The Secretary Problem with Independent Sampling

José Correa	Universidad de Chile
Andrés Cristi	Universidad de Chile
Laurent Feuilloley	Université Lyon 1
Tim Oosterwijk	Maastricht University
Alexandros Tsigonias-Dimitriadis	TUM

MC2 seminar · ENS Lyon

18th March 2021

The classic secretary problem

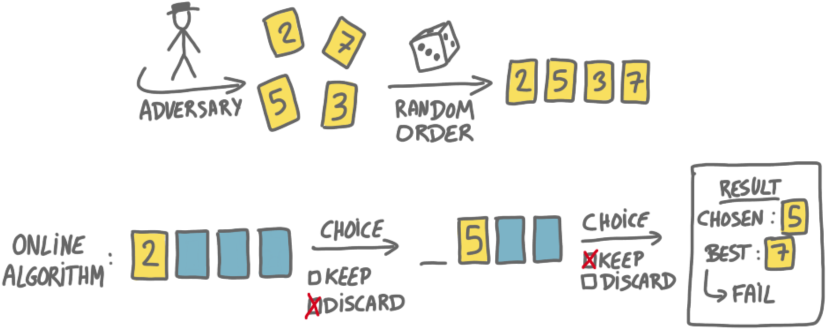
1. An adversary chooses a set of numbers.

The classic secretary problem

1. An adversary chooses a set of numbers.
2. The numbers are placed in a random order

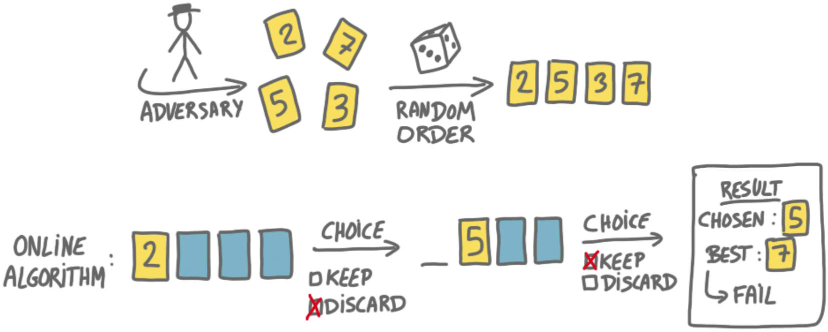
The classic secretary problem

1. An adversary chooses a set of numbers.
2. The numbers are placed in a random order
3. The numbers are presented to the player in that order.



The classic secretary problem

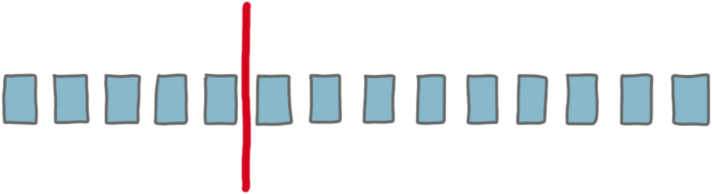
1. An adversary chooses a set of numbers.
2. The numbers are placed in a random order
3. The numbers are presented to the player in that order.
4. Goal: maximize probability of picking the max.



Stopping rule

A stopping rule is a strategy of the following form:

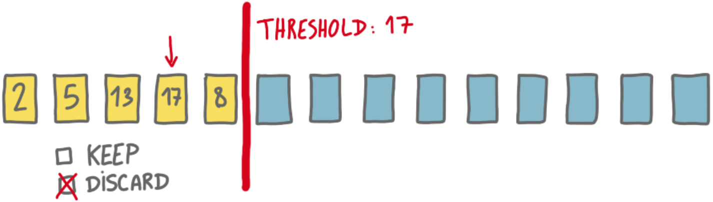
- Discard the first r values, but keep in mind the maximum M.
- Take the first value that is higher than M.



Stopping rule

A stopping rule is a strategy of the following form:

- Discard the first r values, but keep in mind the maximum M.
- Take the first value that is higher than M.



