adversarial bandit problems: the power of randomization

Gábor Lugosi

ICREA and Pompeu Fabra University, Barcelona

on-line prediction

A game between forecaster and environment.

At each round $\mathbf{t}$,

on-line prediction

A game between forecaster and environment.

At each round $\mathbf{t}$,

✳ the forecaster chooses an action $\mathbf{I_t \in \{1, \ldots, N\}}$;

on-line prediction

A game between forecaster and environment.

At each round $\mathbf{t}$,

- the forecaster chooses an action $\mathbf{I_t} \in \{\mathbf{1}, \ldots, \mathbf{N}\}$;

  (actions are often called experts)

on-line prediction

A game between forecaster and environment.

At each round $\mathbf{t}$,

- the forecaster chooses an action $\mathbf{I_t \in \{1, \ldots, N\}}$;

  (actions are often called experts)

- the environment chooses losses $\boldsymbol{\ell_t(1), \ldots, \ell_t(N) \in [0, 1]}$;

on-line prediction

A game between forecaster and environment.

At each round $\mathbf{t}$,

- the forecaster chooses an action $\mathbf{I_t} \in \{\mathbf{1}, \ldots, \mathbf{N}\}$;

  (actions are often called experts)

- the environment chooses losses $\boldsymbol{\ell}_\mathbf{t}(\mathbf{1}), \ldots, \boldsymbol{\ell}_\mathbf{t}(\mathbf{N}) \in [\mathbf{0}, \mathbf{1}]$;

- the forecaster suffers loss $\boldsymbol{\ell}_\mathbf{t}(\mathbf{I_t})$.

on-line prediction

A game between forecaster and environment.

At each round $\mathbf{t}$,

✳ the forecaster chooses an action $\mathbf{I_t \in \{1, \ldots, N\}}$;

(actions are often called experts)

✳ the environment chooses losses $\boldsymbol{\ell_t(1), \ldots, \ell_t(N) \in [0, 1]}$;

✳ the forecaster suffers loss $\boldsymbol{\ell_t(I_t)}$.

The goal is to minimize the average regret

$$\mathbf{R_n = \frac{1}{n} \left(\sum_{t=1}^{n} \ell_t(I_t) - \min_{i \le N} \sum_{t=1}^{n} \ell_t(i) \right)} \ .$$

outcome sequence

Often $\ell_t(i) = \ell(i, y_t)$

where $y_1, \ldots, y_n \in \mathcal{Y}$ is the sequence of outcomes to be predicted.

and $\ell : \{1, \ldots, N\} \times \mathcal{Y} \to [0, 1]$ is a loss function.

simplest example

Is it possible to make regret $\to \mathbf{0}$ for all loss assignments?

simplest example

Is it possible to make regret $\to \mathbf{0}$ for all loss assignments?
Let $\mathbf{N = 2}$ and define, for all $\mathbf{t = 1, \ldots, n}$,

$$\boldsymbol{\ell_t(1)} = \begin{cases} \mathbf{0} & \text{if } \mathbf{I_t = 2} \\ \mathbf{1} & \text{if } \mathbf{I_t = 1} \end{cases}$$

and $\boldsymbol{\ell_t(2) = 1 - \ell_t(1)}$.

simplest example

Is it possible to make regret $\to 0$ for all loss assignments?
Let $N = 2$ and define, for all $t = 1, \ldots, n$,

$$\ell_t(1) = \begin{cases} 0 & \text{if } I_t = 2 \\ 1 & \text{if } I_t = 1 \end{cases}$$

and $\ell_t(2) = 1 - \ell_t(1)$.

Then

$$\sum_{t=1}^{n} \ell_t(I_t) = n \quad \text{and} \quad \min_{i=1,2} \sum_{t=1}^{n} \ell_t(i) \leq \frac{n}{2}$$

simplest example

Is it possible to make regret $\to 0$ for all loss assignments?

Let $N = 2$ and define, for all $t = 1, \ldots, n$,

$$\ell_t(1) = \begin{cases} 0 & \text{if } I_t = 2 \\ 1 & \text{if } I_t = 1 \end{cases}$$

and $\ell_t(2) = 1 - \ell_t(1)$.

Then

$$\sum_{t=1}^{n} \ell_t(I_t) = n \quad \text{and} \quad \min_{i=1,2} \sum_{t=1}^{n} \ell_t(i) \le \frac{n}{2}$$

so

$$R_n \ge \frac{1}{2} \, .$$

randomized prediction

Key to solution: randomization.

At time $\mathbf{t}$, the forecaster chooses a probability distribution
$\mathbf{p_{t-1} = (p_{1,t-1}, \ldots, p_{N,t-1})}$

and chooses action $\mathbf{i}$ with probability $\mathbf{p_{i,t-1}}$.

Simplest model: all losses $\ell_{\mathbf{s}}(\mathbf{i})$, $\mathbf{i = 1, \ldots, N}$, $\mathbf{s < t}$, are observed: full information.

randomized prediction

This and related models have been studied in

- game theory: playing repeated games;
- information theory: gambling and data compression;
- statistics: sequential decisions;
- statistical learning theory: on-line learning;

Hannan and Blackwell

Hannan (1957) and Blackwell (1956) showed that the forecaster has a strategy such that

$$\frac{1}{n}\left(\sum_{t=1}^{n}\ell_t(I_t) - \min_{i\le N}\sum_{t=1}^{n}\ell_t(i)\right) \to 0$$

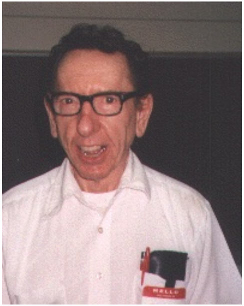
almost surely for all strategies of the environment.

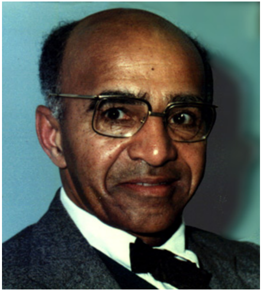
Hannan and Blackwell

Hannan (1957) and Blackwell (1956) showed that the forecaster has a strategy such that

$$\frac{1}{n}\left(\sum_{t=1}^{n}\ell_t(I_t) - \min_{i\le N}\sum_{t=1}^{n}\ell_t(i)\right) \to 0$$

almost surely for all strategies of the environment.





basic ideas

expected loss of the forecaster:

$$\ell_t(p_{t-1}) = \sum_{i=1}^{N} p_{i,t-1}\ell_t(i) = \mathbb{E}_t \ell_t(I_t)$$

basic ideas

expected loss of the forecaster:

$$\ell_t(\mathbf{p}_{t-1}) = \sum_{i=1}^{N} p_{i,t-1}\ell_t(i) = \mathbb{E}_t \ell_t(I_t)$$

