Approximating Pandora's Box with Correlations

Shuchi Chawla[1], ***Evangelia Gergatsouli***[2], Jeremy McMahan[2], Christos Tzamos[2,3]

[1]University of Texas-Austin, [2]University of Wisconsin-Madison, [3]University of Athens

APPROX '23, Atlanta GA, September 2023

A Search Problem

Find the best alternative, with costly information!

$X_1 \sim \mathcal{D}_1$ $X_2 \sim \mathcal{D}_2$ $X_3 \sim \mathcal{D}_3$ $X_4 \sim \mathcal{D}_4$ $X_5 \sim \mathcal{D}_5$

- Information is not free!
- Explore alternatives (open boxes)
- Stop anytime & take best so far

} **Strategy**

A Search Problem

Find the best alternative, with costly information!

42 | $X_2 \sim \mathcal{D}_2$ | 15 | 12 | $X_5 \sim \mathcal{D}_5$

Opening cost: 3
Final option: box 4
Total cost: 12 + 3

- Information is not free!
- Explore alternatives (open boxes)
- Stop anytime & take best so far

} **Strategy**

Goal: find min cost strategy!

Previous Work

Weitzman's algorithm gives the optimal! [Weitz 1979]

- Calculate *reservation value* σ_i for every box[1]
- Search boxes in order of increasing index until:
  - Current min price seen smaller than index of next box

[1](Gittins index)

Previous Work

Weitzman's algorithm gives the optimal! [Weitz 1979]

- Calculate *reservation value* σ_i for every box[1]
- Search boxes in order of increasing index until:
  - Current min price seen smaller than index of next box

Partially Adaptive:
fix order from beginning

[1] (Gittins index)

Previous Work

Weitzman's algorithm gives the optimal! [Weitz 1979]

- Calculate *reservation value* σ_i for every box[1]
- Search boxes in order of increasing index until:
  - Current min price seen smaller than index of next box

Crucial assumption: distributions are **independent**!

What about **correlation?**

[1] (Gittins index)

Previous Work

Weitzman's algorithm gives the optimal! [Weitz 1979]

- Calculate *reservation value* σ_i for every box[1]
- Search boxes in order of increasing index until:
  - Current min price seen smaller than index of next box

Crucial assumption: distributions are **independent**!

What about **correlation?**

Our work: Algorithms for Pandora's Box with Correlations

[1] (Gittins index)

Pandora's Box with Correlations: Setting

Given:

- m scenarios with probabilities p_i
- n boxes with cost 1
- Matrix of values as below
- Pay total opening cost + value chosen per scenario

Goal: Open boxes, stop & pick value
Minimize: Sum of
opening cost + value chosen

	Box 1	**Box 2**	**Box 3**	...	**Box n**
$p_1 \to$ **Scenario 1:**	42	13	15		24
$p_2 \to$ **Scenario 2:**	0	24	94	...	2
⋮			⋮		
$p_m \to$ **Scenario m:**	31	15	9	...	2

Previous Work - Overview

Independent Disributions

- Weitzman's algorithm is the optimal [Weitzman: Econometrica '79]

Correlated Distributions

- 9.22-approx against the PA optimal [Chawla et al.: FOCS '20]
  - LP-based algorithm via reduction to Min Sum Set Cover
  - Learnable from samples
- 4.22-approx against the PA optimal [Gergatsouli, Tzamos: arXiv '23]
  - Extends Weitzman's algorithm
  - Learnable from samples

Previous Work - Overview

Independent Disributions

- Weitzman's algorithm is the optimal [Weitzman: Econometrica '79]

Correlated Distributions

- 9.22-approx against the PA optimal [Chawla et al.: FOCS '20]
  - LP-based algorithm via reduction to Min Sum Set Cover
  - Learnable from samples
- 4.22-approx against the PA optimal [Gergatsouli, Tzamos: arXiv '23]
  - Extends Weitzman's algorithm
  - Learnable from samples

What about the **Fully Adaptive**?

Our Results

Question: How hard is PB with correlations against the Fully Adaptive?

Main Result: Pandora's Box **equivalent** to Uniform Decision Tree

Implications for PB:

- $\tilde{O}(\log m)$-approx poly-time algorithm
- $\tilde{O}(1/\alpha)$-approx running in time $n^{\tilde{O}(m^{\alpha})}$ for any $\alpha \in (0,1)$.
- It is **not** NP-hard to get superconstant approx assuming ETH.

Implications by [Li et al. SODA '20]

Uniform Decision Tree

Given:

- m scenarios with probabilities $1/m$
- n tests with cost 1
- Matrix of results as below
- Pay total tests cost per scenario

Goal: Select tests to identify scenario
Minimize: Sum of test costs

	Test 1	**Test 2**	**Test 3**	...	**Test n**
$1/m \to$ **Scenario** 1:	42	13	15		24
$1/m \to$ **Scenario** 2:	0	24	94	...	2
⋮			⋮		
$1/m \to$ **Scenario** m:	9	4	34	...	131

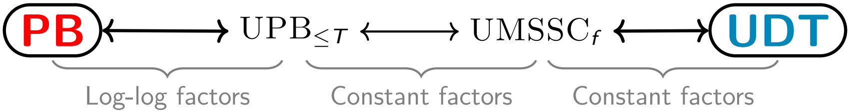
Pandora's Box vs Uniform Decision Tree

	Pandora's Box	**Uniform Decision Tree**
Probabilities	non uniform	uniform
Minimize	opening cost + value chosen	total test cost
Stopping time	depends on algorithm	when scenarios are distinguished

How to transform PB to UDT?

