Journal of Machine Learning Research 12 (2011) 241-266 Submitted 8/10; Revised 12/10; Published 1/11

Online Learning in Case of Unbounded Losses Using Follow the Perturbed Leader Algorithm

Vladimir V. V'yugin VYUGIN@IITP.RU
Institute for Information Transmission Problems
Russian Academy of Sciences
Bol'shoi Karetnyi per. 19
Moscow GSP-4, 127994, Russia

Editor: Manfred Warmuth

Abstract

In this paper the sequential prediction problem with expert advice is considered for the case where losses of experts suffered at each step cannot be bounded in advance. We present some modification of Kalai and Vempala algorithm of following the perturbed leader where weights depend on past losses of the experts. New notions of a volume and a scaled fluctuation of a game are introduced. We present a probabilistic algorithm protected from unrestrictedly large one-step losses. This algorithm has the optimal performance in the case when the scaled fluctuations of one-step losses of experts of the pool tend to zero.

Keywords: prediction with expert advice, follow the perturbed leader, unbounded losses, adaptive learning rate, expected bounds, Hannan consistency, online sequential prediction

1. Introduction

Experts algorithms are used for online prediction or repeated decision making or repeated game playing. Starting with the Weighted Majority Algorithm (WM) of Littlestone and Warmuth (1994) and Aggregating Algorithm of Vovk (1990), the theory of Prediction with Expert Advice has rapidly developed in the recent times. Also, most authors have concentrated on predicting binary sequences and have used specific (usually convex) loss functions, like absolute loss, square and logarithmic loss. A survey can be found in the book of Lugosi and Cesa-Bianchi (2006). Arbitrary losses are less common, and, as a rule, they are supposed to be bounded in advance (see well known Hedge Algorithm of Freund and Schapire 1997, Normal Hedge of Chaudhuri et al. 2009 and other algorithms).

In this paper we consider a different general approach—"Follow the Perturbed Leader – FPL" algorithm, now called Hannan's algorithm, see Hannan (1957), Kalai and Vempala (2003) and Lugosi and Cesa-Bianchi (2006). Under this approach we only choose the decision that has fared the best in the past—the leader. In order to cope with adversary some randomization is implemented by adding a perturbation to the total loss prior to selecting the leader. The goal of the learner's algorithm is to perform almost as well as the best expert in hindsight in the long run. The resulting FPL algorithm has the same performance guarantees as WM-type algorithms for fixed learning rate and bounded one-step losses, save for a factor $\sqrt{2}$. A major advantage of the FPL algorithm is that its analysis remains easy for an adaptive learning rate, in contrast to WM-derivatives (see remark in Hutter and Poland 2004).

Prediction with Expert Advice considered in this paper proceeds as follows. We are asked to perform sequential actions at times $t = 1,2,\dots,T$. At each time step t, experts $i = 1,\dots N$ receive results of their actions in form of their losses s^i_t—arbitrary real numbers.

At the beginning of the step t *Learner*, observing cumulating losses $s^i_{1:t-1} = s^i_1 + \dots + s^i_{t-1}$ of all experts $i = 1,\dots N$, makes a decision to follow one of these experts, say Expert i. At the end of step t *Learner* receives the same loss s^i_t as Expert i at step t and suffers *Learner's* cumulative loss $s_{1:t} = s_{1:t-1} + s^i_t$.

In the traditional framework, we suppose that one-step losses of all experts are bounded, for example, $0 \le s^i_t \le 1$ for all i and t.

Well known simple example of a game with two experts shows that Learner can perform much worse than each expert: let the current losses of two experts on steps $t = 0,1,\dots,6$ be $s^1_{0,1,2,3,4,5,6} = (\frac{1}{2},0,1,0,1,0,1)$ and $s^2_{0.1,2,3,4,5,6} = (0,1,0,1,0,1,0)$. Evidently, "Follow the Leader" algorithm always chooses the wrong prediction.

When the experts one-step losses are bounded, this problem has been solved using randomization of the experts cumulative losses. The method of following the perturbed leader was discovered by Hannan (1957). Kalai and Vempala (2003) rediscovered this method and published a simple proof of the main result of Hannan. They called an algorithm of this type FPL (Following the Perturbed Leader).

