Statistics & Decisions 99, 999– 9999 (2003)

Nonparametric nearest neighbor based empirical portfolio selection strategies

László Györfi, Frederic Udina, Harro Walk

Summary: In recent years optimal portfolio selection strategies for sequential investment have been shown to exist. Although their asymptotical optimality is well established, finite sample properties do need the adjustment of parameters that depend on dimensionality and scale. In this paper we introduce some nearest neighbor based portfolio selectors that solve these problems, and we show that they are also log-optimal for the very general class of stationary and ergodic random processes. The newly proposed algorithm shows very good finite-horizon performance when applied to different markets with different dimensionality or scales without any change: we see it as a very robust strategy.

1 Introduction

In a financial market, on the basis of the past market data, without knowledge of the underlying statistical distribution, a portfolio selection has to be chosen for investment of the current capital in the available assets at the beginning of the new market period. The goal is to find a portfolio selection scheme such that the investor's wealth grows on the average as fast as by the optimum strategy based on the full knowledge of the underlying distribution. Nonparametric statistical methods allow to construct asymptotically optimal strategies for sequential investment in financial markets.

Throughout the paper it is assumed that the vectors of daily price relatives (return vectors) form a stationary and ergodic process. Then a log-optimal rate of growth exists and is achieved with probability one by a strategy based on the knowledge of the underlying distribution (Algoet and Cover [2]). Even in the more realistic case that only the past data are available, with no knowledge of the underlying distribution, selection schemes with log-optimal growth rate exist (Algoet [1]). Such investment schemes are called *universally consistent*. Györfi and Schäfer [7] constructed universally consistent schemes using histograms from nonparametric statistics, and Györfi, Lugosi, and Udina [6] using kernel estimates. In this paper a new universal strategy, called nearest neighbor strategy, is proposed which not only guarantees a log-optimal growth rate of capital for all stationary and

AMS 2000 subject classifications. Primary: 62G10 secondary: 62G05, 62L12, 62M20, 62P05, 91B28
Key words and phrases: sequential investment, universally consistent portfolios, nearest neighbor estimation.

ergodic markets, but also has a good finite-horizon performance in practice, and, as main novelty, is very robust in the sense that no design parameter tuning is needed to guarantee this good finite-horizon performance. The reason may be that nearest neighbor methods can be interpreted as well tractable kernel methods with data-based local choice of bandwidths. In [8], we present a numerical comparison of some empirical portfolio strategies for NYSE and currency exchange data, according to which the nearest neighbor based portfolio selection outperform the histogram and the kernel strategy.

The rest of the paper is organized as follows. In Section 2 the mathematical model is described. In Section 3 a nearest neighbor (NN) based nonparametric sequential investment strategy is introduced and its universal consistency is stated. The proof of this theoretical result (Theorem 3.1) is given in Section 4.

2 Mathematical model

The following stock market model has been investigated, among others, by Algoet and Cover [2]. Further references can be found in Györfi, Lugosi, and Udina [6]. Also the monographs of Cover and Thomas [4], and Luenberger [9] deal with the concept of log-optimality below.

Consider a market of d assets. The evolution of the market in time is represented by a sequence of return vectors $\mathbf{x}_1, \mathbf{x}_2, \dots$ with values in $\mathbb{R}_+^d$, where 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. That is, the j-th component $x_n^{(j)} \geq 0$ of $\mathbf{x}_n$ expresses the ratio of the closing and opening prices of asset j during the n-th trading period.

The investor is allowed to diversify his capital at the beginning of each trading period according to a portfolio vector $\mathbf{b} = (b^{(1)}, \dots b^{(d)})$. The j-th component $b^{(j)}$ of $\mathbf{b}$ denotes the proportion of the investor's capital invested in asset j. Throughout the paper we assume that the portfolio vector $\mathbf{b}$ has nonnegative components with $\sum_{j=1}^d b^{(j)} = 1$. It means that the investor neither consumes money nor deposits new money and that no transaction costs appear. The non-negativity of the components of $\mathbf{b}$ means that short selling and buying stocks on margin are not permitted. Denote by Δ_d the simplex of all vectors $\mathbf{b} \in \mathbb{R}_+^d$ with nonnegative components summing up to one.

Let S_0 denote the investor's initial capital. For the first trading period, the portfolio vector $\mathbf{b}_1$ is constant, usually $(1/d, ...1/d)$. Then at the end of the first trading period the investor's wealth becomes

$$S_1 = S_0 \sum_{j=1}^d b_1^{(j)} x_1^{(j)} = S_0 \langle \mathbf{b}_1 \, , \, \mathbf{x}_1 \rangle \, ,$$

where $\langle \cdot \, , \, \cdot \rangle$ denotes inner product. For $j \leq i$ we abbreviate by $\mathbf{x}_j^i$ the array of market vectors $(\mathbf{x}_j, \dots, \mathbf{x}_i)$. Let S_{n-1} be the wealth at the end of the $n-1$-th trading period, then it is the initial capital for the n-th trading period, for which the portfolio may depend on the past return vectors: $\mathbf{b}_n = \mathbf{b}_n(\mathbf{x}_1^{n-1})$. Therefore we get by induction that

