Empirical log-optimal portfolio selections: a survey.
*

László Györfi György Ottucsák András Urbán

Department of Computer Science and Information Theory
Budapest University of Technology and Economics,
Magyar Tudósok körútja 2.,Budapest,
Hungary, H-1117
{gyorfi,oti,urbi}@szit.bme.hu

Abstract

This paper provides a survey of discrete time, multi period, sequential investment strategies for financial markets. Under memoryless assumption on the underlying process generating the asset prices the Best Constantly Rebalanced Portfolio is studied, called log-optimal portfolio, which achieves the maximal asymptotic average growth rate. Semi-log optimal portfolio selection as a small computational complexity alternative of the log-optimal portfolio selection is studied both theoretically and empirically. For generalized dynamic portfolio selection, when asset prices are generated by a stationary and ergodic process, universally consistent empirical methods are shown. Empirical portfolio selection methods are proposed to handle the proportional transaction cost. The empirical performance of the methods illustrated for NYSE data with and without transaction costs.

*The first author acknowledges the support of the Computer and Automation Research Institute of the Hungarian Academy of Sciences.

Contents

1 Introduction

This paper gives an overview on the investment strategies in financial stock markets inspired by the results of information theory, non-parametric statistics and machine learning. Investment strategies are allowed to use information collected from the past of the market and determine, at the beginning of a trading period, a portfolio, that is, a way to distribute their current capital among the available assets. The goal of the investor is to maximize his wealth in the long run without knowing the underlying distribution generating the stock prices. Under this assumption the asymptotic rate of growth has a well-defined maximum which can be achieved in full knowledge of the underlying distribution generated by the stock prices.

Both static (buy and hold) and dynamic (daily rebalancing) portfolio selections are considered under various assumptions on the behavior of the market process. In case of static portfolio selection, it was shown that every portfolio achieves the maximal growth rate. One can achieve larger growth rate with daily rebalancing. Under memoryless assumption on the underlying process generating the asset prices, the log-optimal portfolio achieves the maximal asymptotic average growth rate, that is the expected value of the logarithm of the return for the best fix portfolio vector. Semi-log optimal portfolio selection as a small computational complexity alternative of the log-optimal portfolio selection is studied both theoretically and empirically. For generalized dynamic portfolio selection, when asset prices are generated by a stationary and ergodic process, universal consistent (empirical) methods that achieve the maximal possible growth rate are shown. Two extensions of the empirical portfolio selection methods are proposed to handle the proportional transaction cost. The empirical performance of the methods illustrated for NYSE data with and without transaction costs.

The rest of the paper is organized as follows. In Section 2 the constantly rebalanced portfolio is introduced, and the properties of log-optimal portfolio selection is analyzed in case of memoryless market. Next, a small computational complexity alternative of the log-optimal portfolio selection, the semi-log optimal portfolio is introduced. In Section 3 the general model of the dynamic portfolio selection is introduced and the basic features of the log-optimal portfolio selection in case of stationary and ergodic market are summarized. Using the principles of nonparametric statistics and machine learning, universal consistent, empirical investment strategies that are able to achieve the maximal asymptotic growth rate are introduced. Experiments on the NYSE data are given in Section 3.7. The possibility of consumption can be included in the model (Section 4). In Section 5 the portfolio selection with proportional transaction costs is analyzed.

1.1 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:

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

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.

1.2 Static portfolio selection

The static portfolio selection is a single period investment strategy. A portfolio vector is denoted by $\mathbf{b} = (b^{(1)}, \dots b^{(d)})$. 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, that means that short selling is not permitted. The set of portfolio vectors is denoted by

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

The aim of static portfolio selection 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 $j = 1, \dots, d$ 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 selection achieves the maximal growth rate $\max_j W^{(j)}$.

2 Constantly rebalanced portfolio selection

One can achieve even higher growth rate for long run investments, if the tuning of the porfolio is allowed *dynamically* after each trading period. 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. An representative example of the dynamic portfolio selection is the constantly rebalanced portfolio (CRP), was introduced and studied by Kelly [30], Latané [32], Breiman [7], Markowitz [35], Finkelstein and Whitley [17], Móri [39], Móri and Székely [42] and Barron and Cover [5]. For a comprehensive survey see also Chapters 6 and 15 in Cover and Thomas [13], and Chapter 15 in Luenberger [33].

Luenberger [33] summarizes the main conclusions as follows:

- "Conclusions about multiperiod investment situations are not mere variations of single-period conclusions – rather they offer *reverse* those earlier conclusions. This makes the subject exiting, both intellectually and in practice. Once the subtleties of multiperiod investment are understood, the reward in terms of enhanced investment performance can be substantial."
- "Fortunately the concepts and the methods of analysis for multiperiod situation build on those of earlier chapters. Internal rate of return, present value, the comparison principle, portfolio design, and lattice and tree valuation all have natural extensions to general situations. But conclusions such as volatility is "bad" or diversification is "good" are no longer universal truths. The story is much more interesting."

In case of CRP we 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. Note that in this case the investor has to rebalance his portfolio after each trading day to "corrigate" the daily price shifts of the invested stocks.

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 $S_0 = 1$.

2.1 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:

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

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 $\{\mathbf{X}_n\}_{-\infty}^{\infty}$,

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

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.

Avoiding (ii), see Section 5. 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 [34] 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

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

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

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

The behavior of the empirical portfolio $\mathbf{b}_n^*$ and its modifications was studied by Móri [40], [41] and by Morvai [43], [44].

The calculation of $\mathbf{b}_n^*$ is a NLP problem, too. Cover [11] introduced an algorithm for calculating $\mathbf{b}_n^*$. An alternative possibility is the software routine DONLP2 of Spelluci [49]. 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.

2.2 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.

The following viewpoint may help explain this at first sight surprising property. First, we write the evolution of the wealth of the stock as follows: let $S_n^{(1)} = 2^{2B(m,\frac{1}{2})-n}$, where $B(n, \frac{1}{2})$ is a binomial distribution random variable with parameters $(n, \frac{1}{2})$ (it is easy to check if we choose $n = 1$ then we return back to the one-step performance of stock). Now we write according to the Moivre-Laplace theorem (a special case of the central limit theorem for binomial distribution):

$$\mathbb{P}\left(\frac{2B(n, \frac{1}{2}) - n}{\sqrt{\mathbf{Var}(2B(n, \frac{1}{2})) - n}} \le x\right) \simeq \phi(x),$$

where $\phi(x)$ is cumulative distribution function of the standard normal distribution. Rearranging the left-hand side we have

$$\begin{aligned} \mathbb{P}\left(\frac{2B(n,\frac{1}{2})-n}{\sqrt{\mathbf{Var}(2B(n,\frac{1}{2}))-n}} \le x\right) &= \mathbb{P}\left(2B(n,\frac{1}{2})-n \le x\sqrt{n}\right) \\ &= \mathbb{P}\left(2^{2B(n,\frac{1}{2})-n} \le 2^{x\sqrt{n}}\right) \\ &= \mathbb{P}\left(S_n^{(1)} \le 2^{x\sqrt{n}}\right) \end{aligned}$$

that is

$$\mathbb{P}\left(S_n^{(1)} \le 2^{x\sqrt{n}}\right) \simeq \phi(x) \ .$$

Now let x_ε choose so that $\phi(x_\varepsilon) = 1-\varepsilon$ then

$$\mathbb{P}\left(S_n^{(1)} \le 2^{x_\varepsilon\sqrt{n}}\right) \simeq 1-\varepsilon$$

and for a fixed $\varepsilon > 0$ let n_0 be so that

$$2^{x\sqrt{n}} < \mathbb{E}S_n^{(1)} = \left(\frac{5}{4}\right)^n$$

for all $n > n_0$ then we have

$$\mathbb{P}\left(S_n^{(1)} \ge \mathbb{E}S_n^{(1)}\right) \le \mathbb{P}\left(S_n^{(1)} \ge 2^{x\sqrt{n}}\right) \simeq \varepsilon.$$

It means that most of the values of $S_n^{(1)}$ are far smaller than its expected value (see in Figure 1).

