Learning Online Algorithms with Distributional Advice

Ilias Diakonikolas Vasilis Kontonis Christos Tzamos

Ali Vakilian Nikos Zarifis

Algorithms with Predictions/Advice

Input:

- Instance **I** of problem **P**,
- Prediction/Advice **A** about **I**

Goal: Design **ALG**(P, I, A) s.t.,

1. If **A** is *accurate*, then *cost* of **ALG** is close to **OPT**(P, I)
2. Otherwise, cost of **ALG** is close to best (classical) algorithm of **P**

Algorithms with Advice

Input:

- Instance **I** of problem **P**,
- Prediction/Advice **A** abo

Goal: Design **ALG**(P, I, A) s

1. If A is *accurate*, then co
2. Otherwise, cost of **ALG**

- Motivated by the success of ML approaches
- Falls in *Beyond Worst-Case Analysis* Framework

Popular approach for Online problems:
The entire input is not available from the start
(Renault & Rosén, 2015; Angelopoulos et al.,2015; Lykouris & Vassilvtiskii, 2018; Purohit et al., 2018; Gollapudi & Panigrahi, 2019; Angelopoulos et al., 2020;Dütting et al., 2020; Lattanzi et al., 2020; Anand et al.,2020; Bamas et al., 2020)

Also, studied for problems in *learning theory*, *data structures*, *streaming and sketching*, *and combinatorial optimization.*

Algorithms with Distributional Advice

Main Contribution: *distributional advice*, instead of *det. Advice*

Algorithms with Distributional Advice

Main Contribution: *distributional advice*, instead of *det. Advice*

Ski-Rental Problem. In each (ski) day, the player decides whether to **rent skis** for this day (costs *1*) or **buy skis** for the rest of the season (costs b).

- ❑ The goal is to minimize the cost paid by the player.
- ❑ **In Classical Online Setting**: algorithm with competitive ratio $\frac{e}{e-1}$

Algorithms with Distributional Advice

Main Contribution: *distributional advice*, instead of *det. Advice*

Ski-Rental Problem. In each (ski) day, the player decides whether to **rent skis** for this day (costs *1*) or **buy skis** for the rest of the season (costs *b*).

- ❑ The goal is to minimize the cost paid by the player.
- ❑ **In Classical Online Setting**: algorithm with competitive ratio $\frac{e}{e-1}$

Deterministic Advice: *predicted number of ski-days* [Purohit et al., 2018]

Algorithms with Distributional Advice

Main Contribution: *distributional advice*, instead of *det. Advice*

Ski-Rental Problem. In each (ski) day, the player decides whether to **rent skis** for this day (costs *1*) or **buy skis** for the rest of the season (costs *b*).

- ❑ The goal is to minimize the cost paid by the player.
- ❑ **In Classical Online Setting**: algorithm with competitive ratio $\frac{e}{e-1}$

Deterministic Advice: *predicted number of ski-days* [Purohit et al., 2018]

However, more natural predictions are of form of **distribution over days**

E.g., uniform distribution over some interval, or normal, exponentials, …

Our Distributional Advice Framework

Setup. an online problem **P**, an unknown distribution **D** on inputs of **P** (inputs to **P** are drawn from D)

Goal: find alg. $\mathcal{A}$ that minimizes $\mathrm{cost}(\mathcal{A}; D) \coloneqq \mathbb{E}_{t\sim D}[\mathrm{cost}(\mathcal{A}; t)]$

Our Distributional Advice Framework

Setup. an online problem **P**, an unknown distribution **D** on inputs of **P** (inputs to **P** are drawn from D)

Goal: find alg. $\mathcal{A}$ that minimizes $\text{cost}(\mathcal{A}; D) \coloneqq \mathbb{E}_{t\sim D}[\text{cost}(\mathcal{A}; t)]$

With Distributional Advice:

- Given an advice family of distributions $\boldsymbol{\mathcal{C}}$
- The goal is to design *α-consistent* and β-robust algorithm $\mathcal{A}_{\mathcal{X}}$ from i.i.d. samples $\mathcal{X} \sim \mathrm{D}$ s.t.
  - If $D \in \mathcal{C}$, then $\text{cost}(\mathcal{A}_{\mathcal{X}}; D) \coloneqq \mathbb{E}_{t\sim D}[\text{cost}(\mathcal{A}; t)] \le \alpha \cdot \mathrm{OPT}_D$
  - Otherwise, $\text{cost}(\mathcal{A}_{\mathcal{X}}; D) \le \beta \cdot \mathrm{OPT}_{\mathbf{ONL}}$ *(the cost of optimal online alg on D)*

Our Results (Ski-Rental)

Observation. For $\alpha < \frac{e}{e-1}$, there exists no algorithm that given any distribution **D**, draws finitely many samples from **D** and returns an α-consistent strategy.

In other words, it is not possible to beat the existing competitive ratio of ski-rental without distributional assumptions.

Our Results (Ski-Rental)

Observation. For $\alpha < \frac{e}{e-1}$, there exists no algorithm that given any distribution **D**, draws finitely many samples from **D** and returns an α-consistent strategy.

In other words, it is not possible to beat the existing competitive ratio of ski-rental without distributional assumptions.

Result 1. For any $\lambda > 1$, there exists an algorithm that draws $\tilde{O}(1/\varepsilon^2)$ samples and outputs a $(\lambda(1+\varepsilon))$*-consistent* and $(\frac{\lambda}{\lambda+1})$*-robust* strategy for ski-rental on log-concave distributions.

Moreover, this sample complexity is essentially optimal: $\Omega(1/\varepsilon^2)$ *samples are necessary to get a* $(1+\varepsilon)$*-consistent strategy.*

Our Results (Ski-Rental)

Observation. For $\alpha < \frac{e}{e-1}$, there exists no algorithm that given any distribution **D**, draws finitely many samples from **D** and returns an α-consistent strategy.

In other words, it is not possible to beat the existing competitive ratio of ski-rental without distributional assumptions.

Result 1. For any $\lambda > 1$, there exists an algorithm that draws $\tilde{O}(1/\varepsilon^2)$ samples and outputs a $(\lambda(1+\varepsilon))$*-consistent* and $(\frac{\lambda}{\lambda+1})$*-robust* strategy for ski-rental on log-concave distributions.

Moreover, this sample complexity is essentially optimal: $\Omega(1/\varepsilon^2)$ *samples are necessary to get a* $(1+\varepsilon)$*-consistent strategy.*

The robustness guarantee is achieved by using a result of [Mahdian et al., 2012].

	$(1+\varepsilon)$-Multiplicative		Consistency and Robustness	ε-Additive‡
	General	Log-Concave		
Ski-Rental	Inapprox.	$\tilde{O}(\varepsilon^{-2})$	$(\lambda(1+\varepsilon))$-*consistent* and $(\frac{\lambda}{\lambda+1})$-*robust*	$\tilde{O}(b^2\varepsilon^{-2})$
Prophet Inequality	Inapprox.	$\tilde{O}(n^3\varepsilon^{-2})$	$(\lambda(1+\varepsilon))$-*consistent* and $(\frac{2\lambda}{\lambda-1})$-*robust*	$\tilde{O}(b^2n^2\varepsilon^{-2})$

‡An algorithm $\mathcal{A}$ achieves ε-additive approximation if $\mathrm{cost}(\mathcal{A}, D) \leq \mathrm{OPT}_D + \varepsilon$

Result. For any input distribution **D**, $O(1/\varepsilon^4)$ conditional samples from **D** suffice to design strategy $\mathcal{A}$ for ski-rental s.t. $\mathrm{cost}(\mathcal{A}; \mathbf{D}) \leq (1+\varepsilon)\mathrm{OPT}_{\mathbf{D}}$

Thank You