$$S_n = S_{n-1} \left\langle \mathbf{b}_n(\mathbf{x}_1^{n-1}) \, , \, \mathbf{x}_n \right\rangle = S_0 \prod_{i=1}^{n} \left\langle \mathbf{b}_i(\mathbf{x}_1^{i-1}) \, , \, \mathbf{x}_i \right\rangle = S_0 \exp \left\{ \sum_{i=1}^{n} \log \left\langle \mathbf{b}_i(\mathbf{x}_1^{i-1}) \, , \, \mathbf{x}_i \right\rangle \right\}.$$

This may be written as $S_0 \exp\{nW_n(\mathbf{B})\}$, where $W_n(\mathbf{B})$ denotes the *average growth rate* of the investment strategy $\mathbf{B} = \{\mathbf{b}_n\}_{n=1}^{\infty}$:

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

The goal is to maximize the wealth $S_n = S_n(\mathbf{B})$ or, equivalently, maximize the average growth rate $W_n(\mathbf{B})$.
We assume that the sequence of return vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots$ are realizations of a random process $\mathbf{X}_1, \mathbf{X}_2, \ldots$ such that $\{\mathbf{X}_n\}_{-\infty}^{\infty}$ is a stationary and ergodic process. Besides a mild moment condition on the log-returns, no other distribution assumptions are made. According to Algoet and Cover [2], for the so-called conditional log-optimum investment strategy $\mathbf{B}^* = \{\mathbf{b}_n^*\}_{n=1}^{\infty}$ defined by

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

one has

$$\limsup_{n\to\infty} \frac{1}{n} \log \frac{S_n}{S_n^*} \le 0 \quad \text{almost surely,}$$

for each competitive strategy $\mathbf{B}$, where $S_n^* = S_n(\mathbf{B}^*)$ and $S_n = S_n(\mathbf{B})$. Furthermore

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

where

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

is the maximal possible growth rate of any investment strategy. The conditional log-optimum investment strategy $\mathbf{B}^*$ depends upon the distribution of the stationary and ergodic process $\{\mathbf{X}_n\}_{n=1}^{\infty}$. Surprisingly, according to Algoet [1], there exists investment strategy $\tilde{\mathbf{B}}$ on the basis of past return data such that

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

i.e., having the same best asymptotic growth rate as $\mathbf{B}^*$, for each stationary and ergodic processes $\{\mathbf{X}_n\}_{-\infty}^{\infty}$. Such investment strategies are called *universally consistent* with respect to a class of all stationary and ergodic processes.
The investment strategy of Györfi and Schäfer's [7] is, as Algoet's [1] strategy, histogram based. At a given time instant n one looks for correspondingly discretized k-tuples $\mathbf{x}_{n-k-j+1}^{n-j}$

of return vectors in the whole history of the market which are identical to the discretized return vectors $\mathbf{x}_{n-k}^{n-1}$. Such time instant $n-j$ is called matching time. Then design a fixed portfolio vector optimizing the return for the trading periods following each matching. For different integer $k > 0$ and histogram design parameter, mix these portfolios (see (3.3) below). Györfi, Lugosi, and Udina [6] modified this strategy by use of kernels ("moving-window"). In both papers, universal consistency of the strategies with respect to the class of all ergodic processes such that $\mathbb{E}\{|\log X^{(j)}|\} < \infty$, for $j = 1, 2, \dots d$, is shown.

3 Nearest neighbor based strategy

Define an infinite array of elementary strategies (the so-called experts) $\mathbf{H}^{(k,\ell)} = \{\mathbf{h}^{(k,\ell)}(\cdot)\}$, where k, ℓ are positive 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

$$\lim_{\ell\to\infty} p_\ell = 0. \tag{3.1}$$

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 ($n > k + \hat{\ell} + 1$) and for each vector $\mathbf{s} = \mathbf{s}_{-k}^{-1}$ of dimension kd introduce the set of the $\hat{\ell}$ nearest neighbor matches:

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

Define the portfolio vector by

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

We define the expert $\mathbf{h}^{(k,\ell)}$ by

$$\mathbf{h}^{(k,\ell)}(\mathbf{x}_1^{n-1}) = \mathbf{b}^{(k,\ell)}(\mathbf{x}_1^{n-1}, \mathbf{x}_{n-k}^{n-1}), \qquad n = 1, 2, \dots \tag{3.2}$$

That is, $\mathbf{h}_n^{(k,\ell)}$ is a fixed portfolio vector according to the return vectors following these nearest neighbors.

