The Space of Online Learning Problems

Yevgeny Seldin

University of Copenhagen

ECML-PKDD-2015 Tutorial

What is Online Learning?

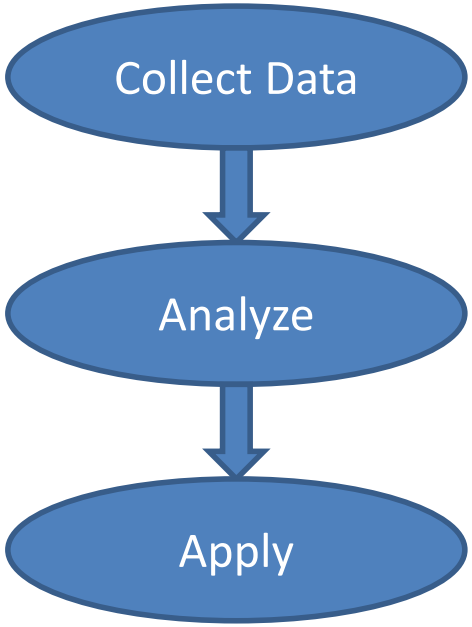
- Subfield of Machine Learning studying problems involving interaction with an environment

Examples

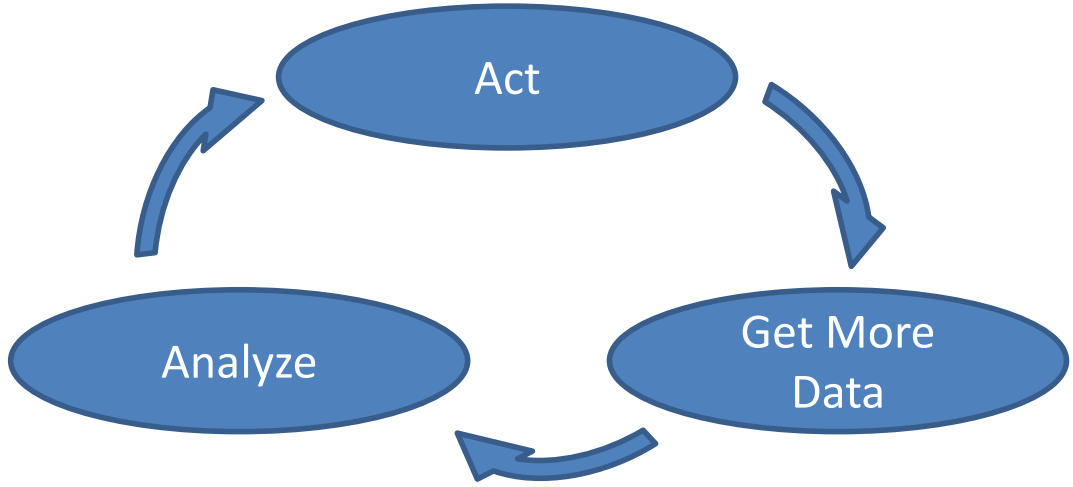
- Investment in the stock market
- Online advertising/personalization/routing/…
- Games
- Robotics
- …

How Online different from “batch”?

Batch Learning



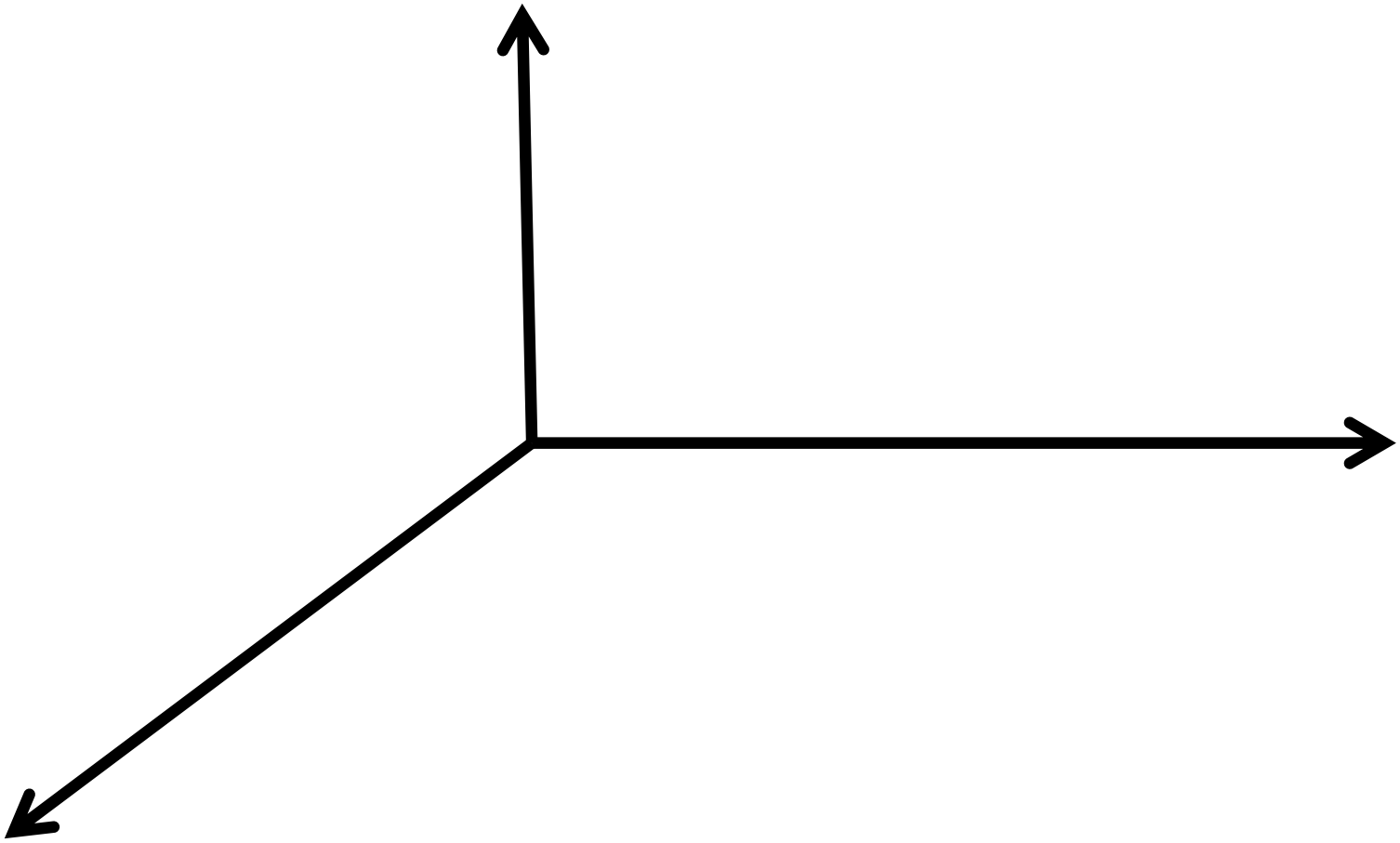
Online Learning



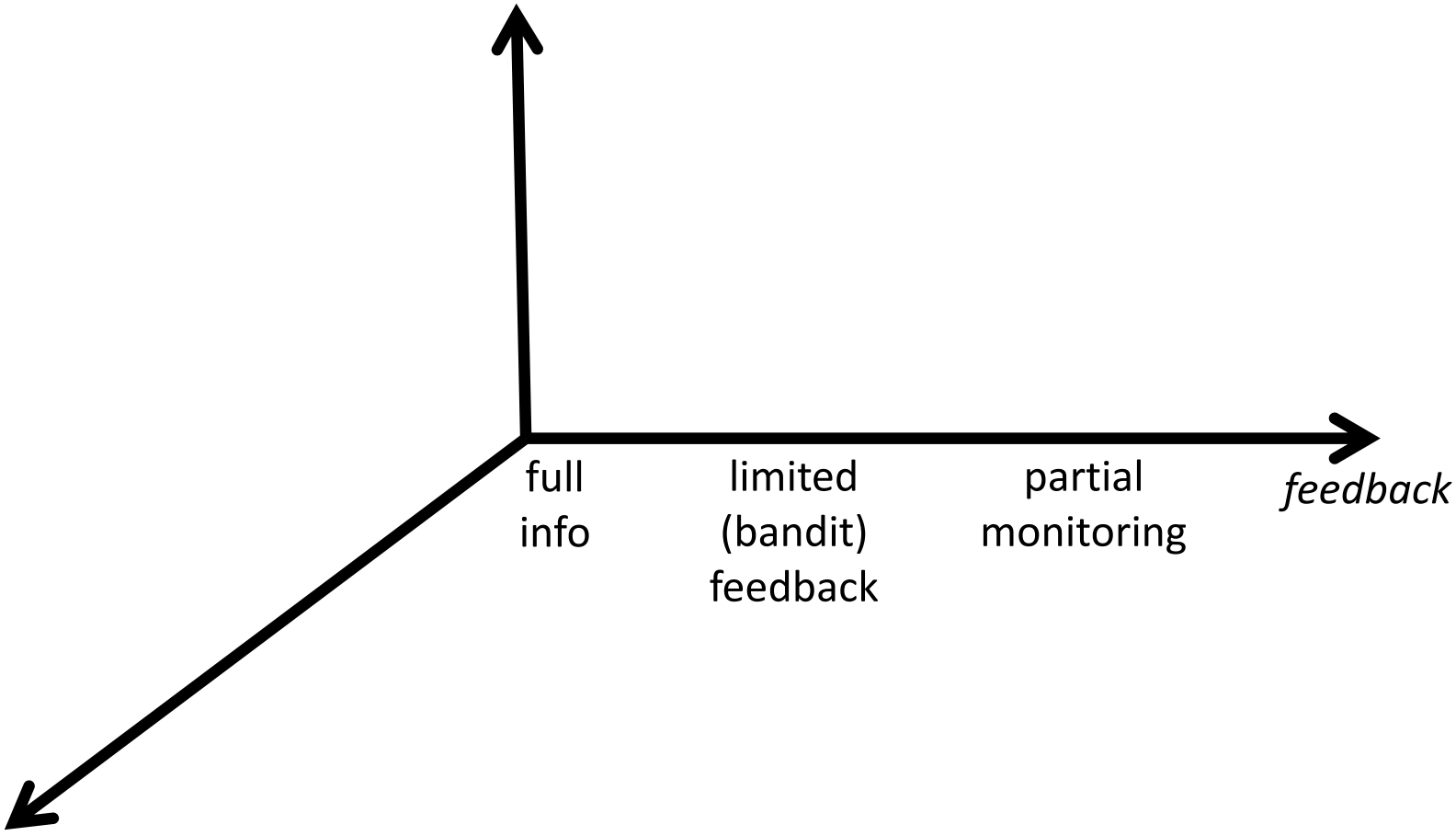
When do we need Online Learning?

- Interactive learning
- “Adversarial” game-theoretic settings
  - No i.i.d. assumptions
- Large-scale data analysis

