Machine learning and portfolio selections.

László Györfi
Department of Computer Science and Information Theory
Budapest University of Technology and Economics
gyorfi@szit.bme.hu

Abstract

In this paper the static portfolio and the dynamic log-optimal portfolio for memoryless and for stationary market processes is studied. We illustrate their performance (average annual yield) for NYSE data.

1 Notations.

Consider a market consisting of d assets. The evolution of the market in time is represented by a sequence of price vectors $\mathbf{s}_1, \mathbf{s}_2, \ldots \in \mathbb{R}_+^d$, where

$$\mathbf{s}_n = (s_n^{(1)}, \ldots, s_n^{(d)})$$

such that the j-th component $s_n^{(j)}$ of $\mathbf{s}_n$ denotes the price of the j-th asset on the n-th trading period. In order to normalize, put $s_0^{(j)} = 1$. $\{\mathbf{s}_n\}$ has exponential trend:

$$s_n^{(j)} = e^{nW_n^{(j)}} \approx e^{nW^{(j)}},$$

with average growth rate (average yield)

$$W_n^{(j)} := \frac{1}{n} \ln s_n^{(j)}$$

and with asymptotic average growth rate

$$W^{(j)} := \lim_{n\to\infty} \frac{1}{n} \ln s_n^{(j)}.$$

In order to apply the usual prediction techniques for time series analysis one has to transform the sequence price vectors $\{\mathbf{s}_n\}$ into a more or less stationary sequence of return vectors $\{\mathbf{x}_n\}$ as follows:

$$\mathbf{x}_n = (x_n^{(1)}, \ldots, x_n^{(d)})$$

such that

$$x_n^{(j)} = \frac{s_n^{(j)}}{s_{n-1}^{(j)}}.$$

Thus, the j-th component $x_n^{(j)}$ of the return vector $\mathbf{x}_n$ denotes the amount obtained after investing a unit capital in the j-th asset on the n-th trading period.

2 Static portfolio.

The static portfolio is a single period investment strategy. A portfolio vector is denoted by $\mathbf{b} = (b^{(1)}, \ldots b^{(d)})$. The j-th component $b^{(j)}$ of $\mathbf{b}$ denotes the proportion of the investor's capital invested in asset j. We assume that the portfolio vector $\mathbf{b}$ has nonnegative components sum up to 1. The set of portfolio vectors is denoted by

$$\Delta_d = \left\{ \mathbf{b} = (b^{(1)}, \ldots, b^{(d)});\ b^{(j)} \geq 0,\ \sum_{j=1}^{d} b^{(j)} = 1 \right\}.$$

The aim of static portfolio is to achieve $\max_{1\leq j\leq d} W^{(j)}$. For static portfolio selection, at time $n = 0$ we distribute the initial capital S_0 according to a fix portfolio vector $\mathbf{b}$, i.e., if S_n denotes the wealth at the trading period n, then

$$S_n = S_0 \sum_{j=1}^{d} b^{(j)} s_n^{(j)}.$$

Apply the following simple bounds

$$S_0 \max_j b^{(j)} s_n^{(j)} \leq S_n \leq d S_0 \max_j b^{(j)} s_n^{(j)}.$$

If $b^{(j)} > 0$ for all $j = 1, \ldots, d$ then these bounds imply that

$$W := \lim_{n\to\infty} \frac{1}{n} \ln S_n = \lim_{n\to\infty} \max_j \frac{1}{n} \ln s_n^{(j)} = \max_j W^{(j)}.$$

Thus, any static portfolio achieves the maximal growth rate $\max_j W^{(j)}$.

3 Dynamic portfolio selection: special case

One can achieve even higher growth rate for long run investments. The constantly rebalanced portfolio (CRP) is a special dynamic portfolio selection. The dynamic portfolio selection is a multi-period investment strategy, where at the beginning of each trading period we rearrange the wealth among the assets. The concept of CRP was introduced and studied by Kelly [24], Latané [25], Breiman [7], Markowitz [28], Finkelstein and Whitley [13], Móri [30], Móri and Székely [33] and Barron and Cover [4]. See also Chapters 6 and 15 in Cover and Thomas [12], and Chapter 15 in Luenberger [26].

Fix a portfolio vector $\mathbf{b} \in \Delta_d$, i.e., we are concerned with a hypothetical investor who neither consumes nor deposits new cash into his portfolio, but reinvest his portfolio each trading period.

Let S_0 denote the investor's initial capital. Then at the beginning of the first trading period $S_0 b^{(j)}$ is invested into asset j, and it results in return $S_0 b^{(j)} x_1^{(j)}$, therefore at the end of the first trading period the investor's wealth becomes

$$S_1 = S_0 \sum_{j=1}^{d} b^{(j)} x_1^{(j)} = S_0 \langle \mathbf{b}\,,\, \mathbf{x}_1 \rangle\,,$$

where $\langle \cdot\,,\, \cdot \rangle$ denotes inner product. For the second trading period, S_1 is the new initial capital

$$S_2 = S_1 \cdot \langle \mathbf{b}\,,\, \mathbf{x}_2 \rangle = S_0 \cdot \langle \mathbf{b}\,,\, \mathbf{x}_1 \rangle \cdot \langle \mathbf{b}\,,\, \mathbf{x}_2 \rangle\,.$$

By induction, for the trading period n the initial capital is S_{n-1}, therefore

$$S_n = S_{n-1} \langle \mathbf{b}\,,\, \mathbf{x}_n \rangle = S_0 \prod_{i=1}^{n} \langle \mathbf{b}\,,\, \mathbf{x}_i \rangle\,.$$

The asymptotic average growth rate of this portfolio selection is

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}\ln S_n &= \lim_{n\to\infty}\left(\frac{1}{n}\ln S_0+\frac{1}{n}\sum_{i=1}^{n}\ln\langle \mathbf{b}\,,\,\mathbf{x}_i\rangle\right)\\ &= \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\ln\langle \mathbf{b}\,,\,\mathbf{x}_i\rangle\,,\end{aligned}$$

therefore without loss of generality one can assume in the sequel that the initial capital $S_0 = 1$.

