

Sequential Forecasting of Individual Sequences

Gilles Stoltz

CNRS – École normale supérieure – HEC Paris

- 1 The model of individual sequences
 - Experts and regret
 - Randomized prediction
- 2 Optimal bounds on the regret in full-information
 - General $\sqrt{n \ln N}$ upper bound on the regret
 - Lower bounds on the regret
 - Refined upper bounds
- 3 Extensions: Aggregation of predictors, limited feedback
 - Aggregation of predictors, application to air-quality forecasting
 - Regret in multi-armed bandit problems
 - Minimizing the regret under partial monitoring

A statistician has to predict a sequence y_1, y_2, \dots of outcomes drawn from a set \mathcal{Y} .

His predictions $\hat{p}_1, \hat{p}_2, \dots$ are picked in a set \mathcal{X} .

Observations and predictions are made in a **sequential fashion** and rely on **no stochastic model**.

At each round, y_t is predicted based on the past,
 $y_1^{t-1} = (y_1, \dots, y_{t-1})$.

The prediction \hat{p}_t is formed before the actual value y_t is revealed and is then compared to it.

Example: weather forecasting, $\mathcal{Y} = \{0, 1\}$ (whether it will rain or not), and $\mathcal{X} = [0, 1]$ (a probability of rain).

To make the problem meaningful, **experts** are called for; they are indexed by $j = 1, \dots, N$.

At each round, expert j outputs a prediction $f_{j,t} = f_{j,t}(y_1^{t-1}) \in \mathcal{X}$.

The statistician now bases his predictions \hat{p}_t on **past outcomes** y_1^{t-1} and on **past and present expert advices**, $f_{j,s}$ with $s = 1, \dots, t$.

The statistician aims at predicting almost as well as the best expert.

Note that the best expert can only be determined **in hindsight** whereas the statistician has to predict in a **sequential fashion**.

The notion of best expert has to be quantified, we do so by means of a loss function $\ell : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$.

Example (weather forecasting, continued): $\mathcal{X} = [0, 1]$, $\mathcal{Y} = \{0, 1\}$ and $\ell(x, y) = |x - y|$.

Given a loss function $\ell : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$, we define the cumulative losses of the statistician and of each expert j ,

$$\widehat{L}_n = \sum_{t=1}^n \ell(\widehat{p}_t, y_t) \quad \text{and} \quad L_{j,n} = \sum_{t=1}^n \ell(f_{j,t}, y_t)$$

The **regret** is the difference between these quantities,

$$R_{j,n} = \widehat{L}_n - L_{j,n} = \sum_{t=1}^n \ell(\widehat{p}_t, y_t) - \sum_{t=1}^n \ell(f_{j,t}, y_t)$$

We ask for prediction strategies such that the per-round regret converges to 0 for all outcome sequences y_1, y_2, \dots ,

$$\frac{1}{n} \max_{j=1, \dots, N} R_{j,n} \longrightarrow 0$$

This is, in general, **not** achievable by a **deterministic** strategy. (See $\mathcal{X} = \mathcal{Y} = \{0, 1\}$ and $\ell(x, y) = \mathbb{I}_{\{x \neq y\}}$.)

This is due to the **worst-case assessment**, which amounts to playing against an adversary reading the statistician's mind.

- 1 The model of individual sequences
 - Experts and regret
 - Randomized prediction
- 2 Optimal bounds on the regret in full-information
 - General $\sqrt{n \ln N}$ upper bound on the regret
 - Lower bounds on the regret
 - Refined upper bounds
- 3 Extensions: Aggregation of predictors, limited feedback
 - Aggregation of predictors, application to air-quality forecasting
 - Regret in multi-armed bandit problems
 - Minimizing the regret under partial monitoring

To surprise the adversary, we resort to an **auxiliary randomization**.

Without any specific assumption (e.g., convexity) on the loss function ℓ , we use this randomization to aggregate expert advices in expectation.

A **randomized strategy** is a sequence of functions; the t -th of them

- associates to past outcomes y_1^{t-1} and past and present experts advices $f_{j,s}$ with $s = 1, \dots, t$,
- a probability distribution $\mathbf{p}_t = (p_{1,t}, \dots, p_{N,t})$ over the experts.

The predictions are formed by drawing an expert index I_t at random according to \mathbf{p}_t , i.e., $\hat{p}_t = f_{I_t,t}$.

Outcomes may be generated independently of the statistician's predictions (**statistical case**, e.g., weather forecasting) or by an adversary reacting to the predictions (**game-theoretic case**). The adversary also has a strategy, then.

The setting may be summarized as a **repeated game** between a statistician and an adversary; the former (resp., the latter) aims at minimizing (resp., maximizing) the regret.

Parameters: prediction set \mathcal{X} , outcome set \mathcal{Y} , loss function $\ell : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$

For each round $t = 1, 2, \dots$,

- experts form their advices $f_{j,t}$;
- the statistician chooses a probability distribution \mathbf{p}_t over $\{1, \dots, N\}$, picks an expert index I_t at random according to it, and predicts $\hat{p}_t = f_{I_t,t}$;
- the adversary chooses simultaneously the outcome y_t ;
- y_t and \hat{p}_t are both revealed and losses may be computed.

Remark: The adversary may choose the experts and be aware of the statistician's strategy (thus, knowing the \mathbf{p}_t when choosing y_t).

- 1 The model of individual sequences
 - Experts and regret
 - Randomized prediction
- 2 Optimal bounds on the regret in full-information
 - General $\sqrt{n \ln N}$ upper bound on the regret
 - Lower bounds on the regret
 - Refined upper bounds
- 3 Extensions: Aggregation of predictors, limited feedback
 - Aggregation of predictors, application to air-quality forecasting
 - Regret in multi-armed bandit problems
 - Minimizing the regret under partial monitoring

Recall that we want to minimize the **regret**

$$\widehat{L}_n - \min_{j=1,\dots,N} L_{j,n} = \sum_{t=1,\dots,n} \ell(f_{I_t,t}, y_t) - \min_{j=1,\dots,N} \sum_{t=1,\dots,n} \ell(f_{j,t}, y_t)$$

and that I_t is drawn at random according to \mathbf{p}_t .

Denote by \mathbb{E}_t the **conditional expectation** at round t ,

$$\mathbb{E}_t [\ell(f_{I_t,t}, y_t)] = \sum_{i=1,\dots,N} p_{i,t} \ell(f_{i,t}, y_t)$$

By **martingale convergence** (when ℓ is bounded),

$$\widehat{L}_n - \bar{L}_n = \sum_{t=1,\dots,n} \ell(f_{I_t,t}, y_t) - \sum_{t=1,\dots,n} \sum_{i=1,\dots,N} p_{i,t} \ell(f_{i,t}, y_t) = o_{\mathbb{P}}(n)$$

We may thus focus on the **expected regret** $\bar{R}_n = \bar{L}_n - \min_{j=1,\dots,N} L_{j,n}$,

$$\bar{R}_n = \sum_{t=1,\dots,n} \sum_{i=1,\dots,N} p_{i,t} \ell(f_{i,t}, y_t) - \min_{j=1,\dots,N} \sum_{t=1,\dots,n} \ell(f_{j,t}, y_t)$$

Since we are interested in convergence rates, we shall be more specific.

Assume ℓ takes values in $[0, 1]$; then **Hoeffding–Azuma** inequality ensures that with probability at least $1 - \delta$,

$$\Delta_n = \widehat{L}_n - \bar{L}_n = \sum_{t=1}^n \ell(f_{I_t, t}, y_t) - \sum_{t=1}^n \sum_{i=1}^N p_{i,t} \ell(f_{i,t}, y_t) \leq \sqrt{\frac{n}{2} \ln \frac{1}{\delta}}$$

As a consequence (as $R_n = \bar{R}_n + \Delta_n$), with probability $1 - \delta$,

$$\begin{aligned} & \sum_{t=1}^n \ell(f_{I_t, t}, y_t) - \min_{j=1, \dots, N} \sum_{t=1}^n \ell(f_{j,t}, y_t) \\ \leq & \left(\sum_{t=1}^n \ell(f_{I_t, t}, y_t) - \sum_{t=1}^n \sum_{i=1}^N p_{i,t} \ell(f_{i,t}, y_t) \right) + \left(\sum_{t=1}^n \sum_{i=1}^N p_{i,t} \ell(f_{i,t}, y_t) - \min_{j=1, \dots, N} \sum_{t=1}^n \ell(f_{j,t}, y_t) \right) \\ \leq & \left(\sum_{t=1}^n \sum_{i=1}^N p_{i,t} \ell(f_{i,t}, y_t) - \min_{j=1, \dots, N} \sum_{t=1}^n \ell(f_{j,t}, y_t) \right) + \sqrt{\frac{n}{2} \ln \frac{1}{\delta}} \end{aligned}$$

The idea of exponential reweighting is to assign a higher probability to better-performing actions (but of course not a probability 1 to the best action—it, in some sense, **smoothes fictitious play**).

Exponentially weighted average predictor

p_1 is uniform and for $t \geq 2$,

$$p_{i,t} = \frac{\exp\left(-\eta \sum_{s=1}^{t-1} \ell(f_{i,s}, y_s)\right)}{\sum_{j=1}^N \exp\left(-\eta \sum_{s=1}^{t-1} \ell(f_{j,s}, y_s)\right)} = \frac{\exp(-\eta L_{i,t-1})}{\sum_{j=1}^N \exp(-\eta L_{j,t-1})}$$

where $\eta > 0$ is a parameter to be tuned.

This strategy was introduced by Vovk '90, Littlestone and Warmuth '94. (See also Fudenberg and Levine '95, Cesa-Bianchi, Freund, Helmbold, Haussler, Schapire, and Warmuth '97, Cesa-Bianchi and Lugosi '99.)

