# Prophet Inequalities

**Jose Correa**
Universidad de Chile

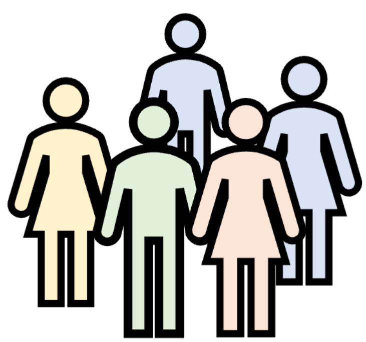
# Motivation

Online platforms, e-commerce, etc

Flexible Model:

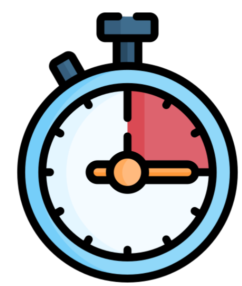
Multiple Goals

Incentives



**Limited data**

Sequential decisions



# Course Overview

1. Classic single-choice problems:

   The classic prophet inequality, secretary problem, prophet secretary problem, etc

2. **Data-driven prophet inequalities:**

   **How can limited amount of data be nearly as useful as full distributional knowledge**

3. Combinatorial Prophet Inequalities

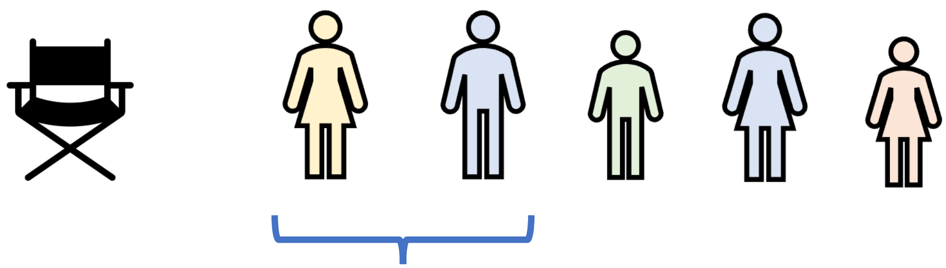
   Many ideas for single choice problems, extend to combinatorial contexts such as k-choice, Matching, hyper graph matching, and beyond

4. Online Combinatorial Auctions

   General Model that encompasses many online selection/allocation problems

# 2. Data-driven prophet inequalities

# Secretary Problem



Candidates come in random order

No values, only pairwise comparisons (there is a total order)

Decide STOP/CONTINUE

We maximize $\mathbb{P}(\text{select the best})$

Optimal algorithm

[Dynkin '63][Ferguson '89]

Skip $\frac{1}{e} \approx 0.367$ fraction of candidates
Then, STOP if best so far

Succeeds w.p. $1/e$

Optimal guarantee and algorithm are the same if

Candidates have i.i.d. values and we maximize $\mathbb{E}(\text{selected candidate})$ (v.s. $\mathbb{E}(\text{best})$ )

[C., Dütting, Fischer, Schewior, EC'19]

# Prophet Inequality

1 item

We are given distributions $F_1, \ldots, F_n$

Agents arrive one by one

At step i: observe $v_i \sim F_i$ (indep.) and decide STOP/CONTINUE

We maximize $\mathbb{E}(v_{\text{stop}})$ and compare with $\mathbb{E}\left(\max_i v_i\right)$

Optimal guarantee: 1/2
[Krengel & Sucheston '77]

Set a threshold (a price)
$T = \frac{1}{2}\mathbb{E}\left(\max_i v_i\right)$
[Kleinberg, Weinberg STOC'12]

If valuations are i.i.d.

**Optimal guarantee is $\approx \mathbf{0.745}$**
[C., Foncea, Hoeksma, Oosterwijk, Vredeveld EC 2017]
Decreasing sequence of thresholds

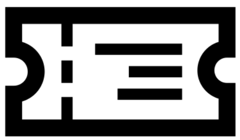
If we maximize $\mathbb{P}(\text{select the best})$

Optimal guarantee is $1/e$
[Allaart & Islas '16]
**i.i.d. case: $\approx \mathbf{0.5801}$**
[Gilbert & Mosteller '66]
Random order: $\approx 0.5801$
[Nuti, IPCO'22]

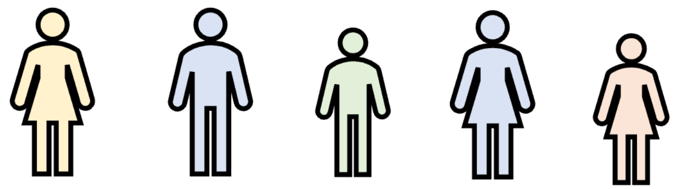
April 2022

April 2023

# The prophet inequality



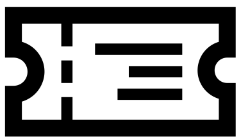
1 item



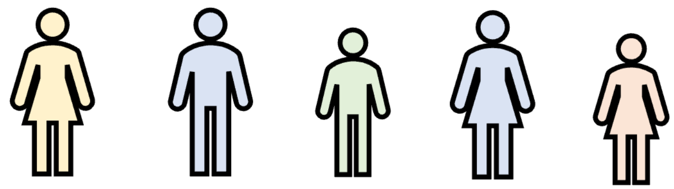
$n$ agents
independent valuations
$v_i \sim F_i$

- ~~We are given $F_1, \ldots, F_n$~~ We are given samples from $F_1, \ldots, F_n$
- Agents arrive sequentially: we observe $v_1 \sim F_1, v_2 \sim F_2, \ldots$ one by one
- We (the Decision Maker) decide stop/continue
- We maximize $\mathbb{E}(v_{\text{stop}})$
- Compare against a prophet that can see realizations in advance and thus gets the optimal social welfare $\mathbb{E}\left(\max_i v_i\right)$

# The prophet inequality



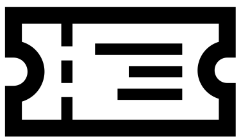
1 item



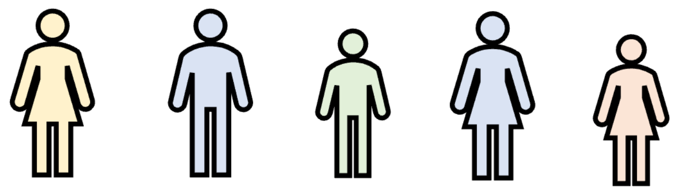
$n$ agents
independent valuations
$v_i \sim F_i$

- With full distributional knowledge we know that $\mathbb{E}(v_{\text{stop}}) \geq \frac{1}{2}\,\mathbb{E}\left(\max_i v_i\right)$
- What if we are only given one sample $s_1, \dots, s_n$ from each $F_1, \dots, F_n$?
- First simple observation: Set $T = \max\{s_1, \dots, s_n\}$, scan the $v_i$'s and stop with first value above $T$. [Azar, Kleinberg, Weinberg SODA 2014]
- $\mathbb{P}$ (max all 2n values is on the $v_i$'s and the second max is on the $s_i$'s ) $\geq \frac{1}{4}$
- Then $\mathbb{E}(v_{\text{stop}}) \geq \frac{1}{4}\,\mathbb{E}\left(\max_i v_i\right)$ Can we do better?