4 Log-optimal portfolio for memoryless market process.

If the market process $\{\mathbf{X}_i\}$ is memoryless, i.e., it is a sequence of independent and identically distributed (i.i.d.) random return vectors then we show that the best constantly rebalanced portfolio (BCRP) is the log-optimal portfolio:

$$\mathbf{b}^* := \arg\max_{\mathbf{b}} \mathbb{E}\{\ln\langle \mathbf{b}\,,\,\mathbf{X}_1\rangle\}.$$

This optimality means that 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 finite $\mathbb{E}\{(\ln\langle \mathbf{b}\,,\,\mathbf{X}_1\rangle)^2\}$ and with capital $S_n = S_n(\mathbf{b})$ and for any memoryless market process $\{\mathbf{X}_n\}_{-\infty}^{\infty}$,

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

and maximal asymptotic average growth rate is

$$\lim_{n\to\infty}\frac{1}{n}\ln S_n^* = W^* := \mathbb{E}\{\ln\langle \mathbf{b}^*\,,\,\mathbf{X}_1\rangle\} \quad \text{almost surely.}$$

The proof of the optimality is a simple consequence of the strong law of large numbers. Introduce the notation

$$W(\mathbf{b}) = \mathbb{E}\{\ln\langle \mathbf{b}\,,\,\mathbf{X}_1\rangle\}.$$

Then

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

The strong law of large numbers implies that

$$
\frac{1}{n}\sum_{i=1}^{n}\left(\ln\langle \mathbf{b}\,,\,\mathbf{X}_i\rangle - \mathbb{E}\{\ln\langle \mathbf{b}\,,\,\mathbf{X}_i\rangle\}\right) \to 0 \quad \text{almost surely,}
$$

therefore

$$
\lim_{n\to\infty}\frac{1}{n}\ln S_n = W(\mathbf{b}) = \mathbb{E}\{\ln\langle \mathbf{b}\,,\,\mathbf{X}_1\rangle\} \quad \text{almost surely.}
$$

Similarly,

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

We have to emphasize the basic conditions of the model: assume that

(i) the assets are arbitrarily divisible, and they are available for buying or for selling in unbounded quantities at the current price at any given trading period,

(ii) there are no transaction costs,

(iii) the behavior of the market is not affected by the actions of the investor using the strategy under investigation.

Concerning condition (ii), Blum and Kalai [5], Iyengar and Cover [23], Iyengar [21], [22], Bobryk and Stettner [6], and Schäfer [36] considered the problem of transaction cost. For memoryless or Markovian market process, optimal strategies have been introduced if the distributions of the market process are known. There is no asymptotically optimal, empirical algorithm taking into account the transaction cost. Condition (iii) means that the market is inefficient.

The principle of log-optimality has the important consequence that

$$S_n(\mathbf{b}) \quad \text{is not close to} \quad \mathbb{E}\{S_n(\mathbf{b})\}.$$

We prove a bit more. The optimality property proved above means that, for any $\delta > 0$, the event

$$\left\{-\delta < \frac{1}{n}\ln S_n(\mathbf{b}) - \mathbb{E}\{\ln\langle \mathbf{b}\,,\,\mathbf{X}_1\rangle\} < \delta\right\}$$

has probability close to 1 if n is large enough. On the one hand, the i.i.d. property implies that

$$\begin{aligned}
& \left\{-\delta < \frac{1}{n}\ln S_n(\mathbf{b}) - \mathbb{E}\{\ln\langle \mathbf{b}\,,\,\mathbf{X}_1\rangle\} < \delta\right\} \\
= & \left\{-\delta + \mathbb{E}\{\ln\langle \mathbf{b}\,,\,\mathbf{X}_1\rangle\} < \frac{1}{n}\ln S_n(\mathbf{b}) < \delta + \mathbb{E}\{\ln\langle \mathbf{b}\,,\,\mathbf{X}_1\rangle\}\right\} \\
= & \left\{e^{n(-\delta+\mathbb{E}\{\ln\langle \mathbf{b}\,,\,\mathbf{X}_1\rangle\})} < S_n(\mathbf{b}) < e^{n(\delta+\mathbb{E}\{\ln\langle \mathbf{b}\,,\,\mathbf{X}_1\rangle\})}\right\},
\end{aligned}$$

therefore

$$S_n(\mathbf{b}) \quad \text{is close to} \quad e^{n\mathbb{E}\{\ln\langle \mathbf{b}\,,\,\mathbf{X}_1\rangle\}}.$$

On the other hand,

$$\mathbb{E}\{S_n(\mathbf{b})\} = \mathbb{E}\left\{\prod_{i=1}^{n}\langle \mathbf{b}\,,\,\mathbf{X}_i\rangle\right\} = \prod_{i=1}^{n}\langle \mathbf{b}\,,\,\mathbb{E}\{\mathbf{X}_i\}\rangle = e^{n\ln\langle \mathbf{b}\,,\,\mathbb{E}\{\mathbf{X}_1\}\rangle}.$$

By Jensen inequality,

$$\ln\langle \mathbf{b}\,,\,\mathbb{E}\{\mathbf{X}_1\}\rangle > \mathbb{E}\{\ln\langle \mathbf{b}\,,\,\mathbf{X}_1\rangle\},$$

therefore

$$S_n(\mathbf{b}) \quad \text{is much less than} \quad \mathbb{E}\{S_n(\mathbf{b})\}.$$

Not knowing this fact, one can apply a naive approach

$$\arg\max_{\mathbf{b}} \mathbb{E}\{S_n(\mathbf{b})\}.$$

Because of

$$\mathbb{E}\{S_n(\mathbf{b})\} = \langle \mathbf{b}\,,\, \mathbb{E}\{\mathbf{X}_1\}\rangle^n\,,$$

this naive approach has the equivalent form

$$\arg\max_{\mathbf{b}} \mathbb{E}\{S_n(\mathbf{b})\} = \arg\max_{\mathbf{b}} \langle \mathbf{b}\,,\, \mathbb{E}\{\mathbf{X}_1\}\rangle\,,$$

which is called the mean approach. It is easy to see that $\arg\max_{\mathbf{b}} \langle \mathbf{b}\,,\, \mathbb{E}\{\mathbf{X}_1\}\rangle$ is a portfolio vector having 1 at the position, where $\mathbb{E}\{\mathbf{X}_1\}$ has the largest component.