Now one forms a "mixture" of all experts using a positive probability distribution $\{q_{k,\ell}\}$ on the set of all pairs (k,ℓ) of positive integers (i. e. such that for all k, ℓ, $q_{k,\ell} > 0$). The investment strategy simply weights the experts $\mathbf{H}^{(k,\ell)}$ according to their past performances and $\{q_{k,\ell}\}$ such that after the nth trading period, the investor's capital becomes

$$S_n = \sum_{k,\ell} q_{k,\ell} S_n(\mathbf{H}^{(k,\ell)}), \tag{3.3}$$

where $S_n(\mathbf{H}^{(k,\ell)})$ is the capital accumulated after n periods when using the portfolio strategy $\mathbf{H}^{(k,\ell)}$ with initial capital $S_0 = 1$. This may easily be achieved by distributing the initial capital $S_0 = 1$ among all experts such that expert $\mathbf{H}^{(k,\ell)}$ trades with initial capital $q_{k,\ell} S_0$.

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.

Theorem 3.1 *Assume (3.1) and that a tie occurs with probability zero. The portfolio scheme* $\mathbf{B}^{\mathrm{NN}}$ *is universally consistent with respect to the class of all stationary and ergodic processes such that* $\mathbb{E}\{|\log X_0^{(j)}|\} < \infty$*, for* $j = 1, 2, \ldots d$.

4 Proofs

The proof of Theorem 3.1 uses the following three auxiliary results. The first is known as Breiman's generalized ergodic theorem [3].

Lemma 4.1 (BREIMAN [3]). *Let* $Z = \{Z_i\}_{-\infty}^{\infty}$ *be a stationary and ergodic process. For each positive integer* i*, let* T^i *denote the operator that shifts any sequence* $\{\ldots, z_{-1}, z_0, z_1, \ldots\}$ *by* i *digits to the left. Let* $f_1, f_2, \ldots$ *be a sequence of real-valued functions such that* $\lim_{n\to\infty} f_n(Z) = f(Z)$ *almost surely (a.s.) for some function* f*. Assume that* $\mathbb{E}\sup_n |f_n(Z)| < \infty$*. Then*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} f_i(T^i Z) = \mathbb{E} f(Z) \qquad \textit{(a.s.)}$$

The next two lemmas are due to Algoet and Cover [2, Theorems 3 and 4].

Lemma 4.2 (ALGOET AND COVER [2]). *Let* $\mathbf{Q}_{n\in\mathcal{N}\cup\{\infty\}}$ *be a family of regular probability distributions over the set* $\mathbb{R}_+^d$ *of all market vectors such that*

$$\mathbb{E}\{|\log U_n^{(j)}|\} < \infty$$

for any coordinate of a random market vector $\mathbf{U}_n = (U_n^{(1)}, \ldots, U_n^{(d)})$ *distributed according to* $\mathbf{Q}_n$*. In addition, let* $\mathbf{B}^*(\mathbf{Q}_n)$ *be the set of all log-optimal portfolios with respect to* $\mathbf{Q}_n$*, that is, the set of all portfolios* $\mathbf{b}$ *that attain* $\max_{\mathbf{b}\in\Delta_d} \mathbb{E}\{\log\langle \mathbf{b}\,,\, \mathbf{U}_n\rangle\}$*. Consider an arbitrary sequence* $\mathbf{b}_n \in \mathbf{B}^*(\mathbf{Q}_n)$*. If*

$$\mathbf{Q}_n \to \mathbf{Q}_\infty \quad \textit{weakly as } n \to \infty$$

then, for $\mathbf{Q}_\infty$*-almost all* $\mathbf{u}$*,*

$$\lim_{n\to\infty} \langle \mathbf{b}_n\,,\, \mathbf{u}\rangle \to \langle \mathbf{b}^*\,,\, \mathbf{u}\rangle$$

where the right-hand side is constant as $\mathbf{b}^*$ *ranges over* $\mathbf{B}^*(\mathbf{Q}_\infty)$.

Lemma 4.3 (ALGOET AND COVER [2]). *Let* $\mathbf{X}$ *be a random market vector defined on a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *satisfying* $\mathbb{E}\{|\log X^{(j)}|\} < \infty$*. If* $\mathcal{F}_k$ *is an increasing sequence of sub-*σ*-fields of* $\mathcal{F}$ *with*

$$\mathcal{F}_k \nearrow \mathcal{F}_\infty \subseteq \mathcal{F}\,,$$

then

$$\mathbb{E}\left\{\max_{\mathbf{b}} \mathbb{E}\left[\log\langle \mathbf{b}\,,\,\mathbf{X}\rangle\,|\mathcal{F}_k\right]\right\} \nearrow \mathbb{E}\left\{\max_{\mathbf{b}} \mathbb{E}\left[\log\langle \mathbf{b}\,,\,\mathbf{X}\rangle\,|\mathcal{F}_\infty\right]\right\}$$

as $k \to \infty$ *where the maximum on the left-hand side is taken over all* $\mathcal{F}_k$*-measurable functions* $\mathbf{b}$ *and the maximum on the right-hand side is taken over all* $\mathcal{F}_\infty$*-measurable functions* $\mathbf{b}$*.*