Stopping rule

A stopping rule is a strategy of the following form:

- Discard the first r values, but keep in mind the maximum M.
- Take the first value that is higher than M.

Stopping rule

Theorem: Optimal strategies are stopping rules.

Stopping rule

Theorem: Optimal strategies are stopping rules.

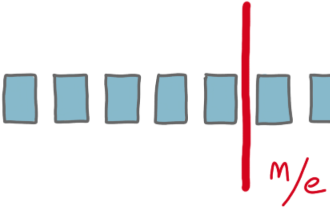
Intuitively:

- Taking the maximum so far is necessary.
- It is reasonable to wait to get some knowledge.
- One should not take the values into account to set of threshold.
- From the observation set, only the maximum is useful.

The optimal stop is n/e

Theorem: The optimal strategy is the stopping rule with $r = 1/e$.

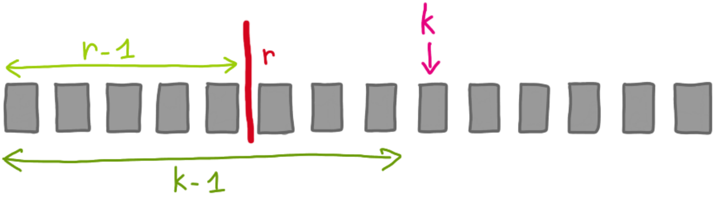
1. Discard the n/e first values but remember the maximum M in this segment.
2. Keep the first value that exceeds M.



The optimal stop is n/e

Theorem: The optimal strategy is the stopping rule with $r = 1/e$.

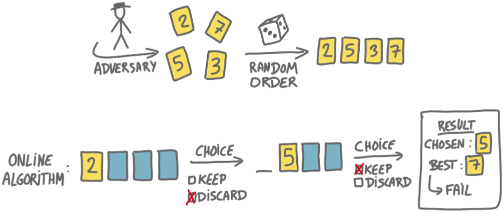
$$
\begin{aligned}
P(\text{win with } r) &= \sum_{k=r}^{n} P(k \text{ is the best}) P(k \text{ is chosen}) | k \text{ is the best}) \\
&= \sum_{k=r}^{n} \frac{1}{n} P(\text{best of } [1, k-1] \text{ appears before } r) \\
&= \sum_{k=r}^{n} \frac{1}{n} \frac{r-1}{k-1} = \frac{r-1}{n} \sum_{k=r}^{n} \frac{1}{k-1} \approx -x \log(x)
\end{aligned}
$$



A bit of history

- Kepler (1613), interviews for marriage
- Cayley (1875), open problem about lottery
- A lot of work in the 50s, 60s and 70s, in the statistics/probability community.
- "A secretary winter" (?)
- Now popular again, especially in TCS, because of its links with algorithmic game theory (auctions, posted prices mechanisms etc.)

Variants of the secretary problem



There are many parameters of the problem that can be changed:

- The order can be random, adversarial, chosen by the player.
- The number of items: player can chose several items, possibly with constraints (matroid)
- The objective: maximize expectation vs probability of picking the max.

Other problems: Prophet inequality, Pandora's box problem.

Question:

What happens if the algorithm
is not completely ignorant about the numbers?

Prior information

1. Distributional information:
   - i.i.d. random variables (Gilbert and Mosteller 66)
   - independent variables from known distributions (Esfandiari et al. 20, Allart and Islas 15)
2. Samples
   - Initiated by Azar, Kleinberg and Weinberg 14.
   - i.i.d variables from unknown distributions (Correa et al. 19+, Rubinstein et al. 20)
   - Known distributions, one sample from each (Kaplan et al. 20, Correa et al. 20).
   - A fraction of the values are sampled (Kaplan et al. 20)

