Practical Aspects of Levin's Neutral Measure

Alexey Chernov

Computer Learning Research Centre and Department of Computer Science
Royal Holloway University of London, UK

Marseille, 02.07.2009

Outline

Uniform Tests

Y is a metric compact, $\mathcal{P}(Y)$ is the set of all measures on Y

A lower semicontinuous function $t\colon Y \times \mathcal{P}(Y) \to \mathbb{R}$ is a uniform test if

$$\forall \mu \in \mathcal{P}(Y) \quad \int_Y t(y, \mu) d\mu(y) \leq 1$$

Uniform Tests

Y is a metric compact, $\mathcal{P}(Y)$ is the set of all measures on Y

A lower semicontinuous function $t \colon Y \times \mathcal{P}(Y) \to \mathbb{R}$ is a uniform test if

$$\forall \mu \in \mathcal{P}(Y) \quad \int_Y t(y, \mu) d\mu(y) \leq 1$$

Example:
Fix any measure ν on a finite set Y.

$t(y, \mu) = \frac{\nu(y)}{\mu(y)}$ is a uniform test.

Neutral Measure

Theorem

Let $t(y, \mu)$ be a uniform test on metric compact Y.
There exists a measure M on Y s.t.

$$\forall y \in Y \quad t(y, M) \leq 1 \, .$$

Neutral Measure

Theorem

Let $t(y, \mu)$ be a uniform test on metric compact Y.
There exists a measure M on Y s.t.

$$\forall y \in Y \quad t(y, M) \leq 1 \, .$$

Neutral measure was introduced (with computability requirements) in:
Leonid Levin, "Uniform Tests of Randomness" (1976).
For details see: Peter Gács, "Lecture notes on descriptional complexity and randomness", Ch. 16.

We will ignore computability issues

Example: Neutral Measure on $\{0,1\}$

$Y=\{0,1\}$, $\mathcal{P}(Y)\cong[0,1]$: $Prob(1)=p$, $Prob(0)=1-p$

Let $t(y,p)$ be lower semicontinuous in p and

$$\forall p \quad \mathbf{E}_p t(y,p) = p\,t(1,p) + (1-p)t(0,p) \le 1$$

There exists p s.t. $t(1,p)\le 1$ and $t(0,p)\le 1$.

Example: Neutral Measure on $\{0,1\}$

$Y = \{0,1\}$, $\mathcal{P}(Y) \cong [0,1]$: $Prob(1) = p$, $Prob(0) = 1 - p$

Let $t(y,p)$ be lower semicontinuous in p and

$$\forall p \quad \mathbf{E}_p t(y,p) = p\, t(1,p) + (1-p)t(0,p) \leq 1$$

There exists p s.t. $t(1,p) \leq 1$ and $t(0,p) \leq 1$.

Proof.

- $\{p \mid t(0,p) \leq 1\} \cup \{p \mid t(1,p) \leq 1\} = [0,1]$
- $0 \in \{p \mid t(0,p) \leq 1\}$ and $1 \in \{p \mid t(1,p) \leq 1\}$
- $\{p \mid t(0,p) \leq 1\}$ and $\{p \mid t(1,p) \leq 1\}$ are closed

Then $\{p \mid t(0,p) \leq 1\} \cap \{p \mid t(1,p) \leq 1\} \neq \emptyset$ □

Outline

Binary Betting Protocol

Initial capital of Skeptic $\mathcal{K}_0 = 1$
For $n = 1, 2, \ldots$

> Forecaster announces $p_n \in [0, 1]$
>
> Skeptic buys $s_n \in \mathbb{R}$ tickets
>
> Reality announces $y_n \in \{0, 1\}$
>
> $\mathcal{K}_n = \mathcal{K}_{n-1} + s_n(y_n - p_n)$

Binary Betting Protocol

Initial capital of Skeptic $\mathcal{K}_0 = 1$
For $n = 1, 2, \dots$

Forecaster announces $p_n \in [0, 1]$

Skeptic buys $s_n \in \mathbb{R}$ tickets

Reality announces $y_n \in \{0, 1\}$

$\mathcal{K}_n = \mathcal{K}_{n-1} + s_n(y_n - p_n)$

Goal: $\mathcal{K}_n$ remains bounded.
Forecaster plays against Skeptic and Reality.

Binary Betting Protocol

Initial capital of Skeptic $\mathcal{K}_0 = 1$
For $n = 1, 2, \ldots$

> Forecaster announces $p_n \in [0, 1]$
>
> Skeptic buys $s_n \in \mathbb{R}$ tickets
>
> Reality announces $y_n \in \{0, 1\}$
>
> $\mathcal{K}_n = \mathcal{K}_{n-1} + s_n(y_n - p_n)$

Goal: $\mathcal{K}_n$ remains bounded.
Forecaster plays against Skeptic and Reality.

Generally unachievable: y_n s.t. $|y_n - p_n| \geq 0.5$, $s_n = \text{sign}(y_n - p_n)$.

