

**Sequential prediction under
imperfect monitoring**

Gábor Lugosi

ICREA and Pompeu Fabra University,
Barcelona

based on joint work with

**Nicolò Cesa-Bianchi, Shie Mannor, and
Gilles Stoltz**

Motivating example: on-line pricing

A seller sells n pieces of a product to n customers.

Customers come one by one.

To customer number t , the seller offers the product at a price $I_t \in [0, 1]$.

Each customer has a maximum price J_t he/she is willing to pay but does not tell it to the seller.

If $J_t \geq I_t$, the product is bought and the seller suffers a “loss” $J_t - I_t$.

If $J_t < I_t$, the product is not bought and the seller’s loss is $c \in [0, 1]$.

On-line pricing

The seller's loss function is

$$\ell(I_t, J_t) = (J_t - I_t)\mathbb{I}_{[I_t \leq J_t]} + c\mathbb{I}_{[I_t > J_t]}$$

The values J_t are arbitrary and may even depend on the seller's past actions.

If the seller knew the “distribution” of J_1, \dots, J_n in advance, he could choose the value p minimizing the total loss

$$\frac{1}{n} \sum_{t=1}^n \ell(p, J_t) .$$

Result: The seller has a (randomized) strategy such that, for all $\delta \in [0, 1]$, with probability at least $1 - \delta$,

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

Randomized prediction

A game between forecaster and environment.

At each round t , the forecaster chooses an action $I_t \in \{1, \dots, N\}$;

the environment chooses an action $J_t \in \mathcal{Y}$;

the forecaster suffers loss $\ell(I_t, J_t) \in [0, 1]$.

The goal is to minimize the *cumulative excess loss*

$$\frac{1}{n} \left(\sum_{t=1}^n \ell(I_t, J_t) - \min_{i \leq N} \sum_{t=1}^n \ell(i, J_t) \right).$$

The forecaster may randomize. At time t chooses a probability distribution

$$\mathbf{p}_t = (p_{1,t}, \dots, p_{N,t})$$

and plays action i with probability $p_{i,t}$.

Actions are often called “experts”.

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;

The simplest model assumes that after each round, the losses $\ell(i, J_t)$ ($i = 1, \dots, N$) are revealed (*full information*).

In this model **Hannan** (1957) showed that the forecaster has a strategy such that

$$\frac{1}{n} \left(\sum_{t=1}^n \ell(I_t, J_t) - \min_{i \leq N} \sum_{t=1}^n \ell(i, J_t) \right) \rightarrow 0$$

almost surely for all strategies of the environment.

Hannan consistency: basic ideas

Obviously, the forecaster must randomize.

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

denotes the “expected” loss of the forecaster.

By martingale convergence,

$$\frac{1}{n} \left(\sum_{t=1}^n \ell(I_t, J_t) - \sum_{t=1}^n \ell(\mathbf{p}_t, J_t) \right) = O_P(n^{-1/2})$$

so it suffices to study

$$\frac{1}{n} \left(\sum_{t=1}^n \ell(\mathbf{p}_t, J_t) - \min_{i \leq N} \sum_{t=1}^n \ell(i, J_t) \right)$$

Weighted average prediction

Idea: assign a higher probability to better-performing actions.

A popular choice is

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

where $\eta > 0$. Then

$$\begin{aligned} \frac{1}{n} \left(\sum_{t=1}^n \ell(\mathbf{p}_t, J_t) - \min_{i \leq N} \sum_{t=1}^n \ell(i, J_t) \right) &\leq \frac{\ln N}{n\eta} + \frac{\eta}{8} \\ &= \sqrt{\frac{\ln N}{2n}} \end{aligned}$$

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

Label efficient prediction

In this variant the forecaster does not see the outcome J_t unless he asks for it, but can do it only $m \ll n$ times.