In his seminal paper Markowitz [27] realized that the mean approach is inadequate, i.e., it is a dangerous portfolio. In order to avoid this difficulty he suggested a diversification, which is called mean-variance portfolio such that

$$\tilde{\mathbf{b}} = \arg\max_{\mathbf{b}:\mathbf{Var}(\langle \mathbf{b}\,,\,\mathbf{X}_1\rangle)\leq\lambda} \langle \mathbf{b}\,,\, \mathbb{E}\{\mathbf{X}_1\}\rangle\,,$$

where $\lambda > 0$ is the risk aversion parameter.

For appropriate choice of λ, the performance (average growth rate) of $\tilde{\mathbf{b}}$ can be close to the performance of the optimal $\mathbf{b}^*$, however, the good choice of λ depends on the (unknown) distribution of the return vector $\mathbf{X}$.

The calculation of $\tilde{\mathbf{b}}$ is a nonlinear programming (NLP) problem, where a linear function is maximized under quadratic constraints.

In order to calculate the log-optimal portfolio $\mathbf{b}^*$, one has to know the distribution of $\mathbf{X}_1$. If this distribution is unknown then the empirical log-optimal portfolio can be defined by

$$\mathbf{b}_n^* = \arg\max_{\mathbf{b}} \frac{1}{n}\sum_{i=1}^{n} \ln \langle \mathbf{b}\,,\, \mathbf{X}_i\rangle$$

with linear constraints

$$\sum_{j=1}^{d} b^{(j)} = 1 \qquad \text{and} \qquad 0 \leq b^{(j)} \leq 1 \quad j = 1, \ldots, d\,.$$

The behavior of the empirical portfolio $\mathbf{b}_n^*$ and its modifications was studied by Móri [31], [32] and by Morvai [34], [35].

The calculation of $\mathbf{b}_n^*$ is a NLP problem, too. Cover [10] introduced an algorithm for calculating $\mathbf{b}_n^*$. An alternative possibility is the software routine DONLP2 of Spelluci [38]. The routine is based on sequential quadratic programming method, which computes sequentially a local solution of NLP by solving a quadratic programming problem and it estimates the global maximum according to these local maximums.

5 Examples for constantly rebalanced portfolio.

Example 1. Consider the example of $d = 2$ and $\mathbf{X} = (X^{(1)}, X^{(2)})$ such that the first component $X^{(1)}$ of the return vector $\mathbf{X}$ is an artificial stock:

$$X^{(1)} = \begin{cases} 2 & \text{with probability } 1/2, \\ 1/2 & \text{with probability } 1/2, \end{cases} \tag{1}$$

and the second component $X^{(2)}$ is the cash:

$$X^{(2)} = 1.$$

Obviously, the cash has zero growth rate. Using the expectation of the first component

$$\mathbb{E}\{X^{(1)}\} = 1/2 \cdot (2 + 1/2) = 5/4 > 1,$$

and the i.i.d. property of the market process, we get that

$$\mathbb{E}\{S_n^{(1)}\} = \mathbb{E}\left\{\prod_{i=1}^{n} X_i^{(1)}\right\} = (5/4)^n, \tag{2}$$

therefore $\mathbb{E}\{S_n^{(1)}\}$ grows exponentially. However, it does not imply that the random variable $S_n^{(1)}$ grows exponentially, too. Let's calculate the growth rate $W^{(1)}$:

$$\begin{aligned} W^{(1)} := \lim_{n\to\infty} \frac{1}{n} \ln S_n^{(1)} &= \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \ln X_i^{(1)} = \mathbb{E}\{\ln X^{(1)}\} \\ &= 1/2 \ln 2 + 1/2 \ln(1/2) = 0, \end{aligned}$$

which means that the first component has zero growth rate, too. Let's calculate the log-optimal portfolio for this return vector, where both components have zero growth rate. The portfolio vector has the form

$$\mathbf{b} = (b, 1-b).$$

Then

$$\begin{aligned} W(\mathbf{b}) &= \mathbb{E}\{\ln \langle \mathbf{b}\,,\, \mathbf{X} \rangle\} \\ &= 1/2\,(\ln(2b + (1-b)) + \ln(b/2 + (1-b))) \\ &= 1/2 \ln[(1+b)(1-b/2)]. \end{aligned}$$

One can check that $W(\mathbf{b})$ has the maximum for $b = 1/2$, so the log-optimal portfolio is

$$\mathbf{b}^* = (1/2, 1/2),$$

and the asymptotic average growth rate is

$$W^* = \mathbb{E}\{\ln \langle \mathbf{b}^*\,,\, \mathbf{X} \rangle\} = 1/2 \ln(9/8) = 0.059,$$

which is a positive growth rate.

Example 2. Consider the example of $d = 3$ and $\mathbf{X} = (X^{(1)}, X^{(2)}, X^{(3)})$ such that the first and the second components of the return vector $\mathbf{X}$ are artificial stocks of form (1), while the third component is the cash. One can show that the log-optimal portfolio is

$$\mathbf{b}^* = (0.46, 0.46, 0.08),$$

and the maximal asymptotic average growth rate is

$$W^* = \mathbb{E}\{\ln \langle \mathbf{b}^*\,,\, \mathbf{X} \rangle\} = 0.112.$$

Example 3. Consider the example of $d > 3$ and $\mathbf{X} = (X^{(1)}, X^{(2)}, \dots, X^{(d)})$ such that the first $d-1$ components of the return vector $\mathbf{X}$ are artificial stocks of form (1), while the last component is the cash. One can show that the log-optimal portfolio is

$$\mathbf{b}^* = (1/(d-1), \dots, 1/(d-1), 0),$$

which means that, for $d > 3$, according to the log-optimal portfolio the cash has zero weight. Let N denote the number of components of $\mathbf{X}$ equal to 2, then N is binomially distributed with parameters $(d-1, 1/2)$, and

$$\ln \langle \mathbf{b}^* , \mathbf{X} \rangle = \ln \left(\frac{2N + (d-1-N)/2}{d-1} \right) = \ln \left(\frac{3N}{2(d-1)} + \frac{1}{2} \right),$$

therefore

$$W^* = \mathbb{E}\{\ln \langle \mathbf{b}^* , \mathbf{X} \rangle\} = \mathbb{E} \left\{ \ln \left(\frac{3N}{2(d-1)} + \frac{1}{2} \right) \right\}.$$

For $d = 4$, the formula implies that the maximal asymptotic average growth rate is

$$W^* = \mathbb{E}\{\ln \langle \mathbf{b}^* , \mathbf{X} \rangle\} = 0.152,$$

while for $d \to \infty$,

$$W^* = \mathbb{E}\{\ln \langle \mathbf{b}^* , \mathbf{X} \rangle\} \to \ln(5/4) = 0.223,$$

which means that

$$S_n \approx e^{nW^*} = (5/4)^n,$$

so with many such stocks

$$S_n \approx \mathbb{E}\{S_n\}$$

(cf. (2)).