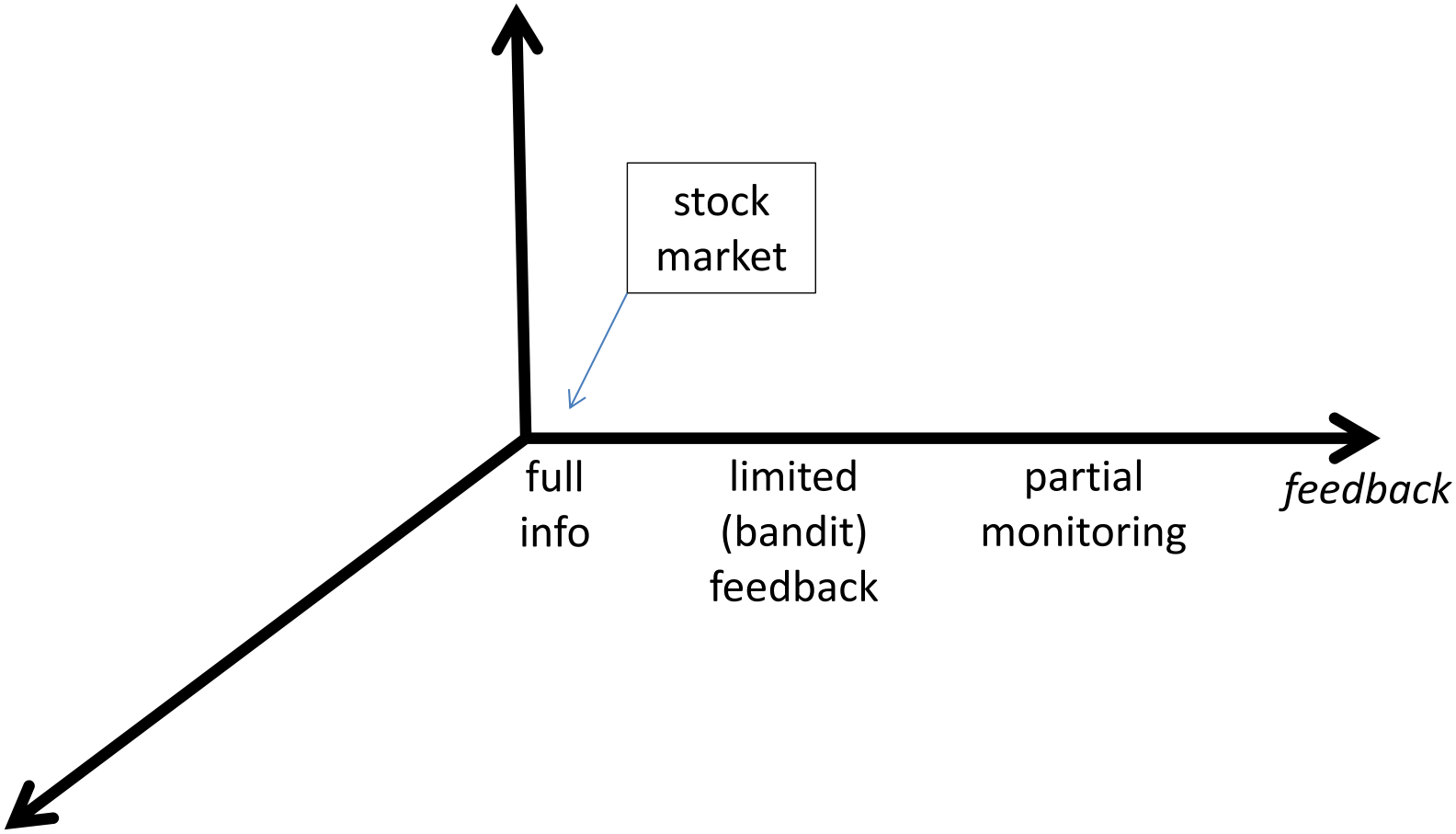
The Space of Online Learning Problems



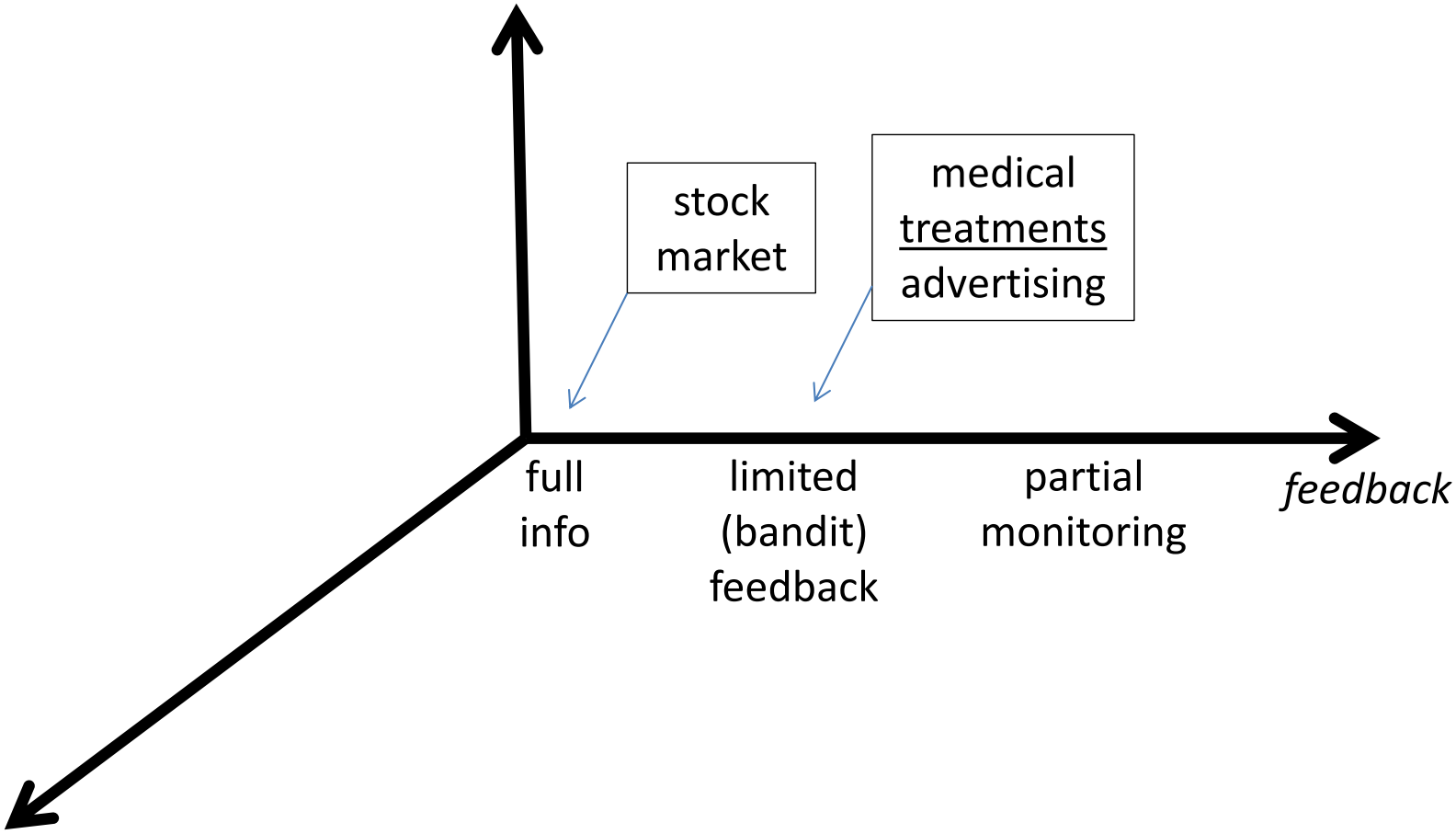
The Space of Online Learning Problems



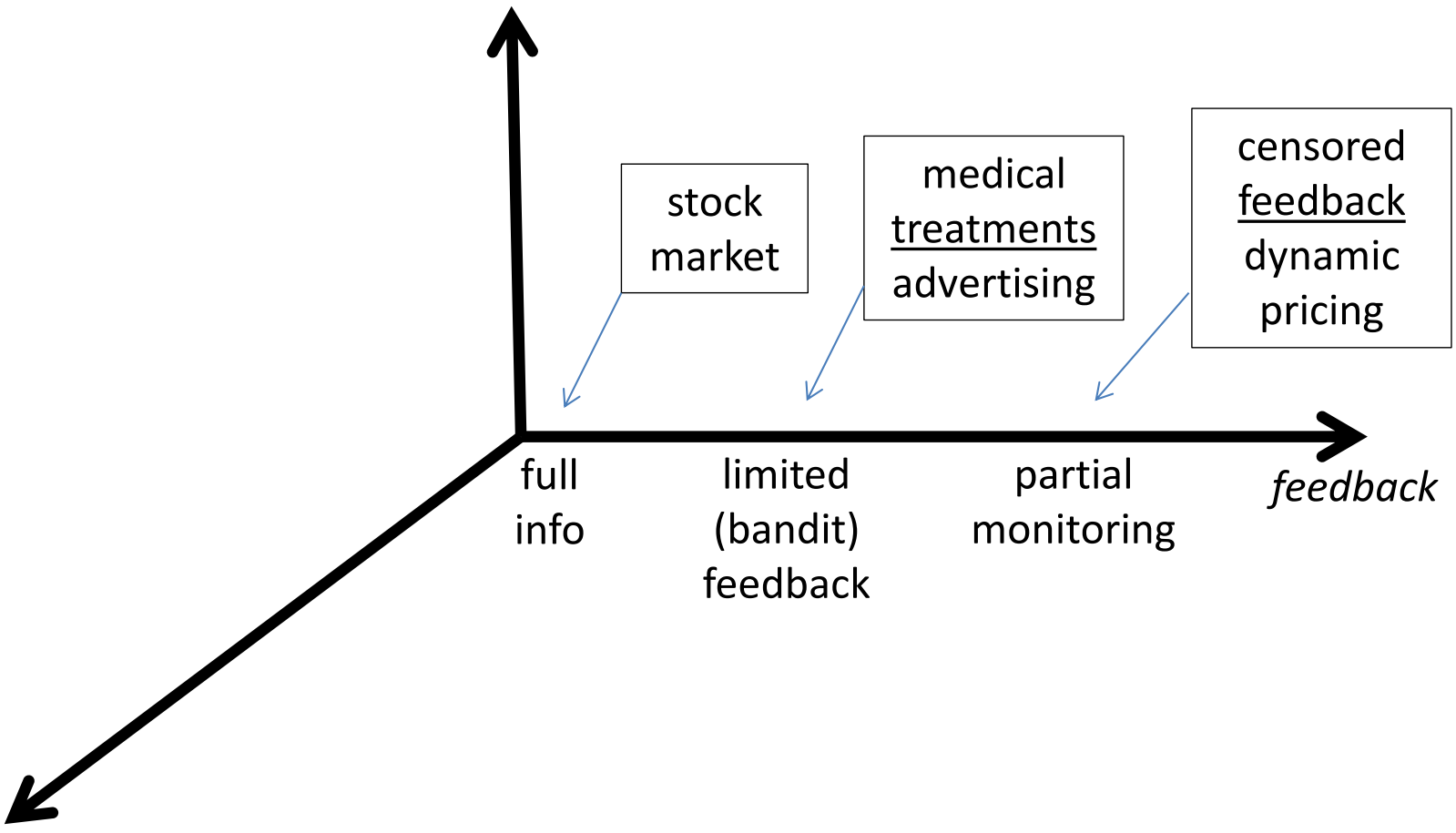
The Space of Online Learning Problems



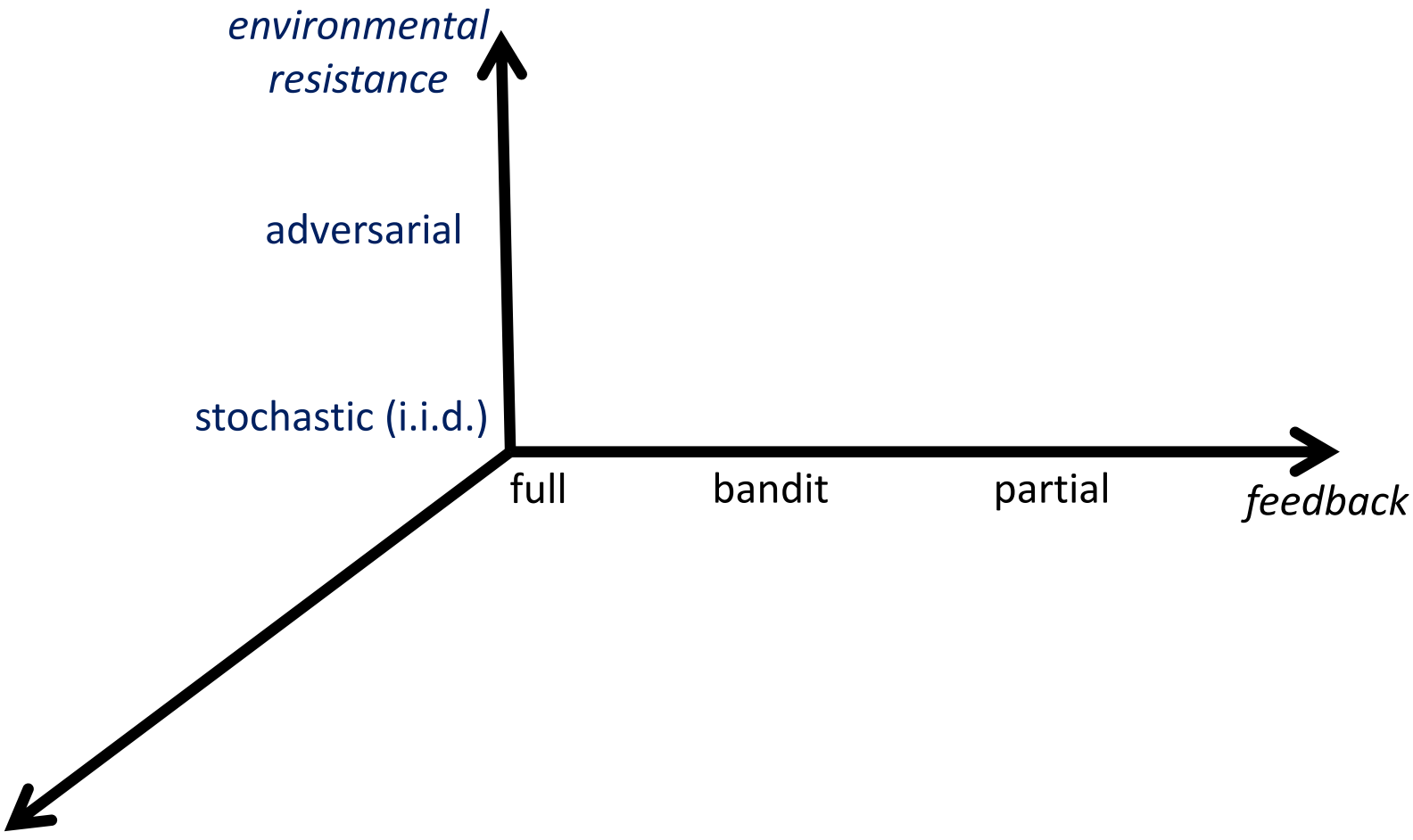
The Space of Online Learning Problems



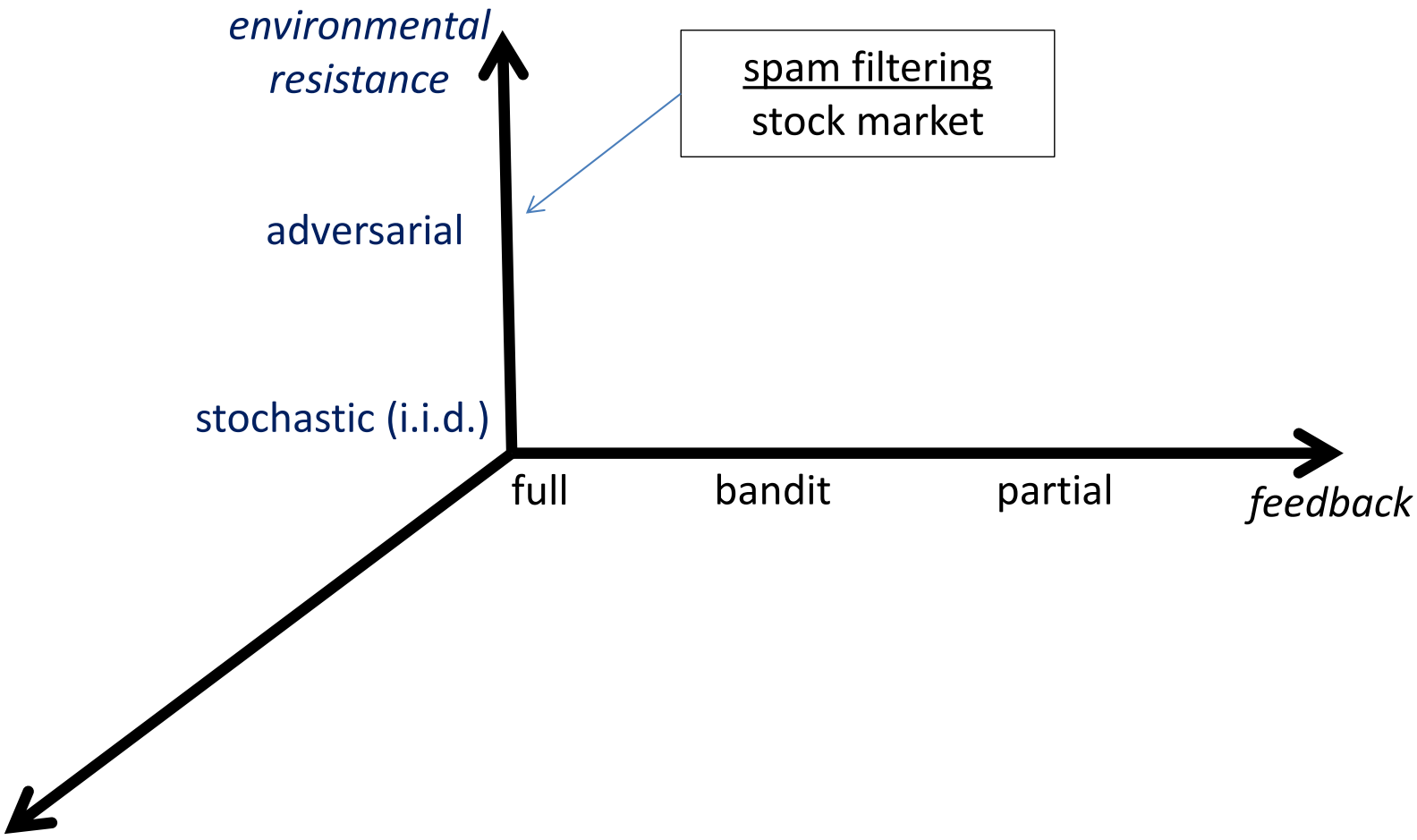
The Space of Online Learning Problems



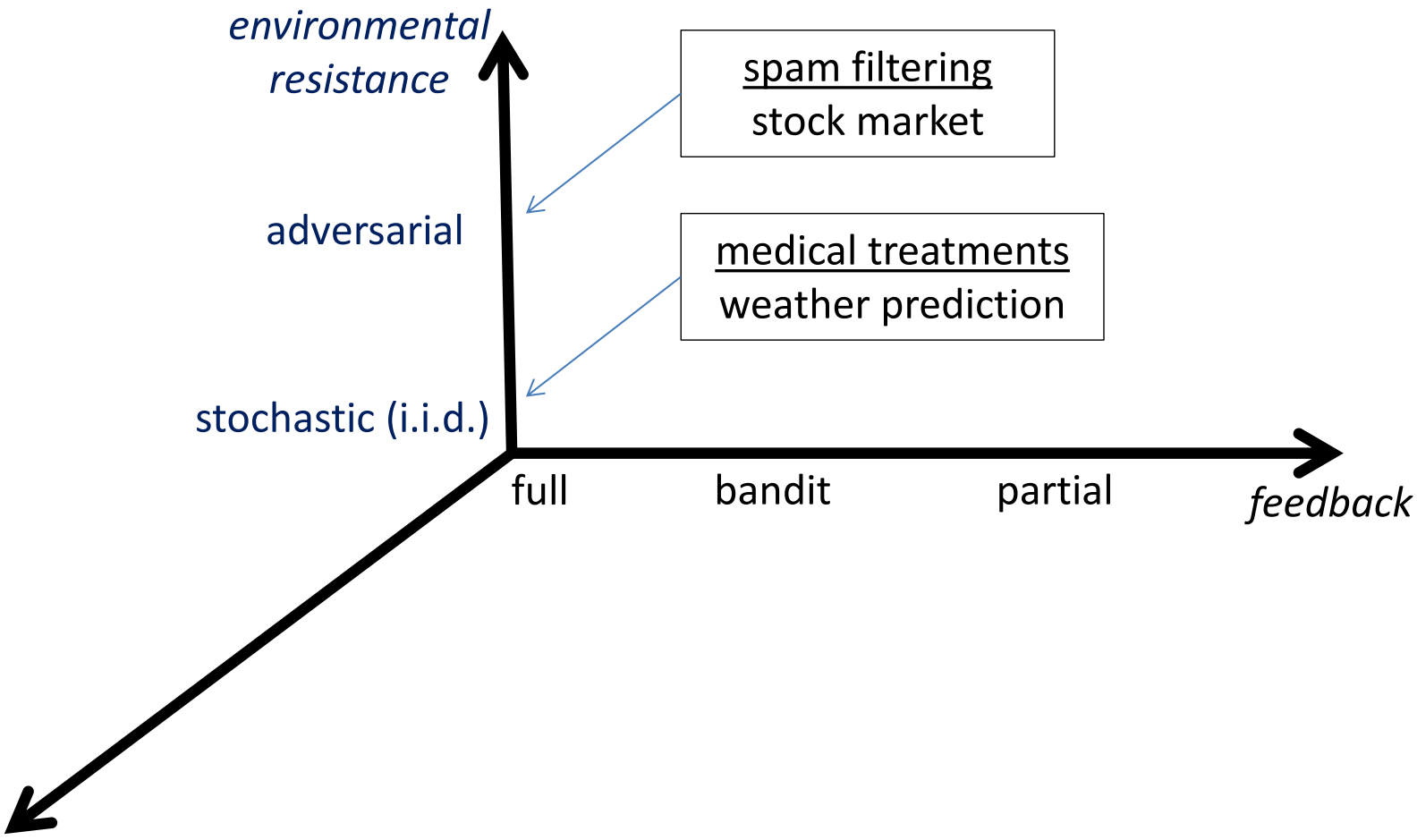
The Space of Online Learning Problems



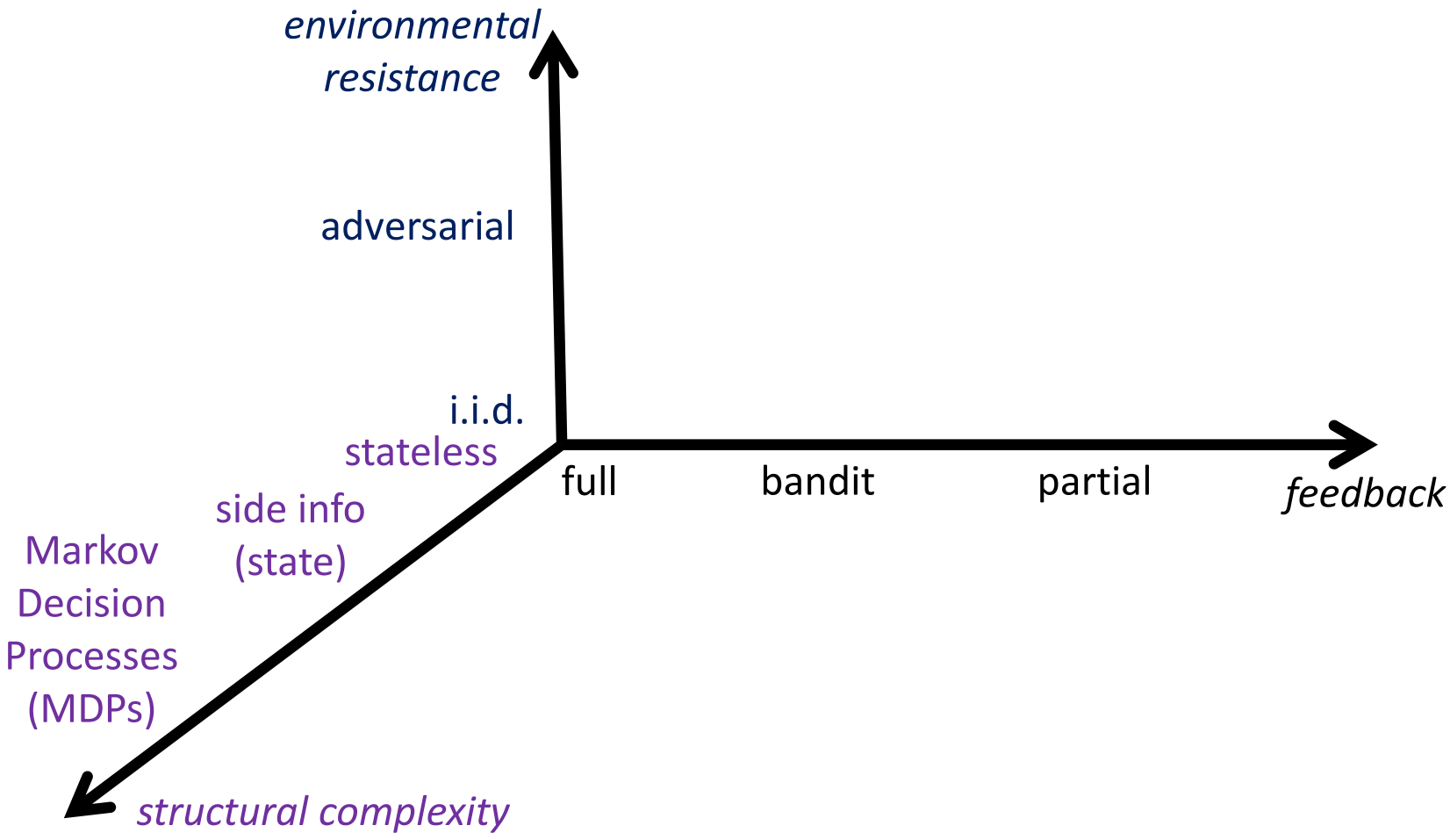
The Space of Online Learning Problems



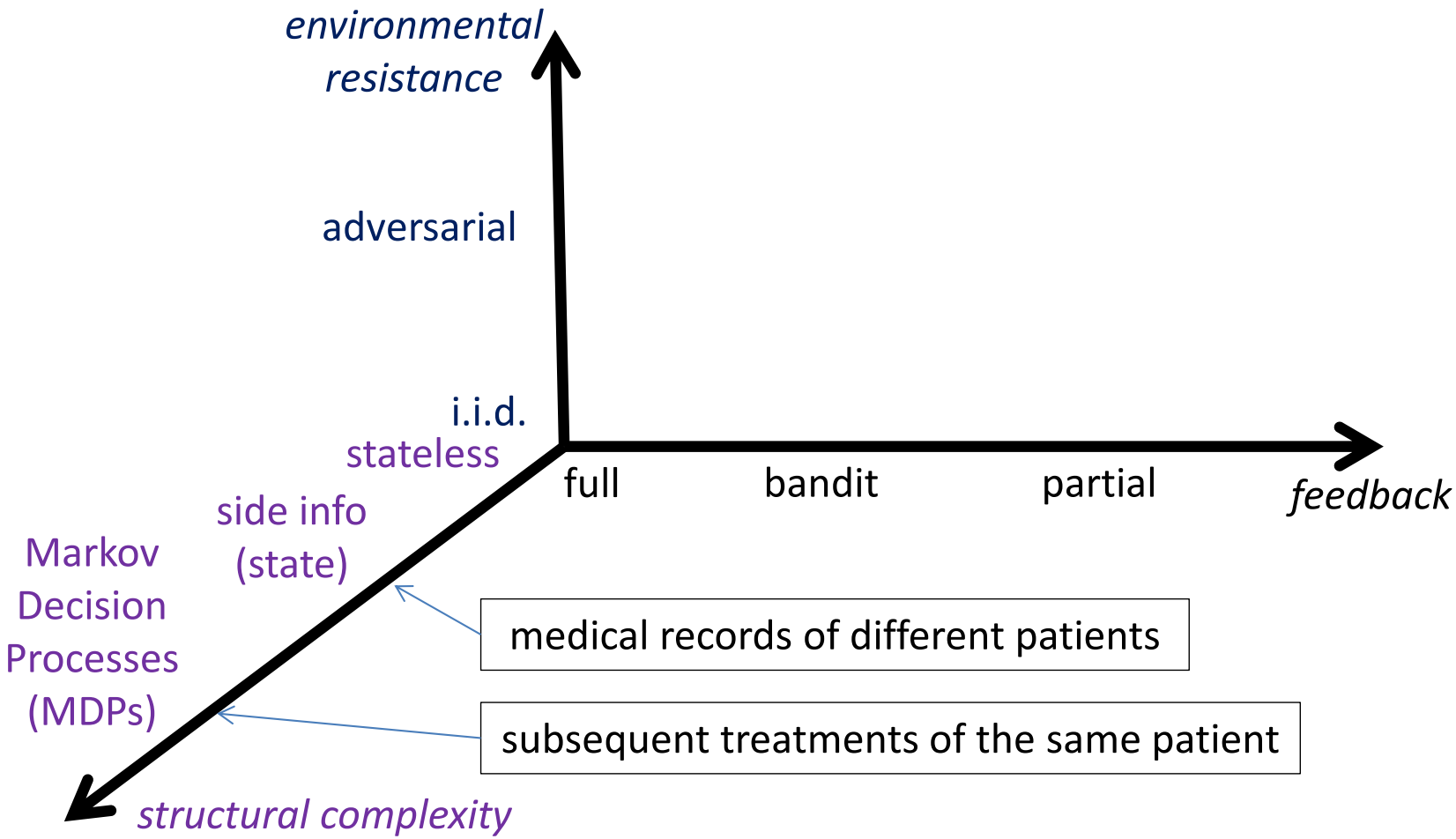
The Space of Online Learning Problems



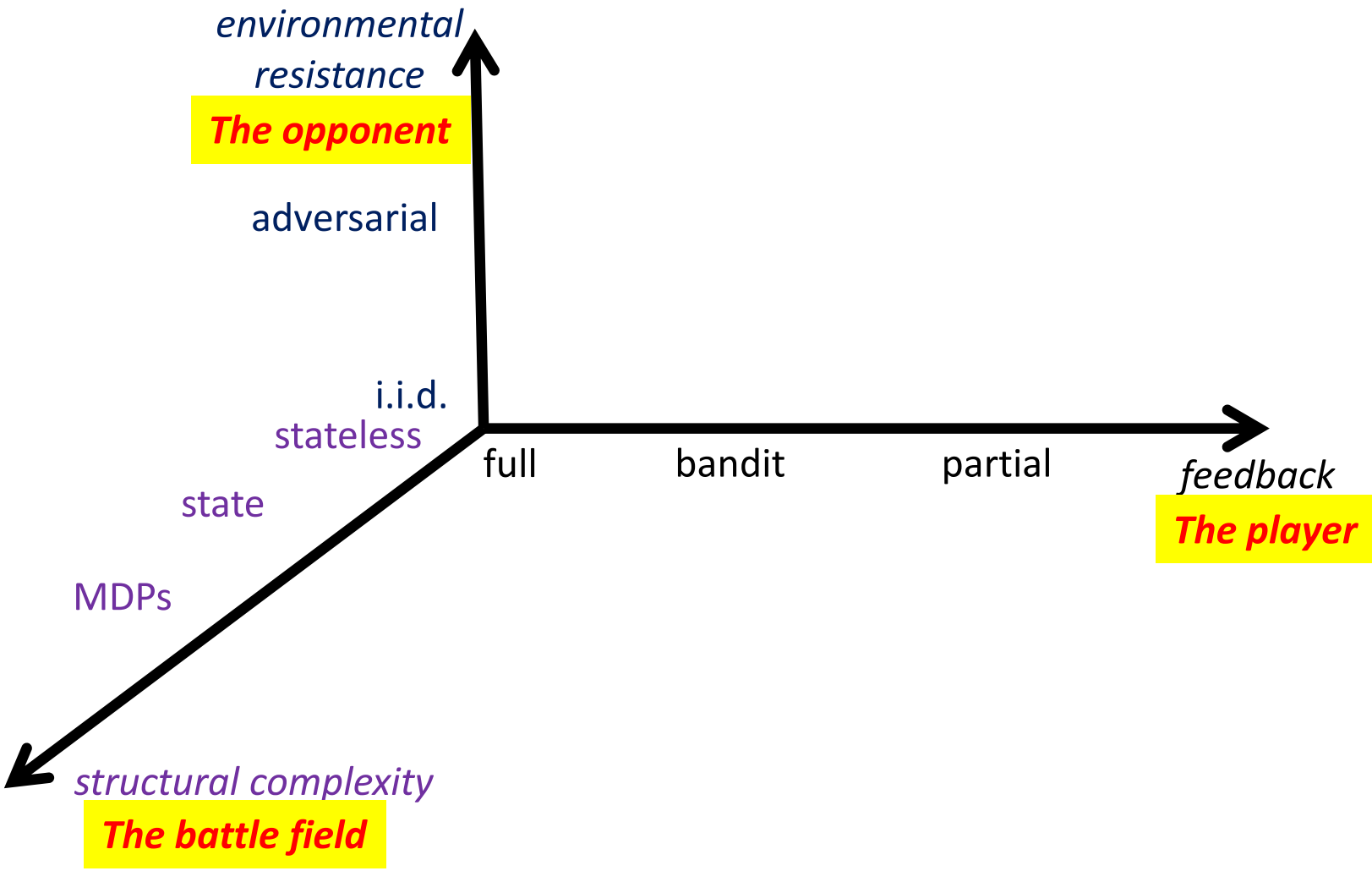
The Space of Online Learning Problems



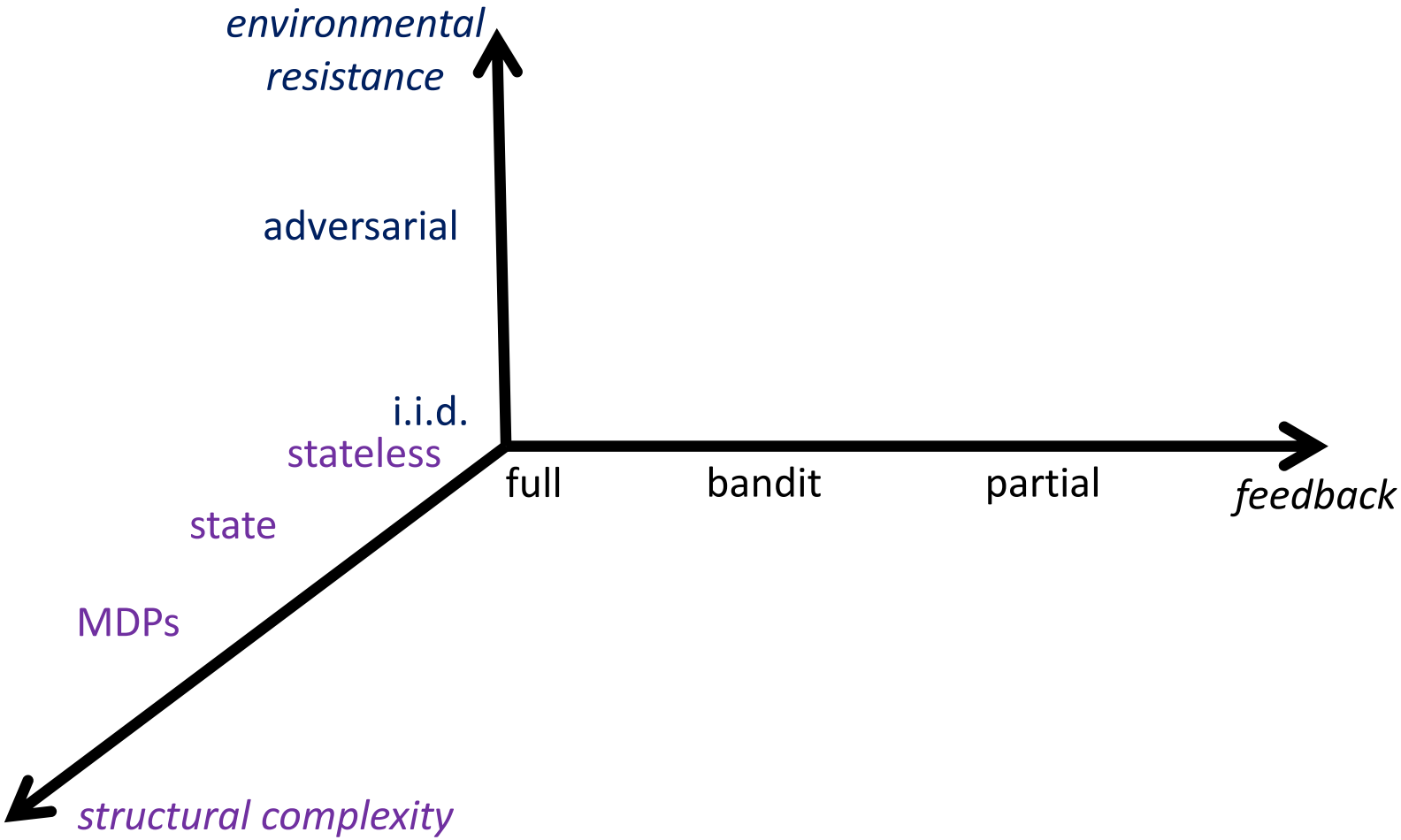
The Space of Online Learning Problems



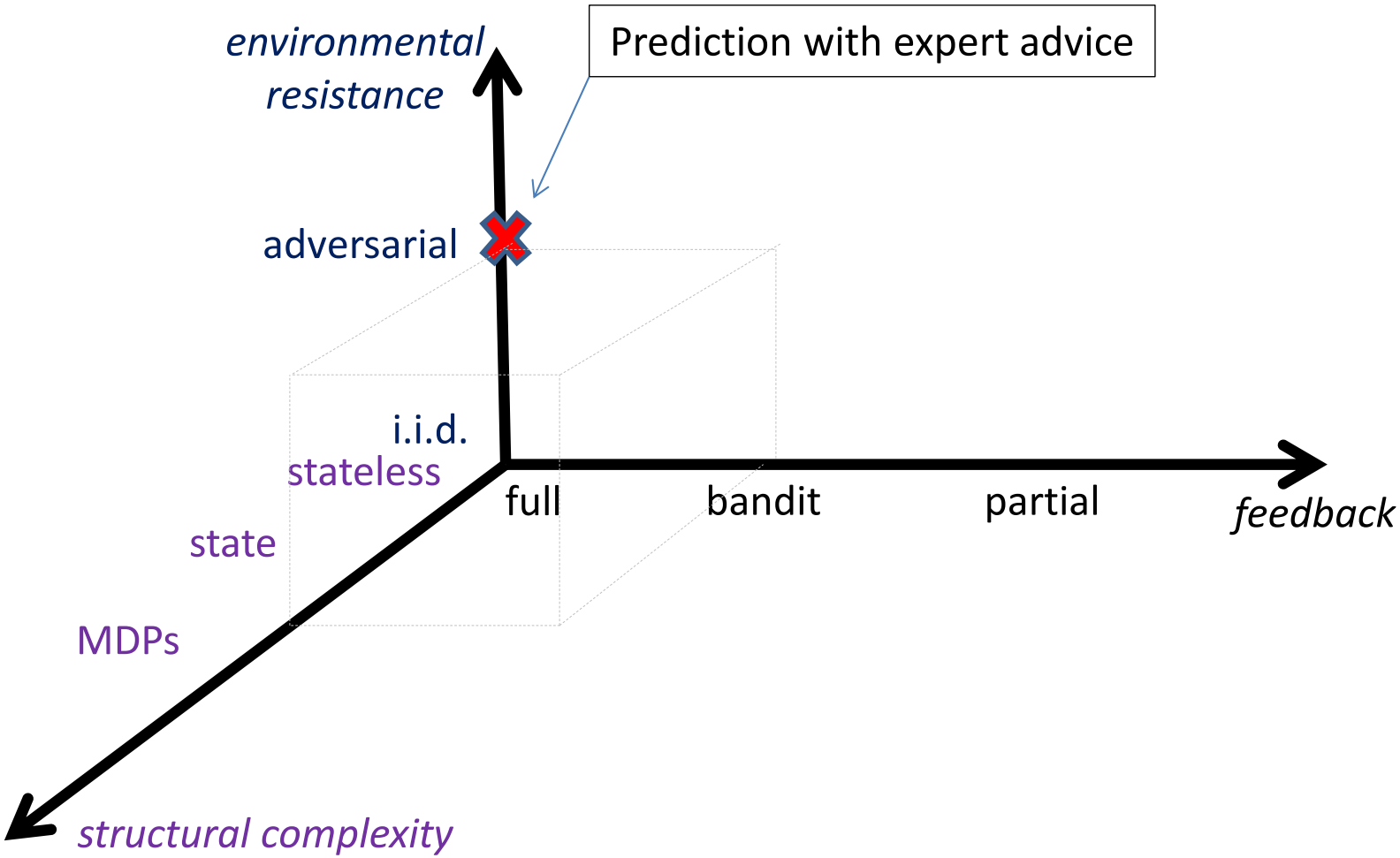
The Space of Online Learning Problems



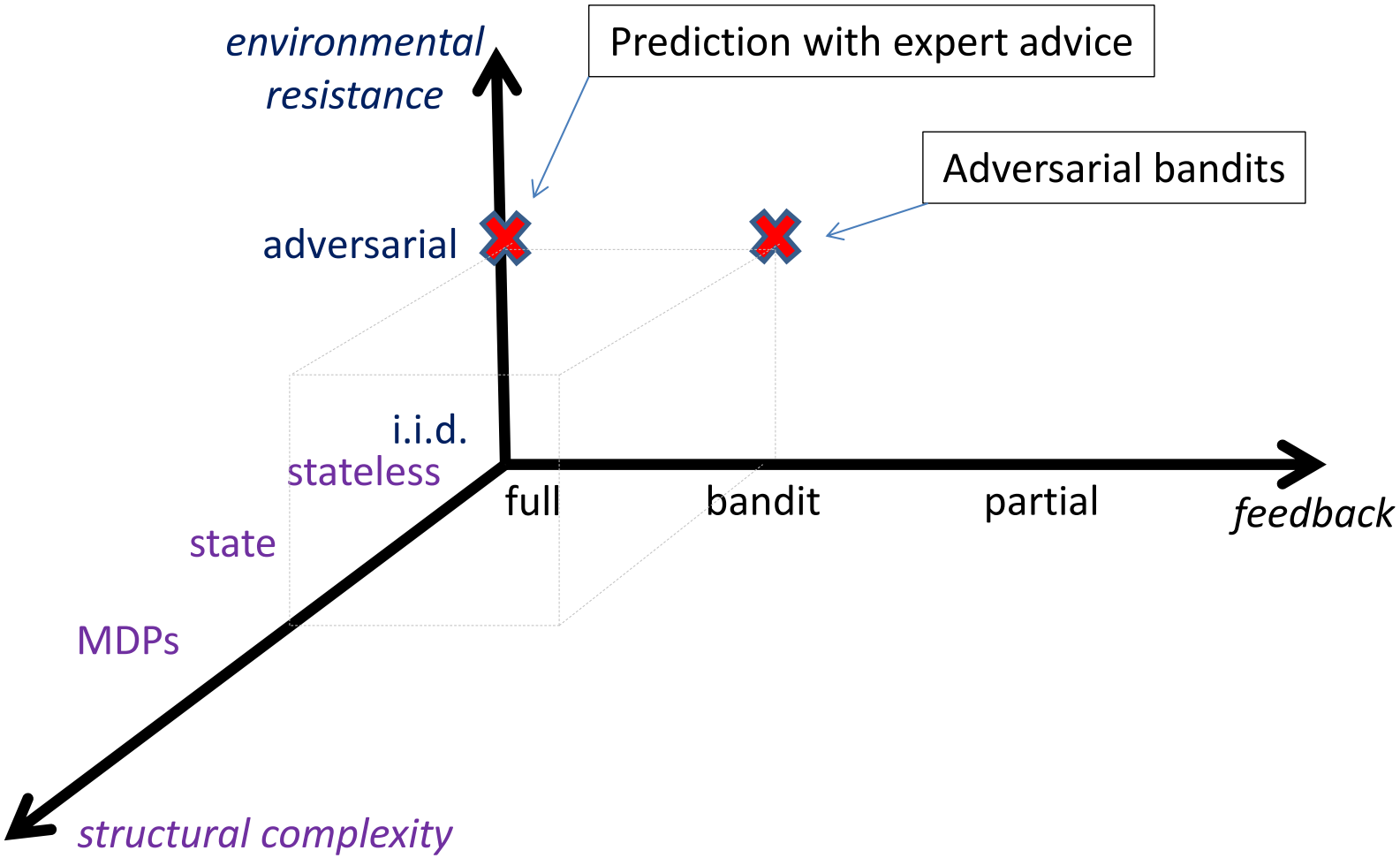
Part I: “classical” algorithms



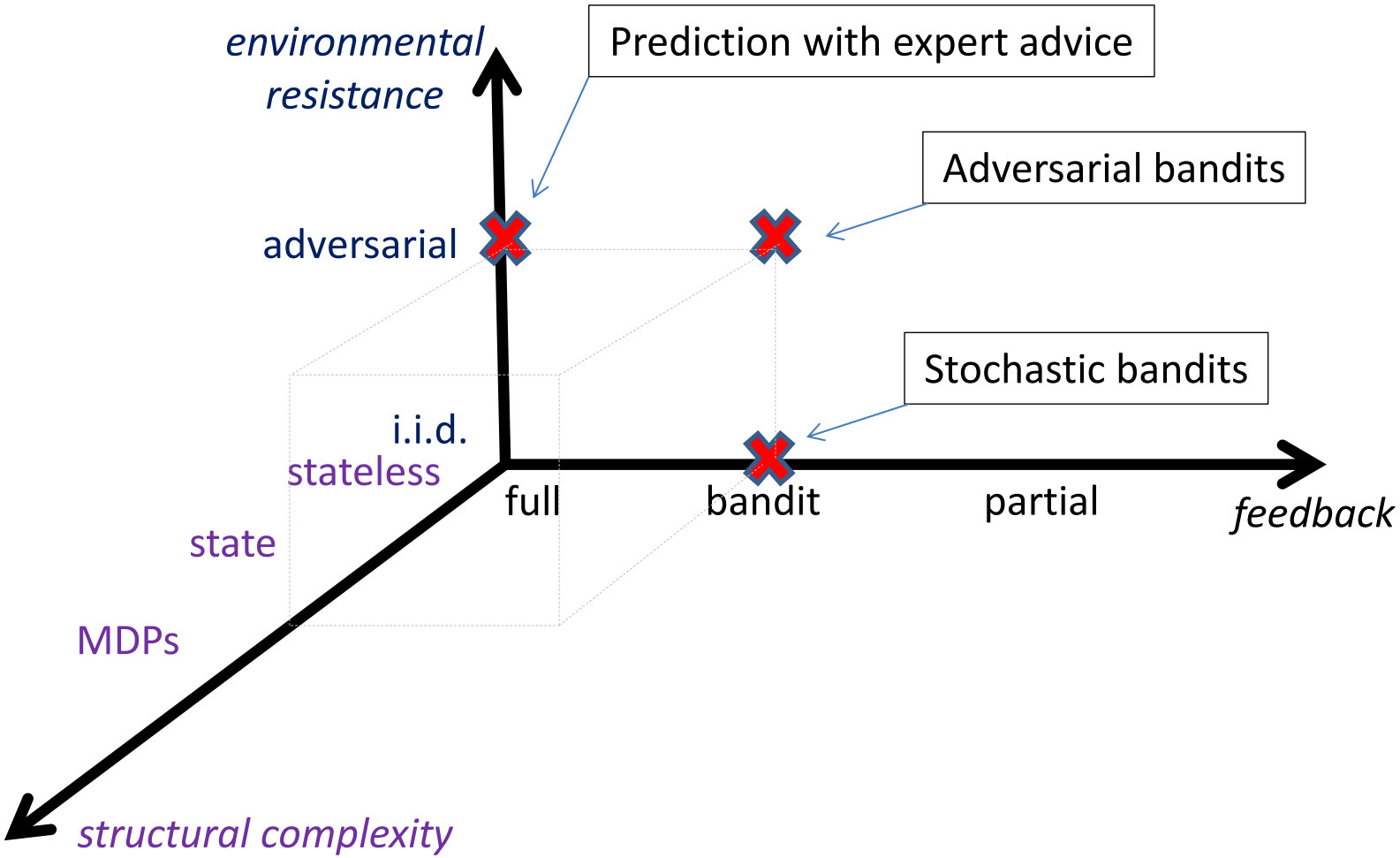
Part I: “classical” algorithms



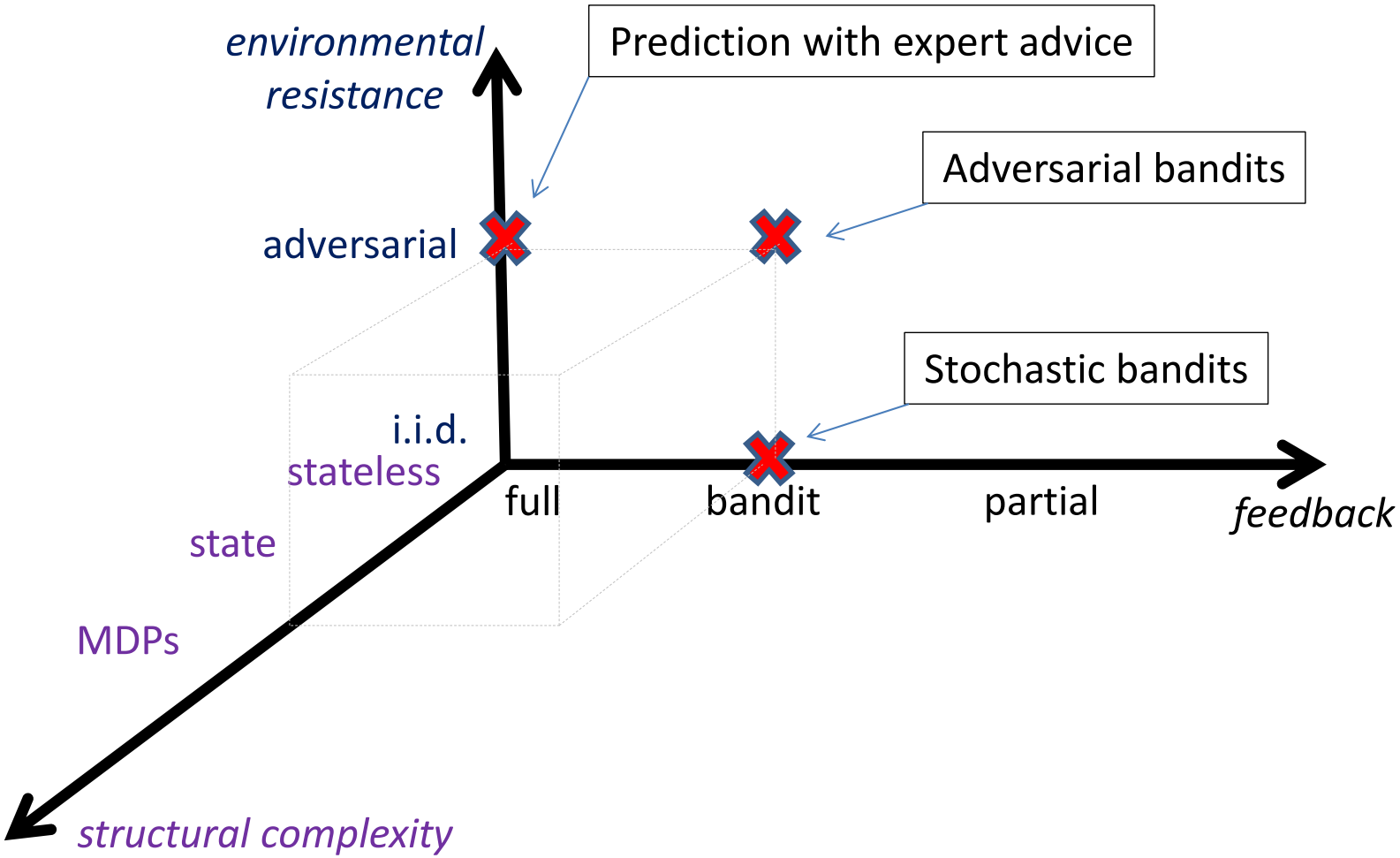
Part I: “classical” algorithms



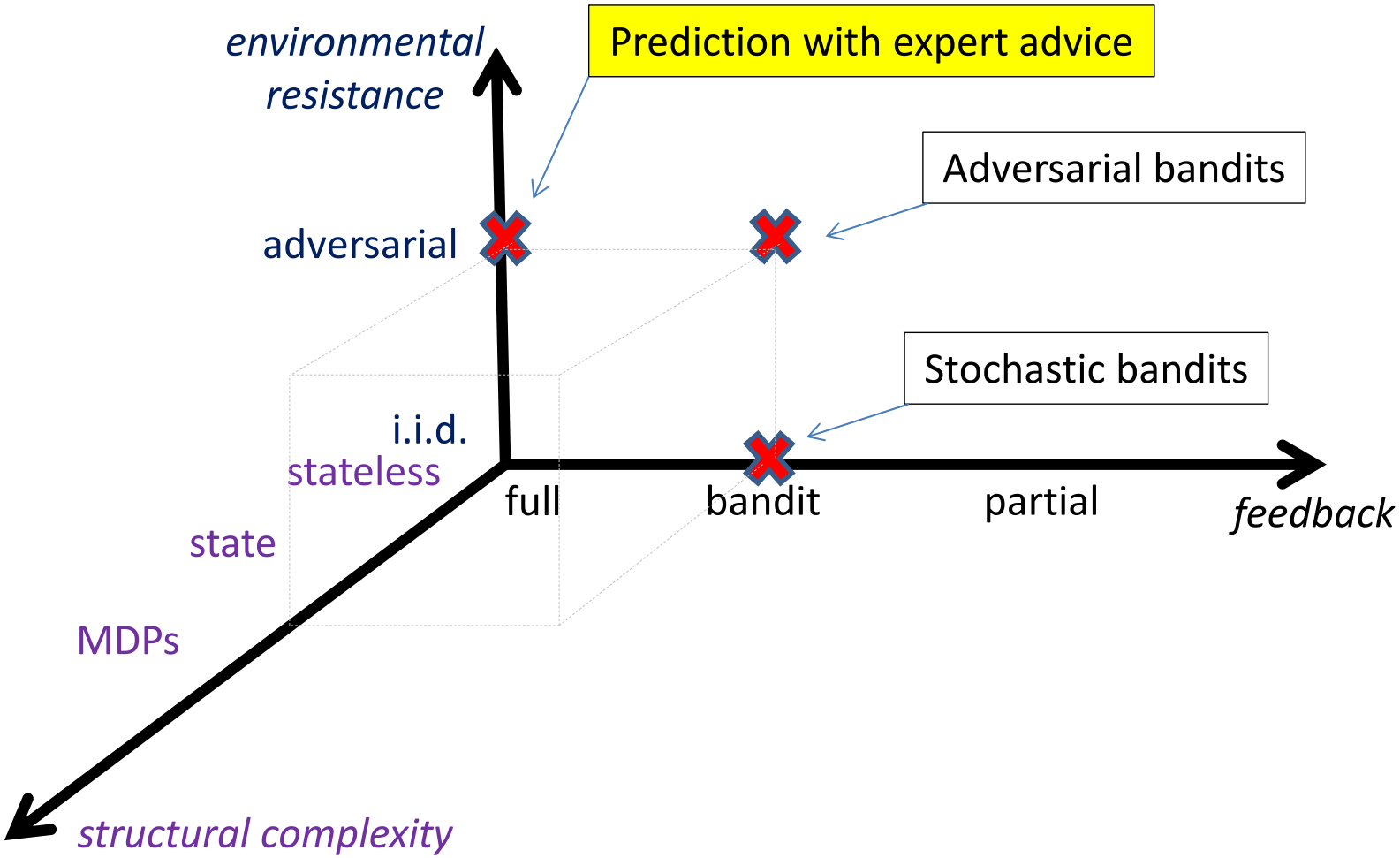
Part I: “classical” algorithms



Part I: “classical” algorithms



Part I: “classical” algorithms



Prediction with Expert Advice

Examples

- Experts = financial advisers
- Experts = different algorithms

Prediction with Expert Advice

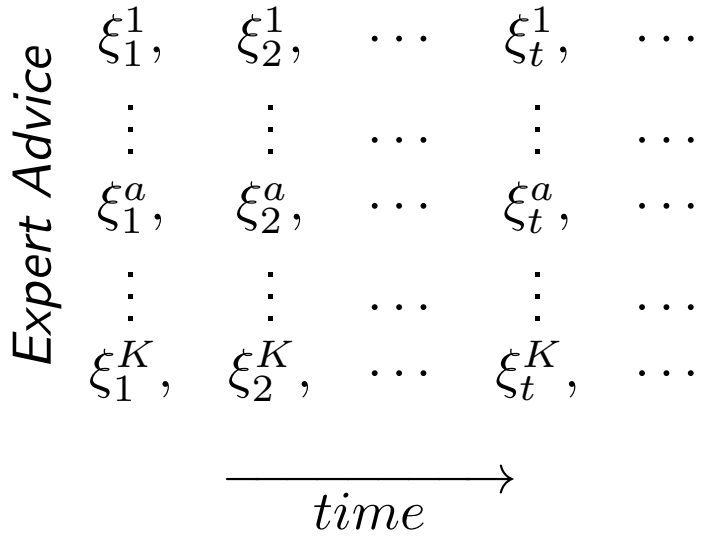
Examples

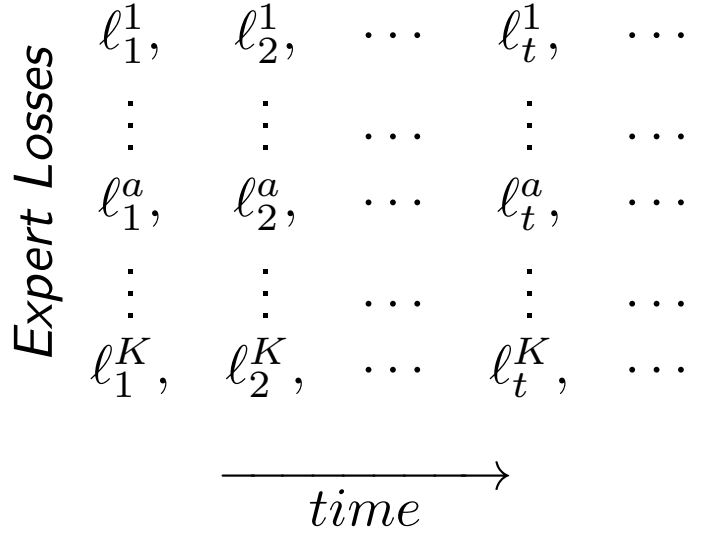
- Experts = financial advisers
- Experts = different algorithms

Game Definition

For $t = 1, 2, \ldots$:

1. Observe advice of K experts
2. Pick an expert A_t to follow
3. Observe $\ell_t^1, \ldots, \ell_t^K$ & suffer $\ell_t^{A_t}$





Prediction with Expert Advice

Examples

- Experts = financial advisers
- Experts = different algorithms

Game Definition

For $t = 1, 2, \ldots$:

1. Observe advice of K experts
2. Pick an expert A_t to follow
3. Observe $\ell_t^1, \ldots, \ell_t^K$ & suffer $\ell_t^{A_t}$

Performance Measure: Regret

$$R_T = \sum_{t=1}^{T} \ell_t^{A_t} - \min_a \left(\sum_{t=1}^{T} \ell_t^a \right)$$

Expert Advice

$$\begin{array}{cccc} \xi_1^1, & \xi_2^1, & \cdots & \xi_t^1, \quad \cdots \\ \vdots & \vdots & \cdots & \vdots \quad \cdots \\ \xi_1^a, & \xi_2^a, & \cdots & \xi_t^a, \quad \cdots \\ \vdots & \vdots & \cdots & \vdots \quad \cdots \\ \xi_1^K, & \xi_2^K, & \cdots & \xi_t^K, \quad \cdots \end{array}$$

$\longrightarrow$ *time*

Expert Losses

$$\begin{array}{cccc} \ell_1^1, & \ell_2^1, & \cdots & \ell_t^1, \quad \cdots \\ \vdots & \vdots & \cdots & \vdots \quad \cdots \\ \ell_1^a, & \ell_2^a, & \cdots & \ell_t^a, \quad \cdots \\ \vdots & \vdots & \cdots & \vdots \quad \cdots \\ \ell_1^K, & \ell_2^K, & \cdots & \ell_t^K, \quad \cdots \end{array}$$

$\longrightarrow$ *time*

Prediction with Expert Advice - Simplification

Examples

- Experts = financial advisers
- Experts = different algorithms

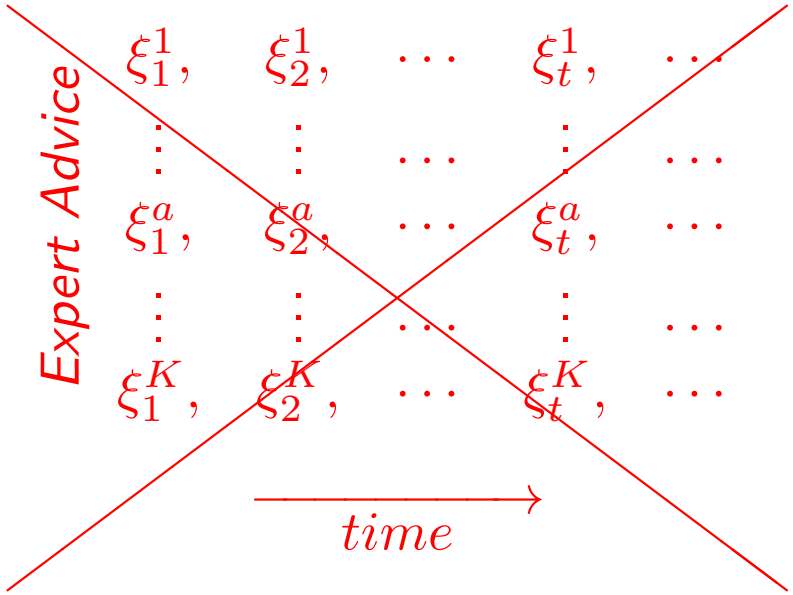
Game Definition

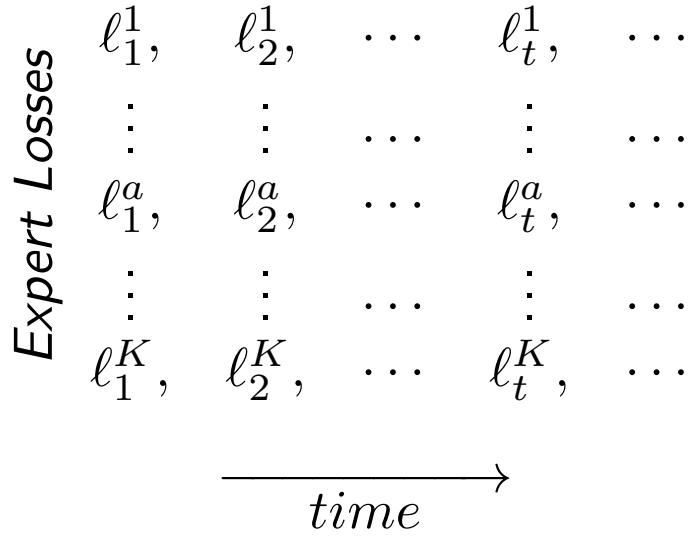
For $t = 1, 2, \ldots$:

1. ~~Observe advice of K experts~~
2. Pick an expert A_t to follow
3. Observe $\ell_t^1, \ldots, \ell_t^K$ & suffer $\ell_t^{A_t}$

Performance Measure: Regret

$$R_T = \sum_{t=1}^{T} \ell_t^{A_t} - \min_a \left(\sum_{t=1}^{T} \ell_t^a \right)$$



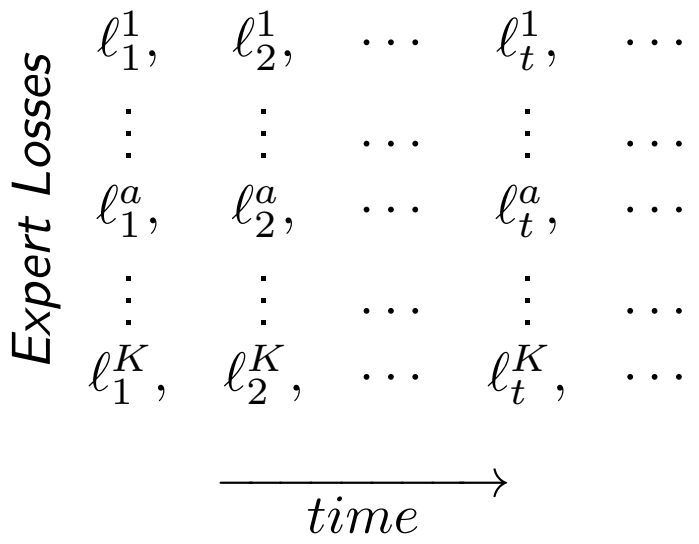


Prediction with Expert Advice

Game Definition

For $t = 1, 2, \ldots$:

1. Pick a row A_t
2. Observe $\ell_t^1, \ldots, \ell_t^K$ & suffer $\ell_t^{A_t}$



Performance Measure: Regret

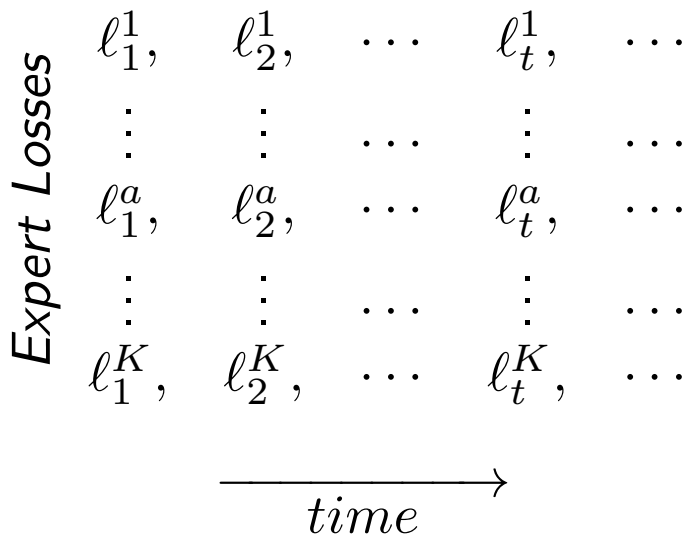
$$R_T = \sum_{t=1}^{T} \ell_t^{A_t} - \min_a \left(\sum_{t=1}^{T} \ell_t^a \right)$$

Prediction with Expert Advice

Game Definition

For $t = 1, 2, \ldots$:

1. Pick a row A_t
2. Observe $\ell_t^1, \ldots, \ell_t^K$ & suffer $\ell_t^{A_t}$



Performance Measure: Regret

$$R_T = \sum_{t=1}^{T} \ell_t^{A_t} - \min_a \left(\sum_{t=1}^{T} \ell_t^a \right)$$

Assumptions

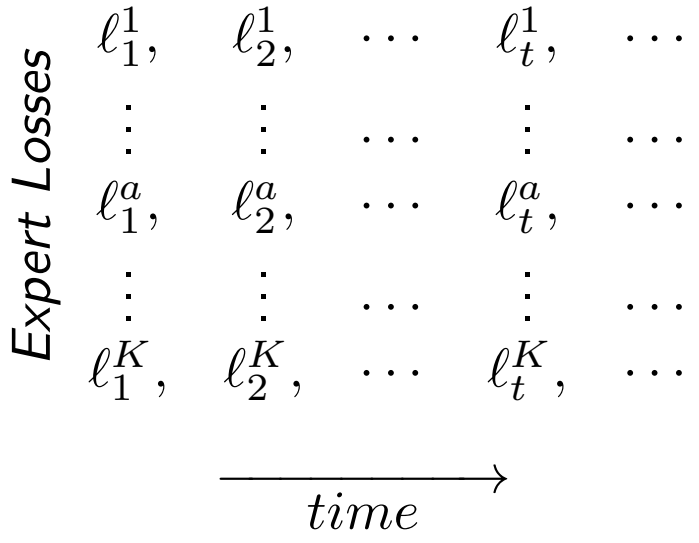
- ℓ_t^a-s are in [0,1] and selected arbitrarily (adversarially)

Prediction with Expert Advice

Game Definition

For $t = 1, 2, \ldots$:

1. Pick a row A_t
2. Observe $\ell_t^1, \ldots, \ell_t^K$ & suffer $\ell_t^{A_t}$



Performance Measure: Regret

$$R_T = \sum_{t=1}^{T} \ell_t^{A_t} - \min_a \left(\sum_{t=1}^{T} \ell_t^a \right)$$

Learning Goal

$R_T = o(T)$

Assumptions

- ℓ_t^a-s are in [0,1] and selected arbitrarily (adversarially)

Prediction with Expert Advice

Game Definition

For $t = 1, 2, \ldots$:

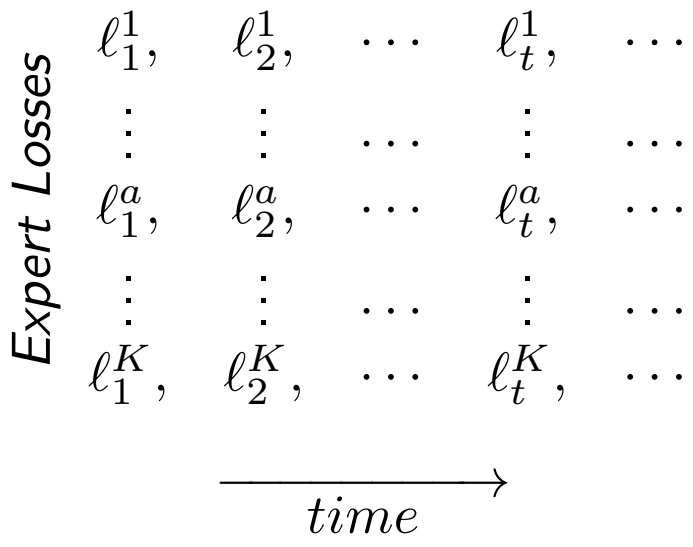
1. Pick a row A_t
2. Observe $\ell_t^1, \ldots, \ell_t^K$ & suffer $\ell_t^{A_t}$

Performance Measure: Regret

$$R_T = \sum_{t=1}^{T} \ell_t^{A_t} - \min_a \left(\sum_{t=1}^{T} \ell_t^a \right)$$

Assumptions

- ℓ_t^a-s are in [0,1] and selected arbitrarily (adversarially)



Learning Goal

$R_T = o(T)$

Why comparing to the best row, not the best path?