Now let's turn back to the original problem and 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

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

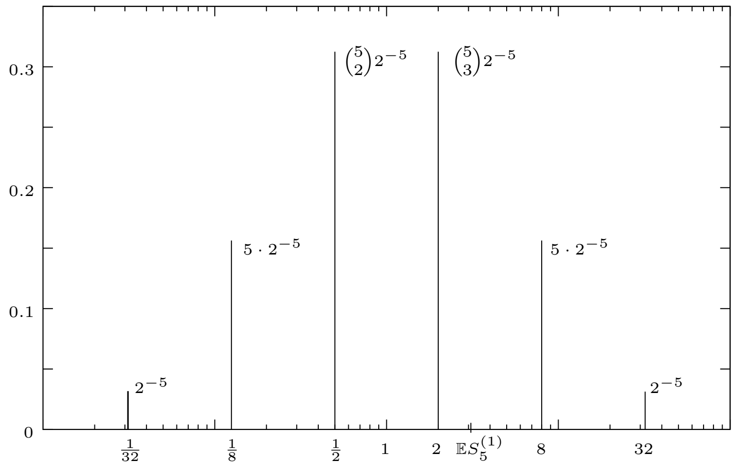

Figure 1: The distribution of $S_n^{(1)}$ in case of $n = 5$

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 $\mathbf{X} = (X^{(1)}, X^{(2)}, \ldots, X^{(d)})$ 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

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

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

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

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 $\mathbf{b}$:

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

The basic property of the Kullback-Leibler divergence is that

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

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

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

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.

Example 5. Sequential St.Petersburg games.
Consider the simple St.Petersburg game, where the player invests 1 dollar and a fair coin is tossed until a tail first appears, ending the game. If the first tail appears in step k then the the payoff X is 2^k and the probability of this event is 2^{-k}:

$$\mathbb{P}\{X = 2^k\} = 2^{-k}.$$

Since $\mathbb{E}\{X\} = \infty$, this game has delicate properties (cf. Aumann [4], Bernoulli [6], Durand [15], Haigh [26], Martin [36], Menger [37], Rieger and Wang [45] and Samuelson [46].) In the literature, usually the repeated St.Petersburg game (called iterated St.Petersburg game, too) means multi-period game such that it is a sequence of simple St.Petersburg games, where in each round the player invest 1 dollar. Let X_n denote the payoff for the n-th simple game. Assume that the sequence $\{X_n\}_{n=1}^{\infty}$ is independent and identically distributed. After n rounds the player's wealth in the repeated game is

$$\tilde{S}_n = \sum_{i=1}^{n} X_i,$$

then

$$\lim_{n\to\infty} \frac{\tilde{S}_n}{n \log_2 n} = 1$$

in probability, where $\log_2$ denotes the logarithm with base 2 (cf. Feller [16]). Moreover,

$$\liminf_{n\to\infty} \frac{\tilde{S}_n}{n \log_2 n} = 1$$

a.s. and

$$\limsup_{n\to\infty} \frac{\tilde{S}_n}{n \log_2 n} = \infty$$

a.s. (cf. Chow and Robbins [10]). Introducing the notation for the largest payoff

$$X_n^* = \max_{1\le i\le n} X_i$$

and for the sum with the largest payoff withheld

$$S_n^* = \tilde{S}_n - X_n^*,$$

one has that

$$\lim_{n\to\infty} \frac{S_n^*}{n \log_2 n} = 1$$

a.s. (cf. Csörgő and Simons [14]). According to the previous results $\tilde{S}_n \approx n \log_2 n$. Next we introduce a multi-period game, called sequential St.Petersburg game, having exponential growth. The sequential St.Petersburg game means that the player starts with initial capital $S_0 = 1$ dollar, and there is an independent sequence of simple St.Petersburg games, and for each simple game the player reinvest his capital. If $S_{n-1}^{(c)}$ is the capital after the $(n-1)$-th simple game then the invested capital is $S_{n-1}^{(c)}(1-c)$, while $S_{n-1}^{(c)}c$ is the proportional cost of the simple game with commission factor $0 < c < 1$. It means that after the n-th round the capital is

$$S_n^{(c)} = S_{n-1}^{(c)}(1-c)X_n = S_0(1-c)^n \prod_{i=1}^{n} X_i = (1-c)^n \prod_{i=1}^{n} X_i.$$

Because of its multiplicative definition, $S_n^{(c)}$ has exponential trend:

$$S_n^{(c)} = e^{nW_n^{(c)}} \approx e^{nW^{(c)}},$$

with average growth rate

$$W_n^{(c)} := \frac{1}{n} \ln S_n^{(c)}$$

and with asymptotic average growth rate

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

Let's calculate the the asymptotic average growth rate. Because of

$$W_n^{(c)} = \frac{1}{n} \ln S_n^{(c)} = \frac{1}{n} \left(n \ln(1-c) + \sum_{i=1}^{n} \ln X_i \right),$$

the strong law of large numbers implies that

$$W^{(c)} = \ln(1-c) + \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \ln X_i = \ln(1-c) + \mathbb{E}\{\ln X_1\}$$

a.s., so $W^{(c)}$ can be calculated via expected log-utility (cf. Kenneth [31]). A commission factor c is called fair if

$$W^{(c)} = 0,$$

so the growth rate of the sequential game is 0. Let's calculate the fair c:

$$\ln(1-c) = -\mathbb{E}\{\ln X_1\} = -\sum_{k=1}^{\infty} k \ln 2 \cdot 2^{-k} = -2\ln 2,$$

i.e.,

$$c = 3/4.$$

Next we study the portfolio game, where a fraction of the capital is invested in the simple fair St.Petersburg game and the rest is kept in cash. This is the model of the constantly rebalanced portfolio (CRP). Fix a portfolio vector $\mathbf{b} = (b, 1-b)$, with $0 \le b \le 1$. Let $S_0 = 1$ denote the player's initial capital. Then at the beginning of the portfolio game $S_0 b = b$ is invested into the fair game, and it results in return $bX_1/4$, while $S_0(1-b) = 1-b$ remains in cash, therefore after the first portfolio game the player's wealth becomes

$$S_1 = S_0(bX_1/4 + (1-b)) = b(X_1/4 - 1) + 1.$$

For the second portfolio game, S_1 is the new initial capital

$$S_2 = S_1(b(X_2/4-1)+1) = (b(X_1/4-1)+1)(b(X_2/4-1)+1).$$

By induction, for n-th portfolio game the initial capital is S_{n-1}, therefore

$$S_n = S_{n-1}(b(X_n/4-1)+1) = \prod_{i=1}^{n}(b(X_i/4-1)+1).$$

The asymptotic average growth rate of this portfolio game is

$$\begin{aligned} W(b) &:= \lim_{n\to\infty} \frac{1}{n} \ln S_n \\ &= \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \ln(b(X_i/4-1)+1) \\ &\to \mathbb{E}\{\ln(b(X_1/4-1)+1)\} \end{aligned}$$

a.s. The function ln is concave, therefore $W(b)$ is concave, too, so $W(0) = 0$ (keep everything in cash) and $W(1) = 0$ (the simple game is fair) imply that for all $0 < b < 1$, $W(b) > 0$. Let's calculate

$$\max_b W(b).$$

b	$W(b)$
0	0
0.1	0.061
0.2	0.084
0.3	0.095
0.4	0.097
0.5	0.093
0.6	0.083
0.7	0.068
0.8	0.049
0.9	0.025
1	0

Table 1: The asymptotic average growth rate of the portfolio game.

We have that

$$\begin{aligned} W(b) &= \sum_{k=1}^{\infty} \ln(b(2^k/4-1)+1) \cdot 2^{-k} \\ &= \ln(1-b/2) \cdot 2^{-1} + \sum_{k=3}^{\infty} \ln(b(2^{k-2}-1)+1) \cdot 2^{-k}. \end{aligned}$$

Table 1 shows some figures on the average growth rate of the portfolio game. If $b = 0.4$ then $W(b) = 0.097$, so if for each game one reinvest 40% of his capital such that the real investment is 10%, while the cost is 30%, then the growth rate is approximately 10%, i.e., the portfolio game with two components of 0 growth rate (fair St.Petersburg game and cash) can result in growth rate of 10%.

