Machine learning and portfolio selections. II.

László (Laci) Györfi[1]

[1]Department of Computer Science and Information Theory
Budapest University of Technology and Economics
Budapest, Hungary

September 22, 2007

e-mail: gyorfi@szit.bme.hu
www.szit.bme.hu/˜gyorfi
www.szit.bme.hu/˜oti/portfolio

Dynamic portfolio selection: general case

$\mathbf{x}_i = (x_i^{(1)}, \dots x_i^{(d)})$ the return vector on day i
$\mathbf{b} = \mathbf{b}_1$ is the portfolio vector for the first day
initial capital S_0

$$S_1 = S_0 \cdot \langle \mathbf{b}_1 \, , \, \mathbf{x}_1 \rangle$$

Dynamic portfolio selection: general case

$\mathbf{x}_i = (x_i^{(1)}, \dots x_i^{(d)})$ the return vector on day i
$\mathbf{b} = \mathbf{b}_1$ is the portfolio vector for the first day
initial capital S_0

$$S_1 = S_0 \cdot \langle \mathbf{b}_1 \, , \, \mathbf{x}_1 \rangle$$

for the second day, S_1 new initial capital, the portfolio vector $\mathbf{b}_2 = \mathbf{b}(\mathbf{x}_1)$

$$S_2 = S_0 \cdot \langle \mathbf{b}_1 \, , \, \mathbf{x}_1 \rangle \cdot \langle \mathbf{b}(\mathbf{x}_1) \, , \, \mathbf{x}_2 \rangle \, .$$

Dynamic portfolio selection: general case

$\mathbf{x}_i = (x_i^{(1)}, \dots x_i^{(d)})$ the return vector on day i
$\mathbf{b} = \mathbf{b}_1$ is the portfolio vector for the first day
initial capital S_0

$$S_1 = S_0 \cdot \langle \mathbf{b}_1 \, , \, \mathbf{x}_1 \rangle$$

for the second day, S_1 new initial capital, the portfolio vector $\mathbf{b}_2 = \mathbf{b}(\mathbf{x}_1)$

$$S_2 = S_0 \cdot \langle \mathbf{b}_1 \, , \, \mathbf{x}_1 \rangle \cdot \langle \mathbf{b}(\mathbf{x}_1) \, , \, \mathbf{x}_2 \rangle \, .$$

nth day a portfolio strategy $\mathbf{b}_n = \mathbf{b}(\mathbf{x}_1, \dots, \mathbf{x}_{n-1}) = \mathbf{b}(\mathbf{x}_1^{n-1})$

$$S_n = S_0 \prod_{i=1}^{n} \left\langle \mathbf{b}(\mathbf{x}_1^{i-1}) \, , \, \mathbf{x}_i \right\rangle = S_0 e^{n W_n(\mathbf{B})}$$

with the average growth rate

$$W_n(\mathbf{B}) = \frac{1}{n} \sum_{i=1}^{n} \ln \left\langle \mathbf{b}(\mathbf{x}_1^{i-1}) \, , \, \mathbf{x}_i \right\rangle .$$

log-optimum portfolio

$\mathbf{X}_1, \mathbf{X}_2, \ldots$ drawn from the vector valued stationary and ergodic process

log-optimum portfolio $\mathbf{B}^* = \{\mathbf{b}^*(\cdot)\}$

$$\mathbf{E}\{\ln \left\langle \mathbf{b}^*(\mathbf{X}_1^{n-1}) \, , \, \mathbf{X}_n \right\rangle \mid \mathbf{X}_1^{n-1}\} = \max_{\mathbf{b}(\cdot)} \mathbf{E}\{\ln \left\langle \mathbf{b}(\mathbf{X}_1^{n-1}) \, , \, \mathbf{X}_n \right\rangle \mid \mathbf{X}_1^{n-1}\}$$

$$\mathbf{X}_1^{n-1} = \mathbf{X}_1, \ldots, \mathbf{X}_{n-1}$$

Optimality

Algoet and Cover (1988): If $S_n^* = S_n(\mathbf{B}^*)$ denotes the capital after day n achieved by a log-optimum portfolio strategy $\mathbf{B}^*$, then for any portfolio strategy $\mathbf{B}$ with capital $S_n = S_n(\mathbf{B})$ and for any process $\{\mathbf{X}_n\}_{-\infty}^{\infty}$,

$$\limsup_{n\to\infty}\left(\frac{1}{n}\ln S_n - \frac{1}{n}\ln S_n^*\right) \leq 0 \quad \text{almost surely}$$

Optimality

Algoet and Cover (1988): If $S_n^* = S_n(\mathbf{B}^*)$ denotes the capital after day n achieved by a log-optimum portfolio strategy $\mathbf{B}^*$, then for any portfolio strategy $\mathbf{B}$ with capital $S_n = S_n(\mathbf{B})$ and for any process $\{\mathbf{X}_n\}_{-\infty}^{\infty}$,

$$\limsup_{n\to\infty}\left(\frac{1}{n}\ln S_n - \frac{1}{n}\ln S_n^*\right) \le 0 \quad \text{almost surely}$$

for stationary ergodic process $\{\mathbf{X}_n\}_{-\infty}^{\infty}$,

$$\lim_{n\to\infty}\frac{1}{n}\ln S_n^* = W^* \quad \text{almost surely},$$

where

$$W^* = \mathbf{E}\left\{\max_{\mathbf{b}(\cdot)}\mathbf{E}\{\ln\left\langle \mathbf{b}(\mathbf{X}_{-\infty}^{-1}),\, \mathbf{X}_0\right\rangle \mid \mathbf{X}_{-\infty}^{-1}\}\right\}$$

is the maximal growth rate of any portfolio.

Martingale difference sequences

for the proof of optimality we use the concept of martingale differences:

Definition

there are two sequences of random variables:

$$\{Z_n\} \qquad \{X_n\}$$

- Z_n is a function of $X_1, \ldots, X_n$,
- $\mathbf{E}\{Z_n \mid X_1, \ldots, X_{n-1}\} = 0$ almost surely.

Then $\{Z_n\}$ is called martingale difference sequence with respect to $\{X_n\}$.

A strong law of large numbers

Chow Theorem: If $\{Z_n\}$ is a martingale difference sequence with respect to $\{X_n\}$ and

$$\sum_{n=1}^{\infty} \frac{\mathbf{E}\{Z_n^2\}}{n^2} < \infty$$

then

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} Z_i = 0 \text{ a.s.}$$

A weak law of large numbers

Lemma: If $\{Z_n\}$ is a martingale difference sequence with respect to $\{X_n\}$ then $\{Z_n\}$ are uncorrelated.

A weak law of large numbers

Lemma: If $\{Z_n\}$ is a martingale difference sequence with respect to $\{X_n\}$ then $\{Z_n\}$ are uncorrelated.
Proof. Put $i < j$.

$$\mathbf{E}\{Z_i Z_j\}$$

A weak law of large numbers

Lemma: If $\{Z_n\}$ is a martingale difference sequence with respect to $\{X_n\}$ then $\{Z_n\}$ are uncorrelated.
Proof. Put $i < j$.

$$\mathbf{E}\{Z_i Z_j\} \quad = \quad \mathbf{E}\{\mathbf{E}\{Z_i Z_j \mid X_1, \dots, X_{j-1}\}\}$$

A weak law of large numbers

Lemma: If $\{Z_n\}$ is a martingale difference sequence with respect to $\{X_n\}$ then $\{Z_n\}$ are uncorrelated.
Proof. Put $i < j$.

$$\begin{aligned} \mathbf{E}\{Z_i Z_j\} &= \mathbf{E}\{\mathbf{E}\{Z_i Z_j \mid X_1, \dots, X_{j-1}\}\} \\ &= \mathbf{E}\{Z_i \mathbf{E}\{Z_j \mid X_1, \dots, X_{j-1}\}\} \end{aligned}$$

A weak law of large numbers

Lemma: If $\{Z_n\}$ is a martingale difference sequence with respect to $\{X_n\}$ then $\{Z_n\}$ are uncorrelated.