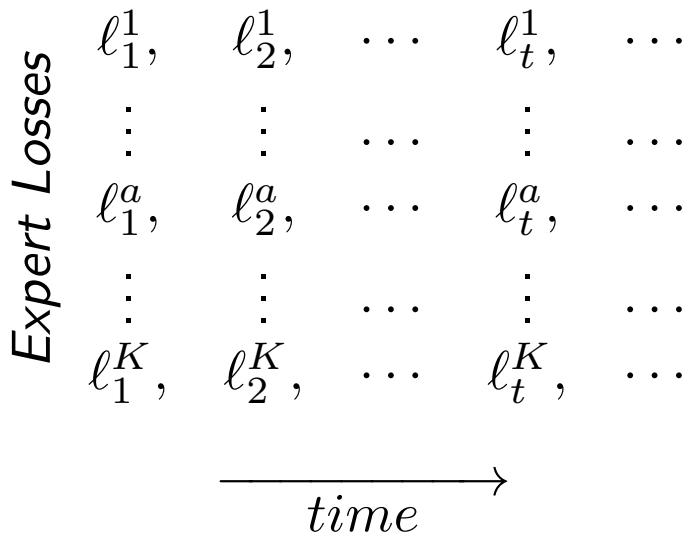
Prediction with Expert Advice

Algorithms

- The algorithm needs a protection against the adversary
- Protection - randomization

Assumptions

- The adversary may know the algorithm, but not the random bits ("oblivious setting" - ℓ_t^a-s are written down before the game starts)



The Hedge Algorithm (a.k.a. Exponential Weights)

[Vovk, 1990, Littlestone & Warmuth, 1994, . . .]

Input: Learning rates $\eta_1 \geq \eta_2 \geq \cdots > 0$

$\forall a : \hat{L}_0(a) = 0$

for $t = 1, 2, \ldots$ **do**

$$\forall a : \; p_t(a) = \frac{e^{-\eta_t \hat{L}_{t-1}(a)}}{\sum_{a'} e^{-\eta_t \hat{L}_{t-1}(a')}}$$

Sample A_t according to p_t and play it

Observe $\ell_t^1, \ldots, \ell_t^K$

$$\forall a : \; \hat{L}_t(a) = \hat{L}_{t-1}(a) + \ell_t^a$$

end

Analysis (simplified for known T and constant η)

Reminder

$$p_t(a) = \frac{e^{-\eta \hat{L}_{t-1}(a)}}{\sum_{a'} e^{-\eta \hat{L}_{t-1}(a')}}$$

$$\hat{L}_t(a) = \hat{L}_{t-1}(a) + \ell_t^a$$

Analysis (simplified for known T and constant η)

Let

$$W_t = \sum_a e^{-\eta \hat{L}_t(a)} = \sum_a e^{-\eta \ell_t^a} e^{-\eta \hat{L}_{t-1}(a)}$$

Reminder

$$p_t(a) = \frac{e^{-\eta \hat{L}_{t-1}(a)}}{\sum_{a'} e^{-\eta \hat{L}_{t-1}(a')}}$$

$$\hat{L}_t(a) = \hat{L}_{t-1}(a) + \ell_t^a$$

Analysis (simplified for known T and constant η)

Let

$$W_t = \sum_a e^{-\eta \hat{L}_t(a)} = \sum_a e^{-\eta \ell_t^a} e^{-\eta \hat{L}_{t-1}(a)}$$

Calculation

$$\frac{W_t}{W_{t-1}} = \sum_a e^{-\eta \ell_t^a} \frac{e^{-\eta \hat{L}_{t-1}(a)}}{\sum_{a'} e^{-\eta \hat{L}_{t-1}(a')}}$$

Reminder

$$p_t(a) = \frac{e^{-\eta \hat{L}_{t-1}(a)}}{\sum_{a'} e^{-\eta \hat{L}_{t-1}(a')}}$$

$$\hat{L}_t(a) = \hat{L}_{t-1}(a) + \ell_t^a$$

Analysis (simplified for known T and constant η)

Let

$$W_t = \sum_a e^{-\eta \hat{L}_t(a)} = \sum_a e^{-\eta \ell_t^a} e^{-\eta \hat{L}_{t-1}(a)}$$

Calculation

$$\begin{aligned} \frac{W_t}{W_{t-1}} &= \sum_a e^{-\eta \ell_t^a} \frac{e^{-\eta \hat{L}_{t-1}(a)}}{\sum_{a'} e^{-\eta \hat{L}_{t-1}(a')}} \\ &= \sum_a e^{-\eta \ell_t^a} p_t(a) \end{aligned}$$

Reminder

$$p_t(a) = \frac{e^{-\eta \hat{L}_{t-1}(a)}}{\sum_{a'} e^{-\eta \hat{L}_{t-1}(a')}}$$

$$\hat{L}_t(a) = \hat{L}_{t-1}(a) + \ell_t^a$$

Analysis (simplified for known T and constant η)

Let

$$W_t = \sum_a e^{-\eta \hat{L}_t(a)} = \sum_a e^{-\eta \ell_t^a} e^{-\eta \hat{L}_{t-1}(a)}$$

Calculation

$$\begin{aligned}\frac{W_t}{W_{t-1}} &= \sum_a e^{-\eta \ell_t^a} \frac{e^{-\eta \hat{L}_{t-1}(a)}}{\sum_{a'} e^{-\eta \hat{L}_{t-1}(a')}} \\ &= \sum_a e^{-\eta \ell_t^a} p_t(a)\end{aligned}$$

Reminder

$$p_t(a) = \frac{e^{-\eta \hat{L}_{t-1}(a)}}{\sum_{a'} e^{-\eta \hat{L}_{t-1}(a')}}$$

$$\hat{L}_t(a) = \hat{L}_{t-1}(a) + \ell_t^a$$

Useful Inequalities

For $x \leq 0$:

$$e^x \leq 1 + x + \frac{1}{2}x^2$$

Analysis (simplified for known T and constant η)

Let

$$W_t = \sum_a e^{-\eta \hat{L}_t(a)} = \sum_a e^{-\eta \ell_t^a} e^{-\eta \hat{L}_{t-1}(a)}$$

Calculation

$$\begin{aligned}
\frac{W_t}{W_{t-1}} &= \sum_a e^{-\eta \ell_t^a} \frac{e^{-\eta \hat{L}_{t-1}(a)}}{\sum_{a'} e^{-\eta \hat{L}_{t-1}(a')}} \\
&= \sum_a e^{-\eta \ell_t^a} p_t(a) \\
&\leq \sum_a \left(1 - \eta \ell_t^a + \frac{1}{2} (\eta \ell_t^a)^2\right) p_t(a)
\end{aligned}$$

Reminder

$$p_t(a) = \frac{e^{-\eta \hat{L}_{t-1}(a)}}{\sum_{a'} e^{-\eta \hat{L}_{t-1}(a')}}$$

$$\hat{L}_t(a) = \hat{L}_{t-1}(a) + \ell_t^a$$

Useful Inequalities

For $x \leq 0$:

$$e^x \leq 1 + x + \frac{1}{2} x^2$$

Analysis (simplified for known T and constant η)

Let

$$W_t = \sum_a e^{-\eta \hat{L}_t(a)} = \sum_a e^{-\eta \ell_t^a} e^{-\eta \hat{L}_{t-1}(a)}$$

Calculation

$$\begin{aligned}
\frac{W_t}{W_{t-1}} &= \sum_a e^{-\eta \ell_t^a} \frac{e^{-\eta \hat{L}_{t-1}(a)}}{\sum_{a'} e^{-\eta \hat{L}_{t-1}(a')}} \\
&= \sum_a e^{-\eta \ell_t^a} p_t(a) \\
&\leq \sum_a \left(1 - \eta \ell_t^a + \frac{1}{2} (\eta \ell_t^a)^2\right) p_t(a) \\
&= 1 - \eta \sum_a \ell_t^a p_t(a) + \frac{\eta^2}{2} \sum_a (\ell_t^a)^2 p_t(a)
\end{aligned}$$

Reminder

$$p_t(a) = \frac{e^{-\eta \hat{L}_{t-1}(a)}}{\sum_{a'} e^{-\eta \hat{L}_{t-1}(a')}}$$

$$\hat{L}_t(a) = \hat{L}_{t-1}(a) + \ell_t^a$$

Useful Inequalities

For $x \leq 0$:

$$e^x \leq 1 + x + \frac{1}{2} x^2$$

Analysis (simplified for known T and constant η)

Let

$$W_t = \sum_a e^{-\eta \hat{L}_t(a)} = \sum_a e^{-\eta \ell_t^a} e^{-\eta \hat{L}_{t-1}(a)}$$

Calculation

$$\begin{aligned}
\frac{W_t}{W_{t-1}} &= \sum_a e^{-\eta \ell_t^a} \frac{e^{-\eta \hat{L}_{t-1}(a)}}{\sum_{a'} e^{-\eta \hat{L}_{t-1}(a')}} \\
&= \sum_a e^{-\eta \ell_t^a} p_t(a) \\
&\leq \sum_a \left(1 - \eta \ell_t^a + \frac{1}{2} (\eta \ell_t^a)^2\right) p_t(a) \\
&= 1 - \eta \sum_a \ell_t^a p_t(a) + \frac{\eta^2}{2} \sum_a (\ell_t^a)^2 p_t(a)
\end{aligned}$$

Reminder

$$p_t(a) = \frac{e^{-\eta \hat{L}_{t-1}(a)}}{\sum_{a'} e^{-\eta \hat{L}_{t-1}(a')}}$$

$$\hat{L}_t(a) = \hat{L}_{t-1}(a) + \ell_t^a$$

Useful Inequalities

For $x \leq 0$:

$$e^x \leq 1 + x + \frac{1}{2} x^2$$

For any x:

$$1 + x \leq e^x$$

Analysis (simplified for known T and constant η)

Let

$$W_t = \sum_a e^{-\eta \hat{L}_t(a)} = \sum_a e^{-\eta \ell_t^a} e^{-\eta \hat{L}_{t-1}(a)}$$

Calculation

$$\begin{aligned}
\frac{W_t}{W_{t-1}} &= \sum_a e^{-\eta \ell_t^a} \frac{e^{-\eta \hat{L}_{t-1}(a)}}{\sum_{a'} e^{-\eta \hat{L}_{t-1}(a')}} \\
&= \sum_a e^{-\eta \ell_t^a} p_t(a) \\
&\leq \sum_a \left(1 - \eta \ell_t^a + \frac{1}{2} (\eta \ell_t^a)^2\right) p_t(a) \\
&= 1 - \eta \sum_a \ell_t^a p_t(a) + \frac{\eta^2}{2} \sum_a (\ell_t^a)^2 p_t(a) \\
&\leq e^{-\eta \sum_a \ell_t^a p_t(a) + \frac{\eta^2}{2} \sum_a (\ell_t^a)^2 p_t(a)}
\end{aligned}$$

Reminder

$$p_t(a) = \frac{e^{-\eta \hat{L}_{t-1}(a)}}{\sum_{a'} e^{-\eta \hat{L}_{t-1}(a')}}$$

$$\hat{L}_t(a) = \hat{L}_{t-1}(a) + \ell_t^a$$

Useful Inequalities

For $x \leq 0$:

$$e^x \leq 1 + x + \frac{1}{2} x^2$$

For any x:

$$1 + x \leq e^x$$

Analysis (simplified for known T and constant η)

From the last slide:

$$W_t = \sum_a e^{-\eta \hat{L}_t(a)}$$

$$\frac{W_t}{W_{t-1}} \leq e^{-\eta \sum_a \ell_t^a p_t(a) + \frac{\eta^2}{2} \sum_a (\ell_t^a)^2 p_t(a)}$$

Analysis (simplified for known T and constant η)

From the last slide:

$$W_t = \sum_a e^{-\eta \hat{L}_t(a)}$$

$$\frac{W_t}{W_{t-1}} \leq e^{-\eta \sum_a \ell_t^a p_t(a) + \frac{\eta^2}{2} \sum_a (\ell_t^a)^2 p_t(a)}$$

Calculation continued:

$$\ln \frac{W_T}{W_0} \leq -\eta \sum_{t=1}^{T} \sum_a \ell_t^a p_t(a) + \frac{\eta^2}{2} \sum_{t=1}^{T} \sum_a (\ell_t^a)^2 p_t(a)$$

Analysis (simplified for known T and constant η)

From the last slide:

$$W_t = \sum_a e^{-\eta \hat{L}_t(a)}$$

$$\frac{W_t}{W_{t-1}} \leq e^{-\eta \sum_a \ell_t^a p_t(a) + \frac{\eta^2}{2} \sum_a (\ell_t^a)^2 p_t(a)}$$

Calculation continued:

$$\ln \frac{W_T}{W_0} \leq -\eta \sum_{t=1}^{T} \sum_a \ell_t^a p_t(a) + \frac{\eta^2}{2} \sum_{t=1}^{T} \sum_a (\ell_t^a)^2 p_t(a)$$

$$\ln \frac{W_T}{W_0} = \ln \frac{\sum_a e^{-\eta \hat{L}_T(a)}}{K}$$

Analysis (simplified for known T and constant η)

From the last slide:

$$W_t = \sum_a e^{-\eta \hat{L}_t(a)}$$

$$\frac{W_t}{W_{t-1}} \leq e^{-\eta \sum_a \ell_t^a p_t(a) + \frac{\eta^2}{2} \sum_a (\ell_t^a)^2 p_t(a)}$$

Calculation continued:

$$\ln \frac{W_T}{W_0} \leq -\eta \sum_{t=1}^{T} \sum_a \ell_t^a p_t(a) + \frac{\eta^2}{2} \sum_{t=1}^{T} \sum_a (\ell_t^a)^2 p_t(a)$$

$$\ln \frac{W_T}{W_0} = \ln \frac{\sum_a e^{-\eta \hat{L}_T(a)}}{K} \geq \ln \frac{\max_a \left(e^{-\eta \hat{L}_T(a)} \right)}{K}$$

Analysis (simplified for known T and constant η)

From the last slide:

$$W_t = \sum_a e^{-\eta \hat{L}_t(a)}$$

$$\frac{W_t}{W_{t-1}} \leq e^{-\eta \sum_a \ell_t^a p_t(a) + \frac{\eta^2}{2} \sum_a (\ell_t^a)^2 p_t(a)}$$

Calculation continued:

$$\ln \frac{W_T}{W_0} \leq -\eta \sum_{t=1}^{T} \sum_a \ell_t^a p_t(a) + \frac{\eta^2}{2} \sum_{t=1}^{T} \sum_a (\ell_t^a)^2 p_t(a)$$

$$\ln \frac{W_T}{W_0} = \ln \frac{\sum_a e^{-\eta \hat{L}_T(a)}}{K} \geq \ln \frac{\max_a \left(e^{-\eta \hat{L}_T(a)} \right)}{K} = -\eta \min_a \left(\hat{L}_T(a) \right) - \ln K$$

Analysis (simplified for known T and constant η)

From the last slide:

$$W_t = \sum_a e^{-\eta \hat{L}_t(a)}$$

$$\frac{W_t}{W_{t-1}} \leq e^{-\eta \sum_a \ell_t^a p_t(a) + \frac{\eta^2}{2} \sum_a (\ell_t^a)^2 p_t(a)}$$

Calculation continued:

$$\ln \frac{W_T}{W_0} \leq -\eta \sum_{t=1}^{T} \sum_a \ell_t^a p_t(a) + \frac{\eta^2}{2} \sum_{t=1}^{T} \sum_a (\ell_t^a)^2 p_t(a)$$

$$\ln \frac{W_T}{W_0} = \ln \frac{\sum_a e^{-\eta \hat{L}_T(a)}}{K} \geq \ln \frac{\max_a \left(e^{-\eta \hat{L}_T(a)} \right)}{K} = -\eta \min_a \left(\hat{L}_T(a) \right) - \ln K$$

$$-\eta \min_a \left(\hat{L}_T(a) \right) - \ln K \leq -\eta \sum_{t=1}^{T} \sum_a \ell_t^a p_t(a) + \frac{\eta^2}{2} \sum_{t=1}^{T} \sum_a (\ell_t^a)^2 p_t(a)$$

Analysis (simplified for known T and constant η)

Calculation Summary

$$\sum_{t=1}^{T} \sum_{a} \ell_t^a p_t(a) \; - \min_a \left(\hat{L}_T(a) \right) \; \leq \frac{\ln K}{\eta} + \frac{\eta}{2} \sum_{t=1}^{T} \sum_{a} \; (\ell_t^a)^2 \; p_t(a)$$

Analysis (simplified for known T and constant η)

Calculation Summary

$$\sum_{t=1}^{T} \underbrace{\sum_{a} \ell_t^a p_t(a)}_{\mathbb{E}\left[\ell_t^{A_t}\right]} - \min_a \left(\hat{L}_T(a)\right) \leq \frac{\ln K}{\eta} + \frac{\eta}{2} \sum_{t=1}^{T} \sum_{a} \left(\ell_t^a\right)^2 p_t(a)$$

Analysis (simplified for known T and constant η)

Calculation Summary

$$\underbrace{\sum_{t=1}^{T} \underbrace{\sum_{a} \ell_t^a p_t(a)}_{\mathbb{E}\left[\ell_t^{A_t}\right]} - \min_a \left(\hat{L}_T(a)\right)}_{\mathbb{E}[R_T]} \leq \frac{\ln K}{\eta} + \frac{\eta}{2} \sum_{t=1}^{T} \sum_{a} \left(\ell_t^a\right)^2 p_t(a)$$

Analysis (simplified for known T and constant η)

Calculation Summary

$$\underbrace{\sum_{t=1}^{T} \underbrace{\sum_{a} \ell_t^a p_t(a)}_{\mathbb{E}\left[\ell_t^{A_t}\right]} - \min_a \left(\hat{L}_T(a)\right)}_{\mathbb{E}[R_T]} \leq \frac{\ln K}{\eta} + \frac{\eta}{2} \sum_{t=1}^{T} \sum_{a} \underbrace{(\ell_t^a)^2}_{\leq 1} p_t(a)$$

Analysis (simplified for known T and constant η)

Calculation Summary

$$\underbrace{\sum_{t=1}^{T} \underbrace{\sum_{a} \ell_t^a p_t(a)}_{\mathbb{E}\left[\ell_t^{A_t}\right]} - \min_a \left(\hat{L}_T(a)\right)}_{\mathbb{E}[R_T]} \leq \frac{\ln K}{\eta} + \frac{\eta}{2} \underbrace{\sum_{t=1}^{T} \sum_{a} \underbrace{(\ell_t^a)^2}_{\leq 1} p_t(a)}_{\leq 1}$$

Analysis (simplified for known T and constant η)

Calculation Summary

$$\underbrace{\sum_{t=1}^{T} \underbrace{\sum_{a} \ell_t^a p_t(a)}_{\mathbb{E}\left[\ell_t^{A_t}\right]} - \min_a \left(\hat{L}_T(a)\right)}_{\mathbb{E}[R_T]} \leq \frac{\ln K}{\eta} + \frac{\eta}{2} \underbrace{\sum_{t=1}^{T} \underbrace{\sum_{a} \underbrace{(\ell_t^a)^2}_{\leq 1} p_t(a)}_{\leq 1}}_{\leq T}$$

Analysis (simplified for known T and constant η)

Calculation Summary

$$\underbrace{\sum_{t=1}^{T} \underbrace{\sum_{a} \ell_t^a p_t(a)}_{\mathbb{E}\left[\ell_t^{A_t}\right]} - \min_a \left(\hat{L}_T(a)\right)}_{\mathbb{E}[R_T]} \leq \frac{\ln K}{\eta} + \frac{\eta}{2} \underbrace{\sum_{t=1}^{T} \underbrace{\sum_{a} \underbrace{(\ell_t^a)^2}_{\leq 1} p_t(a)}_{\leq 1}}_{\leq T}$$

Minimize with respect to η

$$\eta = \sqrt{\frac{2 \ln K}{T}}$$

Analysis (simplified for known T and constant η)

Calculation Summary

$$\underbrace{\sum_{t=1}^{T} \underbrace{\sum_{a} \ell_t^a p_t(a)}_{\mathbb{E}\left[\ell_t^{A_t}\right]} - \min_a \left(\hat{L}_T(a)\right)}_{\mathbb{E}[R_T]} \leq \frac{\ln K}{\eta} + \frac{\eta}{2} \underbrace{\sum_{t=1}^{T} \underbrace{\sum_{a} \underbrace{(\ell_t^a)^2}_{\leq 1} p_t(a)}_{\leq 1}}_{\leq T}$$

Minimize with respect to η

$$\eta = \sqrt{\frac{2 \ln K}{T}}$$

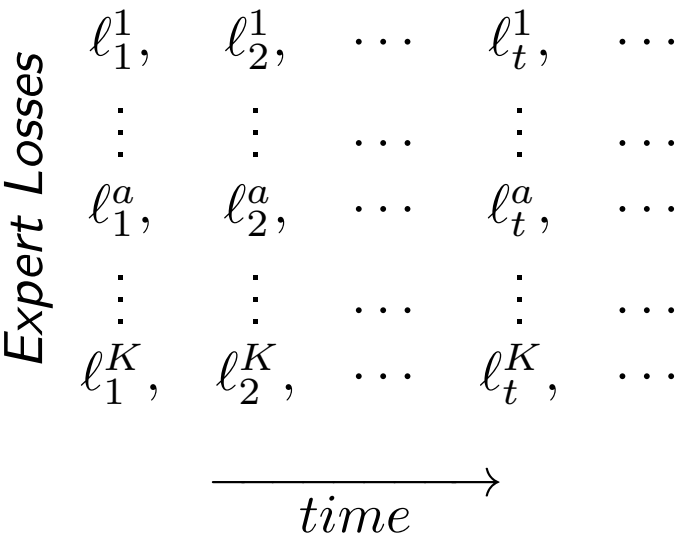
Final Result

$$\mathbb{E}\left[R_T\right] \leq \sqrt{2T \ln K}$$

Lower bound (high-level idea)

Construction

ℓ_t^a-s independent Bernoulli with bias $\frac{1}{2}$



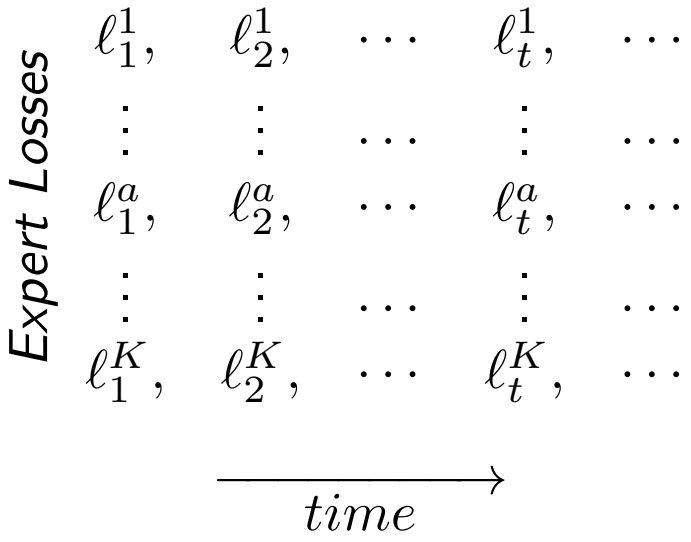
Lower bound (high-level idea)

Construction

ℓ_t^a-s independent Bernoulli with bias $\frac{1}{2}$

Lemma

$$\lim_{K\to\infty, T\to\infty} \frac{T/2 - \mathbb{E}\left[\min_a \left(\hat{L}_T(a)\right)\right]}{\sqrt{\frac{1}{2}T\ln K}} = 1$$



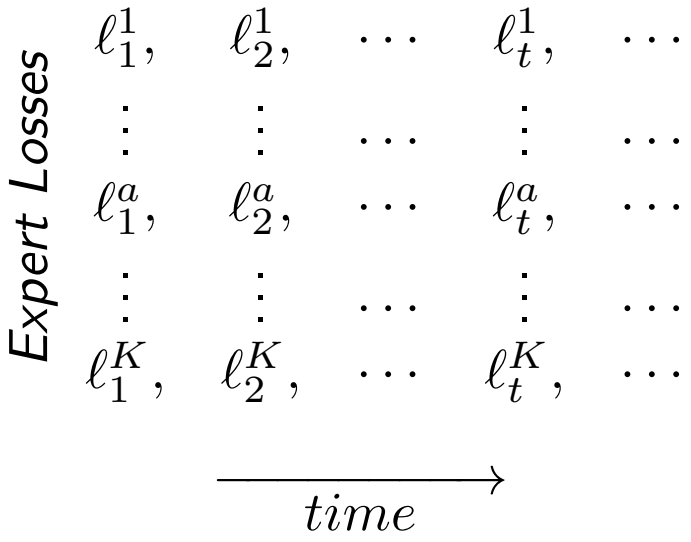
Lower bound (high-level idea)

Construction

ℓ_t^a-s independent Bernoulli with bias $\frac{1}{2}$

Lemma

$$\lim_{K\to\infty, T\to\infty} \frac{\overbrace{T/2 - \mathbb{E}\left[\min_a \left(\hat{L}_T(a)\right)\right]}^{\mathbb{E}[R_T]}}{\sqrt{\frac{1}{2}T\ln K}} = 1$$



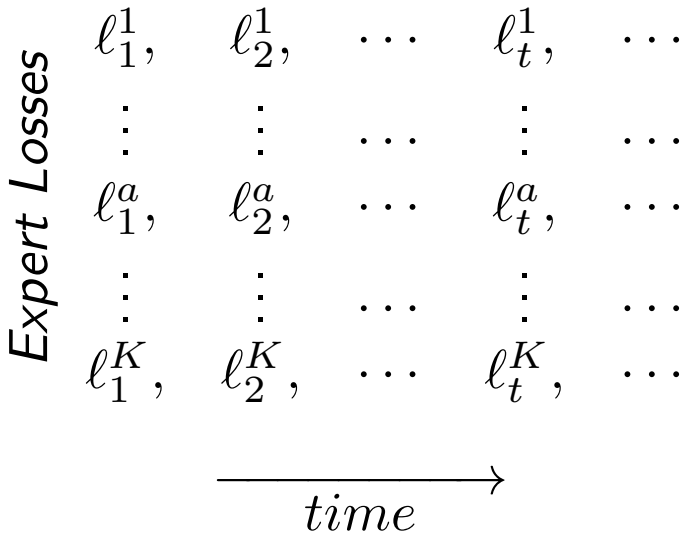
Lower bound (high-level idea)

Construction

ℓ_t^a-s independent Bernoulli with bias $\frac{1}{2}$

Lemma

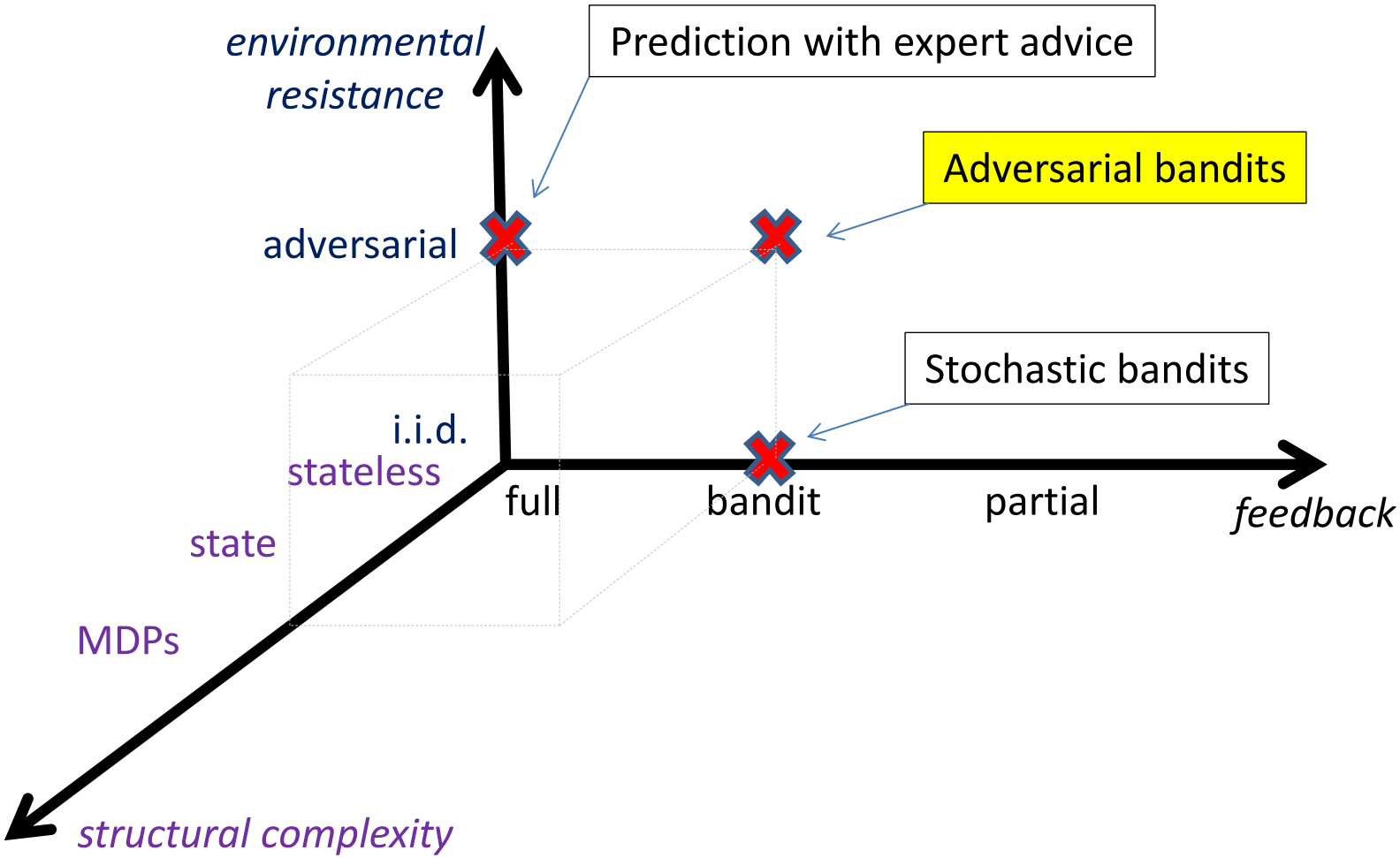
$$\lim_{K\to\infty, T\to\infty} \frac{\overbrace{T/2 - \mathbb{E}\left[\min_a \left(\hat{L}_T(a)\right)\right]}^{\mathbb{E}[R_T]}}{\sqrt{\frac{1}{2}T\ln K}} = 1$$



Conclusion

$$\mathbb{E}\left[R_T\right] = \Omega\left(\sqrt{T\ln K}\right)$$

Part I: “classical” algorithms

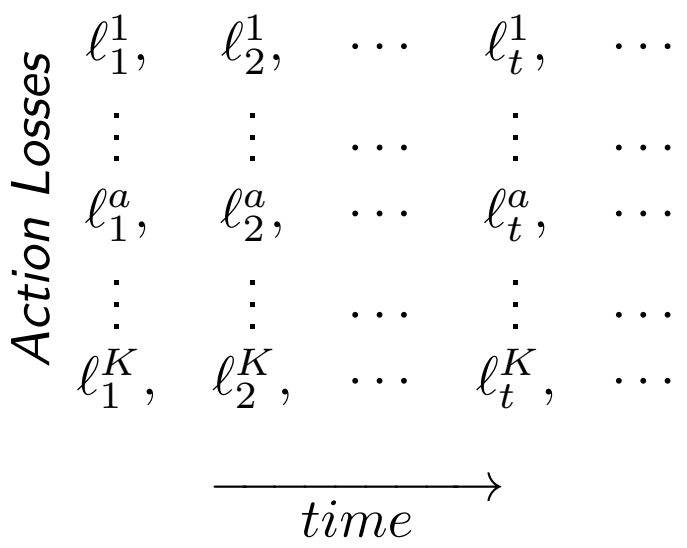


Adversarial Multiarmed Bandits

Game Definition

For $t = 1, 2, \ldots$:

1. Play an action A_t
2. Observe and suffer the loss $\ell_t^{A_t}$



Performance Measure: Regret

$$R_T = \sum_{t=1}^{T} \ell_t^{A_t} - \min_a \left(\sum_{t=1}^{T} \ell_t^a \right)$$

Reminder: The Hedge Algorithm

Input: Learning rates $\eta_1 \geq \eta_2 \geq \cdots > 0$

$\forall a : \hat{L}_0(a) = 0$

for $t = 1, 2, \ldots$ **do**

$$\forall a : \; p_t(a) = \frac{e^{-\eta_t \hat{L}_{t-1}(a)}}{\sum_{a'} e^{-\eta_t \hat{L}_{t-1}(a')}}$$

Sample A_t according to p_t and play it

Observe $\ell_t^1, \ldots, \ell_t^K$

$$\forall a : \; \hat{L}_t(a) = \hat{L}_{t-1}(a) + \ell_t^a$$

end

The EXP3 Algorithm for Adversarial Bandits

[Auer et. al., 2002; Bubeck, 2010]

Input: Learning rates $\eta_1 \geq \eta_2 \geq \cdots > 0$

$\forall a : \tilde{L}_0(a) = 0$

for $t = 1, 2, \ldots$ **do**

$$\forall a : \; p_t(a) = \frac{e^{-\eta_t \tilde{L}_{t-1}(a)}}{\sum_{a'} e^{-\eta_t \tilde{L}_{t-1}(a')}}$$

Sample A_t according to p_t and play it

Observe $\ell_t^{A_t}$

$$\forall a : \; \tilde{\ell}_t^a = \frac{\ell_t^a \mathbb{1}_{\{A_t = a\}}}{p_t(a)} = \begin{cases} \frac{\ell_t^a}{p_t(a)}, & \text{if } A_t = a \\ 0, & \text{otherwise} \end{cases}$$

Importance-weighted sampling

$$\forall a : \; \tilde{L}_t(a) = \tilde{L}_{t-1}(a) + \tilde{\ell}_t^a$$

end

Properties of Importance-Weighted Sampling

Notation

$$\mathbb{E}_t\left[\cdot\right] = \mathbb{E}\left[\cdot | \text{everything up to round } t\right]$$

Expectation

$$\mathbb{E}_t\left[\tilde{\ell}_t^a\right] = \mathbb{E}_t\left[\frac{\ell_t^a \mathbb{1}_{\{A_t=a\}}}{p_t(a)}\right] = \frac{\ell_t^a \overbrace{\mathbb{E}_t\left[\mathbb{1}_{\{A_t=a\}}\right]}^{=p_t(a)}}{p_t(a)} = \ell_t^a$$

Properties of Importance-Weighted Sampling

Second moment

$$\mathbb{E}_t\left[\left(\tilde{\ell}_t^a\right)^2\right] = \frac{\overbrace{\left(\ell_t^a\right)^2}^{\leq 1}\,\mathbb{E}_t\Big[\overbrace{\left(\mathbb{1}_{\{A_t=a\}}\right)^2}^{=\mathbb{1}_{\{A_t=a\}}}\Big]}{p_t(a)^2} \leq \frac{1}{p_t(a)}$$

Properties of Importance-Weighted Sampling

Second moment

$$\mathbb{E}_t\left[\left(\tilde{\ell}_t^a\right)^2\right] = \frac{\overbrace{\left(\ell_t^a\right)^2}^{\leq 1}\,\mathbb{E}_t\left[\overbrace{\left(\mathbb{1}_{\{A_t=a\}}\right)^2}^{=\mathbb{1}_{\{A_t=a\}}}\right]}{p_t(a)^2} \leq \frac{1}{p_t(a)}$$

$$\mathbb{E}_t\left[\sum_a p_t(a)\left(\tilde{\ell}_t^a\right)^2\right] \leq K$$

Analysis (simplified for known T and constant η)

Following the calculations in the analysis of Hedge we have:

$$\sum_{t=1}^{T}\sum_{a}\tilde{\ell}_t^a p_t(a) - \min_a\left(\tilde{L}_T(a)\right) \leq \frac{\ln K}{\eta} + \frac{\eta}{2}\sum_{t=1}^{T}\sum_{a}\left(\tilde{\ell}_t^a\right)^2 p_t(a)$$