2.3 Semi-log-optimal portfolio

Vajda [51] 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 [18]. This program uses Goldfarb-Idnani dual method for solving quadratic programming problems [19]. 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.

Finally we reveal a surprising property of the semi-log optimal portfolio.

Let us use the definition of the function h, then we have that

$$\begin{aligned}\mathbb{E}\{h(\langle \mathbf{b}, \mathbf{x}_1\rangle)\} &= \mathbb{E}\{(\langle \mathbf{b}, \mathbf{x}_1\rangle - 1)\} - \frac{1}{2}\mathbb{E}\{(\langle \mathbf{b}, \mathbf{x}_1\rangle - 1)^2\} \\ &= \mathbb{E}\{\langle \mathbf{b}, \mathbf{x}_1\rangle\} - 1 - \frac{1}{2}\mathbb{E}\{\langle \mathbf{b}, \mathbf{x}_1\rangle^2 - 2\langle \mathbf{b}, \mathbf{x}_1\rangle - 1\} \\ &= -\frac{3}{2} + 2\mathbb{E}\{\langle \mathbf{b}, \mathbf{x}_1\rangle\} - \frac{1}{2}\mathbb{E}\{\langle \mathbf{b}, \mathbf{x}_1\rangle^2\} \\ &= \frac{1}{2} - \frac{1}{2}\mathbb{E}\{(\langle \mathbf{b}, \mathbf{x}_1\rangle - 2)^2\},\end{aligned}$$

therefore

$$\bar{\mathbf{b}} = \arg\max_{\mathbf{b}} \mathbb{E}\{h(\langle \mathbf{b}, \mathbf{x}_1\rangle)\} = \arg\min_{\mathbf{b}} \mathbb{E}\{(\langle \mathbf{b}, \mathbf{x}_1\rangle - 2)^2\}.$$

Thus, $\bar{\mathbf{b}}$ can be considered as the principal component for portfolio selection such that for $\mathbf{b} = \bar{\mathbf{b}}$, $\langle \mathbf{b}, \mathbf{x}_1\rangle$ approximates best the value 2 in mean square sense.

3 Time varying portfolio selection

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

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

with the average growth rate

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

3.1 Log-optimal portfolio for stationary market process

The fundamental limits, determined in Móri [38], 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

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

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 $\{\mathbf{X}_n\}_{-\infty}^{\infty}$,

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

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 [50, 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

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

and the proof is finished.

3.2 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.

3.3 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 [20]. 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

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

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

An example of such an estimate is the *partitioning estimate*. Here one chooses a finite or countably infinite partition $\mathcal{P}_n = \{A_{n,1}, A_{n,2}, \dots\}$ of $\mathbb{R}^d$ consisting of cells $A_{n,j} \subseteq \mathbb{R}^d$ and defines, for $x \in A_{n,j}$, the estimate by averaging Y_i's with the corresponding X_i's in $A_{n,j}$, i.e.,

$$m_n(x) = \frac{\sum_{i=1}^{n} I_{\{X_i \in A_{n,j}\}} Y_i}{\sum_{i=1}^{n} I_{\{X_i \in A_{n,j}\}}} \quad \text{for } x \in A_{n,j}, \tag{4}$$

where I_A denotes the indicator function of set A, so

$$W_{n,i}(x) = \frac{I_{\{X_i \in A_{n,j}\}}}{\sum_{l=1}^{n} I_{\{X_l \in A_{n,j}\}}} \quad \text{for } x \in A_{n,j}.$$

Here and in the following we use the convention $\frac{0}{0} = 0$. In order to have consistency, on the one hand we need that the cells $A_{n,j}$ should be "small", and on the other hand the number of non-zero terms in the denominator of (4) should be "large". These requirements can be satisfied if the sequences of partition $\mathcal{P}_n$ is asymptotically fine, i.e., if

$$\text{diam}(A) = \sup_{x,y \in A} \|x - y\|$$

denotes the diameter of a set such that $||\cdot||$ is the Eucledian norm, then for each sphere S centered at the origin

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

and

$$\lim_{n \to \infty} \frac{|\{j \,:\, A_{n,j} \cap S \neq \emptyset\}|}{n} = 0.$$

For the partition $\mathcal{P}_n$, the most important example is when the cells $A_{n,j}$ are cubes of volume h_n^d. For cubic partition, the consistency conditions above mean that

$$\lim_{n\to\infty} h_n = 0 \quad \text{and} \quad \lim_{n\to\infty} n h_n^d = \infty. \tag{5}$$

The second example of a local averaging estimate is the *Nadaraya–Watson kernel estimate.* Let $K : \mathbb{R}^d \to \mathbb{R}_+$ be a function called the kernel function, and let $h > 0$ be a bandwidth. The kernel estimate is defined by

$$m_n(x) = \frac{\sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) Y_i}{\sum_{i=1}^n K\left(\frac{x-X_i}{h}\right)}, \tag{6}$$

so

$$W_{n,i}(x) = \frac{K\left(\frac{x-X_i}{h}\right)}{\sum_{j=1}^n K\left(\frac{x-X_j}{h}\right)}.$$

Here the estimate is a weighted average of the Y_i, where the weight of Y_i (i.e., the influence of Y_i on the value of the estimate at x) depends on the distance between X_i and x. For the bandwidth $h = h_n$, the consistency conditions are (5). If one uses the so-called naive kernel (or window kernel) $K(x) = I_{\{\|x\|\le 1\}}$, where $I_{\{\cdot\}}$ denotes the indicator function of the events in the brackets, that is, it equals 1 if the event is true and 0 otherwise. Then

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

i.e., one estimates $m(x)$ by averaging Y_i's such that the distance between X_i and x is not greater than h.

Our final example of local averaging estimates is the *k-nearest neighbor* (k-NN) *estimate.* Here one determines the k nearest X_i's to x in terms of distance $\|x - X_i\|$ and estimates $m(x)$ by the average of the corresponding Y_i's. More precisely, for $x \in \mathbb{R}^d$, let

$$(X_{(1)}(x), Y_{(1)}(x)), \dots, (X_{(n)}(x), Y_{(n)}(x))$$

be a permutation of

$$(X_1, Y_1), \dots, (X_n, Y_n)$$

such that

$$\|x - X_{(1)}(x)\| \le \cdots \le \|x - X_{(n)}(x)\|.$$

The k-NN estimate is defined by

$$m_n(x) = \frac{1}{k} \sum_{i=1}^{k} Y_{(i)}(x). \tag{7}$$

Here the weight $W_{ni}(x)$ equals $1/k$ if X_i is among the k nearest neighbors of x, and equals 0 otherwise. If $k = k_n \to \infty$ such that $k_n/n \to 0$ then the k-nearest-neighbor regression estimate is consistent.

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

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

3.4 Histogram based strategy

Next we describe *histogram based strategy* due to Györfi and Schäfer [22] and denote it by $\mathbf{B}^H$. 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 sets. 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 $n > k+1$,

$$\begin{aligned} \mathbf{b}^{(k,\ell)}(\mathbf{x}_1^{n-1}) &= \arg\max_{\mathbf{b} \in \Delta_d} \prod_{i \in J_{k,l,n}} \langle \mathbf{b}\,,\, \mathbf{x}_i \rangle\ , \qquad (8) \\ \text{where } J_{k,l,n} &= \left\{ k < i < n : G_\ell(\mathbf{x}_{i-k}^{i-1}) = G_\ell(\mathbf{x}_{n-k}^{n-1}) \right\}\ , \end{aligned}$$

if $J_{k,l,n} \neq \emptyset$, and uniform $\mathbf{b}_0 = (1/d, \ldots, 1/d)$ otherwise. That is, $\mathbf{b}_n^{(k,\ell)}$ 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 implies that the corresponding regression estimate has large bias,
- large k and large ℓ implies 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 weighting. 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

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

This combination has a simple interpretation:

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

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

$$S_n(\mathbf{B}^H) = \sum_{k,\ell} q_{k,\ell} S_n(\mathbf{B}^{(k,\ell)}). \tag{9}$$

It is shown in [22] that the strategy $\mathbf{B}^H$ is universally consistent with respect to the class of all ergodic processes such that $\mathbb{E}\{|\log X^{(j)}|\} < \infty$, 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,

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

3.5 Kernel based strategy

Györfi, Lugosi, Udina [21] 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 as in (9).

