Universal Algorithm for Trading in Stock Market Based on the Method of Calibration

Vladimir V'yugin

Institute for Information Transmission Problems, Russian Academy of Sciences,
Bol'shoi Karetnyi per. 19, Moscow GSP-4, 127994, Russia
Laboratory of Structural Methods of Data Analysis in Predictive Modeling
at the Moscow Institute of Physics and Technology
email: vyugin@iitp.ru

Abstract. We present a universal method for algorithmic trading in Stock Market which performs asymptotically at least as well as any stationary trading strategy that computes the investment at each step using a continuous function of the side information. In the process of the game, a trader makes decisions using predictions computed by a randomized well-calibrated algorithm. We use Dawid's notion of calibration with more general checking rules and some modification of Kakade and Foster's randomized rounding algorithm for computing the well-calibrated forecasts. The method of randomized calibration is combined with Vovk's method of defensive forecasting in RKHS. Unlike in statistical theory, no stochastic assumptions are made about the stock prices.

1 Introduction

Predicting sequences is the key problem of machine learning and statistics. The learning process proceeds as follows: observing a sequence of outcomes given on-line a forecaster assigns a subjective estimate to future outcomes.

A minimal requirement for testing any prediction algorithm is that it should be calibrated (see Dawid [4]). Dawid gave an informal explanation of calibration for binary outcomes as follows. Let a binary sequence $\omega_1, \omega_2, \dots, \omega_{n-1}$ of outcomes be observed by a forecaster whose task is to give a probability p_n of a future event $\omega_n = 1$. In a typical example, p_n is interpreted as a probability that it will rain. Forecaster is said to be well-calibrated if it rains as often as he leads us to expect. It should rain about 80% of the days for which $p_n = 0.8$, and so on.

A more precise definition is as follows. Let $I(p)$ denote the characteristic function of a subinterval $I \subseteq [0,1]$, i.e., $I(p) = 1$ if $p \in I$, and $I(p) = 0$, otherwise. An infinite sequence of forecasts $p_1, p_2, \dots$ is calibrated for an infinite binary sequence of outcomes $\omega_1\omega_2\dots$ if for characteristic function $I(p)$ of any subinterval of $[0,1]$ the calibration error tends to zero, i.e.,

$$\frac{1}{n}\sum_{i=1}^{n} I(p_i)(\omega_i - p_i) \to 0$$

as $n \to \infty$. The indicator function $I(p_i)$ determines some "checking rule" which selects indices i where we compute the deviation between forecasts p_i and outcomes ω_i.

If the weather acts adversatively, then Oakes [10] shows that any deterministic forecasting algorithm will not always be calibrated.

Foster and Vohra [5] show that calibration is almost surely guaranteed with a randomizing forecasting rule, i.e., where the forecasts p_i are chosen using internal randomization and the forecasts are hidden from the weather until weather makes its decision whether to rain or not. Mannor and Stoltz [9] obtained the upper bound for the rate of convergence of the calibration error. Foster et al. [7] obtained convergence rates which depend on the complexity of the class of checking rules without providing a computationally efficient forecasting algorithms. Vovk ([14], [15], [16]) developed the method of calibration for the case of general RKHS and Banach spaces. Vovk called his method defensive forecasting.

Kakade and Foster and others considered a finite outcome space and a probability distribution as the forecast. In this paper, the outcomes ω_i are real numbers from the unit interval $[0, 1]$ and the forecast p_i is a single real number (which can be an output of a random variable). This setting is closely related to Vovk's defensive forecasting approach.

Several applications of well-calibrated forecasting have been proposed, including convergence to correlated equilibrium (see Foster and Vohra [6]), recovering unknown functional dependencies (see Vovk ([14], [15], [16]) and predictions with expert advice (see Vovk [16]).

In this paper we present a new application of the method of calibration to computational finance. We develop an algorithmic trading strategy that is in some sense always guaranteed to perform well. In competitive analysis, the performance of an algorithm is measured to any trading algorithm from a broad class. Given a particular performance measure, an adaptive algorithm is strongly competitive with a benchmark class of trading algorithms if it achieves the maximum possible regret over all input sequences. Unlike statistical theory, no stochastic assumptions are made about the stock prices.

This line of research in finance was pioneered by Cover (see Cover [1], Cover and Ordentlich[2]) who designed universal portfolio selection algorithms that can provably do well (in terms of their total return) with respect to some benchmark algorithms. Such algorithms are called *universal algorithms*.

We construct in Theorem 1 a universal strategy for algorithmic trading in *Stock Market* which performs asymptotically at least as well as any not "too complex" trading strategy. By "performance" we mean return per unit of currency on an investment.

The empirical results obtained on historical markets provide strong evidence that this type of technical trading can beat some generally accepted trading strategies if transaction costs are ignored. Results of numerical experiments are presented in V'yugin and Trunov [17].

2 Main result

Assume that outcomes (real numbers) $S_1, S_2, \ldots$ that are interpreted as prices of a stock are given online. We assume that they are bounded and scaled such that $0 \le S_i \le 1$ for all i.

We present the process of trading in a stock market in the form of the protocol of a game with players called *traders* (see Fig. 1). We distinguish among *Trader M* which uses a randomized trading strategy and a pool of *Traders D* using stationary trading strategies: any such strategy D is a continuous real function defined on $[0, 1]$.

At the beginning of each step i, traders are given some data z_i relevant to predicting a future price S_i of the stock. We call z_i a *signal* or a piece of *side information*. The real number z_i belongs to $[0, 1]$ and can encode any numerical information. For example, it can even be the future price S_i. There is a restriction for *Traders D*: the functions D must be continuous.

In general, under a strategy we mean an algorithm (possibly randomized) which at each step i of the game outputs the number of units of the financial instrument that you want to buy (if the number is positive or equal to zero) or sell (if it is negative).[1] For *Trader M*, this number is a value of the random variable $\tilde{M}_i$ that is the output of a randomizing algorithm, and for any *Trader D* from the pool, this number equals $D(z_i)$, where D is a continuous function. We suppose that traders can borrow money for buying shares and can incur debt.

At any step i, any *Trader D* from the pool uses only information z_i; he buys (or sells) $D(z_i)$ units of shares. A strategy of this type is called *stationary*.