Analysis (simplified for known T and constant η)

Following the calculations in the analysis of Hedge we have:

$$\sum_{t=1}^{T}\sum_{a}\tilde{\ell}_t^a p_t(a) - \min_a\left(\tilde{L}_T(a)\right) \leq \frac{\ln K}{\eta} + \frac{\eta}{2}\sum_{t=1}^{T}\sum_{a}\left(\tilde{\ell}_t^a\right)^2 p_t(a)$$

Taking expectation on both sides

$$\mathbb{E}\left[\sum_{t=1}^{T}\sum_{a}\tilde{\ell}_t^a p_t(a)\right] - \mathbb{E}\left[\min_a\left(\tilde{L}_T(a)\right)\right] \leq \frac{\ln K}{\eta} + \frac{\eta}{2}\mathbb{E}\left[\sum_{t=1}^{T}\sum_{a}\left(\tilde{\ell}_t^a\right)^2 p_t(a)\right]$$

Analysis (simplified for known T and constant η)

Following the calculations in the analysis of Hedge we have:

$$\sum_{t=1}^{T}\sum_{a}\tilde{\ell}_t^a p_t(a) - \min_a\left(\tilde{L}_T(a)\right) \leq \frac{\ln K}{\eta} + \frac{\eta}{2}\sum_{t=1}^{T}\sum_{a}\left(\tilde{\ell}_t^a\right)^2 p_t(a)$$

Taking expectation on both sides

$$\mathbb{E}\left[\sum_{t=1}^{T}\sum_{a}\tilde{\ell}_t^a p_t(a)\right] - \mathbb{E}\left[\min_a\left(\tilde{L}_T(a)\right)\right] \leq \frac{\ln K}{\eta} + \frac{\eta}{2}\mathbb{E}\left[\sum_{t=1}^{T}\sum_{a}\left(\tilde{\ell}_t^a\right)^2 p_t(a)\right]$$

Propagating the expectations inside ($\mathbb{E}\left[\min(\cdot)\right] \leq \min\mathbb{E}\left[\cdot\right]$)

$$\mathbb{E}\left[\sum_{t=1}^{T}\sum_{a}\mathbb{E}_t\left[\tilde{\ell}_t^a\right] p_t(a)\right] - \min_a\mathbb{E}\left[\tilde{L}_T(a)\right] \leq \frac{\ln K}{\eta} + \frac{\eta}{2}\mathbb{E}\left[\sum_{t=1}^{T}\sum_{a}\mathbb{E}_t\left[\left(\tilde{\ell}_t^a\right)^2\right] p_t(a)\right]$$

Analysis (simplified for known T and constant η)

Following the calculations in the analysis of Hedge we have:

$$\sum_{t=1}^{T}\sum_{a}\tilde{\ell}_t^a p_t(a) - \min_a\left(\tilde{L}_T(a)\right) \leq \frac{\ln K}{\eta} + \frac{\eta}{2}\sum_{t=1}^{T}\sum_{a}\left(\tilde{\ell}_t^a\right)^2 p_t(a)$$

Taking expectation on both sides

$$\mathbb{E}\left[\sum_{t=1}^{T}\sum_{a}\tilde{\ell}_t^a p_t(a)\right] - \mathbb{E}\left[\min_a\left(\tilde{L}_T(a)\right)\right] \leq \frac{\ln K}{\eta} + \frac{\eta}{2}\mathbb{E}\left[\sum_{t=1}^{T}\sum_{a}\left(\tilde{\ell}_t^a\right)^2 p_t(a)\right]$$

Propagating the expectations inside ($\mathbb{E}\left[\min(\cdot)\right] \leq \min\mathbb{E}\left[\cdot\right]$)

$$\mathbb{E}\Bigg[\underbrace{\sum_{t=1}^{T}\sum_{a}\underbrace{\mathbb{E}_t\left[\tilde{\ell}_t^a\right]}_{\ell_t^a} p_t(a)}_{\mathbb{E}\left[\ell_t^{A_t}\right]}\Bigg] - \min_a\underbrace{\mathbb{E}\left[\tilde{L}_T(a)\right]}_{L_T(a)} \leq \frac{\ln K}{\eta} + \frac{\eta}{2}\mathbb{E}\Bigg[\underbrace{\sum_{t=1}^{T}\sum_{a}\mathbb{E}_t\left[\left(\tilde{\ell}_t^a\right)^2\right] p_t(a)}_{\leq K}\Bigg]$$

Analysis (simplified for known T and constant η)

Following the calculations in the analysis of Hedge we have:

$$\sum_{t=1}^{T}\sum_{a}\tilde{\ell}_t^a p_t(a) - \min_a\left(\tilde{L}_T(a)\right) \leq \frac{\ln K}{\eta} + \frac{\eta}{2}\sum_{t=1}^{T}\sum_{a}\left(\tilde{\ell}_t^a\right)^2 p_t(a)$$

Taking expectation on both sides

$$\mathbb{E}\left[\sum_{t=1}^{T}\sum_{a}\tilde{\ell}_t^a p_t(a)\right] - \mathbb{E}\left[\min_a\left(\tilde{L}_T(a)\right)\right] \leq \frac{\ln K}{\eta} + \frac{\eta}{2}\mathbb{E}\left[\sum_{t=1}^{T}\sum_{a}\left(\tilde{\ell}_t^a\right)^2 p_t(a)\right]$$

Propagating the expectations inside ($\mathbb{E}\left[\min(\cdot)\right] \leq \min\mathbb{E}\left[\cdot\right]$)

$$\underbrace{\mathbb{E}\left[\sum_{t=1}^{T}\sum_{a}\mathbb{E}_t\left[\tilde{\ell}_t^a\right]p_t(a)\right] - \min_a\mathbb{E}\left[\tilde{L}_T(a)\right]}_{\mathbb{E}[R_T]} \leq \frac{\ln K}{\eta} + \frac{\eta}{2}\underbrace{\mathbb{E}\left[\sum_{t=1}^{T}\sum_{a}\mathbb{E}_t\left[\left(\tilde{\ell}_t^a\right)^2\right]p_t(a)\right]}_{\leq KT}$$

Analysis (simplified for known T and constant η)

Calculation summary

$$\mathbb{E}\left[R_T\right] \leq \frac{\ln K}{\eta} + \frac{\eta}{2} K T$$

Analysis (simplified for known T and constant η)

Calculation summary

$$\mathbb{E}\left[R_T\right] \leq \frac{\ln K}{\eta} + \frac{\eta}{2} KT$$

Minimize with respect to η

$$\eta = \sqrt{\frac{2 \ln K}{KT}}$$

Analysis (simplified for known T and constant η)

Calculation summary

$$\mathbb{E}\left[R_T\right] \leq \frac{\ln K}{\eta} + \frac{\eta}{2} KT$$

Minimize with respect to η

$$\eta = \sqrt{\frac{2 \ln K}{KT}}$$

Final Result

$$\mathbb{E}\left[R_T\right] \leq \sqrt{2KT \ln K}$$

Analysis (simplified for known T and constant η)

Calculation summary

$$\mathbb{E}\left[R_T\right] \leq \frac{\ln K}{\eta} + \frac{\eta}{2} KT$$

Minimize with respect to η

$$\eta = \sqrt{\frac{2 \ln K}{KT}}$$

Final Result

$$\mathbb{E}\left[R_T\right] \leq \sqrt{2KT \ln K}$$

In comparison with full information we got the extra K factor

Analysis (simplified for known T and constant η)

Calculation summary

$$\mathbb{E}\left[R_T\right] \leq \frac{\ln K}{\eta} + \frac{\eta}{2} KT$$

Minimize with respect to η

$$\eta = \sqrt{\frac{2 \ln K}{KT}}$$

Final Result

$$\mathbb{E}\left[R_T\right] \leq \sqrt{2KT \ln K}$$

It is possible to eliminate $\ln K$ with more sophisticated algorithms

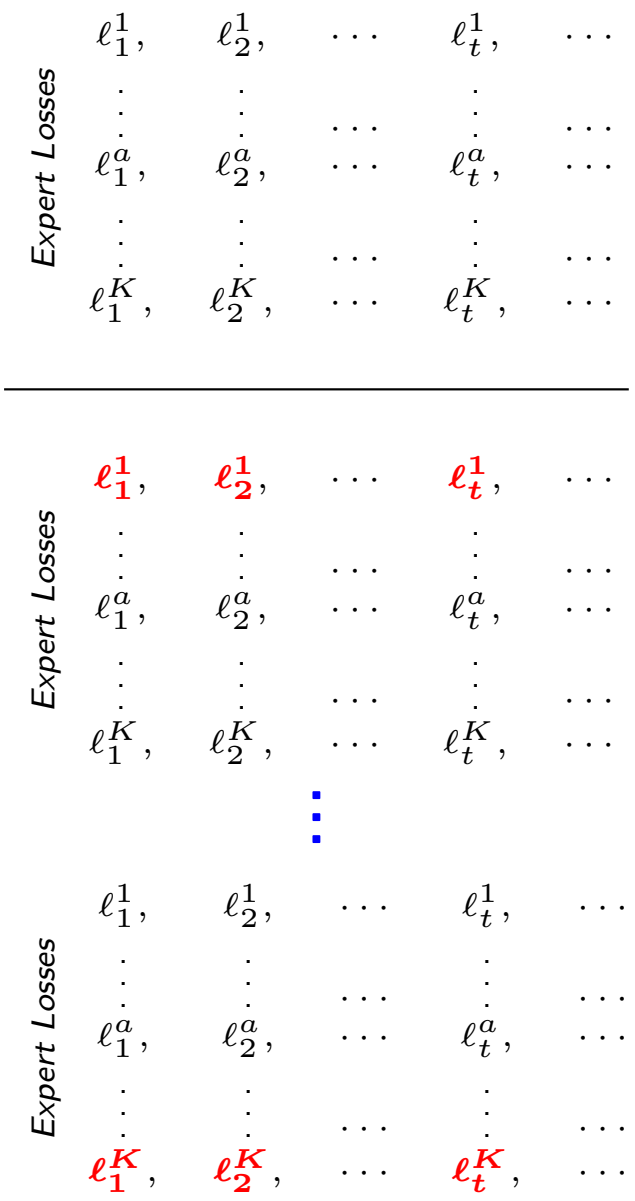
Lower bound (high level idea)

Construct $K+1$ games

0-th game: ℓ_t^a-s Bernoulli with bias $\frac{1}{2}$

i-**th game:** ℓ_t^i-s Bernoulli with bias $\frac{1}{2}+\varepsilon$

For $a \neq i$: ℓ_t^a-s Bernoulli with bias $\frac{1}{2}-\varepsilon$



Lower bound (high level idea)

Construct $K+1$ games

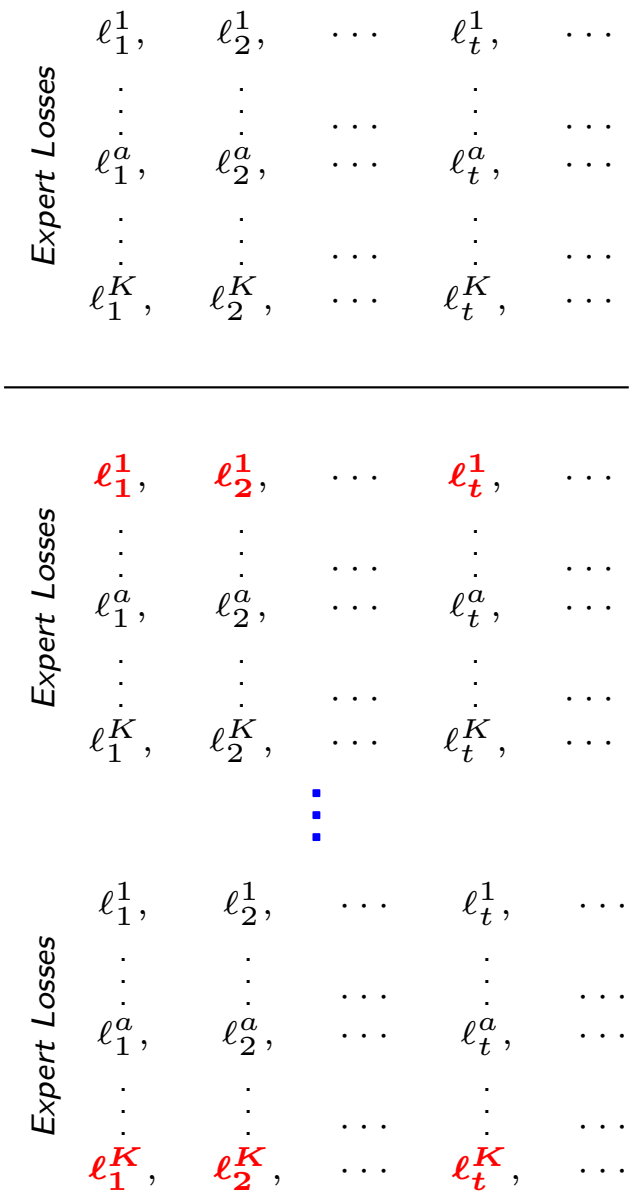
0-th game: ℓ_t^a-s Bernoulli with bias $\frac{1}{2}$

i-th game: ℓ_t^i-s Bernoulli with bias $\frac{1}{2}+\varepsilon$

For $a \neq i$: ℓ_t^a-s Bernoulli with bias $\frac{1}{2}-\varepsilon$

Claim

For small ε, 0-th game is indistinguishable from i-th game based on T observations.



Lower bound (high level idea)

Construct $K+1$ games

0-th game: ℓ_t^a-s Bernoulli with bias $\frac{1}{2}$

i-th game: ℓ_t^i-s Bernoulli with bias $\frac{1}{2}+\varepsilon$

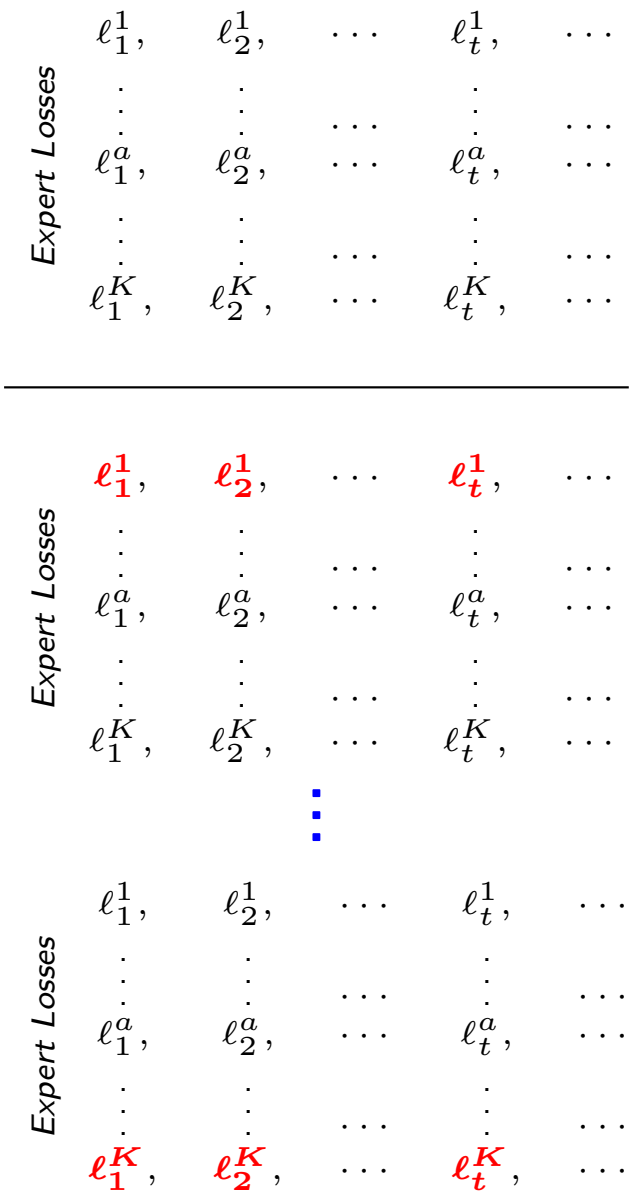
For $a \neq i$: ℓ_t^a-s Bernoulli with bias $\frac{1}{2}-\varepsilon$

Claim

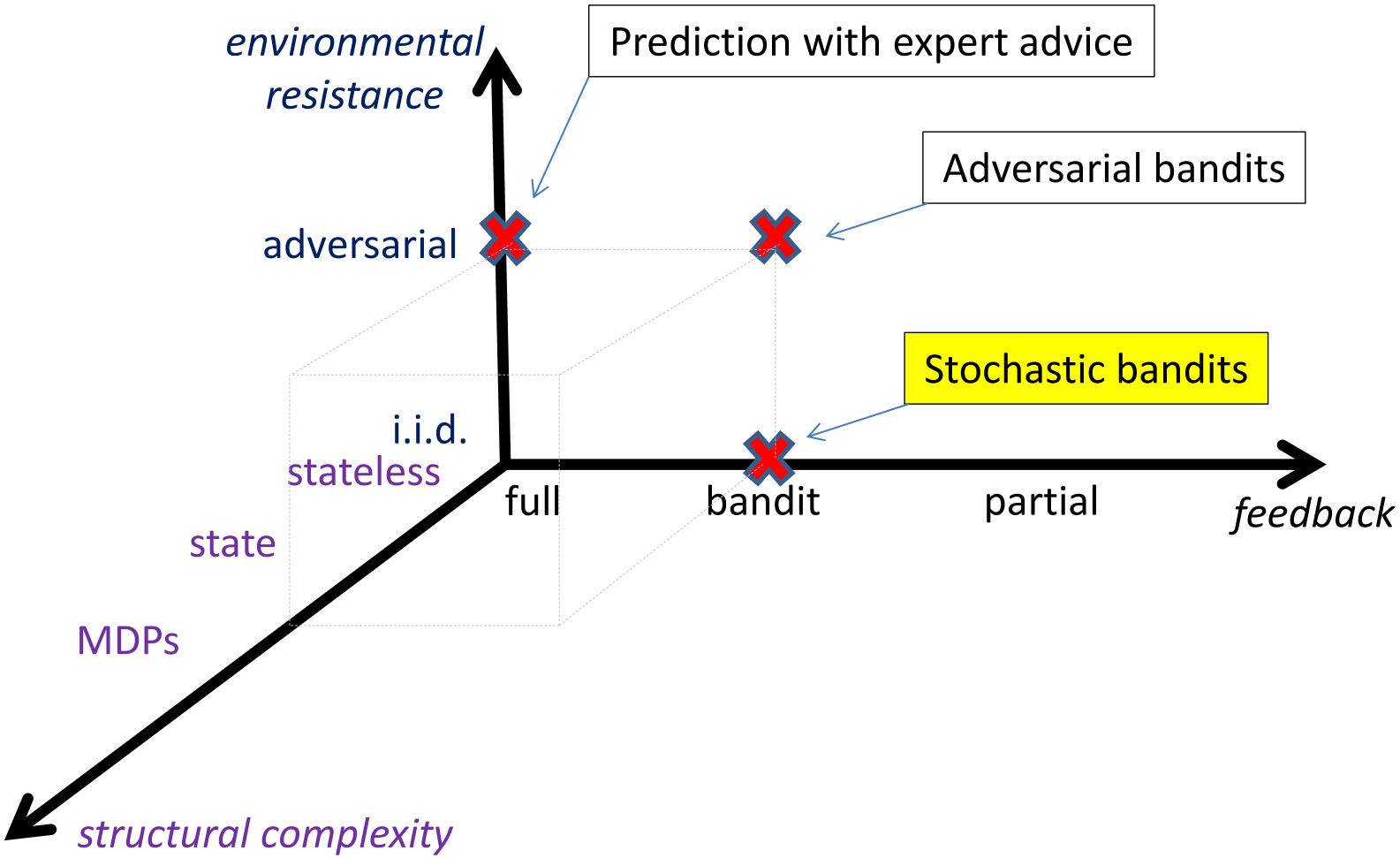
For small ε, 0-th game is indistinguishable from i-th game based on T observations.

For $\varepsilon = \theta\left(\sqrt{K/T}\right)$:

$$\mathbb{E}\left[R_T\right] = \Omega\left(\sqrt{KT}\right)$$



Part I: “classical” algorithms



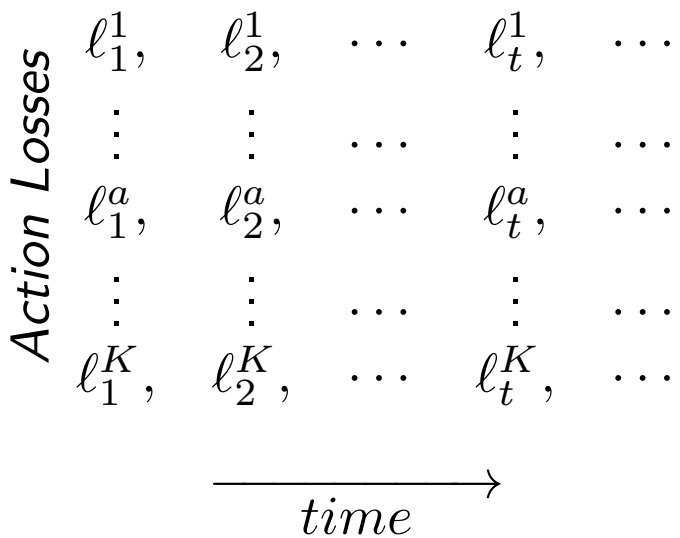
Stochastic Multiarmed Bandits

Game Definition

ℓ_t^a-s independent; $\mathbb{E}\left[\ell_t^a\right] = \mu(a)$

For $t = 1, 2, \ldots$:

1. Play an action A_t
2. Observe and suffer $\ell_t^{A_t}$



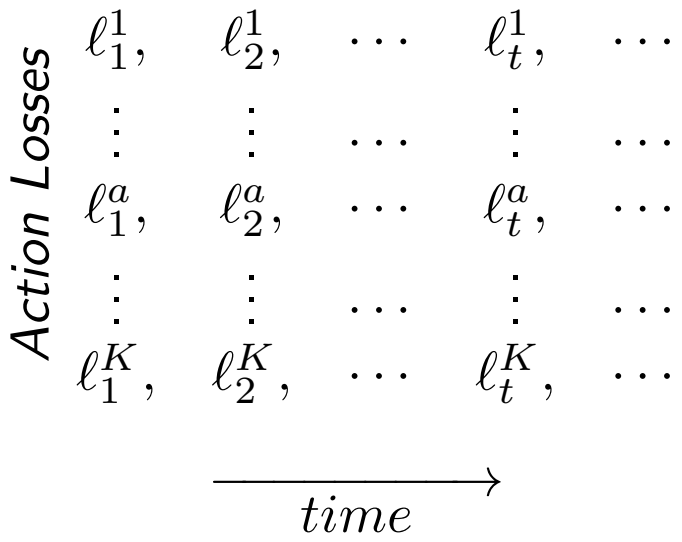
Stochastic Multiarmed Bandits

Game Definition

ℓ_t^a-s independent; $\mathbb{E}\left[\ell_t^a\right] = \mu(a)$

For $t = 1, 2, \ldots$:

1. Play an action A_t
2. Observe and suffer $\ell_t^{A_t}$



Notations

Gaps: $\Delta(a) = \mu(a) - \min_{a'} \mu(a')$

$N_t(a)$ - the number of times a was played up to round t

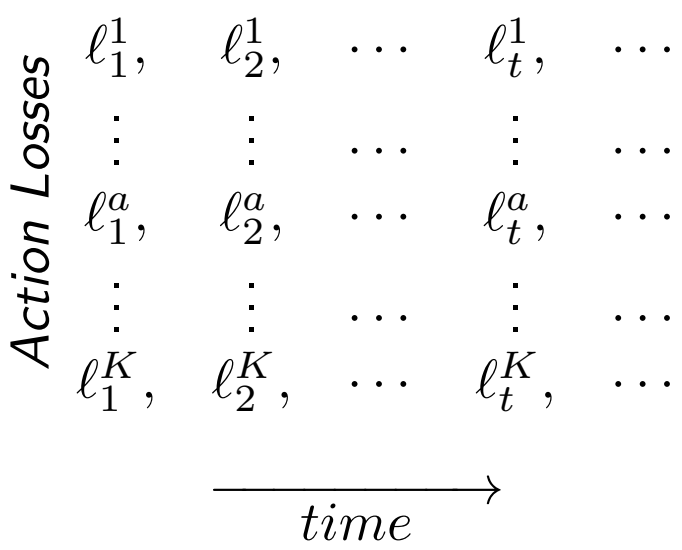
Stochastic Multiarmed Bandits

Game Definition

ℓ_t^a-s independent; $\mathbb{E}\left[\ell_t^a\right] = \mu(a)$

For $t = 1, 2, \ldots$:

1. Play an action A_t
2. Observe and suffer $\ell_t^{A_t}$



Notations

Gaps: $\Delta(a) = \mu(a) - \min_{a'} \mu(a')$

$N_t(a)$ - the number of times a was played up to round t

Performance: Expected Regret

$$\mathbb{E}\left[R_T\right] = \mathbb{E}\left[\sum_{t=1}^{T} \ell_t^{A_t}\right] - T \min_a \mu(a)$$

$$= \sum_a N_T(a)\Delta(a)$$

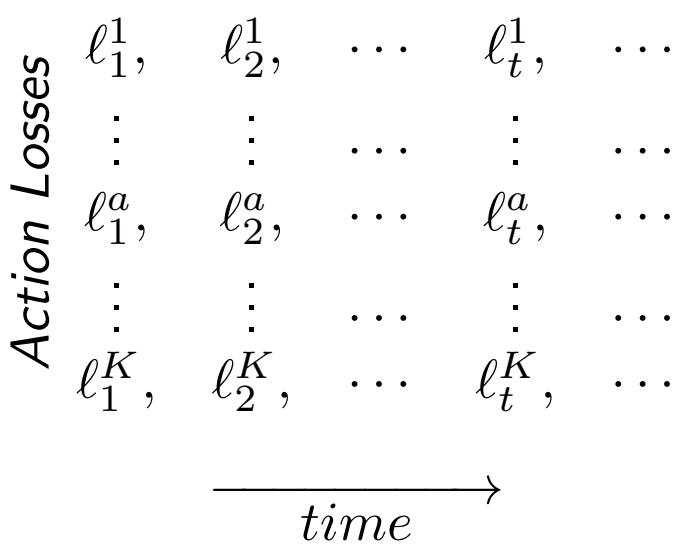
Stochastic Multiarmed Bandits

Game Definition

ℓ_t^a-s independent; $\mathbb{E}\left[\ell_t^a\right] = \mu(a)$

For $t = 1, 2, \ldots$:

1. Play an action A_t
2. Observe and suffer $\ell_t^{A_t}$



Notations

Gaps: $\Delta(a) = \mu(a) - \min_{a'} \mu(a')$

$N_t(a)$ - the number of times a was played up to round t

Performance: Expected Regret

$$\mathbb{E}\left[R_T\right] = \mathbb{E}\left[\sum_{t=1}^{T} \ell_t^{A_t}\right] - T \min_a \mu(a)$$
$$= \sum_a N_T(a)\Delta(a)$$

Historical remark

Originally formulated with gains $r_t^a = 1 - \ell_t^a$

LCB1 (Lower Confidence Bound) Algorithm

Originally UCB1 (Upper Confidence Bound), [Auer et al., 2002]

Initialization: Play each arm once.

for $t = 1, 2, \ldots$ **do**

Let $\hat{L}_t(a)$ - average loss of a up to t

Play $A_t = \arg\min_a \underbrace{\hat{L}_t(a) - \sqrt{\frac{2 \ln t}{N_t(a)}}}_{LCB(a)}$

end

Hoeffding's inequality and LCB

Hoeffding's inequality (simplified)

Let $X_1, \ldots, X_N$ be i.i.d. with $\mathbb{E}[X_i] = \mu$. Then:

$$\mathbb{P}\left\{ \frac{1}{N} \sum_{i=1}^{N} X_i - \mu \geq \sqrt{\frac{2 \ln t}{N}} \right\} \leq \frac{1}{t^4}$$

$$\mathbb{P}\left\{ \mu - \frac{1}{N} \sum_{i=1}^{N} X_i \geq \sqrt{\frac{2 \ln t}{N}} \right\} \leq \frac{1}{t^4}$$

Hoeffding's inequality and LCB

Hoeffding's inequality (simplified)

Let $X_1, \ldots, X_N$ be i.i.d. with $\mathbb{E}[X_i] = \mu$. Then:

$$\mathbb{P}\left\{\frac{1}{N}\sum_{i=1}^{N} X_i - \mu \geq \sqrt{\frac{2\ln t}{N}}\right\} \leq \frac{1}{t^4}$$

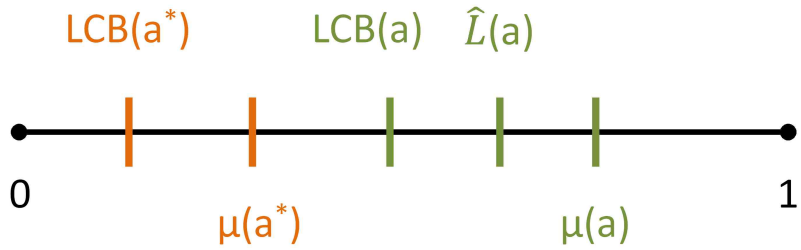
$$\mathbb{P}\left\{\mu - \frac{1}{N}\sum_{i=1}^{N} X_i \geq \sqrt{\frac{2\ln t}{N}}\right\} \leq \frac{1}{t^4}$$

Key properties of LCB (Optimism in the face of uncertainty)

$$LCB(a) = \hat{L}_t(a) - \sqrt{\frac{2\ln t}{N_t(a)}}$$

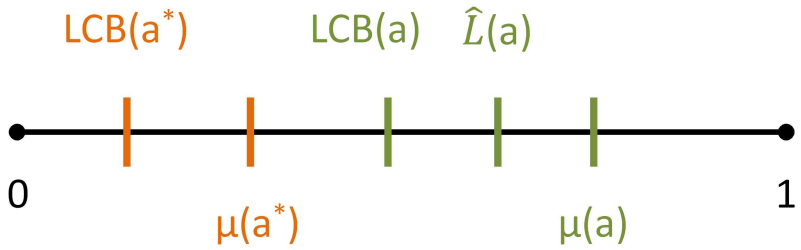
- $\mathbb{E}\left[\hat{L}_t(a)\right] = \mu(a)$
- With prob. $\geq 1 - \frac{1}{t^4}$: $LCB(a) \leq \mu(a)$

LCB1 Analysis Highlights (for two arms, a^* and a)



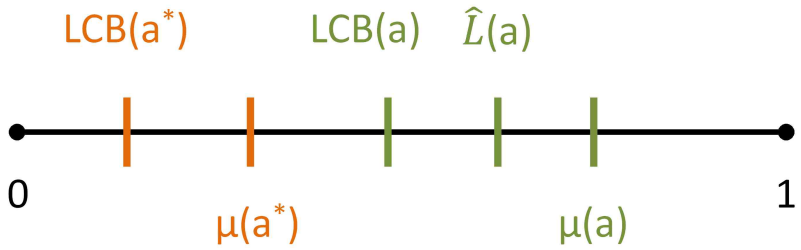
- $\Delta = \mu(a) - \mu(a^*)$

LCB1 Analysis Highlights (for two arms, a^* and a)



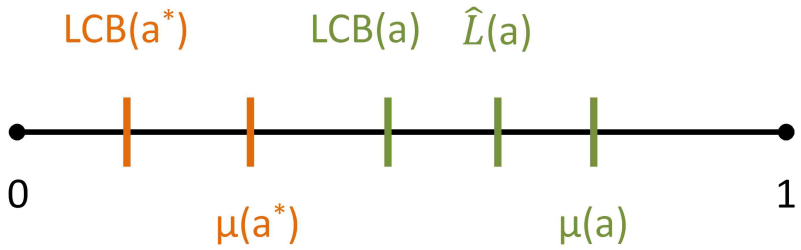
- $\Delta = \mu(a) - \mu(a^*)$
- By Hoeffding, w.h.p.: $\hat{L}_t(a) \geq \mu(a) - \sqrt{\frac{2 \ln t}{N_t(a)}}$

LCB1 Analysis Highlights (for two arms, a^* and a)



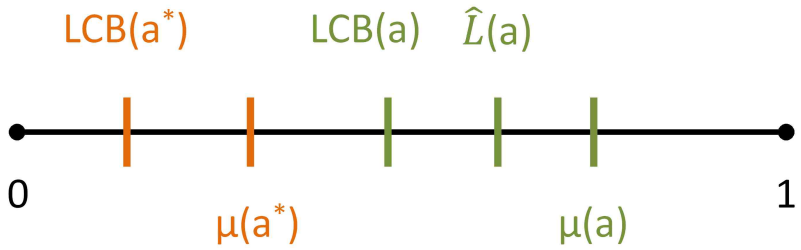
- $\Delta = \mu(a) - \mu(a^*)$
- By Hoeffding, w.h.p.: $\hat{L}_t(a) \geq \mu(a) - \sqrt{\frac{2 \ln t}{N_t(a)}}$
- Which implies: $LCB(a) = \hat{L}_t(a) - \sqrt{\frac{2 \ln t}{N_t(a)}} \geq \mu(a) - 2\sqrt{\frac{2 \ln t}{N_t(a)}}$

LCB1 Analysis Highlights (for two arms, a^* and a)



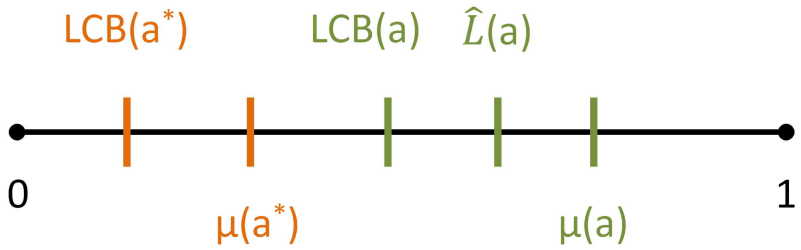
- $\Delta = \mu(a) - \mu(a^*)$
- By Hoeffding, w.h.p.: $\hat{L}_t(a) \geq \mu(a) - \sqrt{\frac{2 \ln t}{N_t(a)}}$
- Which implies: $LCB(a) = \hat{L}_t(a) - \sqrt{\frac{2 \ln t}{N_t(a)}} \geq \mu(a) - 2\sqrt{\frac{2 \ln t}{N_t(a)}}$
- Also, w.h.p.: $LCB(a^*) \leq \mu(a^*)$

LCB1 Analysis Highlights (for two arms, a^* and a)



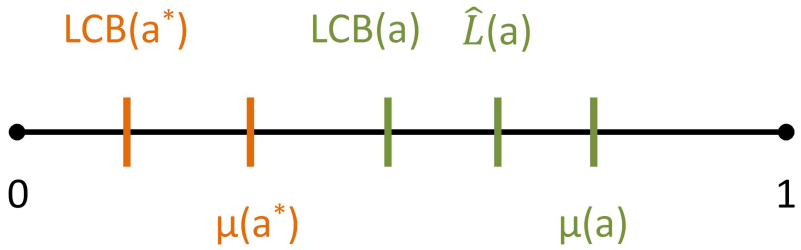
- $\Delta = \mu(a) - \mu(a^*)$
- By Hoeffding, w.h.p.: $\hat{L}_t(a) \geq \mu(a) - \sqrt{\frac{2 \ln t}{N_t(a)}}$
- Which implies: $LCB(a) = \hat{L}_t(a) - \sqrt{\frac{2 \ln t}{N_t(a)}} \geq \mu(a) - 2\sqrt{\frac{2 \ln t}{N_t(a)}}$
- Also, w.h.p.: $LCB(a^*) \leq \mu(a^*)$
- In expectation, the number of rounds on which either of the two confidence bounds fails is bounded by a constant (**careful here**)

LCB1 Analysis Highlights (for two arms, a^* and a)



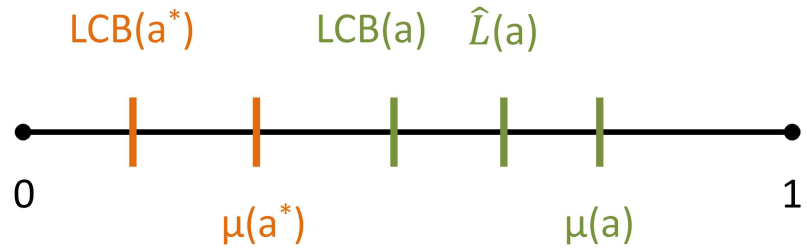
- $\Delta = \mu(a) - \mu(a^*)$
- By Hoeffding, w.h.p.: $\hat{L}_t(a) \geq \mu(a) - \sqrt{\frac{2 \ln t}{N_t(a)}}$
- Which implies: $LCB(a) = \hat{L}_t(a) - \sqrt{\frac{2 \ln t}{N_t(a)}} \geq \mu(a) - 2\sqrt{\frac{2 \ln t}{N_t(a)}}$
- Also, w.h.p.: $LCB(a^*) \leq \mu(a^*)$
- In expectation, the number of rounds on which either of the two confidence bounds fails is bounded by a constant (**careful here**)
- Fix time horizon T

LCB1 Analysis Highlights (for two arms, a^* and a)



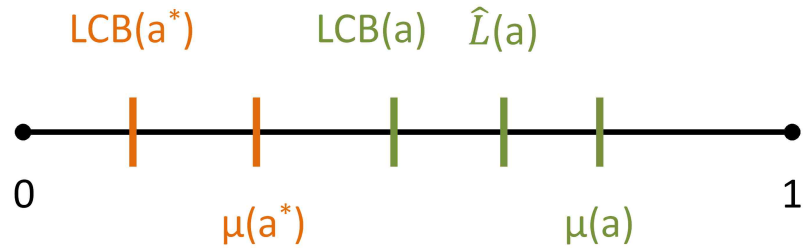
- $\Delta = \mu(a) - \mu(a^*)$
- By Hoeffding, w.h.p.: $\hat{L}_t(a) \geq \mu(a) - \sqrt{\frac{2 \ln t}{N_t(a)}}$
- Which implies: $LCB(a) = \hat{L}_t(a) - \sqrt{\frac{2 \ln t}{N_t(a)}} \geq \mu(a) - 2\sqrt{\frac{2 \ln t}{N_t(a)}}$
- Also, w.h.p.: $LCB(a^*) \leq \mu(a^*)$
- In expectation, the number of rounds on which either of the two confidence bounds fails is bounded by a constant (**careful here**)
- Fix time horizon T
- Once $N_t(a) = \frac{8 \ln T}{\Delta^2}$ we have $2\sqrt{\frac{2 \ln T}{N_t(a)}} = \Delta$

LCB1 Analysis Highlights (for two arms, a^* and a)



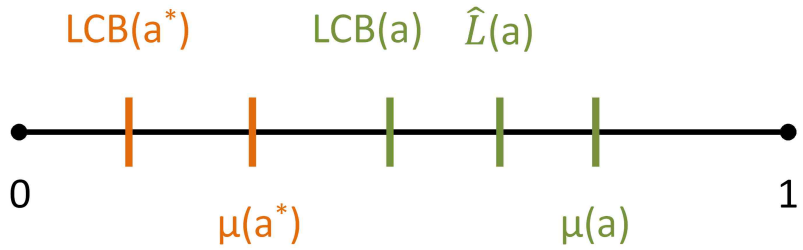
- $\Delta = \mu(a) - \mu(a^*)$
- By Hoeffding, w.h.p.: $\hat{L}_t(a) \geq \mu(a) - \sqrt{\frac{2 \ln t}{N_t(a)}}$
- Which implies: $LCB(a) = \hat{L}_t(a) - \sqrt{\frac{2 \ln t}{N_t(a)}} \geq \mu(a) - 2\sqrt{\frac{2 \ln t}{N_t(a)}}$
- Also, w.h.p.: $LCB(a^*) \leq \mu(a^*)$
- In expectation, the number of rounds on which either of the two confidence bounds fails is bounded by a constant (**careful here**)
- Fix time horizon T
- Once $N_t(a) = \frac{8 \ln T}{\Delta^2}$ we have $2\sqrt{\frac{2 \ln T}{N_t(a)}} = \Delta$
- Thus, $LCB(a) \geq \mu(a) - \Delta = \mu(a^*) \geq LCB(a^*)$

LCB1 Analysis Highlights (for two arms, a^* and a)



- $\Delta = \mu(a) - \mu(a^*)$
- By Hoeffding, w.h.p.: $\hat{L}_t(a) \geq \mu(a) - \sqrt{\frac{2 \ln t}{N_t(a)}}$
- Which implies: $LCB(a) = \hat{L}_t(a) - \sqrt{\frac{2 \ln t}{N_t(a)}} \geq \mu(a) - 2\sqrt{\frac{2 \ln t}{N_t(a)}}$
- Also, w.h.p.: $LCB(a^*) \leq \mu(a^*)$
- In expectation, the number of rounds on which either of the two confidence bounds fails is bounded by a constant (**careful here**)
- Fix time horizon T
- Once $N_t(a) = \frac{8 \ln T}{\Delta^2}$ we have $2\sqrt{\frac{2 \ln T}{N_t(a)}} = \Delta$
- Thus, $LCB(a) \geq \mu(a) - \Delta = \mu(a^*) \geq LCB(a^*)$
- Therefore, in expectation, arm a will be played no more than $\frac{8 \ln T}{\Delta^2} + constant$ number of times

LCB1 Analysis Highlights (for two arms, a^* and a)



- $\Delta = \mu(a) - \mu(a^*)$
- By Hoeffding, w.h.p.: $\hat{L}_t(a) \geq \mu(a) - \sqrt{\frac{2 \ln t}{N_t(a)}}$
- Which implies: $LCB(a) = \hat{L}_t(a) - \sqrt{\frac{2 \ln t}{N_t(a)}} \geq \mu(a) - 2\sqrt{\frac{2 \ln t}{N_t(a)}}$
- Also, w.h.p.: $LCB(a^*) \leq \mu(a^*)$
- In expectation, the number of rounds on which either of the two confidence bounds fails is bounded by a constant (**careful here**)
- Fix time horizon T
- Once $N_t(a) = \frac{8 \ln T}{\Delta^2}$ we have $2\sqrt{\frac{2 \ln T}{N_t(a)}} = \Delta$
- Thus, $LCB(a) \geq \mu(a) - \Delta = \mu(a^*) \geq LCB(a^*)$
- Therefore, in expectation, arm a will be played no more than $\frac{8 \ln T}{\Delta^2} + constant$ number of times
- $\mathbb{E}[R_T] = \Delta N_T(a) \leq \frac{8 \ln T}{\Delta} + constant\Delta$

Lower bound

[Lai & Robbins, 1985]

$$\liminf_{T\to\infty} \frac{\mathbb{E}\left[R_T\right]}{\ln T} \geq \sum_{a:\Delta(a)>0} \frac{\Delta(a)}{\mathcal{K}_{inf}(\nu_a, \mu(a^*))},$$

where $\mathcal{K}_{inf}(\nu_a, \mu(a^*))$ is the minimal KL-divergence between distribution of rewards ν_a of arm a and a suitable distribution with mean lower bounded by $\mu(a^*)$.