By martingale convergence,

$$\frac{1}{n}\left(\sum_{t=1}^{n}\ell_t(I_t) - \sum_{t=1}^{n}\ell_t(\mathbf{p}_{t-1})\right) = O_P(n^{-1/2})$$

basic ideas

expected loss of the forecaster:

$$\ell_t(\mathbf{p}_{t-1}) = \sum_{i=1}^{N} p_{i,t-1}\ell_t(i) = \mathbb{E}_t \ell_t(I_t)$$

By martingale convergence,

$$\frac{1}{n}\left(\sum_{t=1}^{n} \ell_t(I_t) - \sum_{t=1}^{n} \ell_t(\mathbf{p}_{t-1})\right) = O_P(n^{-1/2})$$

so it suffices to study

$$\frac{1}{n}\left(\sum_{t=1}^{n} \ell_t(\mathbf{p}_{t-1}) - \min_{i \le N} \sum_{t=1}^{n} \ell_t(i)\right)$$

weighted average prediction

Idea: assign a higher probability to better-performing actions.
Vovk (1990), Littlestone and Warmuth (1989).

weighted average prediction

Idea: assign a higher probability to better-performing actions.
Vovk (1990), Littlestone and Warmuth (1989).

A popular choice is

$$p_{i,t-1} = \frac{\exp\left(-\eta \sum_{s=1}^{t-1} \ell_s(i)\right)}{\sum_{k=1}^{N} \exp\left(-\eta \sum_{s=1}^{t-1} \ell_s(k)\right)} \quad i = 1, \ldots, N\ .$$

where $\eta > 0$.

weighted average prediction

Idea: assign a higher probability to better-performing actions.
Vovk (1990), Littlestone and Warmuth (1989).

A popular choice is

$$\mathbf{p}_{i,t-1} = \frac{\exp\left(-\eta \sum_{s=1}^{t-1} \ell_s(i)\right)}{\sum_{k=1}^{N} \exp\left(-\eta \sum_{s=1}^{t-1} \ell_s(k)\right)} \quad i = 1, \ldots, N \; .$$

where $\eta > 0$. Then

$$\frac{1}{n}\left(\sum_{t=1}^{n} \ell_t(\mathbf{p}_{t-1}) - \min_{i \le N} \sum_{t=1}^{n} \ell_t(i)\right) = \sqrt{\frac{\ln N}{2n}}$$

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

proof

Let $\mathbf{L_{i,t} = \sum_{s=1}^{t} \ell_s(i)}$ and

$$\mathbf{W_t = \sum_{i=1}^{N} w_{i,t} = \sum_{i=1}^{N} e^{-\eta L_{i,t}}}$$

for $\mathbf{t \geq 1}$, and $\mathbf{W_0 = N}$.

proof

Let $L_{i,t} = \sum_{s=1}^{t} \ell_s(i)$ and

$$W_t = \sum_{i=1}^{N} w_{i,t} = \sum_{i=1}^{N} e^{-\eta L_{i,t}}$$

for $t \geq 1$, and $W_0 = N$. First observe that

$$\begin{aligned}
\ln \frac{W_n}{W_0} &= \ln\left(\sum_{i=1}^{N} e^{-\eta L_{i,n}}\right) - \ln N \\
&\geq \ln\left(\max_{i=1,\ldots,N} e^{-\eta L_{i,n}}\right) - \ln N \\
&= -\eta \min_{i=1,\ldots,N} L_{i,n} - \ln N \,.
\end{aligned}$$

proof

On the other hand, for each $\mathbf{t = 1, \ldots, n}$

$$\begin{aligned}
\ln \frac{W_t}{W_{t-1}} &= \ln \frac{\sum_{i=1}^{N} w_{i,t-1} e^{-\eta \ell_t(i)}}{\sum_{j=1}^{N} w_{j,t-1}} \\
&\le -\eta \frac{\sum_{i=1}^{N} w_{i,t-1} \ell_t(i)}{\sum_{j=1}^{N} w_{j,t-1}} + \frac{\eta^2}{8} \\
&= -\eta \ell_t(p_{t-1}) + \frac{\eta^2}{8}
\end{aligned}$$

by Hoeffding's inequality.

proof

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

$$\begin{aligned}
\ln \frac{W_t}{W_{t-1}} &= \ln \frac{\sum_{i=1}^{N} w_{i,t-1} e^{-\eta \ell_t(i)}}{\sum_{j=1}^{N} w_{j,t-1}} \\
&\le -\eta \frac{\sum_{i=1}^{N} w_{i,t-1} \ell_t(i)}{\sum_{j=1}^{N} w_{j,t-1}} + \frac{\eta^2}{8} \\
&= -\eta \ell_t(p_{t-1}) + \frac{\eta^2}{8}
\end{aligned}$$

by Hoeffding's inequality.

Hoeffding (1963): if $X \in [0, 1]$,

$$\ln \mathbb{E} e^{-\eta X} \le -\eta \mathbb{E} X + \frac{\eta^2}{8}$$

proof

for each $t = 1, \dots, n$

$$\ln \frac{W_t}{W_{t-1}} \le -\eta \ell_t(p_{t-1}) + \frac{\eta^2}{8}$$

Summing over $t = 1, \dots, n$,

$$\ln \frac{W_n}{W_0} \le -\eta \sum_{t=1}^{n} \ell_t(p_{t-1}) + \frac{\eta^2}{8} n \, .$$

Combining these, we get

$$\sum_{t=1}^{n} \ell_t(p_{t-1}) \le \min_{i=1,\dots,N} L_{i,n} + \frac{\ln N}{\eta} + \frac{\eta}{8} n$$

lower bound

The upper bound is optimal: for all predictors,

$$\sup_{n,N,\ell_t(i)} \frac{\sum_{t=1}^{n} \ell_t(I_t) - \min_{i \le N} \sum_{t=1}^{n} \ell_t(i)}{\sqrt{(n/2)\ln N}} \ge 1 \ .$$

lower bound

The upper bound is optimal: for all predictors,

$$\sup_{n,N,\ell_t(i)} \frac{\sum_{t=1}^n \ell_t(I_t) - \min_{i\le N} \sum_{t=1}^n \ell_t(i)}{\sqrt{(n/2)\ln N}} \ge 1 \ .$$

Idea: choose $\ell_t(i)$ to be i.i.d. symmetric Bernoulli coin flips.

lower bound

The upper bound is optimal: for all predictors,

$$\sup_{n,N,\ell_t(i)} \frac{\sum_{t=1}^{n} \ell_t(I_t) - \min_{i \le N} \sum_{t=1}^{n} \ell_t(i)}{\sqrt{(n/2)\ln N}} \ge 1 \;.$$

Idea: choose $\ell_t(i)$ to be i.i.d. symmetric Bernoulli coin flips.

