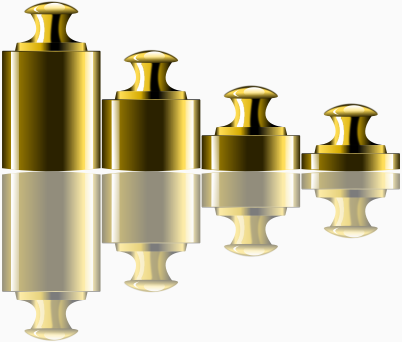
Luckiness in Multi-Scale Online Learning



Wouter M. Koolen

CWI Centrum Wiskunde & Informatica

UNIVERSITY OF TWENTE.

2^{nd} AI & Mathematics workshop
University of Twente
Friday 2^{nd} June, 2023

Advertising

Team effort

Muriel Felipe Pérez-Ortiz
PhD student at CWI

Outline

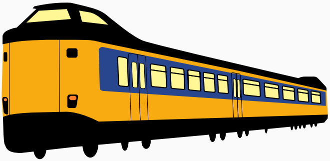
1. Motivation

2. Theory

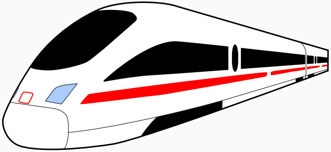
3. Application

Motivating Example

Every week I face the choice



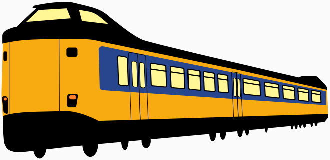
2h $\pm$ 10 min



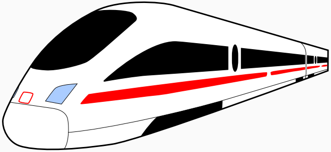
2h $\pm$ 30 min

Motivating Example

Every week I face the choice



2h $\pm$ 10 min



2h $\pm$ 30 min

I want:

- Total travel time $\leq$ best fixed carrier + **small learning overhead**
- By choosing my carrier adaptively (possibly randomised)
- With full information of past service
- Without relying on i.i.d. assumption

Why is this important/interesting

- Fundamental problem with strong connections to
  - martingale deviation inequalities
  - convex optimisation and duality
  - (stochastic) gradient descent
  - **uncertainty quantification**
  - bandit problems (partial information)
  - reinforcement learning
  - game theory (saddle point computation)
  - differential privacy
  - Boosting
  - ...

Why is this important/interesting

- Fundamental problem with strong connections to
  - martingale deviation inequalities
  - convex optimisation and duality
  - (stochastic) gradient descent
  - **uncertainty quantification**
  - bandit problems (partial information)
  - reinforcement learning
  - game theory (saddle point computation)
  - differential privacy
  - Boosting
  - ...
- Theory well-developed for single loss scale. (Freund and Schapire, 1997; De Rooij et al., 2014; Koolen, Grünwald, and Van Erven, 2016)

Why is this important/interesting

- Fundamental problem with strong connections to
  - martingale deviation inequalities
  - convex optimisation and duality
  - (stochastic) gradient descent
  - **uncertainty quantification**
  - bandit problems (partial information)
  - reinforcement learning
  - game theory (saddle point computation)
  - differential privacy
  - Boosting
  - ...
- Theory well-developed for single loss scale. (Freund and Schapire, 1997; De Rooij et al., 2014; Koolen, Grünwald, and Van Erven, 2016)
- Similar treatment for **multi-scale** was lacking.
  - Existing algorithm templates too rigid
  - No multi-scale Bernstein Inequality

Supervised Learning Theory

What is Statistical Learning

- Receive batch of i.i.d. labelled examples
- Output predictor for new data (e.g. by ERM)
- Prove risk bound using **concentration** (PAC, VC dim)

Supervised Learning Theory

What is Statistical Learning

- Receive batch of i.i.d. labelled examples
- Output predictor for new data (e.g. by ERM)
- Prove risk bound using **concentration** (PAC, VC dim)