Example 4. Consider the example of horse racing with d horses in a race. Assume that horse j wins with probability p_j. The payoff is denoted by o_j, which means that investing 1\$ on horse j results in o_j if it wins, otherwise 0\$. Then the return vector is of form

$$\mathbf{X} = (0, \ldots, 0, o_j, 0, \ldots, 0)$$

if horse j wins. For repeated races, it is a constantly rebalanced portfolio problem. Let's calculate the expected log-return:

$$W(\mathbf{b}) = \mathbb{E}\{\ln \langle \mathbf{b} , \mathbf{X} \rangle\} = \sum_{j=1}^{d} p_j \ln(b^{(j)} o_j) = \sum_{j=1}^{d} p_j \ln b^{(j)} + \sum_{j=1}^{d} p_j \ln o_j,$$

therefore

$$\arg\max_{\mathbf{b}} \mathbb{E}\{\ln \langle \mathbf{b}\,,\ \mathbf{X}\rangle\} = \arg\max_{\mathbf{b}} \sum_{j=1}^{d} p_j \ln b^{(j)}.$$

In order to solve the optimization problem

$$\arg\max_{\mathbf{b}} \sum_{j=1}^{d} p_j \ln b^{(j)},$$

we introduce the Kullback-Leibler divergence of the distributions $\mathbf{p}$ and $\mathbf{b}$:

$$KL(\mathbf{p}, \mathbf{b}) = \sum_{j=1}^{d} p_j \ln \frac{p_j}{b^{(j)}}.$$

The basic property of the Kullback-Leibler divergence is that

$$KL(\mathbf{p}, \mathbf{b}) \geq 0,$$

and is equal to zero iff the two distributions are equal. The proof of this property is simple:

$$KL(\mathbf{p}, \mathbf{b}) = -\sum_{j=1}^{d} p_j \ln \frac{b^{(j)}}{p_j} \geq -\sum_{j=1}^{d} p_j \left(\frac{b^{(j)}}{p_j} - 1\right) = -\sum_{j=1}^{d} b^{(j)} + \sum_{j=1}^{d} p_j = 0.$$

This inequality implies that

$$\arg\max_{\mathbf{b}} \sum_{j=1}^{d} p_j \ln b^{(j)} = \mathbf{p}.$$

Surprisingly, the log-optimal portfolio is independent of the payoffs, and

$$W^* = \sum_{j=1}^{d} p_j \ln(p_j o_j).$$

The usual choice of payoffs is

$$o_j = \frac{1}{p_j},$$

and then

$$W^* = 0.$$

It means that, for this choice of payoffs, any gambling strategy has negative growth rate.

6 Semi-log-optimal portfolio.

Vajda [40] suggested an approximation of $\mathbf{b}^*$ and $\mathbf{b}_n^*$ using

$$h(z) := z - 1 - \frac{1}{2}(z-1)^2,$$

which is the second order Taylor expansion of the function $\ln z$ at $z = 1$. Then, the semi-log-optimal portfolio selection is

$$\bar{\mathbf{b}} = \arg\max_{\mathbf{b}} \mathbb{E}\{h(\langle \mathbf{b}\,,\, \mathbf{x}_1 \rangle)\},$$

and the empirical semi-log-optimal portfolio is

$$\bar{\mathbf{b}}_n = \arg\max_{\mathbf{b}} \frac{1}{n}\sum_{i=1}^{n} h(\langle \mathbf{b}\,,\, \mathbf{x}_i \rangle).$$

In order to compute $\mathbf{b}_n^*$, one has to make an optimization over $\mathbf{b}$. In each optimization step the computational complexity is proportional to n. For $\bar{\mathbf{b}}_n$, this complexity can be reduced. We have that

$$\frac{1}{n}\sum_{i=1}^{n} h(\langle \mathbf{b}\,,\, \mathbf{x}_i \rangle) = \frac{1}{n}\sum_{i=1}^{n}(\langle \mathbf{b}\,,\, \mathbf{x}_i \rangle - 1) - \frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}(\langle \mathbf{b}\,,\, \mathbf{x}_i \rangle - 1)^2.$$

If $\mathbf{1}$ denotes the all 1 vector, then

$$\frac{1}{n}\sum_{i=1}^{n} h(\langle \mathbf{b}\,,\, \mathbf{x}_i \rangle) = \langle \mathbf{b}\,,\, \mathbf{m} \rangle - \langle \mathbf{b}\,,\, \mathbf{C}\mathbf{b} \rangle\,,$$

where

$$\mathbf{m} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i - \mathbf{1})$$

and

$$\mathbf{C} = \frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i - \mathbf{1})(\mathbf{x}_i - \mathbf{1})^T.$$

If we calculate the vector $\mathbf{m}$ and the matrix $\mathbf{C}$ beforehand then in each optimization step the complexity does not depend on n, so the running time for calculating $\bar{\mathbf{b}}_n$ is much smaller than for $\mathbf{b}_n^*$. The other advantage of the semi-log-optimal portfolio is that it can be calculated via quadratic programming, which is doable, e.g., using the routine QuadProg++ of Di Gaspero [14]. This program uses Goldfarb-Idnani dual method for solving quadratic programming problems [15]. It easy to see that matrix $\mathbf{C}$ is positive semi-definit, however, the above mentioned dual method is only feasible if $\mathbf{C}$ is positive definite. This difference has not caused any problems in the experiments, but in case of causal empirical strategies sometimes $\mathbf{C}$ is calculated from few data, and so $\mathbf{C}$ is not a full-rank matrix, which implies that $\mathbf{C}$ is only positive semi-definite.