The FPL algorithm outputs prediction of an expert i which minimizes

$$s^i_{1:t-1} - \frac{1}{\varepsilon}\xi^i,$$

where ξ^i, $i = 1,\dots N$, $t = 1,2,\dots$, is a sequence of i.i.d random variables distributed according to the exponential distribution with the density $p(x) = \exp\{-x\}$, and ε is *a learning rate*.

Kalai and Vempala (2003) show that the expected cumulative loss of the FPL algorithm has the upper bound

$$E(s_{1:t}) \le (1+\varepsilon)\min_{i=1,\dots,N} s^i_{1:t} + \frac{\log N}{\varepsilon},$$

where ε is a positive real number such that $0 < \varepsilon < 1$ is a learning rate, N is the number of experts.

Hutter and Poland (2004, 2005) presented a further developments of the FPL algorithm for countable class of experts, arbitrary weights and adaptive learning rate. Also, FPL algorithm is usually considered for bounded one-step losses: $0 \le s^i_t \le 1$ for all i and t. Using a variable learning rate, an optimal upper bound was obtained in Hutter and Poland (2005):

$$E(s_{1:t}) \le \min_{i=1,\dots,N} s^i_{1:t} + 2\sqrt{2T\ln N}.$$

Most papers on prediction with expert advice either consider bounded losses or assume the existence of a specific loss function (see Lugosi and Cesa-Bianchi 2006). We allow losses at any step to be unbounded. The notion of a specific loss function is not used.

The setting allowing unbounded one-step losses do not have wide coverage in literature; we can only refer reader to Allenberg et al. (2006), Cesa-Bianchi et al. (2007) and Poland and Hutter (2005).

Poland and Hutter (2005) have studied the games where one-step losses of all experts at each step t are bounded from above by an increasing sequence B_t given in advance. They presented a learning algorithm which is asymptotically consistent for $B_t = t^{1/16}$.

Allenberg et al. (2006) have considered polynomially bounded one-step losses for a modified version of the Littlestone and Warmuth (1994) algorithm under partial monitoring. In full information case, their algorithm has the expected regret $2\sqrt{N\ln N}(T+1)^{\frac{1}{2}(1+a+\beta)}$ in the case where one-step losses of all experts $i=1,2,\dots N$ at each step t have the bound $(s_t^i)^2 \le t^a$, where $a>0$, and $\beta>0$ is a parameter of the algorithm. They have proved that this algorithm is Hannan consistent if

$$\max_{1\le i\le N}\frac{1}{T}\sum_{t=1}^{T}(s_t^i)^2 < cT^a$$

for all T, where $c>0$ and $0<a<1$.

Cesa-Bianchi et al. (2007) derived a new forecasting strategy for the Weighted Majority algorithm in unbounded setting with regret

$$\sqrt{Q_T\ln N}+M_T\ln N,$$

where $M_T=\max_{1\le i\le N}\max_{1\le t\le T}|s_t^i|$ is the largest absolute value of any loss s_t^i of an expert i at time step T, and $Q_T=\sum\limits_{t=1}^{T}(s_t^{i^{\min}})^2$ is the sum of squared losses of the best at first T steps expert $i^{\min}$. These bounds were improved using cumulative variances of losses (under distributions used in the Weighted Majority algorithm). Cesa-Bianchi et al. (2007) do not study asymptotic consistency of their algorithm.

In this paper we present a sufficient condition for the FPL algorithm to be asymptotically consistent in case where losses are unbounded. In particular, this setting covers a case where loss grows "faster than polynomial, but slower than exponential". We present some modification of Kalai and Vempala (2003) algorithm of following the perturbed leader (FPL) for the case of unrestrictedly large one-step expert losses s_t^i not bounded in advance: $s_t^i\in(-\infty,+\infty)$. This algorithm uses adaptive weights depending on past cumulative losses of the experts.

A motivating example, where losses of the experts cannot be bounded in advance, is given in Section 5.

