Optimal Dating Strategy

Wittawat Jitkrittum

Gatsby Tea Talk

22 Oct 2015

Overview

- A.K.A. secretary problem, marriage problem.
- Well known problem in statistics.
- **Goal**: Select the best out of n candidates.

Conditions:

1. Candidates arrive one at a time in random order.
2. n known.
3. Candidates can be ranked unambiguously.
4. After the interview (or date), each candidate is either accepted or rejected. Irrevocable decision.
5. Decision based only on the relative ranks.

Strategy k

Possibly the only reasonable strategy.

Strategy k (a stopping rule) = "let $k-1$ go by" strategy

1. Sample $k-1$ candidates to get a rough idea about the pool.
2. Reject them all.
3. Pick the next one who is better than all the $k-1$ samples.

- **Main question**: What is the best k?
- k will depend on n.

What is the best k?

- k too small. $k = 3$.

- k too big. $k = 14$. Samples include the global best. Last one selected.

- "Success" := finding the global best (after the $k - 1$ samples).
- Let $S_{n,k}$ be the event of success.

$$\text{Best } k = \arg\max_k p(S_{n,k}).$$

- Find k that maximizes the probability of success.

Example: $n = 3$

- Assume $3 > 2 > 1$. Candidate 3 is the best.
- $3! = 6$ possible arrival patterns.
- Green = best candidate selected. Orange = samples.
  - Tables from http://datagenetics.com/blog/december32012/index.html

$k = 1$ (no sample)

1	2	3
1	3	2
2	1	3
2	3	1
3	1	2
3	2	1

- $p(S_{n,k}) = 2/6$

$k = 2$ (1 sample)

1	2	3
1	3	2
2	1	3
2	3	1
3	1	2
3	2	1

- $p(S_{n,k}) = 3/6$

$k = 3$

1	2	3
1	3	2
2	1	3
2	3	1
3	1	2
3	2	1

- $p(S_{n,k}) = 2/6$

- $k = 2$ gives the highest probability.

Example: $n = 4$

$k = 1$ (no sample)

1	2	3	4
1	2	4	3
1	3	2	4
1	3	4	2
1	4	2	3
1	4	3	2
2	1	3	4
2	1	4	3
2	3	1	4
2	3	4	1
2	4	1	3
2	4	3	1
3	1	2	4
3	1	4	2
3	2	1	4
3	2	4	1
3	4	1	2
3	4	2	1
4	1	2	3
4	1	3	2
4	2	1	3
4	2	3	1
4	3	1	2
4	3	2	1

$p(S_{n,k}) = 6/24$

$k = 2$

1	2	3	4
1	2	4	3
1	3	2	4
1	3	4	2
1	4	2	3
1	4	3	2
2	1	3	4
2	1	4	3
2	3	1	4
2	3	4	1
2	4	1	3
2	4	3	1
3	1	2	4
3	1	4	2
3	2	1	4
3	2	4	1
3	4	1	2
3	4	2	1
4	1	2	3
4	1	3	2
4	2	1	3
4	2	3	1
4	3	1	2
4	3	2	1

$p(S_{n,k}) = 11/24$

$k = 3$

1	2	3	4
1	2	4	3
1	3	2	4
1	3	4	2
1	4	2	3
1	4	3	2
2	1	3	4
2	1	4	3
2	3	1	4
2	3	4	1
2	4	1	3
2	4	3	1
3	1	2	4
3	1	4	2
3	2	1	4
3	2	4	1
3	4	1	2
3	4	2	1
4	1	2	3
4	1	3	2
4	2	1	3
4	2	3	1
4	3	1	2
4	3	2	1

$p(S_{n,k}) = 10/24$

$k = 4$

1	2	3	4
1	2	4	3
1	3	2	4
1	3	4	2
1	4	2	3
1	4	3	2
2	1	3	4
2	1	4	3
2	3	1	4
2	3	4	1
2	4	1	3
2	4	3	1
3	1	2	4
3	1	4	2
3	2	1	4
3	2	4	1
3	4	1	2
3	4	2	1
4	1	2	3
4	1	3	2
4	2	1	3
4	2	3	1
4	3	1	2
4	3	2	1

$p(S_{n,k}) = 6/24$

- $k = 2$ gives the highest probability.
- $k = 1$ or $k = n$ always give uniform probabilities.