Proof. Put $i < j$.

$$\begin{aligned}\mathbf{E}\{Z_i Z_j\} &= \mathbf{E}\{\mathbf{E}\{Z_i Z_j \mid X_1, \ldots, X_{j-1}\}\} \\ &= \mathbf{E}\{Z_i \mathbf{E}\{Z_j \mid X_1, \ldots, X_{j-1}\}\} \\ &= \mathbf{E}\{Z_i \cdot 0\}\end{aligned}$$

A weak law of large numbers

Lemma: If $\{Z_n\}$ is a martingale difference sequence with respect to $\{X_n\}$ then $\{Z_n\}$ are uncorrelated.
Proof. Put $i < j$.

$$\begin{aligned}
\mathbf{E}\{Z_i Z_j\} &= \mathbf{E}\{\mathbf{E}\{Z_i Z_j \mid X_1, \dots, X_{j-1}\}\} \\
&= \mathbf{E}\{Z_i \mathbf{E}\{Z_j \mid X_1, \dots, X_{j-1}\}\} \\
&= \mathbf{E}\{Z_i \cdot 0\} = 0
\end{aligned}$$

A weak law of large numbers

Lemma: If $\{Z_n\}$ is a martingale difference sequence with respect to $\{X_n\}$ then $\{Z_n\}$ are uncorrelated.

Proof. Put $i < j$.

$$\begin{aligned} \mathbf{E}\{Z_i Z_j\} &= \mathbf{E}\{\mathbf{E}\{Z_i Z_j \mid X_1, \dots, X_{j-1}\}\} \\ &= \mathbf{E}\{Z_i \mathbf{E}\{Z_j \mid X_1, \dots, X_{j-1}\}\} \\ &= \mathbf{E}\{Z_i \cdot 0\} = 0 \end{aligned}$$

Corollary

$$\mathbf{E}\left\{\left(\frac{1}{n}\sum_{i=1}^{n} Z_i\right)^2\right\}$$

A weak law of large numbers

Lemma: If $\{Z_n\}$ is a martingale difference sequence with respect to $\{X_n\}$ then $\{Z_n\}$ are uncorrelated.

Proof. Put $i < j$.

$$\begin{aligned}
\mathbf{E}\{Z_i Z_j\} &= \mathbf{E}\{\mathbf{E}\{Z_i Z_j \mid X_1, \ldots, X_{j-1}\}\} \\
&= \mathbf{E}\{Z_i \mathbf{E}\{Z_j \mid X_1, \ldots, X_{j-1}\}\} \\
&= \mathbf{E}\{Z_i \cdot 0\} = 0
\end{aligned}$$

Corollary

$$\mathbf{E}\left\{\left(\frac{1}{n}\sum_{i=1}^{n} Z_i\right)^2\right\} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{E}\{Z_i Z_j\}$$

A weak law of large numbers

Lemma: If $\{Z_n\}$ is a martingale difference sequence with respect to $\{X_n\}$ then $\{Z_n\}$ are uncorrelated.
Proof. Put $i < j$.

$$\begin{aligned} \mathbf{E}\{Z_i Z_j\} &= \mathbf{E}\{\mathbf{E}\{Z_i Z_j \mid X_1, \dots, X_{j-1}\}\} \\ &= \mathbf{E}\{Z_i \mathbf{E}\{Z_j \mid X_1, \dots, X_{j-1}\}\} \\ &= \mathbf{E}\{Z_i \cdot 0\} = 0 \end{aligned}$$

Corollary

$$\begin{aligned} \mathbf{E}\left\{\left(\frac{1}{n}\sum_{i=1}^n Z_i\right)^2\right\} &= \frac{1}{n^2}\sum_{i=1}^n \sum_{j=1}^n \mathbf{E}\{Z_i Z_j\} \\ &= \frac{1}{n^2}\sum_{i=1}^n \mathbf{E}\{Z_i^2\} \end{aligned}$$

A weak law of large numbers

Lemma: If $\{Z_n\}$ is a martingale difference sequence with respect to $\{X_n\}$ then $\{Z_n\}$ are uncorrelated.
Proof. Put $i < j$.

$$\begin{aligned} \mathbf{E}\{Z_i Z_j\} &= \mathbf{E}\{\mathbf{E}\{Z_i Z_j \mid X_1, \dots, X_{j-1}\}\} \\ &= \mathbf{E}\{Z_i \mathbf{E}\{Z_j \mid X_1, \dots, X_{j-1}\}\} \\ &= \mathbf{E}\{Z_i \cdot 0\} = 0 \end{aligned}$$

Corollary

$$\begin{aligned} \mathbf{E}\left\{\left(\frac{1}{n}\sum_{i=1}^n Z_i\right)^2\right\} &= \frac{1}{n^2}\sum_{i=1}^n \sum_{j=1}^n \mathbf{E}\{Z_i Z_j\} \\ &= \frac{1}{n^2}\sum_{i=1}^n \mathbf{E}\{Z_i^2\} \\ &\to 0 \end{aligned}$$

if, for example, $\mathbf{E}\{Z_i^2\}$ is a bounded sequence.

Constructing martingale difference sequence

$\{Y_n\}$ is an arbitrary sequence such that Y_n is a function of $X_1, \ldots, X_n$

Constructing martingale difference sequence

$\{Y_n\}$ is an arbitrary sequence such that Y_n is a function of $X_1, \ldots, X_n$

Put

$$Z_n = Y_n - \mathbf{E}\{Y_n \mid X_1, \ldots, X_{n-1}\}$$

Then $\{Z_n\}$ is a martingale difference sequence:

Constructing martingale difference sequence

$\{Y_n\}$ is an arbitrary sequence such that Y_n is a function of $X_1, \ldots, X_n$
Put

$$Z_n = Y_n - \mathbf{E}\{Y_n \mid X_1, \ldots, X_{n-1}\}$$

Then $\{Z_n\}$ is a martingale difference sequence:

- Z_n is a function of $X_1, \ldots, X_n$,

Constructing martingale difference sequence

$\{Y_n\}$ is an arbitrary sequence such that Y_n is a function of $X_1, \ldots, X_n$

Put

$$Z_n = Y_n - \mathbf{E}\{Y_n \mid X_1, \ldots, X_{n-1}\}$$

Then $\{Z_n\}$ is a martingale difference sequence:

- Z_n is a function of $X_1, \ldots, X_n$,
- $$\mathbf{E}\{Z_n \mid X_1, \ldots, X_{n-1}\}$$

Constructing martingale difference sequence

$\{Y_n\}$ is an arbitrary sequence such that Y_n is a function of $X_1, \ldots, X_n$
Put

$$Z_n = Y_n - \mathbf{E}\{Y_n \mid X_1, \ldots, X_{n-1}\}$$

Then $\{Z_n\}$ is a martingale difference sequence:

- Z_n is a function of $X_1, \ldots, X_n$,
- $$\begin{aligned} & \mathbf{E}\{Z_n \mid X_1, \ldots, X_{n-1}\} \\ = \ & \mathbf{E}\{Y_n - \mathbf{E}\{Y_n \mid X_1, \ldots, X_{n-1}\} \mid X_1, \ldots, X_{n-1}\} \\ = \ & 0 \end{aligned}$$

  almost surely.