An important assumption for the analysis is that the loss function takes bounded values, $\ell : \mathcal{X} \times \mathcal{Y} \rightarrow [0, 1]$.

Theorem

For *all strategies* τ of the opponent player, the expected regret is bounded as

$$\sum_{t=1}^n \sum_{i=1}^N p_{i,t} \ell(f_{i,t}, y_t) - \min_{j=1, \dots, N} \sum_{t=1}^n \ell(f_{j,t}, y_t) \leq \frac{\ln N}{\eta} + \frac{\eta n}{8} = \sqrt{\frac{n}{2} \ln N}$$

with $\eta = \sqrt{8 \ln N / n}$.

Thus, with probability $1 - \delta$, the true regret $R_n \leq \square \sqrt{n \ln(N/\delta)}$.

We now **prove** this bound, and discuss later some improvements (more particularly the **tuning of η**).

In the proof, the following lemma, due to **Hoeffding**, will be the key step:

Lemma

A bounded random variable X (say $0 \leq X \leq 1$) satisfies, for all $s > 0$,

$$-s \mathbb{E}[X] \leq \ln \mathbb{E} \left[e^{-sX} \right] \leq -s \mathbb{E}[X] + \frac{s^2}{8}$$

Recall that $p_{i,t} = w_{i,t-1}/W_{t-1}$, where $W_{t-1} = w_{1,t-1} + \dots + w_{N,t-1}$, $w_{i,0} = 1$, and for $t \geq 2$,

$$w_{i,t-1} = \exp(-\eta L_{i,t-1}) = \exp\left(-\eta \sum_{s=1}^{t-1} \ell(f_{i,s}, y_s)\right)$$

On the one hand,

$$\ln \frac{W_n}{W_0} \geq \ln \frac{\max_{j=1,\dots,N} w_{j,n}}{N} = -\eta \min_{j=1,\dots,N} L_{j,n} - \ln N$$

On the other hand, for $t = 1, \dots, n$,

$$\begin{aligned} \ln \frac{W_t}{W_{t-1}} &= \ln \frac{\sum_{i=1}^N e^{-\eta \ell(f_{i,t}, y_t)} w_{i,t-1}}{\sum_{j=1}^N w_{j,t-1}} = \ln \sum_{i=1}^N p_{i,t} e^{-\eta \ell(f_{i,t}, y_t)} \\ &\leq -\eta \left(\sum_{i=1}^N p_{i,t} \ell(f_{i,t}, y_t) \right) + \frac{\eta^2}{8} \end{aligned}$$

Summing the upper bounds over $t = 1, \dots, n$ and combining with the lower bound,

$$\sum_{t=1}^n \sum_{i=1}^N p_{i,t} \ell(f_{i,t}, y_t) - \min_{j=1,\dots,N} \sum_{t=1}^n \ell(f_{j,t}, y_t) \leq \frac{\ln N}{\eta} + \frac{\eta n}{8}$$

For the **tuning of η** , we can use a “doubling trick” or use all information available by allowing η to depend on time,

$$p_{j,t} = \frac{\exp\left(-\eta \sum_{s=1}^{t-1} \ell(f_{j,s}, y_s)\right)}{\sum_{k=1}^N \exp\left(-\eta \sum_{s=1}^{t-1} \ell(f_{k,s}, y_s)\right)}$$

where $\eta_t = \sqrt{8 \ln N / (t-1)}$.

Auer, Cesa-Bianchi, and Gentile '02 show that the (expected) regret \bar{R}_n of this strategy is less than $1 + 2\sqrt{(n/2) \ln N}$.

These tuning issues can also be overcome by choosing a different **potential function**.

For the **tuning of η** , we can use a “doubling trick” or use all information available by allowing η to depend on time,

$$p_{j,t} = \frac{\exp\left(-\eta_t \sum_{s=1}^{t-1} \ell(f_{j,s}, y_s)\right)}{\sum_{k=1}^N \exp\left(-\eta_t \sum_{s=1}^{t-1} \ell(f_{k,s}, y_s)\right)}$$

where $\eta_t = \sqrt{8 \ln N / (t-1)}$.

Auer, Cesa-Bianchi, and Gentile '02 show that the (expected) regret \bar{R}_n of this strategy is less than $1 + 2\sqrt{(n/2) \ln N}$.

These tuning issues can also be overcome by choosing a different **potential function**.

A **polynomial potential** function is $x \mapsto (\max\{x, 0\})^{p-1} = x_+^{p-1}$.

Instead of playing with

$$p_{j,t} = \frac{\exp\left(\eta\left(\sum_{s=1}^{t-1} \sum_{i=1}^N p_{i,s} \ell(f_{i,s}, y_s) - \sum_{s=1}^{t-1} \ell(f_{j,s}, y_s)\right)\right)}{\sum_{k=1}^N \exp\left(\eta\left(\sum_{s=1}^{t-1} \sum_{i=1}^N p_{i,s} \ell(f_{i,s}, y_s) - \sum_{s=1}^{t-1} \ell(f_{k,s}, y_s)\right)\right)}$$

A **polynomial potential** function is $x \mapsto (\max\{x, 0\})^{p-1} = x_+^{p-1}$.

We let

$$p_{j,t} = \frac{\left(\sum_{s=1}^{t-1} \sum_{i=1}^N p_{i,s} \ell(f_{i,s}, y_s) - \sum_{s=1}^{t-1} \ell(f_{j,s}, y_s) \right)_+^{p-1}}{\sum_{k=1}^N \left(\sum_{s=1}^{t-1} \sum_{i=1}^N p_{i,s} \ell(f_{i,s}, y_s) - \sum_{s=1}^{t-1} \ell(f_{k,s}, y_s) \right)_+^{p-1}}$$

Cesa-Bianchi and Lugosi '03 prove that for $p \geq 2$, the regret is less than $\sqrt{npN^{2/p}} \sim \square \sqrt{n \ln N}$.

They also sketch the link of this strategy to **Blackwell's approachability** theorem, in case the negative orthant is approached.

But it seems that the exponential potential is **easier to use** in setups with **imperfect monitoring**... This is why we henceforth focus on it.

- 1 The model of individual sequences
 - Experts and regret
 - Randomized prediction
- 2 Optimal bounds on the regret in full-information
 - General $\sqrt{n \ln N}$ upper bound on the regret
 - Lower bounds on the regret
 - Refined upper bounds
- 3 Extensions: Aggregation of predictors, limited feedback
 - Aggregation of predictors, application to air-quality forecasting
 - Regret in multi-armed bandit problems
 - Minimizing the regret under partial monitoring

Lower bounds are usually proved and stated with **oblivious opponents** simply choosing i.i.d. outcomes Y_t .

(Thus, a deterministic but hard to predict sequence of outcomes y_1, \dots, y_n could be fixed in advance.)

More precisely, we write

$$\inf_{\sigma} \sup_{\tau} \bar{R}_n(\sigma, \tau) \geq \inf_{\sigma} \sup_{y_1^n \in \mathcal{Y}^n} \bar{R}_n(\sigma, y_1^n) \geq \inf_{\sigma} \sup_{\mathbb{P}} \mathbb{E} \left[\bar{R}_n(\sigma, Y_1^n) \right]$$

where the infimum is over all statistician's strategies σ and the supremum is over all strategies τ of the adversary.

Cesa-Bianchi, Freund, Haussler, Helmbold, Schapire, and Warmuth '97 were the first to prove a lower bound on the **expected regret** \bar{R}_n matching the obtained upper bound.

Theorem

For $\mathcal{Y} = [0, 1]$, there exists a loss function $\ell : \mathbb{N}^* \times \mathcal{Y} \rightarrow \{0, 1\}$ and $\gamma, c > 0$ such that for all $N \geq 2$, $n \geq \gamma \ln N$, and all strategies of the statistician given the constant experts $f_{j,t} \equiv j$,

$$\sup_{y_1, \dots, y_n \in \mathcal{Y}} \left(\sum_{t=1}^n \sum_{i=1}^N p_{i,t} \ell(i, y_t) - \min_{j=1, \dots, N} \sum_{t=1}^n \ell(j, y_t) \right) \geq c \sqrt{n \ln N}$$

(It of course extends to non-expected regret R_n via Hoeffding's inequality.)

A simple but loose argument shows that the minimax regret is larger than $\square \sqrt{n}$. (We deal with the **extra $\sqrt{\ln N}$** term later on.)

We take $\mathcal{X} = \mathcal{Y} = \{0, 1\}$ and consider the loss function $\ell(x, y) = |x - y|$. The two **constant experts** predict respectively 0 and 1 at each round.

Let Y_1, \dots, Y_n be i.i.d. according to a symmetric Bernoulli distribution, that is, $\mathbb{P}(Y_t = 0) = \mathbb{P}(Y_t = 1) = 1/2$.

Then, as shown on the next slide,

$$\mathbb{E}[\bar{L}_n] = \frac{n}{2} \quad \text{and} \quad \mathbb{E}[\min \{L_{0,n}, L_{1,n}\}] = \frac{n}{2} - \square \sqrt{n}$$

and the **expectation** of the regret $\mathbb{E}[\bar{R}_n]$ is more than $\square \sqrt{n}$.

So is therefore \bar{R}_n for **some deterministic** sequence y_1, \dots, y_n .