Binary Defensive Forecasting Protocol

Initial capital of Skeptic $\mathcal{K}_0 = 1$
For $n = 1, 2, \ldots$

Skeptic announces $s_n \colon [0,1] \to \mathbb{R}$

Forecaster announces $p_n \in [0,1]$

Reality announces $y_n \in \{0,1\}$

$\mathcal{K}_n = \mathcal{K}_{n-1} + s_n(p_n)(y_n - p_n)$

Binary Defensive Forecasting Protocol

Initial capital of Skeptic $\mathcal{K}_0 = 1$
For $n = 1, 2, \dots$

Skeptic announces continuous $s_n \colon [0,1] \to \mathbb{R}$

Forecaster announces $p_n \in [0,1]$

Reality announces $y_n \in \{0,1\}$

$\mathcal{K}_n = \mathcal{K}_{n-1} + s_n(p_n)(y_n - p_n)$

Binary Defensive Forecasting Protocol

Initial capital of Skeptic $\mathcal{K}_0 = 1$
For $n = 1, 2, \ldots$

Skeptic announces continuous $s_n \colon [0, 1] \to \mathbb{R}$

Forecaster announces $p_n \in [0, 1]$

Reality announces $y_n \in \{0, 1\}$

$\mathcal{K}_n = \mathcal{K}_{n-1} + s_n(p_n)(y_n - p_n)$

Theorem

Forecaster has a strategy that guarantees $\mathcal{K}_0 \geq \mathcal{K}_1 \geq \mathcal{K}_2 \geq \ldots$

Forecaster's Strategy

s_n is continuous

$$\mathbf{E}_p s_n(p)(y-p) = p s_n(p)(1-p) + (1-p) s_n(p)(0-p) = 0$$

By the Neutral Measure Theorem, $\exists p \, \forall y \in \{0,1\} \;\; s_n(p)(y-p) \leq 0$
Thus $\mathcal{K}_n \leq \mathcal{K}_{n-1}$

Forecaster's Strategy

s_n is continuous

$$\mathbf{E}_p s_n(p)(y-p) = p s_n(p)(1-p) + (1-p) s_n(p)(0-p) = 0$$

By the Neutral Measure Theorem, $\exists p \, \forall y \in \{0,1\} \;\; s_n(p)(y-p) \leq 0$
Thus $\mathcal{K}_n \leq \mathcal{K}_{n-1}$

Explicit algorithm:

- If $s_n(1) > 0$, take $p_n = 1$.
- If $s_n(0) < 0$, take $p_n = 0$.
- If $s_n(0) \geq 0$ and $s_n(1) \leq 0$, take p_n s.t. $s_n(p_n) = 0$.

Thus $s_n(p_n)(y_n - p_n) \leq 0$ for any $y_n \in \{0,1\}$.

Uniformly Good Forecasts

For $n = 1, 2, \ldots$

Forecaster announces $p_n \in [0, 1]$

Reality announces $y_n \in \{0, 1\}$

Sceptic announces a continuous function $f\colon [0, 1] \to \mathbb{R}$

Uniformly Good Forecasts

For $n = 1, 2, \ldots$

Forecaster announces $p_n \in [0, 1]$

Reality announces $y_n \in \{0, 1\}$

Sceptic announces a continuous function $f\colon [0, 1] \to \mathbb{R}$

Forecaster's goal: $\sum_{n=1}^{N} f(p_n)(y_n - p_n)$ is small (grow slowly)

Calibration

y_n is a Bernoulli sequence with $Pr(1) = p$

$$\frac{1}{N}\sum_{n=1}^{N}(y_n - p) \to 0$$

Calibration

y_n is a Bernoulli sequence with $Pr(1) = p$

$$\frac{1}{N}\sum_{n=1}^{N}(y_n - p) \to 0$$

If p_n are different:

For each p take subsequence s.t. $p_n = p$

Calibration

y_n is a Bernoulli sequence with $Pr(1) = p$

$$\frac{1}{N}\sum_{n=1}^{N}(y_n - p) \to 0$$

If p_n are different:

For each p take subsequence s.t. $p_n = p$

To get long (and frequently occurring) subsequences: $p_n \approx p$

$$\frac{1}{N}\sum_{n=1}^{N}\mathbb{I}_{p_n \approx p}(y_n - p_n) \to 0$$

Two Skeptics

Consider a simple version of uniformly good forecasting:
a game with two Skeptics.

At step n, Skeptics buy $s_n^1(p_n)$ and $s_n^2(p_n)$ tickets, respectively.

Their capitals are $\mathcal{K}_n^i = \mathcal{K}_{n-1}^i + s_n^i(p_n)(y_n - p_n)$, $i = 1, 2$.

Goal: both capitals remain small (grow slowly).

Two Skeptics

Consider a simple version of uniformly good forecasting:
a game with two Skeptics.

At step n, Skeptics buy $s_n^1(p_n)$ and $s_n^2(p_n)$ tickets, respectively.

Their capitals are $\mathcal{K}_n^i = \mathcal{K}_{n-1}^i + s_n^i(p_n)(y_n - p_n)$, $i = 1, 2$.

Goal: both capitals remain small (grow slowly).