The full information case is considered in this paper. We analyze the asymptotic consistency of our algorithms using nonstandard scaling. We introduce new notions of *the volume of a game* $v_t=v_0+\sum\limits_{j=1}^{t}\max_i|s_j^i|$ and *the scaled fluctuation* of the game $\mathrm{fluc}(t)=\Delta v_t/v_t$, where $\Delta v_t=v_t-v_{t-1}$ and v_0 is a nonnegative constant.

We show in Theorem 2 that the algorithm of following the perturbed leader with adaptive weights constructed in Section 3 is asymptotically consistent in the mean in the case where $v_t\to\infty$ and $\Delta v_t=o(v_t)$ as $t\to\infty$ with a computable bound. Specifically, if $\mathrm{fluc}(t)\le\gamma(t)$ for all t, where $\gamma(t)$ is a computable function such that $\gamma(t)=o(1)$ as $t\to\infty$, our algorithm has the expected regret

$$2\sqrt{(8+\varepsilon)(1+\ln N)}\sum_{t=1}^{T}(\gamma(t))^{1/2}\Delta v_t,$$

where $\varepsilon>0$ is a parameter of the algorithm.

In case where all losses are nonnegative: $s_t^i\in[0,+\infty)$, we obtain the regret

$$2\sqrt{(2+\varepsilon)(1+\ln N)}\sum_{t=1}^{T}(\gamma(t))^{1/2}\Delta v_t.$$

In particular, this algorithm is asymptotically consistent (in the mean) in a modified sense

$$\limsup_{T\to\infty}\frac{1}{v_T}E(s_{1:T}-\min_{i=1,\ldots N}s^i_{1:T})\le 0, \tag{1}$$

where $s_{1:T}$ is the total loss of our algorithm on steps $1,2,\ldots T$, and $E(s_{1:T})$ is its expectation.

Proposition 1 of Section 2 shows that if the condition $\Delta v_t = o(v_t)$ is violated the cumulative loss of any probabilistic prediction algorithm can be much more than the loss of the best expert of the pool.

In Section 3 we present some sufficient conditions under which our learning algorithm is Hannan consistent.[1]

In Section 4 we consider some special cases of our algorithm and applications for the case of standard time-scaling.

In particular, Corollary 8 of Theorem 2 says that our algorithm is asymptotically consistent (in the modified sense) in the case when one-step losses of all experts at each step t are bounded by t^a, where a is a positive real number. We prove this result under an extra assumption that the volume of the game grows slowly, $\liminf_{t\to\infty} v_t/t^{a+\delta}>0$, where $\delta>0$ is arbitrary. Corollary 8 shows that our algorithm is also Hannan consistent when $\delta>\frac{1}{2}$.

In Section 5 we consider an application of our algorithm for constructing an arbitrage strategy in some game of buying and selling shares of some stock on financial market. We analyze this game in the decision theoretic online learning (DTOL) framework (see Freund and Schapire 1997). We introduce *Learner* that computes weighted average of different strategies with unbounded gains and losses. To change from the follow leader framework to DTOL we derandomize our FPL algorithm.

This paper is an extended version of the ALT 2009 conference paper V'yugin (2009).

2. Games of Prediction with Expert Advice with Unbounded One-step Losses

We consider a game of prediction with expert advice with arbitrary unbounded one-step losses. Following Cesa-Bianchi et al. (2007) we call a game with such losses "signed game" and call these losses "signed losses".

At each step t of the game, all N experts receive one-step losses $s^i_t\in(-\infty,+\infty)$, $i=1,\ldots N$, and the cumulative loss of the ith expert after step t is equal to

$$s^i_{1:t}=s^i_{1:t-1}+s^i_t.$$

A probabilistic learning algorithm of choosing an expert outputs at any step t the probabilities $P\{I_t=i\}$ of following the ith expert given the cumulative losses $s^i_{1:t-1}$ of the experts $i=1,\ldots N$ in hindsight (see Figure 1).