$$\begin{aligned}
& \sup_{\ell_t(i)} \left(\sum_{t=1}^{n} \ell_t(I_t) - \min_{i \le N} \sum_{t=1}^{n} \ell_t(i) \right) \\
& \ge \quad \mathbb{E}\left[\sum_{t=1}^{n} \ell_t(I_t) - \min_{i \le N} \sum_{t=1}^{n} \ell_t(i) \right] \\
& = \quad \frac{n}{2} - \min_{i \le N} B_i
\end{aligned}$$

Where $B_1, \ldots, B_N$ are independent Binomial $(n, 1/2)$.
Use the central limit theorem.

label efficient prediction

The forecaster does not see the outcomes $\ell_t(i)$ unless he asks for them, but can do it only $m \ll n$ times.

label efficient prediction

The forecaster does not see the outcomes $\ell_t(i)$ unless he asks for them, but can do it only $m \ll n$ times.
For each round $t = 1, \ldots, n$,

* the environment chooses the losses $\ell_t(i)$ without revealing them;

label efficient prediction

The forecaster does not see the outcomes $\ell_t(i)$ unless he asks for them, but can do it only $m \ll n$ times.
For each round $t = 1, \ldots, n$,

- the environment chooses the losses $\ell_t(i)$ without revealing them;
- the forecaster chooses p_{t-1} and draws an action $I_t \in \{1, \ldots, N\}$ according to this distribution;

label efficient prediction

The forecaster does not see the outcomes $\ell_t(i)$ unless he asks for them, but can do it only $m \ll n$ times.

For each round $t = 1, \ldots, n$,

- the environment chooses the losses $\ell_t(i)$ without revealing them;
- the forecaster chooses p_{t-1} and draws an action $I_t \in \{1, \ldots, N\}$ according to this distribution;
- the forecaster incurs loss $\ell_t(I_t)$ and each action i incurs loss $\ell_t(i)$. not of revealed to the forecaster!;

label efficient prediction

The forecaster does not see the outcomes $\ell_t(i)$ unless he asks for them, but can do it only $m \ll n$ times.
For each round $t = 1, \ldots, n$,

- the environment chooses the losses $\ell_t(i)$ without revealing them;
- the forecaster chooses p_{t-1} and draws an action $I_t \in \{1, \ldots, N\}$ according to this distribution;
- the forecaster incurs loss $\ell_t(I_t)$ and each action i incurs loss $\ell_t(i)$. not of revealed to the forecaster!;
- the forecaster decides whether he asks for the values of $\ell_t(i)$ if the total number of revealed outcomes up to time $t - 1$ is less than m.

a label efficient forecaster

Idea: ask for values randomly (with probability $\approx \mathbf{m/n}$) and use the weighted average forecaster with the estimated losses.

a label efficient forecaster

Idea: ask for values randomly (with probability $\approx \mathbf{m/n}$) and use the weighted average forecaster with the estimated losses.
Let $\mathbf{Z_t}$ be i.i.d. Bernoulli $\boldsymbol{\epsilon}$ $(\approx \mathbf{m/n})$.
The forecaster asks for $\boldsymbol{\ell_t(i)}$ iff $\mathbf{Z_t = 1}$.

a label efficient forecaster

Idea: ask for values randomly (with probability $\approx \mathbf{m/n}$) and use the weighted average forecaster with the estimated losses.
Let $\mathbf{Z_t}$ be i.i.d. Bernoulli $\epsilon\ (\approx \mathbf{m/n})$.
The forecaster asks for $\ell_{\mathbf{t}}(\mathbf{i})$ iff $\mathbf{Z_t = 1}$. Let

$$\widetilde{\ell}_{\mathbf{t}}(\mathbf{i}) \stackrel{\text{def}}{=} \begin{cases} \ell_{\mathbf{t}}(\mathbf{i})/\varepsilon & \text{if } \mathbf{Z_t = 1}, \\ \mathbf{0} & \text{otherwise.} \end{cases}$$

An unbiased estimate!

a label efficient forecaster

Idea: ask for values randomly (with probability $\approx \mathbf{m/n}$) and use the weighted average forecaster with the estimated losses.
Let $\mathbf{Z_t}$ be i.i.d. Bernoulli ϵ $(\approx \mathbf{m/n})$.
The forecaster asks for $\ell_t(\mathbf{i})$ iff $\mathbf{Z_t = 1}$. Let

$$\widetilde{\ell}_t(\mathbf{i}) \stackrel{\text{def}}{=} \begin{cases} \ell_t(\mathbf{i})/\varepsilon & \text{if } \mathbf{Z_t = 1}, \\ \mathbf{0} & \text{otherwise.} \end{cases}$$

An unbiased estimate!
For each round $\mathbf{t = 1, 2, \ldots, n}$ draw an action from $\{\mathbf{1, \ldots, N}\}$ according to the distribution

$$\mathbf{p_{i,t-1}} = \frac{\exp\left(-\eta \sum_{s=1}^{t-1} \widetilde{\ell}_s(\mathbf{i})\right)}{\sum_{k=1}^{N} \exp\left(-\eta \sum_{s=1}^{t-1} \widetilde{\ell}_s(\mathbf{k})\right)} \quad \mathbf{i = 1, \ldots, N} \; .$$

bound for label efficient prediction

With probability at least $1 - \delta$,

$$\frac{1}{n}\left(\sum_{t=1}^{n} \ell_t(I_t) - \min_{i \le N} \sum_{t=1}^{n} \ell_t(i)\right) \le 9\sqrt{\frac{\ln N + \ln(4/\delta)}{m}} \ .$$

(Cesa-Bianchi, Lugosi, Stoltz, 2005)

bound for label efficient prediction

With probability at least $1-\delta$,

$$\frac{1}{n}\left(\sum_{t=1}^{n}\ell_t(I_t)-\min_{i\le N}\sum_{t=1}^{n}\ell_t(i)\right)\le 9\sqrt{\frac{\ln N+\ln(4/\delta)}{m}}\ .$$

(Cesa-Bianchi, Lugosi, Stoltz, 2005)

Sketch of proof: First bound

$$\sum_{t=1}^{n}\widetilde{\ell}_t(p_{t-1})-\min_{i\le N}\sum_{t=1}^{n}\widetilde{\ell}_t(i)$$

as before.

bound for label efficient prediction

With probability at least $\mathbf{1-\delta}$,

$$\frac{1}{n}\left(\sum_{t=1}^{n}\ell_t(I_t) - \min_{i\le N}\sum_{t=1}^{n}\ell_t(i)\right) \le 9\sqrt{\frac{\ln N + \ln(4/\delta)}{m}} .$$

(Cesa-Bianchi, Lugosi, Stoltz, 2005)