# The prophet inequality



1 item



$n$ agents
independent valuations
$v_i \sim F_i$

- YES! Algorithm is actually ½ competitive [Rubinstein, Wang, Weinberg ITCS 2020]

Amazing! One sample is enough to get the optimal prophet inequality.

- And this gives yet another Algorithm: Take a random threshold $T$ distributed as the $\max\{v_1, \ldots, v_n\}$ so $F_T = \prod F_i$

# Proof

- Take $n$ pairs of arbitrary nonnegative numbers, say $(s_1, v_1), \dots, (s_n, v_n)$
- Call the ordered sequence $a_1 \geq a_2 \geq \cdots \geq a_{2n}$.
- Randomly shuffle each pair assigning each element to L and R w.p. ½

| $L$ | $R$ |
| --- | --- |
| $s_1$ | $v_1$ |

| $L$ | $R$ |
| --- | --- |
| $v_1$ | $s_1$ |

# Proof

| $L$ | $R$ |
|---|---|
| $s_1 \; v_2 \; v_3 \; s_4 \cdots$ | $v_1 \; s_2 \; s_3 \; v_4 \; \cdots$ |

- Call the ordered sequence $a_1 \geq a_2 \geq \cdots \geq a_{2n}$.
- Run ALG on the resulting instance: $T$ = max value in L.

  Stop whenever a value in $R$ surpasses $T$.

  Actually take the weaker algorithm that if $T = a_i$ then it gets $a_{i-1}$
  (except that, if $T = a_1$, it gets 0)

$$\mathbb{P}(OPT = a_i) = \mathbb{P}(\max \text{ in } R = a_i) \approx \frac{1}{2^{i-1}} \times \frac{1}{2}$$

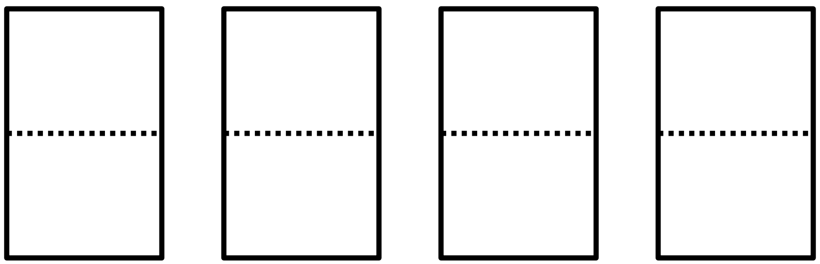
$$\mathbb{P}(ALG = a_i) = \mathbb{P}(T = a_{i+1}) \approx \frac{1}{2^i} \times \frac{1}{2} = \frac{1}{2} \mathbb{P}(OPT = a_i)$$

# Prophet Secretary: The two-sided googol

- The secretary problem is also known as the game of googol.
- An adversary writes arbitrary numbers on $n$ cards and shuffles them.
- The DM flips the cards one by one and has to stop with the max.

# The two-sided game of Googol

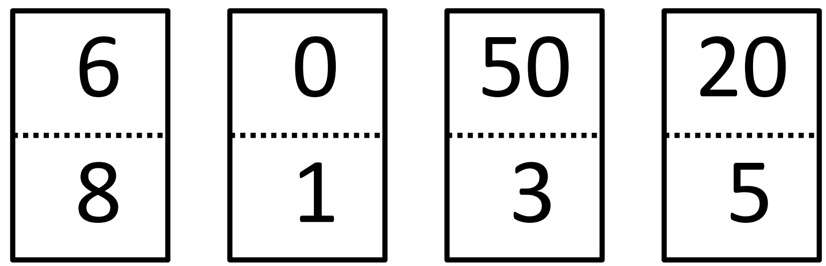
*n* cards with two sides



# The two-sided game of Googol

*n* cards with two sides

Adversary writes numbers on each side (2*n* in total)



# Playing the two-sided game of Googol

Random side revealed

# Playing the two-sided game of Googol

Random side revealed

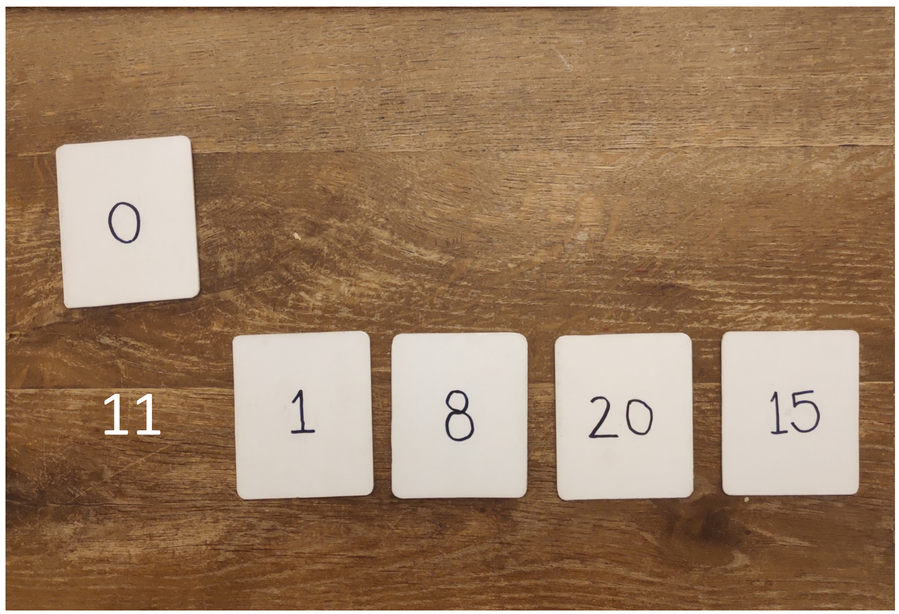
# Playing the two-sided game of Googol

Random side revealed

# Playing the two-sided game of Googol

Random side revealed

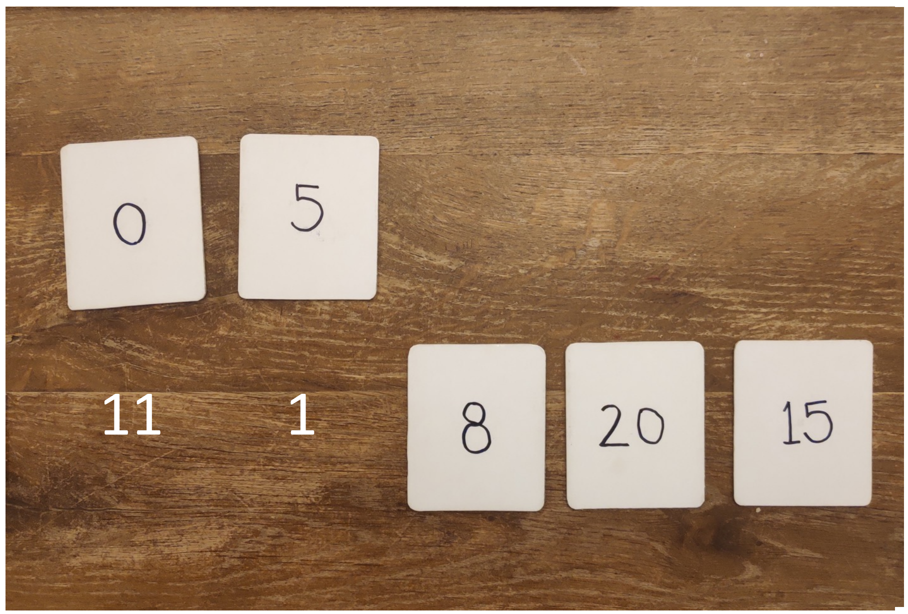
Flip a random card, **ACCEPT** or **CONTINUE**