Let $s_n = \begin{pmatrix} s_n^1 \\ s_n^2 \end{pmatrix}$, $\mathcal{K}_n = \begin{pmatrix} \mathcal{K}_n^1 \\ \mathcal{K}_n^2 \end{pmatrix}$.

$\|\mathcal{K}_n\|^2 = (\mathcal{K}_n^1)^2 + (\mathcal{K}_n^2)^2$.

Two Skeptics: Capital Bound

Initial capital of Skeptics $\mathcal{K}_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

For $n = 1, 2, \dots$

Skeptic announces continuous $s_n \colon [0, 1] \to \mathbb{R}^2$

Forecaster announces $p_n \in [0, 1]$

Reality announces $y_n \in \{0, 1\}$

$\mathcal{K}_n = \mathcal{K}_{n-1} + s_n(p_n)(y_n - p_n)$

Two Skeptics: Capital Bound

Initial capital of Skeptics $\mathcal{K}_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

For $n = 1, 2, \dots$

- Skeptic announces continuous $s_n \colon [0,1] \to \mathbb{R}^2$
- Forecaster announces $p_n \in [0,1]$
- Reality announces $y_n \in \{0,1\}$
- $\mathcal{K}_n = \mathcal{K}_{n-1} + s_n(p_n)(y_n - p_n)$

Theorem

Forecaster has a strategy that guarantees

$$\|\mathcal{K}_n\| = \left\| \sum_{i=1}^{n} s_i(p_i)(y_i - p_i) \right\| \le \sqrt{n} \max_i \max_{p \in [0,1]} \|s_i(p)\|$$

Two Skeptics: Forecaster's Strategy (1)

$$\|\mathcal{K}_n\|^2 = \left\| \sum_{i=1}^{n-1} s_i(p_i)(y_i - p_i) + s_n(p_n)(y_n - p_n) \right\|^2$$

$$= \left\| \sum_{i=1}^{n-1} s_i(p_i)(y_i - p_i) \right\|^2 + \|s_n(p_n)(y_n - p_n)\|^2$$

$$+ 2 \left\langle s_n(p_n)(y_n - p_n), \sum_{i=1}^{n-1} s_i(p_i)(y_i - p_i) \right\rangle$$

$$= \|\mathcal{K}_{n-1}\|^2 + \|s_n(p_n)\|^2 (y_n - p_n)^2$$

$$+ 2 \left\langle s_n(p_n), \sum_{i=1}^{n-1} s_i(p_i)(y_i - p_i) \right\rangle (y_n - p_n)$$

Two Skeptics: Forecaster's Strategy (2)

$$\|\mathcal{K}_n\|^2 = \|\mathcal{K}_{n-1}\|^2 + \|s_n(p_n)\|^2(y_n - p_n)^2$$
$$+2\left\langle s_n(p_n), \textstyle\sum_{i=1}^{n-1} s_i(p_i)(y_i - p_i)\right\rangle (y_n - p_n)$$

By the Neutral Measure Theorem, we can take p_n s.t.

$$\left\langle s_n(p_n), \sum_{i=1}^{n-1} s_i(p_i)(y_i - p_i)\right\rangle (y_n - p_n) \le 0$$

Two Skeptics: Forecaster's Strategy (2)

$$\|\mathcal{K}_n\|^2 = \|\mathcal{K}_{n-1}\|^2 + \|s_n(p_n)\|^2(y_n - p_n)^2$$
$$+2\left\langle s_n(p_n), \textstyle\sum_{i=1}^{n-1} s_i(p_i)(y_i - p_i)\right\rangle (y_n - p_n)$$

By the Neutral Measure Theorem, we can take p_n s.t.

$$\left\langle s_n(p_n), \sum_{i=1}^{n-1} s_i(p_i)(y_i - p_i)\right\rangle (y_n - p_n) \leq 0$$

Thus, $\|\mathcal{K}_n\|^2 \leq \|\mathcal{K}_{n-1}\|^2 + \|s_n(p_n)\|^2$

Finally, $\|\mathcal{K}_n\|^2 \leq n \max_i \max_{p\in[0,1]} \|s_i(p)\|^2$

Two Skeptics: Linear Mixtures

Corollary

The Forecaster strategy above guarantees also that for any $\alpha = (\alpha_1, \alpha_2) \in \mathbb{R}^2$

$$|\alpha_1 \mathcal{K}_n^1 + \alpha_2 \mathcal{K}_n^2| \leq \sqrt{n}\, \|\alpha\| \max_i \max_{p \in [0,1]} \|s_i(p)\|$$

Two Skeptics: Linear Mixtures

Corollary

The Forecaster strategy above guarantees also that for any $\alpha = (\alpha_1, \alpha_2) \in \mathbb{R}^2$

$$|\alpha_1 \mathcal{K}_n^1 + \alpha_2 \mathcal{K}_n^2| \leq \sqrt{n} \, \|\alpha\| \max_i \max_{p \in [0,1]} \|s_i(p)\|$$

Proof.