For *Trader M*, this game is a game with perfect information. For defining the random variable $\tilde{M}_i$, *Trader M* may use all values of S_{j-1} and z_j for $j \le i$, as well as their randomized values.

We use a specific randomization method for real numbers from the unit interval. Given a positive integer number K, divide the interval $[0, 1]$ into subintervals of length $\Delta = 1/K$ with rational endpoints $v_i = i\Delta$, where $i = 0, 1, \ldots, K$. Let V be the set of these points. Any number $p \in [0, 1]$ can be represented as a linear combination of two endpoints of the subinterval containing p: $p = \sum\limits_{v \in V} w_v(p)v = w_{v_{i-1}}(p)v_{i-1} + w_{v_i}(p)v_i$. Define $w_v(p) = 0$ for all other $v \in V$. Let $\tilde{p}$ be v_{i-1} with probability $w_{v_{i-1}}(p)$ or v_i with probability $w_{v_i}(p)$.

We emphasize that in the protocol presented on Fig 1 a value of the random variable $\tilde{M}_i$ is hidden from *Stock Market* when it announces the price S_i; it can only use probabilities of events $\tilde{M}_i = 1$ and $\tilde{M}_i = -1$.

Trader M can buy or sell only one share of the stock. Therefore, in order to compare the performance of the traders we have to standardize the strategies $D(x)$ of *Traders D* from the pool. Recall the norm $\|D\|_\infty = \sup\limits_{0 \le x \le 1} |D(x)|$, where D is a continuous function. We will use $\|D\|_+ = \max\{1, \|D\|_\infty\}$ as a normalization factor.

[1] We assume that the number of units of a financial instrument purchasing by traders may take any real value.

Define $\mathcal{K}_0^D = 0$ and $\mathcal{K}_0^M = 0$.
FOR $i = 1, 2 \ldots$
Stock Market announces a signal $z_i \in [0, 1]$.
Define a precision of random rounding: $\Delta = \Delta_s$, where $n_s < i \le n_{s+1}$ (The sequences n_s and Δ_s, $s = 1, 2, \ldots$, are defined after (22) in the proof of Theorem 2 below).
Compute a forecast p_i using algorithm presented on Fig. 2 with parameter Δ.
Randomize the forecast: $\tilde{p}_i$.
Randomize the past price of the stock: $\tilde{S}_{i-1}$.
Trader M buys $\tilde{M}_i$ shares of the stock by S_{i-1} each, where

$$\tilde{M}_i = \begin{cases} 1 \text{ if } \tilde{p}_i > \tilde{S}_{i-1}, \\ -1 \text{ otherwise.} \end{cases}$$

Trader D buys $D(z_i)$ shares of the stock by S_{i-1} each, where D is an arbitrary continuous function on $[0, 1]$.
Stock Market announces the price S_i of the stock.
Trader M sells $\tilde{M}_i$ shares of the stock by S_i each and updates his cumulative gain: $\mathcal{K}_i^M = \mathcal{K}_{i-1}^M + \tilde{M}_i(S_i - S_{i-1})$.
Trader D sells $D(z_i)$ shares of the stock by S_i each and updates his cumulative gain: $\mathcal{K}_i^D = \mathcal{K}_{i-1}^D + D(z_i)(S_i - S_{i-1})$.
ENDFOR

Fig. 1. Protocol of trading game

The main result of this paper is presented in the following theorem, which says that, with probability 1, the average gain of the universal trading strategy is asymptotically not less than the average gain of any stationary trading strategy from one share of the stock.

Theorem 1. *A randomized algorithm for computing forecasts can be constructed such that for any continuous function D the inequality*

$$\liminf_{n \to \infty} \frac{1}{n} \left(\mathcal{K}_n^M - \|D\|_+^{-1} \mathcal{K}_n^D \right) \ge 0 \tag{1}$$

holds almost surely with respect to the probability distribution generated by the corresponding randomization.

We call any strategy M satisfying (1) *universal*. Note that (1) holds for all continuous D if and only if the inequality:

$$\liminf_{n \to \infty} \frac{1}{n} \left(\mathcal{K}_n^M - \mathcal{K}_n^D \right) \ge 0$$

holds for all continuous D such that $\|D\|_\infty \le 1$.

Since the condition (1) holds for trivial strategy: $D(z) = 0$ for all z, the *Trader's M* strategy is *asymptotically non-risk.*

Corollary 1. *Universal strategy is asymptotically non-risk:*

$$\liminf_{n \to \infty} \frac{\mathcal{K}_n^M}{n} \ge 0 \textit{ almost surely.}$$

A proof of Theorem 1 is given in Section 5, where we construct the corresponding optimal trading strategy based on the well-calibrated forecasts defined in Section 4. In Section 3 we define benchmark classes RKHS.

3 Benchmark class: RKHS

First, we compete the universal trading strategy with stationary trading strategies from a benchmark class $\mathcal{F}$ called RKHS. After that, we approximate any continuous function D using the functions $f \in \mathcal{F}$.

By a kernel function on a set X we mean any function $K(x, y)$ which can be represented as a dot product $K(x, y) = (\Phi(x) \cdot \Phi(y))$, where Φ is a mapping from X to some Hilbert feature space.

The reproducing kernels are of special interest. A Hilbert space $\mathcal{F}$ of real-valued functions on a compact metric space X is called RKHS (Reproducing Kernel Hilbert Space) on X if the evaluation functional $f \to f(x)$ is continuous for each $x \in X$. Let $\|\cdot\|_{\mathcal{F}}$ be the norm in $\mathcal{F}$ and $c_{\mathcal{F}}(x) = \sup\limits_{\|f\|_{\mathcal{F}} \le 1} |f(x)|$. The embedding constant of $\mathcal{F}$ is defined: $c_{\mathcal{F}} = \sup\limits_{x} c_{\mathcal{F}}(x)$. We consider RKHS $\mathcal{F}$ with $c_{\mathcal{F}} < \infty$.

An example of RKHS is the Sobolev space $\mathcal{F} = H^1([0, 1])$, which consists of absolutely continuous functions $f : [0, 1] \to \mathcal{R}$ with $\|f\|_{\mathcal{F}} \le 1$, where $\|f\|_{\mathcal{F}} = \sqrt{\int_0^1 (f(t))^2 dt + \int_0^1 (f'(t))^2 dt}$. For this space, $c_{\mathcal{F}} = \sqrt{\coth 1}$ (see Vovk [14]).