What is Online Learning

- Cancel the i.i.d. assumption ($\Rightarrow$ adversary)
- Algorithm predicts/updates sequentially (e.g. by online gradient descent)
- Prove regret bound using **game theory** (minimax)

Supervised Learning Theory

What is Statistical Learning

- Receive batch of i.i.d. labelled examples
- Output predictor for new data (e.g. by ERM)
- Prove risk bound using **concentration** (PAC, VC dim)

What is Online Learning

- Cancel the i.i.d. assumption ($\Rightarrow$ adversary)
- Algorithm predicts/updates sequentially (e.g. by online gradient descent)
- Prove regret bound using **game theory** (minimax)

We say a **single algorithm** exploits Luckiness if

- Regret bounded (by minimax rate)
- Risk bounded by **fast rate** if i.i.d. with margin

Supervised Learning Theory

What is Statistical Learning

- Receive batch of i.i.d. labelled examples
- Output predictor for new data (e.g. by ERM)
- Prove risk bound using **concentration** (PAC, VC dim)

What is Online Learning

- Cancel the i.i.d. assumption ($\Rightarrow$ adversary)
- Algorithm predicts/updates sequentially (e.g. by online gradient descent)
- Prove regret bound using **game theory** (minimax)

We say a **single algorithm** exploits Luckiness if

- Regret bounded (by minimax rate)
- Risk bounded by **fast rate** if i.i.d. with margin

A learning problem is Multi-Scale if

- range of losses varies wildly between predictions

Philosophical Note

The learner is uncertain about the overall **best predictor**.

Need to maintain uncertainty. Vague: many implementations.

Online learning provides a crisp framework with a scalar objective.

Hence it informs us about optimal/good/appropriate ways to maintain uncertainty.

The answer is **far from** Bayesian (or perhaps profound generalisation)

Outline

1. Motivation
2. Theory
3. Application

Formal Setup

Fix number K of actions with **loss ranges** $\boldsymbol{\sigma} \in [0, \infty)^K$

Protocol

for $t = 1, 2, \ldots$

- Learner picks probability distribution $\boldsymbol{w}_t \in \triangle_K$ on actions
- Adversary sets action losses $\boldsymbol{\ell}_t \in \mathbb{R}^K$ with $|\ell_t^k| \leq \sigma_k$
- Learner incurs expected loss $\boldsymbol{w}_t^\intercal \boldsymbol{\ell}_t$

Formal Setup

Fix number K of actions with **loss ranges** $\boldsymbol{\sigma} \in [0, \infty)^K$

Protocol

for $t = 1, 2, \ldots$

- Learner picks probability distribution $\boldsymbol{w}_t \in \triangle_K$ on actions
- Adversary sets action losses $\boldsymbol{\ell}_t \in \mathbb{R}^K$ with $|\ell_t^k| \leq \sigma_k$
- Learner incurs expected loss $\boldsymbol{w}_t^\intercal \boldsymbol{\ell}_t$

Definition (Regret)

The **regret** after T rounds with respect to action k is

$$R_T^k = \sum_{t=1}^{T} \boldsymbol{w}_t^\intercal \boldsymbol{\ell}_t - \sum_{t=1}^{T} \ell_t^k$$

Formal Setup

Fix number K of actions with **loss ranges** $\boldsymbol{\sigma} \in [0, \infty)^K$

Protocol

for $t = 1, 2, \ldots$

- Learner picks probability distribution $\boldsymbol{w}_t \in \triangle_K$ on actions
- Adversary sets action losses $\boldsymbol{\ell}_t \in \mathbb{R}^K$ with $|\ell_t^k| \leq \sigma_k$
- Learner incurs expected loss $\boldsymbol{w}_t^\intercal \boldsymbol{\ell}_t$

Definition (Regret)

The **regret** after T rounds with respect to action k is

$$R_T^k = \sum_{t=1}^{T} \boldsymbol{w}_t^\intercal \boldsymbol{\ell}_t - \sum_{t=1}^{T} \ell_t^k$$

Question

Can Learner keep $R_T^k \leq \sigma_k \sqrt{T}$?