$$|\alpha_1 \mathcal{K}_n^1 + \alpha_2 \mathcal{K}_n^2| = |\langle \alpha, \mathcal{K}_n \rangle| \leq \|\alpha\| \|\mathcal{K}_n\|$$

□

Skeptic with values in Hilbert Space

$\mathcal{H}$ is any Hilbert space
For $n = 1, 2, \ldots$

Skeptic announces continuous $s_n\colon [0,1] \to \mathcal{H}$
Forecaster announces $p_n \in [0,1]$
Reality announces $y_n \in \{0,1\}$
$\mathcal{K}_n = \mathcal{K}_{n-1} + s_n(p_n)(y_n - p_n)$

Skeptic with values in Hilbert Space

$\mathcal{H}$ is any Hilbert space
For $n = 1, 2, \ldots$

Skeptic announces continuous $s_n \colon [0, 1] \to \mathcal{H}$
Forecaster announces $p_n \in [0, 1]$
Reality announces $y_n \in \{0, 1\}$
$\mathcal{K}_n = \mathcal{K}_{n-1} + s_n(p_n)(y_n - p_n)$

Theorem

Forecaster has a strategy that guarantees for any $\alpha \in \mathcal{H}$

$$\left| \left\langle \alpha, \sum_{i=1}^{n} s_i(p_i)(y_i - p_i) \right\rangle \right| \le \sqrt{n}\, \|\alpha\|_{\mathcal{H}} \max_i \max_{p \in [0,1]} \|s_i(p)\|_{\mathcal{H}}$$

Skeptic with values in Hilbert Space

$\mathcal{H}$ is any Hilbert space, X is a metric compact

For $n = 1, 2, \ldots$

Reality announces $x_n \in X$

Skeptic announces continuous $s_n \colon X \times [0,1] \to \mathcal{H}$

Forecaster announces $p_n \in [0,1]$

Reality announces $y_n \in \{0,1\}$

$\mathcal{K}_n = \mathcal{K}_{n-1} + s_n(x_n, p_n)(y_n - p_n)$

Theorem

Forecaster has a strategy that guarantees for any $\alpha \in \mathcal{H}$

$$\left| \left\langle \alpha, \sum_{i=1}^{n} s_i(x_i, p_i)(y_i - p_i) \right\rangle \right| \leq \sqrt{n} \, \|\alpha\|_{\mathcal{H}} \max_i \max_{(x,p) \in X \times [0,1]} \|s_i(x,p)\|_{\mathcal{H}}$$

Reproducing Kernel Hilbert Spaces

Let $\mathcal{F}$ be a Hilbert Space of functions $X \to \mathbb{R}$

Requirement: $\|f\| \approx 0 \quad \Rightarrow \quad \forall x\, |f(x)| \approx 0$

Reproducing Kernel Hilbert Spaces

Let $\mathcal{F}$ be a Hilbert Space of functions $X \to \mathbb{R}$

Requirement: $\|f\| \approx 0 \quad \Rightarrow \quad \forall x \, |f(x)| \approx 0$

$\mathcal{F}$ is a reproducing kernel Hilbert space on X iff
for any $x \in X$ there exists $\mathbf{k}_x \in \mathcal{F}$ s.t.

$$f(x) = \langle \mathbf{k}_x, f \rangle, \forall f \in \mathcal{F}$$

Reproducing Kernel Hilbert Spaces

Let $\mathcal{F}$ be a Hilbert Space of functions $X \to \mathbb{R}$

Requirement: $\|f\| \approx 0 \quad \Rightarrow \quad \forall x\, |f(x)| \approx 0$

$\mathcal{F}$ is a reproducing kernel Hilbert space on X iff
for any $x \in X$ there exists $\mathbf{k}_x \in \mathcal{F}$ s.t.

$$f(x) = \langle \mathbf{k}_x, f \rangle, \forall f \in \mathcal{F}$$

Example:

$f : [0,1] \to \mathbb{R}$ s.t. $f(0) = 0$

$\langle f, g \rangle = \int_0^1 f'(x) g'(x) dx$

$$\mathbf{k}_x(z) = \begin{cases} z, & z \leq x \\ x, & z > x \end{cases}$$

RKHS-Assessed Probability Forecasts

X is metric compact
For $n = 1, 2, \ldots$

Reality announces $x_n \in X$

Forecaster announces $p_n \in [0, 1]$

Reality announces $y_n \in \{0, 1\}$

RKHS-Assessed Probability Forecasts

X is metric compact
For $n = 1, 2, \ldots$

Reality announces $x_n \in X$
Forecaster announces $p_n \in [0, 1]$
Reality announces $y_n \in \{0, 1\}$

Theorem

Let $\mathcal{F}$ be an RKHS on $X \times [0, 1]$ s.t. $\mathbf{k}_{x,p}$ is continuous in x, p and $C_{\mathcal{F}} = \sup_{x,p} \sup_{\|g\|_{\mathcal{F}}=1} g(x, p)$.
Forecaster has a strategy for $\mathcal{F}$ s.t. for any $f \in \mathcal{F}$

$$\left| \sum_{n=1}^{N} f(x_n, p_n)(y_n - p_n) \right| \leq C_{\mathcal{F}} \|f\|_{\mathcal{F}} \sqrt{N}$$