Let $\mathcal{F}$ be an RKHS on X with the dot product $(f \cdot g)$ for $f, g \in \mathcal{F}$. By Riesz–Fisher theorem, for each $x \in X$ there exists $k_x \in \mathcal{F}$ such that $f(x) = (k_x \cdot f)$.

The reproducing kernel is defined $K(x, y) = (k_x \cdot k_y)$. The main properties of the kernel: 1) $K(x, y) = K(y, x)$ for all $x, y \in X$ (symmetry property); 2) $\sum\limits_{i,j=1}^{k} \alpha_i \alpha_j K(x_i, x_j) \ge 0$ for all k, for all $x_i \in X$, and for all real numbers α_i, where $i = 1, \dots, k$ (positive semidefinite property).

Conversely, kernels define RKHS: any symmetric, positive semidefinite kernel function $K(x, y)$ on X defines some canonical RKHS $\mathcal{F}$ and a mapping $\Phi : X \to \mathcal{F}$ such that $K(x, y) = (\Phi(x) \cdot \Phi(y))$. Also, $c_{\mathcal{F}}(x) = \|k_x\|_{\mathcal{F}} = \|\Phi(x)\|_{\mathcal{F}}$. The mapping $\Phi(x)$ is also called "feature map".

For Sobolev space $H^1([0, 1])$, the reproducing kernel is
$K(t, t') = (\cosh \min(t, t') \cosh \min(1 - t, 1 - t'))/\sinh 1$ (see Vovk[14]).

Well known examples of kernels on $X = [0, 1]$: Gaussian kernel $K(x, y) = \exp\{-\frac{(\bar{x}-\bar{y})^2}{\sigma^2}\}$; $K(t, t') = \cos(2\pi(t - t'))$, $t, t' \in [0, 1]$.

Other examples and details of the kernel theory see in Cristianini and Shawe-Taylor [3], Smola and Scholkopf [11].

4 Well-calibrated forecasting with side information

In this section we present a randomized algorithm for computing well-calibrated forecasts using a side information.

We use tests of calibration or *checking rules* of general type. For any subset $R \subseteq [0,1]^2 = [0,1] \times [0,1]$, define the checking rule $I_R(p,x) = 1$ if $(p,x) \in R$ and $I_R(p,x) = 0$ otherwise. In Section 4 we set $R = \{(p,y) : p > y\}$ or $R = \{(p,y) : p \le y\}$, where $p, y \in [0,1]$.

In the prediction protocol defined on Fig 1, let $S_1, S_2, \ldots$ be a sequence of outcomes and $z_1, z_2, \ldots$ be the corresponding sequences of signals given online.

Let also, $\mathcal{F}$ be an RKHS on $[0,1]$ with a kernel $K_2(z,z')$ and a finite embedding constant $c_{\mathcal{F}}$.

Theorem 2. *For any $\epsilon > 0$, an algorithm for computing forecasts $p_1, p_2, \ldots$ can be constructed such that the following three items hold:*

- *For any n, $R \subseteq [0,1]^2$, and $\delta > 0$, with probability at least $1-\delta$,*

$$\left| \sum_{i=1}^{n} I_R(\tilde{p}_i, \tilde{x}_i)(S_i - \tilde{p}_i) \right| \le 18(c_{\mathcal{F}}^2 + 1)^{\frac{1}{4}} n^{3/4+\epsilon} + \sqrt{\frac{n}{2} \ln \frac{2}{\delta}}, \tag{2}$$

  where $\tilde{p}_i$ is the randomization of p_i, $x_i = S_{i-1}$ and $\tilde{x}_i$ is its randomization.[2]
- *For any $D \in \mathcal{F}$ and n,*

$$\left| \sum_{i=1}^{n} D(z_i)(S_i - p_i) \right| \le \|D\|_{\mathcal{F}} \sqrt{(c_{\mathcal{F}}^2 + 1)n}, \tag{3}$$

  where $z_1, z_2, \ldots$ are signals.
- *For any $R \subseteq [0,1]^2$, with probability 1,*

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} I_R(\tilde{p}_i, \tilde{x}_i)(S_i - \tilde{p}_i) = 0. \tag{4}$$

Proof. First, in Proposition 1, given $\Delta > 0$, we modify a randomized rounding algorithm of Kakade and Foster [8] to construct some Δ-calibrated forecasting algorithm, and combine it with Vovk's [14] defensive forecasting algorithm. After that, we revise it tending $\Delta \to 0$ such that (4) will hold.

Proposition 1. *Under the assumptions of Theorem 2, an algorithm for computing forecasts can be constructed such that the inequality (3) holds for all D from RKHS $\mathcal{F}$ and for all n. Also, for any n, R, and $\delta > 0$, with probability at least $1-\delta$,*

$$\left| \sum_{i=1}^{n} I_R(\tilde{p}_i, \tilde{x}_i)(S_i - \tilde{p}_i) \right| \le \Delta n + \sqrt{\frac{n(c_{\mathcal{F}}^2 + 1)}{\Delta}} + \sqrt{\frac{n}{2} \ln \frac{2}{\delta}}.$$

[2] This theorem can be generalized for the case where $x_i = f(c_i, z_i)$ is k-dimensional vector which is a function of the history $c_i = (\tilde{p}_1, S_1, \ldots, \tilde{p}_{i-1}, S_{i-1})$ and a signal z_i.

Proof. We define a deterministic forecast and after that we randomize it.

The partition $V=\{v_0,\dots,v_K\}$ was defined above. Recall that we round the deterministic forecast p_n to v_{i-1} with probability $w_{v_{i-1}}(p_n)$ and to v_i with probability $w_{v_i}(p_n)$. We also round the number $x_n=S_{n-1}$ to v_{j-1} with probability $w_{v_{j-1}}(x_n)$ and to v_j with probability $w_{v_j}(x_n)$, where $x_n\in[v_{j-1},v_j]$.

Let $W_v(p,x)=w_{v^1}(p)w_{v^2}(x)$, where $v=(v^1,v^2)\in V^2$. Let $\overline{W}p,x)=(W_v(p,x):v\in V^2)$ be the vector of probabilities of rounding. Define the corresponding kernel $K_1(p,x,p',x')=(\overline{W}(p,x)\cdot\overline{W}(p',x'))$.