Setting: $\mathcal{X} = \mathcal{Y} = \{0, 1\}$, loss function $\ell(x, y) = |x - y|$, outcomes (Y_t) i.i.d. symmetric Bernoulli, constant experts 0 and 1

\mathbb{E}_t denoting the conditional expectation w.r.t. Y_1, \dots, Y_{t-1} ,

$$\mathbb{E}_t[\ell(0, Y_t)] = \mathbb{E}_t[Y_t] = 1/2 = \mathbb{E}_t[1 - Y_t] = \mathbb{E}_t[\ell(1, Y_t)]$$

and thus, for **all strategies** σ ,

$$\mathbb{E}_t[p_{0,t} \ell(0, Y_t) + p_{1,t} \ell(1, Y_t)] = (p_{0,t} + p_{1,t})/2 = 1/2$$

which proves that $\mathbb{E}[\bar{L}_n] = n/2$. On the other hand,

$$L_{0,n} = \sum_{t=1}^n Y_t = \frac{n}{2} + \left(\sum_{t=1}^n Y_t - \frac{n}{2} \right), \quad L_{1,n} = \sum_{t=1}^n 1 - Y_t = \frac{n}{2} - \left(\sum_{t=1}^n Y_t - \frac{n}{2} \right)$$

so that

$$\mathbb{E}[\min\{L_{0,n}, L_{1,n}\}] = \frac{n}{2} - \mathbb{E} \left| \sum_{t=1}^n Y_t - \frac{n}{2} \right| = \frac{n}{2} - \Omega(\sqrt{n})$$

by **central limit** arguments.

The **refined central limit argument** of Cesa-Bianchi et al. '97 is as follows.

The constant experts are still identified with the possible predictions, $\mathcal{X} = \{1, \dots, N\}$, and the opponent player chooses the outcomes $Y_1, \dots, Y_n \in \mathcal{Y} = [0, 1]$ uniformly at random.

The loss function is $\ell(j, y) = \lfloor 2^j y \rfloor \bmod 2$, the j -th component of the dyadic expansion of y .

Then, the $\ell(j, Y_t)$ are i.i.d. (when t and j vary) according to a symmetric Bernoulli distribution.

Based on the fact that $\ell(j, Y_t)$ are i.i.d. (when t and j vary) according to a symmetric Bernoulli distribution,

we prove, similarly, as above, that $\mathbb{E}_t \left[\sum_{i=1}^N p_{i,t} \ell(i, Y_t) \right] = 1/2$ for all strategies σ .

Thus the expectation of the regret equals

$$\begin{aligned} \mathbb{E} \left[\sum_{t=1}^n \sum_{i=1}^N p_{i,t} \ell(i, Y_t) - \min_{j=1, \dots, N} \sum_{t=1}^n \ell(j, Y_t) \right] &= \mathbb{E} \left[\max_{j=1, \dots, N} \frac{n}{2} - \sum_{t=1}^n \ell(j, Y_t) \right] \\ &= \frac{1}{2} \mathbb{E} \left[\max_{j=1, \dots, N} \sum_{t=1}^n Z_{j,t} \right] \end{aligned}$$

where the $Z_{j,t}$ are i.i.d. symmetric Rademacher random variables.

We have from the previous slide that

$$\inf_{\sigma} \mathbb{E}[\bar{R}_n(\sigma, Y_1^n)] = \frac{1}{2} \mathbb{E} \left[\max_{j=1, \dots, N} \sum_{t=1}^n Z_{j,t} \right]$$

where the $Z_{j,t}$ are i.i.d. symmetric Rademacher random variables.

But the

$$G_{j,n} = \sum_{t=1}^n Z_{j,t} \stackrel{(d)}{\approx} \sqrt{n} N_j \stackrel{(d)}{=} \sqrt{n} \mathcal{N}(0, 1)$$

are independent random variables, and

$$\lim_{N \rightarrow \infty} \frac{1}{\sqrt{2 \ln N}} \mathbb{E} \left[\max_{j=1, \dots, N} N_j \right] = 1$$

which concludes to

$$\frac{1}{2} \mathbb{E} \left[\max_{j=1, \dots, N} \sum_{t=1}^n Z_{j,t} \right] = (1 + o(1)) \sqrt{\frac{n}{2} \ln N}$$

Not only we have the right orders of magnitude $\sqrt{n \ln N}$ but even the **optimal constant!**

An information-theoretic method, relying on a **Fano**-type lemma, is used in Cesa-Bianchi, Lugosi, and Stoltz '05 to provide an alternative proof with a **non-asymptotic** lower bound.

Lemma (Birgé's version of Fano's lemma)

For N probability distributions $\mathbb{Q}_1, \dots, \mathbb{Q}_N$ on the measurable space Ω and for a partition of the space into $\Omega_1, \dots, \Omega_N$,

$$\min_{j=1, \dots, N} \mathbb{Q}_j(\Omega_j) \leq \max \left\{ \frac{\bar{K}}{\ln(N-1)}, \frac{e}{e+1} \right\}$$

where

$$\bar{K} = \frac{1}{N-1} \sum_{i=2}^N \mathcal{K}(\mathbb{Q}_i, \mathbb{Q}_1)$$

is an average of Kullback-Leibler divergences.

With $\ell(j, y)$ still being the j -th element in the dyadic expansion of y , we may define N probability distributions

$$\mathbb{P}_1, \dots, \mathbb{P}_N$$

on $(Y_1, Y_2, \dots, Y_n, \dots) \in [0, 1]^{\mathbb{N}}$ such that, **under** \mathbb{P}_j ,

- the $\ell(k, Y_t)$ are all independent (when $k = 1, \dots, N$ and $t = 1, \dots, n$ vary)
- the $\ell(j, Y_t)$ has Bernoulli distribution with parameter $1/2 - \varepsilon$,
- the $\ell(k, Y_t)$, for $k \neq j$, have Bernoulli distributions with parameter $1/2$.

Put differently, under \mathbb{P}_j , we should play action j , which is ε -better than all other actions.

We then lower bound the minimax regret as before,

$$\begin{aligned} \inf_{\sigma} \sup_{\tau} \bar{R}_n(\sigma, \tau) &\geq \inf_{\sigma} \max_{k=1, \dots, N} \mathbb{E}_k \left[\sum_{t=1}^n \sum_{i=1}^N p_{i,t} \ell(i, Y_t) - \min_{j=1, \dots, N} \sum_{t=1}^n \ell(j, Y_t) \right] \\ &\geq \inf_{\sigma} \max_{j=1, \dots, N} \mathbb{E}_j \left[\sum_{t=1}^n \sum_{i=1}^N p_{i,t} \ell(i, Y_t) - \sum_{t=1}^n \ell(j, Y_t) \right] \end{aligned}$$

Now, for all t , one has $\mathbb{E}_j[\ell(j, Y_t)] = 1/2 - \varepsilon$ and

$$\mathbb{E}_j \left[\sum_{i=1}^N p_{i,t} \ell(i, Y_t) \right] = 1/2 - \varepsilon \mathbb{E}_j[p_{j,t}] = 1/2 - \varepsilon \mathbb{P}_j \otimes \mathbb{P}_A [I_t = j]$$

(where \mathbb{P}_A denotes the auxiliary randomization the decision-maker has access to). Thus,

$$\max_{j=1, \dots, N} \mathbb{E}_j \left[\sum_{t=1}^n \sum_{i=1}^N p_{i,t} \ell(i, Y_t) - \sum_{t=1}^n \ell(j, Y_t) \right] \geq n\varepsilon \max_{j=1, \dots, N} \left(1 - \frac{1}{n} \sum_{t=1}^n \mathbb{P}_j \otimes \mathbb{P}_A [I_t = j] \right)$$

and we are ready to apply Fano's lemma.

$$\min_{j=1,\dots,N} \frac{1}{n} \sum_{t=1}^n \mathbb{P}_j \otimes \mathbb{P}_A [I_t = j] \leq \max \left\{ \frac{\bar{K}}{\ln(N-1)}, \frac{e}{e+1} \right\}$$

where

$$\bar{K} = \frac{1}{N-1} \sum_{i=2}^N \mathcal{K}(\mathbb{P}_i \otimes \mathbb{P}_A, \mathbb{P}_1 \otimes \mathbb{P}_A) \leq \square n \varepsilon^2$$

for ε small enough.

Therefore,

$$\begin{aligned} \inf_{\sigma} \max_{j=1,\dots,N} \mathbb{E}_j [\bar{R}_n(\sigma, Y_1^n)] &\geq n\varepsilon \max_{j=1,\dots,N} \left(1 - \frac{1}{n} \sum_{t=1}^n \mathbb{P}_j \otimes \mathbb{P}_A [I_t = j] \right) \\ &\geq n\varepsilon \left(1 - \square \frac{n\varepsilon^2}{\ln(N-1)} \right) \end{aligned}$$

and the choice $\varepsilon = \square \sqrt{\ln(N-1)}/\sqrt{n}$ concludes the proof.

$$\min_{j=1,\dots,N} \frac{1}{n} \sum_{t=1}^n \mathbb{P}_j \otimes \mathbb{P}_A [I_t = j] \leq \max \left\{ \frac{\bar{K}}{\ln(N-1)}, \frac{e}{e+1} \right\}$$

where

$$\bar{K} = \frac{1}{N-1} \sum_{i=2}^N \mathcal{K}(\mathbb{P}_i, \mathbb{P}_1) \leq \square n \varepsilon^2$$

for ε small enough.