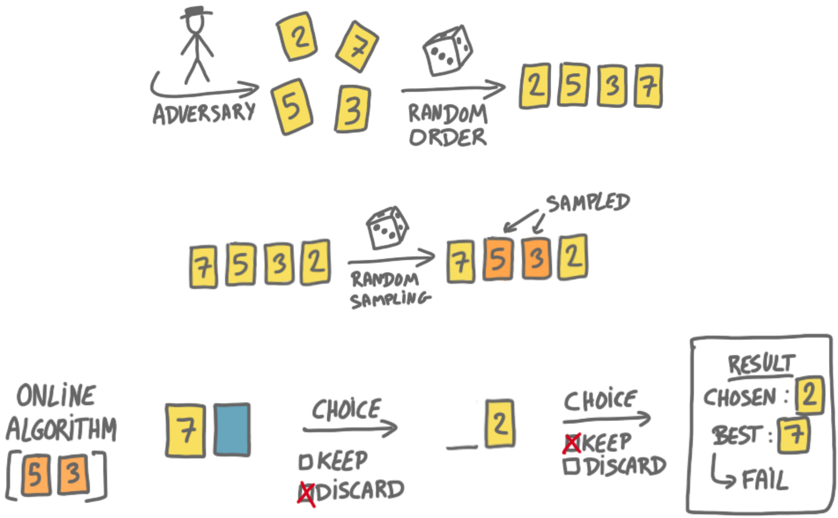
This paper

Problem: Secretary with independent sampling

The same as the classic secretary problem except that:

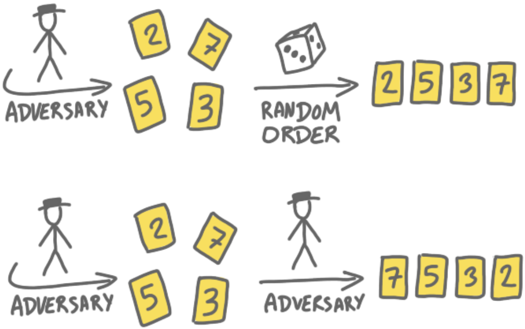
1. After the adversary's choice, every number is sampled with probability p.
2. The algorithm is given the sampled values, before the beginning.
3. The algorithm plays on the rest of the values.

Secretary with independent sampling



Adversarial and random order

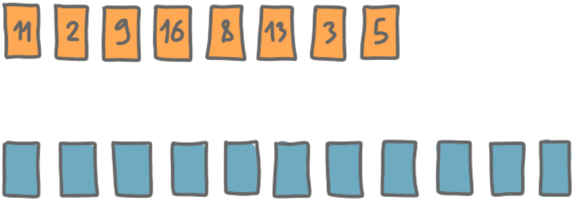
Actually the problem comes in two flavors: either the order is random (like in the classic version) or it is chosen by the adversary.



Heuristics: Max sample threshold

Let's try the following simple heuristics:

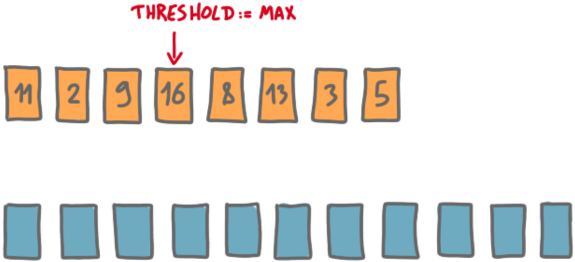
→ Take the maximum sampled value as a threshold.



Heuristics: Max sample threshold

Let's try the following simple heuristics:

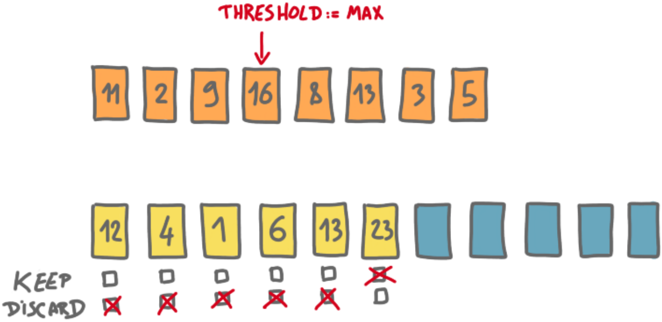
→ Take the maximum sampled value as a threshold.