Sketch of proof: First bound

$$\sum_{t=1}^{n}\widetilde{\ell}_t(p_{t-1}) - \min_{i\le N}\sum_{t=1}^{n}\widetilde{\ell}_t(i)$$

as before. Then use Bernstein-type martingale inequalities to handle

$$\sum_{t=1}^{n}\ell_t(I_t) - \sum_{t=1}^{n}\widetilde{\ell}_t(p_{t-1})$$

and

$$\min_{i\le N}\sum_{t=1}^{n}\widetilde{\ell}_t(i) - \min_{i\le N}\sum_{t=1}^{n}\ell_t(i)$$

lower bound

For any forecaster asking for at most $\mathbf{m}$ values,

$$\sup_{\ell_t(i)\in\{0,1\}} \mathbb{E}\frac{1}{n}\left(\sum_{t=1}^{n}\ell_t(I_t) - \min_{i\leq N}\sum_{t=1}^{n}\ell_t(i)\right) \geq c\sqrt{\frac{\ln N}{m}}\,.$$

lower bound

For any forecaster asking for at most $\mathbf{m}$ values,

$$\sup_{\ell_t(i)\in\{0,1\}} \mathbb{E}\frac{1}{n}\left(\sum_{t=1}^{n}\ell_t(I_t) - \min_{i\le N}\sum_{t=1}^{n}\ell_t(i)\right) \ge c\sqrt{\frac{\ln N}{m}}\ .$$

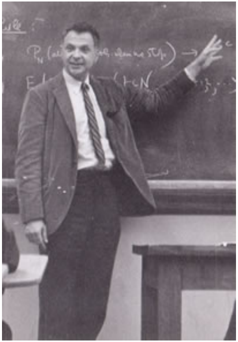
Idea (for $N = 2$): choose the losses randomly (i.i.d.) such that they are either Bernoulli $1/2-\epsilon$ or Bernoulli $1/2+\epsilon$.

multi-armed bandits

The forecaster only observes $\ell_t(I_t)$ but not $\ell_t(i)$ for $i \neq I_t$.

multi-armed bandits

The forecaster only observes $\ell_t(I_t)$ but not $\ell_t(i)$ for $i \neq I_t$.

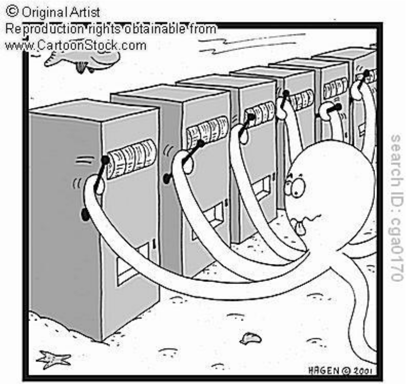


Herbert Robbins (1952).

multi-armed bandits

a one-armed bandit

multi-armed bandits



Compulsive gambling

a multi-armed bandit

multi-armed bandits

Trick: estimate $\ell_t(i)$ by

$$\widetilde{\ell}_t(i) = \frac{\ell_t(I_t)\mathbb{1}_{\{I_t=i\}}}{p_{I_t,t-1}}$$

multi-armed bandits

Trick: estimate $\ell_t(i)$ by

$$\widetilde{\ell}_t(i) = \frac{\ell_t(I_t)\mathbb{1}_{\{I_t=i\}}}{p_{I_t,t-1}}$$

This is an unbiased estimate:

$$\mathbb{E}_t\widetilde{\ell}_t(i) = \sum_{j=1}^{N} p_{j,t-1}\frac{\ell_t(j)\mathbb{1}_{\{j=i\}}}{p_{j,t-1}} = \ell_t(i)$$

multi-armed bandits

Trick: estimate $\ell_t(i)$ by

$$\widetilde{\ell}_t(i) = \frac{\ell_t(I_t)\mathbb{1}_{\{I_t=i\}}}{p_{I_t,t-1}}$$

This is an unbiased estimate:

$$\mathbb{E}_t\widetilde{\ell}_t(i) = \sum_{j=1}^{N} p_{j,t-1}\frac{\ell_t(j)\mathbb{1}_{\{j=i\}}}{p_{j,t-1}} = \ell_t(i)$$

Use the estimated losses to define exponential weights and mix with uniform (Auer, Cesa-Bianchi, Freund, and Schapire, 2002):

$$p_{i,t-1} = (1-\gamma)\underbrace{\frac{\exp\left(-\eta\sum_{s=1}^{t-1}\widetilde{\ell}_s(i)\right)}{\sum_{k=1}^{N}\exp\left(-\eta\sum_{s=1}^{t-1}\widetilde{\ell}_s(k)\right)}}_{\text{exploitation}} + \underbrace{\frac{\gamma}{N}}_{\text{exploration}}$$

multi-armed bandits

$$\mathbb{E}\frac{1}{n}\left(\sum_{t=1}^{n}\ell_t(p_{t-1}) - \min_{i\leq N}\sum_{t=1}^{n}\ell_t(i)\right) = O\left(\sqrt{\frac{N\ln N}{n}}\right) \; ,$$

multi-armed bandits

$$\mathbb{E}\frac{1}{n}\left(\sum_{t=1}^{n}\ell_t(p_{t-1}) - \min_{i\leq N}\sum_{t=1}^{n}\ell_t(i)\right) = O\left(\sqrt{\frac{N\ln N}{n}}\right) ,$$




multi-armed bandits

Lower bound:

$$\sup_{\ell_t(i)} \mathbb{E} \frac{1}{n} \left(\sum_{t=1}^{n} \ell_t(p_{t-1}) - \min_{i \le N} \sum_{t=1}^{n} \ell_t(i) \right) \ge C \sqrt{\frac{N}{n}} \ ,$$

Dependence on N is not logarithmic anymore!

multi-armed bandits

Lower bound:

$$\sup_{\ell_t(i)} \mathbb{E} \frac{1}{n} \left(\sum_{t=1}^{n} \ell_t(p_{t-1}) - \min_{i \le N} \sum_{t=1}^{n} \ell_t(i) \right) \ge C \sqrt{\frac{N}{n}} ,$$

Dependence on N is not logarithmic anymore!

Audibert and Bubeck (2009) constructed a forecaster with

$$\max_{i \le N} \mathbb{E} \frac{1}{n} \left(\sum_{t=1}^{n} \ell_t(p_{t-1}) - \sum_{t=1}^{n} \ell_t(i) \right) = O \left(\sqrt{\frac{N}{n}} \right) ,$$

follow the perturbed leader

$$\mathbf{I_t = \arg\min_{i=1,\ldots,N} \sum_{s=1}^{t-1} \ell_s(i) + Z_{i,t}}$$

where the $\mathbf{Z_{i,t}}$ are random noise variables.

By carefully defining the distribution of $\mathbf{Z_{i,t}}$ one can get similar regret bounds for the full information case, Hannan (1957); Kalai and Vempala (2003).

combinatorial experts

Often the class of experts is very large but has some combinatorial structure. Can the structure be exploited?

combinatorial experts

Often the class of experts is very large but has some combinatorial structure. Can the structure be exploited?