# Playing the two-sided game of Googol

Random side revealed

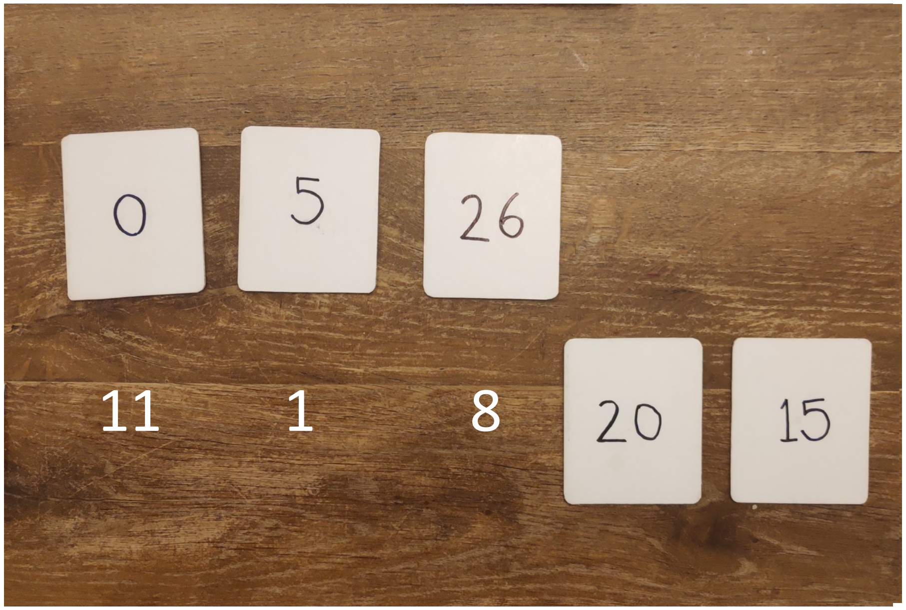
Flip a random card, **ACCEPT** or **CONTINUE**



# Playing the two-sided game of Googol

Random side revealed

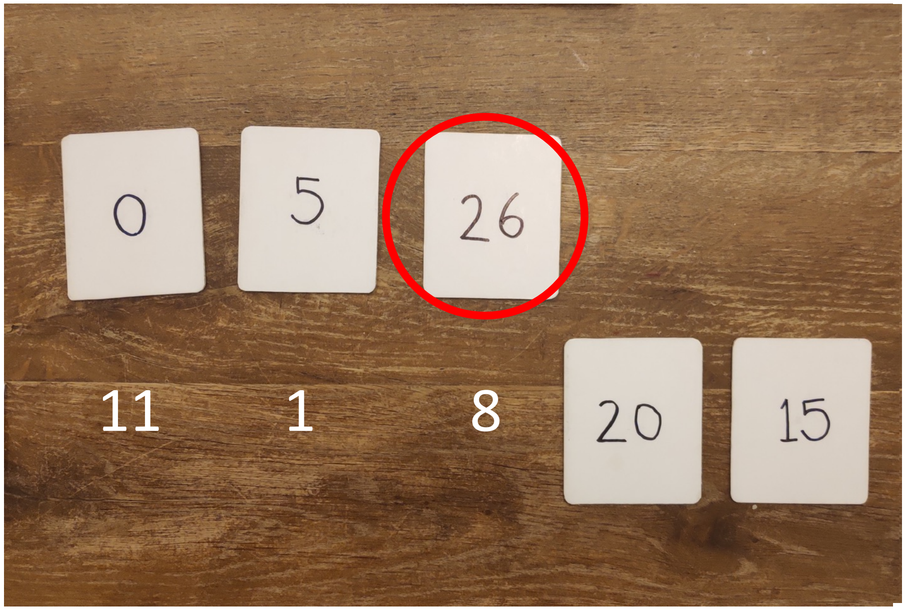
Flip a random card, **ACCEPT** or **CONTINUE**



# Playing the two-sided game of Googol

Random side revealed

Flip a random card, **ACCEPT** or **CONTINUE**




Objective: Maximize expectation of accepted value

Benchmark: Expectation of largest hidden value

# Main results

**Theorem.** There is a stopping rule $ALG^*$ for the two-sided game of Googol such that

$$\mathbb{E}(ALG^*) \geq 0.635 \cdot \mathbb{E}(OPT)$$

[C., Cristi, Epstein, Soto SODA 2020]

Interesting since it took quite some effort to beat $1 - 1/e$ for PS.

Alternative version: Max probab. of taking the maximum hidden value.

**Theorem.** There is a stopping rule $ALG^*$ for the two-sided game of Googol such that

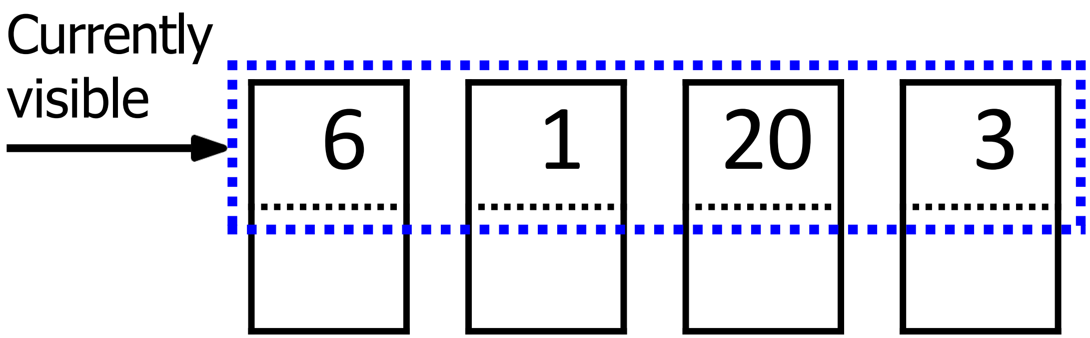
$$\mathbb{P}(ALG^* \text{ wins}) \geq 0.5001$$

And this is almost tight

[Nuti, Vondrak SODA 2023]