Györfi, Lugosi, Udina [21] proved that the portfolio scheme $\mathbf{B}^K = \mathbf{B}$ is universally consistent with respect to the class of all ergodic processes such that $\mathbb{E}\{|\ln X^{(j)}|\} < \infty$, for $j = 1,2,\ldots d$.

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

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

We have that

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

Thus

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

Because of $\lim_{\ell\to\infty} r_{k,\ell} = 0$, we can show that

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

3.6 Nearest neighbor based strategy

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

$$\lim_{\ell\to\infty} p_\ell = 0. \tag{10}$$

Put

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

At a given time instant n, the expert searches for the $\hat{\ell}$ nearest neighbor (NN) 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:

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

Define the expert by

$$\mathbf{b}^{(k,\ell)}(\mathbf{x}_1^{n-1}) = \underset{\mathbf{b}\in\Delta_d}{\arg\max} \prod_{i\in\hat{J}_n^{(k,\ell)}} \langle \mathbf{b}\,,\, \mathbf{x}_i\rangle\,.$$

That is, $\mathbf{b}_n^{(k,\ell)}$ is a fixed portfolio vector according to the returns following these nearest neighbors. These experts are mixed in the same way as in (9).

We say that a 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, and Walk [23] proved the following theorem: assume (10) and that a tie occurs with probability zero, the the portfolio scheme $\mathbf{B}^{NN}$ is universally consistent with respect to the class of all stationary and ergodic processes such that $\mathbb{E}\{|\log X^{(j)}|\} < \infty$, for $j = 1, 2, \ldots d$.

3.7 Numerical results on empirical portfolio selection

This section gives some empirical results on the BCRP selection. 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). More precisely, the data set contains the daily price relatives, that was calculated from the nominal values of the *closing prices* corrected by the dividends and the splits for all trading day. This data set has been used for testing portfolio selection in Cover [12], in Singer [48], in Györfi, Lugosi, Udina [21], in Györfi, Udina, Walk [23] and in Györfi, Urbán, Vajda [24].

- The second data set contains 23 stocks and has length 44 years (11178 trading days ending in 2006) and it was generated same way as the previous data set (it was augmented by the last 22 years).

Our experiment is on the *second data set* such that we left out four small assets (SHERW, KODAK, COMME, KINAR) having small capitalization (less than 10^{10} dollars).

To make the analysis feasible, some simplifying assumptions are used that need to be taken into account. Assume

- 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 (in Section 5 we offer solutions to overcome this problem),

- the behavior of the market is not affected by the actions of the investor using the strategy under investigation.

For the 19 large assets, the average annual yield (AAY) of the fixed uniform portfolio is 14%, while the AAY of the constantly rebalanced uniform portfolio is 15%. Table 2 summarizes the numerical results for 19 large assets. The best asset was MORRIS with AAY 20%. The first column of Table 2 lists the stock's name, the second column shows the 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: 20%.

For the calculation of the optimal portfolio we have developed a recursive GRADIENT ALGORITHM. Introduce the projection P of a vector $\mathbf{b} = (b^{(1)}, \dots b^{(d)})$ to Δ_d:

$$P(\mathbf{b}) = \frac{\mathbf{b}}{\sum_{j=1}^{d} b^{(j)}}.$$

Stock's name	AAY	BCRP	
		log-Grad weights	slog-Grad weights
AHP	13%	0	0
ALCOA	9%	0	0
AMERB	14%	0	0
COKE	14%	0	0
DOW	12%	0	0
DUPONT	9%	0	0
FORD	9%	0	0
GE	13%	0	0
GM	7%	0	0
HP	15%	0.179	0.176
IBM	10%	0	0
INGER	11%	0	0
JNJ	16%	0	0
KIMBC	13%	0	0
MERCK	15%	0	0
MMM	11%	0	0
MORRIS	20%	0.744	0.744
PANDG	13%	0	0
SCHLUM	15%	0.077	0.08
AAY		20%	20%

Table 2: Comparison of the two algorithms for CRPs.

Put

$$W_n(\mathbf{b}) = \frac{1}{n}\sum_{i=1}^{n} \log \langle \mathbf{b}\,,\, \mathbf{x}_i \rangle\,,$$

and let $\mathbf{e}_j$ be the j-th unit vector, i.e., its j-th component is 1, the other components are 0. Choose the initial values

$$\mathbf{b}_0 = (1/d, \dots, 1/d)$$

and

$$V_0 = W_n(\mathbf{b}_0)$$

and a step size $\delta > 0$ (In our experiment we had $\delta = 0.1/d$.)

For $k = 1, 2, ...,$ make the following iteration:

Step 1. Calculate

$$W_n(P(\mathbf{b}_{k-1} + \delta \cdot \mathbf{e}_j)) \quad j = 1, \dots, d.$$

STEP 2. If

$$V_{k-1} \geq \max_j W_n(P(\mathbf{b}_{k-1} + \delta \cdot \mathbf{e}_j))$$

then stop, and the result of the algorithm is $\mathbf{b}_{k-1}$.
Otherwise, put

$$V_k = \max_j W_n(P(\mathbf{b}_{k-1} + \delta \cdot \mathbf{e}_j))$$

and

$$\mathbf{b}_k = P(\mathbf{b}_{k-1} + \delta \cdot \mathbf{e}_{j^*}),$$

where

$$j^* = \arg\max_j W_n(P(\mathbf{b}_{k-1} + \delta \cdot \mathbf{e}_j)).$$

Go to STEP 1.

k / ℓ	1	2	3	4	5
1	31%	30%	24%	21%	26%
2	34%	31%	27%	25%	22%
3	35%	29%	26%	24%	23%
4	35%	30%	30%	32%	27%
5	34%	29%	33%	24%	24%
6	35%	29%	28%	24%	27%
7	33%	29%	32%	23%	23%
8	34%	33%	30%	21%	24%
9	37%	33%	28%	19%	21%
10	34%	29%	26%	20%	24%

Table 3: The average annual yields of the individual experts for the kernel strategy.

One can combine the kernel based portfolio selection and the principle of semi-log-optimal algorithm in Section 2.3, called kernel based semi-log-optimal portfolio (cf. Györfi, Urbán, Vajda [24]). In this section we present some numerical results obtained by applying the kernel based semi-log-optimal algorithm to the 19 large assets of the second NYSE data set.

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

$$r_{k,l}^2 = 0.0002 \cdot d \cdot k + 0.00002 \cdot d \cdot k \cdot \ell,$$

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

Table 3 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 31%. According to Table 2, MORRIS had the best average annual yield, 20%, while the BCRP had average annual yield 20%, so with kernel based semi-log-optimal portfolio we have a spectacular improvement.

Another interesting feature of Table 3 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.

k / ℓ	1	2	3	4	5
50	31%	33%	28%	24%	35%
100	33%	32%	25%	29%	28%
150	38%	33%	26%	32%	27%
200	38%	28%	32%	32%	24%
250	37%	31%	37%	28%	26%
300	41%	35%	35%	30%	29%
350	39%	36%	31%	34%	32%
400	39%	35%	33%	32%	35%
450	39%	34%	34%	35%	37%
500	42%	36%	33%	38%	35%

Table 4: The average annual yields of the individual experts for the nearest neighbor strategy.

We performed some experiments using nearest neighbor strategy. Again, we take a finite array of size $K \times L$ such that $K = 5$ and $L = 10$. Choose the uniform distribution $\{q_{k,\ell}\} = 1/(KL)$ over the experts in use. Table 4 summarizes the average annual yield achieved by each expert at the last period when investing one unit for the nearest neighbor portfolio strategy. Experts are indexed by $k = 1 \ldots 5$ in columns and $\ell = 50, 100, \ldots, 500$ in rows, where ℓ is the number of nearest neighbors. The average annual yield of nearest neighbor portfolio is 35%. Comparing Tables 3 and 4, one can conclude that the nearest neighbor strategy is more robust.