Heuristics: Max sample threshold

Let's try the following simple heuristics:

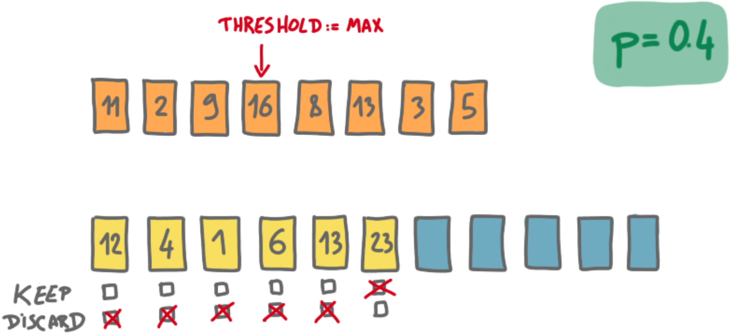
→ Take the maximum sampled value as a threshold.



Heuristics: Max sample threshold

Let's try the following simple heuristics:

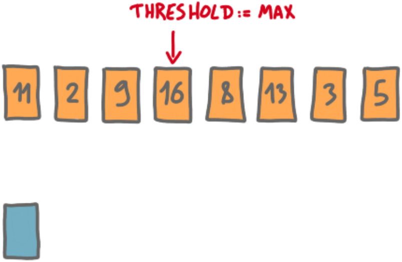
→ Take the maximum sampled value as a threshold.

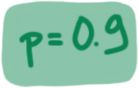


Heuristics: Max sample threshold

Let's try the following simple heuristics:

→ Take the maximum sampled value as a threshold.





Heuristics: Max sample threshold

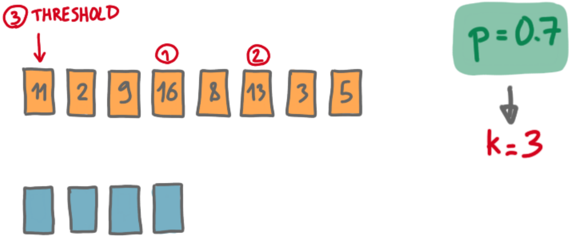
Let's try the following simple heuristics:

→ Take the maximum sampled value as a threshold.

k^{th} max sample strategy

$\rightarrow$ Choose the k^{th} max of the samples as a threshold, where k grows when $p \rightarrow 1$

The optimal $k(p)$ is easy to compute: $k(p) = \left\lfloor \frac{1}{1-p} \right\rfloor$.



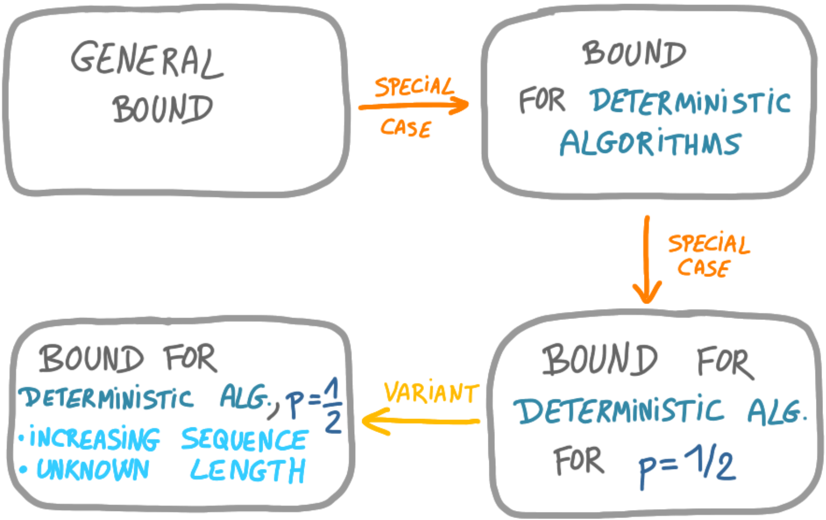
Adversarial order: k^{th}-max is optimal

Theorem:

For the adversarial order, the k^{th}-max strategy is optimal.