Proof of Theorem 3.1. The proof is based on techniques used in related prediction problems, see Györfi and Schäfer [7], Györfi, Lugosi, and Udina [6]. We need to prove that

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

Without loss of generality we may assume $S_0 = 1$, so that

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

Thus

$$\begin{aligned} \liminf_{n\to\infty} W_n(\mathbf{B}) &\geq \liminf_{n\to\infty}\sup_{k,\ell}\left(W_n(\mathbf{H}^{(k,\ell)}) + \frac{\log q_{k,\ell}}{n}\right) \\ &\geq \sup_{k,\ell}\liminf_{n\to\infty}\left(W_n(\mathbf{H}^{(k,\ell)}) + \frac{\log q_{k,\ell}}{n}\right) \\ &= \sup_{k,\ell}\liminf_{n\to\infty} W_n(\mathbf{H}^{(k,\ell)}). \end{aligned} \tag{4.1}$$

The simple argument above shows that the asymptotic rate of growth of the strategy $\mathbf{B}$ is at least as large as the supremum of the rates of growth of all elementary strategies $\mathbf{H}^{(k,\ell)}$. Thus, to estimate $\liminf_{n\to\infty} W_n(\mathbf{B})$, it suffices to investigate the performance of expert $\mathbf{H}^{(k,\ell)}$ on the stationary and ergodic market sequence $\mathbf{X}_0, \mathbf{X}_{-1}, \mathbf{X}_{-2}, \ldots$. First let the integers k, ℓ and the vector $\mathbf{s} = \mathbf{s}_{-k}^{-1} \in \mathbb{R}_+^{dk}$ be fixed.
Fix $p_\ell \in (0,1)$. Put

$$\tilde{\ell} = \lfloor p_\ell j \rfloor.$$

Let $S_{\mathbf{s},r}$ denote the closed sphere centered at $\mathbf{s}$ with radius r. Let the interval

$$R_{k,\ell}(\mathbf{s}) = [r'_{k,\ell}(\mathbf{s}), r''_{k,\ell}(\mathbf{s})]$$

be the set of values $r_{k,\ell}(\mathbf{s})$ such that

$$\mathbb{P}\{\mathbf{X}_{-k}^{-1} \in S_{\mathbf{s},r_{k,\ell}(\mathbf{s})}\} = p_\ell.$$

Since tie occurs with probability zero, such interval exists. Because of (3.1),

$$\lim_{\ell\to\infty} r''_{k,\ell}(\mathbf{s}) = 0.$$

For $j > k + \tilde{\ell} + 1$, introduce the set

$$\begin{aligned} J_{j,\mathbf{s}}^{(k,\ell)} &= \{i;\ -j+k+1 \le i \le 0 \text{ such that } \mathbf{X}_{i-k}^{i-1} \text{ is among the } \tilde{\ell} \text{ NNs of s} \\ &\quad \text{in } \mathbf{X}_{-k}^{-1}, \dots, \mathbf{X}_{-j+1}^{-j+k}\}. \end{aligned}$$

For all Borel sets A, let $\mathbb{P}_{j,\mathbf{s}}^{(k,\ell)}$ denote the (random) measure defined by

$$\mathbb{P}_{j,\mathbf{s}}^{(k,\ell)}\{A\} = \frac{\sum_{i\in J_{j,\mathbf{s}}^{(k,\ell)}} I_{\{\mathbf{X}_i\in A\}}}{|J_{j,\mathbf{s}}^{(k,\ell)}|}.$$

We will show that for all s, with probability one,

$$\mathbb{P}_{j,\mathbf{s}}^{(k,\ell)} \to \mathbb{P}_{\mathbf{X}_0 |\, \|\mathbf{X}_{-k}^{-1}-\mathbf{s}\| \le r_{k,\ell}(\mathbf{s})} = \mathbb{P}_{\mathbf{s}}^{*(k,\ell)} \tag{4.2}$$

with arbitrary $r_{k,\ell}(\mathbf{s}) \in R_{k,\ell}(\mathbf{s})$, as $j \to \infty$ in terms of the weak convergence. To see this, let f be a bounded continuous function defined on $\mathbb{R}_+^d$. Then we prove that

$$\int f(\mathbf{x})\mathbb{P}_{j,\mathbf{s}}^{(k,\ell)}(d\mathbf{x}) \quad\to\quad \int f(\mathbf{x})\mathbb{P}_{\mathbf{s}}^{*(k,\ell)}(dx) \quad \text{almost surely, as } j \to \infty.$$

Notice that

$$\mathbf{X}_{i-k}^{i-1} \text{ is among the } \tilde{\ell} \text{ NNs of s in } \mathbf{X}_{-k}^{-1}, \dots, \mathbf{X}_{-j+1}^{-j+k}$$

if and only if

$$\|\mathbf{X}_{i-k}^{i-1} - \mathbf{s}\| \le \|(\text{the } \tilde{\ell}\text{-th NN of s in } \mathbf{X}_{-k}^{-1}, \dots, \mathbf{X}_{-j+1}^{-j+k}) - \mathbf{s}\|.$$