4 Portfolio selection with consumption

For a real number x, let x^+ be the positive part of x. Assume that at the end of trading period n there is a consumption $c_n \geq 0$. For the trading period n the initial capital is S_{n-1}, therefore

$$S_n = (S_{n-1} \langle \mathbf{b}_n \, , \, \mathbf{x}_n \rangle - c_n)^+ .$$

If $S_j > 0$ for all $j = 1, \ldots, n$ then we show by induction that

$$S_n = S_0 \prod_{i=1}^{n} \langle \mathbf{b}_i \, , \, \mathbf{x}_i \rangle - \sum_{k=1}^{n} c_k \prod_{i=k+1}^{n} \langle \mathbf{b}_i \, , \, \mathbf{x}_i \rangle \, , \tag{11}$$

where the empty product is 1, by definition. For $n = 1$, (11) holds. Assume (11) for $n - 1$:

$$S_{n-1} = S_0 \prod_{i=1}^{n-1} \langle \mathbf{b}_i \, , \, \mathbf{x}_i \rangle - \sum_{k=1}^{n-1} c_k \prod_{i=k+1}^{n-1} \langle \mathbf{b}_i \, , \, \mathbf{x}_i \rangle \, .$$

Then

$$\begin{aligned} S_n &= S_{n-1} \langle \mathbf{b}_n \, , \, \mathbf{x}_n \rangle - c_n \\ &= \left(S_0 \prod_{i=1}^{n-1} \langle \mathbf{b}_i \, , \, \mathbf{x}_i \rangle - \sum_{k=1}^{n-1} c_k \prod_{i=k+1}^{n-1} \langle \mathbf{b}_i \, , \, \mathbf{x}_i \rangle \right) \langle \mathbf{b}_n \, , \, \mathbf{x}_n \rangle - c_n \\ &= S_0 \prod_{i=1}^{n} \langle \mathbf{b}_i \, , \, \mathbf{x}_i \rangle - \sum_{k=1}^{n} c_k \prod_{i=k+1}^{n} \langle \mathbf{b}_i \, , \, \mathbf{x}_i \rangle \, . \end{aligned}$$

One has to emphasize that (11) holds for all n iff $S_n > 0$ for all n, otherwise there is a ruin. In the sequel, we study the average growth rate under no ruin and the probability of ruin.

By definition,

$$\begin{aligned} \mathbb{P}\{ \text{ ruin } \} &= \mathbb{P}\left\{ \bigcup_{n=1}^{\infty} \{S_n = 0\} \right\} \\ &= \mathbb{P}\left\{ \bigcup_{n=1}^{\infty} \left\{ S_0 \prod_{i=1}^{n} \langle \mathbf{b}_i \, , \, \mathbf{x}_i \rangle - \sum_{k=1}^{n} c_k \prod_{i=k+1}^{n} \langle \mathbf{b}_i \, , \, \mathbf{x}_i \rangle \leq 0 \right\} \right\} , \end{aligned}$$

therefore

$$\begin{aligned}
\mathbb{P}\{\text{ ruin }\} &= \mathbb{P}\left\{\bigcup_{n=1}^{\infty}\left\{\prod_{i=1}^{n}\langle\mathbf{b}_i\,,\,\mathbf{x}_i\rangle\left(S_0-\sum_{k=1}^{n}c_k\frac{1}{\prod_{i=1}^{k}\langle\mathbf{b}_i\,,\,\mathbf{x}_i\rangle}\right)\le 0\right\}\right\}\\
&\le \mathbb{P}\left\{\bigcup_{n=1}^{\infty}\left\{\prod_{i=1}^{n}\langle\mathbf{b}_i\,,\,\mathbf{x}_i\rangle\left(S_0-\sum_{k=1}^{\infty}c_k\frac{1}{\prod_{i=1}^{k}\langle\mathbf{b}_i\,,\,\mathbf{x}_i\rangle}\right)\le 0\right\}\right\}\\
&\le \mathbb{P}\left\{S_0\le\sum_{k=1}^{\infty}c_k\frac{1}{\prod_{i=1}^{k}\langle\mathbf{b}_i\,,\,\mathbf{x}_i\rangle}\right\} \qquad (12)
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{P}\{\text{ ruin }\} &= \mathbb{P}\left\{\bigcup_{n=1}^{\infty}\left\{\prod_{i=1}^{n}\langle\mathbf{b}_i\,,\,\mathbf{x}_i\rangle\left(S_0-\sum_{k=1}^{n}c_k\frac{1}{\prod_{i=1}^{k}\langle\mathbf{b}_i\,,\,\mathbf{x}_i\rangle}\right)\le 0\right\}\right\}\\
&\ge \max_n\mathbb{P}\left\{\prod_{i=1}^{n}\langle\mathbf{b}_i\,,\,\mathbf{x}_i\rangle\left(S_0-\sum_{k=1}^{n}c_k\frac{1}{\prod_{i=1}^{k}\langle\mathbf{b}_i\,,\,\mathbf{x}_i\rangle}\right)\le 0\right\}\\
&= \mathbb{P}\left\{S_0\le\sum_{k=1}^{\infty}c_k\frac{1}{\prod_{i=1}^{k}\langle\mathbf{b}_i\,,\,\mathbf{x}_i\rangle}\right\}. \qquad (13)
\end{aligned}$$

(12) and (13) imply that

$$\mathbb{P}\{\text{ ruin }\}=\mathbb{P}\left\{S_0\le\sum_{k=1}^{\infty}c_k\frac{1}{\prod_{i=1}^{k}\langle\mathbf{b}_i\,,\,\mathbf{x}_i\rangle}\right\}.$$

Under no ruin, on the one hand we get the upper bound on the average growth rate

$$\begin{aligned}
W_n &= \frac{1}{n}\ln S_n\\
&= \frac{1}{n}\ln\left(S_0\prod_{i=1}^{n}\langle\mathbf{b}_i\,,\,\mathbf{x}_i\rangle-\sum_{k=1}^{n}c_k\prod_{i=k+1}^{n}\langle\mathbf{b}_i\,,\,\mathbf{x}_i\rangle\right)\\
&\le \frac{1}{n}\ln S_0\prod_{i=1}^{n}\langle\mathbf{b}_i\,,\,\mathbf{x}_i\rangle\\
&= \frac{1}{n}\sum_{i=1}^{n}\ln\langle\mathbf{b}_i\,,\,\mathbf{x}_i\rangle+\frac{1}{n}\ln S_0.
\end{aligned}$$

On the other hand we have the lower bound

$$\begin{aligned} W_n &= \frac{1}{n} \ln S_n \\ &= \frac{1}{n} \ln \left(S_0 \prod_{i=1}^{n} \langle \mathbf{b}_i\,,\,\mathbf{x}_i \rangle - \sum_{k=1}^{n} c_k \prod_{i=k+1}^{n} \langle \mathbf{b}_i\,,\,\mathbf{x}_i \rangle \right) \\ &= \frac{1}{n} \ln \prod_{i=1}^{n} \langle \mathbf{b}_i\,,\,\mathbf{x}_i \rangle \left(S_0 - \sum_{k=1}^{n} c_k \frac{1}{\prod_{i=1}^{k} \langle \mathbf{b}_i\,,\,\mathbf{x}_i \rangle} \right) \\ &\geq \frac{1}{n} \ln \prod_{i=1}^{n} \langle \mathbf{b}_i\,,\,\mathbf{x}_i \rangle \left(S_0 - \sum_{k=1}^{\infty} c_k \frac{1}{\prod_{i=1}^{k} \langle \mathbf{b}_i\,,\,\mathbf{x}_i \rangle} \right) \\ &= \frac{1}{n} \sum_{i=1}^{n} \ln \langle \mathbf{b}_i\,,\,\mathbf{x}_i \rangle + \frac{1}{n} \ln \left(S_0 - \sum_{k=1}^{\infty} c_k \frac{1}{\prod_{i=1}^{k} \langle \mathbf{b}_i\,,\,\mathbf{x}_i \rangle} \right), \end{aligned}$$

therefore under no ruin the asymptotic average growth rate with consumption is the same as without consumption:

$$W_n = \frac{1}{n} \ln S_n \approx \frac{1}{n} \sum_{i=1}^{n} \ln \langle \mathbf{b}_i\,,\,\mathbf{x}_i \rangle\,.$$