The kernel $K_2(z,z')$ can be represented as a dot product in a feature space $K_2(z,z')=(\Phi(z)\cdot\Phi(z'))$. Define

An algorithm for computing the deterministic forecasts $p_1,p_2,\dots$ is presented on Fig. 2 (see also Vovk et al. [13]).

Input parameter: Δ. Define $p_1=1/2$.
FOR $n=1,2\dots$
Define $U_n(p)=\sum\limits_{i=1}^{n-1}(K_1(p,x_n,p_i,x_i)+K_2(z_n,z_i))(S_i-p_i)$.
If $U_n(p)>0$ for all $p\in[0,1]$ then define $p_n=1$;
If $U_n(p)<0$ for all $p\in[0,1]$ then $p_n=0$.
Otherwise, define p_n to be a root of the equation $U_n(p)=0$ (some root exists by the intermediate value theorem).
ENDFOR

Fig. 2. Algorithm for computing deterministic forecasts

Now we continue the proof of the proposition.

Let forecasts $p_1,p_2,\dots$ be computed by the algorithm presented on Fig 2. Since $U(p_n)(S_n-p_n)\le 0$ for all n, we have for any N,

$$0\ge\sum_{n=1}^{N}U_n(p_n)(S_n-p_n)=$$

$$=\sum_{n=1}^{N}\sum_{i=1}^{n-1}(K_1(p_n,x_n,p_i,x_i)+K_2(z_n,z_i))(S_i-p_i)(S_n-p_n)=$$

$$=\frac{1}{2}\sum_{n=1}^{N}\sum_{i=1}^{N}K_1(p_n,x_n,p_i,x_i)(S_i-p_i)(S_n-p_n)-$$

$$-\frac{1}{2}\sum_{n=1}^{N}K_1(p_n,x_n,p_n,x_n)(S_n-p_n)^2+$$

$$+\frac{1}{2}\sum_{n=1}^{N}\sum_{i=1}^{N}K_2(z_n,z_i)(S_i-p_i)(S_n-p_n)-$$

$$-\frac{1}{2}\sum_{n=1}^{N} K_2(z_n,z_n)(S_n-p_n)^2 = \tag{5}$$

$$= \frac{1}{2}\left\|\sum_{n=1}^{N}\overline{W}(p_n,x_n)(S_n-p_n)\right\|^2 - \frac{1}{2}\sum_{n=1}^{N}\|\overline{W}(p_n,x_n)\|^2(S_n-p_n)^2 + \tag{6}$$

$$+\frac{1}{2}\left\|\sum_{n=1}^{N}\Phi(z_n)(S_n-p_n)\right\|_{\mathcal{F}}^2 - \frac{1}{2}\sum_{n=1}^{N}\|\Phi(z_n)\|_{\mathcal{F}}^2(S_n-p_n)^2. \tag{7}$$

In (6), $\|\cdot\|$ is Euclidian norm, and in (7), $\|\cdot\|_{\mathcal{F}}$ is the norm in RKHS $\mathcal{F}$.

Since $(S_n-p_n)^2 \le 1$ for all n and

$$\|(\overline{W}(p_n,x_n)\|^2 = \sum_{v\in V^2}(W_v(p_n,x_n))^2 \le \sum_{v\in V^2} W_v(p_n,x_n) = 1,$$

the subtracted sum of (6) is upper bounded by N.

Since $\|\Phi(z_n)\|_{\mathcal{F}} = c_{\mathcal{F}}(z_n)$ and $c_{\mathcal{F}}(z) \le c_{\mathcal{F}}$ for all z, the subtracted sum of (7) is upper bounded by $c_{\mathcal{F}}^2 N$. As a result we obtain

$$\left\|\sum_{n=1}^{N}\overline{W}(p_n,x_n)(S_n-p_n)\right\| \le \sqrt{(c_{\mathcal{F}}^2+1)N} \tag{8}$$

$$\left\|\sum_{n=1}^{N}\Phi(z_n)(S_n-p_n)\right\|_{\mathcal{F}} \le \sqrt{(c_{\mathcal{F}}^2+1)N} \tag{9}$$

for all N. Let us define $\bar{\mu}_n = \sum_{i=1}^{n}\overline{W}(p_i,x_i)(S_i-p_i)$. By (8), $\|\bar{\mu}_n\| \le \sqrt{(c_{\mathcal{F}}^2+1)n}$ for all n.

Let $\bar{\mu}_n = (\mu_n(v) : v \in V^2)$. By definition of $\bar{\mu}_n$,

$$\mu_n(v) = \sum_{i=1}^{n} W_v(p_i,x_i)(S_i-p_i) \tag{10}$$

for any $v \in V$.

Let $I(p,x)$ be an indicator function. Insert the term $I(v)$ in the sum (10), sum by $v \in V^2$, and exchange the order of summation. Using Cauchy–Schwarz inequality for vectors $\bar{I} = (I(v) : v \in V^2)$, $\bar{\mu}_n = (\mu_n(v) : v \in V^2)$ and Euclidian norm, we obtain

$$\left|\sum_{i=1}^{n}\sum_{v\in V^2} W_v(p_i,x_i)I(v)(S_i-p_i)\right| =$$

$$= \left|\sum_{v\in V^2} I(v)\sum_{i=1}^{n} W_v(p_i,x_i)(S_i-p_i)\right| =$$

$$= |(\bar{I}\cdot\bar{\mu}_n)| \le \|\bar{I}\|\cdot\|\bar{\mu}_n\| \le \sqrt{|V^2|(c_{\mathcal{F}}^2+1)n} \tag{11}$$

for all n, where $|V^2| = (1+\frac{1}{\Delta})^2 \le \left(\frac{2}{\Delta}\right)^2$ is the cardinality of the partition.

Let $\tilde{p}_i$ be a random variable taking values $v \in V$ with probabilities $w_v(p_i)$ (only two of them are nonzero). Recall that $\tilde{x}_i$ is the random variable taking values $v \in V$ with probabilities $w_v(x_i)$. Also, for $v = (v^1, v^2)$, $W_v(p_i, x_i) = w_{v^1}(p_i) w_{v^2}(x_i)$.

