (R e a s s, of co se, a e a ed fo Geo g R e a)

Original Script

If k(n) depends on n and k(n)/n→x∈(0,1) as n→∞ then p n [k(n)]→ x ln(x) as n→∞ .

If k/n→x∈(0,1) as n→∞ then the expression on the right converges to xln(x) as n→∞ .

The graph below shows the true probabilities p n (k) and the limiting values kn ln(kn) as a function of k with n=100 .



4. Once you have determined the final form, take its limit as X→∞ to determine any asymptotic behavior.

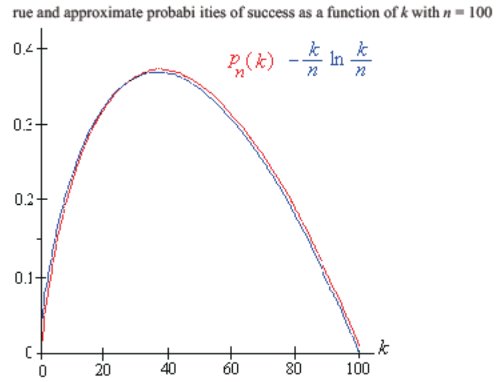
Copied Source

14 If $k(n)$ depe ds o n a d $k(n)/n \to x \in (0,1)$ as $n \to \infty$ e $p_n\,k(n)] \to -x$ (x)as $n \to \infty$

▶ *Proof:*

T e g ap e ows ows e e p o a es $p_n(k)$ a d e g va es $-\frac{k}{n}$ $\frac{k}{n}$) as a fi c o of k w n 100

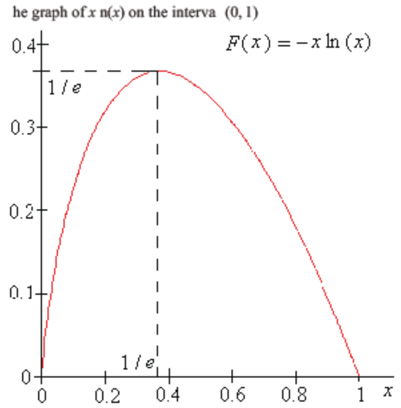
rue and approximate probabi ities of success as a function of k with $n = 100$



Fo e op a s a egy k_n, e e ex s s $x_0 \in (0,1)$ s c a $k_n/n \to x_0$ as $n \to \infty$ T s, $x_0 \in (0,1)$ s e g p opo o of e ca d da es a we ejec o of a d Mo eove , x_0 ax zes $x \mapsto -x$ (x) o $(0,1)$

15 T e ax va e of $-x$ (x) occ s a x_0 $1/e$ a d e ax va e s a so $1/e$

he graph of x n(x) on the interva $(0,1)$



T s, e ag c e $1/e \approx 0.37104$ occ s w ce e p o e Fo a ge n

- O app ox a e op a s a egy s o ejec o of a d e fi s 37% of e ca d da es a d e se ec e fi s ca d da e (fi s e appea s) a s e e a a of e p evo s ca d da es
- O p o a y of fi d g e es ca d da e s a o 0 37

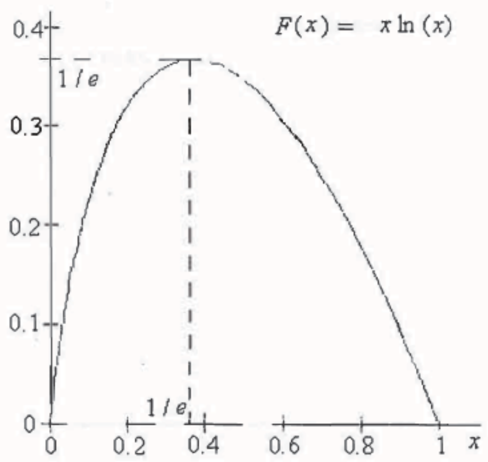
T e a c e "W o So ved e Sec e a y P o e ?" y To Fe g so as a e es g so ca d sc sso of e p o e , c d g spec a o a Jo a es Kep e ay ave sed e op a s a egy o c oose s seco d w fe T e a c e a so d sc sses seve a ge e a za o s of e p o e

Virtual Laboratories > 11 Finite Sampling Models > 1 2 3 4 5 6 7 8 9
Contents Apps Data Sets Biographies External Resources Feedback ©

Search

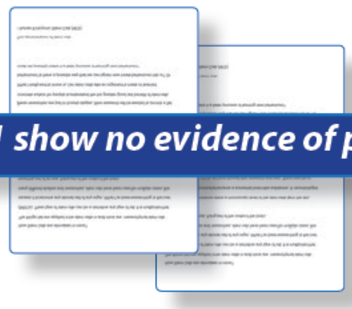
Original Script

For the optimal strategy k n , there exists x 0 □(0,1) such that k n /n→x 0 as n→∞ .
Thus,x (0) □(0,1) is the limiting proportion of the candidates that we reject out of hand.
Moreover, x (0) maximizes x□ xln(x) on (0,1)



Thus, the magic number 1/e≈0.37104 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



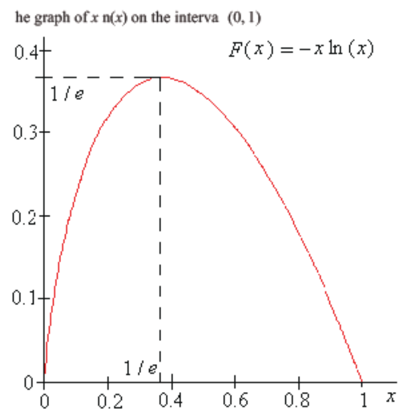
Pages 10-11 show no evidence of plagiarism.

Copied Source

Fo e op a s a egy k_n, e e ex s s $x_0 \in (0,1)$ s c a $k_n/n \to x_0$ as $n \to \infty$ T s, $x_0 \in (0,1)$ s e g p opo o of e ca d da es a we ejec o of a d Mo eove , x_0 ax zes $x \mapsto -x$ (x) o (0, 1)

15 T e ax va e of $-x$ (x) occ s a x_0 1/e a d e ax va e s a so 1/e

he graph of x n(x) on the interva (0, 1)



T s, e ag c e 1/e ≈ 0 37104 occ s w ce e p o e Fo a ge n

- O app ox a e op a s a egy s o ejec o of a d e f s 37% of e ca d da es a d e se ec e f s ca d da e (f s e appea s) a s e e a a of e p ev o s ca d da es
- O p o a y of f d g e es ca d da e s a o 0 37

T e a ce "W o So ved e Sec e a y P o e ?" y To Fe g so as a e es g s o ca d sc ss o of e p o e , c d g spec a o a Jo a es Kep e ay ave sed e op a s a egy o c oose s seco d w fe T e a c e a so d sc sses seve a ge e a za o s of e p o e