Consider the case of constant consumption, i.e., $c_n = c > 0$. Then there is no ruin if

$$S_0 > c \sum_{k=1}^{\infty} \frac{1}{\prod_{i=1}^{k} \langle \mathbf{b}_i\,,\,\mathbf{x}_i \rangle}.$$

Because of the definition of the average growth rate we have that

$$W_k = \frac{1}{k} \ln \prod_{i=1}^{k} \langle \mathbf{b}_i\,,\,\mathbf{x}_i \rangle\,,$$

which implies that

$$\sum_{k=1}^{\infty} \frac{1}{\prod_{i=1}^{k} \langle \mathbf{b}_i\,,\,\mathbf{x}_i \rangle} = \sum_{k=1}^{\infty} e^{-kW_k}.$$

Assume that our portfolio selection is asymptotically optimal, which means that

$$\lim_{n\to\infty} W_n = W^*.$$

Then

$$\sum_{k=1}^{\infty} \frac{1}{\prod_{i=1}^{k} \langle \mathbf{b}_i \,,\, \mathbf{x}_i \rangle} \approx \sum_{k=1}^{\infty} e^{-kW^*} = \frac{e^{-W^*}}{1-e^{-W^*}}.$$

This approximation implies that the ruin probability can be small only if

$$S_0 > c\frac{e^{-W^*}}{1-e^{-W^*}}.$$

A special case of this model is when there is only one risk-free asset:

$$S_n = (S_{n-1}(1+r) - c)^+$$

with some $r > 0$. Obviously, there is no ruin if $S_0 r > c$. It is easy to verify that this assumption can be derived from the general condition if

$$e^{W^*} = 1 + r.$$

The ruin probability can be decreased if the consumptions happen in blocks of size N trading periods. Let S_n denote the wealth at the end of n-th block. Then

$$S_n = \left(S_{n-1} \prod_{j=(n-1)N+1}^{nN} \langle \mathbf{b}_j \,,\, \mathbf{x}_j \rangle - Nc \right)^+ .$$

Similarly to the previous calculations, we can check that under no ruin the average growth rates with and without consumption are the same. Moreover

$$\mathbb{P}\{\text{ ruin }\} = \mathbb{P}\left\{ S_0 \le cN \sum_{k=1}^{\infty} \frac{1}{\prod_{i=1}^{kN} \langle \mathbf{b}_i \,,\, \mathbf{x}_i \rangle} \right\}.$$

This ruin probability is a monotonically decreasing function of N, and for large N the exact condition of no ruin is the same as the approximation in the previous section.

This model can be applied for the analysis of portfolio selection strategies with fixed transaction cost such that c_n is the transaction cost to be paid when change the portfolio $\mathbf{b}_n$ to $\mathbf{b}_{n+1}$. In this case the transaction cost c_n depends on the number of shares involved in the transaction.

Let's calculate c_n. At the end of the n-th trading period and before paying for transaction cost the wealth at asset j is $S_{n-1}b_n^{(j)}x_n^{(j)}$, which means that the number of shares j is

$$m_n^{(j)} = \frac{S_{n-1}b_n^{(j)}x_n^{(j)}}{S_n^{(j)}}.$$

In the model of fixed transaction cost, we assume that $m_n^{(j)}$ is integer. If one changes the portfolio $\mathbf{b}_n$ to $\mathbf{b}_{n+1}$ then the wealth at asset j should be $S_{n-1}\langle \mathbf{b}_n , \mathbf{x}_n \rangle b_{n+1}^{(j)}$, so the number of shares j should be

$$m_{n+1}^{(j)} = \frac{S_{n-1}\langle \mathbf{b}_n , \mathbf{x}_n \rangle b_{n+1}^{(j)}}{S_n^{(j)}}.$$

If $m_{n+1}^{(j)} < m_n^{(j)}$ then we have to sell, and the wealth what we got is

$$\sum_{j=1}^{d}\left(m_n^{(j)} - m_{n+1}^{(j)}\right)^+ S_n^{(j)} = \sum_{j=1}^{d}\left(S_{n-1}b_n^{(j)}x_n^{(j)} - S_{n-1}\langle \mathbf{b}_n , \mathbf{x}_n \rangle b_{n+1}^{(j)}\right)^+ .$$

If $m_{n+1}^{(j)} > m_n^{(j)}$ then we have to buy, and the wealth what we pay is

$$\sum_{j=1}^{d}\left(m_{n+1}^{(j)} - m_n^{(j)}\right)^+ S_n^{(j)} = \sum_{j=1}^{d}\left(S_{n-1}\langle \mathbf{b}_n , \mathbf{x}_n \rangle b_{n+1}^{(j)} - S_{n-1}b_n^{(j)}x_n^{(j)}\right)^+ .$$

Let $C > 0$ be the fixed transaction cost, then the transaction fee is

$$c_n = c_n(\mathbf{b}_{n+1}) = C\sum_{j=1}^{d}\left|m_n^{(j)} - m_{n+1}^{(j)}\right| .$$

The portfolio selection $\mathbf{b}_{n+1}$ is self-financing if

$$\begin{aligned} & \sum_{j=1}^{d}\left(S_{n-1}b_n^{(j)}x_n^{(j)} - S_{n-1}\langle \mathbf{b}_n , \mathbf{x}_n \rangle b_{n+1}^{(j)}\right)^+ \\ \geq\ & \sum_{j=1}^{d}\left(S_{n-1}\langle \mathbf{b}_n , \mathbf{x}_n \rangle b_{n+1}^{(j)} - S_{n-1}b_n^{(j)}x_n^{(j)}\right)^+ + c_n. \end{aligned}$$

$\mathbf{b}_{n+1}$ is an admissible portfolio if $m_{n+1}^{(j)}$ is integer for all j and it satisfies the self-financing condition. The set of admissible portfolios is denoted by $\Delta_{n,d}$.

Taking into account the fixed transaction cost, a kernel based portfolio selection can be defined as follows: choose the radius $r_{k,\ell} > 0$ such that for any fixed k,

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

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

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

if the sum is non-void, and $\mathbf{b}_0 = (1/d, \ldots, 1/d)$ otherwise. 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, $q_{k,\ell} > 0$. The combined strategy $\mathbf{B}$ then arises from weighing the elementary portfolio strategies $\{\mathbf{b}_n^{(k,\ell)}\}$ such that the investor's capital becomes

$$S_n = \sum_{k,\ell} q_{k,\ell} S_n^{(k,\ell)}. \qquad (14)$$

5 Portfolio selection for proportional transaction cost

The problem of growth optimal investment with proportional transaction costs was studied by Cover and Iyengar [29] in horse race markets, also called erodible market. Iyengar [27] investigated growth optimal investment with several assets assuming independent and identically distributed sequence of asset returns. The most far reaching study was Schäfer [47] who considered the maximization of the expected growth rate with several assets when the asset returns are Markovian. Györfi and Vajda [25] extended it to almost sure optimality. All these papers assume the knowledge of the distributions of the market process. In this section we have some experiments on two empirical (data driven) portfolio selections.

Let S_n denote the wealth at the close of market day n, $n = 0, 1, 2, \cdots$, where w.l.o.g. let the investor's initial capital S_0 be 1 dollar. At the beginning of a new market day $n+1$, the investor sets up his new portfolio, i.e. buys/sells stocks according to the actual portfolio vector $\mathbf{b}_{n+1}$. During this rearrangement, he has to pay transaction cost, therefore at the beginning of a new market day $n+1$ the net wealth N_n in the portfolio $\mathbf{b}_{n+1}$ is less than S_n.

Using the above notations the (gross) wealth S_n at the close of market day n is

$$S_n = N_{n-1} \sum_{j=1}^{d} b_n^{(j)} x_n^{(j)} = N_{n-1} \langle \mathbf{b}_n \, , \, \mathbf{x}_n \rangle \, ,$$

where $\langle \cdot \, , \, \cdot \rangle$ denotes inner product.