The performance of this probabilistic algorithm is measured in its *expected regret*

$$E(s_{1:T}-\min_{i=1,\ldots N}s^i_{1:T}),$$

where the random variable $s_{1:T}$ is the cumulative loss of the master algorithm, $s^i_{1:T}$, $i=1,\ldots N$, are the cumulative losses of the experts algorithms and E is the mathematical expectation (with respect

1. This means that (1) holds with probability 1, where E is omitted.

Probabilistic algorithm of choosing an expert.
FOR $t = 1, \dots T$
Given past cumulative losses of the experts $s^i_{1:t-1}$, $i = 1, \dots N$, choose an expert i with probability $P\{I_t = i\}$.
Receive the one-step losses at step t of the expert s^i_t and suffer one-step loss $s_t = s^i_t$ of the master algorithm.
ENDFOR

Figure 1: Probabilistic algorithm of choosing an expert

to the probability distribution generated by probabilities $P\{I_t = i\}$, $i = 1, \dots N$, on the first T steps of the game).

In the case of bounded one-step expert losses, $s^i_t \in [0,1]$, and a convex loss function, the well-known learning algorithms have expected regret $O(\sqrt{T \log N})$ (see Lugosi and Cesa-Bianchi 2006).

A probabilistic algorithm is called *asymptotically consistent* in the mean if

$$\limsup_{T\to\infty} \frac{1}{T} E(s_{1:T} - \min_{i=1,\dots N} s^i_{1:T}) \leq 0. \tag{2}$$

A probabilistic learning algorithm is called *Hannan consistent* if

$$\limsup_{T\to\infty} \frac{1}{T} \left(s_{1:T} - \min_{i=1,\dots N} s^i_{1:T} \right) \leq 0 \tag{3}$$

almost surely, where $s_{1:T}$ is its random cumulative loss.

In this section we study the asymptotical consistency of probabilistic learning algorithms in the case of unbounded one-step losses.

Notice that when $0 \leq s^i_t \leq 1$ all expert algorithms have total loss $\leq T$ on first T steps. This is not true for the unbounded case, and there are no reasons to divide the expected regret (2) by T. We change the standard time scaling (2) and (3) on a new scaling based on a new notion of volume of a game. We modify the definition (2) of the normalized expected regret as follows. Define *the volume* of a game at step t

$$v_t = v_0 + \sum_{j=1}^{t} \max_i |s^i_j|,$$

where v_0 is a nonnegative constant. Evidently, $v_{t-1} \leq v_t$ for all t.

A probabilistic learning algorithm is called *asymptotically consistent* in the mean (in the modified sense) in a game with N experts if

$$\limsup_{T\to\infty} \frac{1}{v_T} E(s_{1:T} - \min_{i=1,\dots N} s^i_{1:T}) \leq 0.$$

A probabilistic algorithm is called Hannan consistent (in the modified sense) if

$$\limsup_{T\to\infty} \frac{1}{v_T} \left(s_{1:T} - \min_{i=1,\dots N} s^i_{1:T} \right) \leq 0 \tag{4}$$

almost surely.

Notice that the notions of asymptotic consistency in the mean and Hannan consistency may be non-equivalent for unbounded one-step losses.

A game is called *non-degenerate* if $v_t \to \infty$ as $t \to \infty$.

Denote $\Delta v_t = v_t - v_{t-1}$. The number

$$\text{fluc}(t) = \frac{\Delta v_t}{v_t} = \frac{\max_i |s_t^i|}{v_t},$$

is called *scaled fluctuation* of the game at the step t.

By definition $0 \le \text{fluc}(t) \le 1$ for all t (put $0/0 = 0$).

The following simple proposition shows that each probabilistic learning algorithm is not asymptotically optimal in some game such that $\text{fluc}(t) \not\to 0$ as $t \to \infty$. For simplicity, we consider the case of two experts and nonnegative losses.

Proposition 1 *For any probabilistic algorithm of choosing an expert and for any* ε *such that* $0 < \varepsilon < 1$ *two experts exist such that* $v_t \to \infty$ *as* $t \to \infty$ *and*

$$\text{fluc}(t) \ge 1 - \varepsilon,$$
$$\frac{1}{v_t} E(s_{1:t} - \min_{i=1,2} s_{1:t}^i) \ge \frac{1}{2}(1 - \varepsilon)$$

for all t.