The game is the following:

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

- (1) the environment chooses the outcome $J_t \in \mathcal{Y}$ without revealing it;
- (2) the forecaster chooses \mathbf{p}_t and draws an action $I_t \in \{1, \dots, N\}$ according to this distribution;
- (3) the forecaster incurs loss $\ell(I_t, J_t)$ and each action i incurs loss $\ell(i, J_t)$, none of these values is revealed to the forecaster;
- (4) the forecaster decides whether he asks for the value of J_t if the total number of revealed outcomes up to time $t - 1$ is less than m .

A label efficient forecaster

The idea is to ask for labels randomly (with probability $\approx m/n$) and use the weighted average forecaster with the estimated losses.

Let Z_t be i.i.d. Bernoulli ϵ ($\approx m/n$).

The forecaster asks for J_t iff $Z_t = 1$.

Let

$$\tilde{\ell}(i, J_t) \stackrel{\text{def}}{=} \begin{cases} \ell(i, J_t)/\epsilon & \text{if } Z_t = 1, \\ 0 & \text{otherwise.} \end{cases}$$

An unbiased estimate!

For each round $t = 1, 2, \dots, n$ draw an action from $\{1, \dots, N\}$ according to the distribution

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

Bound for label efficient prediction

With probability at least $1 - \delta$,

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

Sketch of proof:

First bound

$$\sum_{t=1}^n \tilde{\ell}(\mathbf{p}_t, J_t) - \min_{i \leq N} \sum_{t=1}^n \tilde{\ell}(i, J_t)$$

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

$$\sum_{t=1}^n \ell(I_t, J_t) - \sum_{t=1}^n \tilde{\ell}(\mathbf{p}_t, J_t)$$

and

$$\min_{i \leq N} \sum_{t=1}^n \tilde{\ell}(i, J_t) - \min_{i \leq N} \sum_{t=1}^n \ell(i, J_t)$$

Partial monitoring

In a more general setup the information received by the forecaster after making a prediction I_t is a *feedback* $h(I_t, J_t)$.

We assume that $\mathcal{Y} = \{1, \dots, M\}$.

The matrix of losses is $\mathbf{L} = [\ell(i, j)]_{N \times M}$ (known by the forecaster).

At time t , the forecaster chooses action $I_t \in \{1, \dots, N\}$ and the outcome is $J_t \in \mathcal{Y}$.

The forecaster's loss is $\ell(I_t, J_t)$.

The forecaster only observes the feedback $h(I_t, J_t)$ where $\mathbf{H} = [h(i, j)]_{N \times M}$ is the feedback matrix (with values from a finite set).

Prediction with partial monitoring

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

- (1) the environment chooses the next outcome $J_t \in \mathcal{Y}$ without revealing it;
- (2) the forecaster chooses a probability distribution \mathbf{p}_t and draws an action $I_t \in \{1, \dots, N\}$ according to \mathbf{p}_t ;
- (3) the forecaster incurs loss $\ell(I_t, J_t)$ and each action i incurs loss $\ell(i, J_t)$. None of these values is revealed to the forecaster;
- (4) the feedback $h(I_t, J_t)$ is revealed to the forecaster.

Examples

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

$$\ell(i, j) = \frac{(j - i)\mathbb{I}_{[i \leq j]} + c\mathbb{I}_{[i > j]}}{N} .$$

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

$$h(i, j) = a\mathbb{I}_{[i \leq j]} + b\mathbb{I}_{[i > j]} , \quad i, j = 1, \dots, N .$$

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

Examples

Apple tasting. $N = M = 2$.

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

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

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

Label efficient prediction. $N = 3, M = 2$.

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

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

A general predictor

A forecaster first proposed by Piccolboni and Schindelhauer.

Crucial assumption: \mathbf{H} can be encoded such that there exists an $N \times N$ matrix $\mathbf{K} = [k(i, j)]_{N \times N}$ such that

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

Thus,

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

Then we may estimate the losses by

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

A general predictor

Observe

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

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

Let

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

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

Performance bound

For all $\delta \in [0, 1]$, with probability at least $1 - \delta$,

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

where C depends on \mathbf{K} .