Optimality

log-optimum portfolio $\mathbf{B}^* = \{\mathbf{b}^*(\cdot)\}$

$$\mathbf{E}\{\ln \left\langle \mathbf{b}^*(\mathbf{X}_1^{n-1}) \, , \, \mathbf{X}_n \right\rangle \mid \mathbf{X}_1^{n-1}\} = \max_{\mathbf{b}(\cdot)} \mathbf{E}\{\ln \left\langle \mathbf{b}(\mathbf{X}_1^{n-1}) \, , \, \mathbf{X}_n \right\rangle \mid \mathbf{X}_1^{n-1}\}$$

Optimality

log-optimum portfolio $\mathbf{B}^* = \{\mathbf{b}^*(\cdot)\}$

$$\mathbf{E}\{\ln \left\langle \mathbf{b}^*(\mathbf{X}_1^{n-1}) \, , \, \mathbf{X}_n \right\rangle \mid \mathbf{X}_1^{n-1}\} = \max_{\mathbf{b}(\cdot)} \mathbf{E}\{\ln \left\langle \mathbf{b}(\mathbf{X}_1^{n-1}) \, , \, \mathbf{X}_n \right\rangle \mid \mathbf{X}_1^{n-1}\}$$

If $S_n^* = S_n(\mathbf{B}^*)$ denotes the capital after day n achieved by a log-optimum portfolio strategy $\mathbf{B}^*$, then for any portfolio strategy $\mathbf{B}$ with capital $S_n = S_n(\mathbf{B})$ and for any process $\{\mathbf{X}_n\}_{-\infty}^{\infty}$,

$$\limsup_{n\to\infty} \left(\frac{1}{n} \ln S_n - \frac{1}{n} \ln S_n^* \right) \le 0 \quad \text{almost surely}$$

Proof of optimality

$$\frac{1}{n}\ln S_n = \frac{1}{n}\sum_{i=1}^{n}\ln\left\langle \mathbf{b}(\mathbf{X}_1^{i-1}), \mathbf{X}_i\right\rangle$$

Proof of optimality

$$
\begin{aligned}
\frac{1}{n}\ln S_n &= \frac{1}{n}\sum_{i=1}^{n}\ln\left\langle \mathbf{b}(\mathbf{X}_1^{i-1}), \mathbf{X}_i\right\rangle \\
&= \frac{1}{n}\sum_{i=1}^{n}\mathbf{E}\{\ln\left\langle \mathbf{b}(\mathbf{X}_1^{i-1}), \mathbf{X}_i\right\rangle \mid \mathbf{X}_1^{i-1}\} \\
&+ \frac{1}{n}\sum_{i=1}^{n}\left(\ln\left\langle \mathbf{b}(\mathbf{X}_1^{i-1}), \mathbf{X}_i\right\rangle - \mathbf{E}\{\ln\left\langle \mathbf{b}(\mathbf{X}_1^{i-1}), \mathbf{X}_i\right\rangle \mid \mathbf{X}_1^{i-1}\}\right)
\end{aligned}
$$

Proof of optimality

$$\frac{1}{n}\ln S_n = \frac{1}{n}\sum_{i=1}^{n}\ln\left\langle \mathbf{b}(\mathbf{X}_1^{i-1}), \mathbf{X}_i\right\rangle$$

$$= \frac{1}{n}\sum_{i=1}^{n}\mathbf{E}\{\ln\left\langle \mathbf{b}(\mathbf{X}_1^{i-1}), \mathbf{X}_i\right\rangle \mid \mathbf{X}_1^{i-1}\}$$

$$+ \frac{1}{n}\sum_{i=1}^{n}\left(\ln\left\langle \mathbf{b}(\mathbf{X}_1^{i-1}), \mathbf{X}_i\right\rangle - \mathbf{E}\{\ln\left\langle \mathbf{b}(\mathbf{X}_1^{i-1}), \mathbf{X}_i\right\rangle \mid \mathbf{X}_1^{i-1}\}\right)$$

and

$$\frac{1}{n}\ln S_n^* = \frac{1}{n}\sum_{i=1}^{n}\mathbf{E}\{\ln\left\langle \mathbf{b}^*(\mathbf{X}_1^{i-1}), \mathbf{X}_i\right\rangle \mid \mathbf{X}_1^{i-1}\}$$

$$+ \frac{1}{n}\sum_{i=1}^{n}\left(\ln\left\langle \mathbf{b}^*(\mathbf{X}_1^{i-1}), \mathbf{X}_i\right\rangle - \mathbf{E}\{\ln\left\langle \mathbf{b}^*(\mathbf{X}_1^{i-1}), \mathbf{X}_i\right\rangle \mid \mathbf{X}_1^{i-1}\}\right)$$

Universally consistent portfolio

These limit relations give rise to the following definition:

Definition

An empirical (data driven) portfolio strategy $\mathbf{B}$ is called **universally consistent with respect to a class $\mathcal{C}$ of stationary and ergodic processes $\{\mathbf{X}_n\}_{-\infty}^{\infty}$**, if for each process in the class,

$$\lim_{n\to\infty} \frac{1}{n} \ln S_n(\mathbf{B}) = W^* \quad \text{almost surely.}$$

Empirical portfolio selection

$$\mathbf{E}\{\ln\left\langle \mathbf{b}^*(\mathbf{X}_1^{n-1}),\, \mathbf{X}_n\right\rangle \mid \mathbf{X}_1^{n-1}\} = \max_{\mathbf{b}(\cdot)} \mathbf{E}\{\ln\left\langle \mathbf{b}(\mathbf{X}_1^{n-1}),\, \mathbf{X}_n\right\rangle \mid \mathbf{X}_1^{n-1}\}$$

Empirical portfolio selection

$$\mathbf{E}\{\ln \left\langle \mathbf{b}^*(\mathbf{X}_1^{n-1}) \, , \, \mathbf{X}_n \right\rangle \mid \mathbf{X}_1^{n-1}\} = \max_{\mathbf{b}(\cdot)} \mathbf{E}\{\ln \left\langle \mathbf{b}(\mathbf{X}_1^{n-1}) \, , \, \mathbf{X}_n \right\rangle \mid \mathbf{X}_1^{n-1}\}$$

fixed integer $k > 0$

$$\mathbf{E}\{\ln \left\langle \mathbf{b}(\mathbf{X}_1^{n-1}) \, , \, \mathbf{X}_n \right\rangle \mid \mathbf{X}_1^{n-1}\} \approx \mathbf{E}\{\ln \left\langle \mathbf{b}(\mathbf{X}_{n-k}^{n-1}) \, , \, \mathbf{X}_n \right\rangle \mid \mathbf{X}_{n-k}^{n-1}\}$$

Empirical portfolio selection

$$\mathbf{E}\{\ln\left\langle \mathbf{b}^*(\mathbf{X}_1^{n-1}),\, \mathbf{X}_n\right\rangle \mid \mathbf{X}_1^{n-1}\} = \max_{\mathbf{b}(\cdot)} \mathbf{E}\{\ln\left\langle \mathbf{b}(\mathbf{X}_1^{n-1}),\, \mathbf{X}_n\right\rangle \mid \mathbf{X}_1^{n-1}\}$$

fixed integer $k > 0$

$$\mathbf{E}\{\ln\left\langle \mathbf{b}(\mathbf{X}_1^{n-1}),\, \mathbf{X}_n\right\rangle \mid \mathbf{X}_1^{n-1}\} \approx \mathbf{E}\{\ln\left\langle \mathbf{b}(\mathbf{X}_{n-k}^{n-1}),\, \mathbf{X}_n\right\rangle \mid \mathbf{X}_{n-k}^{n-1}\}$$

and

$$\mathbf{b}^*(\mathbf{X}_1^{n-1}) \approx \mathbf{b}_k(\mathbf{X}_{n-k}^{n-1}) = \arg\max_{\mathbf{b}(\cdot)} \mathbf{E}\{\ln\left\langle \mathbf{b}(\mathbf{X}_{n-k}^{n-1}),\, \mathbf{X}_n\right\rangle \mid \mathbf{X}_{n-k}^{n-1}\}$$

Empirical portfolio selection

$$\mathbf{E}\{\ln \left\langle \mathbf{b}^*(\mathbf{X}_1^{n-1}) \, , \, \mathbf{X}_n \right\rangle \mid \mathbf{X}_1^{n-1}\} = \max_{\mathbf{b}(\cdot)} \mathbf{E}\{\ln \left\langle \mathbf{b}(\mathbf{X}_1^{n-1}) \, , \, \mathbf{X}_n \right\rangle \mid \mathbf{X}_1^{n-1}\}$$

fixed integer $k > 0$

$$\mathbf{E}\{\ln \left\langle \mathbf{b}(\mathbf{X}_1^{n-1}) \, , \, \mathbf{X}_n \right\rangle \mid \mathbf{X}_1^{n-1}\} \approx \mathbf{E}\{\ln \left\langle \mathbf{b}(\mathbf{X}_{n-k}^{n-1}) \, , \, \mathbf{X}_n \right\rangle \mid \mathbf{X}_{n-k}^{n-1}\}$$

and

$$\mathbf{b}^*(\mathbf{X}_1^{n-1}) \approx \mathbf{b}_k(\mathbf{X}_{n-k}^{n-1}) = \arg\max_{\mathbf{b}(\cdot)} \mathbf{E}\{\ln \left\langle \mathbf{b}(\mathbf{X}_{n-k}^{n-1}) \, , \, \mathbf{X}_n \right\rangle \mid \mathbf{X}_{n-k}^{n-1}\}$$

because of stationarity

$$\begin{aligned} \mathbf{b}_k(\mathbf{x}_1^k) &= \arg\max_{\mathbf{b}(\cdot)} \mathbf{E}\{\ln \left\langle \mathbf{b}(\mathbf{x}_1^k) \, , \, \mathbf{X}_{k+1} \right\rangle \mid \mathbf{X}_1^k = \mathbf{x}_1^k\} \\ &= \arg\max_{\mathbf{b}} \mathbf{E}\{\ln \left\langle \mathbf{b} \, , \, \mathbf{X}_{k+1} \right\rangle \mid \mathbf{X}_1^k = \mathbf{x}_1^k\}, \end{aligned}$$