Proof. Given a probabilistic algorithm of choosing an expert and ε such that $0 < \varepsilon < 1$, define recursively one-step losses s_t^1 and s_t^2 of expert 1 and expert 2 at any step $t = 1, 2, \ldots$ as follows. By $s_{1:t}^1$ and $s_{1:t}^2$ denote the cumulative losses of these experts incurred at steps $\le t$, let v_t be the corresponding volume, where $t = 1, 2, \ldots$.

Define $v_0 = 1$ and $M_t = 4v_{t-1}/\varepsilon$ for all $t \ge 1$. For $t \ge 1$, define $s_t^1 = M_t$ and $s_t^2 = 0$ if $P\{I_t = 1\} \ge \frac{1}{2}$, and define $s_t^1 = 0$ and $s_t^2 = M_t$ otherwise.

Let s_t be one-step loss of the master algorithm and $s_{1:t}$ be its cumulative loss at step $t \ge 1$. We have

$$E(s_{1:t}) \ge E(s_t) = s_t^1 P\{I_t = 1\} + s_t^2 P\{I_t = 2\} \ge \frac{1}{2} M_t$$

for all $t \ge 1$. Also, since $v_t = v_{t-1} + M_t = (1 + 4/\varepsilon) v_{t-1}$ and $\min_i s_{1:t}^i \le v_{t-1}$, the normalized expected regret of the master algorithm is bounded from below

$$\frac{1}{v_t} E(s_{1:t} - \min_i s_{1:t}^i) \ge \frac{2/\varepsilon - 1}{1 + 4/\varepsilon} \ge \frac{1}{2}(1 - \varepsilon).$$

for all t. By definition

$$\text{fluc}(t) = \frac{M_t}{v_{t-1} + M_t} = \frac{1}{1 + \varepsilon/4} \ge 1 - \varepsilon$$

for all t. △

Proposition 1 shows that we should impose some restrictions of asymptotic behavior of $\text{fluc}(t)$ to prove the asymptotic consistency of a probabilistic algorithm.

3. Follow the Perturbed Leader Algorithm with Adaptive Weights

In this section we construct the FPL algorithm with adaptive weights protected from unbounded one-step losses.

Let $\gamma(t)$ be a computable non-increasing real function such that $0 < \gamma(t) < 1$ for all t. In that follows we usually assume $\gamma(t) \to 0$ as $t \to \infty$; for example, $\gamma(t) = 1/(t+c)^{\delta}$, where $c > 0$ and $\delta > 0$. Let also a be a positive real number. Define

$$\alpha_t = \frac{1}{2}\left(1 - \frac{\ln\frac{a(1+\ln N)}{2(e^{4/a}-1)}}{\ln\gamma(t)}\right) \text{ and} \tag{5}$$

$$\mu_t = a(\gamma(t))^{\alpha_t} = \sqrt{\frac{2a(e^{4/a}-1)}{(1+\ln N)}}(\gamma(t))^{1/2} \tag{6}$$

for all t, where $e = 2.72\dots$ is the base of the natural logarithm.[2]

Without loss of generality we suppose that $\gamma(t) < \min\{A, A^{-1}\}$ for all t, where

$$A = \frac{2(e^{4/a}-1)}{a(1+\ln N)}.$$

We can obtain this choosing an appropriate value of the initial constant v_0. Then $0 < \alpha_t < 1$ for all t.

We consider an FPL algorithm with a variable learning rate

$$\varepsilon_t = \frac{1}{\mu_t v_{t-1}}, \tag{7}$$

where μ_t is defined by (6) and the volume v_{t-1} depends on experts actions on steps $< t$. By definition $v_t \geq v_{t-1}$ and $\mu_t \leq \mu_{t-1}$ for $t = 1,2,\dots$. Also, by definition $\mu_t \to 0$ as $t \to \infty$ if $\gamma(t) \to 0$.

Let $\xi_t^1, \dots \xi_t^N$, $t = 1,2,\dots$, be a sequence of i.i.d random variables distributed exponentially with the density $p(x) = \exp\{-x\}$. In what follows we omit the lower index t.