First Step

Consider Follow-the-Regularised-Leader (**FTRL**) template

$$\boldsymbol{w}_t \;=\; \operatorname*{argmin}_{\boldsymbol{w} \in \triangle_K} \; \langle \boldsymbol{w}, \boldsymbol{L}_{t-1} \rangle + D_{\boldsymbol{\eta}}(\boldsymbol{w}, \boldsymbol{u})$$

with **cumulative losses** $\boldsymbol{L}_t = \sum_{s=1}^t \boldsymbol{\ell}_s$ and multi-scale entropy

$$D_{\boldsymbol{\eta}}(\boldsymbol{w}, \boldsymbol{u}) \;=\; \sum_k \frac{w_k \ln \frac{w_k}{u_k} - w_k + u_k}{\eta_k}.$$

First Step

Consider Follow-the-Regularised-Leader (**FTRL**) template

$$\boldsymbol{w}_t \;=\; \underset{\boldsymbol{w} \in \triangle_K}{\operatorname{argmin}} \;\; \langle \boldsymbol{w}, \boldsymbol{L}_{t-1} \rangle + D_{\boldsymbol{\eta}}(\boldsymbol{w}, \boldsymbol{u})$$

with **cumulative losses** $\boldsymbol{L}_t = \sum_{s=1}^{t} \boldsymbol{\ell}_s$ and multi-scale entropy

$$D_{\boldsymbol{\eta}}(\boldsymbol{w}, \boldsymbol{u}) \;=\; \sum_k \frac{w_k \ln \frac{w_k}{u_k} - w_k + u_k}{\eta_k}.$$

Theorem (Bubeck et al., 2019)

FTRL with learning rate $\eta_k = \frac{1}{\sigma_k}\sqrt{\frac{2 \ln K}{T}}$ *has regret bounded by*

$$R_T^k \;\leq\; \sigma_k \sqrt{T \ln K}.$$

Matching worst-case regret lower bound.

Are we there yet?

Luckiness?

What if losses $\boldsymbol{\ell}_1, \boldsymbol{\ell}_2, \dots$ turn out to be i.i.d. after all?

Are we there yet?

Luckiness?

What if losses $\ell_1, \ell_2, \ldots$ turn out to be i.i.d. after all?

Actually, the worst-case lower bound example is of that kind.

Are we there yet?

Luckiness?

What if losses $\boldsymbol{\ell}_1, \boldsymbol{\ell}_2, \dots$ turn out to be i.i.d. after all?

Actually, the worst-case lower bound example is of that kind.

Luckiness with Margin?

Losses sampled i.i.d. $\boldsymbol{\ell}_t \sim \mathbb{P}$, where mean loss vector $\boldsymbol{\mu} = \mathbb{E}[\boldsymbol{\ell}]$ exhibits **positive gap** $\Delta = \mu_{(2)} - \mu_{(1)} > 0$

Are we there yet?

Luckiness?

What if losses $\boldsymbol{\ell}_1, \boldsymbol{\ell}_2, \ldots$ turn out to be i.i.d. after all?

Actually, the worst-case lower bound example is of that kind.

Luckiness with Margin?

Losses sampled i.i.d. $\boldsymbol{\ell}_t \sim \mathbb{P}$, where mean loss vector $\boldsymbol{\mu} = \mathbb{E}[\boldsymbol{\ell}]$ exhibits **positive gap** $\Delta = \mu_{(2)} - \mu_{(1)} > 0$

Still no cigar. $\Omega(\sqrt{T})$ regret.

Are we there yet?

Luckiness?

What if losses $\boldsymbol{\ell}_1, \boldsymbol{\ell}_2, \ldots$ turn out to be i.i.d. after all?

Actually, the worst-case lower bound example is of that kind.

Luckiness with Margin?

Losses sampled i.i.d. $\boldsymbol{\ell}_t \sim \mathbb{P}$, where mean loss vector $\boldsymbol{\mu} = \mathbb{E}[\boldsymbol{\ell}]$ exhibits **positive gap** $\Delta = \mu_{(2)} - \mu_{(1)} > 0$

Still no cigar. $\Omega(\sqrt{T})$ regret.