The rate of proportional transaction cost (commission factor) levied on one asset is denoted by $0 < c < 1$, i.e. the sale of 1 dollar worth of asset i nets only $1 - c$ dollars, and similarly we take into account the purchase of an asset such that the purchase of 1 dollar's worth of asset i costs c dollars. It is not hard to see that gross wealth S_n decomposes to the sum of the net wealth and cost the following - self-financing - way

$$\begin{aligned} N_n &= S_n - \sum_{j=1}^{d} c \left(b_n^{(j)} x_n^{(j)} N_{n-1} - b_{n+1}^{(j)} N_n \right)^+ - \sum_{j=1}^{d} c \left(b_{n+1}^{(j)} N_n - b_n^{(j)} x_n^{(j)} N_{n-1} \right)^+ \\ &= S_n - c \sum_{j=1}^{d} \left| b_n^{(j)} x_n^{(j)} N_{n-1} - b_{n+1}^{(j)} N_n \right| , \end{aligned}$$

or equivalently

$$S_n = N_n + c \sum_{j=1}^{d} \left| b_n^{(j)} x_n^{(j)} N_{n-1} - b_{n+1}^{(j)} N_n \right| .$$

Dividing both sides by S_n and introducing ratio

$$w_n = \frac{N_n}{S_n},$$

$0 < w_n < 1$, we get

$$1 = w_n + c \sum_{j=1}^{d} \left| \frac{b_n^{(j)} x_n^{(j)}}{\langle \mathbf{b}_n \, , \, \mathbf{x}_n \rangle} - b_{n+1}^{(j)} w_n \right| . \tag{15}$$

Equation (15) is used in the sequel. Examining this cost equation, it turns out, that for arbitrary portfolio vectors $\mathbf{b}_n$, $\mathbf{b}_{n+1}$, and return vector $\mathbf{x}_n$ there exists a unique cost factors $w_n \in [0, 1)$, i.e. the portfolio is self financing. The value of cost factor w_n at day n is determined by portfolio vectors $\mathbf{b}_n$ and $\mathbf{b}_{n+1}$ as well as by return vector $\mathbf{x}_n$, i.e.

$$w_n = w(\mathbf{b}_n, \mathbf{b}_{n+1}, \mathbf{x}_n),$$

for some function w. If we want to rearrange our portfolio substantially, then our net wealth decreases more considerably, however, it remains positive. Note, also, that the cost does not restricts the set of new portfolio vectors, i.e. the optimization algorithm searches for optimal vector $\mathbf{b}_{n+1}$ within the whole simplex Δ_d. The value of the cost factor ranges between

$$\frac{1-c}{1+c} \leq w_n \leq 1.$$

Starting with an initial wealth $S_0 = 1$ and $w_0 = 1$, wealth S_n at the closing time of the n-th market day becomes

$$S_n = N_{n-1}\langle \mathbf{b}_n \, , \, \mathbf{x}_n \rangle = w_{n-1} S_{n-1} \langle \mathbf{b}_n \, , \, \mathbf{x}_n \rangle = \prod_{i=1}^{n} w(\mathbf{b}_{i-1}, \mathbf{b}_i, \mathbf{x}_{i-1}) \langle \mathbf{b}_i \, , \, \mathbf{x}_i \rangle \, .$$

Introduce the notation

$$g(\mathbf{b}_{i-1}, \mathbf{b}_i, \mathbf{x}_{i-1}, \mathbf{x}_i) = \ln(w(\mathbf{b}_{i-1}, \mathbf{b}_i, \mathbf{x}_{i-1}) \langle \mathbf{b}_i \, , \, \mathbf{x}_i \rangle),$$

then the average growth rate becomes

$$\begin{aligned} \frac{1}{n} \ln S_n &= \frac{1}{n} \sum_{i=1}^{n} \ln(w(\mathbf{b}_{i-1}, \mathbf{b}_i, \mathbf{x}_{i-1}) \langle \mathbf{b}_i \, , \, \mathbf{x}_i \rangle) \\ &= \frac{1}{n} \sum_{i=1}^{n} g(\mathbf{b}_{i-1}, \mathbf{b}_i, \mathbf{x}_{i-1}, \mathbf{x}_i). \end{aligned}$$

Our aim is to maximize the average growth rate.

In the sequel $\mathbf{x}_i$ will be random variable and is denoted by $\mathbf{X}_i$. Let's use the decomposition

$$\frac{1}{n} \ln S_n = I_n + J_n,$$

where

$$I_n = \frac{1}{n} \sum_{i=1}^{n} (g(\mathbf{b}_{i-1}, \mathbf{b}_i, \mathbf{X}_{i-1}, \mathbf{X}_i) - \mathbb{E}\{g(\mathbf{b}_{i-1}, \mathbf{b}_i, \mathbf{X}_{i-1}, \mathbf{X}_i) | \mathbf{X}_1^{i-1}\})$$

and

$$J_n = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\{g(\mathbf{b}_{i-1}, \mathbf{b}_i, \mathbf{X}_{i-1}, \mathbf{X}_i) | \mathbf{X}_1^{i-1}\}.$$

I_n is an average of martingale differences, which, under general conditions on the support of the distribution of $\mathbf{X}$, converges to 0 almost surely. Thus, the asymptotic maximization of the average growth rate $\frac{1}{n} \ln S_n$ is equivalent to the maximization of J_n.

Algorithm 1. For transaction cost, one may apply the portfolio $\mathbf{b}_n^*(\mathbf{X}_{n-1})$ or its empirical approximation. For example, we may apply the kernel based log-optimal portfolio selection introduced by Györfi, Lugosi and Udina [21] as follows: Define an infinite array of experts $\mathbf{B}^{(\ell)} = \{\mathbf{b}^{(\ell)}(\cdot)\}$, where ℓ is a positive integer. For fixed positive integer ℓ, choose the radius $r_\ell > 0$ such that

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

Then, for $n > 1$, define the expert $\mathbf{b}^{(\ell)}$ as follows. Put

$$\mathbf{b}_n^{(\ell)} = \arg\max_{\mathbf{b} \in \Delta_d} \sum_{\{i<n: \|\mathbf{x}_{i-1} - \mathbf{x}_{n-1}\| \le r_\ell\}} \ln \langle \mathbf{b} \, , \, \mathbf{x}_i \rangle \ , \tag{16}$$

if the sum is non-void, and $\mathbf{b}_0 = (1/d, \ldots, 1/d)$ otherwise, where $\|\cdot\|$ denotes the Euclidean norm.

These experts are aggregated (mixed) as follows: let $\{q_\ell\}$ be a probability distribution over the set of all positive integers ℓ such that for all ℓ, $q_\ell > 0$. Consider two types of aggregations:

- Here the initial capital $S_0 = 1$ is distributed among the expert according to the distribution $\{q_\ell\}$, and the expert makes the portfolio selection and pays for transaction cost individually. If $S_n(\mathbf{B}^{(\ell)})$ is the capital accumulated by the elementary strategy $\mathbf{B}^{(\ell)}$ after n periods when starting with an initial capital $S_0 = 1$, then, after period n, the investor's wealth after period n, aggregations with the wealth:

$$S_n = \sum_\ell q_\ell S_n(\mathbf{B}^{(\ell)}). \tag{17}$$

- Here $S_n(\mathbf{B}^{(\ell)})$ is again the capital accumulated by the elementary strategy $\mathbf{B}^{(\ell)}$ after n periods when starting with an initial capital $S_0 = 1$, but it is virtual figure, i.e., the experts make no trading, its wealth is just the base of aggregation. Then, after period n, the investor's aggregated portfolio becomes

$$\mathbf{b}_n = \frac{\sum_\ell q_\ell S_{n-1}(\mathbf{B}^{(\ell)}) \mathbf{b}_n^{(\ell)}}{\sum_\ell q_\ell S_{n-1}(\mathbf{B}^{(\ell)})}. \tag{18}$$

Moreover, the investor's capital is

$$S_n = S_{n-1}\langle \mathbf{b}_n \, , \, \mathbf{x}_n \rangle w(\mathbf{b}_{n-1}, \mathbf{b}_n, \mathbf{x}_{n-1}),$$

so only the aggregated portfolio pays for the transaction cost.