Lower bound

[Lai & Robbins, 1985]

$$\liminf_{T\to\infty} \frac{\mathbb{E}[R_T]}{\ln T} \geq \sum_{a:\Delta(a)>0} \frac{\Delta(a)}{\mathcal{K}_{inf}(\nu_a, \mu(a^*))},$$

where $\mathcal{K}_{inf}(\nu_a, \mu(a^*))$ is the minimal KL-divergence between distribution of rewards ν_a of arm a and a suitable distribution with mean lower bounded by $\mu(a^*)$.

Simplified

- $\mathcal{K}_{inf}(\nu_a, \mu(a^*)) \geq \frac{1}{2\Delta(a)^2}$
- When ℓ_t^a Bernoulli with $\mu(a)$ close to $\frac{1}{2}$,

$$\mathcal{K}_{inf}(\nu_a, \mu(a^*)) \approx \frac{1}{2\Delta(a)^2} \text{ and } \mathbb{E}[R_T] = \theta\left(\sum_{a:\Delta(a)>0} \frac{\ln T}{\Delta(a)}\right)$$

Other popular algorithms

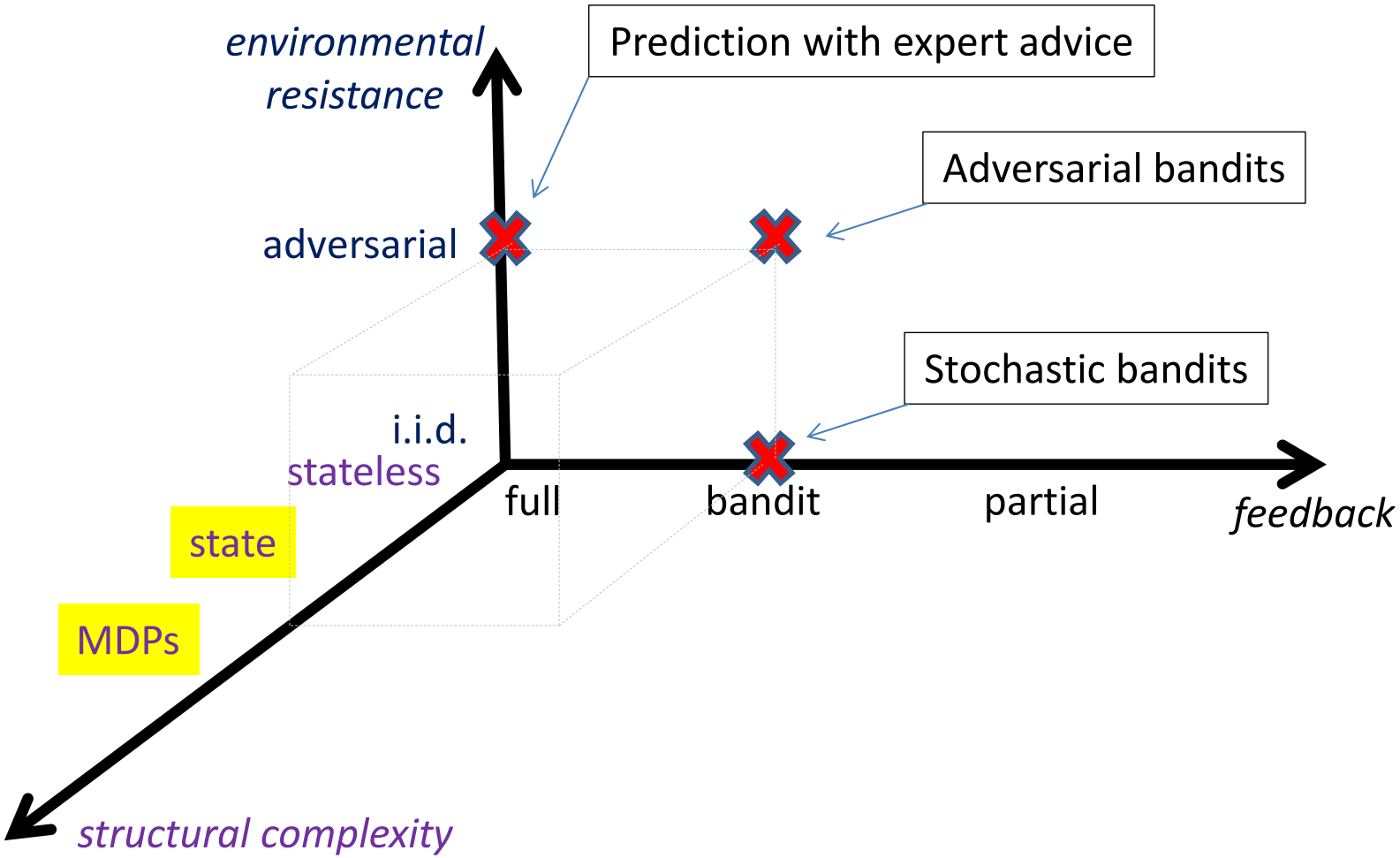
KL-UCB

- Cappé, Garivier, Maillard, Munos, Stoltz. Kullback-Leibler Upper Confidence Bounds for Optimal Sequential Allocation. *Annals of Statistics*, 2013
- Replaces Hoeffding's inequality with a tighter KL concentration inequality
- Matches the lower bound

Thompson Sampling

- Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, **1933**
- Kaufmann, Korda, Munos. Thompson sampling: an asymptotically optimal finite time analysis. *ALT*, 2012
- A Bayesian playing strategy
- Matches the lower bound

Part I: “classical” algorithms



Bandits with Side Information

A.k.a. Contextual Bandits

Version 1: Multiarmed Bandits with Expert Advice

For $t = 1, 2, \ldots$:

1. Observe advice of N experts in a form of N distributions $p_{t,h}(\cdot)$ over K arms, where $h \in \{1, \ldots, N\}$ indexes the experts
2. Play one arm, observe and suffer the loss of that arm

Bandits with Side Information

A.k.a. Contextual Bandits

Version 1: Multiarmed Bandits with Expert Advice

For $t = 1, 2, \ldots$:

1. Observe advice of N experts in a form of N distributions $p_{t,h}(\cdot)$ over K arms, where $h \in \{1, \ldots, N\}$ indexes the experts
2. Play one arm, observe and suffer the loss of that arm

Algorithm: EXP4 [Auer et al., 2002]

For $t = 1, 2, \ldots$:

- Mix the advice into $p_t(a) \propto \sum_h p_{t,h}(a) e^{-\eta_t \tilde{L}_{t-1}(h)}$
- Pick an arm A_t according to $p_t(a)$
- Use importance-weighted sampling to track $\tilde{L}_t(h)$-s

Bandits with Side Information

A.k.a. Contextual Bandits

Version 1: Multiarmed Bandits with Expert Advice

For $t = 1, 2, \ldots$:

1. Observe advice of N experts in a form of N distributions $p_{t,h}(\cdot)$ over K arms, where $h \in \{1, \ldots, N\}$ indexes the experts
2. Play one arm, observe and suffer the loss of that arm

Algorithm: EXP4 [Auer et al., 2002]

For $t = 1, 2, \ldots$:

- Mix the advice into $p_t(a) \propto \sum_h p_{t,h}(a) e^{-\eta_t \tilde{L}_{t-1}(h)}$
- Pick an arm A_t according to $p_t(a)$
- Use importance-weighted sampling to track $\tilde{L}_t(h)$-s

Regret Bound

$$\mathbb{E}[R_T] = O\left(\sqrt{KT \ln N}\right)$$

Bandits with Side Information

A.k.a. Contextual Bandits

Version 2: Multiarmed bandits with side info

For $t = 1, 2, \ldots$:

1. Observe side info (= state) S_t
2. Play arm A_t, observe and suffer the loss $\ell(A_t, S_t)$

Bandits with Side Information

A.k.a. Contextual Bandits

Version 2: Multiarmed bandits with side info

For $t = 1, 2, \ldots$:

1. Observe side info (= state) S_t
2. Play arm A_t, observe and suffer the loss $\ell(A_t, S_t)$

Expert advice is a special case of side info

Side info = the advice vector

Bandits with Side Information

A.k.a. Contextual Bandits

Version 2: Multiarmed bandits with side info

For $t = 1, 2, \ldots$:

1. Observe side info (= state) S_t
2. Play arm A_t, observe and suffer the loss $\ell(A_t, S_t)$

Expert advice is a special case of side info

Side info = the advice vector

Inverse reduction for finite state space $\mathcal{S}$

Experts = the set of all possible functions $h : \mathcal{S} \to \{1, \ldots, K\}$

$N = K^{|\mathcal{S}|}$

$\mathbb{E}[R_T] = O\left(\sqrt{KT|\mathcal{S}| \ln K}\right)$

Bandits with Side Information

A.k.a. Contextual Bandits

Version 2: Multiarmed bandits with side info

For $t = 1, 2, \ldots$:

1. Observe side info (= state) S_t
2. Play arm A_t, observe and suffer the loss $\ell(A_t, S_t)$

Expert advice is a special case of side info

Side info = the advice vector

Inverse reduction for finite state space $\mathcal{S}$

Experts = the set of all possible functions $h : \mathcal{S} \to \{1, \ldots, K\}$

$N = K^{|\mathcal{S}|}$

$\mathbb{E}[R_T] = O\left(\sqrt{KT|\mathcal{S}| \ln K}\right)$

Structural complexity

$\ln N = |\mathcal{S}| \ln K$

Markov Decision Processes (MDPs)

Game definition

Start from state S_1

For $t = 1, 2, \ldots$:

1. Play an action A_t
2. Observe and suffer loss $\ell(A_t, S_t)$
3. Transfer to state $S_{t+1} \sim p(S_{t+1}|A_t, S_t)$

Major difference with bandits with side info

S_{t+1} depends on A_t

Markov Decision Processes (MDPs)

Game definition

Start from state S_1

For $t = 1, 2, \ldots$:

1. Play an action A_t
2. Observe and suffer loss $\ell(A_t, S_t)$
3. Transfer to state $S_{t+1} \sim p(S_{t+1}|A_t, S_t)$

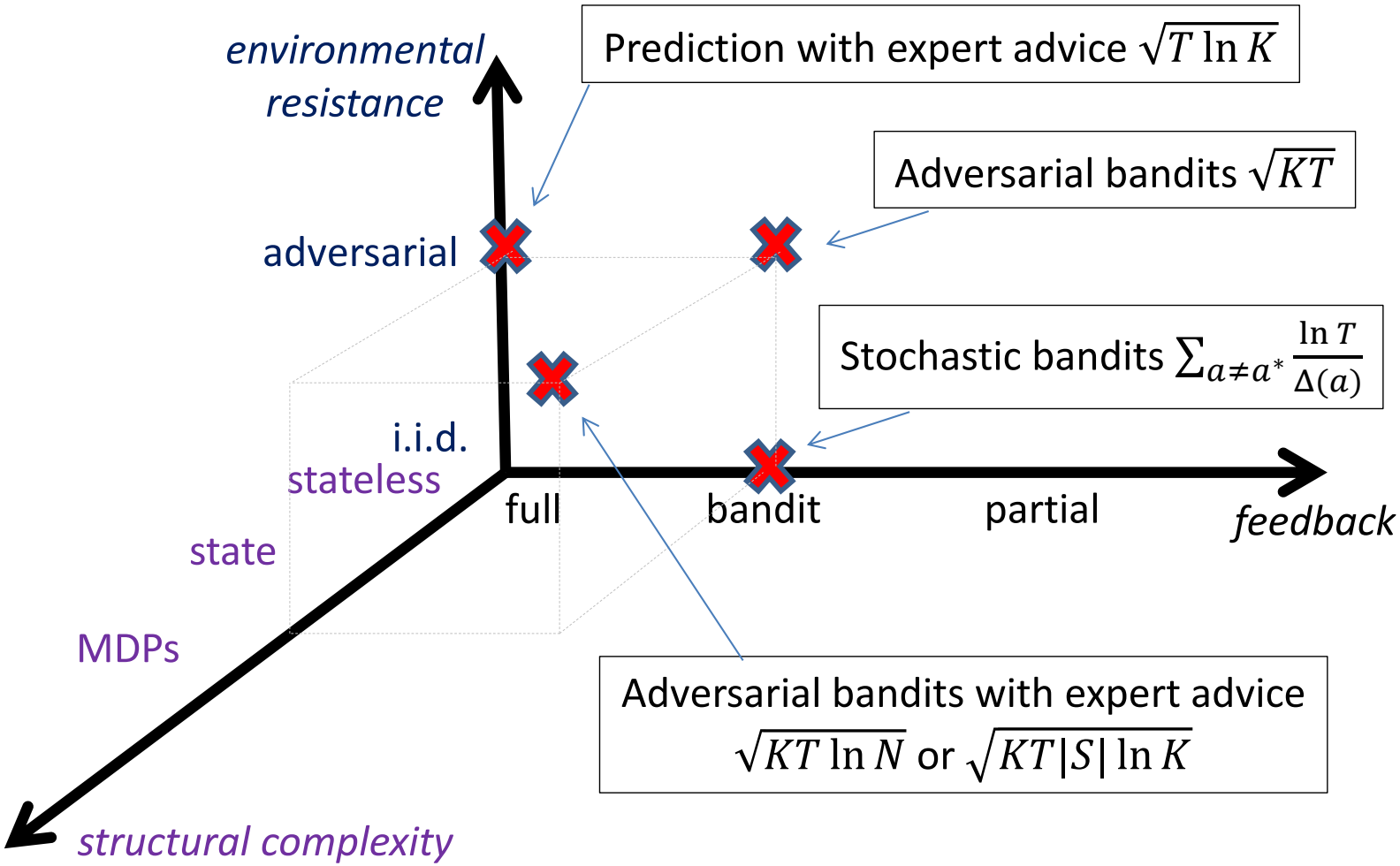
Major difference with bandits with side info

S_{t+1} depends on A_t

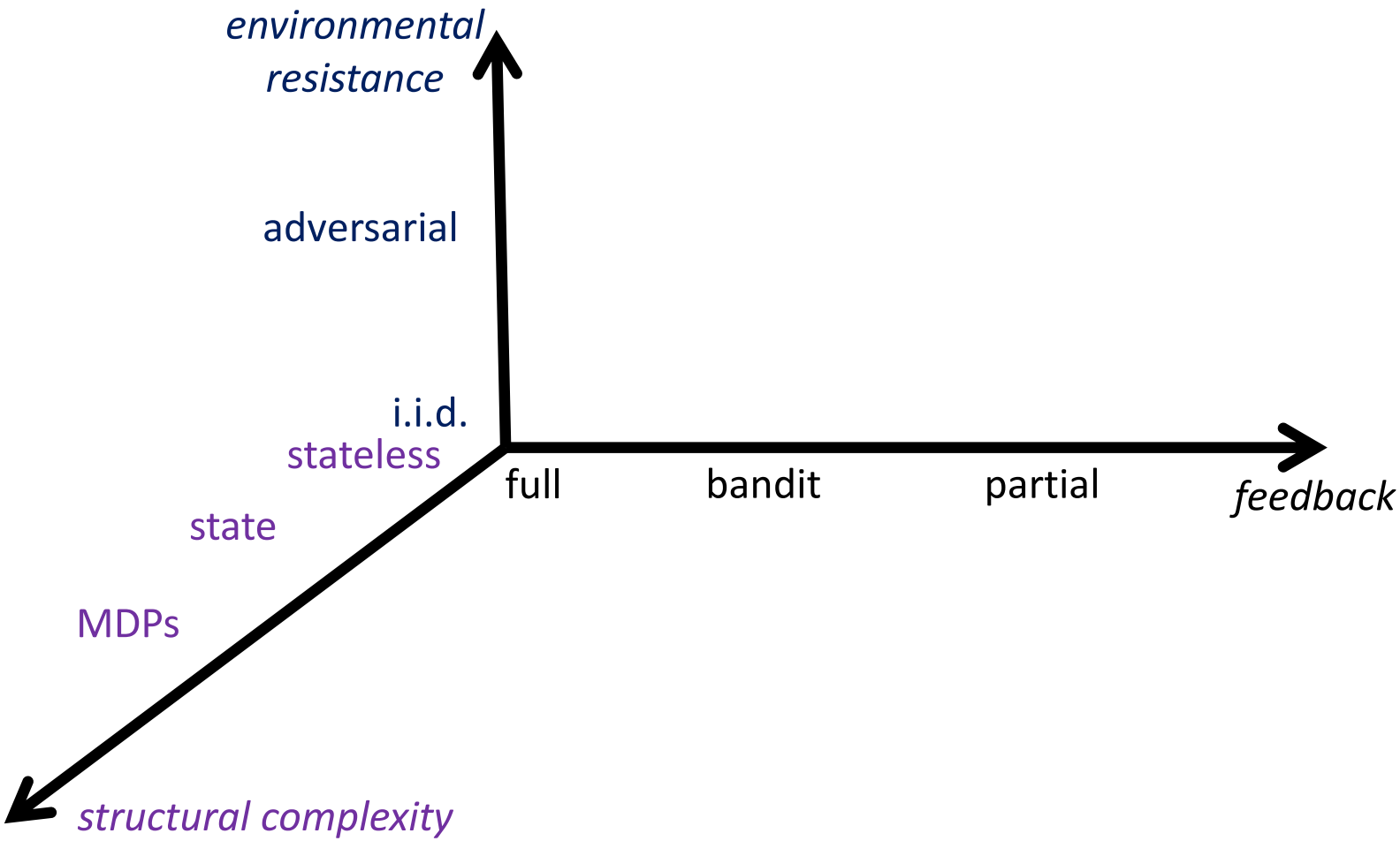
Complexity of MDPs

1. The size of the state space (same as in bandits with side info)
2. Mixing time

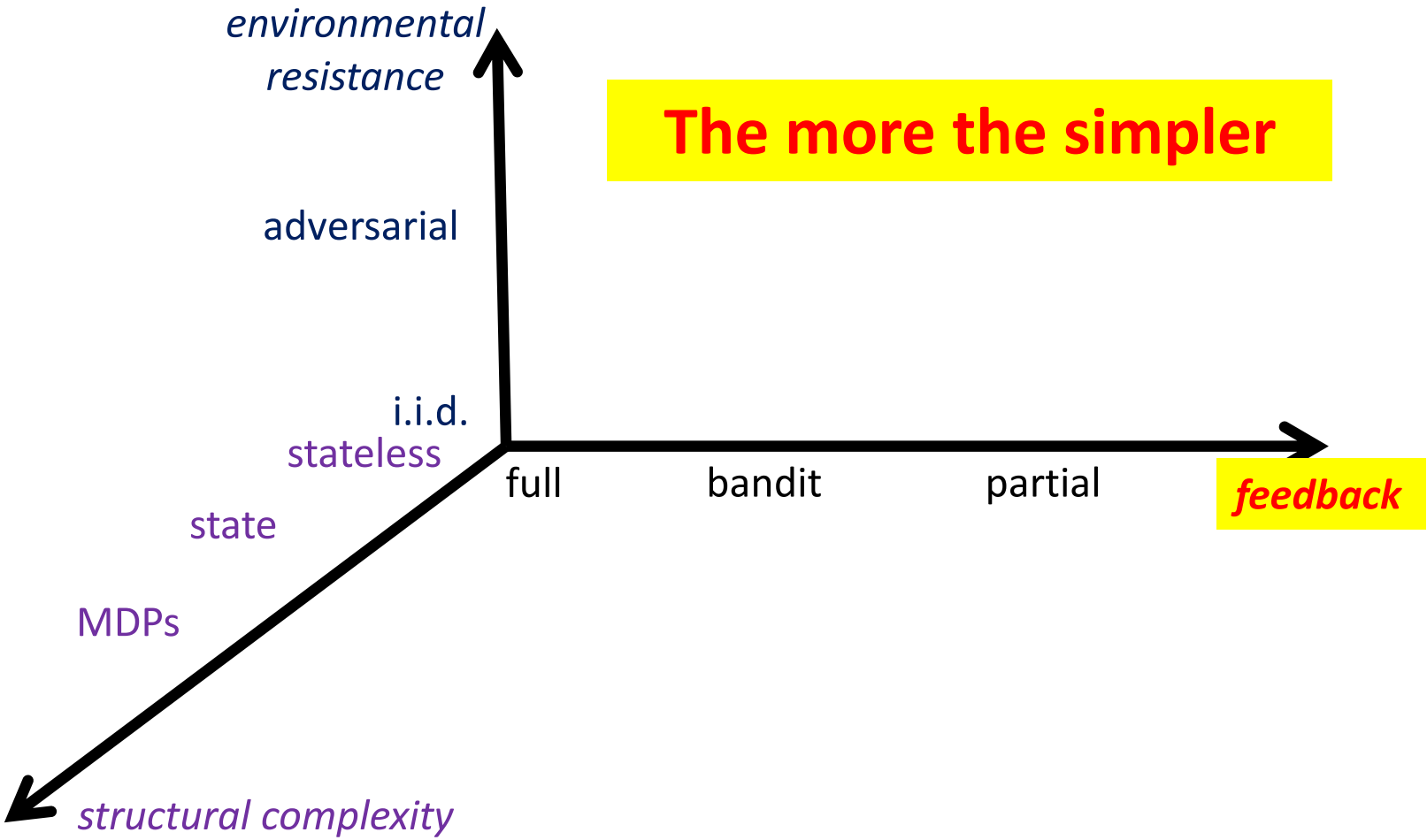
“Classical” algorithms summary



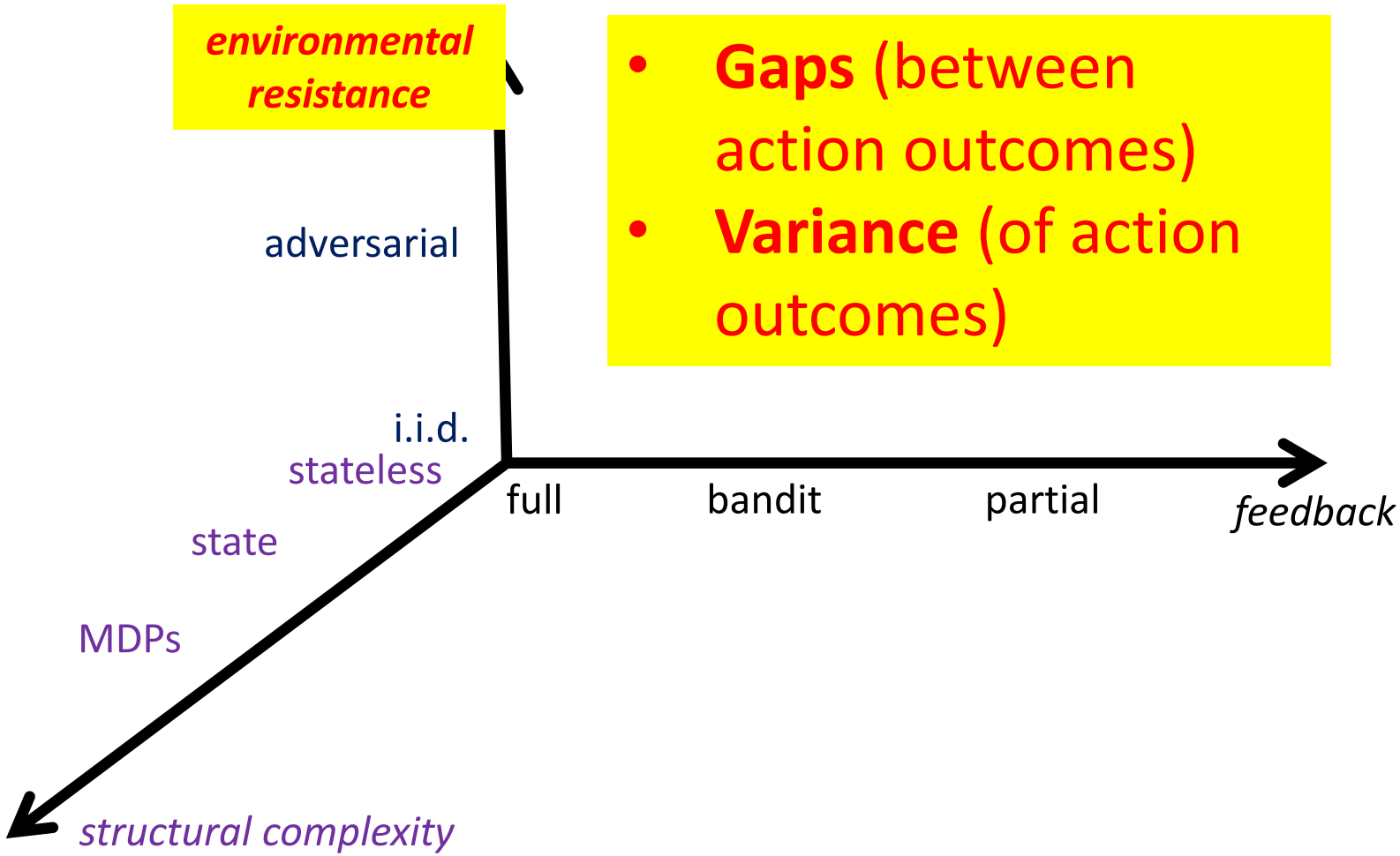
Simplicities along the axes



Simplicities along the axes

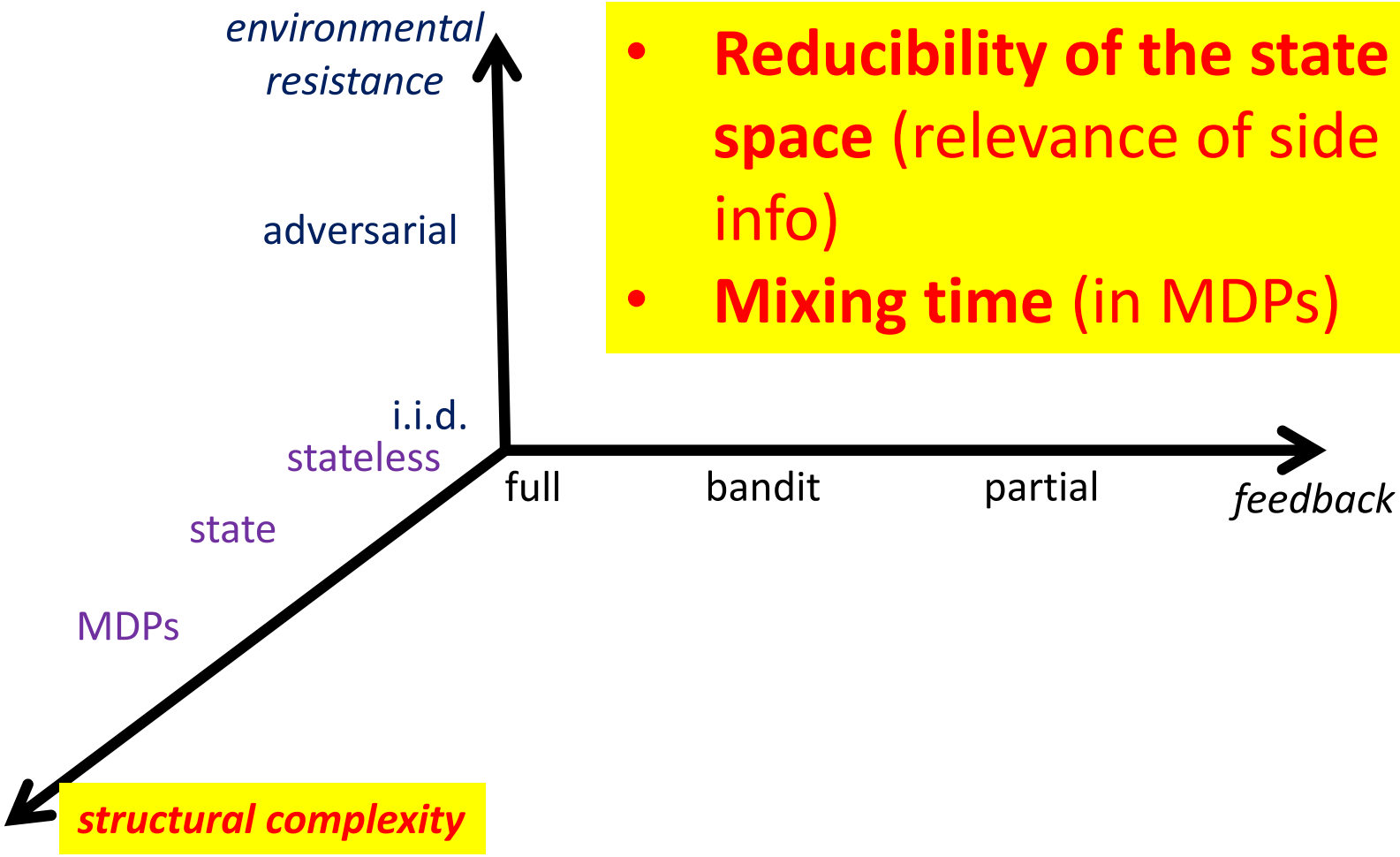


Simplicities along the axes



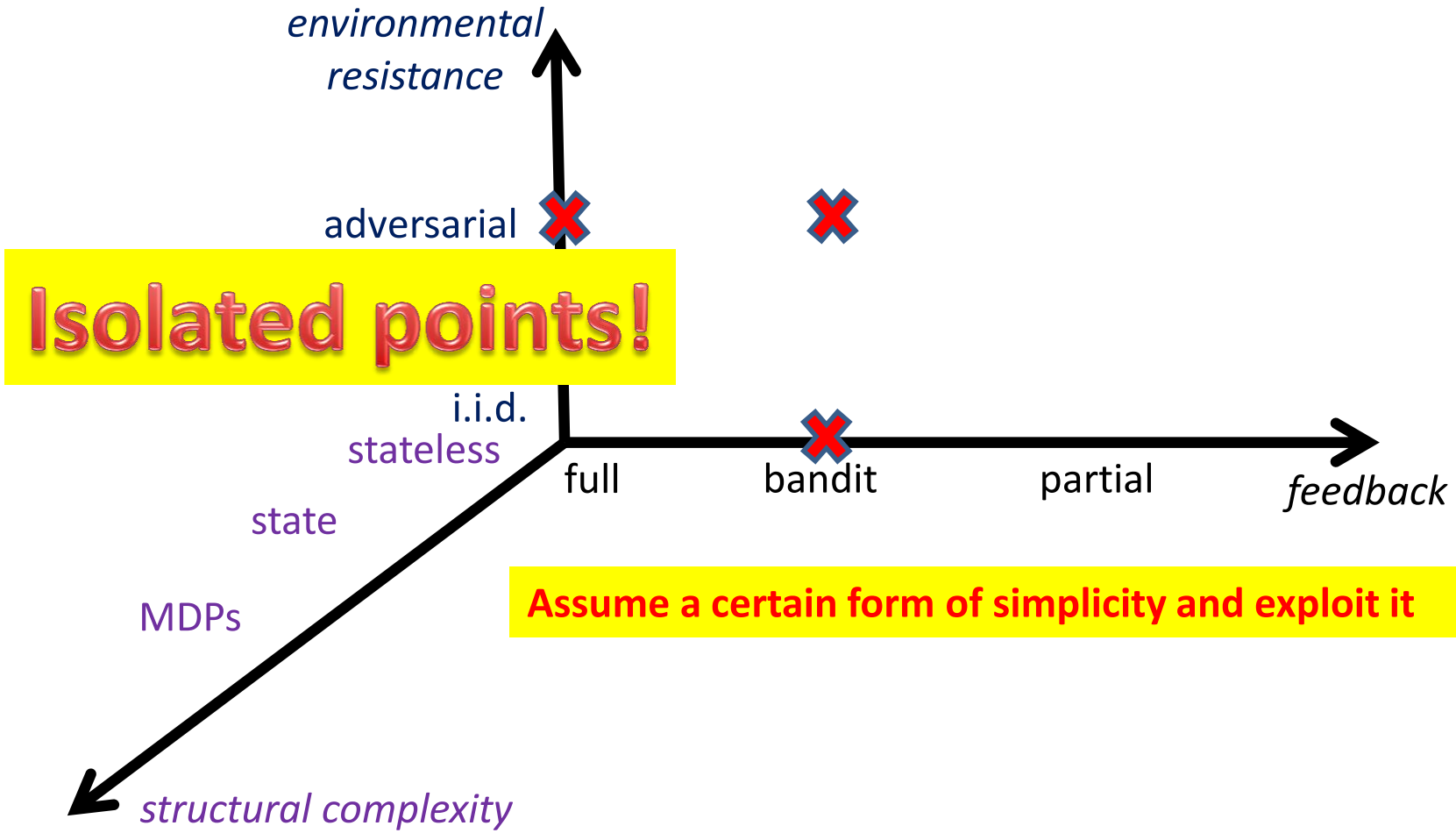
- **Gaps** (between action outcomes)
- **Variance** (of action outcomes)

Simplicities along the axes

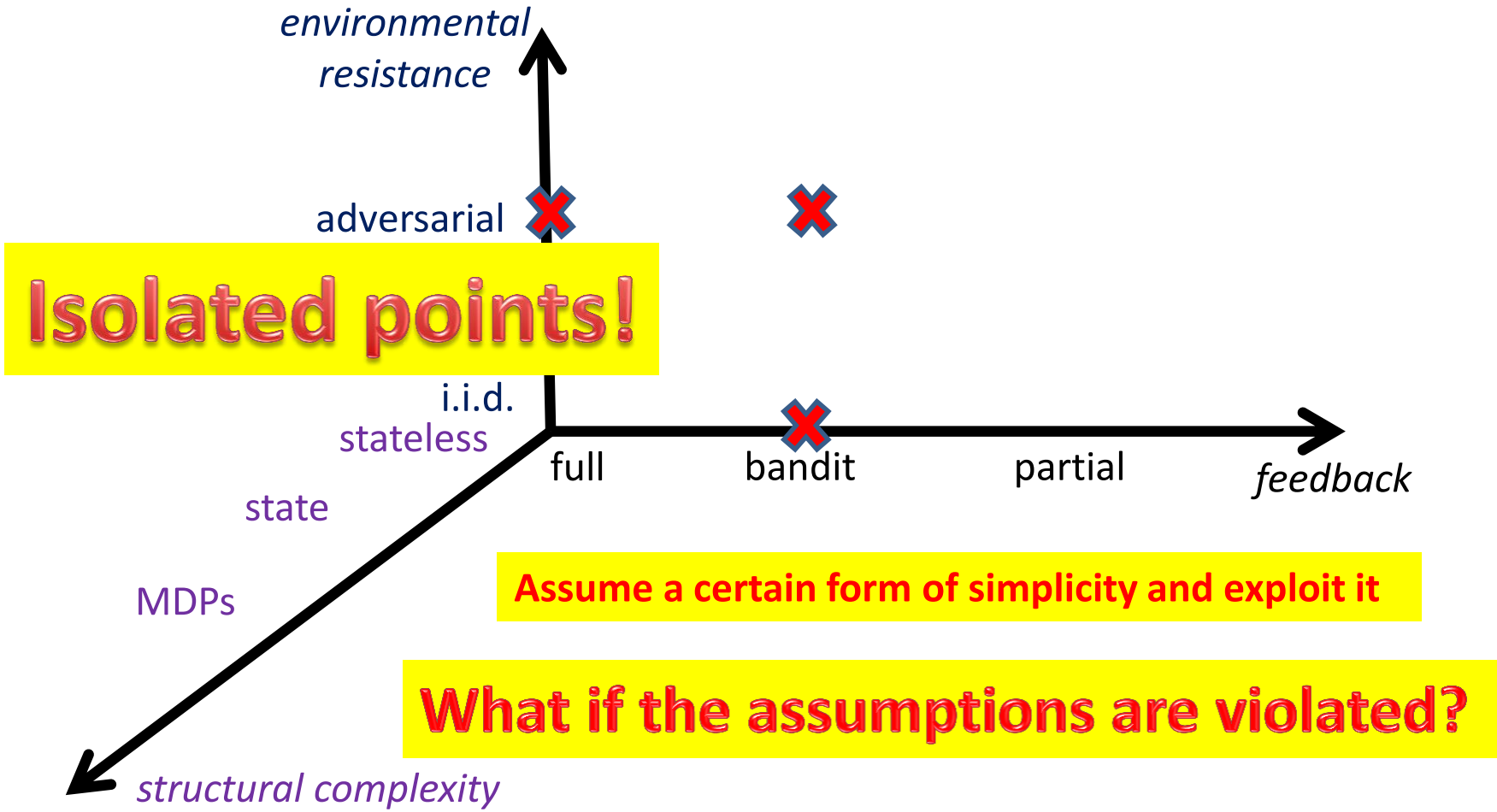


- **Reducibility of the state space** (relevance of side info)
- **Mixing time** (in MDPs)