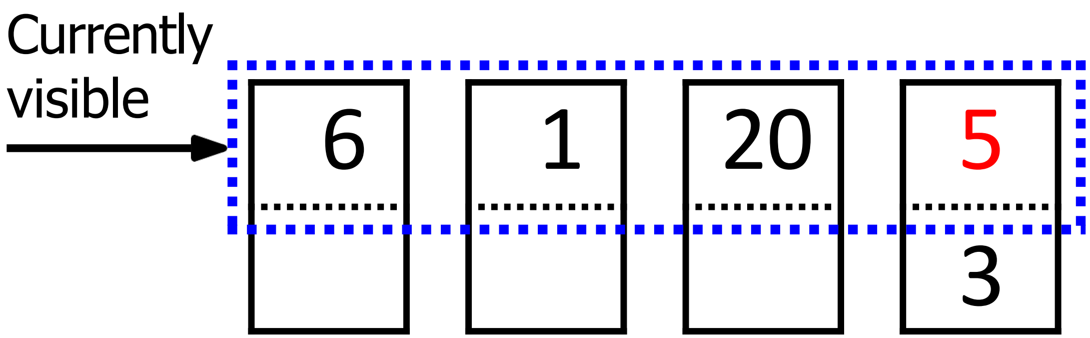
# Basic algorithm 1: Open window

Stop in first value that is maximum among all currently visible values



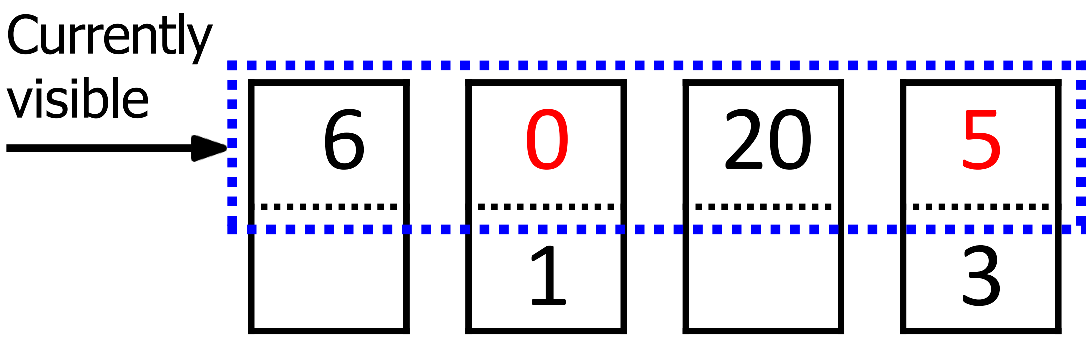
# Basic algorithm 1: Open window

Stop in first value that is maximum among all currently visible values



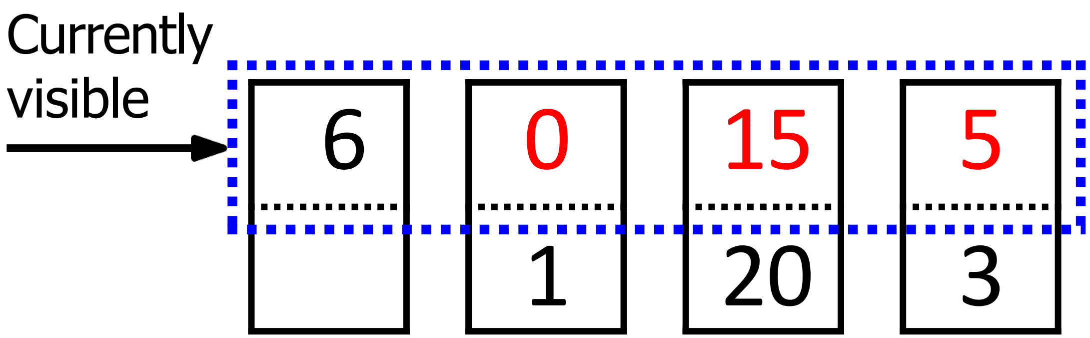
# Basic algorithm 1: Open window

Stop in first value that is maximum among all currently visible values



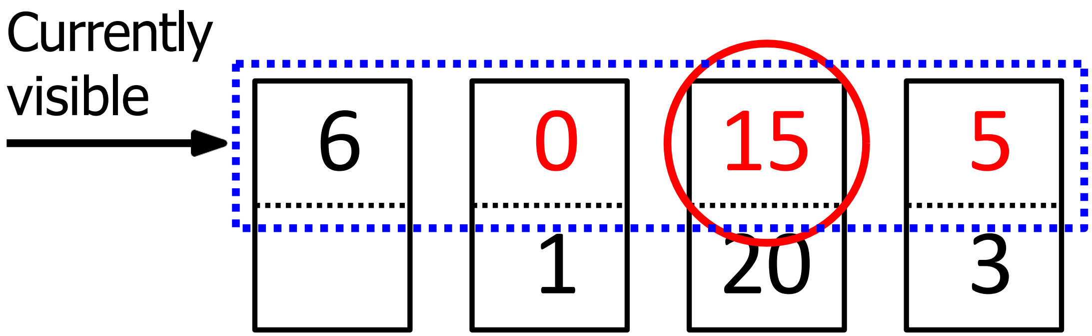
# Basic algorithm 1: Open window

Stop in first value that is maximum among all currently visible values



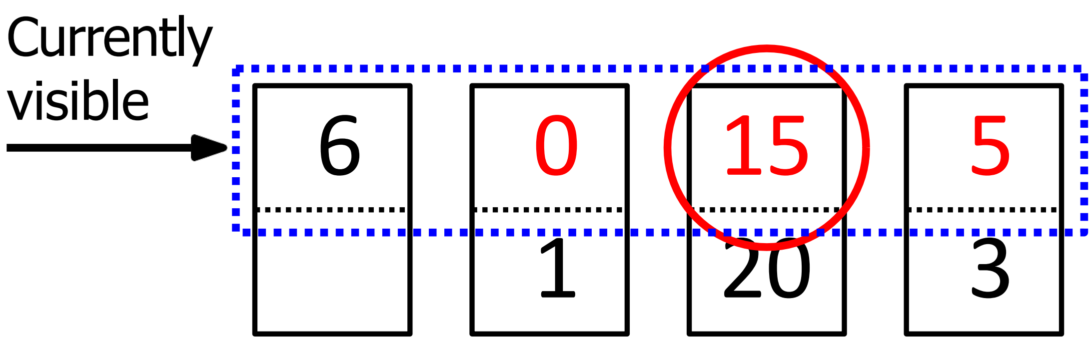
# Basic algorithm 1: Open window

Stop in first value that is maximum among all currently visible values

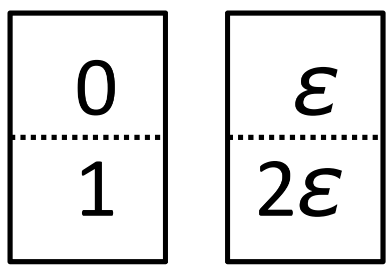


# Basic algorithm 1: Open window

Stop in first value that is maximum among all currently visible values



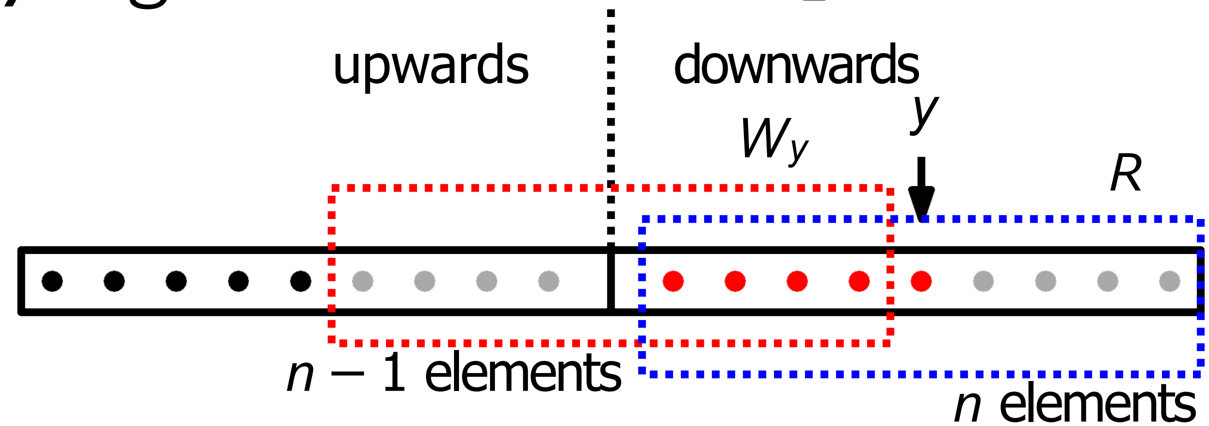

Bad instance:




$\mathrm{E}(OPT) = 1/2 + O(\varepsilon)$

$\mathrm{E}(ALG_1) = 1/4 + O(\varepsilon)$

# 1/2 guarantee for $\text{ALG}_1$




$$\mathbb{P}(OPT = y | R = S \dot\cup \{y\}) \leq 2 \cdot \mathbb{P}(ALG_1 = y | y \in R, W_y = S)$$

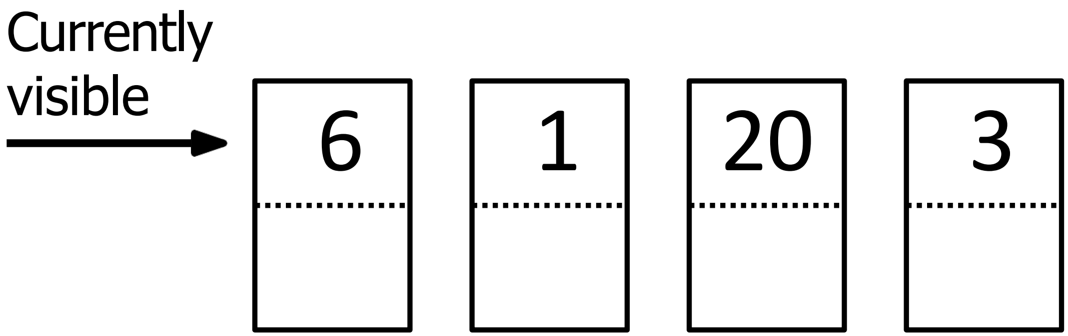
Indicator function

- 1 if $y > \max S$
  0 otherwise

w.p. 1/2, max $S$ lands facing upwards (left)

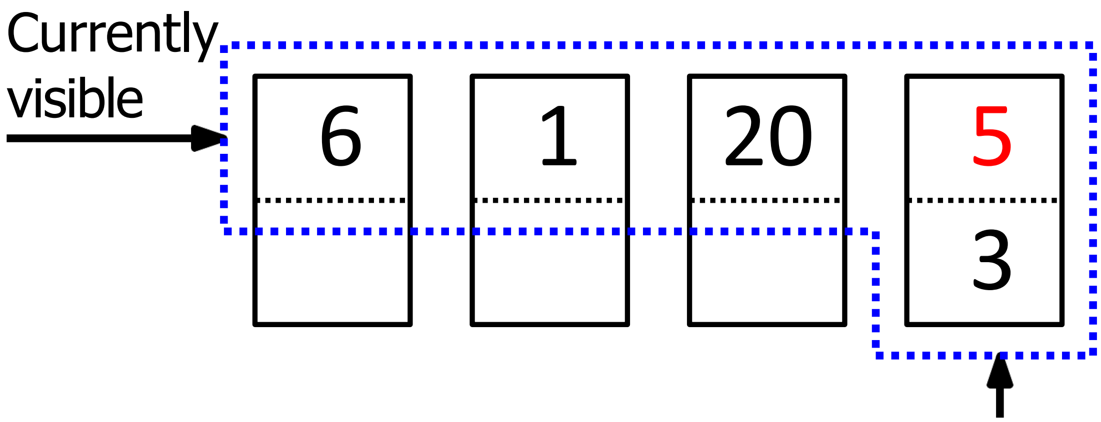
# Basic algorithm 2: Closed window

Stop in first value that is maximum among all currently visible values **& other side of card**



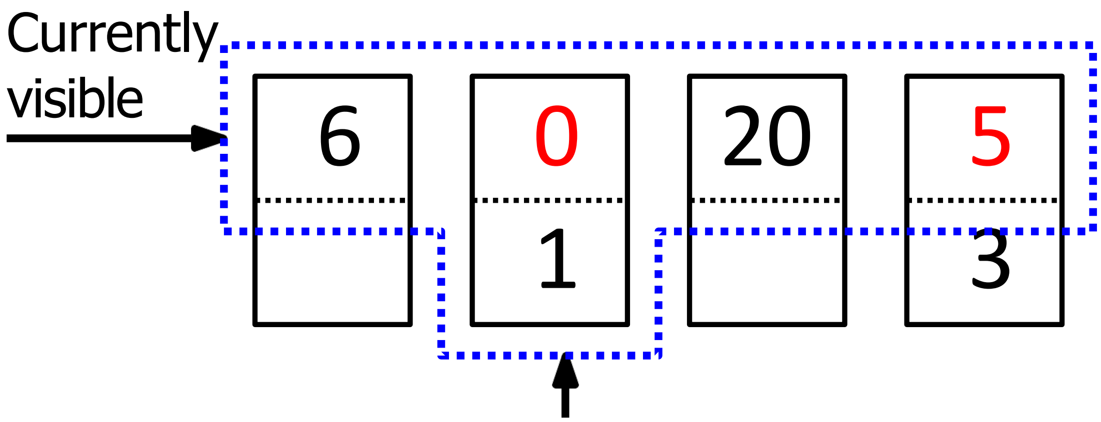
# Basic algorithm 2: Closed window

Stop in first value that is maximum among all currently visible values **& other side of card**



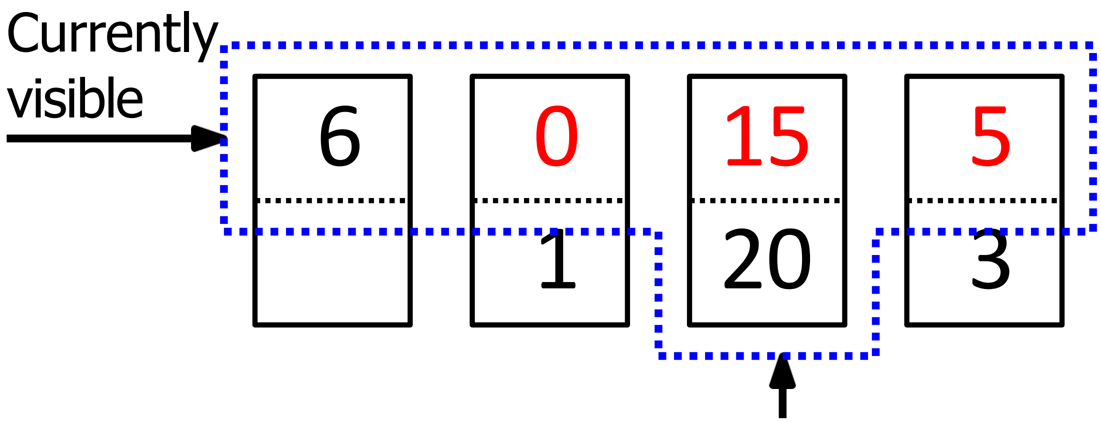
# Basic algorithm 2: Closed window

Stop in first value that is maximum among all currently visible values **& other side of card**