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

This solves the dynamic pricing problem.

Whenever Hannan consistency is achievable, a version of this predictor works and attains rate $O(n^{-1/3})$.

Optimality

The example of label efficient prediction shows that the rate $O(n^{-1/3})$ is not improvable, in general.

Bandit problems. In this case $\mathbf{H} = \mathbf{L}$ so \mathbf{K} is the identity matrix.

The forecaster becomes

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

suggested by Auer, Cesa-Bianchi, Freund, and Schapire.

They show that a carefully modified version achieves a faster $O(n^{-1/2})$ rate (as in the full information case).

General problem

S is a finite set of signals. The feedback matrix is $H : \{1, \dots, N\} \times \{1, \dots, M\} \rightarrow \mathcal{P}(S)$.

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

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

Target

Define

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

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

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

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

\mathcal{F} is the set of “observable” vectors of signal distributions Δ .

The key quantity is

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

λ is convex in \mathbf{p} and concave in Δ .

Rustichini's theorem

If $\bar{\mathbf{q}}_n$ is the empirical distribution of J_1, \dots, J_n , even with the knowledge of $H(\cdot, \bar{\mathbf{q}}_n)$ we cannot hope to do better than $\min_{\mathbf{p}} \lambda(\mathbf{p}, H(\cdot, \bar{\mathbf{q}}_n))$.

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

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1, \dots, n} \ell(I_t, J_t) - \min_{\mathbf{p}} \lambda(\mathbf{p}, H(\cdot, \bar{\mathbf{q}}_n)) \leq 0$$

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

It is non-constructive.

No convergence rates.

Deterministic feedback, depending on the outcome

Consider first the special case

$H(I_t, J_t) = h(J_t)$, deterministic.

Introduce a forecaster motivated by the gradient-based strategies (see Cesa-Bianchi and Lugosi, 2006, Section 2.5).

The forecaster uses a sub-gradient of $\lambda(\cdot, \delta_{h(J_t)})$. $I_t = i$ with probability

$$p_{i,t} = \frac{e^{-\eta \sum_{s=1}^{t-1} (\tilde{\ell}(\mathbf{p}_s, \delta_{h(J_s)}))_i}}{\sum_{j=1}^N e^{-\eta \sum_{s=1}^{t-1} (\tilde{\ell}(\mathbf{p}_s, \delta_{h(J_s)}))_j}},$$

where $(\tilde{\ell}(\mathbf{p}_s, \delta_{h(J_s)}))_i$ is the i -th component of a sub-gradient $\tilde{\ell}(\mathbf{p}_s, \delta_{h(J_s)}) \in \nabla \lambda(\mathbf{p}_s, \delta_{h(J_s)})$ of the convex function $\lambda(\cdot, \delta_{h(J_s)})$.

$$\frac{1}{n} \sum_{t=1}^n \ell(I_t, J_t) - \lambda(\mathbf{p}, H(\bar{\mathbf{q}}_n)) = O(\sqrt{(\ln N)/n})$$

Random feedback, depending on the outcome

We still assume $H(I_t, J_t) = H(J_t)$, but $H(j)$ is a distribution over signals.

At time t the forecaster observes s_t drawn from $H(J_t)$.

$$H(\bar{q}_n) = \frac{1}{n} \sum_{t=1}^n H(J_t)$$

needs to be estimated.

Idea: a lazy strategy. Group together $m \ll n$ time rounds and use the same mixed strategy.