7 Numerical comparison of log-optimal and semi-log-optimal portfolios.

Table 1 summarizes the numerical results for NYSE data. At the web page `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 (5651 trading days ending in 1985). This data set has been used for testing portfolio selection in Cover [11], in Singer [37], in Györfi, Lugosi, Udina [17], in Györfi, Udina, Walk [19] and in Györfi, Urbán, Vajda [20].
- The second data set contains 23 stocks and has length 44 years (11178 trading days ending in 2006).

Our experiment is on the second data set. The first column of Table 1 lists the stock's name, the second column shows the average annual yield (AAY). The third and the fourth columns present the weights of the stocks (the components of the portfolio vector) using the log-optimal and semi-log-optimal algorithms. Surprisingly, the two portfolio vectors are almost the same, according to next-to-the-last row the growth rates are the same, and only four stocks have non-trivial (non-zero) weights.

An additional interesting feature of these results is that the KINAR with the smallest growth rate (5.9%) is included in the portfolio. The last row shows the running times of the algorithms. There were no difference

Stock's name	AAY	BCRP	
		log-NLP weights	semi-log-NLP weights
AHP	18.6%	0	0
ALCOA	13.1%	0	0
AMERB	20.2%	0	0
COKE	20.9%	0	0
COMME	26.1%	0.3028	0.2962
DOW	17.0%	0	0
DUPONT	12.4%	0	0
FORD	13.7%	0	0
GE	18.7%	0	0
GM	9.9%	0	0
HP	21.8%	0.0100	0.0317
IBM	14.1%	0	0
INGER	16.0%	0	0
JNJ	23.8%	0	0
KIMBC	18.9%	0	0
KINAR	5.9%	0.2175	0.2130
KODAK	8.9%	0	0
MERCK	21.7%	0	0
MMM	15.8%	0	0
MORRIS	30.1%	0.4696	0.4590
PANDG	18.8%	0	0
SCHLUM	22.2%	0	0
SHERW	18.2%	0	0
AAY		35.2%	35.2%
running time (sec)		9002	3

Table 1: Comparison of the two algorithms for CRPs.

between the obtained average annual yield of the methods, however, the running time is significantly shorter for semi-log-optimal algorithm.

8 Dynamic portfolio selection: general case

For a general dynamic portfolio selection, the portfolio vector may depend on the past data. As before, $\mathbf{x}_i = (x_i^{(1)}, \dots x_i^{(d)})$ denotes the return vector on trading period i. Let $\mathbf{b} = \mathbf{b}_1$ be the portfolio vector for the first trading period. For initial capital S_0, we get that

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

For the second trading period, S_1 is new initial capital, the portfolio vector is $\mathbf{b}_2 = \mathbf{b}(\mathbf{x}_1)$, and

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

For the nth trading period, a portfolio vector is $\mathbf{b}_n = \mathbf{b}(\mathbf{x}_1, \dots, \mathbf{x}_{n-1}) = \mathbf{b}(\mathbf{x}_1^{n-1})$ and

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

with the average growth rate

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

9 Log-optimal portfolio for stationary market process.

The fundamental limits, determined in Móri [29], in Algoet and Cover [3], and in Algoet [1, 2], reveal that the so-called *log-optimum portfolio* $\mathbf{B}^* = \{\mathbf{b}^*(\cdot)\}$ is the best possible choice. More precisely, on trading period n let $\mathbf{b}^*(\cdot)$ be such that

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

If $S_n^* = S_n(\mathbf{B}^*)$ denotes the capital achieved by a log-optimum portfolio strategy $\mathbf{B}^*$, after n trading periods, then for any other investment strategy $\mathbf{B}$ with capital $S_n = S_n(\mathbf{B})$ and with

$$\sup_n \mathbb{E}\left\{ (\ln \left\langle \mathbf{b}_n(\mathbf{X}_1^{n-1}) \, , \, \mathbf{X}_n \right\rangle)^2 \right\} < \infty,$$

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

$$\limsup_{n\to\infty} \frac{1}{n} \ln \frac{S_n}{S_n^*} \leq 0 \quad \text{almost surely} \tag{3}$$

and

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

where

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

is the maximal possible growth rate of any investment strategy. (Note that for memoryless markets $W^* = \max_{\mathbf{b}} \mathbb{E}\{\ln \langle \mathbf{b}\, , \, \mathbf{X}_0 \rangle\}$ which shows that in this case the log-optimal portfolio is a constantly rebalanced portfolio.)

For the proof of this optimality we use the concept of martingale differences:

Definition 1 *There are two sequences of random variables $\{Z_n\}$ and $\{X_n\}$ such that*

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

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

For martingale difference sequences, there is a strong law of large numbers: If $\{Z_n\}$ is a martingale difference sequence with respect to $\{X_n\}$ and

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

then

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

(cf. Chow [9], see also Stout [39, Theorem 3.3.1]).

In order to be self-contained, for martingale differences, we prove a weak law of large numbers. We show that if $\{Z_n\}$ is a martingale difference sequence with respect to $\{X_n\}$ then $\{Z_n\}$ are uncorrelated. Put $i < j$, then

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

It implies that

$$\mathbb{E}\left\{\left(\frac{1}{n}\sum_{i=1}^{n} Z_i\right)^2\right\} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}\{Z_i Z_j\} = \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{E}\{Z_i^2\} \to 0$$

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

One can construct martingale difference sequence as follows: let $\{Y_n\}$ be an arbitrary sequence such that Y_n is a function of $X_1, \dots, X_n$. Put

$$Z_n = Y_n - \mathbb{E}\{Y_n \mid X_1, \dots, X_{n-1}\}.$$

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

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

Now we can prove of optimality of the log-optimal portfolio: introduce the decomposition

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

The last average is an average of martingale differences, so it tends to zero a.s. Similarly,

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

Because of the definition of the log-optimal portfolio we have that

$$\mathbb{E}\{\ln \left\langle \mathbf{b}(\mathbf{X}_1^{i-1}) \, , \, \mathbf{X}_i \right\rangle \mid \mathbf{X}_1^{i-1}\} \leq \mathbb{E}\{\ln \left\langle \mathbf{b}^*(\mathbf{X}_1^{i-1}) \, , \, \mathbf{X}_i \right\rangle \mid \mathbf{X}_1^{i-1}\},$$

and the proof is finished.