Therefore,

$$\begin{aligned} \inf_{\sigma} \max_{j=1,\dots,N} \mathbb{E}_j [\bar{R}_n(\sigma, Y_1^n)] &\geq n\varepsilon \max_{j=1,\dots,N} \left(1 - \frac{1}{n} \sum_{t=1}^n \mathbb{P}_j \otimes \mathbb{P}_A [I_t = j] \right) \\ &\geq n\varepsilon \left(1 - \square \frac{n\varepsilon^2}{\ln(N-1)} \right) \end{aligned}$$

and the choice $\varepsilon = \square \sqrt{\ln(N-1)}/\sqrt{n}$ concludes the proof.

$$\min_{j=1,\dots,N} \frac{1}{n} \sum_{t=1}^n \mathbb{P}_j \otimes \mathbb{P}_A [I_t = j] \leq \max \left\{ \frac{\bar{K}}{\ln(N-1)}, \frac{e}{e+1} \right\}$$

where

$$\bar{K} = \frac{1}{N-1} \sum_{i=2}^N \mathcal{K}(\mathbb{P}_i, \mathbb{P}_1) \leq \square n \varepsilon^2$$

for ε small enough.

Therefore,

$$\begin{aligned} \inf_{\sigma} \max_{j=1,\dots,N} \mathbb{E}_j [\bar{R}_n(\sigma, Y_1^n)] &\geq n\varepsilon \max_{j=1,\dots,N} \left(1 - \frac{1}{n} \sum_{t=1}^n \mathbb{P}_j \otimes \mathbb{P}_A [I_t = j] \right) \\ &\geq n\varepsilon \left(1 - \square \frac{n\varepsilon^2}{\ln(N-1)} \right) \end{aligned}$$

and the choice $\varepsilon = \square \sqrt{\ln(N-1)}/\sqrt{n}$ concludes the proof.

We have proved the following.

Theorem (Cesa-Bianchi, Lugosi, and Stoltz '05)

For $N \geq 2$, $n \geq 20 \ln(N - 1)$ and *any (randomized) strategy* σ , there exists a sequence y_1, \dots, y_n of outcomes such that for the constant experts and the loss function used above,

$$\begin{aligned} \sup_{y_1^n \in \mathcal{Y}^n} \max_{j=1, \dots, N} \sum_{t=1}^n \sum_{i=1}^N p_{i,t} \ell(i, y_t) - \sum_{t=1}^n \ell(j, y_t) \\ \geq \frac{\sqrt{e}}{(1+e)\sqrt{5(1+e)}} \sqrt{n \ln(N-1)} \end{aligned}$$

- 1 The model of individual sequences
 - Experts and regret
 - Randomized prediction
- 2 Optimal bounds on the regret in full-information
 - General $\sqrt{n \ln N}$ upper bound on the regret
 - Lower bounds on the regret
 - Refined upper bounds
- 3 Extensions: Aggregation of predictors, limited feedback
 - Aggregation of predictors, application to air-quality forecasting
 - Regret in multi-armed bandit problems
 - Minimizing the regret under partial monitoring

We have seen so far that for losses **known to lie in $[0, 1]$** , the exponentially weighted average strategy

- at round t only requires the knowledge of the $\ell(f_{i,s}, y_s)$, for $s \leq t - 1$ and $i = 1, \dots, N$,
- and ensures that the expected regret \bar{R}_n is always less than $\sqrt{n \ln N}$, which is its minimax order of magnitude.

One possible extension is to **specify** the general $\sqrt{n \ln N}$ rate by making it **data-dependent**.

For instance, Freund and Schapire '97 proved that

$$\bar{R}_n \leq \square \sqrt{L_n^* \ln N} + \square \ln N$$

where

$$L_n^* = \min_{j=1, \dots, N} L_{j,n}$$

is the cumulative loss of the best expert. (But **what about R_n ?**)

Cesa-Bianchi, Mansour, and Stoltz '07 deal both with unknown payoff ranges $[a, b]$ and refined bounds.

By a careful choice of η , we can be **adaptive** in all parameters (a , b , and n); and get, without any previous knowledge,

$$\bar{R}_n \leq \square \sqrt{V_n \ln N} + \square (b - a) \ln N$$

where V_n is a variance term,

$$V_n = \sum_{t=1}^n \sum_{i=1}^N p_{i,t} \left(\ell(f_{i,t}, y_t) - \sum_{k=1}^N p_{k,t} \ell(f_{k,t}, y_t) \right)^2$$

A **concentration** argument using **Bernstein's inequality** yields that with probability at least $1 - \delta$,

$$R_n \leq \bar{R}_n + \square \sqrt{V_n \ln \frac{1}{\delta}} + \dots \leq \square \sqrt{V_n \ln \frac{N}{\delta}} + \dots$$

To get a meaningful bound, we have to **solve** the inequation for the regret.

We finally get an improvement for small or large losses; for all strategies of the adversary and with probability at least $1 - \delta$,

$$R_n = \sum_{t=1, \dots, n} \ell(f_{I_t, t}, y_t) - \min_{j=1, \dots, N} \sum_{t=1, \dots, n} \ell(f_{j, t}, y_t) \\ \leq \square \sqrt{\frac{L_n^* (Mn - L_n^*)}{n}} \ln \frac{N}{\delta} + \dots$$

where

$$L_n^* = \min_{j=1, \dots, N} L_{j, n}$$

is the cumulative loss of the best expert and

$$M = \max_{j, t} |\ell(f_{j, t}, y_t)|$$

is the actual range of the losses.

- 1 The model of individual sequences
 - Experts and regret
 - Randomized prediction
- 2 Optimal bounds on the regret in full-information
 - General $\sqrt{n \ln N}$ upper bound on the regret
 - Lower bounds on the regret
 - Refined upper bounds
- 3 Extensions: Aggregation of predictors, limited feedback
 - Aggregation of predictors, application to air-quality forecasting
 - Regret in multi-armed bandit problems
 - Minimizing the regret under partial monitoring

Extension to convex losses: We now want to perform almost as well as the best convex combination of the experts.

We assume that decision set \mathcal{X} is **convex** (recall that the experts and the statistician pick their predictions in \mathcal{X}).

In addition, for all $y \in \mathcal{Y}$, the loss functions $\ell(\cdot, y) : \mathcal{X} \rightarrow \mathbb{R}$ are **convex**.

At each round, the statistician outputs a convex combination $\mathbf{p}_t = (p_{1,t}, \dots, p_{N,t})$ of the expert advices,

$$\hat{\mathbf{p}}_t = \sum_{i=1}^N p_{i,t} \mathbf{f}_{i,t}$$

The regret is now defined w.r.t. all possible convex combinations,

$$R_n = \sum_{t=1}^n \ell \left(\sum_{i=1}^N p_{i,t} \mathbf{f}_{i,t}, y_t \right) - \min_{\mathbf{q}} \sum_{t=1}^n \ell \left(\sum_{i=1}^N q_i \mathbf{f}_{i,t}, y_t \right)$$

$$\begin{aligned}
R_n &= \sum_{t=1}^n \ell \left(\sum_{i=1}^N p_{i,t} f_{i,t}, y_t \right) - \min_{\mathbf{q}} \sum_{t=1}^n \ell \left(\sum_{i=1}^N q_i f_{i,t}, y_t \right) \\
&\leq \max_{\mathbf{q}} \sum_{t=1}^n \nabla \ell \left(\sum_{i=1}^N p_{i,t} f_{i,t}, y_t \right) \cdot (\mathbf{p}_t - \mathbf{q}) \\
&\leq \sum_{t=1}^n \nabla \ell \left(\sum_{i=1}^N p_{i,t} f_{i,t}, y_t \right) \cdot \mathbf{p}_t - \min_{j=1, \dots, N} \sum_{t=1}^n \left(\nabla \ell \left(\sum_{i=1}^N p_{i,t} f_{i,t}, y_t \right) \right)_j
\end{aligned}$$

Thus, defining \mathbf{p}_t as an exponentially weighted average (over the sums of the **components of the sub-gradients**) leads to a regret

$$R_n \leq B \sqrt{n \ln N}$$

where B is a bound on the supremum norm of the sub-gradients.

This leads to a forecaster called EG. Note that it is **deterministic**.

Theoretical application: Oracle inequalities

We consider a new framework, with a risk function $Q : \Theta \times \mathcal{Z} \rightarrow \mathbb{R}$ convex in its first argument, where Θ is the simplex of \mathbb{R}^d .

Observations $Z_1^n = (Z_1, \dots, Z_n)$ (i.i.d. or even simply **stationary**) are available, and we aim at constructing a rule $\bar{\theta}_n = \bar{\theta}_n(Z_1^n)$ such that

$$\mathbb{E} [Q(\bar{\theta}_n, Z)] \leq \min_{\theta \in \Theta} \mathbb{E} [Q(\theta, Z)] + \Delta_n$$

where $\Delta_n = o(1)$ and expectations \mathbb{E} are taken w.r.t. Z_1^n and Z (independent of and with same distribution as the Z_1^n).

The results and (direct) techniques in **Juditsky, Nazin, Tsybakov, and Vayatis '05** are strongly related to the ones described on the next slide and usually lead to improved constant factors.