Empirical portfolio selection

$$\mathbf{E}\{\ln \left\langle \mathbf{b}^*(\mathbf{X}_1^{n-1}) \, , \, \mathbf{X}_n \right\rangle \mid \mathbf{X}_1^{n-1}\} = \max_{\mathbf{b}(\cdot)} \mathbf{E}\{\ln \left\langle \mathbf{b}(\mathbf{X}_1^{n-1}) \, , \, \mathbf{X}_n \right\rangle \mid \mathbf{X}_1^{n-1}\}$$

fixed integer $k > 0$

$$\mathbf{E}\{\ln \left\langle \mathbf{b}(\mathbf{X}_1^{n-1}) \, , \, \mathbf{X}_n \right\rangle \mid \mathbf{X}_1^{n-1}\} \approx \mathbf{E}\{\ln \left\langle \mathbf{b}(\mathbf{X}_{n-k}^{n-1}) \, , \, \mathbf{X}_n \right\rangle \mid \mathbf{X}_{n-k}^{n-1}\}$$

and

$$\mathbf{b}^*(\mathbf{X}_1^{n-1}) \approx \mathbf{b}_k(\mathbf{X}_{n-k}^{n-1}) = \arg\max_{\mathbf{b}(\cdot)} \mathbf{E}\{\ln \left\langle \mathbf{b}(\mathbf{X}_{n-k}^{n-1}) \, , \, \mathbf{X}_n \right\rangle \mid \mathbf{X}_{n-k}^{n-1}\}$$

because of stationarity

$$\begin{aligned} \mathbf{b}_k(\mathbf{x}_1^k) &= \arg\max_{\mathbf{b}(\cdot)} \mathbf{E}\{\ln \left\langle \mathbf{b}(\mathbf{x}_1^k) \, , \, \mathbf{X}_{k+1} \right\rangle \mid \mathbf{X}_1^k = \mathbf{x}_1^k\} \\ &= \arg\max_{\mathbf{b}} \mathbf{E}\{\ln \left\langle \mathbf{b} \, , \, \mathbf{X}_{k+1} \right\rangle \mid \mathbf{X}_1^k = \mathbf{x}_1^k\}, \end{aligned}$$

which is the maximization of the regression function

$$m_{\mathbf{b}}(\mathbf{x}_1^k) = \mathbf{E}\{\ln \left\langle \mathbf{b} \, , \, \mathbf{X}_{k+1} \right\rangle \mid \mathbf{X}_1^k = \mathbf{x}_1^k\}$$

Regression function

Y real valued
X observation vector

Regression function

Y real valued
X observation vector
Regression function

$$m(x) = \mathbf{E}\{Y \mid X = x\}$$

i.i.d. data: $D_n = \{(X_1, Y_1), \dots, (X_n, Y_n)\}$

Regression function

Y real valued
X observation vector
Regression function

$$m(x) = \mathbf{E}\{Y \mid X = x\}$$

i.i.d. data: $D_n = \{(X_1, Y_1), \ldots, (X_n, Y_n)\}$
Regression function estimate

$$m_n(x) = m_n(x, D_n)$$

Regression function

Y real valued
X observation vector
Regression function

$$m(x) = \mathbf{E}\{Y \mid X = x\}$$

i.i.d. data: $D_n = \{(X_1, Y_1), \dots, (X_n, Y_n)\}$
Regression function estimate

$$m_n(x) = m_n(x, D_n)$$

local averaging estimates

$$m_n(x) = \sum_{i=1}^{n} W_{ni}(x; X_1, \dots, X_n) Y_i$$

Regression function

Y real valued
X observation vector
Regression function

$$m(x) = \mathbf{E}\{Y \mid X = x\}$$

i.i.d. data: $D_n = \{(X_1, Y_1), \dots, (X_n, Y_n)\}$
Regression function estimate

$$m_n(x) = m_n(x, D_n)$$

local averaging estimates

$$m_n(x) = \sum_{i=1}^{n} W_{ni}(x; X_1, \dots, X_n) Y_i$$

L. Györfi, M. Kohler, A. Krzyzak, H. Walk (2002) *A Distribution-Free Theory of Nonparametric Regression*, Springer-Verlag, New York.

Correspondence

$$X \quad \sim \quad \mathbf{X}_1^k$$

Correspondence

$$
\begin{aligned}
X &\sim \mathbf{X}_1^k \\
Y &\sim \ln \langle \mathbf{b}\,,\, \mathbf{X}_{k+1} \rangle
\end{aligned}
$$

Correspondence

$$\begin{aligned} X &\sim \mathbf{X}_1^k \\ Y &\sim \ln \langle \mathbf{b}\,,\, \mathbf{X}_{k+1} \rangle \\ m(x) = \mathbf{E}\{Y \mid X = x\} &\sim m_{\mathbf{b}}(\mathbf{x}_1^k) = \mathbf{E}\{\ln \langle \mathbf{b}\,,\, \mathbf{X}_{k+1} \rangle \mid \mathbf{X}_1^k = \mathbf{x}_1^k\} \end{aligned}$$

Partitioning regression estimate

Partition $\mathcal{P}_n = \{A_{n,1}, A_{n,2} \ldots \}$

Partitioning regression estimate

Partition $\mathcal{P}_n = \{A_{n,1}, A_{n,2} \dots \}$
$A_n(x)$ is the cell of the partition $\mathcal{P}_n$ into which x falls

Partitioning regression estimate

Partition $\mathcal{P}_n = \{A_{n,1}, A_{n,2} \dots\}$
$A_n(x)$ is the cell of the partition $\mathcal{P}_n$ into which x falls

$$m_n(x) = \frac{\sum_{i=1}^n Y_i I_{[X_i \in A_n(x)]}}{\sum_{i=1}^n I_{[X_i \in A_n(x)]}}$$

Partitioning regression estimate

Partition $\mathcal{P}_n = \{A_{n,1}, A_{n,2} \dots\}$
$A_n(x)$ is the cell of the partition $\mathcal{P}_n$ into which x falls

$$m_n(x) = \frac{\sum_{i=1}^n Y_i I_{[X_i \in A_n(x)]}}{\sum_{i=1}^n I_{[X_i \in A_n(x)]}}$$

Let G_n be the quantizer corresponding to the partition $\mathcal{P}_n$:
$G_n(x) = j$ if $x \in A_{n,j}$.

Partitioning regression estimate

Partition $\mathcal{P}_n = \{A_{n,1}, A_{n,2} \dots\}$
$A_n(x)$ is the cell of the partition $\mathcal{P}_n$ into which x falls

$$m_n(x) = \frac{\sum_{i=1}^n Y_i I_{[X_i \in A_n(x)]}}{\sum_{i=1}^n I_{[X_i \in A_n(x)]}}$$

Let G_n be the quantizer corresponding to the partition $\mathcal{P}_n$:
$G_n(x) = j$ if $x \in A_{n,j}$.
the set of matches

$$I_n(x) = \{i \le n : \; G_n(x) = G_n(X_i)\}$$

Then

$$m_n(x) = \frac{\sum_{i \in I_n(x)} Y_i}{|I_n(x)|}.$$

Partitioning-based portfolio selection

fix $k, \ell = 1, 2, \ldots$
$\mathcal{P}_\ell = \{A_{\ell,j}, j = 1, 2, \ldots, m_\ell\}$ finite partitions of $\mathbb{R}^d$,

Partitioning-based portfolio selection

fix $k, \ell = 1, 2, \dots$
$\mathcal{P}_\ell = \{A_{\ell,j}, j = 1, 2, \dots, m_\ell\}$ finite partitions of $\mathbb{R}^d$,
G_ℓ be the corresponding quantizer: $G_\ell(\mathbf{x}) = j$, if $\mathbf{x} \in A_{\ell,j}$.