Moreover

$$\|(\text{the } \tilde{\ell}\text{-th NN of s in } \mathbf{X}_{-k}^{-1}, \dots, \mathbf{X}_{-j+1}^{-j+k}) - \mathbf{s}\|$$

tends to the set $R_{k,\ell}(\mathbf{s})$ ($j \to \infty$) a.s. by the ergodic theorem in context of empirical measures, thus almost uniformly by Egorov's theorem. Therefore, for arbitrary $\epsilon > 0$ and $\delta > 0$ an i_0 exists such that with probability $\ge 1-\delta$ for $-i \ge i_0$ the following implications hold:

$$\|\mathbf{X}_{i-k}^{i-1} - \mathbf{s}\| \le r'_{k,\ell}(\mathbf{s}) - \epsilon$$

implies that

$$\mathbf{X}_{i-k}^{i-1} \text{ is among the } \tilde{\ell} \text{ NNs of s in } \mathbf{X}_{-k}^{-1}, \dots, \mathbf{X}_{-j+1}^{-j+k},$$

which implies that

$$\|\mathbf{X}_{i-k}^{i-1}-\mathbf{s}\| \le r''_{k,\ell}(\mathbf{s})+\epsilon.$$

Introduce the sets

$$\bar{J}_{j,\mathbf{s}}^{(k,\ell)} = \{i;\ -j+k+1\le i\le 0,\ \|\mathbf{X}_{i-k}^{i-1}-\mathbf{s}\|\le r'_{k,\ell}(\mathbf{s})-\epsilon\}$$

and

$$\tilde{J}_{j,\mathbf{s}}^{(k,\ell)} = \{i;\ -j+k+1\le i\le 0,\ \|\mathbf{X}_{i-k}^{i-1}-\mathbf{s}\|\le r''_{k,\ell}(\mathbf{s})+\epsilon\}.$$

Without loss of generality, assume that $f\ge 0$. The ergodic theorem implies that

$$\lim_{j\to\infty}\frac{\frac{1}{j-k}\sum_{i\in\bar{J}_{j,\mathbf{s}}^{(k,\ell)}} f(\mathbf{X}_i)}{\frac{1}{j-k}|\tilde{J}_{j,\mathbf{s}}^{(k,\ell)}|}=\frac{\mathbb{E}\{f(\mathbf{X}_0)I_{\{\|\mathbf{X}_{-k}^{-1}-\mathbf{s}\|\le r'_{k,\ell}(\mathbf{s})-\epsilon\}}\}}{\mathbb{P}\{\|\mathbf{X}_{-k}^{-1}-\mathbf{s}\|\le r''_{k,\ell}(\mathbf{s})+\epsilon\}}$$

a.s. and with probability $\ge 1-\delta$

$$\begin{aligned}\frac{\mathbb{E}\{f(\mathbf{X}_0)I_{\{\|\mathbf{X}_{-k}^{-1}-\mathbf{s}\|\le r'_{k,\ell}(\mathbf{s})-\epsilon\}}\}}{\mathbb{P}\{\|\mathbf{X}_{-k}^{-1}-\mathbf{s}\|\le r''_{k,\ell}(\mathbf{s})+\epsilon\}} &\le \liminf_{j\to\infty}\frac{\frac{1}{j-k}\sum_{i\in J_{j,\mathbf{s}}^{(k,\ell)}} f(\mathbf{X}_i)}{\frac{1}{j-k}|J_{j,\mathbf{s}}^{(k,\ell)}|}\\ &\le \limsup_{j\to\infty}\frac{\frac{1}{j-k}\sum_{i\in J_{j,\mathbf{s}}^{(k,\ell)}} f(\mathbf{X}_i)}{\frac{1}{j-k}|J_{j,\mathbf{s}}^{(k,\ell)}|}\\ &\le \lim_{j\to\infty}\frac{\frac{1}{j-k}\sum_{i\in\tilde{J}_{j,\mathbf{s}}^{(k,\ell)}} f(\mathbf{X}_i)}{\frac{1}{j-k}|\bar{J}_{j,\mathbf{s}}^{(k,\ell)}|}\\ &= \frac{\mathbb{E}\{f(\mathbf{X}_0)I_{\{\|\mathbf{X}_{-k}^{-1}-\mathbf{s}\|\le r''_{k,\ell}(\mathbf{s})+\epsilon\}}\}}{\mathbb{P}\{\|\mathbf{X}_{-k}^{-1}-\mathbf{s}\|\le r'_{k,\ell}(\mathbf{s})-\epsilon\}}\end{aligned}$$

a.s. by ergodic theorem. $\epsilon\to 0$ yields that with probability $\ge 1-\delta$

$$\lim_{j\to\infty}\frac{\frac{1}{j-k}\sum_{i\in J_{j,\mathbf{s}}^{(k,\ell)}} f(\mathbf{X}_i)}{\frac{1}{j-k}|J_{j,\mathbf{s}}^{(k,\ell)}|}=\frac{\mathbb{E}\{f(\mathbf{X}_0)I_{\{\|\mathbf{X}_{-k}^{-1}-\mathbf{s}\|\le r_{k,\ell}(\mathbf{s})\}}\}}{\mathbb{P}\{\|\mathbf{X}_{-k}^{-1}-\mathbf{s}\|\le r_{k,\ell}(\mathbf{s})\}}$$

for arbitrary $r_{k,\ell}(\mathbf{s})\in R_{k,\ell}(\mathbf{s})$. Thus a.s.