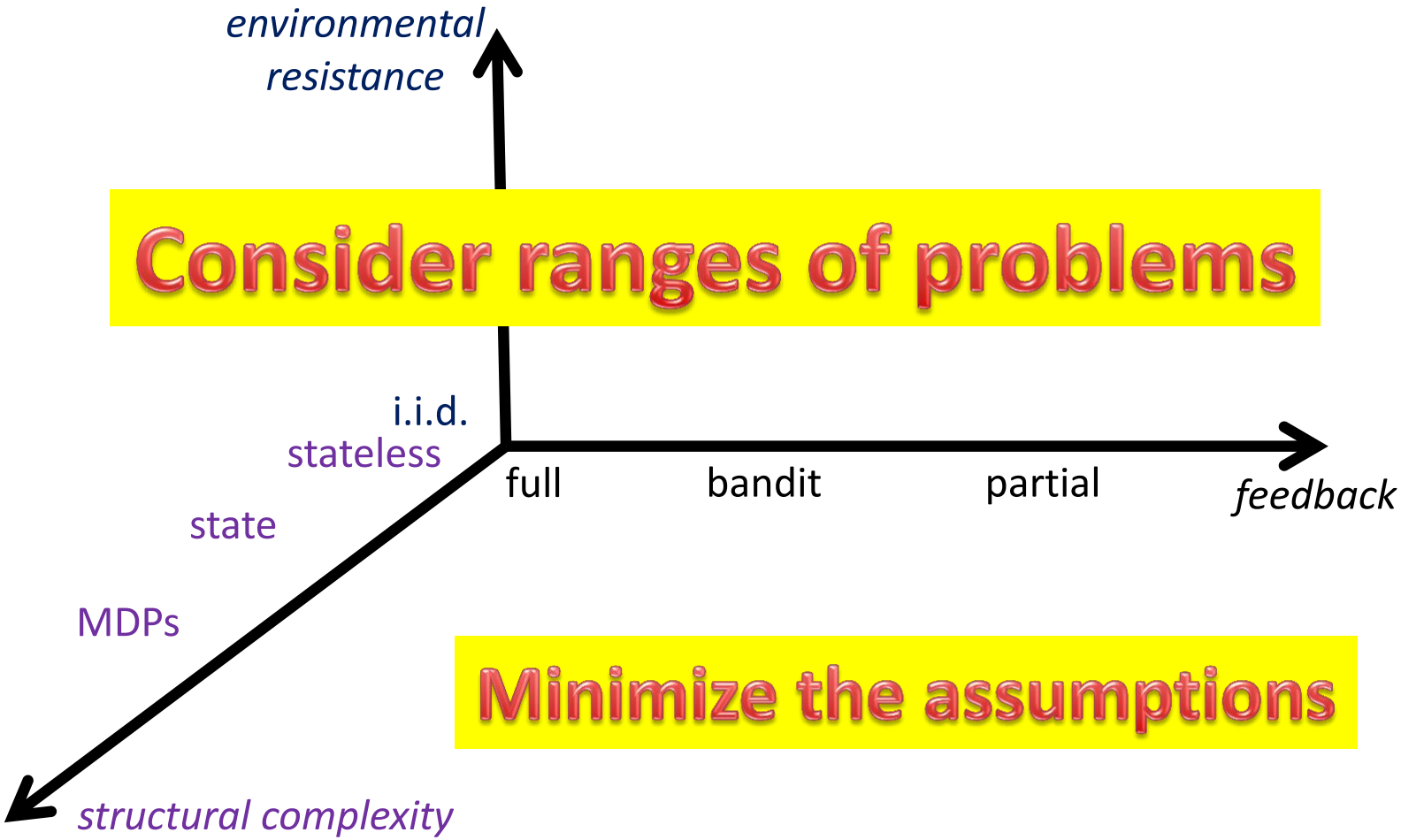
“Classical” algorithms



“Classical” algorithms

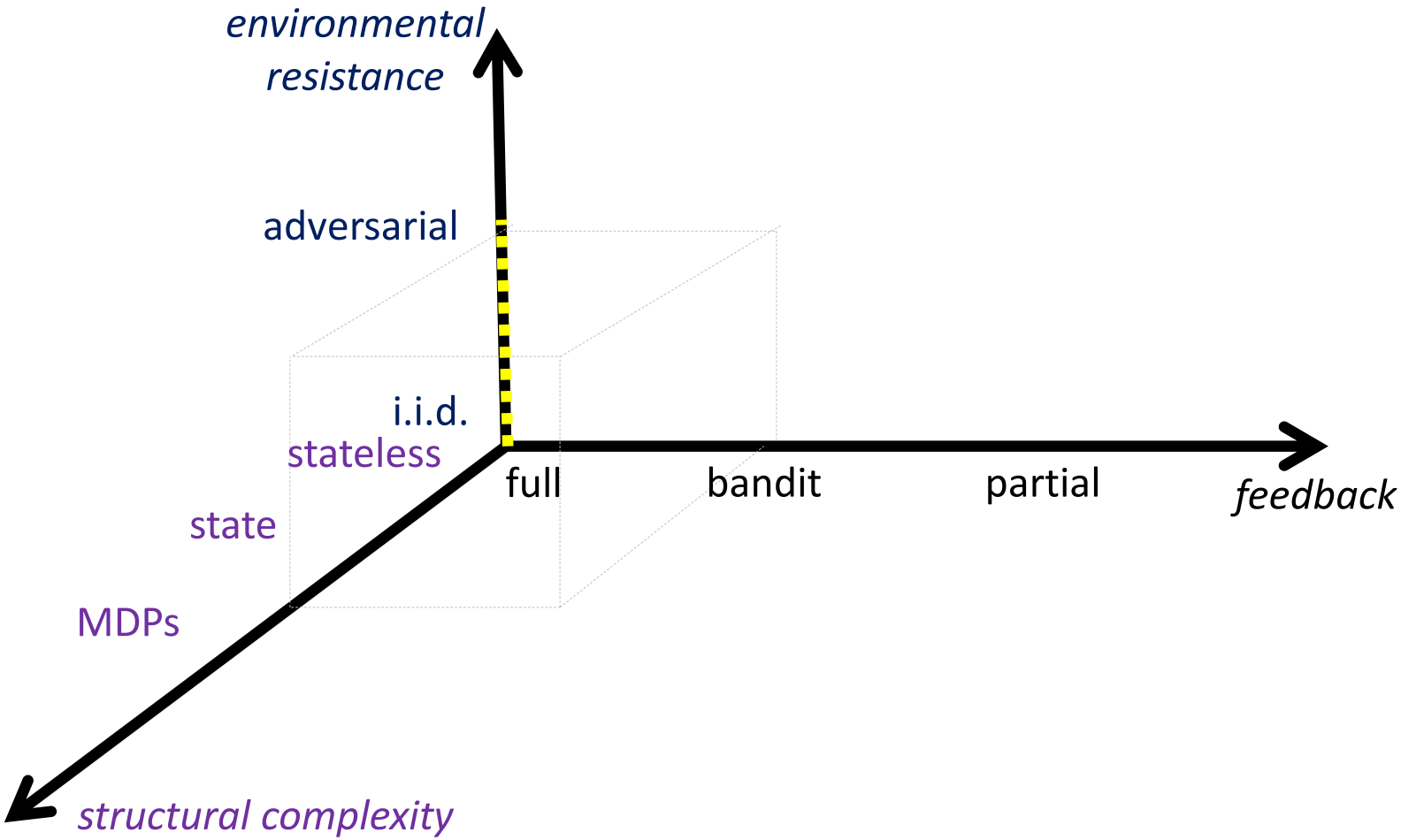


The New Generation



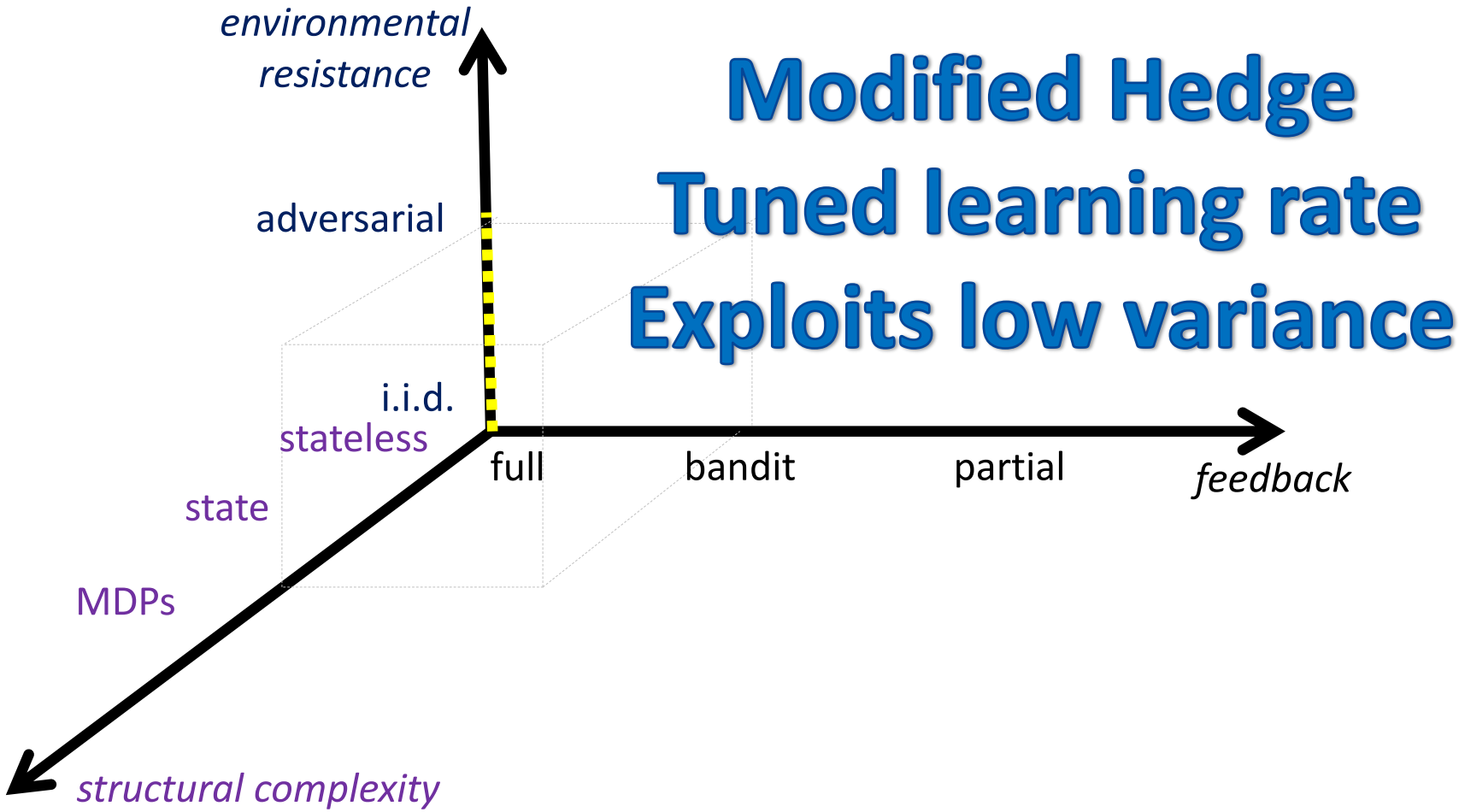
Environmental resistance in full info

[Koolen & van Erven, COLT, 2015, Luo & Schapire, COLT, 2015, Wintenberger, 2015, van Erven, Kotłowski & Warmuth, COLT, 2014, Gaillard, Stoltz & van Erven, COLT 2014, …]



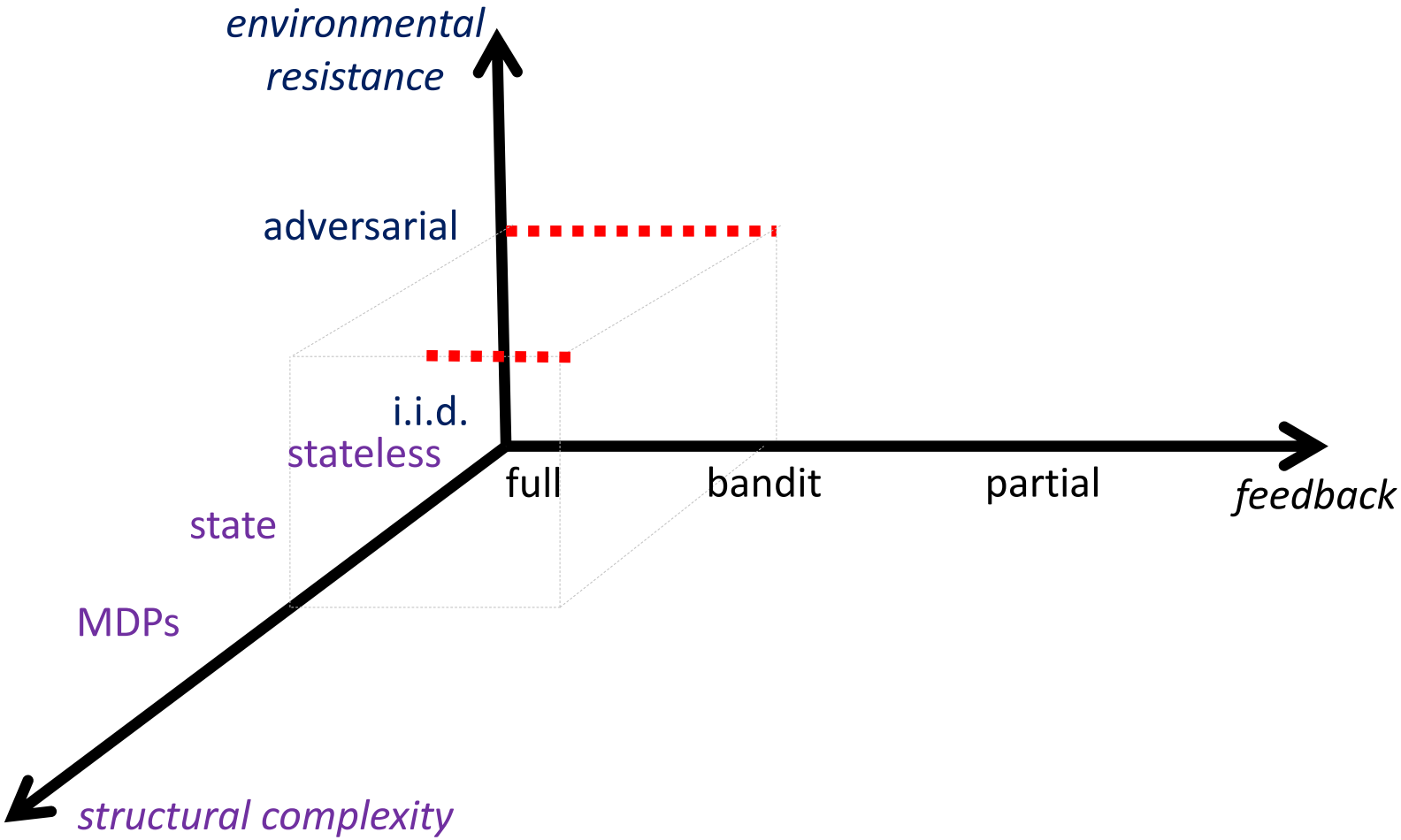
Environmental resistance in full info

[Koolen & van Erven, COLT, 2015, Luo & Schapire, COLT, 2015, Wintenberger, 2015, van Erven, Kotłowski & Warmuth, COLT, 2014, Gaillard, Stoltz & van Erven, COLT 2014, …]



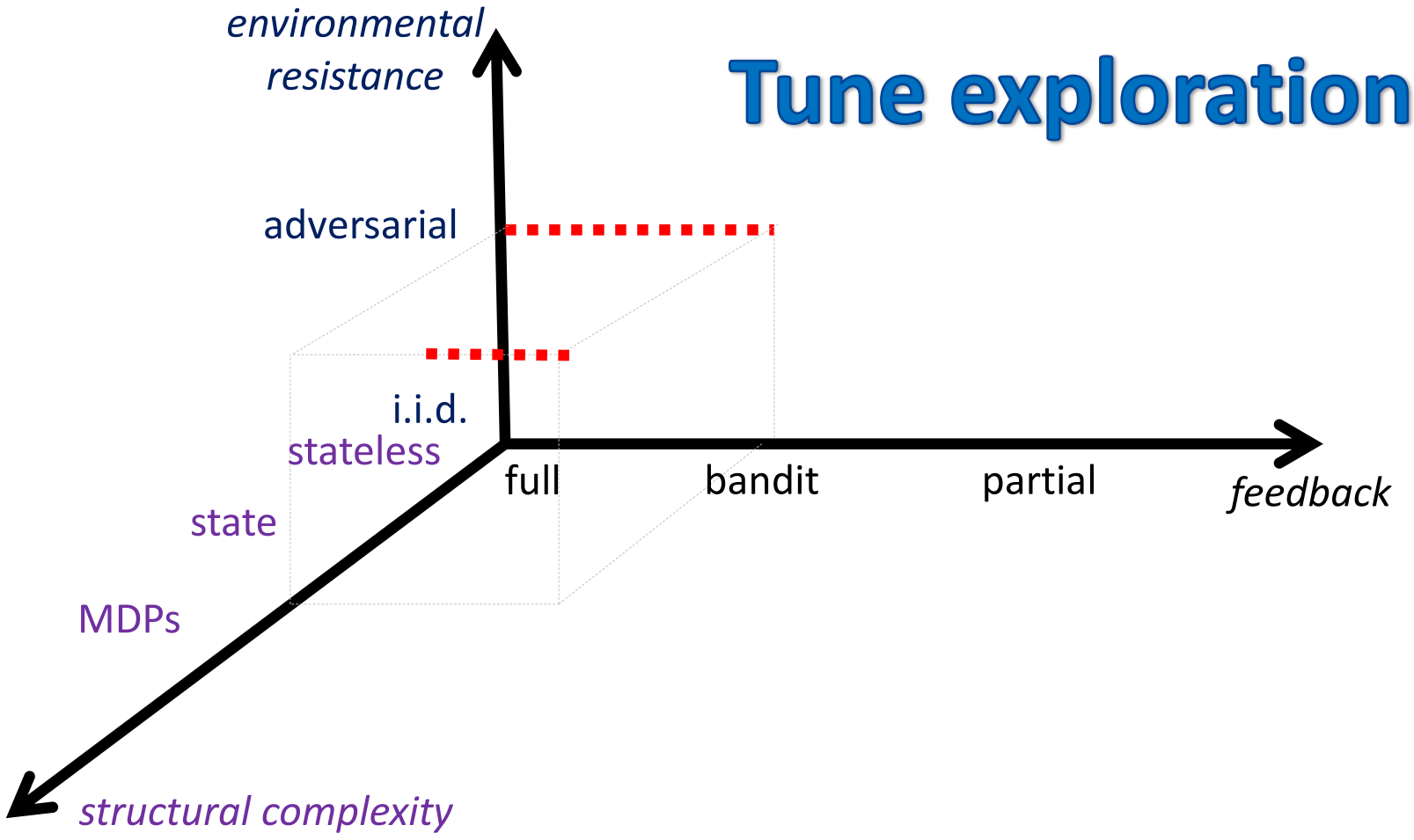
Prediction with limited advice

[Seldin, Bartlett, Cramer, Abbasi-Yadkori, ICML, 2014, Kale, COLT, 2014]



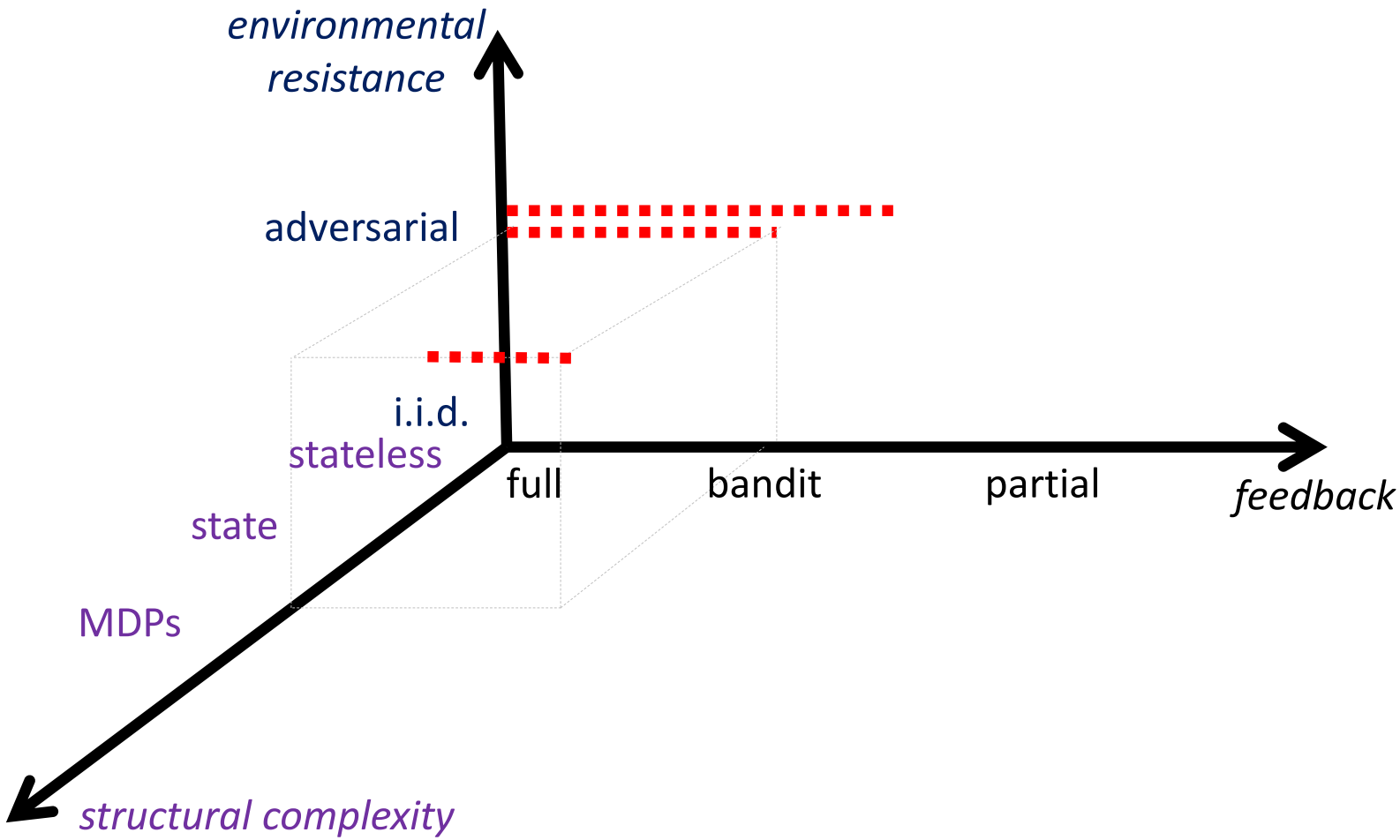
Prediction with limited advice

[Seldin, Bartlett, Cramer, Abbasi-Yadkori, ICML, 2014, Kale, COLT, 2014]



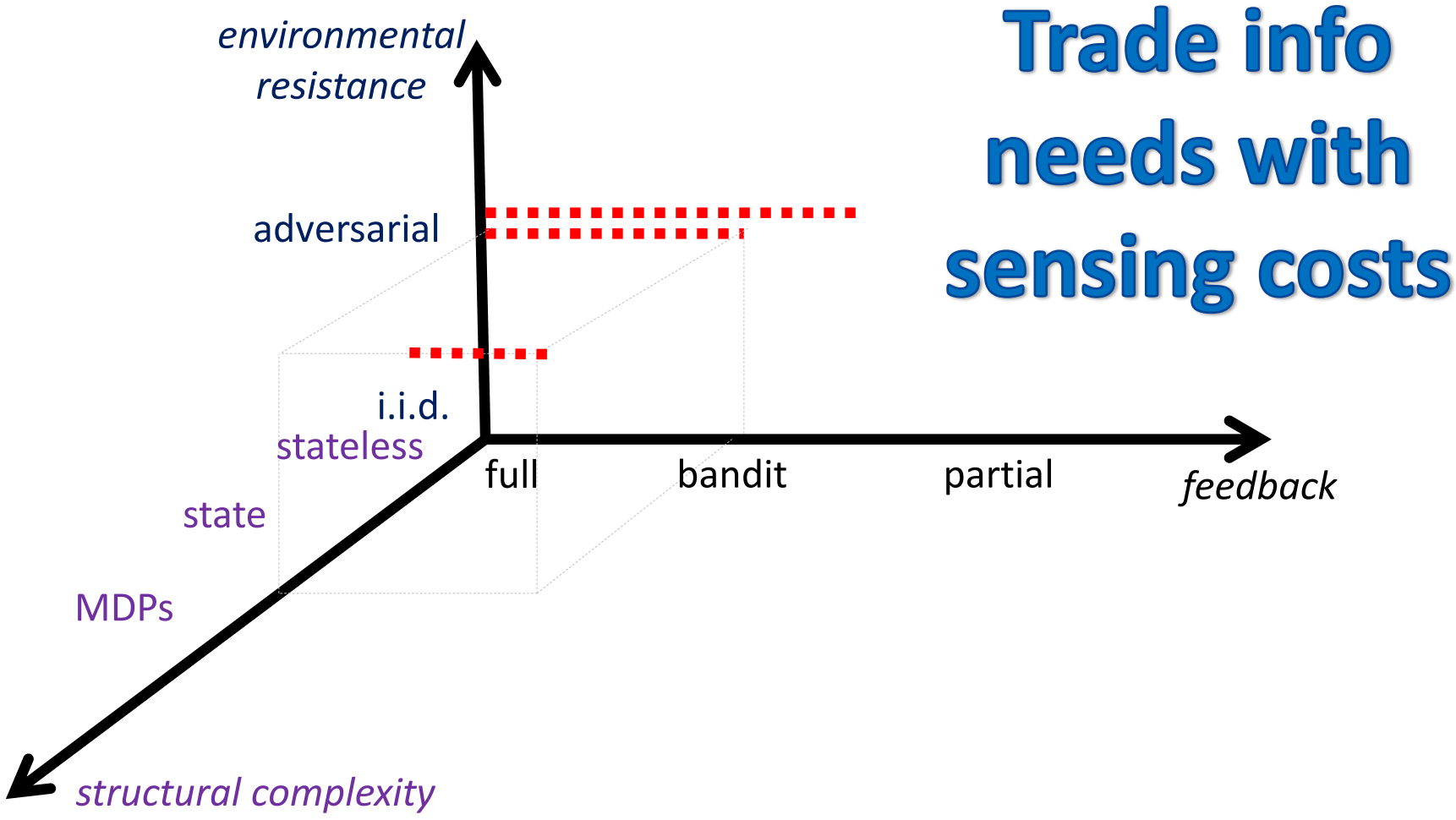
Bandits with paid observations

[Seldin, Bartlett, Cramer, Abbasi-Yadkori, ICML, 2014]



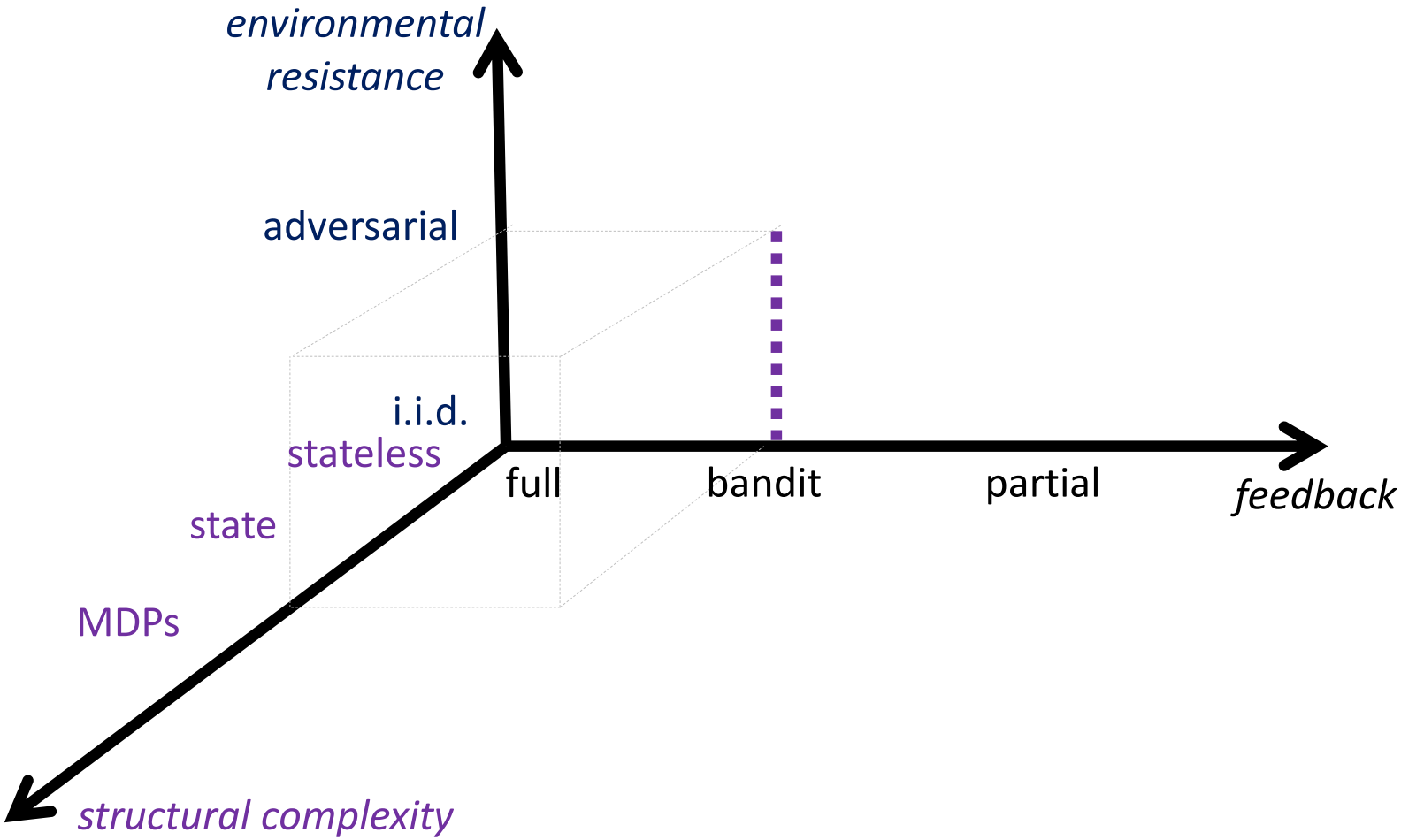
Bandits with paid observations

[Seldin, Bartlett, Cramer, Abbasi-Yadkori, ICML, 2014]



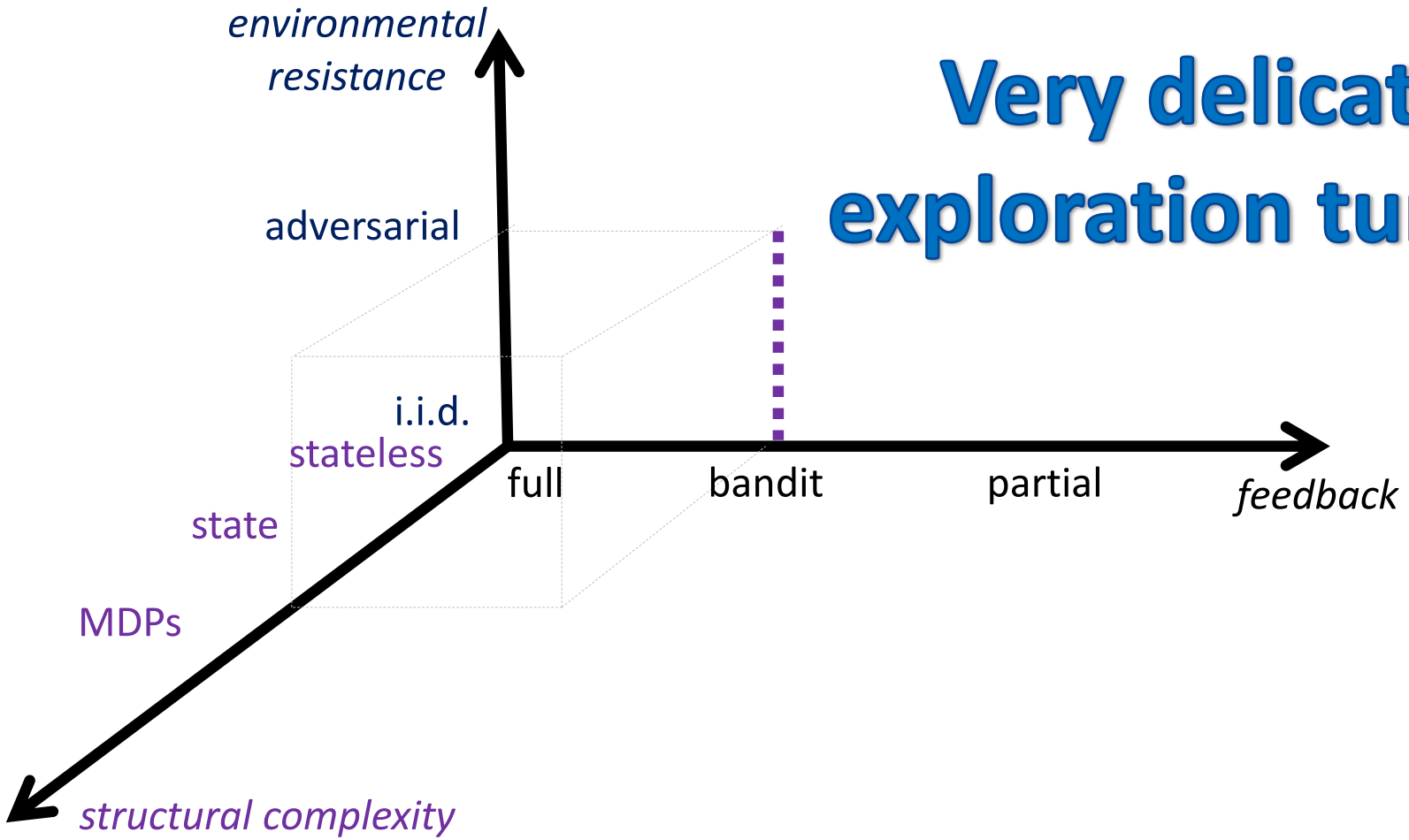
Contaminated stochastic bandits

[Seldin & Slivkins, ICML, 2014]



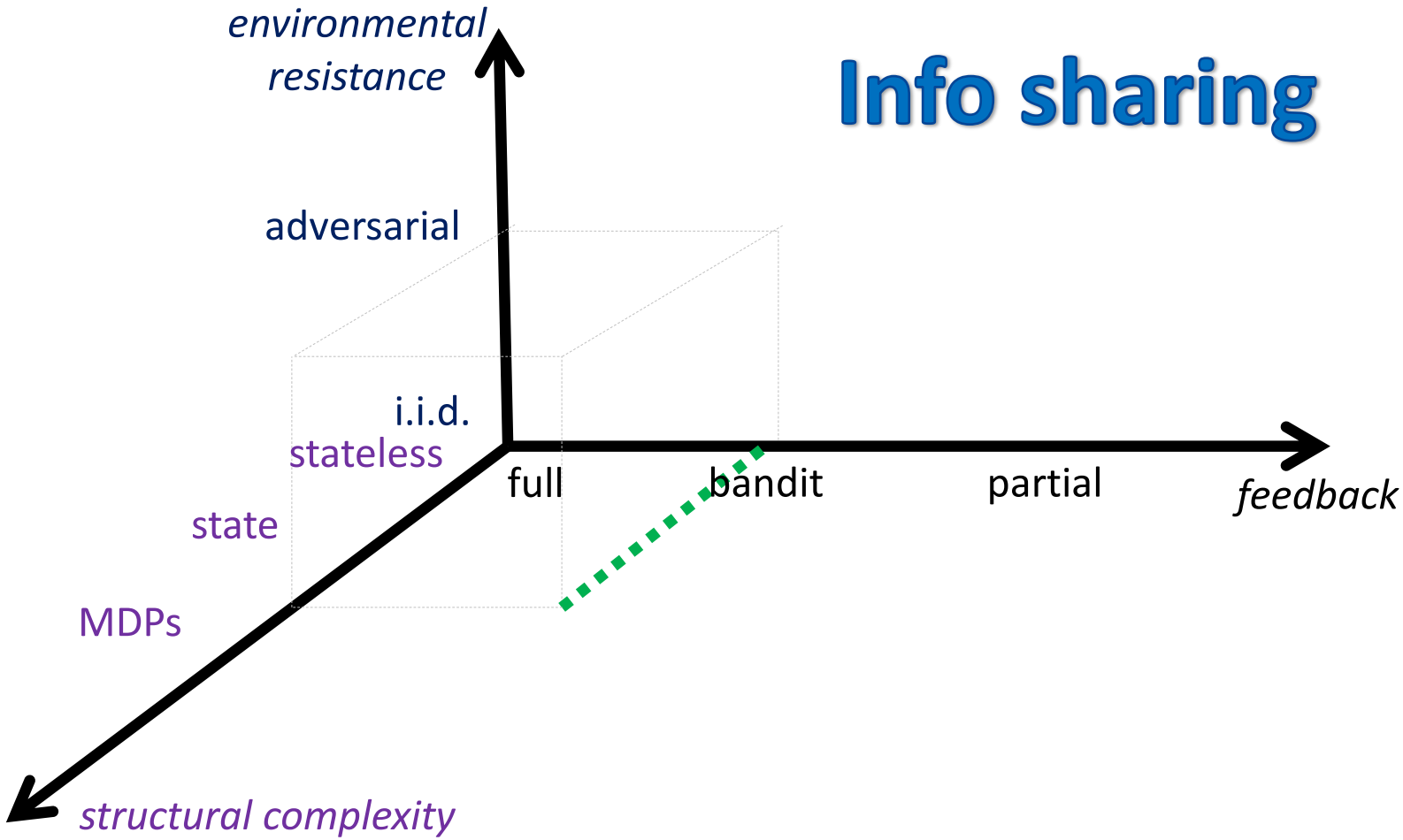
Contaminated stochastic bandits

[Seldin & Slivkins, ICML, 2014]