10 Empirical portfolio selection

The optimality relations proved above give rise to the following definition:

Definition 2 *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,*

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

It is not at all obvious that such universally consistent portfolio strategy exists. The surprising fact that there exists a strategy, universal with respect to the class of all stationary and ergodic processes was proved by Algoet [1].

Most of the papers dealing with portfolio selections assume that the distributions of the market process are known. If the distributions are unknown then one can apply a two stage splitting scheme.

1: In the first time period the investor collects data, and estimates the corresponding distributions. In this period there is no any investment.

2: In the second time period the investor derives strategies from the distribution estimates and performs the investments.

In the sequel we show that there is no need to make any splitting, one can construct sequential algorithms such that the investor can make trading during the whole time period, i.e., the estimation and the portfolio selection is made on the whole time period.

Let's recapitulate the definition of log-optimal portfolio:

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

For a fixed integer $k > 0$ large enough, we expect that

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

and

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

Because of stationarity

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

which is the maximization of the regression function

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

Thus, a possible way for asymptotically optimal empirical portfolio selection is that, based on the past data, sequentially estimate the regression function $m_{\mathbf{b}}(\mathbf{x}_1^k)$, and choose the portfolio vector, which maximizes the regression function estimate.

11 Regression function estimation

Briefly summarize the basics of nonparametric regression function estimation. Concerning the details we refer to the book of Györfi, Kohler, Krzyzak and Walk [16]. Let Y be a real valued random variable, and let X denote a random vector. The regression function is the conditional expectation of Y given X:

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

If the distribution of (X, Y) is unknown then one has to estimate the regression function from data. The data is a sequence of i.i.d. copies of (X, Y):

$$D_n = \{(X_1, Y_1), \ldots, (X_n, Y_n)\}.$$

The regression function estimate is of form

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

An important class of estimates is the local averaging estimates

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

where usually the weights $W_{ni}(x; X_1, \ldots, X_n)$ are non-negative and sum up to 1. Moreover, $W_{ni}(x; X_1, \ldots, X_n)$ is relatively large if x is close to X_i, otherwise it is zero.

We use the following correspondence between the general regression estimation and portfolio selection:

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

12 Histogram based portfolio selection

The histogram or partitioning regression estimate is a local averaging estimate. Let $\mathcal{P}_n = \{A_{n,1}, A_{n,2} \ldots\}$ be a partition of the vector space, where X takes values. $A_{n,j}$ are called cells. Usual examples of cells are the cubes. If $A_n(x)$ is the cell of the partition $\mathcal{P}_n$ into which x falls then the partitioning regression estimate is defined by

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

where $I_{[\cdot]}$ stands for the indicator function.

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

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

then the histogram estimate has the form

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

Next we describe the *histogram based portfolio selection* introduced by Györfi and Schäfer [18] denote it by $\mathbf{B}^H$. $\mathbf{B}^H$ is constructed as follows. We first define an infinite array of elementary strategies (the so-called *experts*) $\mathbf{B}^{(k,\ell)} = \{\mathbf{b}^{(k,\ell)}(\cdot)\}$, indexed by the positive integers $k, \ell = 1, 2, \ldots$. Each expert $\mathbf{B}^{(k,\ell)}$ is determined by a period length k and by a partition $\mathcal{P}_\ell = \{A_{\ell,j}\}$, $j = 1, 2, \ldots, m_\ell$ of $\mathbb{R}_+^d$ into m_ℓ disjoint cells. To determine its portfolio on the nth trading period, expert $\mathbf{B}^{(k,\ell)}$ looks at the market vectors $\mathbf{x}_{n-k}, \ldots, \mathbf{x}_{n-1}$ of the last k periods, discretizes this kd-dimensional vector by means of the partition $\mathcal{P}_\ell$, and determines the portfolio vector which is optimal for those past trading periods whose preceding k trading periods have identical discretized market vectors to the present one. Formally, let G_ℓ be the discretization function corresponding to the partition $\mathcal{P}_\ell$, that is,

$$G_\ell(\mathbf{x}) = j, \text{ if } \mathbf{x} \in A_{\ell,j} \ .$$

With some abuse of notation, for any n and $\mathbf{x}_1^n \in \mathbb{R}^{dn}$, we write $G_\ell(\mathbf{x}_1^n)$ for the sequence $G_\ell(\mathbf{x}_1), \ldots, G_\ell(\mathbf{x}_n)$. Then define the expert $\mathbf{B}^{(k,\ell)} = \{\mathbf{b}^{(k,\ell)}(\cdot)\}$ by writing, for each $n > k+1$,

$$\mathbf{b}^{(k,\ell)}(\mathbf{x}_1^{n-1}) = \underset{\mathbf{b} \in \Delta_d}{\arg\max} \sum_{\left\{k<i<n: G_\ell(\mathbf{x}_{i-k}^{i-1}) = G_\ell(\mathbf{x}_{n-k}^{n-1})\right\}} \ln \langle \mathbf{b} \, , \, \mathbf{x}_i \rangle \ , \tag{4}$$

if the sum is non-void, and uniform $\mathbf{b}_0 = (1/d, \ldots, 1/d)$ otherwise. That is, $\mathbf{h}_n^{(k,\ell)}$ discretizes the sequence $\mathbf{x}_1^{n-1}$ according to the partition $\mathcal{P}_\ell$, and browses through all past appearances of the last seen discretized string $G_\ell(\mathbf{x}_{n-k}^{n-1})$ of length k. Then it designs a fixed portfolio vector optimizing the return for the trading periods following each occurrence of this string.

The problem left is how to choose k, ℓ. There are two extreme cases:

- small k or small ℓ imply that the corresponding regression estimate has large bias,

- large k and large ℓ imply that usually there are few matching, which results in large variance.

The good, data dependent choice of k and ℓ is doable borrowing current techniques from machine learning. In machine learning setup k and ℓ are considered as parameters of the estimates, called experts. The basic idea of machine learning is the combination of the experts. The combination is an aggregated estimate, where an expert has large weight if its past performance is good (cf. Cesa-Bianchi and Lugosi [8]).