$$\lim_{j\to\infty}\frac{\frac{1}{j-k}\sum_{i\in J_{j,\mathbf{s}}^{(k,\ell)}} f(\mathbf{X}_i)}{\frac{1}{j-k}|J_{j,\mathbf{s}}^{(k,\ell)}|}=\mathbb{E}\{f(\mathbf{X}_0)\mid\ \|\mathbf{X}_{-k}^{-1}-\mathbf{s}\|\le r_{k,\ell}(\mathbf{s})\},$$

and (4.2) is proved. Recall that by definition, $\mathbf{b}^{(k,\ell)}(\mathbf{X}_{1-j}^{-1},\mathbf{s})$ is a log-optimal portfolio with respect to the probability measure $\mathbb{P}_{j,\mathbf{s}}^{(k,\ell)}$. Let $\mathbf{b}^*_{k,\ell}(\mathbf{s})$ denote a log-optimal portfolio with respect to the limit distribution $\mathbb{P}_{\mathbf{s}}^{*(k,\ell)}$. Then, using Lemma 4.2, we infer from (4.2) that, as j tends to infinity, we have the almost sure convergence

$$\lim_{j\to\infty}\left\langle \mathbf{b}^{(k,\ell)}(\mathbf{X}_{1-j}^{-1},\mathbf{s})\,,\ \mathbf{x}_0\right\rangle=\left\langle \mathbf{b}^*_{k,\ell}(\mathbf{s})\,,\ \mathbf{x}_0\right\rangle$$

for $\mathbb{P}_{\mathbf{s}}^{*(k,\ell)}$-almost all $\mathbf{x}_0$ and hence for $\mathbb{P}_{\mathbf{X}_0}$-almost all $\mathbf{x}_0$. Since $\mathbf{s}$ was arbitrary, we obtain

$$\lim_{j\to\infty}\left\langle \mathbf{b}^{(k,\ell)}(\mathbf{X}_{1-j}^{-1},\mathbf{X}_{-k}^{-1})\,,\,\mathbf{x}_0\right\rangle = \left\langle b_{k,\ell}^{*}(\mathbf{X}_{-k}^{-1})\,,\,\mathbf{x}_0\right\rangle \quad \text{(a.s.)} \tag{4.3}$$

Next we apply Lemma 4.1 for the function

$$f_i(\mathbf{x}_{-\infty}^{\infty}) = \log\left\langle \mathbf{h}^{(k,\ell)}(\mathbf{x}_{1-i}^{-1})\,,\,\mathbf{x}_0\right\rangle = \log\left\langle \mathbf{b}^{(k,\ell)}(\mathbf{x}_{1-i}^{-1},\mathbf{x}_{-k}^{-1})\,,\,\mathbf{x}_0\right\rangle$$

defined on $\mathbf{x}_{-\infty}^{\infty} = (\ldots,\mathbf{x}_{-1},\mathbf{x}_0,\mathbf{x}_1,\ldots)$. Note that

$$f_i(\mathbf{X}_{-\infty}^{\infty}) = \left|\log\left\langle \mathbf{h}^{(k,\ell)}(\mathbf{X}_{1-i}^{-1})\,,\,\mathbf{X}_0\right\rangle\right| \le \sum_{j=1}^{d}\left|\log X_0^{(j)}\right| \,,$$

which has finite expectation, and

$$f_i(\mathbf{X}_{-\infty}^{\infty}) \to \left\langle \mathbf{b}_{k,\ell}^{*}(\mathbf{X}_{-k}^{-1})\,,\,\mathbf{X}_0\right\rangle$$

almost surely as $i\to\infty$, by (4.3). As $n\to\infty$, Lemma 4.1 yields

$$\begin{aligned} W_n(\mathbf{H}^{(k,\ell)}) &= \frac{1}{n}\sum_{i=1}^{n} f_i(T^i\mathbf{X}_{-\infty}^{\infty}) \\ &= \frac{1}{n}\sum_{i=1}^{n}\log\left\langle \mathbf{h}^{(k,\ell)}(\mathbf{X}_1^{i-1})\,,\,\mathbf{X}_i\right\rangle \\ &\to \mathbb{E}\left\{\log\left\langle \mathbf{b}_{k,\ell}^{*}(\mathbf{X}_{-k}^{-1})\,,\,\mathbf{X}_0\right\rangle\right\} \\ &\stackrel{\text{def}}{=} \epsilon_{k,\ell} \quad \text{(a.s.)} \end{aligned}$$