From now on we consider the optimization of investment with transaction costs. If the market process $\{\mathbf{X}_i\}$ is a *stationary and first order Markov process* then, for any portfolio selection $\{\mathbf{b}_i\}$, we have that

$$\begin{aligned} & \mathbb{E}\{g(\mathbf{b}_{i-1}, \mathbf{b}_i, \mathbf{X}_{i-1}, \mathbf{X}_i) | \mathbf{X}_1^{i-1}\} \\ = \ & \mathbb{E}\{\ln(w(\mathbf{b}_{i-1}, \mathbf{b}_i, \mathbf{X}_{i-1}) \langle \mathbf{b}_i \, , \, \mathbf{X}_i \rangle) | \mathbf{X}_1^{i-1}\} \\ = \ & \ln w(\mathbf{b}_{i-1}, \mathbf{b}_i, \mathbf{X}_{i-1}) + \mathbb{E}\{\ln \langle \mathbf{b}_i \, , \, \mathbf{X}_i \rangle \, | \mathbf{X}_1^{i-1}\} \\ = \ & \ln w(\mathbf{b}_{i-1}, \mathbf{b}_i, \mathbf{X}_{i-1}) + \mathbb{E}\{\ln \langle \mathbf{b}_i \, , \, \mathbf{X}_i \rangle \, | \mathbf{b}_i, \mathbf{X}_{i-1}\} \\ \stackrel{\text{def}}{=} \ & v(\mathbf{b}_{i-1}, \mathbf{b}_i, \mathbf{X}_{i-1}), \end{aligned}$$

therefore the maximization of the average growth rate $\frac{1}{n} \ln S_n$ is asymptotically equivalent to the maximization of

$$J_n = \frac{1}{n} \sum_{i=1}^{n} v(\mathbf{b}_{i-1}, \mathbf{b}_i, \mathbf{X}_{i-1}).$$

This maximization is a dynamic programming problem.

Algorithm 2. We may introduce a suboptimal solution, called *naive portfolio*, by a one-step optimization as follows: put $\mathbf{b}_1 = \{1/d, \dots, 1/d\}$ and for $n \geq 1$,

$$\mathbf{b}_n^{(\ell)} = \arg\max_{\mathbf{b} \in \Delta_d} \sum_{\{i<n: \|\mathbf{x}_{i-1} - \mathbf{x}_{n-1}\| \leq r_\ell\}} \left(\ln \langle \mathbf{b} \, , \, \mathbf{x}_i \rangle + \ln w(\mathbf{b}_{n-1}, \mathbf{b}, \mathbf{x}_{n-1}) \right), \quad (19)$$

if the sum is non-void, and $\mathbf{b}_0 = (1/d, \dots, 1/d)$ otherwise. These elementary portfolios are mixed as in (17) or (18).

Obviously, this portfolio has no global optimality property.

Next we present some numerical results for transaction cost obtained by applying the kernel based semi-log-optimal algorithm to the 19 large assets of the second NYSE data set as in Section 3.7. The proposed empirical portfolio selection algorithms use an infinite set of experts. Here we take a finite set of size L. In the experiment we selected $L = 10$. Choose the uniform distribution $\{q_\ell\} = 1/L$ over the experts in use, and the radius

$$r_\ell^2 = 0.0002 \cdot d + 0.00002 \cdot d \cdot \ell, \qquad \text{for } \ell = 1, \dots, L \, .$$

Table 5 summarizes the average annual yield achieved by each expert at the last period when investing one unit for the kernel-based log-optimal portfolio. Experts are indexed by $\ell = 1 \dots 10$ in rows. The second column

ℓ	$c = 0$	Algorithm 1	Algorithm 2
1	31%	-22%	18%
2	34%	-22%	10%
3	35%	-24%	9 %
4	35%	-23%	14%
5	34%	-21%	13%
6	35%	-19%	13%
7	33%	-20%	12%
8	34%	-18%	8 %
9	37%	-17%	6 %
10	34%	-18%	11%
Aggregation with wealth (17)	35%	**-19%**	**13%**
Aggregation with portfolio (18)	35%	**-15%**	**17%**

Table 5: The average annual yields of the individual experts for kernel strategy and of the aggregations with $c = 0.0015$.

ℓ	$c = 0$	Algorithm 1	Algorithm 2
50	31%	-35%	-14%
100	33%	-33%	3%
150	38%	-29%	3%
200	38%	-28%	9%
250	37%	-28%	9%
300	41%	-26%	7%
350	39%	-26%	9%
400	39%	-26%	10%
450	39%	-25%	14%
500	42%	-23%	14%
Aggregation with wealth (17)	39%	**-25%**	**11%**
Aggregation with portfolio (18)	39%	**-23%**	**11%**

Table 6: The average annual yields of the individual experts for nearest neighbor strategy and of the aggregations with $c = 0.0015$.

contains the average annual yields of experts for kernel based log-optimal portfolio if there is no transaction cost, and in this case the results of the two aggregations are the same: 35%. Mention that, out of the 19 assets, MORRIS had the best average annual yield, 20%, so, for no transaction cost, with kernel based log-optimal portfolio we have a spectacular improvement. The third and fourth columns contain the average annual yields of experts for kernel based log-optimal portfolio if the commission factor is $c = 0.0015$. Notice that the growth rate of the Algorithm 1 is negative, and the growth rate of the Algorithm 2 is poor, too, it is less than the growth rate of the best asset, and the results of aggregations are different.

We have got similar results for nearest neighbor strategy (cf. Table 6). On the one hand these 19 assets may be too risk averse to offer "good" growth rate of the wealth. On the other hand these results have revealed that the proper handling of the transaction cost is still an open question and an important direction of the further research. If the market process is first order Markov and one knows the conditional distributions then Györfi and Vajda [25] introduced a.s. optimal strategies, however, it is unknown how to construct their data driven versions.

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] Aumann, R. J. The St. Petersburg paradox: A discussion of some recent comments. *Journal of Economic Theory*, 14:443–445, 1977.

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

[6] Bernoulli, D. Originally published in 1738; translated by L. Sommer. Exposition of a new theory on the measurement of risk. *Econometrica*, 22:22–36, 1954.

[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] Chow, Y. S. and Robbins, H. On sums of independent random variables with infinite moments and "fair" games. *Proc. Nat. Acad. Sci. USA*, 47:330–335, 1961.

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

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

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

[14] Csörgő, S. and Simons, G. A strong law of large numbers for trimmed sums, with applications to generalized StPetersburg games. *Statistics and Probability Letters*, 26:65–73, 1996.

[15] Durand, D. Growth stocks and the Petersburg paradox. *The Journal of Finance*, 12:348–363, 1957.

[16] Feller, W. Note on the law of large numbers and "fair" games. *Ann. Math. Statist.*, 16:301–304, 1945.

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

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

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

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

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

[22] 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.

[23] Györfi, L., Udina, F. and Walk, H. Nonparametric nearest-neighbor-based sequential investment strategies, *Statistics and Decisions*, (in print), 2008.
http://www.szit.bme.hu/˜oti/portfolio/articles/NN.pdf

[24] 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.

[25] Györfi, L. and Vajda, I. On the growth of sequential portfolio selection with transaction cost, in *Algorithmic Learning Theory*, Freund, Y., Györfi, L., Turán, G., Zeugmann, Th. (Eds.), Lecture Notes in Computer Science, Vol. 5254, Springer Verlag, pp. 108-122, 2008.

[26] Haigh, J. *Taking Chances*. Oxford,UK: Oxford University Press, 1999.

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

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

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

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

[31] Kenneth, A. J. The use of unbounded utility functions in expected-utility maximization: Response. *Quarterly Journal of Economics*, 88:136–138, 1974.

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

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

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

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

[36] Martin, R. *The St. Petersburg Paradox.* The Stanford Encyclopedia of Philosophy (Fall 2004 Edition). Ed. E. N. Zalta. Stanford, California: Stanford University, 2004.

[37] Menger, K. Das Unsicherheitsmoment in der Wertlehre Betrachtungen im Anschlu an das sogenannte Petersburger Spiel. *Zeitschrift fr Nationalkonomie*, 5:459–485, 1934.

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

[39] 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.

[40] 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.

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

[42] 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.

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

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

[45] Rieger, M. O. and Mei Wang. Cumulative prospect theory and the St. Petersburg paradox. *Economic Theory*, 28:665–679, 2006.

[46] Samuelson, P. The St. Petersburg paradox as a divergent double limit. *International Economic Review*, 1:31–37, 1960.

[47] 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

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

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

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

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