RKHS-Assessed Forecasts: Proof

Take "Hilbert-valued" Sceptic

$$s_n(x_n, p_n) = \mathbf{k}_{x_n,p_n}$$

Forecaster can guarantee that for any $f \in \mathcal{F}$

$$\left| \sum_{n=1}^{N} \langle f, \mathbf{k}_{x_n,p_n} \rangle (y_n - p_n) \right| \leq \sqrt{N} \|f\|_{\mathcal{F}} \max_{(x,p) \in X \times [0,1]} \|\mathbf{k}_{x,p}\|_{\mathcal{F}}$$

$\langle f, \mathbf{k}_{x_n,p_n} \rangle = f(x_n, p_n)$ for any $f \in \mathcal{F}$.

$\|\mathbf{k}_{x,p}\|_{\mathcal{F}} \leq C_{\mathcal{F}}$

General Asymptotic Calibration

X and Y are compact metric spaces

For $n = 1, 2, \ldots$

Reality announces $x_n \in X$

Forecaster announces $P_n \in \mathcal{P}(Y)$

Reality announces $y_n \in Y$

General Asymptotic Calibration

X and Y are compact metric spaces
For $n = 1, 2, \dots$

Reality announces $x_n \in X$
Forecaster announces $P_n \in \mathcal{P}(Y)$
Reality announces $y_n \in Y$

Theorem

Forecaster has a strategy s.t. for any continuous $f\colon X \times \mathcal{P}(Y) \times Y \to \mathbb{R}$

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \left(f(x_n, P_n, y_n) - \int_Y f(x_n, P_n, y) P_n(dy) \right) = 0$$

Function Approximation

X is metric compact, $B > 0$
For $n = 1, 2, \ldots$

> Reality announces $x_n \in X$
> Forecaster announces $\gamma_n \in [-B, B]$
> Reality announces $y_n \in [-B, B]$

Function Approximation

X is metric compact, $B > 0$
For $n = 1, 2, \ldots$

Reality announces $x_n \in X$
Forecaster announces $\gamma_n \in [-B, B]$
Reality announces $y_n \in [-B, B]$

Theorem

Let $\mathcal{F}$ be an RKHS on X.
Forecaster has a strategy that guarantees for any $f \in \mathcal{F}$

$$\sum_{n=1}^{N}(y_n - \gamma_n)^2 \leq \sum_{n=1}^{N}(y_n - f(x_n))^2 + O(B^2\|f\|_{\mathcal{F}}\sqrt{N})$$

References

Results by Volodya Vovk, some in coauthorship with Ilia Nouretdinov, Glenn Shafer, and Akimichi Takemura.

Recent experimental results by Brian Burford, Fedor Zhdanov.

For more details, see

`http://onlineprediction.net/`

`http://vovk.net/df/index.html`

Outline

Prediction with Expert Advice: Scheme

	Prediction	
Expert 1	π^1	
⋮	⋮	
Expert K	π^K	
Learner		

Prediction with Expert Advice: Scheme

	Prediction	
Expert 1	π^1	
⋮	⋮	
Expert K	π^K	
Learner	π	

Prediction with Expert Advice: Scheme

	Prediction	Loss
Expert 1	π^1	$\lambda(\pi^1, \omega)$
$\vdots$	$\vdots$	$\vdots$
Expert K	π^K	$\lambda(\pi^K, \omega)$
Learner	π	$\lambda(\pi, \omega)$

ω is the outcome

Goal: Learner's loss is not greater than the loss of any Expert

Prediction with Expert Advice: Scheme

	Prediction	Loss
Expert 1	π^1	$\lambda(\pi^1, \omega)$
$\vdots$	$\vdots$	$\vdots$
Expert K	π^K	$\lambda(\pi^K, \omega)$
Learner	π	$\lambda(\pi, \omega)$

ω is the outcome

Goal: Learner's loss is not greater than the loss of any Expert

At each step N, for any k,

$$\sum_{n=1}^{N} \lambda(\pi_n, \omega_n) \leq \sum_{n=1}^{N} \lambda(\pi_n^k, \omega_n) + \text{something small}$$

Prediction with Expert Advice: Protocol

Outcome space Ω, $|\Omega| < \infty$.
Experts $1, 2, \ldots$ (finitely or infinitely many)
Loss function $\lambda\colon \mathcal{P}(\Omega) \times \Omega \to [0, \infty]$

$L_0^k = 0$, $L_0 = 0$
For $n = 1, 2, \ldots$

- Experts announce $\pi_n^k \in \mathcal{P}(\Omega)$.
- Learner announces $\pi_n \in \mathcal{P}(\Omega)$.
- Reality announces $\omega_n \in \Omega$.
- $L_n^k = L_{n-1}^k + \lambda(\pi_n^k, \omega_n), \quad L_n = L_{n-1} + \lambda(\pi_n, \omega_n)$.

One can consider non-probabilistic predictions as well.