The trick is to do as if the observations Z_1^n were only available sequentially and apply, for instance, the EG forecaster, which yields a sequence $\hat{\theta}_1, \dots, \hat{\theta}_n$ of predictions (**almost surely**)

$$\sum_{t=1}^n Q(\hat{\theta}_t, Z_t) - \min_{\theta \in \Theta} \sum_{t=1}^n Q(\theta, Z_t) \leq \square \|\nabla Q\|_{\infty} \sqrt{n \ln N}$$

As a reminder, for $t = 1, \dots, n$ and $j = 1, \dots, d$,

$$\hat{\theta}_{j,t} = \frac{\exp\left(-\eta_t \sum_{s=1}^{t-1} \left(\nabla Q(\hat{\theta}_s, Z_s)\right)_j\right)}{\sum_{k=1}^d \exp\left(-\eta_t \sum_{s=1}^{t-1} \left(\nabla Q(\hat{\theta}_s, Z_s)\right)_k\right)}$$

The trick is to do as if the observations Z_1^n were only available sequentially and apply, for instance, the EG forecaster, which yields a sequence $\hat{\theta}_1, \dots, \hat{\theta}_n$ of predictions (**almost surely**)

$$\sum_{t=1}^n Q(\hat{\theta}_t, Z_t) - \min_{\theta \in \Theta} \sum_{t=1}^n Q(\theta, Z_t) \leq \square \|\nabla Q\|_{\infty} \sqrt{n \ln N}$$

We take **expectations** and use that Z_t is independent of $\hat{\theta}_t$,

$$\mathbb{E} \left[\sum_{t=1}^n Q(\hat{\theta}_t, Z) \right] - \min_{\theta \in \Theta} \mathbb{E} \left[\sum_{t=1}^n Q(\theta, Z) \right] \leq \square \|\nabla Q\|_{\infty} \sqrt{n \ln N}$$

We now define

$$\bar{\theta}_n = \frac{\hat{\theta}_1 + \dots + \hat{\theta}_n}{n}$$

and get by **Jensen's inequality** a **simple version** of the desired result,

$$\mathbb{E} [Q(\bar{\theta}_n, Z)] \leq \min_{\theta \in \Theta} \mathbb{E} [Q(\theta, Z)] + \square \|\nabla Q\|_{\infty} \sqrt{\frac{\ln N}{n}}$$

Empirical application (1): Prediction of electrical load and consumption

See Yannig Goude's PhD thesis (EDF & Paris-Sud University, '08).

We focus on a different example, but similar techniques are used and lead to comparable results.

Empirical application (2): Air-quality forecasting

We construct 48 base forecasters by choosing **for each a model** based on different physical and chemical sub-models of the propagation of ozone and related components, different numerical approximation schemes of the involved PDEs, and so on.

Instead of **selecting** a good model, we apply brute force techniques and consider several of them, which we **combine**.

A **network** \mathcal{S} of stations is available and each model $j = 1, \dots, 48$ outputs a prediction $f_{j,t}^s$ for the ozone peak at each station s and day t , which is then compared to the actual peak y_t^s .

The statistician chooses at each round a convex combination \mathbf{p}_t of the experts' predictions to be used at **all stations** (for the sake of interpretability).

The criterion is RMSE, which amounts to considering the **convex** loss functions

$$\ell(\mathbf{p}_t, (y_t^s)_{s \in \mathcal{S}_t}) = \sum_{s \in \mathcal{S}_t} \left(\sum_{i=1}^{48} p_{i,t} f_{i,t}^s - y_t^s \right)^2$$

where \mathcal{S}_t is the subset of active stations at day t .

The figures below show that **all** experts are useful.

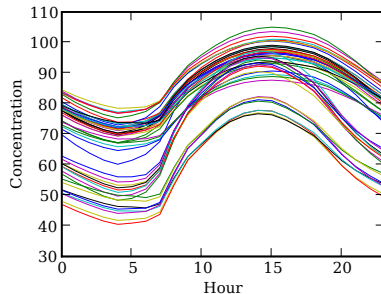
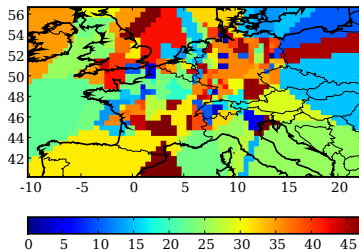


Figure: **Left:** Colored map of Europe according to the local best expert index. **Right:** Daily mean prediction profiles (averaged over prediction period and space, in $\mu\text{g}/\text{m}^3$).

The RMSE of a method with cumulative loss L_n up to round n is

$$\sqrt{\frac{L_n}{\sum_{t=1}^n |\mathcal{S}_t|}}$$

We indicate the RMSE (in $\mu\text{g}/\text{m}^3$) of

- the **mean** of the 48 predictors' predictions,
- the **best single** predictor among $j = 1, \dots, 48$
- the **best convex** combination \mathbf{p} (in the simplex) of the 48 predictors,
- the **best linear** prediction \mathbf{u} (among all vectors of \mathbb{R}^{48}) of the 48 predictors,
- the **prescient** forecaster that would know the y_t^s before predicting them and is allowed to pick any linear prediction.

Mean	Best single	Best convex	Best linear	Prescient
24.41	22.43	21.45	19.24	11.99

Mean	Best single	Best convex	Best linear	Prescient
24.41	22.43	21.45	19.24	11.99

We implemented about 20 different forecasters and focus on two efficient families, **EG**-type and ridge regression-type forecasters.

Ridge regression is a classical forecaster for square loss, using the penalized best linear combination so far. (Penalization in terms of ℓ^2 -norm.)

EG	Trunc. EG	Disc. EG	Ridge	Trunc. Ridge	Disc. Ridge
21.47	21.37	21.31	20.77	20.03	19.45

“Trunc.” stands for **truncated**; we only sum up past losses over a fixed window.

“Disc.” is for **discounting**; we give a higher weight to most recent losses.

The **best convex** combination is always beaten, and even the **best linear** is so (by the discounted version of ridge regression)!

The forecasters do **not** concentrate on one expert. The figures below show how much and how quick the weights can change over time.

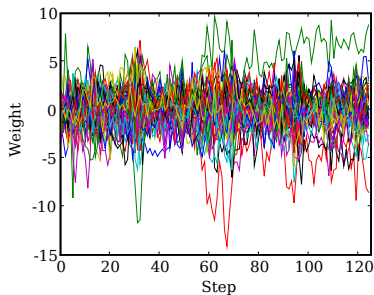
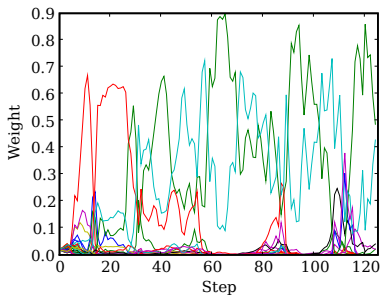


Figure: **Left:** Weights for EG over time. **Right:** Weights for Discounted Ridge Regression over time.

- 1 The model of individual sequences
 - Experts and regret
 - Randomized prediction
- 2 Optimal bounds on the regret in full-information
 - General $\sqrt{n \ln N}$ upper bound on the regret
 - Lower bounds on the regret
 - Refined upper bounds
- 3 Extensions: Aggregation of predictors, limited feedback
 - Aggregation of predictors, application to air-quality forecasting
 - Regret in multi-armed bandit problems
 - Minimizing the regret under partial monitoring

So far, we have been acting as **statisticians**. Now, we shall be **game-theoretists**.

For instance, **losses** are to be replaced by **payoffs**!

A base finite game against Nature is repeated.

- A decision-maker takes actions l_1, l_2, \dots from a **finite** set $\mathcal{X} = \{1, \dots, N\}$.
- The opponent player (Nature, a given adversary, the environment) selects the outcomes $y_1, y_2, \dots \in \mathcal{Y}$. (The **outcome space** \mathcal{Y} is arbitrary.)
- The payoff function is $r : \mathcal{X} \times \mathcal{Y} \rightarrow [0, 1]$.

That is, at each round $t = 1, 2, \dots$, the opponent player chooses a **vector**

$$(r(1, y_t), \dots, r(N, y_t)) = r(\cdot, y_t)$$

and the decision-maker chooses (simultaneously) a **component**.

To assess the performance of a strategy,

- we still fix the realized sequence of outcomes y_1, y_2, \dots
- and compare the sequence I_1, I_2, \dots of actions chosen by the decision-maker to constant sequences of pure actions j, j, \dots

That is, we compare $\hat{X}_n = \sum_{t=1}^n r(I_t, y_t)$ to the $X_{j,n} = \sum_{t=1}^n r(j, y_t)$.

Definition

The **(Hannan) regret** R_n is defined as the maximal difference of these cumulative payoffs,

$$R_n = \max_{j=1, \dots, N} R_{j,n} = \max_{j=1, \dots, N} X_{j,n} - \hat{X}_n = \max_{j=1, \dots, N} \sum_{t=1}^n r(j, y_t) - \sum_{t=1}^n r(I_t, y_t)$$

We want to control R_n a.s. This is already possible by the previous techniques.

The notion of regret studied here corresponds to external regret without activation functions.

(Various refined notions of regret have been introduced, see Lehrer '03 for **activation functions**, Foster and Vohra '97, Fudenberg and Levine '99, for **internal regret**).

Here, we fix this simple notion of regret and focus on the **information available** to the decision-maker.

We show that the amount of information determines the **minimax orders** of magnitude of the regret in n (number of game rounds) and N (number of actions).

Strategies presented at the beginning of the talk may be applied and show that the regret is less than $\sqrt{n \ln(N/\delta)}$ with probability at least $1 - \delta$.

This is proved via a control of the **expected regret** \bar{R}_n by $\sqrt{n \ln N}$.

More generally,

Theorem

The **minimax**^a orders of the regret are

- $R_n^* \sim \square \sqrt{n \ln N}$ in case of full monitoring,
- $R_n^* \sim \square \sqrt{nN \ln N}$ in multi-armed bandit problems,
- $R_n^* \sim \square n^{2/3} \psi(N)$ under partial monitoring (with deterministic signals).

^ai.e., $R_n^* = \inf_{\sigma} \sup_{\tau} E \bar{R}_n$ where \bar{R}_n is the (conditionally) expected regret

We shall present **simple strategies** that achieve these rates.

The **multi-armed bandit** problem is played as follows.