Therefore, by (4.1) we have

$$\liminf_{n\to\infty} W_n(\mathbf{B}) \ge \sup_{k,\ell}\epsilon_{k,\ell} \ge \sup_{k}\liminf_{\ell}\epsilon_{k,\ell} \qquad \text{(a.s.)}$$

and it suffices to show that the right-hand side is at least W^*. The rest of the proof is similar to the end of the proof in [6], so the reader may skip it.
To this end, define, for Borel sets $A, B \subset \mathbb{R}_+^d$,

$$m_A(\mathbf{z}) = \mathbb{P}\{\mathbf{X}_0 \in A | \mathbf{X}_{-k}^{-1} = \mathbf{z}\}$$

and

$$\mu_k(B) = \mathbb{P}\{\mathbf{X}_{-k}^{-1} \in B\}.$$

Then for any $\mathbf{s} \in \text{support}(\mu_k)$, and for all A,

$$\begin{aligned} \mathbb{P}_{\mathbf{s}}^{*(k,\ell)}(A) &= \mathbb{P}\left\{\mathbf{X}_0 \in A \,\middle|\, \|\mathbf{X}_{-k}^{-1} - \mathbf{s}\| \le r_{k,\ell}(\mathbf{s})\right\} \\ &= \frac{\mathbb{P}\{\mathbf{X}_0 \in A, \|\mathbf{X}_{-k}^{-1} - \mathbf{s}\| \le r_{k,\ell}(\mathbf{s})\}}{\mathbb{P}\{\|\mathbf{X}_{-k}^{-1} - \mathbf{s}\| \le r_{k,\ell}(\mathbf{s})\}} \\ &= \frac{1}{\mu_k(S_{\mathbf{s},r_{k,\ell}(\mathbf{s})})}\int_{S_{\mathbf{s},r_{k,\ell}(\mathbf{s})}} m_A(\mathbf{z})\mu_k(d\mathbf{z}) \\ &\to m_A(\mathbf{s}) = \mathbb{P}\{\mathbf{X}_0 \in A | \mathbf{X}_{-k}^{-1} = \mathbf{s}\} \end{aligned}$$

as $\ell \to \infty$ and for μ_k-almost all s by the Lebesgue density theorem (see [5, Lemma 24.5]), and therefore

$$\mathbb{P}^{*(k,\ell)}_{\mathbf{X}_{-k}^{-1}}(A) \to \mathbb{P}\{\mathbf{X}_0 \in A | \mathbf{X}_{-k}^{-1}\}$$

as $\ell \to \infty$ for all A.
Thus, using Lemma 4.2 again, we have

$$\begin{aligned}
\liminf_{\ell} \epsilon_{k,\ell} &= \lim_{\ell} \epsilon_{k,\ell} \\
&= \mathbb{E}\left\{\log \left\langle \mathbf{b}_k^*(\mathbf{X}_{-k}^{-1}) \,,\, \mathbf{X}_0 \right\rangle\right\} \\
&\quad \text{(where } \mathbf{b}_k^*(\cdot) \text{ is the log-optimum portfolio with respect} \\
&\quad \text{to the conditional probability } \mathbb{P}\{X_0 \in A | \mathbf{X}_{-k}^{-1}\}) \\
&= \mathbb{E}\left\{\mathbb{E}\left\{ \log \left\langle \mathbf{b}_k^*(\mathbf{X}_{-k}^{-1}) \,,\, \mathbf{X}_0 \right\rangle \middle| \, \mathbf{X}_{-k}^{-1} \right\}\right\} \\
&= \mathbb{E}\left\{ \max_{\mathbf{b}(\cdot)} \mathbb{E}\left\{ \log \left\langle \mathbf{b}(\mathbf{X}_{-k}^{-1}) \,,\, \mathbf{X}_0 \right\rangle \middle| \, \mathbf{X}_{-k}^{-1} \right\}\right\} \\
&\stackrel{\text{def}}{=} \epsilon_k^* \,.
\end{aligned}$$

To finish the proof we appeal to the sub-martingale convergence theorem. First note that the sequence

$$Y_k \stackrel{\text{def}}{=} \mathbb{E}\left\{ \log \left\langle \mathbf{b}_k^*(\mathbf{X}_{-k}^{-1}) \,,\, \mathbf{X}_0 \right\rangle \middle| \, \mathbf{X}_{-k}^{-1} \right\} = \max_{\mathbf{b}(\cdot)} \mathbb{E}\left\{ \log \left\langle \mathbf{b}(\mathbf{X}_{-k}^{-1}) \,,\, \mathbf{X}_0 \right\rangle \middle| \, \mathbf{X}_{-k}^{-1} \right\}$$

of random variables forms a sub-martingale, that is, $\mathbb{E}\left\{Y_{k+1} | \mathbf{X}_{-k}^{-1}\right\} \geq Y_k$. To see this, note that