Partitioning-based portfolio selection

fix $k, \ell = 1, 2, \ldots$
$\mathcal{P}_\ell = \{A_{\ell,j}, j = 1, 2, \ldots, m_\ell\}$ finite partitions of $\mathbb{R}^d$,
G_ℓ be the corresponding quantizer: $G_\ell(\mathbf{x}) = j$, if $\mathbf{x} \in A_{\ell,j}$.
$G_\ell(\mathbf{x}_1^n) = G_\ell(\mathbf{x}_1), \ldots, G_\ell(\mathbf{x}_n)$,

Partitioning-based portfolio selection

fix $k, \ell = 1, 2, \ldots$
$\mathcal{P}_\ell = \{A_{\ell,j}, j = 1, 2, \ldots, m_\ell\}$ finite partitions of $\mathbb{R}^d$,
G_ℓ be the corresponding quantizer: $G_\ell(\mathbf{x}) = j$, if $\mathbf{x} \in A_{\ell,j}$.
$G_\ell(\mathbf{x}_1^n) = G_\ell(\mathbf{x}_1), \ldots, G_\ell(\mathbf{x}_n)$,
the set of matches:

$$J_n = \{k < i < n : G_\ell(\mathbf{x}_{i-k}^{i-1}) = G_\ell(\mathbf{x}_{n-k}^{n-1})\}$$

Partitioning-based portfolio selection

fix $k, \ell = 1, 2, \dots$
$\mathcal{P}_\ell = \{A_{\ell,j}, j = 1, 2, \dots, m_\ell\}$ finite partitions of $\mathbb{R}^d$,
G_ℓ be the corresponding quantizer: $G_\ell(\mathbf{x}) = j$, if $\mathbf{x} \in A_{\ell,j}$.
$G_\ell(\mathbf{x}_1^n) = G_\ell(\mathbf{x}_1), \dots, G_\ell(\mathbf{x}_n)$,
the set of matches:

$$J_n = \{k < i < n : G_\ell(\mathbf{x}_{i-k}^{i-1}) = G_\ell(\mathbf{x}_{n-k}^{n-1})\}$$

$$\mathbf{b}^{(k,\ell)}(\mathbf{x}_1^{n-1}) = \arg\max_{\mathbf{b}} \sum_{i \in J_n} \ln \langle \mathbf{b} \, , \, \mathbf{x}_i \rangle$$

if the set I_n is non-void, and $\mathbf{b}_0 = (1/d, \dots, 1/d)$ otherwise.

Elementary portfolios

for fixed $k, \ell = 1, 2, \ldots,$
$\mathbf{B}^{(k,\ell)} = \{\mathbf{b}^{(k,\ell)}(\cdot)\}$, are called elementary portfolios

Elementary portfolios

for fixed $k, \ell = 1, 2, \ldots$,
$\mathbf{B}^{(k,\ell)} = \{\mathbf{b}^{(k,\ell)}(\cdot)\}$, are called elementary portfolios

That is, $\mathbf{b}_n^{(k,\ell)}$ quantizes the sequence $\mathbf{x}_1^{n-1}$ according to the partition $\mathcal{P}_\ell$, and browses through all past appearances of the last seen quantized string $G_\ell(\mathbf{x}_{n-k}^{n-1})$ of length k.
Then it designs a fixed portfolio vector according to the returns on the days following the occurence of the string.

Combining elementary portfolios

How to choose k, ℓ

- small k or small ℓ: large bias
- large k and large ℓ: few matching, large variance

Combining elementary portfolios

How to choose k, ℓ

- small k or small ℓ: large bias
- large k and large ℓ: few matching, large variance

Machine learning: combination of experts

N. Cesa-Bianchi and G. Lugosi, *Prediction, Learning, and Games*. Cambridge University Press, 2006.

Exponential weighing

combine the elementary portfolio strategies $\mathbf{B}^{(k,\ell)} = \{\mathbf{b}_n^{(k,\ell)}\}$

Exponential weighing

combine the elementary portfolio strategies $\mathbf{B}^{(k,\ell)} = \{\mathbf{b}_n^{(k,\ell)}\}$

let $\{q_{k,\ell}\}$ be a probability distribution on the set of all pairs (k,ℓ) such that for all k,ℓ, $q_{k,\ell} > 0$.

Exponential weighing

combine the elementary portfolio strategies $\mathbf{B}^{(k,\ell)} = \{\mathbf{b}_n^{(k,\ell)}\}$

let $\{q_{k,\ell}\}$ be a probability distribution on the set of all pairs (k, ℓ) such that for all k, ℓ, $q_{k,\ell} > 0$.

for $\eta > 0$ put

$$w_{n,k,\ell} = q_{k,\ell} e^{\eta \ln S_{n-1}(\mathbf{B}^{(k,\ell)})}$$

Exponential weighing

combine the elementary portfolio strategies $\mathbf{B}^{(k,\ell)} = \{\mathbf{b}_n^{(k,\ell)}\}$

let $\{q_{k,\ell}\}$ be a probability distribution on the set of all pairs (k,ℓ) such that for all k,ℓ, $q_{k,\ell} > 0$.

for $\eta > 0$ put

$$w_{n,k,\ell} = q_{k,\ell} e^{\eta \ln S_{n-1}(\mathbf{B}^{(k,\ell)})}$$

for $\eta = 1$,

$$w_{n,k,\ell} = q_{k,\ell} e^{\ln S_{n-1}(\mathbf{B}^{(k,\ell)})} = q_{k,\ell} S_{n-1}(\mathbf{B}^{(k,\ell)})$$

Exponential weighing

combine the elementary portfolio strategies $\mathbf{B}^{(k,\ell)} = \{\mathbf{b}_n^{(k,\ell)}\}$
let $\{q_{k,\ell}\}$ be a probability distribution on the set of all pairs (k,ℓ) such that for all k,ℓ, $q_{k,\ell} > 0$.
for $\eta > 0$ put

$$w_{n,k,\ell} = q_{k,\ell} e^{\eta \ln S_{n-1}(\mathbf{B}^{(k,\ell)})}$$

for $\eta = 1$,

$$w_{n,k,\ell} = q_{k,\ell} e^{\ln S_{n-1}(\mathbf{B}^{(k,\ell)})} = q_{k,\ell} S_{n-1}(\mathbf{B}^{(k,\ell)})$$

and

$$v_{n,k,\ell} = \frac{w_{n,k,\ell}}{\sum_{i,j} w_{n,i,j}}.$$

Exponential weighing

combine the elementary portfolio strategies $\mathbf{B}^{(k,\ell)} = \{\mathbf{b}_n^{(k,\ell)}\}$
let $\{q_{k,\ell}\}$ be a probability distribution on the set of all pairs (k, ℓ) such that for all k, ℓ, $q_{k,\ell} > 0$.
for $\eta > 0$ put

$$w_{n,k,\ell} = q_{k,\ell} e^{\eta \ln S_{n-1}(\mathbf{B}^{(k,\ell)})}$$

for $\eta = 1$,

$$w_{n,k,\ell} = q_{k,\ell} e^{\ln S_{n-1}(\mathbf{B}^{(k,\ell)})} = q_{k,\ell} S_{n-1}(\mathbf{B}^{(k,\ell)})$$

and

$$v_{n,k,\ell} = \frac{w_{n,k,\ell}}{\sum_{i,j} w_{n,i,j}}.$$

the combined portfolio $\mathbf{b}$:

$$\mathbf{b}_n(\mathbf{x}_1^{n-1}) = \sum_{k,\ell} v_{n,k,\ell} \mathbf{b}_n^{(k,\ell)}(\mathbf{x}_1^{n-1}).$$