Parameters (known to both players): number N of actions, number M of outcomes, payoff function r (taking values in $[0, 1]$)

For each round $t = 1, 2, \dots$,

- the environment chooses the next outcome $y_t \in \{1, \dots, M\}$ without revealing it;
- the forecaster chooses a probability distribution \mathbf{p}_t and draws an action $I_t \in \{1, \dots, N\}$ according to this distribution;
- the forecaster receives a reward $r(I_t, y_t)$ and each action i gets a reward $r(i, y_t)$;
- **only his own reward $r(I_t, y_t)$** is revealed to the forecaster.

The regret is to be made small almost surely,

$$R_n = \max_{j=1, \dots, N} X_{j,n} - \bar{X}_n = o(n) \quad \text{a.s.}$$

The key idea is to **estimate** the unobserved payoffs and to form exponentially weighted averages on these estimates.

The estimates are

$$\tilde{r}_{j,t} = \frac{r(I_t, y_t)}{p_{I_t,t}} \mathbb{I}_{[I_t=j]}$$

We still denote by \mathbb{E}_t the **conditional expectation** at round t .

The estimates above are (conditionally) **unbiased**, since I_t is distributed as \mathbf{p}_t ,

$$\mathbb{E}_t [\tilde{r}_{j,t}] = \mathbb{E}_t \left[\frac{r(j, y_t)}{p_{j,t}} \mathbb{I}_{[I_t=j]} \right] = \frac{r(j, y_t)}{p_{j,t}} p_{j,t} = r(j, y_t)$$

We now perform exponentially weighted averages on these unbiased estimates. (Well, almost...)

We use the estimated payoffs $\tilde{r}_{j,t} = \frac{r(l_t, y_t)}{p_{l_t,t}} \mathbb{1}_{[l_t=j]}$

Exponentially weighted average predictor

p_1 is uniform and for $t \geq 2$,

$$p_{i,t} = (1 - \gamma) \frac{\exp\left(\eta \sum_{s=1}^{t-1} \tilde{r}(i, y_s)\right)}{\sum_{j=1}^N \exp\left(\eta \sum_{s=1}^{t-1} \tilde{r}(j, y_s)\right)} + \frac{\gamma}{N}$$

where $\eta, \gamma > 0$ are parameters to be tuned.

This forecaster has a $O(n^{2/3})$ regret.

The mixing is needed to bound from below the $p_{i,t}$, which in turn, bounds from above the conditional variances of the estimators $\tilde{r}_{j,t}$.

We use the estimated payoffs $\tilde{r}_{j,t} = \frac{r(l_t, y_t)}{p_{l_t,t}} \mathbb{1}_{[l_t=j]}$

Exponentially weighted average predictor

p_1 is uniform and for $t \geq 2$,

$$p_{i,t} = (1 - \gamma) \frac{\exp\left(\eta \sum_{s=1}^{t-1} \tilde{r}(i, y_s)\right)}{\sum_{j=1}^N \exp\left(\eta \sum_{s=1}^{t-1} \tilde{r}(j, y_s)\right)} + \frac{\gamma}{N}$$

where $\eta, \gamma > 0$ are parameters to be tuned.

This forecaster has a $O(n^{2/3})$ regret.

The mixing is needed to bound from below the $p_{i,t}$, which in turn, **bounds** from above the **conditional variances** of the estimators $\tilde{r}_{j,t}$.

A more efficient version use **shifted** versions of the estimated payoffs.

Exponentially weighted average predictor

\mathbf{p}_1 is uniform and for $t \geq 2$,

$$p_{i,t} = (1 - \gamma) \frac{\exp\left(\eta \sum_{s=1}^{t-1} \tilde{r}(i, y_s) + \frac{\beta}{p_{i,t}}\right)}{\sum_{j=1}^N \exp\left(\eta \sum_{s=1}^{t-1} \tilde{r}(j, y_s) + \frac{\beta}{p_{j,t}}\right)} + \frac{\gamma}{N}$$

where $\eta, \gamma, \beta > 0$ are parameters to be tuned.

For properly chosen parameters η, γ, β , and with probability at least $1 - \delta$, the regret of this forecaster is

$$R_n = \max_{j=1, \dots, N} X_{j,n} - \bar{X}_n \leq 6 \sqrt{nN \ln \frac{N}{\delta}} + \frac{\ln N}{2}$$

Auer, Cesa-Bianchi, Freund, and Schapire '02 showed a **lower bound** for the expected regret.

Theorem

For $\mathcal{Y} = [0, 1]$, there exists a payoff function $r : \mathbb{N} \times \mathcal{Y} \rightarrow \{0, 1\}$ such that for all $N \geq 2$ and $n \geq 1$, and all strategies suited to bandit settings,

$$\bar{R}_n = \max_{j=1, \dots, N} X_{j,n} - \bar{X}_n \geq \frac{1}{20} \min \left\{ \sqrt{nN}, n \right\}$$

The proof relies on **Pinsker's** inequality.

Open question: We suspect that the minimax order is $\sqrt{nN \ln N}$, but this has been **open** for more 10 years now!

- 1 The model of individual sequences
 - Experts and regret
 - Randomized prediction
- 2 Optimal bounds on the regret in full-information
 - General $\sqrt{n \ln N}$ upper bound on the regret
 - Lower bounds on the regret
 - Refined upper bounds
- 3 Extensions: Aggregation of predictors, limited feedback
 - Aggregation of predictors, application to air-quality forecasting
 - Regret in multi-armed bandit problems
 - Minimizing the regret under partial monitoring

In the general case, the received feedback may be even **less informative** than the received payoff.

The formulation of this general problem is as follows.

A base finite **game against Nature** is to be repeated.

It is parameterized by

- the strategy sets $\mathcal{X} = \{1, \dots, N\}$ and $\mathcal{Y} = \{1, \dots, M\}$ of the decision-maker and Nature,
- a payoff function $r : \mathcal{X} \times \mathcal{Y} \rightarrow [0, 1]$ for the decision-maker,
- a (finite) set \mathcal{S} of possible **signals**, $\Delta(\mathcal{S})$ denoting the set of probability distributions on \mathcal{S} ,
- a **feedback** function $H : \mathcal{X} \times \mathcal{Y} \rightarrow \Delta(\mathcal{S})$.

The repeated zero-sum game against Nature is played as follows.

Parameters (known to both players): number N of actions, number M of outcomes, payoff function r , random feedback function H

For each round $t = 1, 2, \dots$,

- Nature chooses the next outcome $y_t \in \{1, \dots, M\}$ without revealing it;
- the forecaster chooses a probability distribution \mathbf{p}_t and draws an action $I_t \in \{1, \dots, N\}$ according to this distribution;
- the forecaster receives reward $r(I_t, y_t)$ and each action i gets reward $r(i, y_t)$, where **none** of these values **is revealed** to the forecaster;
- **only a feedback** s_t drawn at random according to $H(I_t, y_t)$ is revealed to the forecaster.

Example: Dynamic pricing.

A vendor sells **T-shirts** on the Internet and chooses prices in

$$\mathcal{X} = \{9.90, 14.90, 19.90, 24.90, 29.90, \dots, 99.90\}$$

To the t -th customer, she/he offers the T-shirt at a **price** l_t .

Customers connect one by one to his web site.

Each of them has in mind a **maximal price** $y_t \in \mathcal{X} = \mathcal{Y}$ she/he is willing to pay – but does not tell it to the vendor.

When $y_t \geq l_t$, the product is bought and the vendor suffers a loss of earnings $r(l_t, y_t) = l_t - y_t$.

Otherwise, no deal takes place and the loss equals a fixed c (accounting for all her/his charges), $r(l_t, y_t) = -c$.

The feedback is thus $H(l_t, y_t) = \delta_{\mathbb{I}_{[y_t \geq l_t]}}$.

We want the **same average payoff** as the the **best constant price**.

Full monitoring: the outcomes are revealed, $H(i, j) = \delta_j$

Bandit games: the obtained payoffs are revealed, $H(i, j) = \delta_{r(i, j)}$

Noisy binary observations: $\mathcal{X} = \mathcal{Y} = \{0, 1\}$, the matrix representations of the feedback and payoff functions are

$$\begin{bmatrix} (1 - \varepsilon_0, \varepsilon_0) & (\varepsilon_1, 1 - \varepsilon_1) \\ (1 - \varepsilon_0, \varepsilon_0) & (\varepsilon_1, 1 - \varepsilon_1) \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Label-efficient prediction: 3 actions, 2 outcomes, the matrix representations of the feedback and payoff functions are

$$\begin{bmatrix} \delta_a & \delta_b \\ \delta_a & \delta_a \\ \delta_a & \delta_a \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

A **trade-off** needs to be done between getting **information** (action 1) and getting **rewards** (actions 2 and 3).

The feedback and payoff matrices are

$$\begin{bmatrix} (1, 0) & (1/2, 1/2) & (0, 1) \\ (1, 0) & (1/2, 1/2) & (0, 1) \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 5 & 0 & 0 \\ 0 & 3 & 1 \end{bmatrix}$$

If the **averaged observed distribution of signals** is close to $\Delta = (1/2, 1/2)$, the forecaster does not know whether it follows

- from a constant choice of outcome 2, $\mathbf{q}_2 = (0, 1, 0)$;
- outcomes 1 and 3 played equally often, $\mathbf{q}_{13} = (1/2, 0, 1/2)$;
- or any possible mixing of both, $\mathbf{q} = \alpha \mathbf{q}_2 + (1 - \alpha) \mathbf{q}_{13}$.

Action 2 is the optimal one against \mathbf{q}_2 (average payoff of 3 vs. 0).

One should play action 1 against \mathbf{q}_{13} (payoff 2.5 vs. 0.5).

In the **mixing case**, the best action **depends** on the mixing parameter α .