The most successful combination is the exponential weighing. Combine the elementary portfolio strategies $\mathbf{B}^{(k,\ell)} = \{\mathbf{b}_n^{(k,\ell)}\}$ as follows: 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 $\eta > 0$, introduce the exponential weights

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

For $\eta = 1$, it means that

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

and

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

The combined portfolio $\mathbf{b}$ is defined by

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

This combination has a simple interpretation:

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

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

$$S_n(\mathbf{B}) = \sum_{k,\ell} q_{k,\ell} S_n(\mathbf{B}^{(k,\ell)}). \tag{5}$$

It is shown in [18] that the strategy $\mathbf{B}^H$ is universally consistent with respect to the class of all ergodic processes such that $\mathbb{E}\{|\ln X^{(j)}|\} < \infty$, for all $j = 1, 2, \ldots, d$ under the following two conditions on the partitions used in the discretization:

(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 $S \subset \mathbb{R}^d$ centered at the origin,

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

Sketch of the proof: Because of the fundamental limit (3), we have to prove that

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

We have 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

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

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

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

13 Kernel based portfolio selection

The kernel regression estimate is defined by a kernel function $K(x) \geq 0$ and by a bandwidth $h > 0$ such that

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

For the naive (window) kernel function $K(x) = I_{\{\|x\|\leq 1\}}$,

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

Györfi, Lugosi, Udina [17] introduced *kernel-based portfolio selection* strategies. Define an infinite array of experts $\mathbf{B}^{(k,\ell)} = \{\mathbf{b}^{(k,\ell)}(\cdot)\}$, where k, ℓ are positive integers. For fixed positive integers k, ℓ, choose the radius $r_{k,\ell} > 0$ such that for any fixed k,

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

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

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

if the sum is non-void, and $\mathbf{b}_0 = (1/d,\ldots,1/d)$ otherwise. These experts are mixed the same way as in the case of the histogram-based strategy (cf. (5)).

Györfi, Lugosi, Udina [17] proved that the portfolio scheme $\mathbf{B}^K$ is universally consistent with respect to the class of all ergodic processes such that $\mathbb{E}\{|\ln X^{(j)}|\} < \infty$, for $j = 1, 2, \ldots d$.

14 Nearest neighbor based portfolio selection

For an integer $k > 0$, the k-nearest neighbor (NN) regression estimate is a local averaging regression estimate, too,

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

such that 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.

Györfi, Udina, Walk [19] introduced the *nearest-neighbor-based portfolio selection* $\mathbf{B}^{NN}$ as follows. Define an infinite array of experts $\mathbf{B}^{(k,\ell)} = \{\mathbf{b}^{(k,\ell)}(\cdot)\}$, where $0 < k$, ℓ are integers. Just like before, k is the window length of the near past, and for each ℓ choose $p_\ell \in (0,1)$ such that

$$\lim_{\ell\to\infty} p_\ell = 0. \tag{6}$$

Put

$$\hat{\ell} = \lfloor p_\ell n \rfloor .$$

At a given time instant n, the expert searches for the $\hat{\ell}$ nearest neighbor matches in the past. For fixed positive integers k, ℓ ($n > k+\hat{\ell}+1$) introduce the set of the $\hat{\ell}$ nearest neighbor matches:

$$\begin{aligned}\hat{J}_n^{(k,\ell)} \quad = \quad & \{i;\, k+1 \leq i \leq n \text{ such that } \mathbf{x}_{i-k}^{i-1} \text{ is among the } \hat{\ell} \text{ NNs of } \mathbf{x}_{n-k}^{n-1} \\ & \text{in } \mathbf{x}_1^k, \ldots, \mathbf{x}_{n-k-1}^{n-2}\}.\end{aligned}$$

Define the expert $\mathbf{b}^{(k,\ell)}$ by

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

if the sum is non-void, and $\mathbf{b}_0 = (1/d, \ldots, 1/d)$ otherwise. That is, $\mathbf{b}_n^{(k,\ell)}$ is a fixed portfolio vector according to the returns following these nearest neighbors. These experts are mixed the same way as before (cf. (5)).

We say that tie occurs with probability zero if for any vector $\mathbf{s} = \mathbf{s}_1^k$ the random variable

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

has continuous distribution function.

Györfi, Udina, Walk [19] proved the following: Assume (6) and that tie occurs with probability zero. The portfolio scheme $\mathbf{B}^{NN}$ is universally consistent with respect to the class of all ergodic processes such that $\mathbb{E}\{|\ln X^{(j)}|\} < \infty$, for $j = 1, 2, \ldots d$.

15 Numerical results on empirical portfolio selection

One can combine the kernel based portfolio selection and the principle of semi-log-optimal algorithm in Section 6, called kernel based semi-log-optimal portfolio (cf. Györfi, Urbán, Vajda [20]). In this section we present some numerical results obtained by applying the kernel based semi-log-optimal algorithm to the second NYSE data set as in Section 7.

k / ℓ	1	2	3	4	5
1	29.4%	27.8%	22.9%	23.9%	23.7%
2	201.0%	124.6%	98.3%	36.1%	90.8%
3	114.2%	62.9%	38.8%	91.4%	31.6%
4	172.5%	155.0%	100.6%	162.3%	50.8%
5	233.5%	170.2%	166.6%	171.1%	107.7%
6	245.4%	216.2%	176.9%	182.0%	143.0%
7	261.8%	211.0%	181.2%	165.6%	158.7%
8	229.4%	189.2%	171.0%	138.8%	131.3%
9	219.3%	172.8%	162.4%	118.7%	116.0%
10	210.6%	151.5%	131.8%	103.4%	110.9%

Table 2: The average annual yields of the individual experts.

The proposed empirical portfolio selection algorithms use an infinite array of experts. In practice we take a finite array of size $K \times L$. In our experiment we selected $K = 5$ and $L = 10$. Choose the uniform distribution $\{q_{k,\ell}\} = 1/(KL)$ over the experts in use, and the radius

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

($k = 1, \ldots, K$ and $\ell = 1, \ldots, L$).

Table 2 summarizes the average annual yield achieved by each expert at the last period when investing one unit for the kernel-based semi-log-optimal portfolio. Experts are indexed by $k = 1 \ldots 5$ in columns and $\ell = 1 \ldots 10$ in rows. The average annual yield of kernel based semi-log-optimal portfolio is 222.7%. According to Table 1, MORRIS had the best average annual yield, 30.1%, while the BCRP had average annual yield 35.2%, so with kernel based semi-log-optimal portfolio we have a spectacular improvement.