We suppose without loss of generality that $s_0^i = v_0 = 0$ for all i and $\varepsilon_0 = \infty$.

The FPL algorithm PROT is defined on Figure 2.

Let $s_{1:T} = \sum_{t=1}^{T} s_t^{I_t}$ be the cumulative loss of the FPL algorithm on steps $\leq T$.

The following theorem shows that if the game is non-degenerate and $\Delta v_t = o(v_t)$ as $t \to \infty$ with a computable bound then the FPL-algorithm with variable learning rate (7) is asymptotically consistent.

We suppose that the experts are oblivious, that is, they do not use in their work random actions of the learning algorithm. The inequality (9) of Theorem 2 below is reformulated and proved for non-oblivious experts at the end this section.

Theorem 2 *Assume that a computable non-increasing real function* $\gamma(t)$ *exists such that* $0 < \gamma(t) < 1$ *and*

$$\text{fluc}(t) \leq \gamma(t) \tag{8}$$

2. The choice of the optimal value of α_t will be explained later. It will be obtained by minimization of the corresponding member of the sum (36).

FPL algorithm PROT.
FOR $t = 1, \dots T$
Choose an expert with the minimal perturbed cumulated loss on steps $< t$

$$I_t = \mathrm{argmin}_{i=1,2,\dots N} \left\{ s^i_{1:t-1} - \frac{1}{\varepsilon_t} \xi^i \right\}.$$

Receive one-step losses s^i_t for experts $i = 1, \dots, N$, define ε_{t+1} by (7), and

$$v_t = v_{t-1} + \max_i s^i_t.$$

Receive one-step loss $s_t = s^{I_t}_t$ of the master algorithm.
ENDFOR

Figure 2: FPL algorithm PROT

for all t. Then for any $\varepsilon > 0$ the expected cumulated loss of the FPL *algorithm* PROT *with variable learning rate (7), where a parameter $a > 0$ depends on ε, is bounded:*

$$E(s_{1:T}) \le \min_i s^i_{1:T} + 2\sqrt{(8+\varepsilon)(1+\ln N)} \sum_{t=1}^{T} (\gamma(t))^{1/2} \Delta v_t \tag{9}$$

for all t.[3]

In case of nonnegative unbounded losses $s^i_t \in [0, +\infty)$ we have a bound

$$E(s_{1:T}) \le \min_i s^i_{1:T} + 2\sqrt{(2+\varepsilon)(1+\ln N)} \sum_{t=1}^{T} (\gamma(t))^{1/2} \Delta v_t. \tag{10}$$

Let also, the game be non-degenerate and $\gamma(t) \to 0$ as $t \to \infty$. Then the algorithm PROT *is asymptotically consistent in the mean*

$$\limsup_{T\to\infty} \frac{1}{v_T} E(s_{1:T} - \min_{i=1,\dots N} s^i_{1:T}) \le 0. \tag{11}$$

Proof. The proof of this theorem follows the proof-scheme of Kalai and Vempala (2003) and Hutter and Poland (2004).

Let α_t be a sequence of real numbers defined by (5); recall that $0 < \alpha_t < 1$ for all t.

The analysis of optimality of the FPL algorithm is based on an intermediate predictor IFPL (Infeasible FPL) (see Figure 3) with the learning rate ε'_t defined by (12).

The IFPL algorithm predicts under the knowledge of $s^i_{1:t}$, $i = 1, \dots N$, which may not be available at beginning of step t. Using unknown value of ε'_t is the main distinctive feature of our version of IFPL.

The expected one-step and cumulated losses of the FPL and IFPL algorithms at steps t and T are denoted

$$l_t = E(s^{I_t}_t) \text{ and } r_t = E(s^{J_t}_t),$$

3. In case of bounded losses when $\Delta v_t = 1$ we have $v_t = t$ and $\gamma(t) = 1/t$. In this case the regret in the bound (9) has a standard form $O(\sqrt{T \ln N})$.