$$S_n(\mathbf{B}) \quad = \quad \prod_{i=1}^{n} \left\langle \mathbf{b}_i(\mathbf{x}_1^{i-1}) \, , \, \mathbf{x}_i \right\rangle$$

$$
\begin{aligned}
S_n(\mathbf{B}) &= \prod_{i=1}^{n} \left\langle \mathbf{b}_i(\mathbf{x}_1^{i-1})\,,\, \mathbf{x}_i \right\rangle \\
&= \prod_{i=1}^{n} \frac{\sum_{k,\ell} w_{i,k,\ell} \left\langle \mathbf{b}_i^{(k,\ell)}(\mathbf{x}_1^{i-1})\,,\, \mathbf{x}_i \right\rangle}{\sum_{k,\ell} w_{i,k,\ell}}
\end{aligned}
$$

$$
\begin{aligned}
S_n(\mathbf{B}) &= \prod_{i=1}^{n} \left\langle \mathbf{b}_i(\mathbf{x}_1^{i-1}) \, , \, \mathbf{x}_i \right\rangle \\
&= \prod_{i=1}^{n} \frac{\sum_{k,\ell} w_{i,k,\ell} \left\langle \mathbf{b}_i^{(k,\ell)}(\mathbf{x}_1^{i-1}) \, , \, \mathbf{x}_i \right\rangle}{\sum_{k,\ell} w_{i,k,\ell}} \\
&= \prod_{i=1}^{n} \frac{\sum_{k,\ell} q_{k,\ell} S_{i-1}(\mathbf{B}^{(k,\ell)}) \left\langle \mathbf{b}_i^{(k,\ell)}(\mathbf{x}_1^{i-1}) \, , \, \mathbf{x}_i \right\rangle}{\sum_{k,\ell} q_{k,\ell} S_{i-1}(\mathbf{B}^{(k,\ell)})}
\end{aligned}
$$

$$
\begin{aligned}
S_n(\mathbf{B}) &= \prod_{i=1}^{n} \left\langle \mathbf{b}_i(\mathbf{x}_1^{i-1}) \,,\, \mathbf{x}_i \right\rangle \\
&= \prod_{i=1}^{n} \frac{\sum_{k,\ell} w_{i,k,\ell} \left\langle \mathbf{b}_i^{(k,\ell)}(\mathbf{x}_1^{i-1}) \,,\, \mathbf{x}_i \right\rangle}{\sum_{k,\ell} w_{i,k,\ell}} \\
&= \prod_{i=1}^{n} \frac{\sum_{k,\ell} q_{k,\ell} S_{i-1}(\mathbf{B}^{(k,\ell)}) \left\langle \mathbf{b}_i^{(k,\ell)}(\mathbf{x}_1^{i-1}) \,,\, \mathbf{x}_i \right\rangle}{\sum_{k,\ell} q_{k,\ell} S_{i-1}(\mathbf{B}^{(k,\ell)})} \\
&= \prod_{i=1}^{n} \frac{\sum_{k,\ell} q_{k,\ell} S_i(\mathbf{B}^{(k,\ell)})}{\sum_{k,\ell} q_{k,\ell} S_{i-1}(\mathbf{B}^{(k,\ell)})}
\end{aligned}
$$

$$
\begin{aligned}
S_n(\mathbf{B}) &= \prod_{i=1}^{n} \left\langle \mathbf{b}_i(\mathbf{x}_1^{i-1})\,,\,\mathbf{x}_i \right\rangle \\
&= \prod_{i=1}^{n} \frac{\sum_{k,\ell} w_{i,k,\ell} \left\langle \mathbf{b}_i^{(k,\ell)}(\mathbf{x}_1^{i-1})\,,\,\mathbf{x}_i \right\rangle}{\sum_{k,\ell} w_{i,k,\ell}} \\
&= \prod_{i=1}^{n} \frac{\sum_{k,\ell} q_{k,\ell} S_{i-1}(\mathbf{B}^{(k,\ell)}) \left\langle \mathbf{b}_i^{(k,\ell)}(\mathbf{x}_1^{i-1})\,,\,\mathbf{x}_i \right\rangle}{\sum_{k,\ell} q_{k,\ell} S_{i-1}(\mathbf{B}^{(k,\ell)})} \\
&= \prod_{i=1}^{n} \frac{\sum_{k,\ell} q_{k,\ell} S_i(\mathbf{B}^{(k,\ell)})}{\sum_{k,\ell} q_{k,\ell} S_{i-1}(\mathbf{B}^{(k,\ell)})} \\
&= \sum_{k,\ell} q_{k,\ell} S_n(\mathbf{B}^{(k,\ell)}),
\end{aligned}
$$

The strategy $\mathbf{B} = \mathbf{B}^H$ then arises from weighing the elementary portfolio strategies $\mathbf{B}^{(k,\ell)} = \{\mathbf{b}_n^{(k,\ell)}\}$ such that the investor's capital becomes

$$S_n(\mathbf{B}) = \sum_{k,\ell} q_{k,\ell} S_n(\mathbf{B}^{(k,\ell)}).$$

Theorem

Assume that

(a) the sequence of partitions is nested, that is, any cell of $\mathcal{P}_{\ell+1}$ is a subset of a cell of $\mathcal{P}_\ell$, $\ell = 1, 2, \ldots$;

(b) if $\mathrm{diam}(A) = \sup_{\mathbf{x},\mathbf{y}\in A} \|\mathbf{x} - \mathbf{y}\|$ denotes the diameter of a set, then for any sphere S centered at the origin

$$\lim_{\ell\to\infty} \max_{j: A_{\ell,j}\cap S\neq\emptyset} \mathrm{diam}(A_{\ell,j}) = 0 .$$

Then the portfolio scheme $\mathbf{B}^H$ defined above is universally consistent with respect to the class of all ergodic processes such that $\mathbf{E}\{|\ln X^{(j)}| < \infty$, for $j = 1, 2, \ldots, d$.

L. Györfi, D. Schäfer (2003) "Nonparametric prediction", in *Advances in Learning Theory: Methods, Models and Applications*, J. A. K. Suykens, G. Horváth, S. Basu, C. Micchelli, J. Vandevalle (Eds.), IOS Press, NATO Science Series, pp. 341-356.
www.szit.bme.hu/˜gyorfi/histog.ps

Proof

We have to prove that

$$\liminf_{n\to\infty} W_n(\mathbf{B}) = \liminf_{n\to\infty} \frac{1}{n} \ln S_n(\mathbf{B}) \geq W^* \quad \text{a.s.}$$

W.l.o.g. we may assume $S_0 = 1$, so that

$$W_n(\mathbf{B}) = \frac{1}{n} \ln S_n(\mathbf{B})$$

Proof

We have to prove that

$$\liminf_{n\to\infty} W_n(\mathbf{B}) = \liminf_{n\to\infty} \frac{1}{n} \ln S_n(\mathbf{B}) \geq W^* \quad \text{a.s.}$$

W.l.o.g. we may assume $S_0 = 1$, so that

$$\begin{aligned} W_n(\mathbf{B}) &= \frac{1}{n} \ln S_n(\mathbf{B}) \\ &= \frac{1}{n} \ln \left(\sum_{k,\ell} q_{k,\ell} S_n(\mathbf{B}^{(k,\ell)}) \right) \end{aligned}$$

Proof

We have to prove that

$$\liminf_{n\to\infty} W_n(\mathbf{B}) = \liminf_{n\to\infty} \frac{1}{n} \ln S_n(\mathbf{B}) \geq W^* \quad \text{a.s.}$$

W.l.o.g. we may assume $S_0 = 1$, so that

$$\begin{aligned} W_n(\mathbf{B}) &= \frac{1}{n} \ln S_n(\mathbf{B}) \\ &= \frac{1}{n} \ln \left(\sum_{k,\ell} q_{k,\ell} S_n(\mathbf{B}^{(k,\ell)}) \right) \\ &\geq \frac{1}{n} \ln \left(\sup_{k,\ell} q_{k,\ell} S_n(\mathbf{B}^{(k,\ell)}) \right) \end{aligned}$$

Proof

We have to prove that

$$\liminf_{n\to\infty} W_n(\mathbf{B}) = \liminf_{n\to\infty} \frac{1}{n} \ln S_n(\mathbf{B}) \geq W^* \quad \text{a.s.}$$