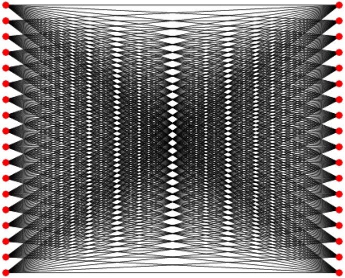
examples:

path planning. At each time instance, the forecaster chooses a path in a graph between two fixed nodes. Each edge has an associated loss. Loss of a path is the sum of the losses over the edges in the path.

N is huge!!!

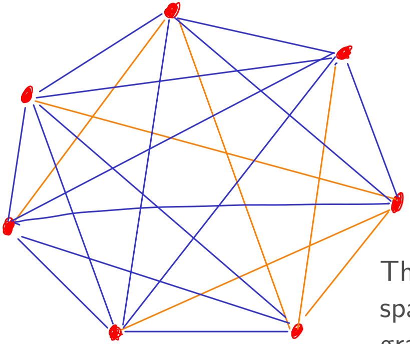
assignments: learning permutations

Given a complete bipartite graph $\mathbf{K_{m,m}}$, the forecaster chooses a perfect matching. The loss is the sum of the losses over the edges.



Helmbold and Warmuth (2007): full information case.

spanning trees



The forecaster chooses a spanning tree in the complete graph $\mathbf{K_m}$. The cost is the sum of the losses over the edges.

combinatorial bandits

Two models.

(Easy.) Losses of the components of the chosen object are observed separately. (György, Lugosi, Ottucsák, 2007.)

(Interesting.) Only total loss of the chosen object is observed.
(Awerbuch and Kleinberg, 2004;
McMahan and Blum, 2004;
Dani, Hayes, and Kakade, 2008;
Abernethy, Hazan, and Rakhlin, 2008;
Bartlett, Dani, Hayes, Kakade, and Tewari, 2008;
Cesa-Bianchi and Lugosi, 2009.)

challenges

Performance bounds: Is $\mathbf{O(n^{-1/2}\text{poly}(d))}$ regret achievable for the bandit problem?

Algorithmic: How can one draw a random object from the exponentially weighted distribution in polynomial time?

model

$\mathcal{S} = \{\mathbf{v_1}, \ldots, \mathbf{v_N}\} \subset \mathbb{R}^{\mathbf{d}}$ is a collection of objects (experts).

Denote $\mathbf{B} = \max_{\mathbf{v_k}} \|\mathbf{v_k}\|_2$.

At every time $\mathbf{t} = 1, 2, \ldots$, the opponent chooses a loss vector $\boldsymbol{\ell}_\mathbf{t} \in \mathbb{R}^{\mathbf{d}}$.

We assume $\boldsymbol{\ell}_\mathbf{t}(\mathbf{k}) = \boldsymbol{\ell}_\mathbf{t}^\top \mathbf{v_k} \in [-1, 1]$ for all $\mathbf{v_k} \in \mathcal{S}$.

linear bandit problem

For $\mathbf{t = 1, 2, \ldots}$,

- ✳ opponent chooses $\boldsymbol{\ell_t} \in \mathbb{R}^{\mathbf{d}}$
- ✳ Forecaster chooses $\mathbf{K_t} \in \{\mathbf{1, \ldots, N}\}$
- ✳ The cost $\boldsymbol{\ell_t}(\mathbf{K_t}) = \boldsymbol{\ell_t}^\top \mathbf{v}_{\mathbf{K_t}}$ is revealed.

The forecaster's goal is to control the expected regret

$$\mathbb{E}\widehat{\mathbf{L}}_{\mathbf{n}} - \min_{\mathbf{k}=1,\ldots,\mathbf{N}} \mathbf{L_n}(\mathbf{k}) = \sum_{\mathbf{t}=1}^{\mathbf{n}} \mathbb{E}\boldsymbol{\ell_t}(\mathbf{K_t}) - \min_{\mathbf{k}=1,\ldots,\mathbf{N}} \sum_{\mathbf{t}=1}^{\mathbf{n}} \boldsymbol{\ell_t}(\mathbf{k}) \; .$$

Expectation is with respect to the forecaster's internal randomization.

weighted average forecaster

At time $\mathbf{t}$ assign a weight $\mathbf{w_{t,i}}$ to each $\mathbf{i = 1, \ldots, d}$.

The weight of each $\mathbf{v_k} \in \mathcal{S}$ is

$$\overline{\mathbf{w}}_{\mathbf{t}}(\mathbf{k}) = \prod_{\mathbf{i:v_k(i)=1}} \mathbf{w_{t,i}} \ .$$

weighted average forecaster

At time $\mathbf{t}$ assign a weight $\mathbf{w_{t,i}}$ to each $\mathbf{i = 1, \ldots, d}$.

The weight of each $\mathbf{v_k} \in \mathcal{S}$ is

$$\overline{\mathbf{w}}_{\mathbf{t}}(\mathbf{k}) = \prod_{\mathbf{i:v_k(i)=1}} \mathbf{w_{t,i}} \ .$$

Let $\mathbf{q_{t-1}(k) = \overline{w}_{t-1}(k) / \sum_{k=1}^{N} \overline{w}_{t-1}(k)}$.

weighted average forecaster

At time $\mathbf{t}$ assign a weight $\mathbf{w_{t,i}}$ to each $\mathbf{i = 1, \ldots, d}$.

The weight of each $\mathbf{v_k} \in \mathcal{S}$ is

$$\overline{\mathbf{w}}_{\mathbf{t}}(\mathbf{k}) = \prod_{\mathbf{i:v_k(i)=1}} \mathbf{w_{t,i}} \; .$$

Let $\mathbf{q_{t-1}(k) = \overline{w}_{t-1}(k) / \sum_{k=1}^{N} \overline{w}_{t-1}(k)}$.

At each time $\mathbf{t}$, draw $\mathbf{K_t}$ from the distribution

$$\mathbf{p_{t-1}(k) = (1-\gamma)q_{t-1}(k) + \gamma\mu(k)}$$

where μ is a fixed distribution on $\mathcal{S}$ and $\gamma > \mathbf{0}$. Here

$$\mathbf{w_{t,i} = \exp\left(-\eta\,\widetilde{L}_{t,i}\right)}$$

where $\mathbf{\widetilde{L}_{t,i} = \widetilde{\ell}_{1,i} + \cdots + \widetilde{\ell}_{t,i}}$ and $\widetilde{\ell}_{\mathbf{t,i}}$ is an estimated loss.

loss estimates

Dani, Hayes, and Kakade (2008).

Define the scaled incidence vector

$$\mathbf{X_t} = \ell_\mathbf{t}(\mathbf{K_t})\mathbf{V_{K_t}}$$

where $\mathbf{K_t}$ is distributed according to $\mathbf{p_{t-1}}$.

loss estimates

Dani, Hayes, and Kakade (2008).
Define the scaled incidence vector

$$\mathbf{X_t} = \boldsymbol{\ell_t}(\mathbf{K_t})\mathbf{V_{K_t}}$$

where $\mathbf{K_t}$ is distributed according to $\mathbf{p_{t-1}}$.