For any i, the mathematical expectation of the random variable $I(\tilde{p}_i, \tilde{x}_i)(S_i - \tilde{p}_i)$ is equal to

$$E(I(\tilde{p}_i, \tilde{x}_i)(S_i - \tilde{p}_i)) = \sum_{v \in V^2} W_v(p_i, x_i) I(v)(S_i - v^1), \tag{12}$$

where $v = (v^1, v^2)$. By Azuma–Hoeffding inequality (see (24) below), for any n and $\delta > 0$, with probability $1 - \delta$,

$$\left| \sum_{i=1}^{n} I(\tilde{p}_i, \tilde{x}_i)(S_i - \tilde{p}_i) - \sum_{i=1}^{n} E(I(\tilde{p}_i, \tilde{x}_i)(S_i - \tilde{p}_i)) \right| \le \sqrt{\frac{n}{2} \ln \frac{2}{\delta}}. \tag{13}$$

By method of rounding

$$\left| \sum_{v \in V^2} W_v(p_i, x_i) I(v)(S_i - p_i) - \sum_{v \in V^2} W_v(p_i, x_i) I(v)(S_i - v^1) \right| \le \Delta$$

for all i, where $v = (v^1, v^2)$. Summing (12) over $i = 1, \ldots, n$ and using the inequality (11), we obtain

$$\begin{aligned} \left| \sum_{i=1}^{n} E(I(\tilde{p}_i, \tilde{x}_i)(S_i - \tilde{p}_i)) \right| &= \\ = \left| \sum_{i=1}^{n} \sum_{v \in V^2} W_v(p_i, x_i) I(v)(S_i - v^1) \right| &\le \\ \le \Delta n + \sqrt{(c_{\mathcal{F}}^2 + 1) n / \Delta^2} & \end{aligned} \tag{14}$$

for all n. By (13) and (14), with probability $1 - \delta$,

$$\left| \sum_{i=1}^{n} I(\tilde{p}_i, \tilde{x}_i)(S_i - \tilde{p}_i) \right| \le \Delta n + 2\sqrt{(c_{\mathcal{F}}^2 + 1) n / \Delta^2} + \sqrt{\frac{n}{2} \ln \frac{2}{\delta}}. \tag{15}$$

By Cauchy–Schwarz inequality

$$\begin{aligned} \left| \sum_{n=1}^{N} D(z_n)(S_n - p_n) \right| = \left| \sum_{n=1}^{N} (S_n - p_n)(D \cdot \Phi(z_n)) \right| = \\ \left| \left(\sum_{n=1}^{N} (S_n - p_n) \Phi(z_n) \cdot D \right) \right| \le \left\| \sum_{n=1}^{N} (S_n - p_n) \Phi(z_n) \right\|_{\mathcal{F}} \cdot \|D\|_{\mathcal{F}} \le \\ \le \|D\|_{\mathcal{F}} \sqrt{(c_{\mathcal{F}}^2 + 1) N}. \end{aligned}$$

Proposition is proved. $\triangle$

Now we turn to the proof of Theorem 2.

The expression $\Delta n + 2\sqrt{(c_{\mathcal{F}}^2+1)n/\Delta^2}$ from (14) and (15) takes its minimal value when $\Delta = \sqrt{2}(c_{\mathcal{F}}^2+1)^{\frac{1}{4}}n^{-\frac{1}{4}}$. In this case, the right-hand side of the inequality (14) is equal to

$$\Delta n + 2\sqrt{n(c_{\mathcal{F}}^2+1)/\Delta^2} = 2\Delta n = 2\sqrt{2}(c_{\mathcal{F}}^2+1)^{\frac{1}{4}}n^{\frac{3}{4}}. \tag{16}$$

In what follows we use the upper bound $2\Delta n$ in (14).

To prove the bound (2) choose a monotonic sequence of real numbers $\Delta_1 > \Delta_2 > \ldots$ such that $\Delta_s \to 0$ as $s \to \infty$. We also define an increasing sequence of positive integer numbers $n_1 < n_2 < \ldots$ For any s, we use on steps $n_s \le n < n_{s+1}$ the randomization grid of $[0,1]$ defined by subintervals of length Δ_s.

We start our sequences from $n_1 = 1$ and $\Delta_1 = 1$. Also, define the numbers $n_2, n_3, \ldots$ such that the inequality

$$\left|\sum_{i=1}^{n} E(I(\tilde{p}_i,\tilde{x}_i)(S_i-\tilde{p}_i))\right| \le 4(s+1)\Delta_s n \tag{17}$$

holds for all $n_s \le n \le n_{s+1}$ and for all $s \ge 1$.

We define this sequence by mathematical induction on s. Suppose that n_s $(s \ge 1)$ is defined such that the inequality

$$\left|\sum_{i=1}^{n} E(I(\tilde{p}_i,\tilde{x}_i)(S_i-\tilde{p}_i))\right| \le 4s\Delta_{s-1} n \tag{18}$$

holds for all $n_{s-1} \le n \le n_s$, and the inequality

$$\left|\sum_{i=1}^{n_s} E(I(\tilde{p}_i,\tilde{x}_i)(S_i-\tilde{p}_i))\right| \le 4s\Delta_s n_s \tag{19}$$

also holds.

Let us define n_{s+1}. Consider all forecasts $\tilde{p}_i$ defined by the algorithm given above for the discretization $\Delta = \Delta_{s+1}$. We do not use first n_s of these forecasts (more correctly we will use them only in bounds (20) and (21); denote these forecasts $\mathbf{\hat{p}_1}, \ldots, \mathbf{\hat{p}_{n_s}}$. We add the forecasts $\tilde{p}_i$ for $i > n_s$ to the forecasts defined before this step of induction (for n_s). Let n_{s+1} be such that the inequality

$$\begin{aligned}\left|\sum_{i=1}^{n_{s+1}} E(I(\tilde{p}_i,\tilde{x}_i)(S_i-\tilde{p}_i))\right| \le \left|\sum_{i=1}^{n_s} E(I(\tilde{p}_i,\tilde{x}_i)(S_i-\tilde{p}_i))\right| + \\ + \left|\sum_{i=n_s+1}^{n_{s+1}} E(I(\tilde{p}_i,\tilde{x}_i)(S_i-\tilde{p}_i)) + \sum_{i=1}^{n_s} E(I(\mathbf{\hat{p}_i},\tilde{x}_i)(S_i-\mathbf{\hat{p}_i}))\right| + \\ + \left|\sum_{i=1}^{n_s} E(I(\mathbf{\hat{p}_i},\tilde{x}_i)(S_i-\mathbf{\hat{p}_i}))\right| \le 4(s+1)\Delta_{s+1} n_{s+1}\end{aligned} \tag{20}$$

holds. Here the first sum of the right-hand side of the inequality (20) is bounded by $4s\Delta_s n_s$ – by the induction hypothesis (19). The second and third sums are bounded by $2\Delta_{s+1}n_{s+1}$ and by $2\Delta_{s+1}n_s$, respectively, where $\Delta = \Delta_{s+1}$ is defined such that (16) holds. This follows from (14) and by choice of n_s.