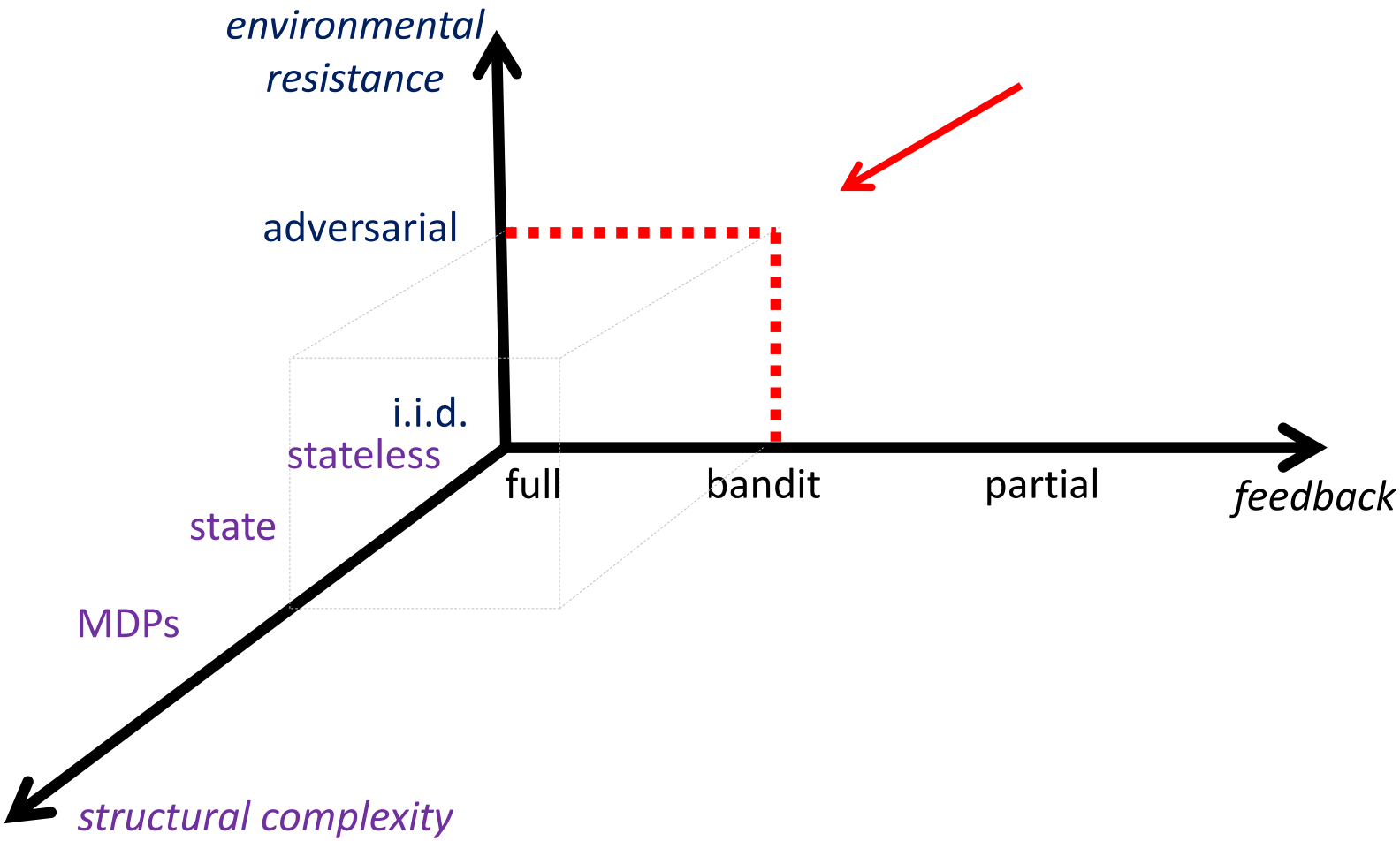


Filtering of relevant side info

[Seldin, Auer, Laviolette, Shawe-Taylor, Ortner, NIPS, 2011]



In details


environmental
resistance
adversarial
i.i.d.
stateless
state
MDPs
structural complexity
full
bandit
partial
feedback

Putting all in one language

$$
\begin{array}{cccc}
\ell_1^1, & \ell_2^1, & \dots & \ell_t^1, & \dots \\
\vdots & \vdots & \dots & \vdots & \dots \\
\ell_1^a, & \ell_2^a, & \dots & \ell_t^a, & \dots \\
\vdots & \vdots & \dots & \vdots & \dots \\
\ell_1^K, & \ell_2^K, & \dots & \ell_t^K, & \dots
\end{array}
$$

Putting all in one language

$$\begin{array}{cccc} \ell_1^1, & \ell_2^1, & \dots & \ell_t^1, & \dots \\ \vdots & \vdots & \dots & \vdots & \dots \\ \ell_1^a, & \ell_2^a, & \dots & \ell_t^a, & \dots \\ \vdots & \vdots & \dots & \vdots & \dots \\ \ell_1^K, & \ell_2^K, & \dots & \ell_t^K, & \dots \end{array}$$

Feedback

- Expert Advice: K/K
- Bandits: $1/K$

Putting all in one language

$$
\begin{array}{cccc}
\ell_1^1, & \ell_2^1, & \dots & \ell_t^1, & \dots \\
\vdots & \vdots & \dots & \vdots & \dots \\
\ell_1^a, & \ell_2^a, & \dots & \ell_t^a, & \dots \\
\vdots & \vdots & \dots & \vdots & \dots \\
\ell_1^K, & \ell_2^K, & \dots & \ell_t^K, & \dots
\end{array}
$$

Feedback

- Expert Advice: K/K
- Limited Advice: M/K
- Bandits: $1/K$

Putting all in one language

$$\begin{array}{cccc} \ell_1^1, & \ell_2^1, & \dots & \ell_t^1, & \dots \\ \vdots & \vdots & & \vdots & \\ \ell_1^a, & \ell_2^a, & \dots & \ell_t^a, & \dots \\ \vdots & \vdots & & \vdots & \\ \ell_1^K, & \ell_2^K, & \dots & \ell_t^K, & \dots \end{array}$$

Feedback

- Expert Advice: K/K
- Limited Advice: M/K
- Bandits: $1/K$
- Paid Observations: $0/K$

Putting all in one language

$$\begin{array}{cccc} \ell_1^1, & \ell_2^1, & \cdots & \ell_t^1, & \cdots \\ \vdots & \vdots & \cdots & \vdots & \cdots \\ \ell_1^a, & \ell_2^a, & \cdots & \ell_t^a, & \cdots \\ \vdots & \vdots & \cdots & \vdots & \cdots \\ \ell_1^K, & \ell_2^K, & \cdots & \ell_t^K, & \cdots \end{array}$$

Feedback

- Expert Advice: K/K
- Limited Advice: M/K
- Bandits: $1/K$
- Paid Observations: $0/K$

Loss generation

- Adversarial (deterministic)
- Stochastic ($\mathbb{E}\left[\ell_t^a\right] = \mu(a)$)

Putting all in one language

$$\begin{array}{cccc} \ell_1^1, & \ell_2^1, & \dots & \ell_t^1, & \dots \\ \vdots & \vdots & \dots & \vdots & \dots \\ \ell_1^a, & \ell_2^a, & \dots & \ell_t^a, & \dots \\ \vdots & \vdots & \dots & \vdots & \dots \\ \ell_1^K, & \ell_2^K, & \dots & \ell_t^K, & \dots \end{array}$$

Feedback

- Expert Advice: K/K
- Limited Advice: M/K
- Bandits: $1/K$
- Paid Observations: $0/K$

Loss generation

- Adversarial (deterministic)
- Contaminated stochastic
- Stochastic ($\mathbb{E}\left[\ell_t^a\right] = \mu(a)$)

Putting all in one language

$$\begin{array}{cccc} \ell_1^1, & \ell_2^1, & \cdots & \ell_t^1, & \cdots \\ \vdots & \vdots & \cdots & \vdots & \cdots \\ \ell_1^a, & \ell_2^a, & \cdots & \ell_t^a, & \cdots \\ \vdots & \vdots & \cdots & \vdots & \cdots \\ \ell_1^K, & \ell_2^K, & \cdots & \ell_t^K, & \cdots \end{array}$$

Feedback

- Expert Advice: K/K
- Limited Advice: M/K
- Bandits: $1/K$
- Paid Observations: $0/K$

Loss generation

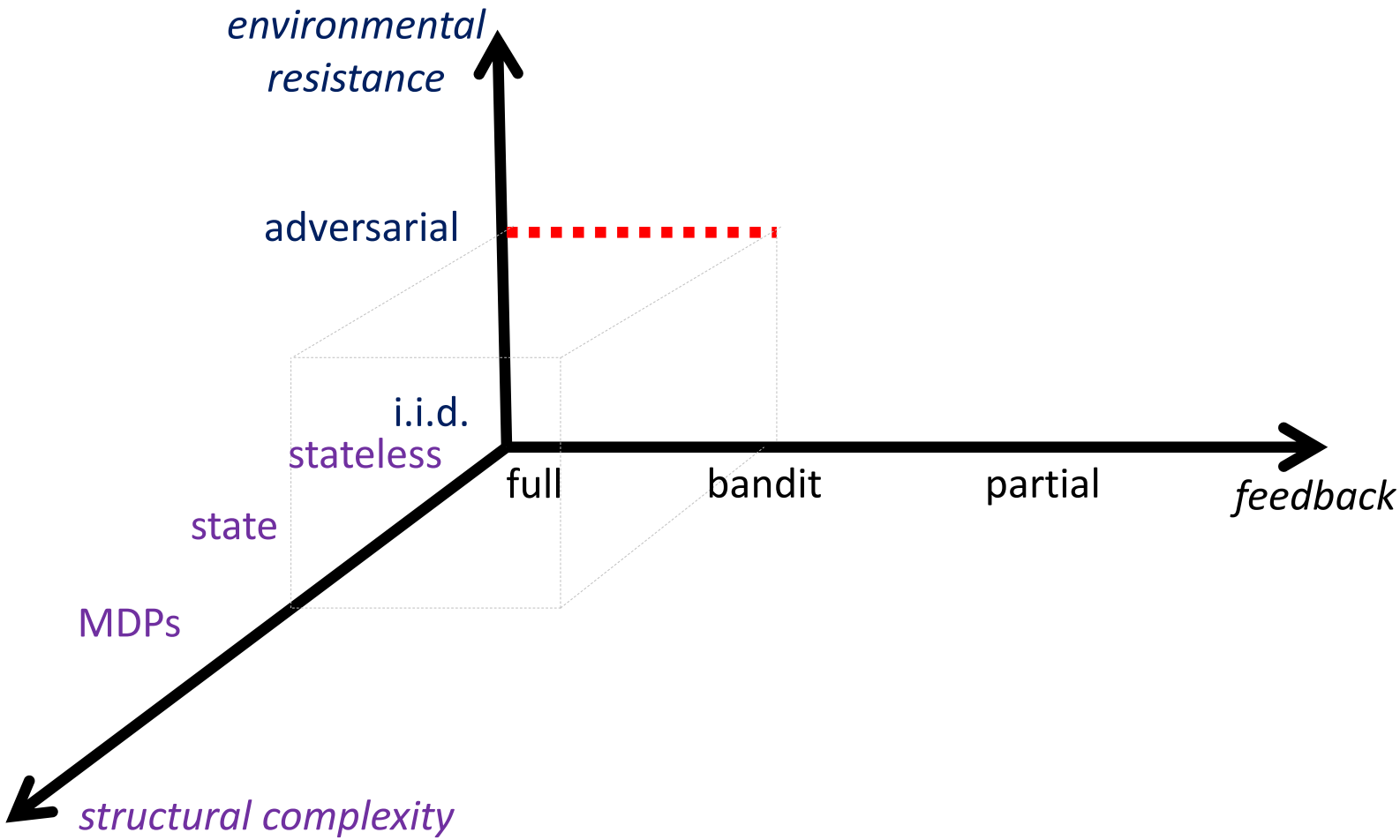
- Adversarial (deterministic)
- Contaminated stochastic
- Stochastic ($\mathbb{E}\left[\ell_t^a\right] = \mu(a)$)

Regret

$$\mathbb{E}\left[R_T\right] = \mathbb{E}\left[\sum_{t=1}^{T} \ell_t^{A_t}\right] - \min_a \mathbb{E}\left[\sum_{t=1}^{T} \ell_t^a\right]$$

Prediction with limited advice

[Seldin, Bartlett, Cramer, Abbasi-Yadkori, ICML, 2014]



Prediction with Limited Advice

Motivation

We can observe the advice of M out of K experts for $M \leq K$

Examples

Experts are computationally-expensive functions (or algorithms) and we have a constraint on the response time

Experts are humans that have to be paid

Prediction with Limited Advice

Notations

$\mathcal{O}_t \subseteq \{1, \ldots, K\}$ - the set of observed experts
$|\mathcal{O}_t| = M_t$ - the number of observed experts
$1 \leq M_t \leq N$

Game Definition

For $t = 1, 2, \ldots$:

1. Pick $(\mathcal{O}_t, A_t)$, such that $A_t \in \mathcal{O}_t$ and follow the advice of A_t
2. Observe ℓ_t^a for $a \in \mathcal{O}_t$ and suffer $\ell_t^{A_t}$

General Picture

	Prediction with Expert Advice	Prediction with Limited Advice	Bandits
Observations	$\ell_t^1, \ldots, \ell_t^K$ $(M = K)$	$\{\ell_t^a \mid a \in \mathcal{O}_t, \lvert\mathcal{O}_t\rvert = M\}$	$\ell_t^{A_t}$ $(M = 1)$
Regret Upper Bound	$O\left(\sqrt{T \ln K}\right)$	???	$O\left(\sqrt{KT}\right)$
Regret Lower Bound	$\Omega\left(\sqrt{T \ln K}\right)$	???	$\Omega\left(\sqrt{KT}\right)$

General Picture

	Prediction with Expert Advice	Prediction with Limited Advice	Bandits
Observations	$\ell_t^1, \ldots, \ell_t^K$ $(M = K)$	$\{\ell_t^a \mid a \in \mathcal{O}_t, \lvert\mathcal{O}_t\rvert = M\}$	$\ell_t^{A_t}$ $(M = 1)$
Regret Upper Bound	$O\left(\sqrt{T \ln K}\right)$	$O\left(\sqrt{\frac{K}{M} T \ln K}\right)$	$O\left(\sqrt{KT}\right)$
Regret Lower Bound	$\Omega\left(\sqrt{T \ln K}\right)$	$\Omega\left(\sqrt{\frac{K}{M} T}\right)$	$\Omega\left(\sqrt{KT}\right)$

General Picture

	Prediction with Expert Advice	Prediction with Limited Advice	Bandits
Observations	$\ell_t^1, \ldots, \ell_t^K$ $(M = K)$	$\{\ell_t^a \mid a \in \mathcal{O}_t, \lvert\mathcal{O}_t\rvert = M\}$	$\ell_t^{A_t}$ $(M = 1)$
Regret Upper Bound	$O\left(\sqrt{T \ln K}\right)$	$O\left(\sqrt{\frac{K}{M} T \ln K}\right)$	$O\left(\sqrt{KT}\right)$
Regret Lower Bound	$\Omega\left(\sqrt{T \ln K}\right)$	$\Omega\left(\sqrt{\frac{K}{M} T}\right)$	$\Omega\left(\sqrt{KT}\right)$

- The $(\ln K)$ gaps can be closed

General Picture

	Prediction with Expert Advice	Prediction with Limited Advice	Bandits
Observations	$\ell_t^1, \ldots, \ell_t^K$ $(M = K)$	$\{\ell_t^a \mid a \in \mathcal{O}_t, \lvert\mathcal{O}_t\rvert = M\}$	$\ell_t^{A_t}$ $(M = 1)$
Regret Upper Bound	$O\left(\sqrt{T \ln K}\right)$	$O\left(\sqrt{\frac{K}{M} T \ln K}\right)$	$O\left(\sqrt{KT}\right)$
Regret Lower Bound	$\Omega\left(\sqrt{T \ln K}\right)$	$\Omega\left(\sqrt{\frac{K}{M} T}\right)$	$\Omega\left(\sqrt{KT}\right)$

- The $(\ln K)$ gaps can be closed
- For time-dependent M_t the regret is $O\left(\sqrt{K\left(\sum_{t=1}^{T} \frac{1}{M_t}\right) \ln K}\right)$

Reminder: Hedge Algorithm (Exponential Weights)

Input: Learning rates $\eta_1 \geq \eta_2 \geq \cdots > 0$

$\forall a : \hat{L}_0(a) = 0$

for $t = 1, 2, \ldots$ **do**

$$\forall a : \; p_t(a) = \frac{e^{-\eta_t \hat{L}_{t-1}(a)}}{\sum_{a'} e^{-\eta_t \hat{L}_{t-1}(a')}}$$

Sample A_t according to p_t and play it

Observe $\ell_t^1, \ldots, \ell_t^K$

$$\forall a : \; \hat{L}_t(a) = \hat{L}_{t-1}(a) + \ell_t^a$$

end

Reminder: EXP3 Algorithm

Input: Learning rates $\eta_1 \geq \eta_2 \geq \cdots > 0$

$\forall a : \tilde{L}_0(a) = 0$

for $t = 1, 2, \ldots$ **do**

$$\forall a : \; p_t(a) = \frac{e^{-\eta_t \tilde{L}_{t-1}(a)}}{\sum_{a'} e^{-\eta_t \tilde{L}_{t-1}(a')}}$$

Sample A_t according to p_t and play it

Observe $\ell_t^{A_t}$

$$\forall a : \; \tilde{\ell}_t^a = \frac{\ell_t^a \mathbb{1}_{\{A_t = a\}}}{p_t(a)} = \begin{cases} \frac{\ell_t^a}{p_t(a)}, & \text{if } A_t = a \\ 0, & \text{otherwise} \end{cases}$$

Importance-weighted sampling

$$\forall a : \; \tilde{L}_t(a) = \tilde{L}_{t-1}(a) + \tilde{\ell}_t^a$$

end

Algorithm for Prediction with Limited Advice

Input: $M_1, M_2, \ldots$ and learning rates $\eta_1 \geq \eta_2 \geq \ldots$

$\forall a : \tilde{L}_0(a) = 0$

for $t = 1, 2, \ldots$ **do**

$$\forall a : \; p_t(a) = \frac{e^{-\eta_t \tilde{L}_{t-1}(a)}}{\sum_{a'} e^{-\eta_t \tilde{L}_{t-1}(a')}}$$

Sample A_t according to p_t and play it ($A_t \in \mathcal{O}_t$)

Sample $M_t - 1$ additional experts (rest of $\mathcal{O}_t$) uniformly

Observe ℓ_t^a for $a \in \mathcal{O}_t$.

$$\forall a : \; \tilde{\ell}_t^a = \frac{\ell_t^a \mathbb{1}_{\{a \in \mathcal{O}_t\}}}{p_t(a) + (1 - p_t(a)) \frac{M_t - 1}{N - 1}}$$

$$\forall a : \; \tilde{L}_t(a) = \tilde{L}_{t-1}(a) + \tilde{\ell}_t^a$$

end

Analysis idea

Upper bound

By the analysis of Hedge/EXP3:

$$\mathbb{E}\left[R_T\right] \leq \mathbb{E}\left[\sum_{t=1}^{T} \frac{\eta_t}{2} \sum_{a=1}^{K} p_t(a) \left(\tilde{\ell}_t^a\right)^2\right] + \frac{\ln K}{\eta_T}$$

Analysis idea

Upper bound

By the analysis of Hedge/EXP3:

$$\mathbb{E}\left[R_T\right] \leq \mathbb{E}\left[\sum_{t=1}^{T} \frac{\eta_t}{2} \sum_{a=1}^{K} p_t(a) \left(\tilde{\ell}_t^a\right)^2\right] + \frac{\ln K}{\eta_T}$$

And we have:

$$\mathbb{E}_t\left[\sum_{a=1}^{K} p_t(a) \left(\tilde{\ell}_t^a\right)^2\right] \leq \frac{K}{M_t}$$

Analysis idea

Upper bound

By the analysis of Hedge/EXP3:

$$\mathbb{E}\left[R_T\right] \leq \mathbb{E}\left[\sum_{t=1}^{T} \frac{\eta_t}{2} \sum_{a=1}^{K} p_t(a) \left(\tilde{\ell}_t^a\right)^2\right] + \frac{\ln K}{\eta_T}$$

And we have:

$$\mathbb{E}_t\left[\sum_{a=1}^{K} p_t(a) \left(\tilde{\ell}_t^a\right)^2\right] \leq \frac{K}{M_t}$$

By tuning η_t:

$$\mathbb{E}\left[R_T\right] \leq O\left(\sqrt{\left(\sum_{t=1}^{T} \frac{1}{M_t}\right) K \ln K}\right) = O\left(\sqrt{\frac{K}{M} T \ln K}\right)$$

Analysis idea

Upper bound

By the analysis of Hedge/EXP3:

$$\mathbb{E}\left[R_T\right] \leq \mathbb{E}\left[\sum_{t=1}^{T} \frac{\eta_t}{2} \sum_{a=1}^{K} p_t(a) \left(\tilde{\ell}_t^a\right)^2\right] + \frac{\ln K}{\eta_T}$$

And we have:

$$\mathbb{E}_t\left[\sum_{a=1}^{K} p_t(a) \left(\tilde{\ell}_t^a\right)^2\right] \leq \frac{K}{M_t}$$

By tuning η_t:

$$\mathbb{E}\left[R_T\right] \leq O\left(\sqrt{\left(\sum_{t=1}^{T} \frac{1}{M_t}\right) K \ln K}\right) = O\left(\sqrt{\frac{K}{M} T \ln K}\right)$$

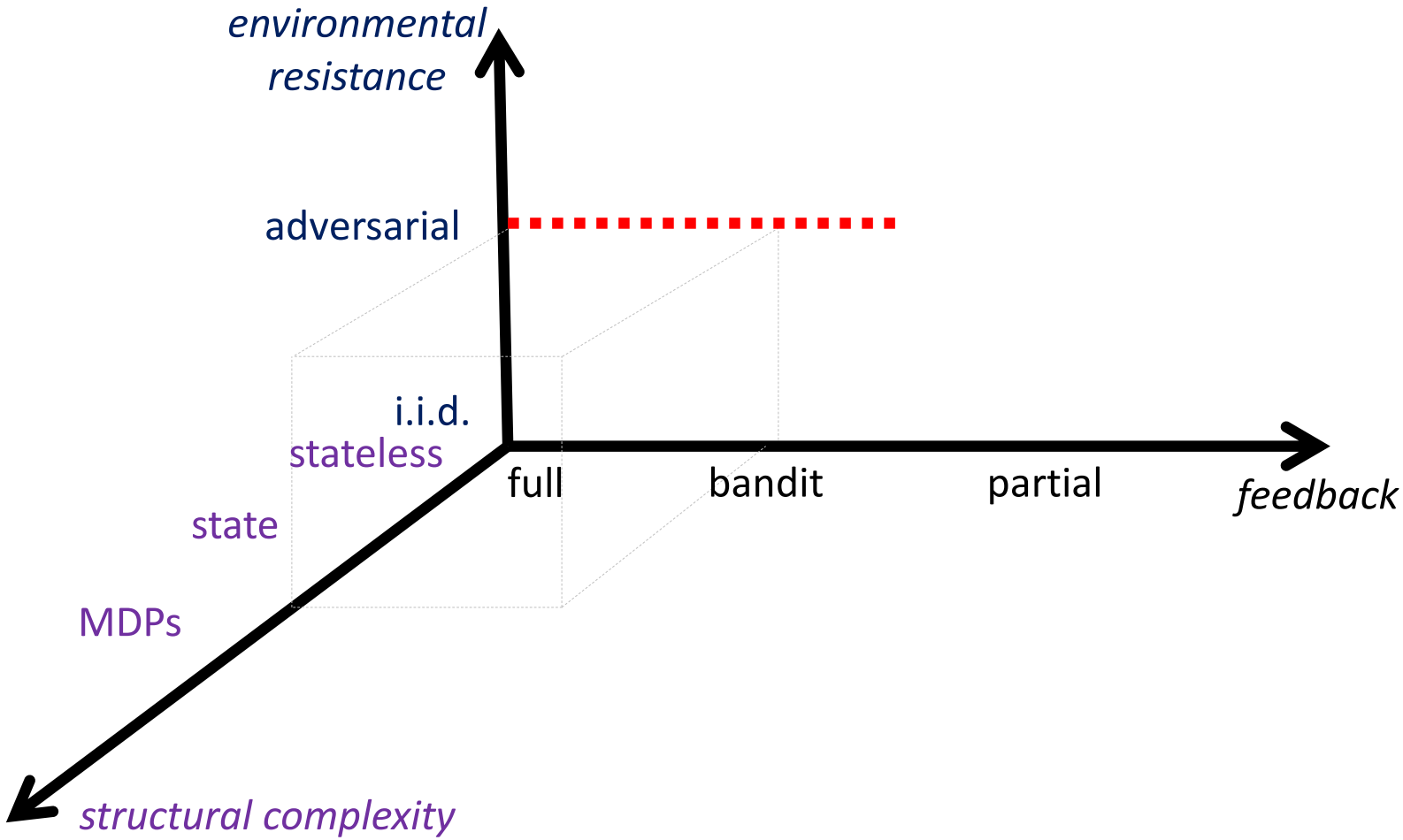
Lower bound

Similar to bandits (indistinguishability of K games), just with MT observations

$$\mathbb{E}\left[R_T\right] = \Omega\left(\sqrt{\frac{K}{M} T}\right)$$

Bandits with paid observations

[Seldin, Bartlett, Cramer, Abbasi-Yadkori, ICML, 2014]



Multiarmed Bandits with Paid Observations

Motivation

How to deal with a situation when we have to pay for observations?

The loss of any arm can be observed, but each observation has a known cost $c_t(a)$

Example

Signing contracts with service providers

$c_t(a)$ - inspection cost

Multiarmed Bandits with Paid Observations

Notations

A_t - the arm played

$\mathcal{O}_t \subseteq \{1, \ldots, K\}$ - the set of observed arms

Game Definition

For $t = 1, 2, \ldots$:

1. Pick $(A_t, \mathcal{O}_t)$ and play A_t (A_t is not necessarily in $\mathcal{O}_t$)
2. Observe ℓ_t^a for $a \in \mathcal{O}_t$ and suffer $\ell_t^{A_t} + \sum_{a \in \mathcal{O}_t} c_t(a)$

Multiarmed Bandits with Paid Observations

Notations

A_t - the arm played

$\mathcal{O}_t \subseteq \{1, \ldots, K\}$ - the set of observed arms

Game Definition

For $t = 1, 2, \ldots$:

1. Pick $(A_t, \mathcal{O}_t)$ and play A_t (A_t is not necessarily in $\mathcal{O}_t$)
2. Observe ℓ_t^a for $a \in \mathcal{O}_t$ and suffer $\ell_t^{A_t} + \sum_{a \in \mathcal{O}_t} c_t(a)$

Performance Measure: Cost-sensitive Expected Regret

$$\mathbb{E}\left[R_T^c\right] = \underbrace{\mathbb{E}\left[\sum_{t=1}^{T} \ell_t^{A_t}\right] - \min_a \left(\sum_{t=1}^{T} \ell_t^a\right)}_{\mathbb{E}[R_T]} + \mathbb{E}\left[\sum_{t=1}^{T} \sum_{a \in \mathcal{O}_t} c_t(a)\right]$$

Lower bound

Assume the algorithm makes MT observations and $c_t(a) = c$:

$$\mathbb{E}\left[R_T^c\right] = \mathbb{E}\left[R_T\right] + cMT$$

Lower bound

Assume the algorithm makes MT observations and $c_t(a) = c$:

$$\mathbb{E}\left[R_T^c\right] = \mathbb{E}\left[R_T\right] + cMT = \Omega\left(\sqrt{\frac{K}{M}T}\right) + cMT$$

Lower bound

Assume the algorithm makes MT observations and $c_t(a) = c$:

$$\mathbb{E}\left[R_T^c\right] = \mathbb{E}\left[R_T\right] + cMT = \Omega\left(\sqrt{\frac{K}{M}T}\right) + cMT \geq \Omega\left(\left(cK\right)^{1/3} T^{2/3}\right)$$

Algorithm

$\forall a : \tilde{L}_0(a) = 0$

for $t = 1, 2, \ldots$ **do**

$$\forall a : \; p_t(a) = \frac{e^{-\eta_t \tilde{L}_{t-1}(a)}}{\sum_{a'} e^{-\eta_t \tilde{L}_{t-1}(a')}}$$

Sample A_t according to p_t and play it

$\forall a$: Query the loss of a with probability

$$q_t(a) = \min\left(1, \sqrt{\frac{\eta_t p_t(a)}{2c_t(a)}}\right)$$

Trade-off between relative arm quality $p_t(a)$ & observation cost $c_t(a)$

$$\forall a : \; \tilde{\ell}_t^a = \frac{\ell_t^a}{q_t(a)} \mathbb{1}_{\{A_t = a\}}$$

$$\forall a : \; \tilde{L}_t(a) = \tilde{L}_{t-1}(a) + \tilde{\ell}_t^a$$

end

The learning rate η_t is tuned based on $p_1(\cdot), \ldots, p_{t-1}(\cdot)$ and $c_1(\cdot), \ldots, c_{t-1}(\cdot)$

Analysis

By the analysis of Hedge/EXP3:

$$\mathbb{E}\left[R_T\right] \leq \frac{\ln K}{\eta_T} + \mathbb{E}\left[\sum_{t=1}^{T} \frac{\eta_t}{2} \sum_{a=1}^{K} p_t(a) \left(\tilde{\ell}_t^a\right)^2\right]$$

Analysis

By the analysis of Hedge/EXP3:

$$\mathbb{E}\left[R_T\right] \leq \frac{\ln K}{\eta_T} + \mathbb{E}\left[\sum_{t=1}^{T} \frac{\eta_t}{2} \sum_{a=1}^{K} p_t(a) \left(\tilde{\ell}_t^a\right)^2\right]$$

And:

$$\mathbb{E}_t\left[\left(\tilde{\ell}_t^a\right)^2\right] \leq \frac{1}{q_t(a)}$$

Analysis

By the analysis of Hedge/EXP3:

$$\mathbb{E}\left[R_T\right] \leq \frac{\ln K}{\eta_T} + \mathbb{E}\left[\sum_{t=1}^{T} \frac{\eta_t}{2} \sum_{a=1}^{K} p_t(a) \left(\tilde{\ell}_t^a\right)^2\right]$$

And:

$$\mathbb{E}_t\left[\left(\tilde{\ell}_t^a\right)^2\right] \leq \frac{1}{q_t(a)}$$

Thus:

$$\mathbb{E}\left[R_T^c\right] \leq \frac{\ln K}{\eta_T} + \mathbb{E}\left[\sum_{t=1}^{T} \frac{\eta_t}{2} \sum_{a=1}^{K} \frac{p_t(a)}{q_t(a)} + c_t(a) q_t(a)\right]$$

Analysis

By the analysis of Hedge/EXP3:

$$\mathbb{E}\left[R_T\right] \leq \frac{\ln K}{\eta_T} + \mathbb{E}\left[\sum_{t=1}^{T} \frac{\eta_t}{2} \sum_{a=1}^{K} p_t(a) \left(\tilde{\ell}_t^a\right)^2\right]$$

And:

$$\mathbb{E}_t\left[\left(\tilde{\ell}_t^a\right)^2\right] \leq \frac{1}{q_t(a)}$$

Thus:

$$\mathbb{E}\left[R_T^c\right] \leq \frac{\ln K}{\eta_T} + \mathbb{E}\left[\sum_{t=1}^{T} \frac{\eta_t}{2} \sum_{a=1}^{K} \frac{p_t(a)}{q_t(a)} + c_t(a) q_t(a)\right]$$

We have to minimize:

$$\sum_{a=1}^{K} \left(\frac{\eta_t p_t(a)}{2 q_t(a)} + c_t(a) q_t(a)\right)$$

Analysis

By the analysis of Hedge/EXP3:

$$\mathbb{E}\left[R_T\right] \leq \frac{\ln K}{\eta_T} + \mathbb{E}\left[\sum_{t=1}^{T} \frac{\eta_t}{2} \sum_{a=1}^{K} p_t(a) \left(\tilde{\ell}_t^a\right)^2\right]$$

And:

$$\mathbb{E}_t\left[\left(\tilde{\ell}_t^a\right)^2\right] \leq \frac{1}{q_t(a)}$$

Thus:

$$\mathbb{E}\left[R_T^c\right] \leq \frac{\ln K}{\eta_T} + \mathbb{E}\left[\sum_{t=1}^{T} \frac{\eta_t}{2} \sum_{a=1}^{K} \frac{p_t(a)}{q_t(a)} + c_t(a) q_t(a)\right]$$

We have to minimize:

$$\sum_{a=1}^{K} \left(\frac{\eta_t p_t(a)}{2 q_t(a)} + c_t(a) q_t(a)\right)$$

This is achieved by

$$q_t(a) = \min\left\{1, \sqrt{\frac{\eta_t p_t(a)}{2 c_t(a)}}\right\}$$

Results

Simplified Upper Bound for $c_t(a) = c$

$$R_T^c \lesssim (32c \ln K)^{1/3} \left(\sum_{t=1}^{T} \underbrace{\sum_{a=1}^{K} \sqrt{p_t(a)}}_{1 \leq \cdots \leq K} \right)^{2/3} + 2\sqrt{T \ln K}$$

Worst case:

$$R_T^c \leq (32cK \ln K)^{1/3} T^{2/3} + 2\sqrt{T \ln K}$$

Favorable case (one dominating arm):

$$R_T^c \to (32c \ln K)^{1/3} T^{2/3} + 2\sqrt{T \ln K}$$

General Upper Bound

$$R_T^c \lesssim (32 \ln K)^{1/3} \left(\sum_{t=1}^{T} \underbrace{\sum_{a=1}^{K} \sqrt{p_t(a) c_t(a)}}_{\sqrt{c_t(a')} \leq \cdots \leq \sqrt{\sum_{a=1}^{K} c_t(a)}} \right)^{2/3} + 2\sqrt{T \ln K}$$

Worst case:

$$R_T^c \lesssim (32 \ln K)^{1/3} \left(\sum_{t=1}^{T} \sqrt{\sum_{a=1}^{K} c_t(a)} \right)^{2/3} + 2\sqrt{T \ln K}$$

Favorable case (one dominating arm $h^\star$):

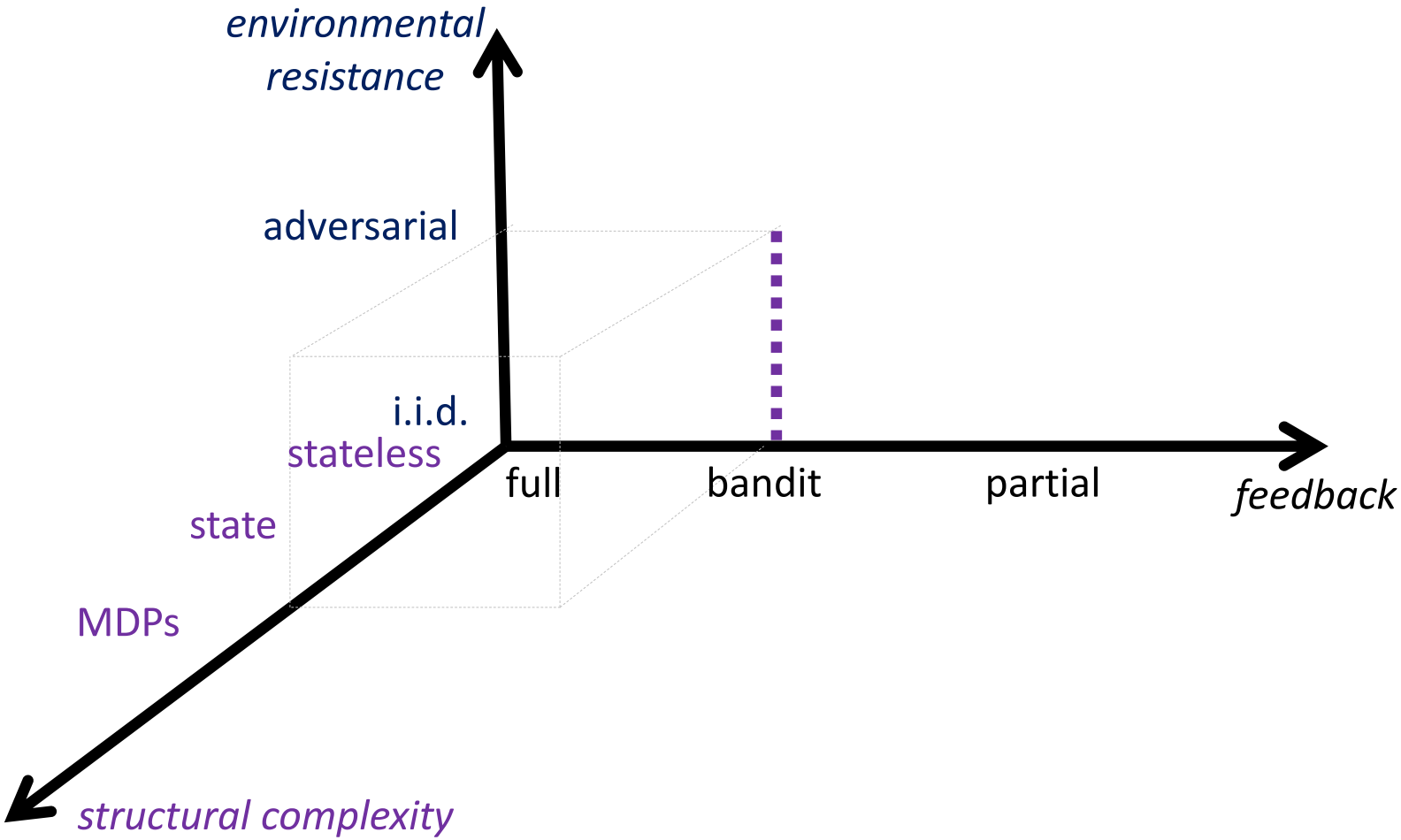
$$R_T^c \to (32 \ln K)^{1/3} \left(\sum_{t=1}^{T} \sqrt{c_t(a^\star)} \right)^{2/3} + 2\sqrt{T \ln K}$$

Bandits with Paid Observations Summary

- Adaptation to the cost of information gathering
- Balance between problem complexity and information cost
  $q_t(a) = \min\left(1, \sqrt{\frac{\eta_t p_t(a)}{2c_t(a)}}\right)$
- Automatic tuning of the learning rate η_t

Stochastic and Adversarial bandits

[Seldin & Slivkins, ICML, 2014]



Loss Generation Models

Adversarial Regime

ℓ_t^a-s are picked by an adversary in an arbitrary way

Stochastic Regime

ℓ_t^a-s are drawn independently at random, so that $\mathbb{E}\left[\ell_t^a\right] = \mu(a)$
$\Delta(a) = \mu(a) - \min_{a'}\{\mu(a')\}$ - the gap
$\Delta = \min_{a:\Delta(a)>0}\{\Delta(a)\}$ - the minimal gap

Loss Generation Models

Adversarial Regime

ℓ_t^a-s are picked by an adversary in an arbitrary way

Stochastic Regime

ℓ_t^a-s are drawn independently at random, so that $\mathbb{E}\left[\ell_t^a\right] = \mu(a)$
$\Delta(a) = \mu(a) - \min_{a'}\{\mu(a')\}$ - the gap
$\Delta = \min_{a:\Delta(a)>0}\{\Delta(a)\}$ - the minimal gap

Moderately Contaminated Stochastic Regime[NEW]

A stochastic regime, where the adversary can contaminate

- up to $t\Delta(a)/4$ entries for suboptimal actions
- up to $t\Delta/4$ entries for optimal actions

Loss Generation Models

Adversarial Regime

ℓ_t^a-s are picked by an adversary in an arbitrary way

Stochastic Regime

ℓ_t^a-s are drawn independently at random, so that $\mathbb{E}\left[\ell_t^a\right] = \mu(a)$
$\Delta(a) = \mu(a) - \min_{a'}\{\mu(a')\}$ - the gap
$\Delta = \min_{a:\Delta(a)>0}\{\Delta(a)\}$ - the minimal gap

Moderately Contaminated Stochastic Regime[NEW]

A stochastic regime, where the adversary can contaminate

- up to $t\Delta(a)/4$ entries for suboptimal actions
- up to $t\Delta/4$ entries for optimal actions

Adversarial Regime with a Gap[NEW]

Let $\lambda_t(a) = \sum_{s=1}^{t} \ell_s^a$
There exists a consistent minimizer a_τ^* of $\lambda_t(a)$ for all $t \geq \tau$
$\Delta(\tau, a) = \min_{t\geq\tau}\left\{\frac{1}{t}\left(\lambda_t(a) - \lambda_t(a_\tau^*)\right)\right\}$ - deterministic gap

Can we have one algorithm
that performs "well" in all the regimes?