W.l.o.g. we may assume $S_0 = 1$, so that

$$\begin{aligned} W_n(\mathbf{B}) &= \frac{1}{n} \ln S_n(\mathbf{B}) \\ &= \frac{1}{n} \ln \left(\sum_{k,\ell} q_{k,\ell} S_n(\mathbf{B}^{(k,\ell)}) \right) \\ &\geq \frac{1}{n} \ln \left(\sup_{k,\ell} q_{k,\ell} S_n(\mathbf{B}^{(k,\ell)}) \right) \\ &= \frac{1}{n} \sup_{k,\ell} \left(\ln q_{k,\ell} + \ln S_n(\mathbf{B}^{(k,\ell)}) \right) \end{aligned}$$

Proof

We have to prove that

$$\liminf_{n\to\infty} W_n(\mathbf{B}) = \liminf_{n\to\infty} \frac{1}{n} \ln S_n(\mathbf{B}) \geq W^* \quad \text{a.s.}$$

W.l.o.g. we may assume $S_0 = 1$, so that

$$\begin{aligned}
W_n(\mathbf{B}) &= \frac{1}{n} \ln S_n(\mathbf{B}) \\
&= \frac{1}{n} \ln \left(\sum_{k,\ell} q_{k,\ell} S_n(\mathbf{B}^{(k,\ell)}) \right) \\
&\geq \frac{1}{n} \ln \left(\sup_{k,\ell} q_{k,\ell} S_n(\mathbf{B}^{(k,\ell)}) \right) \\
&= \frac{1}{n} \sup_{k,\ell} \left(\ln q_{k,\ell} + \ln S_n(\mathbf{B}^{(k,\ell)}) \right) \\
&= \sup_{k,\ell} \left(W_n(\mathbf{B}^{(k,\ell)}) + \frac{\ln q_{k,\ell}}{n} \right).
\end{aligned}$$

Thus

$$\liminf_{n\to\infty} W_n(\mathbf{B}) \quad \geq \quad \liminf_{n\to\infty} \sup_{k,\ell} \left(W_n(\mathbf{B}^{(k,\ell)}) + \frac{\ln q_{k,\ell}}{n} \right)$$

Thus

$$\begin{aligned}\liminf_{n\to\infty} W_n(\mathbf{B}) &\geq \liminf_{n\to\infty} \sup_{k,\ell} \left(W_n(\mathbf{B}^{(k,\ell)}) + \frac{\ln q_{k,\ell}}{n} \right) \\ &\geq \sup_{k,\ell} \liminf_{n\to\infty} \left(W_n(\mathbf{B}^{(k,\ell)}) + \frac{\ln q_{k,\ell}}{n} \right)\end{aligned}$$

Thus

$$
\begin{aligned}
\liminf_{n\to\infty} W_n(\mathbf{B}) &\geq \liminf_{n\to\infty} \sup_{k,\ell} \left(W_n(\mathbf{B}^{(k,\ell)}) + \frac{\ln q_{k,\ell}}{n} \right) \\
&\geq \sup_{k,\ell} \liminf_{n\to\infty} \left(W_n(\mathbf{B}^{(k,\ell)}) + \frac{\ln q_{k,\ell}}{n} \right) \\
&= \sup_{k,\ell} \liminf_{n\to\infty} W_n(\mathbf{B}^{(k,\ell)})
\end{aligned}
$$

Thus

$$\begin{aligned}
\liminf_{n\to\infty} W_n(\mathbf{B}) &\geq \liminf_{n\to\infty} \sup_{k,\ell} \left(W_n(\mathbf{B}^{(k,\ell)}) + \frac{\ln q_{k,\ell}}{n} \right) \\
&\geq \sup_{k,\ell} \liminf_{n\to\infty} \left(W_n(\mathbf{B}^{(k,\ell)}) + \frac{\ln q_{k,\ell}}{n} \right) \\
&= \sup_{k,\ell} \liminf_{n\to\infty} W_n(\mathbf{B}^{(k,\ell)}) \\
&= \sup_{k,\ell} \epsilon_{k,\ell}
\end{aligned}$$

Since the partitions $\mathcal{P}_\ell$ are nested, we have that

$$\sup_{k,\ell} \epsilon_{k,\ell} = \lim_{k\to\infty} \lim_{l\to\infty} \epsilon_{k,\ell} = W^*.$$

Kernel regression estimate

Kernel function $K(x) \geq 0$
Bandwidth $h > 0$

$$m_n(x) = \frac{\sum_{i=1}^{n} Y_i K\left(\frac{x-X_i}{h}\right)}{\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)}$$

Naive (window) kernel function $K(x) = I_{\{\|x\| \leq 1\}}$

$$m_n(x) = \frac{\sum_{i=1}^{n} Y_i I_{\{\|x-X_i\| \leq h\}}}{\sum_{i=1}^{n} I_{\{\|x-X_i\| \leq h\}}}$$

Kernel-based portfolio selection

choose the radius $r_{k,\ell} > 0$ such that for any fixed k,

$$\lim_{\ell \to \infty} r_{k,\ell} = 0.$$

Kernel-based portfolio selection

choose the radius $r_{k,\ell} > 0$ such that for any fixed k,

$$\lim_{\ell \to \infty} r_{k,\ell} = 0.$$

for $n > k + 1$, define the expert $\mathbf{b}^{(k,\ell)}$ by

$$\mathbf{b}^{(k,\ell)}(\mathbf{x}_1^{n-1}) = \arg\max_{\mathbf{b}} \sum_{\left\{k<i<n:\|\mathbf{x}_{i-k}^{i-1}-\mathbf{x}_{n-k}^{n-1}\|\le r_{k,\ell}\right\}} \ln \langle \mathbf{b} \, , \, \mathbf{x}_i \rangle \, ,$$

if the sum is non-void, and $\mathbf{b}_0 = (1/d, \dots, 1/d)$ otherwise.

Theorem

The kernel-based portfolio scheme is universally consistent with respect to the class of all ergodic processes such that $\mathbf{E}\{|\ln X^{(j)}| < \infty$, for $j = 1, 2, \ldots, d$.

L. Györfi, G. Lugosi, F. Udina (2006) "Nonparametric kernel-based sequential investment strategies", *Mathematical Finance*, 16, pp. 337-357
www.szit.bme.hu/~gyorfi/kernel.pdf

k-nearest neighbor (NN) regression estimate

$$m_n(x) = \sum_{i=1}^{n} W_{ni}(x; X_1, \ldots, X_n) Y_i.$$

W_{ni} is $1/k$ if X_i is one of the k nearest neighbors of x among $X_1, \ldots, X_n$, and W_{ni} is 0 otherwise.

Nearest-neighbor-based portfolio selection

choose $p_\ell \in (0,1)$ such that $\lim_{\ell\to\infty} p_\ell = 0$
for fixed positive integers k,ℓ ($n > k + \hat{\ell} + 1$) introduce the set of the $\hat{\ell} = \lfloor p_\ell n \rfloor$ nearest neighbor matches:

$$\hat{J}_n^{(k,\ell)} = \{i;\ k+1 \le i \le n \text{ such that } \mathbf{x}_{i-k}^{i-1} \text{ is among the } \hat{\ell} \text{ NNs of } \mathbf{x}_{n-k}^{n-1} \text{ in } \mathbf{x}_1^k, \dots, \mathbf{x}_{n-k}^{n-1}\}.$$

Nearest-neighbor-based portfolio selection

choose $p_\ell \in (0,1)$ such that $\lim_{\ell\to\infty} p_\ell = 0$
for fixed positive integers k,ℓ ($n > k + \hat{\ell} + 1$) introduce the set of the $\hat{\ell} = \lfloor p_\ell n \rfloor$ nearest neighbor matches:

$$\hat{J}_n^{(k,\ell)} = \{i;\ k+1 \le i \le n \text{ such that } \mathbf{x}_{i-k}^{i-1} \text{ is among the } \hat{\ell} \text{ NNs of } \mathbf{x}_{n-k}^{n-1} \text{ in } \mathbf{x}_1^k, \dots, \mathbf{x}_{n-k}^{n-1}\}.$$

Define the portfolio vector by

$$\mathbf{b}^{(k,\ell)}(\mathbf{x}_1^{n-1}) = \arg\max_{\mathbf{b}} \sum_{\left\{i \in \hat{J}_n^{(k,\ell)}\right\}} \ln \langle \mathbf{b}\,,\, \mathbf{x}_i \rangle$$

if the sum is non-void, and $\mathbf{b}_0 = (1/d, \dots, 1/d)$ otherwise.