Strong Contrast

For same-scale case there is a single algorithm with

- Worst-case regret $\sqrt{T \ln K}$ (matching lower bound)
- Stochastic+gap regret $O(1/\Delta)$, a constant(!)
- Interpolates spectrum by data-dependent $\sqrt{V_T \ln K}$ bound.

Main Result

Muscada Algorithm

$$\boldsymbol{w}_t \;:=\; \operatorname*{argmin}_{\boldsymbol{w} \in \triangle_K} \; \langle \boldsymbol{w}, \boldsymbol{L}_{t-1} + \boldsymbol{\mu}_{t-1} \rangle + D_{\boldsymbol{\eta}_{t-1}}(\boldsymbol{w}, \boldsymbol{u})$$

where

$$\mu_t^k = \sigma_k \sqrt{v_t \ln K}$$

$$v_t = 4 \sum_{s=1}^{t} \frac{\mathrm{var}_{\tilde{\boldsymbol{w}}_s}(\boldsymbol{\ell}_s)}{\langle \tilde{\boldsymbol{w}}_s, \boldsymbol{\sigma}^2 \rangle} \qquad \text{with} \qquad \tilde{w}_t^k \propto w_t^k \eta_{t-1}^k$$

$$\eta_t^k = \frac{1}{\sigma_k} \sqrt{\frac{2 \ln K}{v_t}}$$

Main Result

Muscada Algorithm

$$\boldsymbol{w}_t := \operatorname*{argmin}_{\boldsymbol{w} \in \triangle_K} \langle \boldsymbol{w}, \boldsymbol{L}_{t-1} + \boldsymbol{\mu}_{t-1} \rangle + D_{\boldsymbol{\eta}_{t-1}}(\boldsymbol{w}, \boldsymbol{u})$$

where

$$\mu_t^k = \sigma_k \sqrt{v_t \ln K}$$

$$v_t = 4 \sum_{s=1}^{t} \frac{\mathrm{var}_{\tilde{\boldsymbol{w}}_s}(\boldsymbol{\ell}_s)}{\langle \tilde{\boldsymbol{w}}_s, \boldsymbol{\sigma}^2 \rangle} \quad \text{with} \quad \tilde{w}_t^k \propto w_t^k \eta_{t-1}^k$$

$$\eta_t^k = \frac{1}{\sigma_k} \sqrt{\frac{2 \ln K}{v_t}}$$

Theorem (Main Result)

Muscada guarantees $R_T^k \leq \mu_T^k$.

- *Sharpens worst-case regret bound of Bubeck et al. 2019 as* $v_t \leq t$.
- *In i.i.d. setting with gap* Δ, *expected regret is* **constant** $\mathbb{E}[\mu_T^{k^*}] \leq \frac{K\sigma_{\max}^2}{\Delta}$

Anatomy of the Muscada weights

$$w_{t+1}^{k} = \frac{1}{K} \exp\left(-\frac{L_t^k + \lambda_t^*}{\sigma_k \sqrt{v_t}}\right)$$

where λ_t^* ensures normalisation

Anatomy of the Muscada weights

$$w_{t+1}^{k} = \frac{1}{K} \exp\left(-\frac{L_t^k + \lambda_t^*}{\sigma_k \sqrt{v_t}} \right)$$

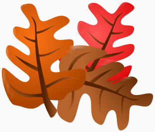
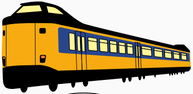
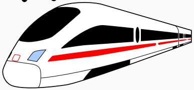
where λ_t^* ensures normalisation

- More loss, less weight
- Evidence in loss decays with "time" v_t
- Loss and normalisation λ_t^* affects large scales σ_k less.

Outline

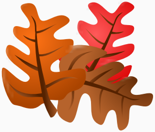
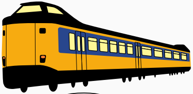
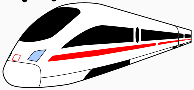
Saddle Point Computation

Now suppose I am paranoid about travel time.