The feedback and payoff matrices are

$$\begin{bmatrix} (1, 0) & (1/2, 1/2) & (0, 1) \\ (1, 0) & (1/2, 1/2) & (0, 1) \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 5 & 0 & 0 \\ 0 & 3 & 1 \end{bmatrix}$$

If the **averaged observed distribution of signals** is close to $\Delta = (1/2, 1/2)$, the forecaster does not know whether it follows

- from a constant choice of outcome 2, $\mathbf{q}_2 = (0, 1, 0)$;
- outcomes 1 and 3 played equally often, $\mathbf{q}_{13} = (1/2, 0, 1/2)$;
- or any possible mixing of both, $\mathbf{q} = \alpha\mathbf{q}_2 + (1 - \alpha)\mathbf{q}_{13}$.

The rewards against $\mathbf{q} = \alpha\mathbf{q}_2 + (1 - \alpha)\mathbf{q}_{13}$ equal

Action 1:		5 α /2
Action 2:	3(1 - α) + α /2	= 3 - 5 α /2

The worst α is obtained by equating the two payoffs. We should think that **Nature** is indeed **playing this worst α** .

Summary:

Denote by \mathbf{p} our mixed distribution, $H(\mathbf{q})$ the distribution of signals induced by \mathbf{q} (independent of our actions here), and Δ the distribution of signals that results from the outcomes y_1, y_2, \dots, y_n .

Then the **best quantity** the forecaster can get **for sure given** $\Delta = (1/2, 1/2)$ equals

$$\begin{aligned} \min_{\alpha} \max \{ r(\delta_1, \mathbf{q}), r(\delta_2, \mathbf{q}) \} &= \min_{\mathbf{q}: H(\mathbf{q})=\Delta} \max_{\mathbf{p}} r(\mathbf{p}, \mathbf{q}) \\ &= \max_{\mathbf{p}} \min_{\mathbf{q}: H(\mathbf{q})=\Delta} r(\mathbf{p}, \mathbf{q}) \end{aligned}$$

by application of the minmax theorem.

The forecaster should play any element of

$$\operatorname{argmax}_{\mathbf{p}} \min_{\mathbf{q}: H(\mathbf{q})=\Delta} r(\mathbf{p}, \mathbf{q})$$

(these are also **equalizers**).

In full monitoring or in the **multi-armed bandit** model, the aim of the forecaster is to have a cumulative reward

$$\sum_{t=1}^n r(I_t, y_t)$$

close to (i.e., within $o(n)$ of) the cumulative reward of the best **fixed action**,

$$\max_{j=1, \dots, N} \sum_{t=1}^n r(j, y_t) = n \max_{\mathbf{p}} r(\mathbf{p}, \bar{\mathbf{q}}_n)$$

where $\bar{\mathbf{q}}_n$ is the empirical distribution of y_1, \dots, y_n .

Here, this quantity can in general **not be achieved**.

The previous slide and the next two ones explain why this is so and introduce the optimal goal.

We **extend linearly** r and H .

For probability distributions \mathbf{p} and \mathbf{q} on $\{1, \dots, N\}$ and $\{1, \dots, M\}$, we define

$$r(\mathbf{p}, \mathbf{q}) = \sum_{i,j} p_i q_j r(i, j)$$

$$\text{and } H(\cdot, \mathbf{q}) = \sum_{j=1, \dots, N} q_j \begin{bmatrix} H(1, j) \\ H(2, j) \\ \vdots \\ H(N, j) \end{bmatrix} \in (\Delta(\mathcal{S}))^N$$

Denote by \mathcal{F} the set of the Δ that may be written as $H(\cdot, \mathbf{q})$ for some \mathbf{q} .

The **target function** ρ is defined as

$$\rho(\mathbf{p}, \Delta) = \min_{\mathbf{q}: H(\cdot, \mathbf{q}) = \Delta} r(\mathbf{p}, \mathbf{q})$$

Recall that $\bar{\mathbf{q}}_n$ the empirical distribution of y_1, \dots, y_n .

The previous slides explained why, even with the knowledge of $\hat{\Delta}_n = H(\cdot, \bar{\mathbf{q}}_n)$ (i.e., **in hindsight**), we cannot hope to do better than $\max_{\mathbf{p}} \rho(\mathbf{p}, H(\cdot, \bar{\mathbf{q}}_n))$.

The following theorem shows that the latter is indeed achievable.

Theorem (Rustichini '99)

The decision-maker has a strategy σ such that for all strategies τ of the opponent player,

$$\limsup_{n \rightarrow \infty} \max_{\mathbf{p}} \rho(\mathbf{p}, H(\cdot, \bar{\mathbf{q}}_n)) - \frac{1}{n} \sum_{t=1, \dots, n} r(I_t, y_t) \leq 0 \quad \mathbb{P}_{\sigma, \tau}\text{-a.s.}$$

Definition

The **Rustichini regret** R_n under partial monitoring is defined as

$$R_n = n \max_{\mathbf{p}} \rho(\mathbf{p}, H(\cdot, \bar{\mathbf{q}}_n)) - \sum_{t=1, \dots, n} r(I_t, y_t)$$

The strategies σ such that $R_n = o(n) \mathbb{P}_{\sigma, \tau}$ -a.s. against all strategies τ are said **Rustichini consistent**.

Rustichini's proof relies on an approachability theorem for a continuum of types (see Mertens, Sorin, and Zamir '94). It is neither constructive nor indicates convergence rates.

We deal with both issues (and provide a simpler proof).

Two special cases had been dealt with so far,

- the case when the feedback depends **only on the outcome**, see Mannor and Shimkin '03;
- the **Hannan-consistent** case when

$$n \max_{\mathbf{p}} \rho(\mathbf{p}, H(\cdot, \bar{\mathbf{q}}_n)) = \max_{j=1, \dots, N} \sum_{t=1}^n r(j, y_t) ;$$

this corresponds, for instance, to **multi-armed bandit**, label-efficient prediction, noisy binary prediction (provided that $\varepsilon_1 + \varepsilon_0 \neq 1$);

see Piccolboni and Schindelhauer '01, Cesa-Bianchi, Lugosi, and Stoltz '06, where a **necessary and sufficient condition** for Hannan consistency is proposed and a $O(n^{2/3})$ rate for the regret of an explicit forecaster is exhibited and proved to be optimal.

Our new and complete solution relies on **several layers** of techniques, namely,

- the use of a lazy strategy that **groups rounds** together, to be able to **estimate** the original **feedback** distributions,
- which in turns allows **estimation** of the (pessimistic lower bounds on) unobserved **payoffs**;
- as well as some classical **exploration–exploitation** trade-off and **linearized** upper bounds on the quantities at hand by using sub-gradients of concave functions.

We now give some more details...

The forecaster only sees the signals $s_t \sim H(I_t, y_t)$ and aims at **reconstructing** $H(\cdot, \bar{\mathbf{q}}_n) = (H(\cdot, y_1) + \dots + H(\cdot, y_n))/n$.

The key is to find an unbiased estimate for each $H(i, y_t)$. For instance,

$$\hat{h}_{i,t} = \frac{\delta_{s_t}}{p_{i,t}} \mathbb{I}_{I_t=i}$$

is such that

$$\mathbb{E}_t [\hat{h}_{i,t}] = \frac{1}{p_{i,t}} \mathbb{E}_t [\delta_{s_t} \mathbb{I}_{I_t=i}] = \frac{1}{p_{i,t}} \mathbb{E}_t [H(I_t, y_t) \mathbb{I}_{I_t=i}] = \frac{1}{p_{i,t}} p_{i,t} H(i, y_t) = H(i, y_t)$$

Thus, for all m large enough and b , with Π projection onto \mathcal{F}

$$\hat{\Delta}^b = \Pi \left(\frac{1}{m} \sum_{t=bm+1}^{(b+1)m} [\hat{h}_{i,t}]_{i=1,\dots,N} \right) \quad \text{estimates well} \quad \Delta^b = \frac{1}{m} \sum_{t=bm+1}^{(b+1)m} H(\cdot, y_t)$$

The forecaster only sees the signals $s_t \sim H(I_t, y_t)$ and aims at **reconstructing** $H(\cdot, \bar{\mathbf{q}}_n) = (H(\cdot, y_1) + \dots + H(\cdot, y_n))/n$.

The key is to find an unbiased estimate for each $H(i, y_t)$. For instance,

$$\hat{h}_{i,t} = \frac{\delta_{s_t}}{p_{i,t}} \mathbb{I}_{I_t=i}$$

is such that

$$\mathbb{E}_t [\hat{h}_{i,t}] = \frac{1}{p_{i,t}} \mathbb{E}_t [\delta_{s_t} \mathbb{I}_{I_t=i}] = \frac{1}{p_{i,t}} \mathbb{E}_t [H(I_t, y_t) \mathbb{I}_{I_t=i}] = \frac{1}{p_{i,t}} p_{i,t} H(i, y_t) = H(i, y_t)$$

Thus, for all m large enough and b , with Π projection onto \mathcal{F}

$$\hat{\Delta}^b = \Pi \left(\frac{1}{m} \sum_{t=bm+1}^{(b+1)m} [\hat{h}_{i,t}]_{i=1, \dots, N} \right) \quad \text{estimates well} \quad \Delta^b = \frac{1}{m} \sum_{t=bm+1}^{(b+1)m} H(\cdot, y_t)$$

We continue by indicating some analytical properties of ρ , where we recall that for $\Delta \in \mathcal{F}$,

$$\rho(\mathbf{p}, \Delta) = \min_{\mathbf{q}: H(\cdot, \mathbf{q}) = \Delta} r(\mathbf{p}, \mathbf{q})$$

Thus, ρ is **concave** in its first argument and **convex** in the second argument.

In addition, it can be shown that ρ is **uniformly Lipschitz** in its second argument.