Prediction with Expert Advice: Protocol

Outcome space Ω, $|\Omega| < \infty$.
Experts $1, 2, \ldots$ (finitely or infinitely many)
Loss function $\lambda \colon \mathcal{P}(\Omega) \times \Omega \to [0, \infty]$

$L_0^k = 0$, $L_0 = 0$
For $n = 1, 2, \ldots$

> Experts announce $\pi_n^k \in \mathcal{P}(\Omega)$.
>
> Learner announces $\pi_n \in \mathcal{P}(\Omega)$.
>
> Reality announces $\omega_n \in \Omega$.
>
> $L_n^k = L_{n-1}^k + \lambda(\pi_n^k, \omega_n), \qquad L_n = L_{n-1} + \lambda(\pi_n, \omega_n).$

Learner plays against Experts and Reality.
Goal: $L_n \leq L_n^k + Const(k)$, for all k and n.

One can consider non-probabilistic predictions as well.

Prediction with Expert Advice: Bound

Theorem

Let w^k be arbitrary weights of Experts, $w^k \geq 0$, $\sum w^k \leq 1$.
If λ is η-mixable, then Learner has a strategy that guarantees for all n and for all k that

$$L_n \leq L_n^k + \frac{1}{\eta} \ln \frac{1}{w^k} .$$

The bound is in a sense optimal.

λ is η-mixable iff for any $\pi^k \in \mathcal{P}(\Omega)$ and any weights w^k

$$\exists \pi \in \mathcal{P}(\Omega)\ \forall \omega \quad \mathrm{e}^{-\eta\lambda(\pi,\omega)} \geq \sum_{k=1}^{K} w^k \mathrm{e}^{-\eta\lambda(\pi^k,\omega)} .$$

Example: Log Loss

Logarithmic loss:

$$\lambda(\pi, \omega) = \ln \frac{1}{\pi(\omega)}$$

Log loss is 1-mixable

Example: Log Loss

Logarithmic loss:

$$\lambda(\pi, \omega) = \ln \frac{1}{\pi(\omega)}$$

Log loss is 1-mixable

Let $\mu^k(\omega_n|\omega_1 \ldots \omega_{n-1}) = \pi_n^k(\omega_n)$, $M(\omega_n|\omega_1 \ldots \omega_{n-1}) = \pi_n(\omega_n)$.
Then $L_N^k = \sum_{n=1}^N \ln \frac{1}{\pi_n^k(\omega_n)} = \ln \frac{1}{\mu^k(\omega_1 \ldots \omega_N)}$, $L_N = \ln \frac{1}{M(\omega_1 \ldots \omega_N)}$
and the bound of the theorem reduces to

$$\mu^k(\omega_1 \ldots \omega_N) \leq \frac{1}{w^k} M(\omega_1 \ldots \omega_N)$$

Example: Square (Brier) Loss

Square (Brier) loss:

$$\lambda(\pi,\omega) = (1-\pi(\omega))^2 + \sum_{\omega'\neq\omega} \big(\pi(\omega')\big)^2$$

Also 1-mixable

For $\Omega = \{0,1\}$, $\quad \lambda(\pi,\omega) = 2\big(1-\pi(\omega)\big)^2$

$$2\sum_{n=1}^{N}\big(1-M(\omega_n|\omega_1\ldots\omega_{n-1})\big)^2 \leq 2\sum_{n=1}^{N}\big(1-\mu^k(\omega_n|\omega_1\ldots\omega_{n-1})\big)^2 + \ln\frac{1}{w^k}$$

Real Bookmakers

Recent experiments by Fedor Zhdanov.

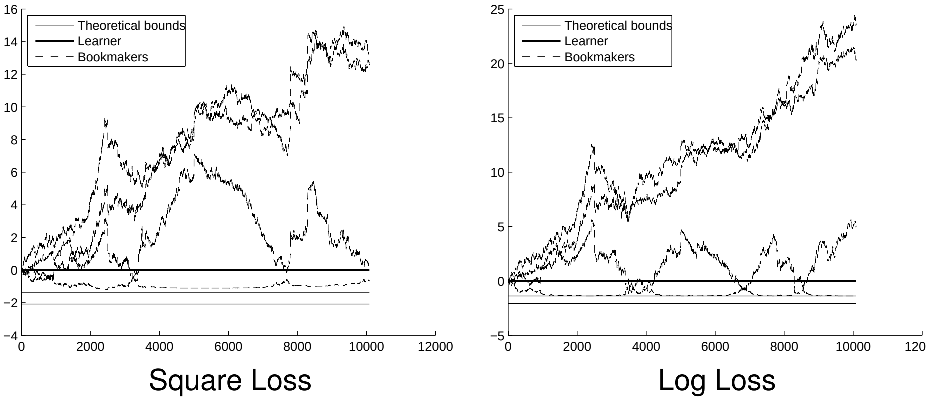
Data:
4 bookmakers, odds for $\sim$ 10000 tennis matches (2 outcomes)
8 bookmakers, odds for $\sim$ 9000 football matches (3 outcomes)