IFPL algorithm.
FOR $t = 1, \dots T$
Define the learning rate

$$\varepsilon'_t = \frac{1}{\mu_t v_t}, \text{ where } \mu_t = a(\gamma(t))^{\alpha_t}, \tag{12}$$

where v_t is the volume of the game at step t and α_t is defined by (5).
Choose an expert with the minimal perturbed cumulated loss on steps $\le t$

$$J_t = \text{argmin}_{i=1,2,\dots N}\{s^i_{1:t} - \frac{1}{\varepsilon'_t}\xi^i\}.$$

Receive the one step loss $s^{J_t}_t$ of the IFPL algorithm.
ENDFOR

Figure 3: IFPL algorithm

$$l_{1:T} = \sum_{t=1}^{T} l_t \text{ and } r_{1:T} = \sum_{t=1}^{T} r_t,$$

respectively, where $s^{I_t}_t$ is the one-step loss of the FPL algorithm at step t and $s^{J_t}_t$ is the one-step loss of the IFPL algorithm, and E denotes the mathematical expectation. Recall that $I_t = \text{argmin}_i\{s^i_{1:t-1} - \frac{1}{\varepsilon_t}\xi^i\}$ and $J_t = \text{argmin}_i\{s^i_{1:t} - \frac{1}{\varepsilon'_t}\xi^i\}$.

Lemma 3 *The cumulated expected losses of the* FPL *and* IFPL *algorithms with learning rates defined by (7) and (12) satisfy the inequality*

$$l_{1:T} \le r_{1:T} + 2(e^{4/a} - 1)\sum_{t=1}^{T}(\gamma(t))^{1-\alpha_t}\Delta v_t \tag{13}$$

for all T, where α_t is defined by (5).

Proof. Let $c_1, \dots c_N$ be arbitrary nonnegative real numbers. For any $1 \le j \le N$ define

$$m_j = \min_{i \ne j}\{s^i_{1:t-1} - \frac{1}{\varepsilon_t}c_i\},$$
$$m'_j = \min_{i \ne j}\{s^i_{1:t} - \frac{1}{\varepsilon'_t}c_i\}.$$

Assume that these minima are achieved at $i = j_1$ and $i = j_2$ correspondingly:

$$m_j = s^{j_1}_{1:t-1} - \frac{1}{\varepsilon_t}c_{j_1},$$
$$m'_j = s^{j_2}_{1:t} - \frac{1}{\varepsilon'_t}c_{j_2} = s^{j_2}_{1:t-1} + s^{j_2}_t - \frac{1}{\varepsilon'_t}c_{j_2}$$

for some j_1 and j_2. By definition $j_1 \neq j$ and $j_2 \neq j$. Then we have

$$m_j = s^{j_1}_{1:t-1} - \frac{1}{\varepsilon_t} c_{j_1} \leq s^{j_2}_{1:t-1} - \frac{1}{\varepsilon_t} c_{j_2} \leq$$

$$\leq s^{j_2}_{1:t-1} + s^{j_2}_t + \Delta\nu_t - \frac{1}{\varepsilon_t} c_{j_2} = \tag{14}$$

$$= s^{j_2}_{1:t} + \Delta\nu_t - \frac{1}{\varepsilon'_t} c_{j_2} + \left(\frac{1}{\varepsilon'_t} - \frac{1}{\varepsilon_t}\right) c_{j_2} =$$

$$= m'_j + \Delta\nu_t + \left(\frac{1}{\varepsilon'_t} - \frac{1}{\varepsilon_t}\right) c_{j_2}. \tag{15}$$

We add $\Delta\nu_t$ to the right-hand side of the inequality (14) since the term $s^{j_2}_t$ may be negative in case of signed losses. In case of nonnegative losses we need not to do this.

Comparing conditional probabilities

$$P\{I_t = j | \xi^i = c_i, i \neq j\} \text{ and } P\{J_t = j | \xi^i = c_i i \neq j\}$$

is the core of the proof of the lemma.