Let $\mathbf{P_{t-1}} = \mathbb{E}\left[\mathbf{V_{K_t}}\,\mathbf{V_{K_t}^\top}\right]$ be the $\mathbf{d} \times \mathbf{d}$ correlation matrix.
Hence

$$\mathbf{P_{t-1}(i,j)} = \sum_{\mathbf{k\,:\,v_k(i)=v_k(i)=1}} \mathbf{p_{t-1}(k)}\ .$$

Similarly, let $\mathbf{Q_{t-1}}$ and $\mathbf{M}$ be the correlation matrices of $\mathbb{E}\left[\mathbf{V}\,\mathbf{V}^\top\right]$ when $\mathbf{V}$ has law, $\mathbf{q_{t-1}}$ and $\boldsymbol{\mu}$. Then

$$\mathbf{P_{t-1}(i,j)} = (\mathbf{1} - \gamma)\mathbf{Q_{t-1}(i,j)} + \gamma\,\mathbf{M(i,j)}\ .$$

loss estimates

Dani, Hayes, and Kakade (2008).

Define the scaled incidence vector

$$\mathbf{X_t} = \ell_t(\mathbf{K_t})\mathbf{V_{K_t}}$$

where $\mathbf{K_t}$ is distributed according to $\mathbf{p_{t-1}}$.

Let $\mathbf{P_{t-1}} = \mathbb{E}\left[\mathbf{V_{K_t}} \, \mathbf{V_{K_t}^\top}\right]$ be the $\mathbf{d} \times \mathbf{d}$ correlation matrix.

Hence

$$\mathbf{P_{t-1}(i,j)} = \sum_{\mathbf{k : v_k(i) = v_k(i) = 1}} \mathbf{p_{t-1}(k)} \ .$$

Similarly, let $\mathbf{Q_{t-1}}$ and $\mathbf{M}$ be the correlation matrices of $\mathbb{E}\left[\mathbf{V}\,\mathbf{V}^\top\right]$ when $\mathbf{V}$ has law, $\mathbf{q_{t-1}}$ and μ. Then

$$\mathbf{P_{t-1}(i,j)} = (1-\gamma)\mathbf{Q_{t-1}(i,j)} + \gamma\,\mathbf{M(i,j)} \ .$$

The vector of loss estimates is defined by

$$\widetilde{\ell}_t = \mathbf{P_{t-1}^+ X_t}$$

where $\mathbf{P_{t-1}^+}$ is the pseudo-inverse of $\mathbf{P_{t-1}}$.

key properties

- $\mathbf{M}\,\mathbf{M}^{+}\mathbf{v} = \mathbf{v}$ for all $\mathbf{v} \in \mathcal{S}$.
- $\mathbf{Q_{t-1}}$ is positive semidefinite for every $\mathbf{t}$.
- $\mathbf{P_{t-1}}\,\mathbf{P}^{+}_{\mathbf{t-1}}\mathbf{v} = \mathbf{v}$ for all $\mathbf{t}$ and $\mathbf{v} \in \mathcal{S}$.

By definition,

$$\mathbb{E}_{\mathbf{t}}\,\mathbf{X_t} = \mathbf{P_{t-1}}\,\boldsymbol{\ell}_{\mathbf{t}}$$

and therefore

$$\mathbb{E}_{\mathbf{t}}\,\widetilde{\boldsymbol{\ell}}_{\mathbf{t}} = \mathbf{P}^{+}_{\mathbf{t-1}}\mathbb{E}_{\mathbf{t}}\,\mathbf{X_t} = \boldsymbol{\ell}_{\mathbf{t}}$$

An unbiased estimate!

performance bound

The regret of the forecaster satisfies

$$\frac{1}{n}\left(\mathbb{E}\widehat{L}_n - \min_{k=1,\ldots,N} L_n(k)\right) \le 2\sqrt{\left(\frac{2B^2}{d\lambda_{\min}(M)}+1\right)\frac{d\ln N}{n}}\;.$$

where

$$\lambda_{\min}(M) = \min_{x\in \mathrm{span}(\mathcal{S}):\|x\|=1} x^T M x > 0$$

is the smallest "relevant" eigenvalue of M. (Cesa-Bianchi and Lugosi, 2009.)

Large $\lambda_{\min}(M)$ is needed to make sure no $|\widetilde{\ell}_{t,i}|$ is too large.

performance bound

Other bounds:

$\mathbf{B}\sqrt{\mathbf{d}\ln\mathbf{N}/\mathbf{n}}$ (Dani, Hayes, and Kakade). No condition on $\boldsymbol{\mathcal{S}}$. Sampling is over a barycentric spanner.

$\mathbf{d}\sqrt{(\theta\ln\mathbf{n})/\mathbf{n}}$ (Abernethy, Hazan, and Rakhlin). Computationally efficient.

eigenvalue bounds

$$\lambda_{\min}(\mathbf{M}) = \min_{\mathbf{x} \in \mathrm{span}(\mathcal{S}): \|\mathbf{x}\|=1} \mathbb{E}(\mathbf{V}, \mathbf{x})^2 \ .$$

where $\mathbf{V}$ has distribution μ over $\mathcal{S}$.

In many cases it suffices to take μ uniform.

multitask bandit problem

The decision maker acts in $\mathbf{m}$ games in parallel.
In each game, the decision maker selects one of $\mathbf{R}$ possible actions.
After selecting the $\mathbf{m}$ actions, the sum of the losses is observed.

$$\lambda_{\min} = \frac{1}{R}$$

$$\max_k \mathbb{E}\left[\widehat{L}_n - L_n(k)\right] \le 2m\sqrt{3nR\ln R}\ .$$

The price of only observing the sum of losses is a factor of $\mathbf{m}$.

Generating a random joint action can be done in polynomial time.

assignments

Perfect matchings of $\mathbf{K_{m,m}}$.

At each time one of the $\mathbf{N = m!}$ perfect matchings of $\mathbf{K_{m,m}}$ is selected.

$$\lambda_{\min}(\mathbf{M}) = \frac{1}{\mathbf{m-1}}$$

$$\max_{\mathbf{k}} \mathbb{E}\left[\widehat{\mathbf{L}}_{\mathbf{n}} - \mathbf{L}_{\mathbf{n}}(\mathbf{k})\right] \leq \mathbf{2m}\sqrt{\mathbf{3n\ln(m!)}} \ .$$

Only a factor of $\mathbf{m}$ worse than naive full-information bound.