Odds can be transformed to probabilities

There is no obvious choice for the loss function
Use just two popular losses: log loss and square loss

Tennis Odds: “Own” Loss

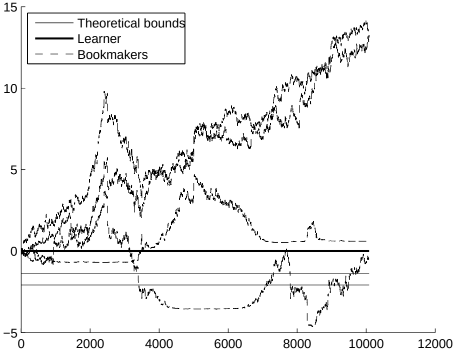
Graphs of the regret $L_n^k - L_n$ ($\geq -\ln 4$, theoretically)



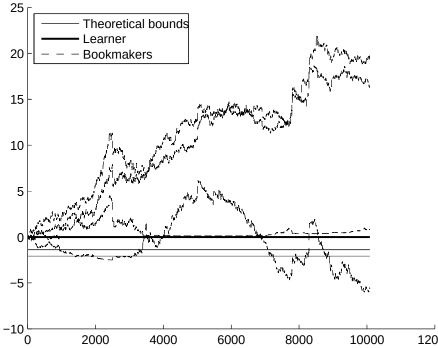
Learner optimizes for the same loss function

Tennis Odds: "Foreign" Loss

Graphs of the regret $L_n^k - L_n$ (no theoretical bound)



Square Loss



Log Loss

Learner optimizes for the other loss function

Multiobjective PEA: Scheme

	Prediction	Loss
Expert 1	π^1	$\lambda^1(\pi^1,\omega),\ldots,\lambda^M(\pi^1,\omega)$
$\vdots$	$\vdots$	$\vdots$
Expert K	π^K	$\lambda^1(\pi^K,\omega),\ldots,\lambda^M(\pi^K,\omega)$
Learner	π	$\lambda^1(\pi,\omega),\ldots,\lambda^M(\pi,\omega)$

$$\sum_{n=1}^{N}\lambda^1(\pi_n,\omega_n)\le\sum_{n=1}^{N}\lambda^1(\pi_n^k,\omega_n)+\text{something small}$$

$$\vdots$$

$$\sum_{n=1}^{N}\lambda^M(\pi_n,\omega_n)\le\sum_{n=1}^{N}\lambda^M(\pi_n^k,\omega_n)+\text{something small}$$

Multiobjective PEA: Protocol

Outcome space Ω, $|\Omega| < \infty$.
Experts $1, 2, 3, \ldots$ (finitely or infinitely many)
Loss functions $\lambda^1 \colon \mathcal{P}(\Omega) \times \Omega \to [0, \infty]$, $\lambda^2 \colon \mathcal{P}(\Omega) \times \Omega \to [0, \infty]$,...

$L_0^{k,m} = 0$, $L_0^m = 0$
For $n = 1, 2, \ldots$

Experts announce $\pi_n^k \in \mathcal{P}(\Omega)$.

Learner announces $\pi_n \in \mathcal{P}(\Omega)$.

Reality announces $\omega_n \in \Omega$.

$L_n^{k,m} = L_{n-1}^{k,m} + \lambda^m(\pi_n^k, \omega_n)$, $\quad L_n^m = L_{n-1}^m + \lambda^m(\pi_n, \omega_n)$.

Learner plays against Experts and Reality.

Goal: $L_n^m \leq L_n^{k,m} + Const(k, m)$, for all k, m and n.

Multiobjective PEA: Bound

Theorem

Let $w^{k,m}$ be arbitrary weights of Experts and loss functions, $w^{k,m} \geq 0$, $\sum w^{k,m} \leq 1$.
If each λ^m is η^m-mixable and strictly proper, then Learner has a strategy that guarantees for all n and for all k and m that

$$L_n^m \leq L_n^{k,m} + \frac{1}{\eta^m} \ln \frac{1}{w^{k,m}} .$$

Multiobjective PEA: Bound

Theorem

Let $w^{k,m}$ be arbitrary weights of Experts and loss functions, $w^{k,m} \geq 0$, $\sum w^{k,m} \leq 1$.
If each λ^m is η^m-mixable and strictly proper, then Learner has a strategy that guarantees for all n and for all k and m that

$$L_n^m \leq L_n^{k,m} + \frac{1}{\eta^m} \ln \frac{1}{w^{k,m}} .$$

Strictly Proper Loss Functions

λ is proper if for any $\pi, \pi' \in \mathcal{P}(\Omega)$, $\pi \neq \pi'$

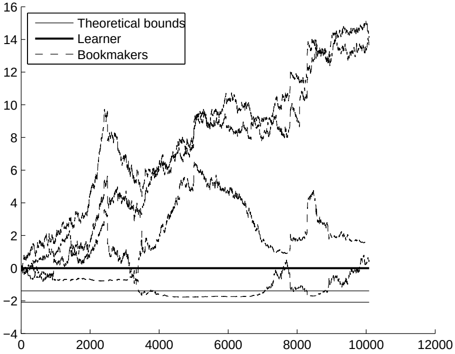
$$\mathbf{E}_\pi \lambda(\pi, \omega) < \mathbf{E}_\pi \lambda(\pi', \omega)\,, \quad \text{where } \mathbf{E}_\pi \lambda(\pi', \omega) = \sum_\omega \pi(\omega)\lambda(\pi, \omega)$$

Motivation:
if $\omega \sim \pi$ then $\mathbf{E}_\pi \lambda(\pi', \omega)$ is the expected loss for prediction π'.