		
	10	90
	0	-10

Saddle Point Computation

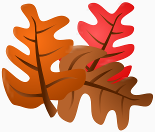
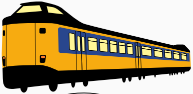
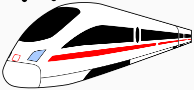
Now suppose I am paranoid about travel time.

		
	10	90
	0	-10

What is the saddle point? (Hint: it is pure)

Saddle Point Computation

Now suppose I am paranoid about travel time.

		
	10	90
	0	-10

What is the saddle point? (Hint: it is pure)

Multi-scale:  and  of scale 90 while  and  of scale 10.

In General

Problem

Given (large) payoff matrix M. Compute an ϵ-equilibrium $(\boldsymbol{p}, \boldsymbol{q})$:

$$\max_j \boldsymbol{p}^\intercal M \boldsymbol{e}_j - \min_i \boldsymbol{e}_i^\intercal M \boldsymbol{q} \;\leq\; \epsilon.$$

Popular approach (Freund and Schapire, 1999)

Run online learners $\boldsymbol{p}_t$ and $\boldsymbol{q}_t$ on loss vectors $M\boldsymbol{q}_t$ and $-M^\intercal \boldsymbol{p}_t$.

In General

Problem

Given (large) payoff matrix M. Compute an ϵ-equilibrium $(\boldsymbol{p}, \boldsymbol{q})$:

$$\max_j \boldsymbol{p}^\intercal M \boldsymbol{e}_j - \min_i \boldsymbol{e}_i^\intercal M \boldsymbol{q} \ \leq \ \epsilon.$$

Popular approach (Freund and Schapire, 1999)

Run online learners $\boldsymbol{p}_t$ and $\boldsymbol{q}_t$ on loss vectors $M\boldsymbol{q}_t$ and $-M^\intercal \boldsymbol{p}_t$.

Question

Does multi-scale knowledge help?

In General

Problem

Given (large) payoff matrix M. Compute an ϵ-equilibrium $(\boldsymbol{p}, \boldsymbol{q})$:

$$\max_j \boldsymbol{p}^\intercal M \boldsymbol{e}_j - \min_i \boldsymbol{e}_i^\intercal M \boldsymbol{q} \;\le\; \epsilon.$$

Popular approach (Freund and Schapire, 1999)

Run online learners $\boldsymbol{p}_t$ and $\boldsymbol{q}_t$ on loss vectors $M\boldsymbol{q}_t$ and $-M^\intercal \boldsymbol{p}_t$.

Question

Does multi-scale knowledge help?

Yes, sub-optimality gap improves from $\sigma_{\text{max}}/\sqrt{T}$ to $\sigma_{\text{saddle-point}}/\sqrt{T}$.

With optimism (Rakhlin and Sridharan, 2013), empirically $\sigma_{\text{saddle-point}}/T$.

Thanks!

References i

Bubeck, S., N. R. Devanur, Z. Huang, and R. Niazadeh (2019). "Multi-scale Online Learning: Theory and Applications to Online Auctions and Pricing". In: **Journal of Machine Learning Research** 20.62, pp. 1–37.

Freund, Y. and R. E. Schapire (1999). "Adaptive game playing using multiplicative weights". In: **Games and Economic Behavior** 29.1-2, pp. 79–103.

— (Aug. 1997). "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting". en. In: **Journal of Computer and System Sciences** 55.1, pp. 119–139.

Koolen, W. M., P. D. Grünwald, and T. van Erven (Dec. 2016). "Combining Adversarial Guarantees and Stochastic Fast Rates in Online Learning". In: **Advances in Neural Information Processing Systems (NeurIPS) 29**, pp. 4457–4465.

Rakhlin, S. and K. Sridharan (2013). “Optimization, Learning, and Games with Predictable Sequences”. In: **Advances in Neural Information Processing Systems**. Vol. 26. Curran Associates, Inc.

de Rooij, S., T. van Erven, P. D. Grünwald, and W. M. Koolen (Apr. 2014). “Follow the Leader If You Can, Hedge If You Must”. In: **Journal of Machine Learning Research** 15, pp. 1281–1316.