Sum of weights (partition function) is the permanent of a non-negative matrix. Sampling may be done by a FPAS of Jerrum, Sinclair, and Vigoda (2004).

spanning trees

In a network of $\mathbf{m}$ nodes, the cost of communication between two nodes joined by edge $\mathbf{e}$ is $\ell_t(\mathbf{e})$ at time $\mathbf{t}$. At each time a minimal connected subnetwork (a spanning tree) is selected. The goal is to minimize the total cost. $\mathbf{N = m^{m-2}}$.

$$\lambda_{\min}(\mathbf{M}) = \frac{1}{m} - O\left(\frac{1}{m^2}\right) .$$

The entries of $\mathbf{M}$ are

$$\begin{aligned}
\mathbb{P}\{V_i = 1\} &= \frac{2}{m} \\
\mathbb{P}\left\{V_i = 1,\, V_j = 1\right\} &= \frac{3}{m^2} \qquad \text{if } i \sim j \\
\mathbb{P}\left\{V_i = 1,\, V_j = 1\right\} &= \frac{4}{m^2} \qquad \text{if } i \not\sim j .
\end{aligned}$$

spanning trees

Propp and Wilson (1998) define an exact sampling algorithm. Expected running time is the average hitting time of the Markov chain defined by the edge weights $\mathbf{w_t(e) = \exp\left(-\eta\, \widetilde{L}_t(e)\right)}$.

stars

At each time a central node of a network of $\mathbf{m}$ nodes is selected. Cost is the total cost of the edges adjacent to the node.

$$\lambda_{\min} \geq 1 - O\left(\frac{1}{m}\right) \ .$$

cut sets

A balanced cut in $\mathbf{K_{2m}}$ is the collection of all edges between a set of $\mathbf{m}$ vertices and its complement. Each balanced cut has $\mathbf{m^2}$ edges and there are $\mathbf{N} = \binom{2m}{m}$ balanced cuts.

$$\lambda_{\min}(M) = \frac{1}{4} - O\left(\frac{1}{m^2}\right) .$$

Choosing from the exponentially weighted average distribution is equivalent to sampling from ferromagnetic Ising model. FPAS by Randall and Wilson (1999).

hamiltonian cycles

A Hamiltonian cycle in $\mathbf{K_m}$ is a cycle that visits each vertex exactly once and returns to the starting vertex. $\mathbf{N = (m-1)!}$

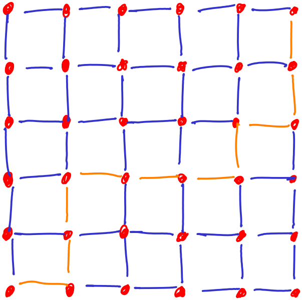
$$\boldsymbol{\lambda_{\min} \geq \frac{2}{m}}$$

Efficient computation is hopeless.

sampling paths

In all these examples μ is uniform over $\mathcal{S}$.

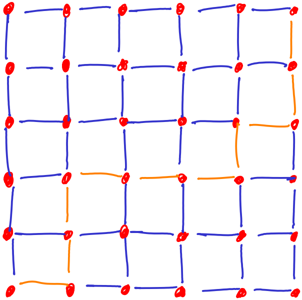
For path planning it does not always work.



sampling paths

In all these examples μ is uniform over $\mathcal{S}$.

For path planning it does not always work.



What is the optimal choice of μ?
What is the optimal way of exploration?

prediction with partial monitoring

For each round $\mathbf{t = 1, \ldots, n}$,

- the environment chooses the next outcome $\mathbf{J_t} \in \{\mathbf{1, \ldots, M}\}$ without revealing it;
- the forecaster chooses a probability distribution $\mathbf{p_t}$ and draws an action $\mathbf{I_t} \in \{\mathbf{1, \ldots, N}\}$ according to $\mathbf{p_t}$;
- the forecaster incurs loss $\boldsymbol{\ell}(\mathbf{I_t}, \mathbf{J_t})$ and each action $\mathbf{i}$ incurs loss $\boldsymbol{\ell}(\mathbf{i}, \mathbf{J_t})$. None of these values is revealed to the forecaster;
- the feedback $\mathbf{h}(\mathbf{I_t}, \mathbf{J_t})$ is revealed to the forecaster.

$\mathbf{H} = [\mathbf{h}(\mathbf{i}, \mathbf{j})]_{\mathbf{N \times M}}$ is the feedback matrix.

$\mathbf{L} = [\boldsymbol{\ell}(\mathbf{i}, \mathbf{j})]_{\mathbf{N \times M}}$ is the loss matrix.

examples

Dynamic pricing. Here $\mathbf{M = N}$, and $\mathbf{L = [\ell(i,j)]_{N\times N}}$ where

$$\ell(i,j) = \frac{(j-i)\mathbb{1}_{\{i\le j\}} + c\mathbb{1}_{\{i>j\}}}{N} \; .$$

and $h(i,j) = \mathbb{1}_{\{i>j\}}$ or

$$h(i,j) = a\mathbb{1}_{\{i\le j\}} + b\mathbb{1}_{\{i>j\}} \; , \quad i,j = 1,\ldots,N \; .$$

examples

Dynamic pricing. Here $\mathbf{M = N}$, and $\mathbf{L = [\ell(i,j)]_{N \times N}}$ where

$$\ell(i,j) = \frac{(j-i)\mathbb{1}_{\{i \le j\}} + c\mathbb{1}_{\{i > j\}}}{N} \; .$$

and $\mathbf{h(i,j)} = \mathbb{1}_{\{i > j\}}$ or

$$h(i,j) = a\mathbb{1}_{\{i \le j\}} + b\mathbb{1}_{\{i > j\}} \; , \quad i,j = 1,\ldots,N \; .$$

Multi-armed bandit problem. The only information the forecaster receives is his own loss: $\mathbf{H = L}$.

examples

Apple tasting. $\mathbf{N} = \mathbf{M} = \mathbf{2}$.

$$\mathbf{L} = \begin{bmatrix} \mathbf{0} & \mathbf{1} \\ \mathbf{1} & \mathbf{0} \end{bmatrix}$$

$$\mathbf{H} = \begin{bmatrix} \mathbf{a} & \mathbf{a} \\ \mathbf{b} & \mathbf{c} \end{bmatrix} .$$

The predictor only receives feedback when he chooses the second action.

examples

Apple tasting. $\mathbf{N} = \mathbf{M} = \mathbf{2}$.

$$\mathbf{L} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

$$\mathbf{H} = \begin{bmatrix} a & a \\ b & c \end{bmatrix} .$$

The predictor only receives feedback when he chooses the second action.