$$\begin{aligned}
\mathbb{E}\left\{Y_{k+1} | \mathbf{X}_{-k}^{-1}\right\} &= \mathbb{E}\left\{\mathbb{E}\left\{ \log \left\langle \mathbf{b}_{k+1}^*(\mathbf{X}_{-k-1}^{-1}) \,,\, \mathbf{X}_0 \right\rangle \middle| \, \mathbf{X}_{-k-1}^{-1} \right\} \middle| \, \mathbf{X}_{-k}^{-1} \right\} \\
&\geq \mathbb{E}\left\{\mathbb{E}\left\{ \log \left\langle \mathbf{b}_{k}^*(\mathbf{X}_{-k}^{-1}) \,,\, \mathbf{X}_0 \right\rangle \middle| \, \mathbf{X}_{-k-1}^{-1} \right\} \middle| \, \mathbf{X}_{-k}^{-1} \right\} \\
&= \mathbb{E}\left\{ \log \left\langle \mathbf{b}_{k}^*(\mathbf{X}_{-k}^{-1}) \,,\, \mathbf{X}_0 \right\rangle \middle| \, \mathbf{X}_{-k-1}^{-1} \right\} \\
&= Y_k \,.
\end{aligned}$$

This sequence is bounded by

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

which has a finite expectation. The sub-martingale convergence theorem (see, e.g., Stout [10]) implies that this sub-martingale is convergent almost surely, and $\sup_k \epsilon_k^*$ is finite. In particular, by the submartingale property, ϵ_k^* is a bounded increasing sequence, so that

$$\sup_k \epsilon_k^* = \lim_{k\to\infty} \epsilon_k^* \,.$$

Applying Lemma 4.3 with the σ-algebras

$$\sigma\left(\mathbf{X}_{-k}^{-1}\right) \nearrow \sigma\left(\mathbf{X}_{-\infty}^{-1}\right)$$

yields

$$\begin{aligned}
\sup_k \epsilon_k^* &= \lim_{k\to\infty} \mathbb{E}\left\{ \max_{\mathbf{b}(\cdot)} \mathbb{E}\left\{ \log \left\langle \mathbf{b}(\mathbf{X}_{-k}^{-1})\,,\, \mathbf{X}_0 \right\rangle \middle| \, \mathbf{X}_{-k}^{-1} \right\} \right\} \\
&= \mathbb{E}\left\{ \max_{\mathbf{b}(\cdot)} \mathbb{E}\left\{ \log \left\langle \mathbf{b}(\mathbf{X}_{-\infty}^{-1})\,,\, \mathbf{X}_0 \right\rangle \middle| \, \mathbf{X}_{-\infty}^{-1} \right\} \right\} \\
&= W^*
\end{aligned}$$

and the proof of the theorem is finished. ∎

Acknowledgments

The first author acknowledges the support of the Computer and Automation Research Institute and the Research Group of Informatics and Electronics of the Hungarian Academy of Sciences. The work of the second author was supported by the Spanish Ministry of Science and Technology and FEDER, grant MTM2006-05650 We thank Michael Greenacre for his careful and annotated reading.

References

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

[2] P. Algoet and T. Cover. Asymptotic optimality asymptotic equipartition properties of log-optimum investments. *Annals of Probability* **16** (1988) 876–898.

[3] L. Breiman. The individual ergodic theorem of information theory. *Annals of Mathematical Statistics* **28** (1957) 809–811. Correction. *Annals of Mathematical Statistics*, **31** (1960) 809–810.

[4] T. Cover and J. Thomas. *Elements of Information Theory*. John Wiley and Sons, New York (1991).

[5] L. Györfi, M. Kohler, A. Krzyżak, and H. Walk. *A Distribution-Free Theory of Nonparametric Regression*. Springer, New York (2002).

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

[7] L. Györfi and D. Schäfer. 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*, pages 339–354. IOS Press, NATO Science Series (2003).

[8] L. Györfi, F. Udina, and H. Walk. Experiments on universal portfolio selection using data from real markets. (submitted). http://tukey.upf.es/papers/NNexp.pdf.

[9] D. G. Luenberger *Investment Science*. Oxford University Press, (1998).

[10] W. F. Stout. *Almost Sure Convergence*. Academic Press, New York (1974).

László Györfi
Department of Computer Science and Information Theory,
Budapest University of Technology and Economics,
1521 Stoczek u. 2, Budapest, Hungary.
e-mail: gyorfi@szit.bme.hu

Frederic Udina
Department of Economics and Business,
Universitat Pompeu Fabra,
Ramon Trias Fargas 25-27, 08005 Barcelona, Spain.
e-mail: udina@upf.es

Harro Walk
Institute of Stochastics and Applications,
Universität Stuttgart,
Pfaffenwaldring 57, D-70569 Stuttgart, Germany.
e-mail: walk@mathematik.uni-stuttgart.de