Proof shape: Prove that the performance of the algorithm matches an "upper" bound.

Proof roadmap



Special case: increasing values

Special case: increasing values

One restriction: no knowledge of the length.

Fact: This variant of the problem is equivalent to the last yellow card game.

Last yellow card game:

- There are n cards, with n unknown.
- Every card is colored blue with probability p and yellow otherwise.
- The player is given the number of blue cards.
- The cards are revealed one by one, and the player has to stop on the last yellow card.

Special case: increasing values

One restriction: no knowledge of the length.

Fact: This variant of the problem is equivalent to the last yellow card game.

Last yellow card game:

- There are n cards, with n unknown.
- Every card is colored blue with probability p and yellow otherwise.
- The player is given the number of blue cards.
- The cards are revealed one by one, and the player has to stop on the last yellow card.

Special case: increasing values

One restriction: no knowledge of the length.

Fact: This variant of the problem is equivalent to the last yellow card game.

Last yellow card game:

- There are n cards, with n unknown.
- Every card is colored blue with probability p and yellow otherwise.
- The player is given the number of blue cards.
- The cards are revealed one by one, and the player has to stop on the last yellow card.

Special case: increasing values

One restriction: no knowledge of the length.

Fact: This variant of the problem is equivalent to the last yellow card game.

Last yellow card game:

- There are n cards, with n unknown.
- Every card is colored blue with probability p and yellow otherwise.
- The player is given the number of blue cards.
- The cards are revealed one by one, and the player has to stop on the last yellow card.

Special case: increasing values

One restriction: no knowledge of the length.

Fact: This variant of the problem is equivalent to the last yellow card game.

Last yellow card game:

- There are n cards, with n unknown.
- Every card is colored blue with probability p and yellow otherwise.
- The player is given the number of blue cards.
- The cards are revealed one by one, and the player has to stop on the last yellow card.

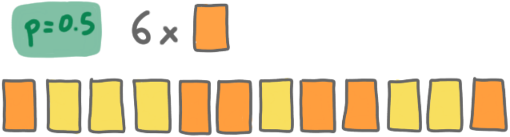
Special case: increasing values

One restriction: no knowledge of the length.

Fact: This variant of the problem is equivalent to the last yellow card game.

Last yellow card game:

- There are n cards, with n unknown.
- Every card is colored blue with probability p and yellow otherwise.
- The player is given the number of blue cards.
- The cards are revealed one by one, and the player has to stop on the last yellow card.



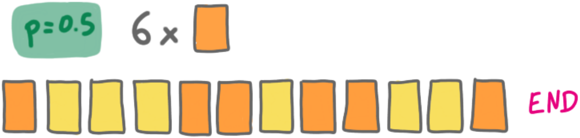
Special case: increasing values

One restriction: no knowledge of the length.

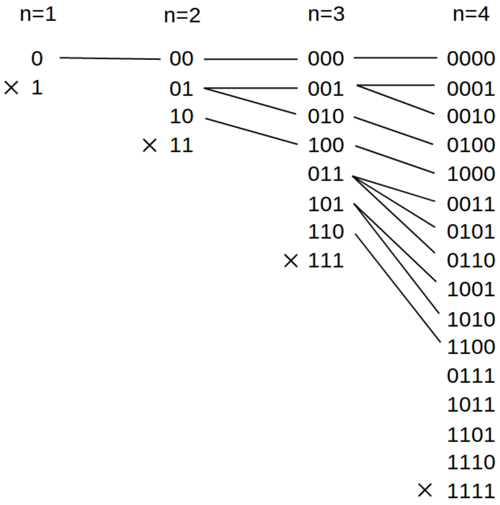
Fact: This variant of the problem is equivalent to the last yellow card game.