(without prior knowledge of the regime type)

Classical Results

Adversarial Regime

Lower bound - $\Omega\left(\sqrt{Kt}\right)$ [Auer et. al., 1995]

EXP3 - $O\left(\sqrt{Kt\ln K}\right)$ [Auer et. al., 2002]

INF - $O\left(\sqrt{Kt}\right)$ [Audibert & Bubeck, 2009]

Stochastic Regime

Lower bound - $\Omega\left(\sum_{a:\Delta(a)>0}\frac{\ln t}{\Delta(a)}\right)$ [Lai & Robbins, 1985]

UCB1 - $O\left(\sum_{a:\Delta(a)>0}\frac{\ln t}{\Delta(a)}\right)$ [Auer et. al., 2002]

KL-UCB, Thompson sampling, EwS, ... - $O\left(\sum_{a:\Delta(a)>0}\frac{\ln t}{\Delta(a)}\right)$

Classical Results

Adversarial Regime

Lower bound - $\Omega\left(\sqrt{Kt}\right)$ [Auer et. al., 1995]

EXP3 - $O\left(\sqrt{Kt\ln K}\right)$ [Auer et. al., 2002]

INF - $O\left(\sqrt{Kt}\right)$ [Audibert & Bubeck, 2009]

Stochastic Regime

Lower bound - $\Omega\left(\sum_{a:\Delta(a)>0}\frac{\ln t}{\Delta(a)}\right)$ [Lai & Robbins, 1985]

UCB1 - $O\left(\sum_{a:\Delta(a)>0}\frac{\ln t}{\Delta(a)}\right)$ [Auer et. al., 2002]

KL-UCB, Thompson sampling, EwS, ... - $O\left(\sum_{a:\Delta(a)>0}\frac{\ln t}{\Delta(a)}\right)$

- Algorithms for the stochastic regime are inapplicable in the adversarial regime (linear regret)
- Algorithms for the adversarial regime are suboptimal in the stochastic regime

SAO

[Bubeck & Slivkins, 2012]

+ $O\left(\sqrt{TK}\,(\ln T)^{3/2}\ln K\right)$ - in the adversarial regime
+ $O\left(\frac{K}{\Delta}\,(\ln T)^{2}\ln K\right)$ - in the stochastic regime
- Does not cover the intermediate regimes
- Relatively complicated and unnatural for the problem
- Relies on the knowledge of time horizon T
- Based on one-time irreversible transition from stochastic to adversarial operation mode

The EXP3++ Algorithm

[Seldin & Slivkins, 2014]

+ Simple and natural generalization of the EXP3 algorithm
+ $O\left(\sqrt{Kt\ln K}\right)$ regret in the adversarial regime
+ $O\left(\sum_{a:\Delta(a)>0}\frac{(\ln t)^3}{\Delta(a)}\right)$ regret in the stochastic regime
+ $O\left(\sum_{a:\Delta(a)>0}\frac{(\ln t)^3}{\Delta(a)}\right)$ regret in the moderately contaminated stochastic regime
+ $O\left(\min_\tau\left\{\tau+\sum_{a:\Delta(\tau,a)>0}\frac{(\ln t)^3}{\Delta(\tau,a)}\right\}\right)$ regret in the adversarial regime with a gap

Reminder: EXP3

Control lever: $\eta_t \left(= \sqrt{\frac{\ln K}{tK}} \right)$

$\forall h : \tilde{L}_0(h) = 0$

for $t = 1, 2, \ldots$ **do**

$$\forall a : \; p_t(a) = \frac{e^{-\eta_t \tilde{L}_{t-1}(a)}}{\sum_{a'} e^{-\eta_t \tilde{L}_{t-1}(a')}}$$

Sample A_t according to p_t and play it. Observe and suffer $\ell_t^{A_t}$

$$\forall a : \; \tilde{\ell}_t^a = \frac{\ell_t^{A_t} \mathbb{1}_{\{A_t = a\}}}{p_t(h)}$$

Importance-weighted sampling

$$\forall a : \; \tilde{L}_t(a) = \tilde{L}_{t-1}(a) + \tilde{\ell}_t^a$$

end

The EXP3++ Algorithm

$\forall h : \tilde{L}_0(h) = 0$

Control levers: η_t and $\varepsilon_t(a)$-s

for $t = 1, 2, \ldots$ **do**

$$\forall a : \; p_t(a) = \frac{e^{-\eta_t \tilde{L}_{t-1}(a)}}{\sum_{a'} e^{-\eta_t \tilde{L}_{t-1}(a')}}$$

$$\forall a : \; \tilde{p}_t(a) = \left(1 - \sum_{a'} \varepsilon_t(a')\right) p_t(a) + \varepsilon_t(a)$$

Sample A_t according to $\tilde{p}_t$ and play it. Observe and suffer $\ell_t^{A_t}$

$$\forall a : \; \tilde{\ell}_t^a = \frac{\ell_t^{A_t} \mathbb{1}_{\{A_t = a\}}}{\tilde{p}_t(h)}$$

$$\forall a : \; \tilde{L}_t(a) = \tilde{L}_{t-1}(a) + \tilde{\ell}_t^a$$

end

Analysis

Adversarial Regime

$$\tilde{p}_t(a) = \left(1 - \sum_a \varepsilon_t(a)\right) p_t(a) + \varepsilon_t(a)$$

For $\varepsilon_t(a) = O\left(\sqrt{\frac{\ln K}{Kt}}\right)$:

$$\mathbb{E}\left[R_T\right] = O\left(\sqrt{KT \ln K}\right)$$

($\mathbb{E}\left[R_T\right]$ is unaffected by $\varepsilon_t(a)$)

Analysis

Stochastic Regime

Properties of Importance-Weighted Sampling

- $\mathbb{E}\left[\tilde{\ell}_t^a\right] = \mathbb{E}\left[\ell_t^a\right] = \mu(a)$
- $t\mu(a) - \tilde{L}_t(a) = \sum_{s=1}^{t}\left(\mu(a) - \tilde{\ell}_s^a\right)$ is a martingale
- Instantaneous variance: $\mathbb{E}_t\left[\left(\mu(a) - \tilde{\ell}_t^a\right)^2\right] \leq \frac{1}{\tilde{p}_t(a)} \leq \frac{1}{\varepsilon_t(a)}$
- Cumulative variance over t rounds: $\nu_t(a) \approx \frac{t}{\varepsilon_t(a)}$

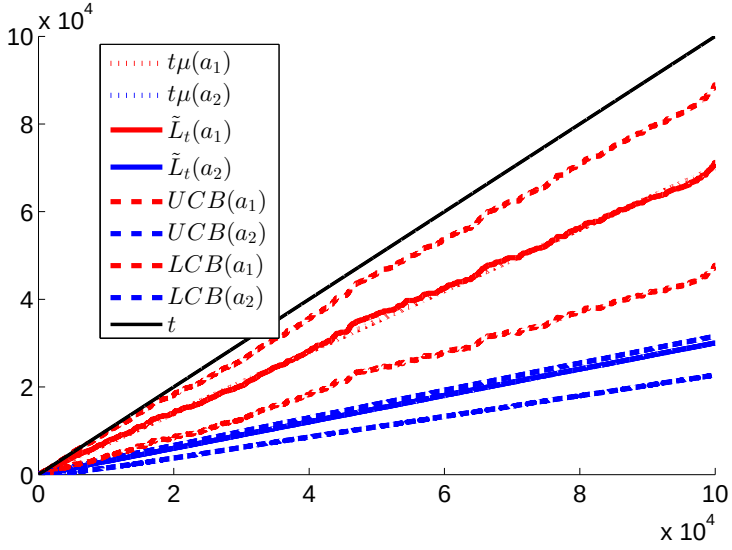
The Fundamental Trade-off of the Algorithm

Stochastic Regime

By Benstein's inequality, w.p. $\geq 1 - \frac{1}{t}$:

$$\left| t\mu(a) - \tilde{L}_t(a) \right| \leq \sqrt{2\nu_t(a) \ln t} + \frac{\ln t}{3\varepsilon_t}$$

$$\approx \sqrt{\frac{2t \ln t}{\varepsilon_t(a)}} + \frac{\ln t}{3\varepsilon_t}$$



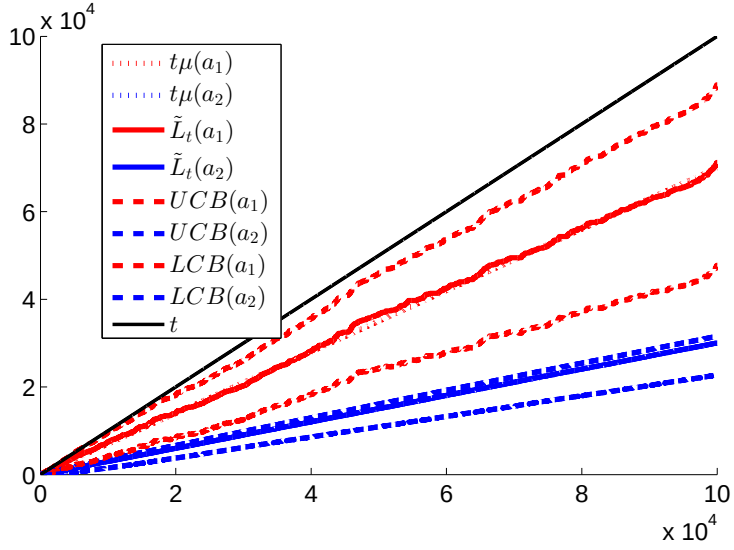
The Fundamental Trade-off of the Algorithm

Stochastic Regime

By Benstein's inequality, w.p. $\geq 1 - \frac{1}{t}$:

$$\left| t\mu(a) - \tilde{L}_t(a) \right| \leq \sqrt{2\nu_t(a) \ln t} + \frac{\ln t}{3\varepsilon_t}$$

$$\approx \sqrt{\frac{2t \ln t}{\varepsilon_t(a)}} + \frac{\ln t}{3\varepsilon_t}$$



- For separation of arms

$$\sqrt{\frac{2t \ln t}{\varepsilon_t(a)}} = O\left(t\Delta(a)\right) \quad \Rightarrow \quad \varepsilon_t(a) = \Omega\left(\frac{1}{t\Delta(a)^2}\right)$$

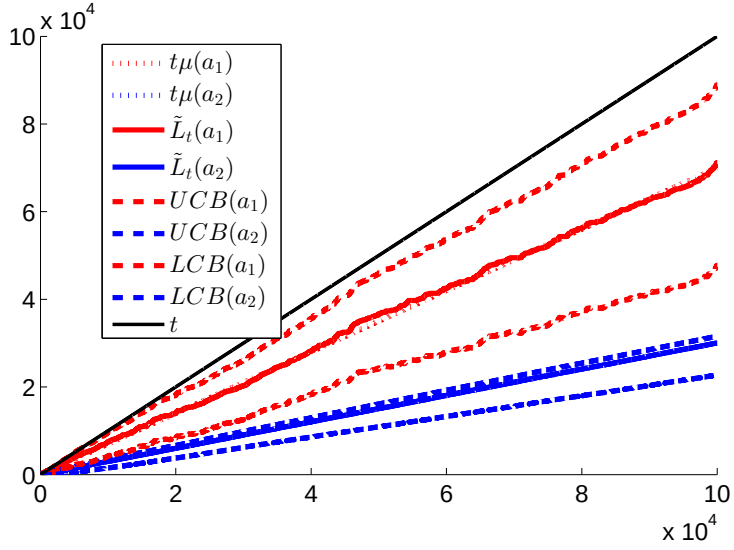
The Fundamental Trade-off of the Algorithm

Stochastic Regime

By Benstein's inequality, w.p. $\geq 1 - \frac{1}{t}$:

$$\left| t\mu(a) - \tilde{L}_t(a) \right| \leq \sqrt{2\nu_t(a) \ln t} + \frac{\ln t}{3\varepsilon_t}$$

$$\approx \sqrt{\frac{2t \ln t}{\varepsilon_t(a)}} + \frac{\ln t}{3\varepsilon_t}$$



- For separation of arms

$$\sqrt{\frac{2t \ln t}{\varepsilon_t(a)}} = O\left(t\Delta(a)\right) \quad \Rightarrow \quad \varepsilon_t(a) = \Omega\left(\frac{1}{t\Delta(a)^2}\right)$$

- $N_t(a) \geq \sum_{s=1}^{t} \varepsilon_s(a) \quad \Rightarrow \quad \varepsilon_t(a) = O\left(\frac{1}{t\Delta(a)^2}\right)$

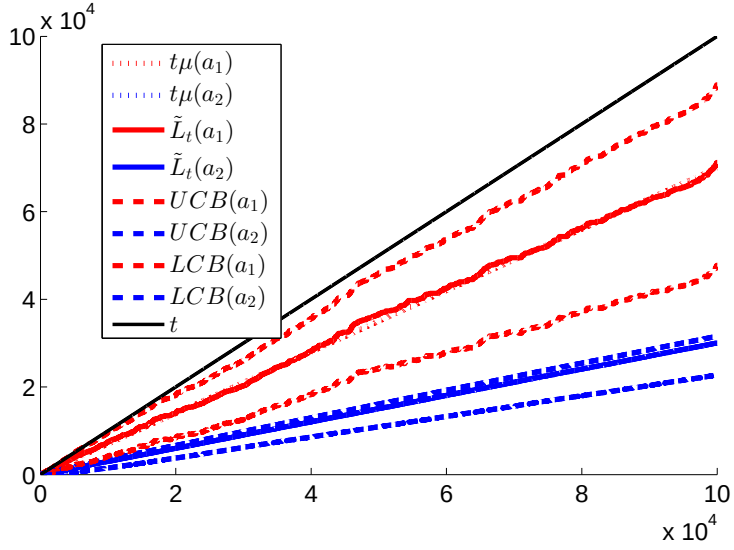
The Fundamental Trade-off of the Algorithm

Stochastic Regime

By Benstein's inequality, w.p. $\geq 1 - \frac{1}{t}$:

$$\left| t\mu(a) - \tilde{L}_t(a) \right| \leq \sqrt{2\nu_t(a) \ln t} + \frac{\ln t}{3\varepsilon_t}$$

$$\approx \sqrt{\frac{2t \ln t}{\varepsilon_t(a)}} + \frac{\ln t}{3\varepsilon_t}$$



- For separation of arms
$$\sqrt{\frac{2t \ln t}{\varepsilon_t(a)}} = O\left(t\Delta(a)\right) \Rightarrow \varepsilon_t(a) = \Omega\left(\frac{1}{t\Delta(a)^2}\right)$$
- $N_t(a) \geq \sum_{s=1}^{t} \varepsilon_s(a) \Rightarrow \varepsilon_t(a) = O\left(\frac{1}{t\Delta(a)^2}\right)$
- We take $\varepsilon_t(a) = \frac{18(\ln t)^2}{t\hat{\Delta}_t(a)^2}$ and show that $\hat{\Delta}_t(a) \to \Delta(a)$

Main Results

- $\eta_t = \frac{1}{2}\sqrt{\frac{\ln K}{tK}}$ and $\varepsilon_t(a) = O\left(\sqrt{\frac{\ln K}{tK}}\right)$

  $\Rightarrow O\left(\sqrt{tK \ln K}\right)$ regret in the adversarial regime

Main Results

- $\eta_t = \frac{1}{2}\sqrt{\frac{\ln K}{tK}}$ and $\varepsilon_t(a) = O\left(\sqrt{\frac{\ln K}{tK}}\right)$

  $\Rightarrow O\left(\sqrt{tK \ln K}\right)$ regret in the adversarial regime

Let $\hat{\Delta}_t(a)$ be empirical estimate of $\Delta(a)$

- $\varepsilon_t(a) = \frac{18(\ln t)^2}{t\hat{\Delta}_t(a)^2}$ and $\eta_t \geq \frac{1}{2}\sqrt{\frac{\ln K}{tK}}$

  $\Rightarrow O\left(\frac{(\ln t)^3}{\Delta(a)}\right)$ regret in the stochastic regime,
  moderately contaminated stochastic,
  adversarial with a gap

Main Results

- $\eta_t = \frac{1}{2}\sqrt{\frac{\ln K}{tK}}$ and $\varepsilon_t(a) = O\left(\sqrt{\frac{\ln K}{tK}}\right)$

  $\Rightarrow O\left(\sqrt{tK \ln K}\right)$ regret in the adversarial regime

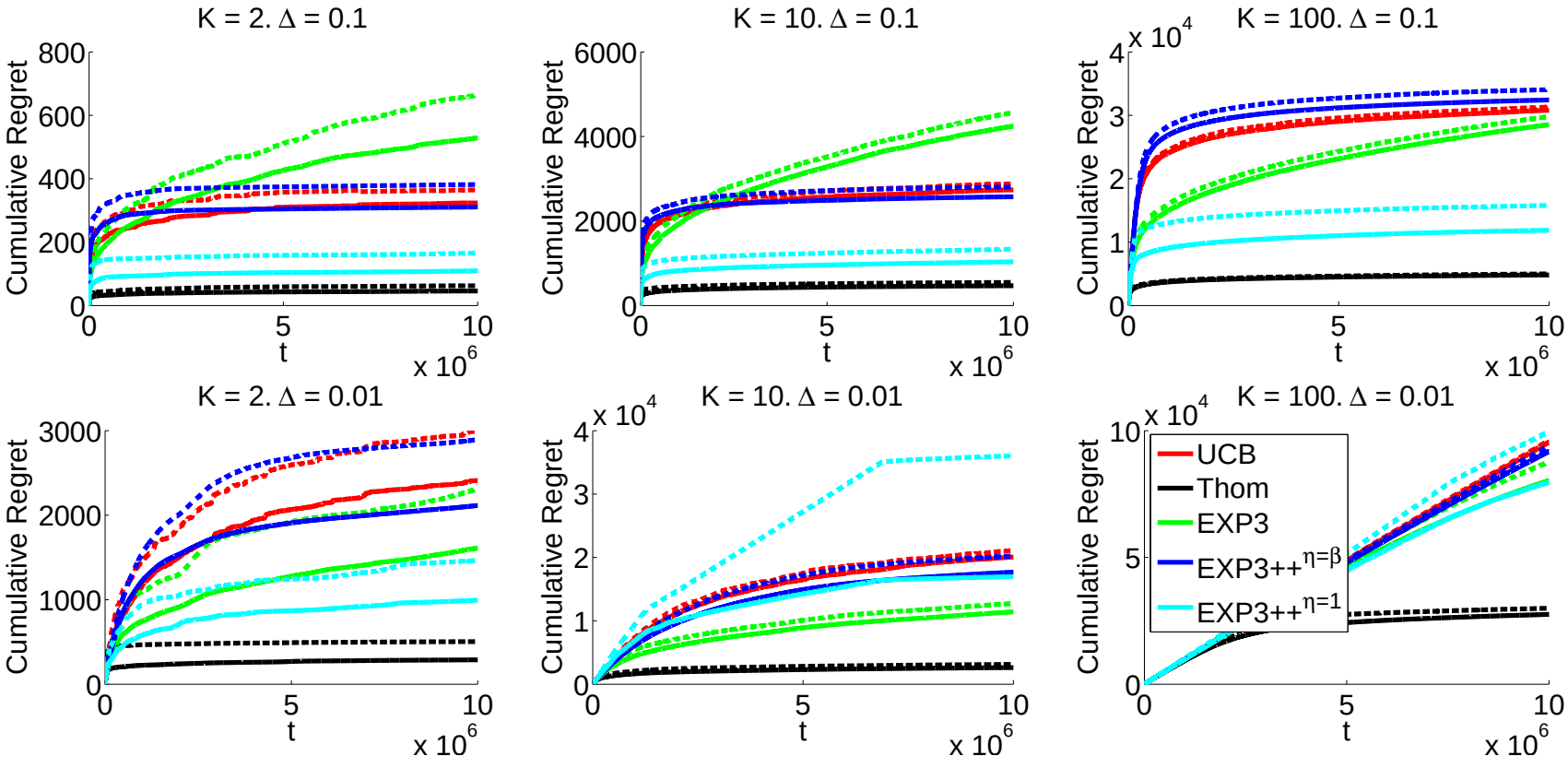
Let $\hat{\Delta}_t(a)$ be empirical estimate of $\Delta(a)$

- $\varepsilon_t(a) = \frac{18(\ln t)^2}{t\hat{\Delta}_t(a)^2}$ and $\eta_t \geq \frac{1}{2}\sqrt{\frac{\ln K}{tK}}$

  $\Rightarrow O\left(\frac{(\ln t)^3}{\Delta(a)}\right)$ regret in the stochastic regime, moderately contaminated stochastic, adversarial with a gap

- $\eta_t = \frac{1}{2}\sqrt{\frac{\ln K}{tK}}$ and $\varepsilon_t(a) = \frac{18(\ln t)^2}{t\hat{\Delta}_t(a)^2} \Rightarrow$ good for all four regimes

Experiments in the Stochastic Regime



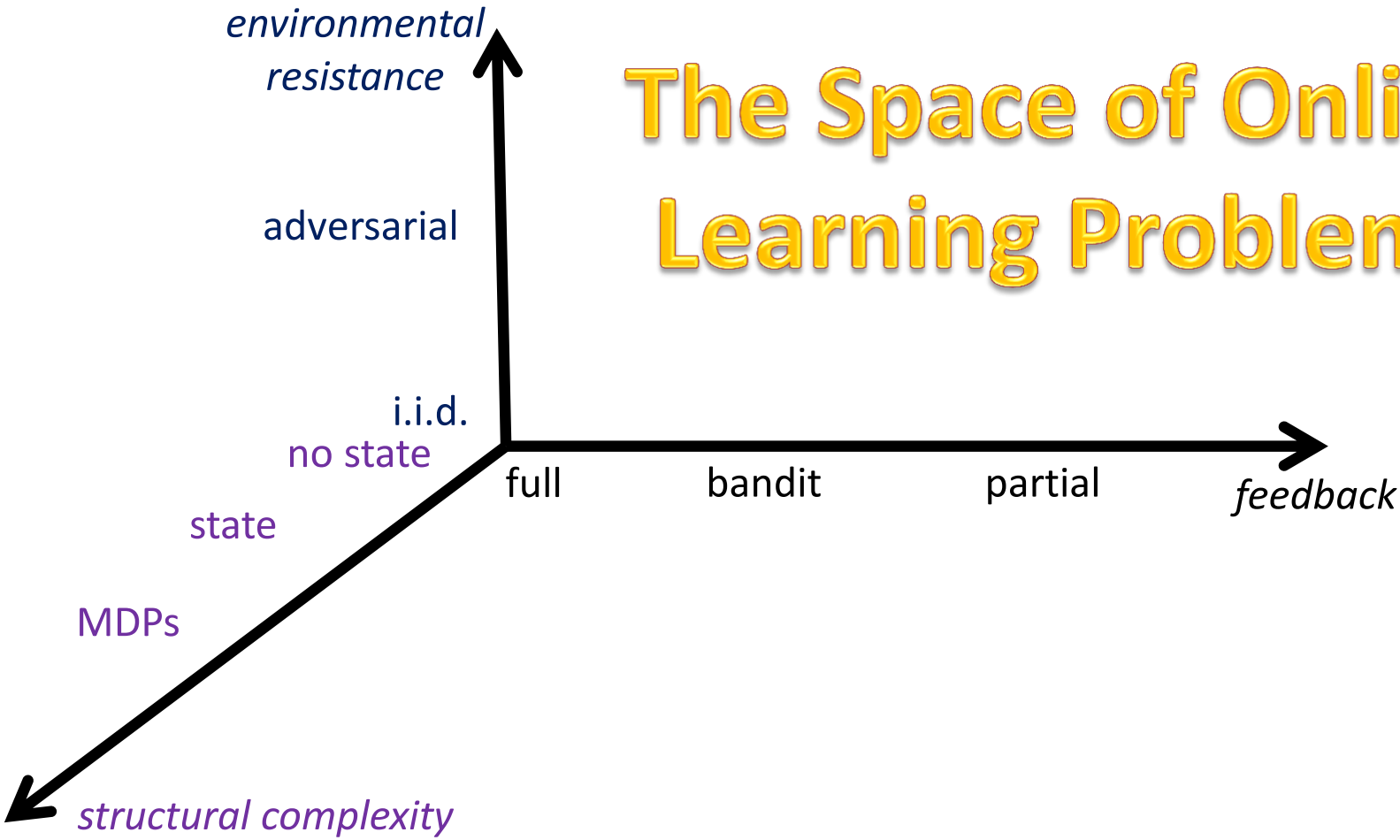
EXP3++ Summary

- EXP3++ simple and natural extension of EXP3
- Two control levers η_t and $\varepsilon_t(a)$-s
- Almost optimal performance in both stochastic and adversarial regimes
- "Logarithmic" regret in two new regimes
  - Moderately contaminated stochastic regime
  - Adversarial regime with a gap
- In the stochastic regime empirically comparable to UCB1

Punch Line

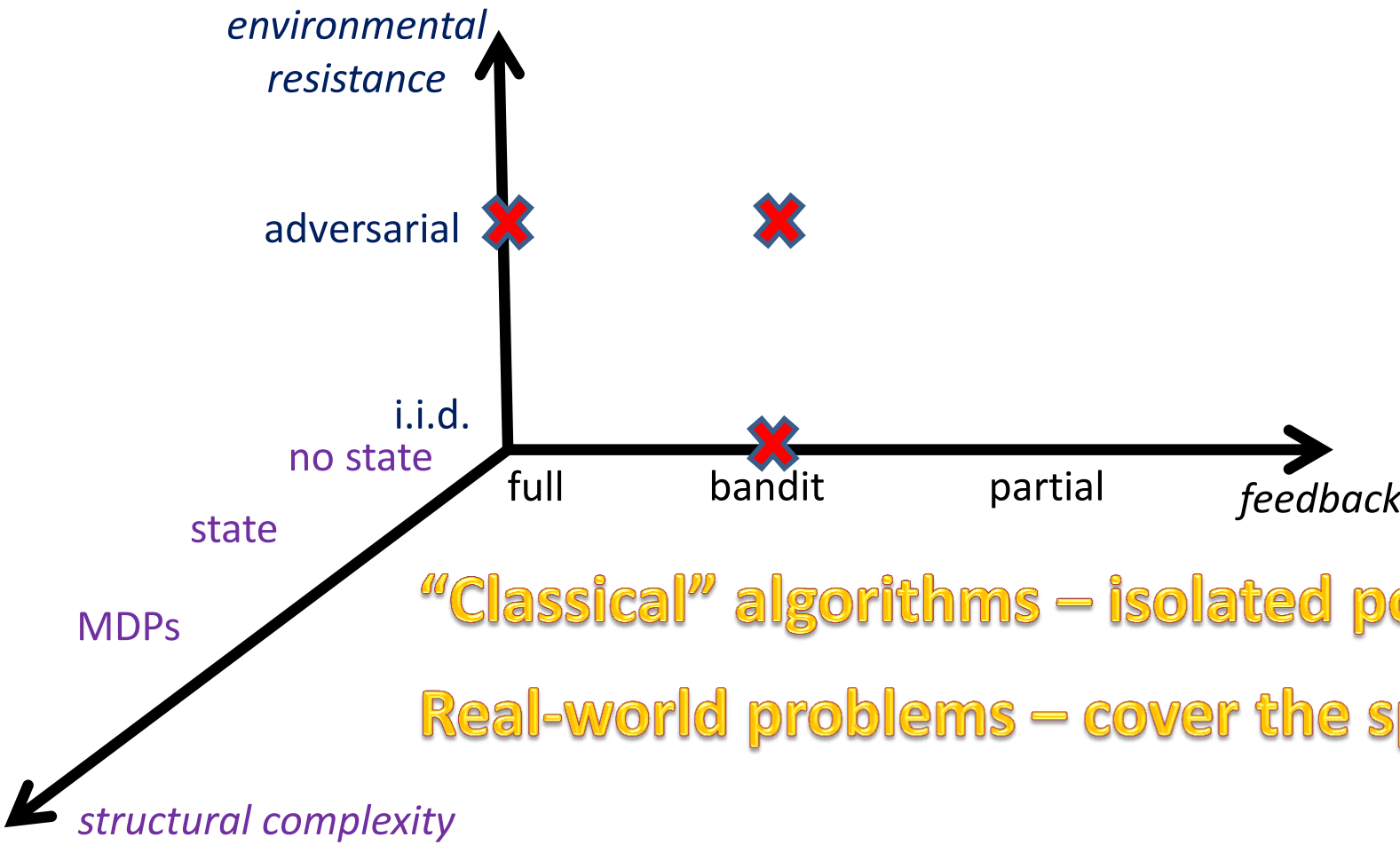
EXP3++ is a powerful tool for exploiting the gaps in a variety of regimes without compromising on the worst-case performance!

Summary



The Space of Online Learning Problems

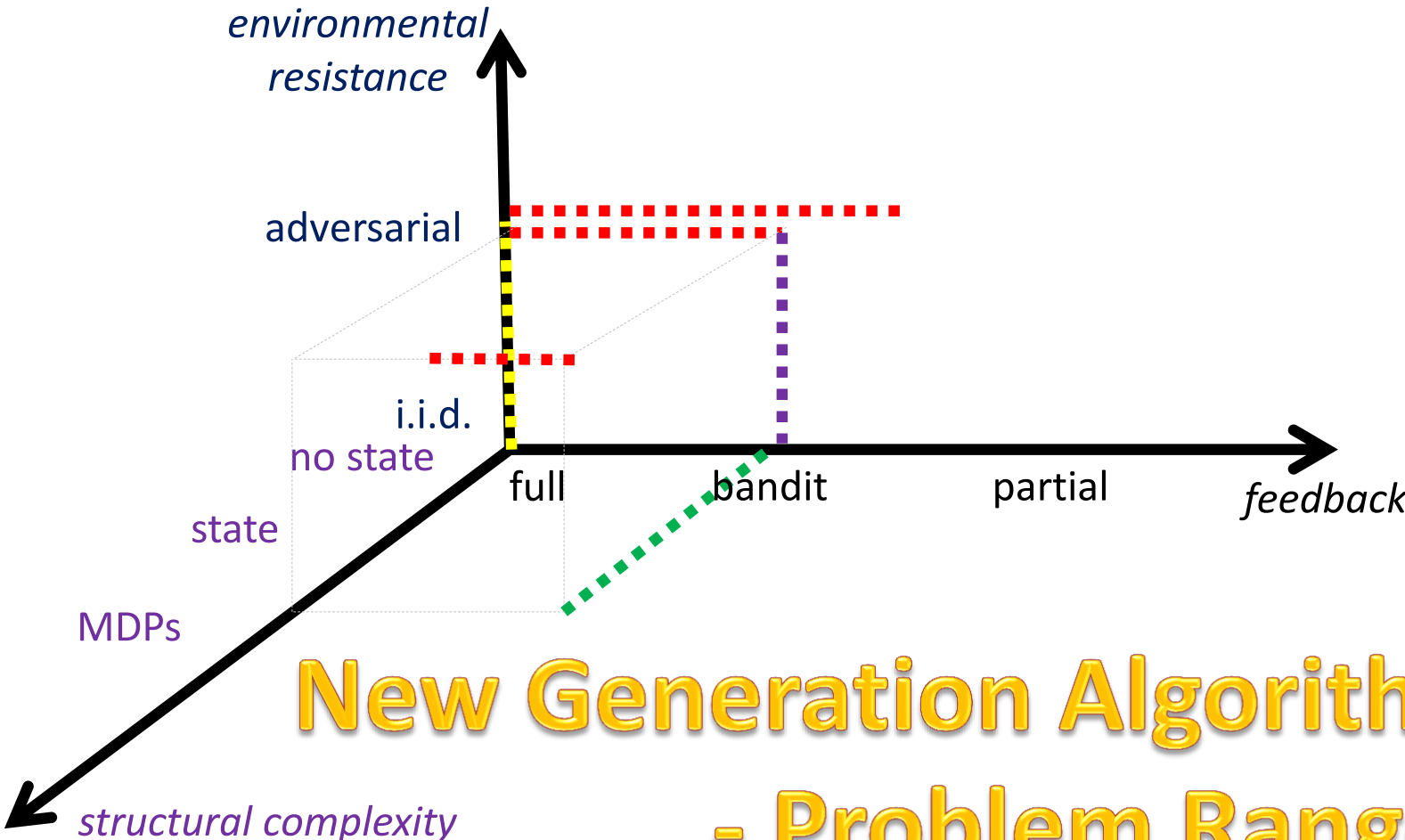
Summary



"Classical" algorithms – isolated points

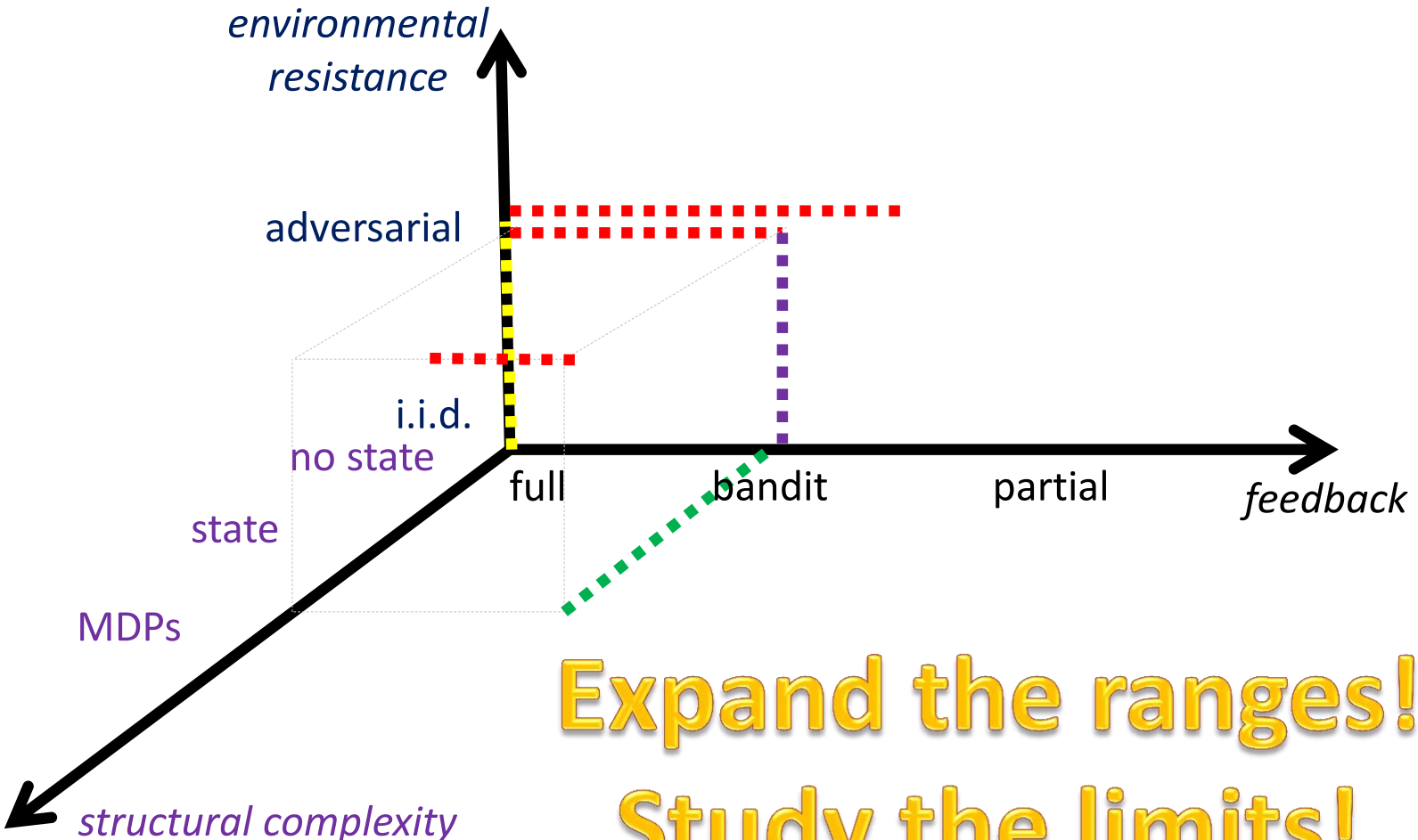
Real-world problems – cover the space

Summary



New Generation Algorithms - Problem Ranges

Future Work



Expand the ranges!

Study the limits!

Other popular problems we have not touched

- Linear bandits
- Combinatorial bandits
- Dueling bandits
- And many, many more variations ...

Further Reading Part 1

- Nicolò Cesa-Bianchi and Gábor Lugosi. Prediction, learning, and games. *Cambridge University Press*, 2006
- Shai Shalev-Shwartz. Online learning and online convex optimization. *Foundations and Trends in Machine Learning*, 2012
- Peter Auer, Nicolò Cesa-Bianchi, Yoav Freund, and Robert E. Schapire. The nonstochastic multiarmed bandit problem. *SIAM Journal on Computing*, 2002
- Peter Auer, Nicolò Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine Learning*, 2002
- Sébastien Bubeck and Nicolò Cesa-Bianchi. Regret Analysis of Stochastic and Nonstochastic Multi-armed Bandit Problems. In *Foundations and Trends in Machine Learning*, 2012

Further Reading Part 2

- Wouter M. Koolen and Tim Van Erven. Second-order quantile methods for experts and combinatorial games. *COLT*, 2015
- Haipeng Luo and Robert E. Schapire. Achieving all with no parameters: AdaNormalHedge. *COLT*, 2015
- Yevgeny Seldin and Aleksandrs Slivkins. One practical algorithm for both stochastic and adversarial bandits. *ICML*, 2014
- Yevgeny Seldin, Peter L. Bartlett, Koby Crammer, and Yasin Abbasi-Yadkori. Prediction with limited advice and multiarmed bandits with paid observations. *ICML*, 2014
- Satyen Kale. Multiarmed bandits with limited expert advice. *COLT*, 2014
- Yevgeny Seldin, Peter Auer, François Laviolette, John Shawe-Taylor, and Ronald Ortner. PAC-Bayesian analysis of contextual bandits. *NIPS*, 2011