Roadmap of Results

Idea: connect **P**andora's **B**ox with **U**niform **D**ecision **T**ree via similar problems



- $\mathrm{UPB}_{\leq T}$: Uniform Pandora's Box with outside option T
- UMSSC_f: Uniform Min Sum Set Cover (with feedback)

A Related Problem: $\mathrm{UPB}_{\leq T}$

Uniform Pandora's Box with outside option

Given: m scenarios, n boxes

- Outside option T
- Can stop if: value $\leq T$
- Can pay outside option T and stop

Goal: Open boxes to cover all scenarios
Minimize: Sum of opening costs
(+outside option)

	Box 1	Box 2	Box 3	...	Box n	Option T
$1/m \to$ **Scenario** 1:	42	13	47		24	T
$1/m \to$ **Scenario** 2:	0	124	94	...	2	T
⋮			⋮			
$1/m \to$ **Scenario** m:	10	91	9	...	65	T

Table: Example for $T = 42$

A Related Problem: UMSSC_f

Uniform Min Sum Set Cover with feedback

Given: m sets (scenarios), n elements (boxes)

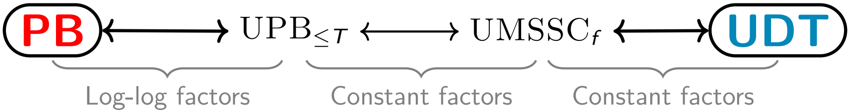
- ∞: Element does **not** belong to set
- 0: Element belongs to set

Goal: Select elements to cover all sets
Minimize: Sum of covering times

	Element 1	**Element** 2	**Element** 3	...	**Element** n
$1/m \to$ **Set** 1	0_3	∞_9	∞_{42}		0_{19}
$1/m \to$ **Set** 2	0_{99}	0_{23}	∞_{57}	...	∞_{67}
$\vdots$			$\vdots$		
$1/m \to$ **Set** m	∞_{13}	0_{12}	0_9	$\dots_{24}$	∞_{21}

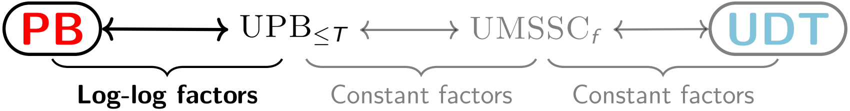
Roadmap of Results

Idea: connect **P**andora's **B**ox with **U**niform **D**ecision **T**ree via similar problems



- $\mathrm{UPB}_{\leq T}$: Uniform Pandora's Box with outside option T
- UMSSC_f: Uniform Min Sum Set Cover (with feedback)

Reduction Overview



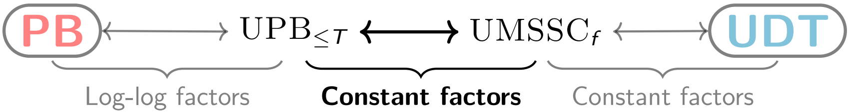
Reduction Description

- Choose T_i s.t. enough scenarios have value $< T_i$
- run UPB_{T_i}
- repeat by only keeping scenarios that chose outside option T_i

Intuition

- Make probabilities uniform
- Stopping time is simplified
  - either find value $< T$
  - or choose T and stop

Reduction Overview



Reduction Description

- Create T copies of sets for each scenario in $\mathrm{UPB}_{\leq T}$
- simulate outside option T

Intuition

- Remove outside option T
- Stopping time is clear: need to **cover** all scenarios

Reduction Overview

PB $\longleftrightarrow$ $\mathrm{UPB}_{\leq T}$ $\longleftrightarrow$ UMSSC_f $\longleftrightarrow$ **UDT**

Log-log factors — Constant factors — **Constant factors**

Reduction Description

- Add feedback info to every element
- When element (test) is chosen, we get information about scenario realized

Intuition

- Change objective to **distinguish** scenarios instead of just **covering**

Conclusion

Question

Pandora's Box with Correlations against the Fully Adaptive

Our Result

Main Result: Pandora's Box **equivalent** to Uniform Decision Tree

- Connections to well studied problems (Decision Tree, Min Sum Set Cover)
- Reductions still hold with arbitrary box costs
- Mixture of distributions: constant approximation in $n^{m^2/\varepsilon}$ time

Conclusion

Question

Pandora's Box with Correlations against the Fully Adaptive

Our Result

Main Result: Pandora's Box **equivalent** to Uniform Decision Tree

- Connections to well studied problems (Decision Tree, Min Sum Set Cover)
- Reductions still hold with arbitrary box costs
- Mixture of distributions: constant approximation in $n^{m^2/\varepsilon}$ time

Thank you!