Success probability of strategy k

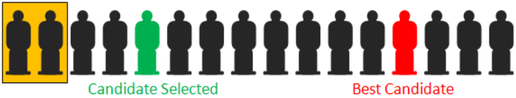
$$p(S_{n,k}) = \sum_{i=1}^{n} p(\text{i selected} \cap \text{ i best}) = \sum_{i=1}^{n} p(\text{i selected} \mid \text{i best})p(\text{i best})$$

$$= \frac{1}{n}\left[\sum_{i=1}^{k-1} \overbrace{p(\text{i selected} \mid \text{i best})}^{\text{0. best is in the samples. See 1)}} + \sum_{i=k}^{n} \overbrace{p(\text{i selected} \mid \text{i best})}^{\text{see 2)}}\right]$$

$$= \frac{1}{n}\sum_{i=k}^{n} p(\text{best up to i-1 is in the samples} \mid \text{i best})$$

$$= \frac{1}{n}\sum_{i=k}^{n} \frac{k-1}{i-1}$$

1) 

2) 

As $n \to \infty$

$$p(S_{n,k}) = \frac{1}{n}\sum_{i=k}^{n}\frac{k-1}{i-1} = \frac{k-1}{n}\sum_{i=k}^{n}\frac{n}{i-1}\frac{1}{n}$$
$$= \left(\frac{k}{n}-\frac{1}{n}\right)\sum_{i=k}^{n}\left(\frac{i}{n}-\frac{1}{n}\right)^{-1}\frac{1}{n}$$

- Let $t := i/n$
- Let $x := k/n$ = proportion of candidates to be used as samples
- $dt := 1/n$
- As $n \to \infty$, the sum above is a Riemann sum.

$$p(S_{n,k}) = \left(x-\frac{1}{n}\right)\sum_{t=k/n}^{n/n}\left(t-\frac{1}{n}\right)^{-1}\frac{1}{n}$$
$$\to x\int_{x}^{1}\frac{1}{t}\,dt = -x\log(x).$$

- $-\frac{\partial x \log x}{\partial x} = 0 \implies x = e^{-1} \approx 0.3679$
- What is the best k? **Answer:** $k = n/e \approx 37\% \times n$

Practical for dating?

- Say I will live for another 50 years.
- 3 years for a partner. So $n = 50/3 \approx 17$.
- $k = 17/e = 6.25 \approx 6$.
- Spend $3 \times 6 = 18$ years checking the 6 "samples".
- Then dump them all.
-
- ...
-

Practical for dating?

- Say I will live for another 50 years.
- 3 years for a partner. So $n = 50/3 \approx 17$.
- $k = 17/e = 6.25 \approx 6$.
- Spend $3 \times 6 = 18$ years checking the 6 "samples".
- Then dump them all.
- Pick the next one that is better than all the 6 people (assuming that I will still be attractive and can be picky).
- ...
-

Practical for dating?

- Say I will live for another 50 years.
- 3 years for a partner. So $n = 50/3 \approx 17$.
- $k = 17/e = 6.25 \approx 6$.
- Spend $3 \times 6 = 18$ years checking the 6 "samples".
- Then dump them all.
- Pick the next one that is better than all the 6 people (assuming that I will still be attractive and can be picky).
- ... All assuming that I can find the first one to begin with.
-

Practical for dating?

- Say I will live for another 50 years.
- 3 years for a partner. So $n = 50/3 \approx 17$.
- $k = 17/e = 6.25 \approx 6$.
- Spend $3 \times 6 = 18$ years checking the 6 "samples".
- Then dump them all.
- Pick the next one that is better than all the 6 people (assuming that I will still be attractive and can be picky).
- ... All assuming that I can find the first one to begin with.
- Conclusion: Not practical.

References I

- Took many images from: http://datagenetics.com/blog/december32012/index.html
- Job queue image: https://www.quora.com/Probability-statistics-1/What-are-the-most-interesting-or-popular-probability-puzzles-in-which-the-intuition-is-contrary-to-the-solution
- http://www.math.uah.edu/stat/urn/Secretary.html
- Thomas S. Ferguson. Statist. Sci. Volume 4, Number 3 (1989), 282-289.