Warning!

The next slide is even more technical than the previous ones.

You might want to take a 2-minute long **nap**.

By martingale convergence, it suffices to deal with the **expected regret**; for all \mathbf{p} ,

$$n \rho(\mathbf{p}, H(\cdot, \bar{\mathbf{q}}_n)) - \sum_{t=1}^n r(\mathbf{p}_t, y_t)$$

By martingale convergence, it suffices to deal with the **expected regret**; for all \mathbf{p} ,

$$n \rho(\mathbf{p}, H(\cdot, \bar{\mathbf{q}}_n)) - \sum_{t=1}^n r(\mathbf{p}_t, y_t) \leq \sum_{b=0}^B \left(m \rho(\mathbf{p}, \Delta^b) - \sum_{t=bm+1}^{(b+1)m} r(\mathbf{p}_t, y_t) \right)$$

by linearity of r , **convexity** of ρ in its second argument, and using

$$\Delta^b = \frac{1}{m} \sum_{t=bm+1}^{(b+1)m} H(\cdot, y_t)$$

By martingale convergence, it suffices to deal with the **expected regret**; for all \mathbf{p} ,

$$\begin{aligned} n \rho(\mathbf{p}, H(\cdot, \bar{\mathbf{q}}_n)) - \sum_{t=1}^n r(\mathbf{p}_t, y_t) &\leq \sum_{b=0}^B \left(m \rho(\mathbf{p}, \Delta^b) - \sum_{t=bm+1}^{(b+1)m} r(\mathbf{p}_t, y_t) \right) \\ &= m \sum_{b=0}^B \left(\rho(\mathbf{p}, \Delta^b) - r\left(\mathbf{p}^b, \frac{1}{m} \sum_{t=bm+1}^{(b+1)m} \delta_{y_t}\right) \right) \end{aligned}$$

provided that for all $t = bm + 1, \dots, (b + 1)m$, there exists \mathbf{p}^b such that $\mathbf{p}_t = \mathbf{p}^b$ (i.e., the strategy is **lazy**)

By martingale convergence, it suffices to deal with the **expected regret**; for all \mathbf{p} ,

$$\begin{aligned} n \rho(\mathbf{p}, H(\cdot, \bar{\mathbf{q}}_n)) - \sum_{t=1}^n r(\mathbf{p}_t, y_t) &\leq \sum_{b=0}^B \left(m \rho(\mathbf{p}, \Delta^b) - \sum_{t=bm+1}^{(b+1)m} r(\mathbf{p}_t, y_t) \right) \\ &= m \sum_{b=0}^B \left(\rho(\mathbf{p}, \Delta^b) - r\left(\mathbf{p}^b, \frac{1}{m} \sum_{t=bm+1}^{(b+1)m} \delta_{y_t}\right) \right) \\ &\leq m \sum_{b=0}^B \left(\rho(\mathbf{p}, \Delta^b) - \rho(\mathbf{p}^b, \Delta^b) \right) \end{aligned}$$

using

$$\mathbf{q}^b = \frac{1}{m} \sum_{t=bm+1}^{(b+1)m} \delta_{y_t} \quad \text{implying} \quad \Delta^b = H(\cdot, \mathbf{q}^b)$$

and the **definition** of ρ as a **minimum**, $\rho(\mathbf{p}^b, \Delta^b) \leq r(\mathbf{p}^b, \mathbf{q}^b)$

By martingale convergence, it suffices to deal with the **expected regret**; for all \mathbf{p} ,

$$\begin{aligned} n \rho(\mathbf{p}, H(\cdot, \bar{\mathbf{q}}_n)) - \sum_{t=1}^n r(\mathbf{p}_t, y_t) &\leq \sum_{b=0}^B \left(m \rho(\mathbf{p}, \Delta^b) - \sum_{t=bm+1}^{(b+1)m} r(\mathbf{p}_t, y_t) \right) \\ &= m \sum_{b=0}^B \left(\rho(\mathbf{p}, \Delta^b) - r\left(\mathbf{p}^b, \frac{1}{m} \sum_{t=bm+1}^{(b+1)m} \delta_{y_t}\right) \right) \\ &\leq m \sum_{b=0}^B \left(\rho(\mathbf{p}, \Delta^b) - \rho(\mathbf{p}^b, \Delta^b) \right) \end{aligned}$$

Then, first, estimation is performed using **uniform Lipschitzness**,

$$m \sum_{b=0}^B \left(\rho(\mathbf{p}, \Delta^b) - \rho(\mathbf{p}^b, \Delta^b) \right) \leq m \sum_{b=0}^B \left(\rho(\mathbf{p}, \hat{\Delta}^b) - \rho(\mathbf{p}^b, \hat{\Delta}^b) \right) + \dots$$

By martingale convergence, it suffices to deal with the **expected regret**; for all \mathbf{p} ,

$$\begin{aligned} n \rho(\mathbf{p}, H(\cdot, \bar{\mathbf{q}}_n)) - \sum_{t=1}^n r(\mathbf{p}_t, y_t) &\leq \sum_{b=0}^B \left(m \rho(\mathbf{p}, \Delta^b) - \sum_{t=bm+1}^{(b+1)m} r(\mathbf{p}_t, y_t) \right) \\ &= m \sum_{b=0}^B \left(\rho(\mathbf{p}, \Delta^b) - r\left(\mathbf{p}^b, \frac{1}{m} \sum_{t=bm+1}^{(b+1)m} \delta_{y_t}\right) \right) \\ &\leq m \sum_{b=0}^B \left(\rho(\mathbf{p}, \Delta^b) - \rho(\mathbf{p}^b, \Delta^b) \right) \end{aligned}$$

Then, first, estimation is performed using **uniform Lipschitzness**,

$$\begin{aligned} m \sum_{b=0}^B \left(\rho(\mathbf{p}, \Delta^b) - \rho(\mathbf{p}^b, \Delta^b) \right) &\leq m \sum_{b=0}^B \left(\rho(\mathbf{p}, \hat{\Delta}^b) - \rho(\mathbf{p}^b, \hat{\Delta}^b) \right) + \dots \\ &\leq m \sum_{b=0}^B \nabla \rho(\mathbf{p}^b, \hat{\Delta}^b) \cdot (\mathbf{p} - \mathbf{p}^b) + \dots \end{aligned}$$

and second, a linear upper bound is considered using **concavity**.

Summary:

The expected regret is **linearly** upper bounded as

$$n \rho(\mathbf{p}, H(\cdot, \bar{\mathbf{q}}_n)) - \sum_{t=1}^n r(\mathbf{p}_t, y_t) \leq m \sum_{b=0}^B \nabla \rho(\mathbf{p}^b, \hat{\Delta}^b) \cdot (\mathbf{p} - \mathbf{p}^b) + \dots$$

and can thus be made small using exponentially weighted averages.

We also need(ed)

- the strategy to be **lazy**,
- a **mixing** with the uniform distribution as in the exploration–exploitation trade-off.

Both tricks are used to ensure an efficient estimation of the Δ^b .

As a consequence, the forecaster is as follows.

Parameters: Integer $m \geq 1$, real numbers $\eta, \gamma > 0$

Initialization: $\mathbf{w}^0 = (1, \dots, 1)$

For each round $t = 1, 2, \dots$,

- 1 if $bm + 1 \leq t < (b + 1)m$ for some integer b , choose the distribution $\mathbf{p}_t = \mathbf{p}^b = (1 - \gamma)\tilde{\mathbf{p}}^b + \gamma\mathbf{u}$, where $\tilde{\mathbf{p}}^b$ is defined component-wise as

$$\tilde{p}_k^b = \frac{w_k^b}{\sum_{j=1}^N w_j^b}$$

and \mathbf{u} denotes the uniform distribution, $\mathbf{u} = (1/N, \dots, 1/N)$;

- 2 draw an action I_t from $\{1, \dots, N\}$ according to it;
- 3 if $t = (b + 1)m$ for some integer b , perform the update

$$w_k^{b+1} = w_k^b e^{\eta(\nabla\rho(\mathbf{p}^b, \hat{\Delta}^b))_k} \quad \text{for each } k = 1, \dots, N,$$

where for all $\Delta \in \mathcal{F}$, $\nabla\rho(\cdot, \Delta)$ is a sub-gradient of $\rho(\cdot, \Delta)$ and $\hat{\Delta}^b$ is defined in the previous slides.

Two steps of the forecaster require some attention,

- **sub-gradients** of concave functions (actually, piecewise linear functions) need to be computed,
- **ℓ^2 -projections** onto the convex set \mathcal{F} are also required (when we compute the estimates $\hat{\Delta}^b$).

Both can be done efficiently, in time **polynomial** in N and $|\mathcal{S}|$.

Theorem (Lugosi, Mannor, and Stoltz '08)

The regret

$$R_n = n \max_{\mathbf{p}} \rho(\mathbf{p}, H(\cdot, \bar{\mathbf{q}}_n)) - \sum_{t=1, \dots, n} r(I_t, y_t)$$

is bounded by

- $O(n^{4/5})$ in the most general case,
- $O(n^{3/4})$ in the case of random signals depending on outcome only,
- $O(n^{2/3})$ in the case of deterministic signals,
- $O(n^{1/2})$ in the case of deterministic signals depending on outcome only.

Cesa-Bianchi, Lugosi, and Stoltz '06 show with label-efficient prediction that the $n^{2/3}$ rate is optimal (and that the $n^{1/2}$ rate is optimal as well has been known for a decade now).

Two main problems remain **open**,

- **minimax orders** of magnitude of the Rustichini regret in case of random signals (should we improve the upper or the lower bounds?),
- the extension to **continuous** probability distributions for the signals (of course, without quantization!).