Theorem

If for any vector $\mathbf{s} = \mathbf{s}_1^k$ the random variable

$$\|\mathbf{X}_1^k - \mathbf{s}\|$$

has continuous distribution function, then the nearest-neighbor portfolio scheme is universally consistent with respect to the class of all ergodic processes such that $\mathbf{E}\{|\ln X^{(j)}|\} < \infty$, for $j = 1, 2, \ldots d$.

NN is robust, there is no scaling problem

Theorem

If for any vector $\mathbf{s} = \mathbf{s}_1^k$ the random variable

$$\|\mathbf{X}_1^k - \mathbf{s}\|$$

has continuous distribution function, then the nearest-neighbor portfolio scheme is universally consistent with respect to the class of all ergodic processes such that $\mathbf{E}\{|\ln X^{(j)}|\} < \infty$, for $j = 1, 2, \ldots d$.

NN is robust, there is no scaling problem

L. Györfi, F. Udina, H. Walk (2006) "Nonparametric nearest neighbor based empirical portfolio selection strategies", (submitted), www.szit.bme.hu/~gyorfi/NN.pdf

Semi-log-optimal portfolio

empirical log-optimal:

$$h^{(k,\ell)}(\mathbf{x}_1^{n-1}) = \arg\max_{\mathbf{b}} \sum_{i \in J_n} \ln \langle \mathbf{b} \, , \, \mathbf{x}_i \rangle$$

Semi-log-optimal portfolio

empirical log-optimal:

$$h^{(k,\ell)}(\mathbf{x}_1^{n-1}) = \arg\max_{\mathbf{b}} \sum_{i \in J_n} \ln \langle \mathbf{b} \, , \, \mathbf{x}_i \rangle$$

Taylor expansion: $\ln z \approx h(z) = z - 1 - \frac{1}{2}(z-1)^2$

Semi-log-optimal portfolio

empirical log-optimal:

$$h^{(k,\ell)}(\mathbf{x}_1^{n-1}) = \arg\max_{\mathbf{b}} \sum_{i \in J_n} \ln \langle \mathbf{b} , \mathbf{x}_i \rangle$$

Taylor expansion: $\ln z \approx h(z) = z - 1 - \frac{1}{2}(z-1)^2$ empirical semi-log-optimal:

$$\tilde{h}^{(k,\ell)}(\mathbf{x}_1^{n-1}) = \arg\max_{\mathbf{b}} \sum_{i \in J_n} h(\langle \mathbf{b} , \mathbf{x}_i \rangle) = \arg\max_{\mathbf{b}} \{\langle \mathbf{b} , \mathbf{m} \rangle - \langle \mathbf{b} , \mathbf{C}\mathbf{b} \rangle\}$$

Semi-log-optimal portfolio

empirical log-optimal:

$$h^{(k,\ell)}(\mathbf{x}_1^{n-1}) = \arg\max_{\mathbf{b}} \sum_{i \in J_n} \ln \langle \mathbf{b}\,,\, \mathbf{x}_i \rangle$$

Taylor expansion: $\ln z \approx h(z) = z - 1 - \frac{1}{2}(z-1)^2$ empirical semi-log-optimal:

$$\tilde{h}^{(k,\ell)}(\mathbf{x}_1^{n-1}) = \arg\max_{\mathbf{b}} \sum_{i \in J_n} h(\langle \mathbf{b}\,,\, \mathbf{x}_i \rangle) = \arg\max_{\mathbf{b}} \{\langle \mathbf{b}\,,\, \mathbf{m} \rangle - \langle \mathbf{b}\,,\, \mathbf{C}\mathbf{b} \rangle\}$$

smaller computational complexity: quadratic programming

Semi-log-optimal portfolio

empirical log-optimal:

$$h^{(k,\ell)}(\mathbf{x}_1^{n-1}) = \arg\max_{\mathbf{b}} \sum_{i \in J_n} \ln \langle \mathbf{b}\,,\, \mathbf{x}_i \rangle$$

Taylor expansion: $\ln z \approx h(z) = z - 1 - \frac{1}{2}(z-1)^2$ empirical semi-log-optimal:

$$\tilde{h}^{(k,\ell)}(\mathbf{x}_1^{n-1}) = \arg\max_{\mathbf{b}} \sum_{i \in J_n} h(\langle \mathbf{b}\,,\, \mathbf{x}_i \rangle) = \arg\max_{\mathbf{b}} \{\langle \mathbf{b}\,,\, \mathbf{m} \rangle - \langle \mathbf{b}\,,\, \mathbf{C}\mathbf{b} \rangle\}$$

smaller computational complexity: quadratic programming

L. Györfi, A. Urbán, I. Vajda (2007) "Kernel-based semi-log-optimal portfolio selection strategies", *International Journal of Theoretical and Applied Finance*, 10, pp. 505-516. www.szit.bme.hu/~gyorfi/semi.pdf

Conditions of the model:

Assume that

- the assets are arbitrarily divisible,
- the assets are available in unbounded quantities at the current price at any given trading period,
- there are no transaction costs,
- the behavior of the market is not affected by the actions of the investor using the strategy under investigation.

NYSE data sets

At www.szit.bme.hu/~oti/portfolio there are two benchmark data set from NYSE:

- The first data set consists of daily data of 36 stocks with length 22 years.
- The second data set contains 23 stocks and has length 44 years.

NYSE data sets

At www.szit.bme.hu/~oti/portfolio there are two benchmark data set from NYSE:

- The first data set consists of daily data of 36 stocks with length 22 years.
- The second data set contains 23 stocks and has length 44 years.

Our experiment is on the second data set.

Experiments on average annual yields (AAY)

Kernel based semi-log-optimal portfolio selection with $k = 1, \ldots, 5$ and $l = 1, \ldots, 10$

$$r_{k,l}^2 = 0.0001 \cdot d \cdot k \cdot \ell,$$

Experiments on average annual yields (AAY)

Kernel based semi-log-optimal portfolio selection with $k = 1, \ldots, 5$ and $l = 1, \ldots, 10$

$$r_{k,l}^2 = 0.0001 \cdot d \cdot k \cdot \ell,$$

AAY of kernel based semi-log-optimal portfolio is 128%

Experiments on average annual yields (AAY)

Kernel based semi-log-optimal portfolio selection with $k = 1, \ldots, 5$ and $l = 1, \ldots, 10$

$$r_{k,l}^2 = 0.0001 \cdot d \cdot k \cdot \ell,$$

AAY of kernel based semi-log-optimal portfolio is 128%

double the capital

Experiments on average annual yields (AAY)

Kernel based semi-log-optimal portfolio selection with $k = 1, \ldots, 5$ and $l = 1, \ldots, 10$

$$r_{k,l}^2 = 0.0001 \cdot d \cdot k \cdot \ell,$$

AAY of kernel based semi-log-optimal portfolio is 128%
double the capital
MORRIS had the best AAY, 20%

Experiments on average annual yields (AAY)

Kernel based semi-log-optimal portfolio selection with $k = 1, \dots, 5$ and $l = 1, \dots, 10$

$$r_{k,l}^2 = 0.0001 \cdot d \cdot k \cdot \ell,$$

AAY of kernel based semi-log-optimal portfolio is 128%
double the capital
MORRIS had the best AAY, 20%
the BCRP had average AAY 24%

The average annual yields of the individual experts.

k / ℓ	1	2	3	4	5
1	20%	19%	16%	16%	16%
2	118%	77%	62%	24%	58%
3	71%	41%	26%	58%	21%
4	103%	94%	63%	97%	34%
5	134%	102%	100%	102%	67%
6	140%	125%	105%	108%	87%
7	148%	123%	107%	99%	96%
8	132%	112%	102%	85%	81%
9	127%	103%	98%	74%	72%
10	123%	92%	81%	65%	69%