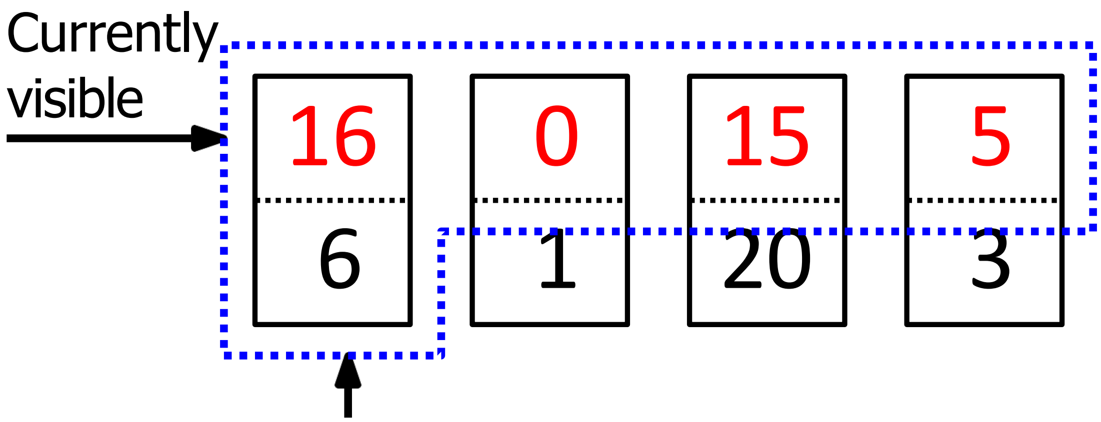
# Basic algorithm 2: Closed window

Stop in first value that is maximum among all currently visible values **& other side of card**



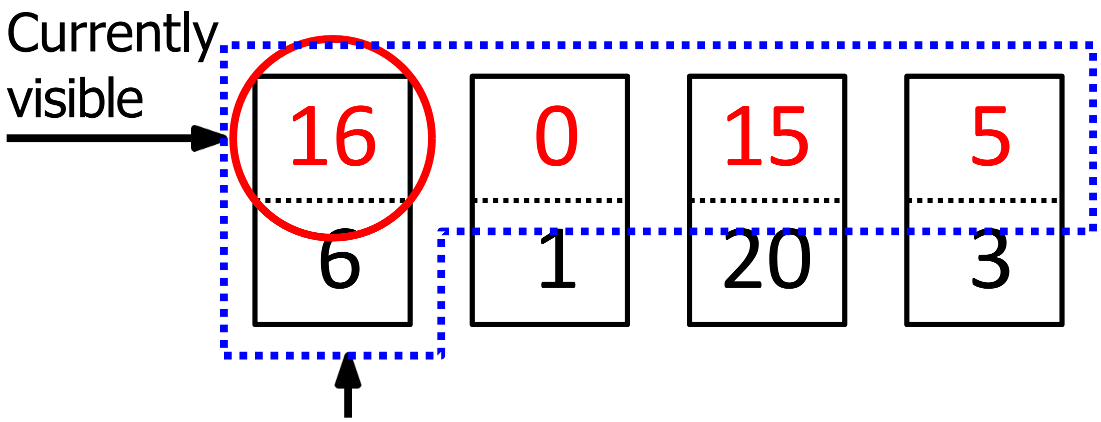
# Basic algorithm 2: Closed window

Stop in first value that is maximum among all currently visible values **& other side of card**



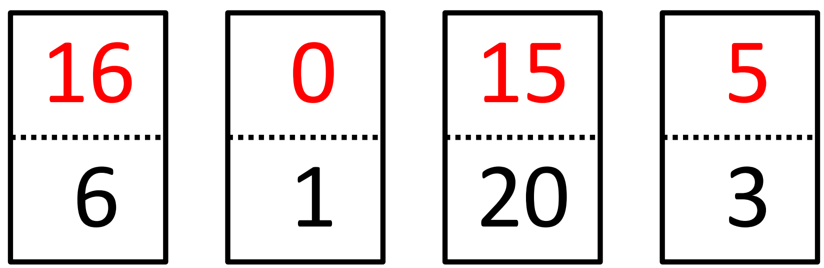
# Basic algorithm 2: Closed window

Stop in first value that is maximum among all currently visible values **& other side of card**



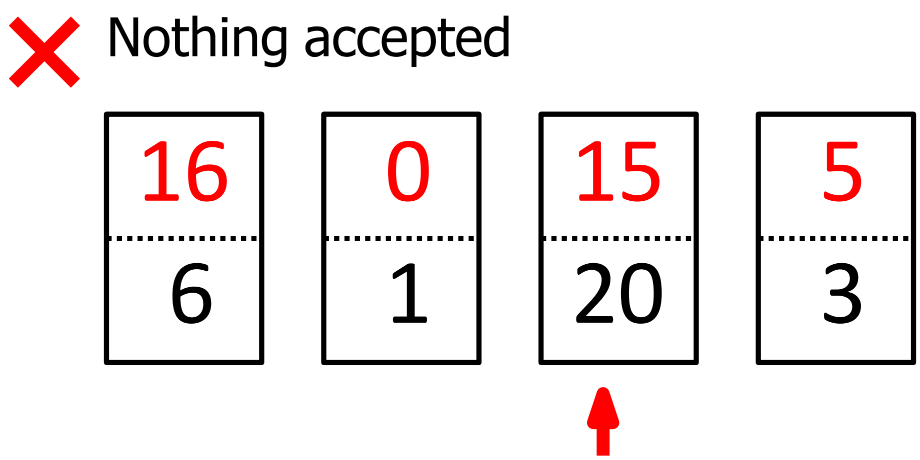
# Basic algorithm 3: Full window

Stop in first value larger than all values seen so far

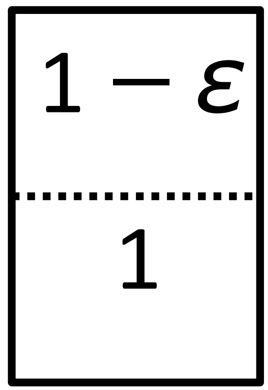


# Basic algorithm 3: Full window

Stop in first value larger than all values seen so far



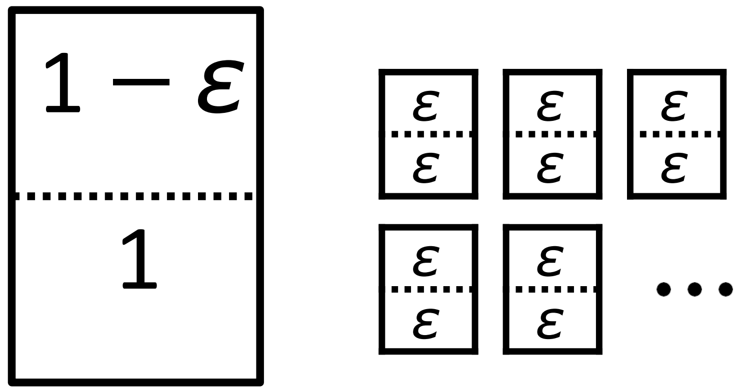
# Bad instance for $ALG_2$ and $ALG_3$




$\mathrm{E}(OPT) = 1 - O(\varepsilon)$

If ALG sets other side of card as threshold, then $\mathrm{E}(ALG) \leq 1/2 + O(\varepsilon)$