The induction hypothesis (19) is valid for

$$n_{s+1} \geq \frac{2s\Delta_s + \Delta_{s+1}}{\Delta_{s+1}(2s+1)} n_s.$$

Similarly,

$$\left|\sum_{i=1}^{n} E(I(\tilde{p}_i, \tilde{x}_i)(S_i - \tilde{p}_i))\right| \leq \left|\sum_{i=1}^{n_s} E(I(\tilde{p}_i, \tilde{x}_i)(S_i - \tilde{p}_i))\right| +$$
$$+\left|\sum_{i=n_s+1}^{n} E(I(\tilde{p}_i, \tilde{x}_i)(S_i - \tilde{p}_i)) + \sum_{i=1}^{n_s} E(I(\hat{\mathbf{p}}_\mathbf{i}, \tilde{x}_i)(S_i - \hat{\mathbf{p}}_\mathbf{i}))\right| +$$
$$+\left|\sum_{i=1}^{n_s} E(I(\hat{p}_i, \tilde{x}_i)(S_i - \hat{\mathbf{p}}_\mathbf{i}))\right| \leq 4(s+1)\Delta_s n \tag{21}$$

for $n_s < n \leq n_{s+1}$. Here the first sum of the right-hand inequality (20) is also bounded by $4s\Delta_s n_s \leq 4s\Delta_s n$ – by the induction hypothesis (19). The second and the third sums are bounded by $2\Delta_{s+1}n \leq 2\Delta_s n$ and by $2\Delta_{s+1}n_s \leq 2\Delta_s n$, respectively. This follows from (14) and from choice of Δ_s. The induction hypothesis (18) is valid.

By (17) for any s

$$\left|\sum_{i=1}^{n} E(I(\tilde{p}_i, \tilde{x}_i)(S_i - \tilde{p}_i))\right| \leq 4(s+1)\Delta_s n \tag{22}$$

for all $n \geq n_s$ if Δ_s satisfies the condition $\Delta_{s+1} \leq \Delta_s(1 - \frac{1}{s+2})$ for all s.

We show now that sequences n_s and Δ_s satisfying all the conditions above exist. Let $\epsilon > 0$ and $M = \lceil 2/\epsilon \rceil$, where $\lceil r \rceil$ is the least integer number $\geq r$. Define $n_s = (s+M)^M$ and $\Delta_s = (c_{\mathcal{F}}^2 + 1)^{\frac{1}{4}} n_s^{-\frac{1}{4}}$. Easy to verify that all requirements for n_s and Δ_s given above are satisfied for all $s \geq s_0$, where s_0 is sufficiently large. We redefine $n_i = n_{s_0}$ for all $1 \leq i \leq s_0$. Then all these requirements hold for these i trivially.

We have in (22) for all $n_s \leq n < n_{s+1}$

$$4(s+1)\Delta_s n \leq 4(s+M)\Delta_s n_{s+1} =$$
$$= 4\sqrt{2}(c_{\mathcal{F}}^2 + 1)^{\frac{1}{4}}(s+M)(s+M+1)^M(s+M)^{-\frac{M}{4}} \leq$$
$$\leq 18(c_{\mathcal{F}}^2 + 1)^{\frac{1}{4}} n_s^{\frac{3}{4}+2/M} \leq$$
$$\leq 18(c_{\mathcal{F}}^2 + 1)^{\frac{1}{4}} n^{\frac{3}{4}+\epsilon}.$$

Therefore, we obtain

$$\left|\sum_{i=1}^{n} E(I(\tilde{p}_i, \tilde{x}_i)(S_i - \tilde{p}_i))\right| \leq 18(c_{\mathcal{F}}^2 + 1)^{\frac{1}{4}} n^{\frac{3}{4}+\epsilon} \tag{23}$$

for all n. Azuma–Hoeffding inequality says that for any $\gamma > 0$

$$Pr\left\{\left|\frac{1}{n}\sum_{i=1}^{n} V_i\right| > \gamma\right\} \le 2e^{-2n\gamma^2} \tag{24}$$

for all n, where V_i are martingale–differences. We define $V_i = I(\tilde{p}_i, \tilde{x}_i)(S_i - \tilde{p}_i) - E(I(\tilde{p}_i, \tilde{x}_i)(S_i - \tilde{p}_i))$ and $\gamma = \sqrt{\frac{1}{2n}\ln\frac{2}{\delta}}$, where $\delta > 0$.

Combining (23) with (24), we obtain that for any n and $\delta > 0$, with probability $1 - \delta$,

$$\left|\sum_{i=1}^{n} I(\tilde{p}_i, \tilde{x}_i)(S_i - \tilde{p}_i)\right| \le 18(c_{\mathcal{F}}^2 + 1)^{\frac{1}{4}} n^{\frac{3}{4}+\epsilon} + \sqrt{\frac{n}{2}\ln\frac{2}{\delta}}.$$

The asymptotic relation (4) can be proved using the Borel–Cantelli lemma. This proof is similar to the final part of the proof of Theorem 1 below. Theorem 2 is proved. $\triangle$

5 Proof of Theorem 1

At any step i we compute the deterministic forecast p_i defined in Section 4 and its randomization to $\tilde{p}_i$ using the parameters $M = \lceil 2/\epsilon \rceil$, $\Delta = \Delta_s = \sqrt{2}(c_{\mathcal{F}} + 1)^{\frac{1}{4}}(s + M)^{-\frac{M}{4}}$ and $n_s = (s + M)^M$, where $n_s \le i < n_{s+1}$. Let also $\tilde{S}_{i-1}$ be the randomized past price S_{i-1}. In Theorem 2, we have $z_i = S_{i-1}$ and $\tilde{z}_i = \tilde{S}_{i-1}$.