Last yellow card game:

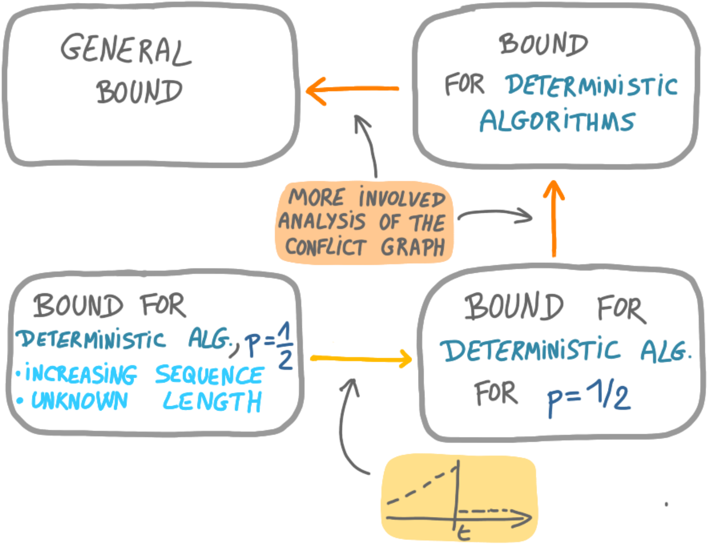
- There are n cards, with n unknown.
- Every card is colored blue with probability p and yellow otherwise.
- The player is given the number of blue cards.
- The cards are revealed one by one, and the player has to stop on the last yellow card.



Conflict graph



Proof roadmap



Adversarial order: k^{th}-max is optimal

Theorem:

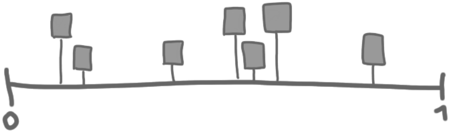
For the adversarial order, the k^{th}-max strategy is optimal.

Random order

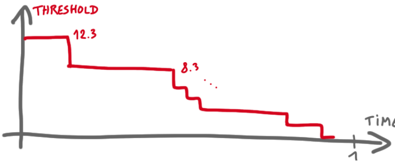
→ Again we can design an optimal algorithm, but via a completely different technique.

Proof technique

1. Reformulate as random arrivals.



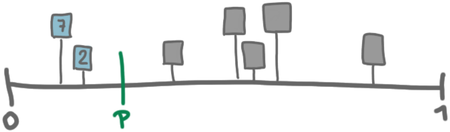
2. Any optimal strategy can be seen as a decreasing collection of thresholds.



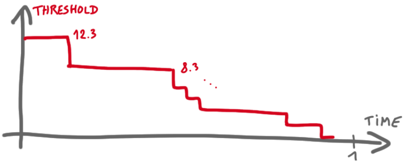
3. The optimal thresholds are the solutions to an easy separable concave optimization problem.

Proof technique

1. Reformulate as random arrivals.



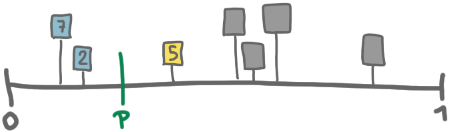
2. Any optimal strategy can be seen as a decreasing collection of thresholds.



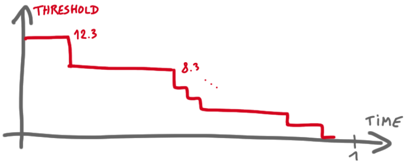
3. The optimal thresholds are the solutions to an easy separable concave optimization problem.

Proof technique

1. Reformulate as random arrivals.



2. Any optimal strategy can be seen as a decreasing collection of thresholds.



3. The optimal thresholds are the solutions to an easy separable concave optimization problem.

Thank you for watching!