# Bad instance for $ALG_2$ and $ALG_3$




$\mathrm{E}(OPT) = 1 - O(\varepsilon)$

If ALG sets other side of card as threshold, then $\mathrm{E}(ALG) \leq 1/2 + O(\varepsilon)$

# Combined algorithm

- $ALG^*$: run $ALG_1$ w.p. $\alpha$, run $ALG_2$ w.p. $\beta$ and run $ALG_3$ w.p. $1-\alpha-\beta$

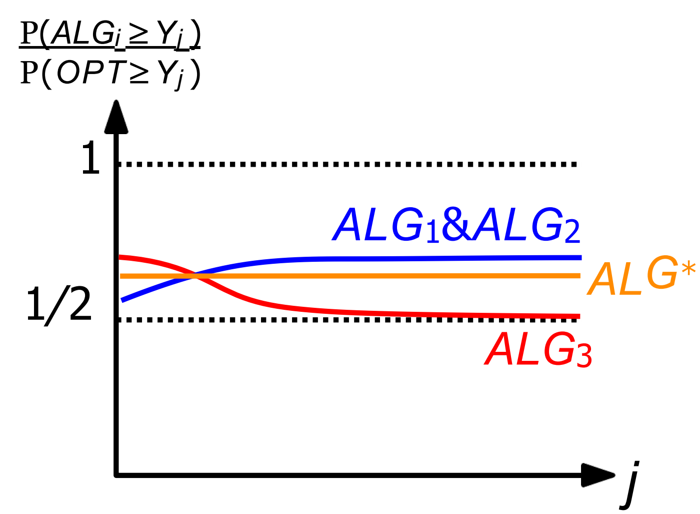
$$\frac{\mathrm{P}(ALG \geq x)}{\mathrm{P}(OPT \geq x)} \geq c,\ \forall x \geq 0 \Rightarrow \frac{\mathrm{E}(ALG)}{\mathrm{E}(OPT)} \geq c$$

Denote numbers in cards $Y_1 > Y_2 > \cdots > Y_{2n}$

We can compute $\mathrm{P}(OPT \geq Y_j), \forall j$

$\mathrm{P}(ALG_i \geq Y_j), \forall j, i$

+ Extra parameter $k$ (technical)



# Two-sided googol

**Prophet-secretary.** $n$ independent realizations of known distributions $F_1,\ldots,F_n$ arrive sequentially in random order. Decide when to stop in order to maximize expected value.

**Data-driven version:** Distributions are unknown. Access only to one independent sample of each on beforehand.

If adversary draws numbers from distributions $F_1,\ldots,F_n$ (two of each), we obtain prophet-secretary with samples

# Two-sided googol

- Implies a factor 0.635 for prophet secretary

  → Improves upon previous $1 - \frac{1}{e} + \frac{1}{400} \approx 0.634$ which took effort

  [Azar, Chiplunkar, Kaplan EC 2018]

  → Different sampling idea [Kaplan, Naori, Raz SODA 2020]

  → Best known for PS is 0.669 [C., Saona, Ziliotto, SODA 2019]

- Many open questions

  →What is the best algorithm?

  →What happens in two-sided googol if we can choose the order of observation?

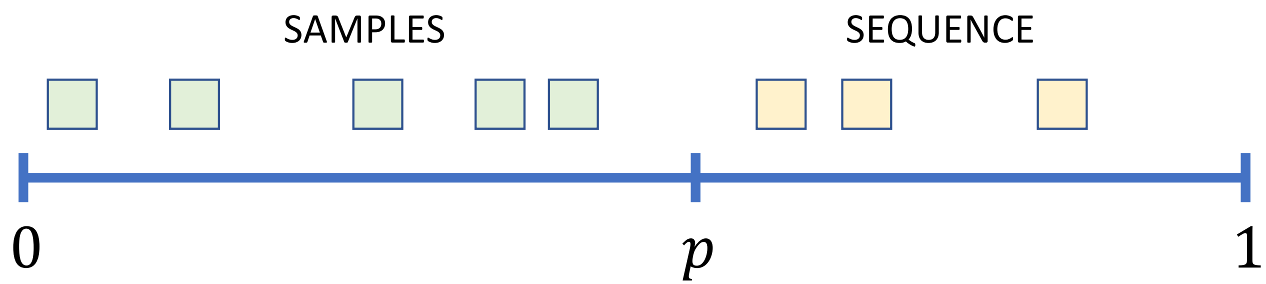
  →What about k-sided? Can we obtain the best possible algorithm for PS this way?

# Independent sampling model

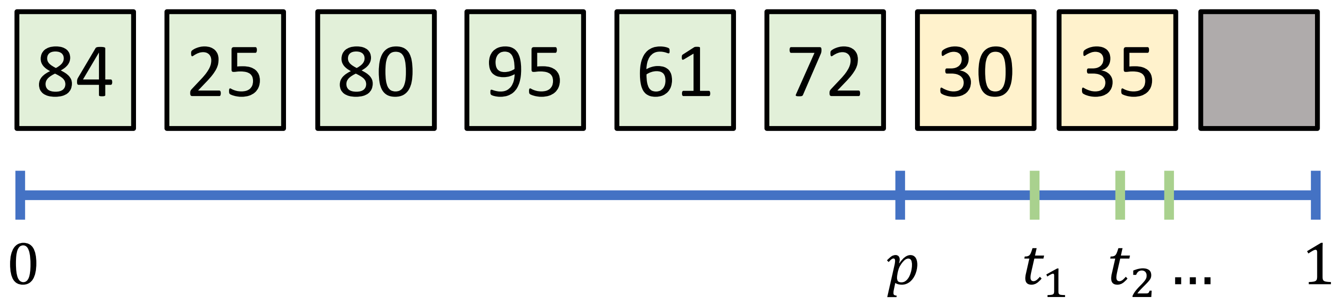
- Set of N **unknown** values is fixed, a probability $p \in [0,1)$ is given
- Each value is in SAMPLES with probability $p$, independently. Otherwise, in the SEQUENCE.
- We observe the SAMPLES and then, one by one, the values in the SEQUENCE, in uniform random order.

# Uniform[0,1] arrivals

- Values arrive at an independent $\text{Uniform}[0,1]$ time
- We start playing at time $p$



# Maximize $\mathbb{P}(\text{select the best})$




W.l.o.g. we can look at ordinal algorithms
[Moran, Snir, Manber '85]

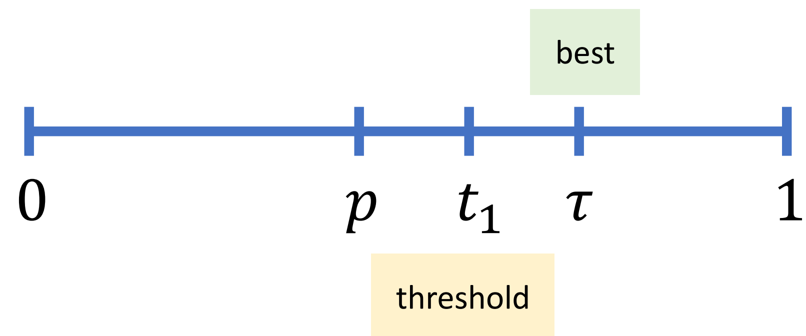
For an ordinal algorithm: the probability that a best-so-far is the best depends only in its **overall rank** and **how many elements are left**.

A **time-thresholds** algorithm achieves the best-possible guarantee.