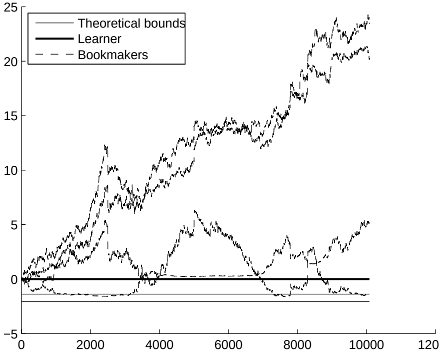
A proper loss function encourages to predict the true probability distribution
$\Rightarrow$ predictions are really (not formally) probabilistic

Tennis Odds: Two Losses

Graphs of the regret $L_n^k - L_n$ ($\geq -\ln 8$, theoretically)



Square Loss



Log Loss

Learner optimizes for both loss functions

Multiobjective PEA: Proof Idea

Lemma

If λ is strictly proper and η-mixable, then $\lambda(\pi, \omega)$ is a continuous function of π and for any $\pi, \pi' \in \mathcal{P}(\Omega)$

$$\mathbf{E}_\pi \frac{e^{\eta\lambda(\pi,\omega)}}{e^{\eta\lambda(\pi',\omega)}} = \sum_\omega \pi(\omega) \frac{e^{\eta\lambda(\pi,\omega)}}{e^{\eta\lambda(\pi',\omega)}} \leq 1\,.$$

Example: for log loss $\lambda(\pi, \omega) = \ln \frac{1}{\pi(\omega)}$, $\eta = 1$,

$$\sum_\omega \pi(\omega) \frac{e^{\lambda(\pi,\omega)}}{e^{\lambda(\pi',\omega)}} = \sum_\omega \pi(\omega) \frac{\frac{1}{\pi(\omega)}}{\frac{1}{\pi'(\omega)}} = \sum_\omega \pi'(\omega) = 1$$

Multiobjective PEA: Algorithm

At step N:

$$f_N(\pi,\omega) = \sum_{k,m} \left(w^{k,m} \prod_{n=1}^{N-1} \frac{\mathrm{e}^{\eta^m \lambda^m(\pi_n,\omega_n)}}{\mathrm{e}^{\eta^m \lambda^m(\pi_n^k,\omega_n)}} \right) \frac{\mathrm{e}^{\eta^m \lambda^m(\pi,\omega)}}{\mathrm{e}^{\eta^m \lambda^m(\pi_N^k,\omega)}}$$

Find $\pi \in \mathcal{P}(\Omega)$ s.t.

$$f_N(\pi,\omega) \leq 1$$

for all ω.
π exists by the Neutral Measure Theorem.

Learner predicts $\pi_N = \pi$.

Multiobjective PEA: Proof of the Bound

For any N,

$$\sum_{k,m}\left(w^{k,m}\prod_{n=1}^{N}\frac{\mathrm{e}^{\eta^m\lambda^m(\pi_n,\omega_n)}}{\mathrm{e}^{\eta^m\lambda^m(\pi_n^k,\omega_n)}}\right)\leq 1\,.$$

Thus,

$$\frac{\mathrm{e}^{\sum_{n=1}^{N}\eta^m\lambda^m(\pi_n,\omega_n)}}{\mathrm{e}^{\sum_{n=1}^{N}\eta^m\lambda^m(\pi_n^k,\omega_n)}}\leq\frac{1}{w^{k,m}}\,.$$

Finally,

$$L_N^m\leq L_N^{k,m}+\ln\frac{1}{w^{k,m}}$$

Possible Future Work

For any measures μ^k, we can (non-constructively) find a (non-computable) measure M s.t.

$$\mu^k(\omega_1 \dots \omega_N) \leq \frac{1}{w^k} M(\omega_1 \dots \omega_N)$$

and

$$\sum_{n=1}^{N} \left(1 - M(\omega_n | \omega_1 \dots \omega_{n-1})\right)^2 \leq \sum_{n=1}^{N} \left(1 - \mu^k(\omega_n | \omega_1 \dots \omega_{n-1})\right)^2 + \frac{1}{2} \ln \frac{1}{w^k}$$

Is it possible to find a semi-enumerable M for all computable μ?

References

For more details, see

http://onlineprediction.net/

A. Chernov, Y. Kalnishkan, F. Zhdanov, V. Vovk. Supermartingales in Prediction with Expert Advice. ALT 2008.

A. Chernov, V. Vovk. Prediction with expert evaluators' advice. http://arxiv.org/abs/0902.4127