The following upper bound directly follows from the method of discretization:

$$\left|\sum_{i=1}^{n} I(\tilde{p}_i > \tilde{S}_{i-1})(\tilde{S}_{i-1} - S_{i-1})\right| \le \sum_{t=0}^{s}(n_{t+1} - n_t)\Delta_t \le$$
$$\le 4(c_{\mathcal{F}}^2 + 1)^{\frac{1}{4}} n_s^{\frac{3}{4}+\epsilon} \le 4(c_{\mathcal{F}}^2 + 1)^{\frac{1}{4}} n^{\frac{3}{4}+\epsilon}, \tag{25}$$

where $n_s \le n < n_{s+1}$. Here $I(p > S) = 1$ if $p > S$ and $I(p > S) = 0$ otherwise.

Let $D(z)$ be an arbitrary function from the RKHS $\mathcal{F}$. Clearly, the bound (25) holds if we replace $I(\tilde{p}_i > \tilde{S}_{i-1})$ with $\|D\|_+^{-1} D(z_i)$.

First, we give the proof for the case, where $D(x) \ge 0$ for all x and $\tilde{M}_i^+ = \max\{\tilde{M}_i, 0\}$, where $\tilde{M}_i$ is defined on Fig 1. We use the notation

$$\nu_1(n) = 4(c_{\mathcal{F}}^2 + 1)^{\frac{1}{4}} n^{\frac{3}{4}+\epsilon}, \tag{26}$$

$$\nu_2(n) = 18n^{\frac{3}{4}+\epsilon}(c_{\mathcal{F}}^2 + 1)^{\frac{1}{4}} + \sqrt{\frac{n}{2}\ln\frac{2}{\delta}}. \tag{27}$$

$$\nu_3(n) = \sqrt{(c_{\mathcal{F}}^2 + 1)n} \tag{28}$$

All sums below are considered for $i = 1, \ldots n$. Also, we use the Azuma–Hoeffding inequality (24).

For any $\delta > 0$, the following chain of equalities and inequalities is valid with probability $1-\delta$:

$$\mathcal{K}_n^M = \sum_{i=1}^n \tilde{M}_i^+(S_i - S_{i-1}) = \sum_{\tilde{p}_i > \tilde{S}_{i-1}} (S_i - S_{i-1}) =$$

$$= \sum_{\tilde{p}_i > \tilde{S}_{i-1}} (S_i - \tilde{p}_i) + \sum_{\tilde{p}_i > \tilde{S}_{i-1}} (\tilde{p}_i - \tilde{S}_{i-1}) + \sum_{\tilde{p}_i > \tilde{S}_{i-1}} (\tilde{S}_{i-1} - S_{i-1}) \geq \tag{29}$$

$$\geq \sum_{\tilde{p}_i > \tilde{S}_{i-1}} (\tilde{p}_i - \tilde{S}_{i-1}) - \nu_1(n) - \nu_2(n) \geq \tag{30}$$

$$\geq \|D\|_+^{-1} \sum_{i=1}^n D(z_i)(\tilde{p}_i - \tilde{S}_{i-1}) - \nu_1(n) - \nu_2(n) =$$

$$= \|D\|_+^{-1} \sum_{i=1}^n D(z_i)(p_i - S_{i-1}) + \|D\|_+^{-1} \sum_{i=1}^n D(z_i)(\tilde{p}_i - p_i) -$$

$$-\|D\|_+^{-1} \sum_{i=1}^n D(z_i)(\tilde{S}_{i-1} - S_{i-1}) - \nu_1(n) - \nu_2(n) \geq \tag{31}$$

$$\geq \|D\|_+^{-1} \sum_{i=1}^n D(z_i)(p_i - S_{i-1}) - 3\nu_1(n) - \nu_2(n) = \tag{32}$$

$$= \|D\|_+^{-1} \sum_{i=1}^n D(z_i)(S_i - S_{i-1}) - \|D\|_+^{-1} \sum_{i=1}^n D(z_i)(S_i - p_i) -$$

$$-3\nu_1(n) - \nu_2(n) \geq \tag{33}$$

$$\geq \|D\|_+^{-1} \sum_{i=1}^n D(z_i)(S_i - S_{i-1}) - 3\nu_1(n) - \nu_2(n) - \|D\|_+^{-1}\|D\|_{\mathcal{F}}\nu_3(n) =$$

$$= \|D\|_+^{-1}\mathcal{K}_n^D - 3\nu_1(n) - \nu_2(n) - \|D\|_+^{-1}\|D\|_{\mathcal{F}}\nu_3(n). \tag{34}$$

To pass from (29) to (30), inequality (2) of Theorem 2 and the bound (25) were used, and so the terms (26) and (27) were subtracted. To pass from (31) to (32), the bound (25) was twice applied to intermediate terms, and so the term (25) was subtracted twice. To pass from (32) to (33), inequality (3) of Theorem 2 was used, and so the term (28) was subtracted.

In general case, we represent $\tilde{M}_i = \tilde{M}_i^+ + \tilde{M}_i^-$, where $\tilde{M}_i^+ = \max\{\tilde{M}_i, 0\}$ and $\tilde{M}_i^- = \min\{\tilde{M}_i, 0\}$, and $\tilde{M}_i$ is defined on Fig 1. Also, define $D = D^+ + D^-$, where $D^+ = \max\{D, 0\}$ and $D^- = \min\{D, 0\}$. After that, we obtain (34) for any pair M_i^+, D^+ and M_i^-, D^- separately and add the results.