# Example: single time-threshold

$$\int_{\max\{p,t_1\}}^{1} \frac{\max\{p, t_1\}}{\tau}\, d\tau$$

$$= -\max\{p, t_1\} \cdot \ln(\max\{p, t_1\})$$



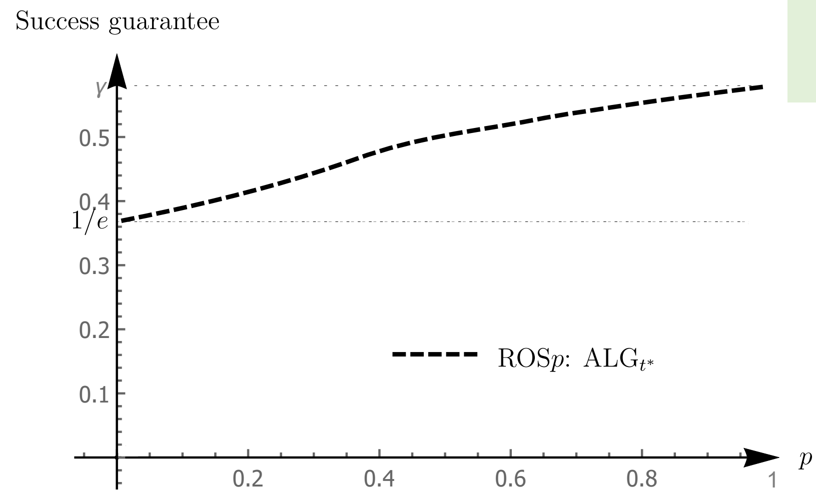

Concave, maximized at $t_1 = 1/e$ for all $p$

If we have a time threshold for each rank,

$$\sum_{i=1}^{\infty} p^{i-1} \cdot \left(1 - \max\{p, t_i\} - \int_{\max\{p,t_i\}}^{1} \sum_{j=1}^{i} \frac{\tau - \max\{p, t_i\}}{\tau^j}\, d\tau\right)$$

Turns out to be separable and concave!

| | |
|---|---|
| $t_1^* \approx 0.3678794$ | $t_6^* \approx 0.8709762$ |
| $t_2^* \approx 0.6422006$ | $t_7^* \approx 0.8887973$ |
| $t_3^* \approx 0.7518116$ | $t_8^* \approx 0.9022956$ |
| $t_4^* \approx 0.8101810$ | $t_9^* \approx 0.9128731$ |
| $t_5^* \approx 0.8463645$ | $t_{10}^* \approx 0.9213851$ |



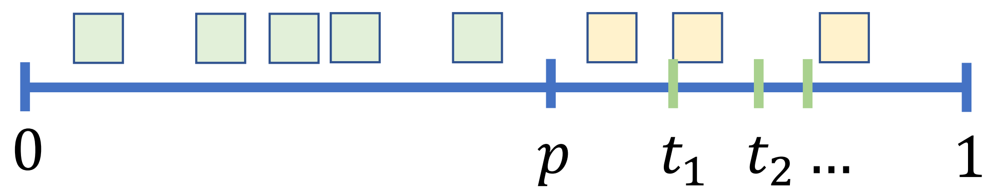

Gilbert & Mosteller's $\approx 0.5801$

[C., Cristi, Feuilloley, Oosterwijk, Tsiagonias-Dimitradis SODA 2021]

# Maximize $\mathbb{E}(v_{\text{stop}})$

Again we consider only ordinal algorithms (not necessarily w.l.og.)

We first assume numbers $Y_1, \dots, Y_N$ **are known**




**Linear Program**

$$\max_{ALG} \sum_{j=1}^{N} Y_j \cdot \mathbb{P}\big(ALG(Y) \text{ selects } Y_j\big)$$
$$\text{s.t.} \quad ALG \text{ is feasible}$$

**Limit Problem**

$$\sup_{t=(t_i)_{i\in\mathbb{N}}} \; Y_1 - \sum_{k\geq 1} (Y_k - Y_{k+1})\big(1 - F_k(t)\big)$$
$$\text{s.t.} \qquad p \leq t_i \leq t_{i+1} \leq 1 \quad \forall i \geq 1$$

Where $F_k(t) = \mathbb{P}\big(ALG_t(Y) \geq Y_j\big)$

$$\sup_{t=(t_i)_{i\in\mathbb{N}}} \inf_{Y_1\geq Y_2\geq\cdots} \frac{Y_1 - \sum_{k\geq 1}(Y_k - Y_{k+1})\big(1-F_k(t)\big)}{\mathbb{E}\big(OPT(Y)\big)}$$

s.t. $\quad p \leq t_i \leq t_{i+1} \leq 1 \quad \forall i \geq 1$

Where $F_k(t) = \mathbb{P}\big(ALG_t(Y) \geq Y_j\big)$

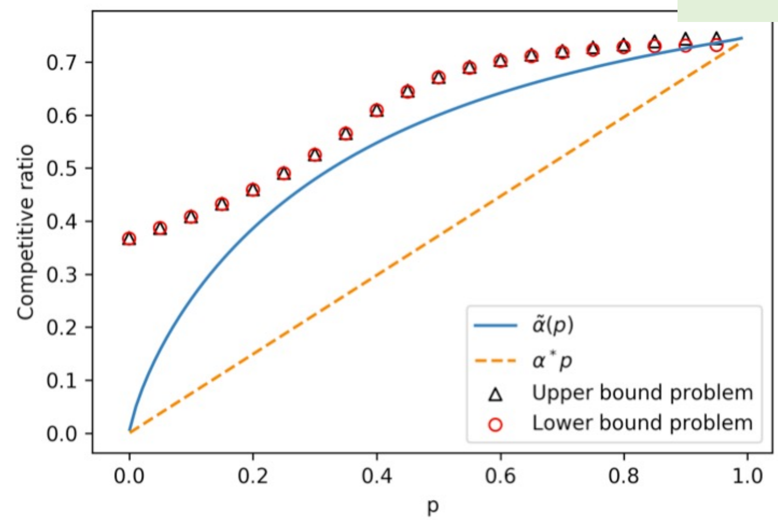
$$\sup_{t=(t_i)_{i\in\mathbb{N}}} \inf_{k\geq 1} \frac{F_k(t)}{1-p^k}$$

s.t. $\quad p \leq t_i \leq t_{i+1} \leq 1 \quad \forall i \geq 1$

Where $F_k(t) = \mathbb{P}\big(ALG_t(Y) \geq Y_j\big)$

[C., Cristi, Epstein Soto MOR 2023]

Best for i.i.d. $\approx 0.745$




$O\left(\frac{N}{\varepsilon}\right)$ samples are sufficient to get almost optimal guarantee of $0.745 - \varepsilon$

Improves upon previous bound of $O\left(\frac{N}{\varepsilon^6}\right)$

[Rubinstein, Wang, Weinberg, ITCS'20]

# Summary

- 1/2 PI w. one sample per distribution [Rubinstein, Wang, Weinberg ITCS 2020]
- Two-sided googol (prophet secretary with single sample)
  - 0.635 for expectation [Correa, Cristi, Epstein, Soto SODA 2020]
  - 0.5001 for probability of selecting the best [Nuti, Vondrak SODA 2023]
- Best i.i.d. PI with any number of samples [Correa, Cristi, Epstein, Soto MOR 2023]