In the b -th group one can calculate

$$\frac{1}{m} \sum_{t=bm+1}^{(b+1)m} \delta_{s_t}$$

and project it to the set \mathcal{F} of feasible distributions:

$$\widehat{\Delta}^b = \Pi \left(\frac{1}{m} \sum_{t=bm+1}^{(b+1)m} \delta_{s_t} \right).$$

If

$$\Delta^b = \frac{1}{m} \sum_{t=bm+1}^{(b+1)m} H(J_t) \quad (\in \mathcal{F})$$

then $\widehat{\Delta}^b \approx \Delta^b$ by (vector valued) martingale convergence.

The strategy

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

1. If $bm + 1 \leq t < (b + 1)m$ for some integer b , choose the distribution $\mathbf{p}_t = \mathbf{p}^b$ given by

$$p_{k,t} = p_k^b = \frac{w_k^b}{\sum_{j=1}^N w_j^b}$$

and draw an action I_t from $\{1, \dots, N\}$ according to it;

2. if $t = (b + 1)m$ for some integer b , perform the update

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

where for all Δ , $\tilde{\ell}(\cdot, \Delta)$ is a sub-gradient of $\lambda(\cdot, \Delta)$.

Performance

By optimizing parameters $m \approx \sqrt{n}$ and $\eta \approx n^{-1/4} \sqrt{\ln N}$, we get the regret bound

$$\begin{aligned} \frac{1}{n} \sum_{t=1, \dots, n} r(I_t, J_t) - \min_{\mathbf{p}} \lambda(\mathbf{p}, H(\cdot, \bar{\mathbf{q}}_n)) \\ = O(n^{-1/4} \sqrt{\ln(nN/\delta)}) \end{aligned}$$

which holds with probability $> 1 - \delta$.

The general case

Now the random feedback depends on the action–outcome pairs: $H(I_t, J_t)$.

Again, we need to estimate the (unobserved) $H(\cdot, \bar{\mathbf{q}}_n)$. Let

$$\hat{h}_{i,t} = \frac{\delta_{st}}{p_{i,t}} \mathbb{I}_{[I_t=i]} .$$

Then $\hat{h}_{i,t}$ is conditionally unbiased:

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

Define

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

This will be close to

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

provided $p_{i,t}$ is not too small.

The strategy

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

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

where for all $\Delta \in \mathcal{F}$, $\tilde{\ell}(\cdot, \Delta)$ is a
sub-gradient of $\lambda(\cdot, \Delta)$.

Performance

By choosing $m \approx n^{3/5}$, $\gamma \approx n^{-1/5}$, and $\eta \approx n^{-1/5} \sqrt{\ln N}$, we obtain

$$\begin{aligned} \frac{1}{n} \sum_{t=1, \dots, n} \ell(I_t, J_t) - \min_{\mathbf{p}} \lambda(\mathbf{p}, H(\cdot, \bar{\mathbf{q}}_n)) \\ = O(n^{-1/5} N \sqrt{\ln(nN/\delta)}) \end{aligned}$$

which holds with probability $> 1 - \delta$.

For deterministic feedback this can be improved to $O(n^{-1/3} N^{2/3} \sqrt{\ln(1/\delta)})$.

In the deterministic case the rates (as a function of n) cannot be improved.

Remarks

The strategies involve computation of l_2 projections to a convex set and computation of (sub)gradients of piecewise linear concave functions.

These can be done in time polynomial in N and $|S|$.

Are the obtained rates optimal in the case of random feedback?

Hannan consistency

Our strategies are *Hannan consistent* whenever

$$\min_{\mathbf{p}} \lambda(\mathbf{p}, H(\cdot, \mathbf{q})) = \min_{i=1, \dots, N} \ell(i, \mathbf{q}) .$$

For example, Hannan consistency is possible if there is a matrix \mathbf{K} such that $\mathbf{R} = \mathbf{KH}$.

We have a different sufficient condition (independent of the rewards):

If H doesn't have identical columns, then for all \mathbf{q} ,

$$\min_{\mathbf{p}} \lambda(\mathbf{p}, H(\cdot, \mathbf{q})) = \min_{i=1, \dots, N} \ell(i, \mathbf{q}) .$$