Therefore, for any $D \in \mathcal{F}$ a constant $c > 0$ exists such that for any n, with probability $1-\delta$,

$$\mathcal{K}_n^M \geq \|D\|_+^{-1}\mathcal{K}_n^D - cn^{\frac{3}{4}+\epsilon} - \sqrt{\frac{n}{2}\ln\frac{2}{\delta}}. \tag{35}$$

Inequality (1) will follow from (35). We apply Borel–Cantelli lemma and the Hoeffding inequality. Denote $\gamma = \sqrt{\frac{1}{2n}\ln\frac{2}{\delta}}$. Then $\delta = 2e^{-2n\gamma^2}$. Rewrite (35) in the form

$$\frac{1}{n}\mathcal{K}_n^M - \|D\|_+^{-1}\frac{1}{n}\mathcal{K}_n^D \geq -cn^{-\frac{1}{4}+\epsilon} - \gamma. \tag{36}$$

According to (35), for any n and $\gamma > 0$, inequality (36) is violated with probability $2e^{-2n\gamma^2}$. Since the series $\sum\limits_{n=1}^{\infty} e^{-2n\gamma^2}$ converges, inequality (36) for a fixed γ can be violated no more than for finitely many different n. By the Borel–Cantelli lemma the event

$$\liminf_{n\to\infty}\frac{1}{n}\left(\mathcal{K}_n^M - \|D\|_+^{-1}\mathcal{K}_n^D\right) \geq 0$$

holds almost surely. Theorem 1 is proved for any $D \in \mathcal{F}$.

Using a universal kernel and the corresponding canonical universal RKHS, we can extend our asymptotic results for all continuous stationary trading strategies D. An RKHS $\mathcal{F}$ on X is universal if X is a compact metric space and every continuous function f on X can be approximated arbitrarily well in the metric $\|\cdot\|_\infty$ by a function from $\mathcal{F}$: for any $\epsilon > 0$ there exists $D \in \mathcal{F}$ such that $\sup\limits_{x\in X}|f(x) - D(x)| \leq \epsilon$ (see Steinwart [12], Definition 4).

We use $X = [0,1]$. The Sobolev space $\mathcal{F} = H^1([0,1])$ is the universal RKHS (see Steinwart [12], Vovk [14]).

Existence of a universal RKHS on $[0,1]$ implies a full version of Theorem 1. This result directly follows from inequality (36) and from the possibility to approximate any continuous function f on $[0,1]$ arbitrarily closely by a function D from the universal RKHS $\mathcal{F}$.

The universal consistency property (1) is strictly asymptotic and says nothing about finite data sequences. The convergence bound (35) was obtained for functions from narrower RKHS classes.

6 Conclusion

This impressive efficiency of the trading strategy $\tilde{M}_i$ can be explained by the restrictive power of continuous functions. A continuous stationary trading strategy D cannot respond sufficiently quickly to information about changes of the value of a future price S_i. The optimal trading strategy $\tilde{M}_i$ is a discontinuous function, though it is applied to the random variables.

A positive argument in favor of the requirement of continuity of D is that it is natural to compete only with computable trading strategies, and continuity is often regarded as a necessary condition for computability (Brouwer's "continuity principle").

If D is allowed to be discontinuous, we cannot prove (1) in general case. Moreover, we can prove that for any randomizing trading strategy $\tilde{M}_i$, which randomly takes values 1 and -1 with probabilities varying with time, a (discontinuous) function $D(z)$ also taking values 1 and -1 and the sequences of signals

z_i and outcomes S_i, $i = 1, 2, \ldots$, exist in the protocol presented on Fig. 1 such that with probability one:

$$\limsup_{n\to\infty} \frac{1}{n}\left(\mathcal{K}_n^M - \frac{1}{2}\mathcal{K}_n^D\right) \le 0.$$

We can also prove that (1) remain valid in case of discontinuous D if we use randomized signal $\tilde{z}_i$, i.e., if we use the value $D(\tilde{z}_i)$ in the protocol presented on Fig. 1.

This research was partially supported by Russian foundation for fundamental research: 13-01-00521.

References

1. Cover, T.: Universal portfolios. Mathematical Finance. **1** (1991) 1-29
2. Cover, T., Ordentlich, E.: Universal portfolio with side information. IEEE Transaction on Information Theory. **42** (1996) 348–363
3. Cristianini, N,, Shawe-Taylor, J.: An Introduction to Support Vector Machines and other kernel-based learning methods. Cambridge University Press. Cambridge (2000)
4. Dawid, A.P.: The well-calibrated Bayesian [with discussion]. J. Am. Statist. Assoc. **77** (1982) 605–613
5. Foster, D.P., Vohra, R.: Asymptotic calibration. Biometrika. **85** (1998) 379–390
6. Foster, D.P., Vohra, R.: Calibrated learning and correlated equilibrium. Games and Economic Behavior. **21(1-2)** (1997) 40–55
7. Foster, D.P., Rakhlin, A., Sridharan, K., Tewari, A.: Complexity-based approach to calibration with checking rules. Journal of Machine Learning Research - Proceedings Track. **19** (2011) 293–314
8. Kakade, S.M., Foster, D.P.: Deterministic calibration and Nash equilibrium. LNCS **3120** (John Shawe Taylor and Yoram Singer. ed) (2004) 33–48
9. Mannor, S., Stoltz, G.: A geometric proof of calibration. Mathematics of Operations Research. **35(4)** (2010) 721–727
10. Oakes, D.: Self-Calibrating Priors Do not Exist [with discussion]. J. Am. Statist. Assoc. **80** (1985) 339–342
11. Scholkopf, B., Smola, A.: Learning with Kernels. MIT Press. Cambridge MA (2002)
12. Steinwart, I.: On the influence of the kernel on the consistency of support vector machines. Journal of Machine Learning Research **2** (2001) 67-93
13. Vovk, V., Takemura, A., Shafer, G.: Defensive forecasting. Proceedings of the 10th International Workshop on Artificial Intelligence and Statistics (ed. by R. G. Cowell and Z. Ghahramani) – Cambridge UK: Society for Artificial Intelligence and Statistics (2005) 365–372
14. Vovk, V.: On-line regression competitive with reproducing kernel Hilbert spaces (extended abstract). TAMS Lecture Notes in Computer Science – Berlin: Springer. **3959** (2006) 452–463
15. Vovk, V.: Predictions as statements and decisions. Theoretical Computer Science. **405(3)** (2008) 285–296
16. Vovk, V.: Defensive Forecasting for Optimal Prediction with Expert Advice. arXiv:0708.1503v1 (2007)
17. V'yugin, V., Trunov, V.: Universal algorithmic trading. Journal of Investment Strategies. **2 (1)** (2012/13) 63–88