Another interesting feature of Table 2 is that for any fixed ℓ, the best k is equal to 1, so as far as empirical portfolio is concerned the Markovian modelling is appropriate.

References

[1] Algoet, P. Universal schemes for prediction, gambling, and portfolio selection. *Annals of Probability*, 20:901–941, 1992.

[2] Algoet, P. The strong law of large numbers for sequential decisions under uncertainty. *IEEE Transactions on Information Theory*, 40:609–634, 1994.

[3] Algoet, P. and Cover, T. M. Asymptotic optimality asymptotic equipartition properties of log-optimum investments. *Annals of Probability*, 16:876–898, 1988.

[4] Barron, A. R. and Cover, T.M. A bound on the financial value of information. *IEEE Transactions on Information Theory*, 34:1097–1100, 1988.

[5] Blum, A. and Kalai, A. Universal portfolios with and without transaction costs. *Machine Learning*, 35:193–205, 1999.

[6] Bobryk, R. V. and Stettner, L. Discrete time portfolio selection with proportional transaction costs. *Probability and Mathematical Statistics*, 19:235-248, 1999.

[7] Breiman, L. Optimal gambling systems for favorable games. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.*, 1:65–78, Univ. California Press, Berkeley, 1961.

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

[9] Chow, Y. S. Local convergence of martingales and the law of large numbers. *Annals of Mathematical Statistics*, 36:552–558, 1965.

[10] Cover, T. An algorithm for maximazing expected log investment return. *IEEE Transactions on Information Theory*, 30:369–373, 1984.

[11] Cover, T. Universal portfolios. *Mathematical Finance*, 1:1–29, 1991.

[12] Cover, T. and Thomas, J. *Elements of Information Theory.* John Wiley and Sons, 1991.

[13] Finkelstein, M. and Whitley, R. Optimal strategies for repeated games. *Advances in Applied Probability*, 13:415–428, 1981.

[14] Di Gaspero, L. QuadProg++. *Luca Di Gaspero's homesite*, 2006, http://www.diegm.uniud.it/digaspero/

[15] Goldfarb, D. and Idnani, A. A numerically stable dual method for solving strictly quadratic programs. *Mathematical Programming*, 27:1–33, 1983.

[16] Györfi, L., Kohler, M., Krzyżak, A. and Walk, H. *A Distribution-Free Theory of Nonparametric Regression.* Springer Verlag, 2002.

[17] Györfi, L., Lugosi, G. and Udina, F. Nonparametric kernel based sequential investment strategies. *Mathematical Finance*, 16:337–357, 2006.

[18] Györfi, L. and Schäfer, D. Nonparametric prediction. In J. A. K. Suykens, G. Horváth, S. Basu, C. Micchelli, and J. Vandevalle, editors, *Advances in Learning Theory: Methods, Models and Applications*, pp. 339–354. IOS Press, NATO Science Series, 2003.

[19] Györfi, L., Udina, F. and Walk, H. Nonparametric nearest-neighbor-based sequential investment strategies, (submitted for publications), 2007.
http://www.szit.bme.hu/˜oti/portfolio/articles/NN.pdf

[20] Györfi, L., Urbán, A. and Vajda, I. Kernel-based semi-log-optimal empirical portfolio selection strategies. *Interantional Journal of Theoretical and Applied Finance*, 10:505-516, 2007.

[21] Iyengar, G. Discrete time growth optimal investment with costs. *Working Paper*, 2002.
http://www.columbia.edu/ gi10/Papers/stochastic.pdf

[22] Iyengar, G. Universal investment in markets with transaction costs. *Mathematical Finance*, 15:359-371, 2005.

[23] Iyengar, G. and Cover, T. Growths optimal investment in horse race markets with costs. *IEEE Transactions on Information Theory*, 46:2675-2683, 2000.

[24] Kelly, J. L. A new interpretation of information rate. *Bell System Technical Journal*, 35:917–926, 1956.

[25] Latané, H. A. Criteria for choice among risky ventures. *Journal of Political Economy*, 38:145–155, 1959.

[26] Luenberger, D. G. *Investment Science.* Oxford University Press, 1998.

[27] Markowitz, H. Portfolio selection. *Journal of Finance*, 7:77–91, 1952.

[28] Markowitz, H. Investment for long run: new evidence for an old rule. *Journal of Finance*, 31:1273–1286, 1976.

[29] Móri, T. F. On favourable stochastic games. *Annales Univ. Sci. Budapest. R. Eötvös Nom., Sect. Comput.*, 3:99–103, 1982.

[30] Móri, T. F. I-divergence geometry of distributions and stochastic games. In *Proc. of the 3rd Pannonian Symp. on Math. Stat.*, Eds. J. Mogyoródi et al, Reidel, Dordrecht, pages 231–238, 1984.

[31] Móri, T. F. Asymptotic properties of the empirical strategy in favourable stochastic games. In *Limit Theorems in Probability and Statistics*, Eds. P. Révész et al, North-Holland, pages 777–790, 1984.

[32] Móri, T. F. Is the empirical strategy optimal? *Statistics and Decisions*, 4:45–60, 1986.

[33] Móri, T. F. and Székely, G. J. How to win if you can? In *Limit Theorems in Probability and Statistics*, Eds. P. Révész et al, North-Holland, pages 791–806, 1984.

[34] Morvai, G. Empirical log-optimal portfolio selection. *Problems of Control and Information Theory*, 20:453–463, 1991.

[35] Morvai, G. Portfolio choice based on the empirical distribution. *Kybernetika*, 28:484–493, 1992.

[36] Schäfer, D. *Nonparametric Estimation for Finantial Investment under Log-Utility.* PhD Dissertation, Mathematical Institute, University Stuttgart, Aachen: Shaker Verlag, 2002.
http://www.szit.bme.hu/˜oti/portfolio/articles/dominik.pdf

[37] Singer, Y. Switching portfolios. *International Journal of Neural Systems*, 84:445–455, 1997.

[38] Spellucci, P. DONLP2, C Version 7/99. *Netlib*, 1999,
`http://www.netlib.org/`

[39] Stout. W. F. *Almost sure convergence.* Academic Press, New York, 1974.

[40] Vajda, I. Analysis of semi-log-optimal investment strategies, in *Prague Stochastics 2006*, M. Huskova, M. Janzura (Eds.), pages 719-727, MATFYZPRESS, Prague, 2006.