Label efficient prediction. $\mathbf{N} = \mathbf{3}$, $\mathbf{M} = \mathbf{2}$.

$$\mathbf{L} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}$$

$$\mathbf{H} = \begin{bmatrix} a & b \\ c & c \\ c & c \end{bmatrix} .$$

a general predictor

A forecaster first proposed by Piccolboni and Schindelhauer (2001).
Crucial assumption: $\mathbf{H}$ can be encoded such that there exists an $\mathbf{N} \times \mathbf{N}$ matrix $\mathbf{K} = [\mathbf{k}(\mathbf{i},\mathbf{j})]_{\mathbf{N}\times\mathbf{N}}$ such that

$$\mathbf{L} = \mathbf{K} \cdot \mathbf{H} \ .$$

Thus,

$$\ell(\mathbf{i},\mathbf{j}) = \sum_{\mathbf{l}=1}^{\mathbf{N}} \mathbf{k}(\mathbf{i},\mathbf{l})\mathbf{h}(\mathbf{l},\mathbf{j}) \ .$$

Then we may estimate the losses by

$$\widetilde{\ell}(\mathbf{i},\mathbf{J_t}) = \frac{\mathbf{k}(\mathbf{i},\mathbf{I_t})\mathbf{h}(\mathbf{I_t},\mathbf{J_t})}{\mathbf{p_{I_t,t}}} \ .$$

a general predictor

Observe

$$\begin{aligned}\mathbb{E}_t \widetilde{\ell}(i, J_t) &= \sum_{k=1}^{N} p_{k,t} \frac{k(i,k)h(k,J_t)}{p_{k,t}} \\ &= \sum_{k=1}^{N} k(i,k)h(k,J_t) = \ell(i, J_t)\ ,\end{aligned}$$

$\widetilde{\ell}(i, J_t)$ is an unbiased estimate of $\ell(i, J_t)$.
Let

$$p_{i,t} = (1-\gamma)\frac{e^{-\eta \widetilde{L}_{i,t-1}}}{\sum_{k=1}^{N} e^{-\eta \widetilde{L}_{k,t-1}}} + \frac{\gamma}{N}$$

where $\widetilde{L}_{i,t} = \sum_{s=1}^{t} \widetilde{\ell}(i, J_t)$.

performance bound

With probability at least $1-\delta$,

$$\frac{1}{n}\sum_{t=1}^{n}\ell(I_t,J_t) - \min_{i=1,\ldots,N}\frac{1}{n}\sum_{t=1}^{n}\ell(i,J_t) \le Cn^{-1/3}N^{2/3}\sqrt{\ln(N/\delta)}\ .$$

where C depends on K. (Cesa-Bianchi, Lugosi, Stoltz (2006))

performance bound

With probability at least $1-\delta$,

$$\frac{1}{n}\sum_{t=1}^{n}\ell(I_t,J_t)-\min_{i=1,\ldots,N}\frac{1}{n}\sum_{t=1}^{n}\ell(i,J_t) \leq Cn^{-1/3}N^{2/3}\sqrt{\ln(N/\delta)}\,.$$

where C depends on K. (Cesa-Bianchi, Lugosi, Stoltz (2006))

Hannan consistency is achieved with rate $O(n^{-1/3})$ whenever $L = K \cdot H$.

This solves the dynamic pricing problem.

performance bound

With probability at least $\mathbf{1-\delta}$,

$$\frac{1}{n}\sum_{t=1}^{n}\ell(I_t,J_t) - \min_{i=1,\dots,N}\frac{1}{n}\sum_{t=1}^{n}\ell(i,J_t) \le \quad Cn^{-1/3}N^{2/3}\sqrt{\ln(N/\delta)}\;.$$

where $\mathbf{C}$ depends on $\mathbf{K}$. (Cesa-Bianchi, Lugosi, Stoltz (2006))

Hannan consistency is achieved with rate $\mathbf{O(n^{-1/3})}$ whenever $\mathbf{L = K \cdot H}$.

This solves the dynamic pricing problem.

Bartók, Pál, and Szepesvári (2010): if $\mathbf{M = 2}$, only possible rates are $\mathbf{n^{-1/2}, n^{-1/3}, 1}$

imperfect monitoring: a general framework

$\mathbf{S}$ is a finite set of signals.

Feedback matrix: $\mathbf{H} : \{1, \ldots, N\} \times \{1, \ldots, M\} \to \mathcal{P}(\mathbf{S})$.

For each round $\mathbf{t} = 1, 2 \ldots, n$,

- the environment chooses the next outcome $J_t \in \{1, \ldots, M\}$ without revealing it;
- the forecaster chooses p_t and draws an action $I_t \in \{1, \ldots, N\}$ according to it;
- the forecaster receives loss $\ell(I_t, J_t)$ and each action $\mathbf{i}$ suffers loss $\ell(\mathbf{i}, J_t)$, none of these values is revealed to the forecaster;
- a feedback s_t drawn at random according to $\mathbf{H}(I_t, J_t)$ is revealed to the forecaster.

target

Define

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

$$H(\cdot, \mathbf{q}) = (H(1, \mathbf{q}), \ldots, H(N, \mathbf{q}))$$

where $H(i, \mathbf{q}) = \sum_j q_j H(i, j)$.

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

$\mathcal{F}$ is the set of "observable" vectors of signal distributions $\boldsymbol{\Delta}$.
The key quantity is

$$\rho(\mathbf{p}, \boldsymbol{\Delta}) = \max_{\mathbf{q} \,:\, H(\cdot, \mathbf{q}) = \boldsymbol{\Delta}} \ell(\mathbf{p}, \mathbf{q})$$

ρ is convex in $\mathbf{p}$ and concave in $\boldsymbol{\Delta}$.

rustichini's theorem

The value of the base one-shot game is

$$\min_{\mathbf{p}} \max_{\mathbf{q}} \ell(\mathbf{p}, \mathbf{q}) = \min_{\mathbf{p}} \max_{\boldsymbol{\Delta} \in \mathcal{F}} \rho(\mathbf{p}, \boldsymbol{\Delta})$$

If $\overline{\mathbf{q}}_{\mathbf{n}}$ is the empirical distribution of $\mathbf{J_1, \ldots, J_n}$, even with the knowledge of $\mathbf{H}(\cdot, \overline{\mathbf{q}}_{\mathbf{n}})$ we cannot hope to do better than $\min_{\mathbf{p}} \rho(\mathbf{p}, \mathbf{H}(\cdot, \overline{\mathbf{q}}_{\mathbf{n}}))$.

Rustichini (1999) proved that there exists a strategy such that for all strategies of the opponent, almost surely,

$$\limsup_{\mathbf{n} \to \infty} \left(\frac{1}{\mathbf{n}} \sum_{\mathbf{t}=1,\ldots,\mathbf{n}} \ell(\mathbf{I_t}, \mathbf{J_t}) - \min_{\mathbf{p}} \rho\left(\mathbf{p}, \mathbf{H}(\cdot, \overline{\mathbf{q}}_{\mathbf{n}})\right) \right) \leq \mathbf{0}$$

rustichini's theorem

Rustichini's proof relies on an approachability theorem for a continuum of types (Mertens, Sorin, and Zamir, 1994).

It is non-constructive.

It does not imply any convergence rate.

Lugosi, Mannor, and Stoltz (2008) construct efficiently computable strategies that guarantee fast rates of convergence.