The following chain of equalities and inequalities is valid:

$$P\{I_t = j | \xi^i = c_i, i \neq j\} =$$

$$= P\{s^j_{1:t-1} - \frac{1}{\varepsilon_t}\xi^j \leq m_j | \xi^i = c_i, i \neq j\} =$$

$$= P\{\xi^j \geq \varepsilon_t (s^j_{1:t-1} - m_j) | \xi^i = c_i, i \neq j\} =$$

$$= P\{\xi^j \geq \varepsilon'_t (s^j_{1:t-1} - m_j) + (\varepsilon_t - \varepsilon'_t)(s^j_{1:t-1} - m_j) | \xi^i = c_i, i \neq j\} \leq \tag{16}$$

$$\leq P\{\xi^j \geq \varepsilon'_t (s^j_{1:t-1} - m_j) +$$

$$+ (\varepsilon_t - \varepsilon'_t)(s^j_{1:t-1} - s^{j_2}_{1:t-1} + \frac{1}{\varepsilon_t} c_{j_2}) | \xi^i = c_i, i \neq j\} = \tag{17}$$

$$= P\{\xi^j \geq \varepsilon'_t (s^j_{1:t-1} - m_j) +$$

$$+ (\varepsilon_t - \varepsilon'_t)(s^j_{1:t-1} - s^{j_2}_{1:t-1}) + (\varepsilon_t - \varepsilon'_t)\frac{1}{\varepsilon_t} c_{j_2} | \xi^i = c_i, i \neq j\} \leq \tag{18}$$

$$\leq P\{\xi^j \geq \varepsilon'_t (s^j_{1:t} - s^j_t - m'_j - \Delta\nu_t - \left(\frac{1}{\varepsilon'_t} - \frac{1}{\varepsilon_t}\right) c_{j_2}) +$$

$$+ (\varepsilon_t - \varepsilon'_t)(s^j_{1:t-1} - s^{j_2}_{1:t-1}) + (\varepsilon_t - \varepsilon'_t)\frac{1}{\varepsilon_t} c_{j_2} | \xi^i = c_i, i \neq j\} = \tag{19}$$

$$= P\{\xi^j \geq \varepsilon'_t (s^j_{1:t} - m'_j) +$$

$$+ (\varepsilon_t - \varepsilon'_t)(s^j_{1:t-1} - s^{j_2}_{1:t-1}) - \varepsilon'_t (s^j_t + \Delta\nu_t) | \xi^i = c_i, i \neq j\} =$$

$$= P\{\xi^j > \frac{1}{\mu_t \nu_t}(s^j_{1:t} - m'_j) +$$

$$+ \left(\frac{1}{\mu_t \nu_{t-1}} - \frac{1}{\mu_t \nu_t}\right)(s^j_{1:t-1} - s^{j_2}_{1:t-1}) - \frac{s^j_t + \Delta\nu_t}{\mu_t \nu_t} | \xi^i = c_i, i \neq j\} \leq \tag{20}$$

$$\leq P\{\xi^j > \frac{1}{\mu_t \nu_t}(s^j_{1:t} - m'_j) +$$

$$+\frac{\Delta v_t}{\mu_t v_t}\frac{(s^j_{1:t-1}-s^{j_2}_{1:t-1})}{v_{t-1}}-\frac{2\Delta v_t}{\mu_t v_t}|\xi^i=c_i, i\neq j\}\le \quad (21)$$

$$\le \exp\left\{\frac{\Delta v_t}{\mu_t v_t}\left|\frac{s^j_{1:t-1}-s^{j_2}_{1:t-1}}{v_{t-1}}-2\right|\right\}\times \quad (22)$$

$$\times P\{J_t=j|\xi^i=c_i, i\neq j\}. \quad (23)$$

Here the inequality (16)–(17) follows from (14) and $\varepsilon_t\ge\varepsilon'_t$. The inequality (18–(19) follows from (15). We have used in transition from (20) to (21) the equality $v_t-v_{t-1}=\Delta v_t$ and the inequality $|s^j_t|\le\Delta v_t$ for all j and t. We have used in transition from (21) to (22)–(23) the inequality $P\{\xi\ge a+b\}\le e^{|b|}P\{\xi\ge a\}$ for any random variable ξ distributed according to the exponential law, where a and b are arbitrary real numbers.[4]

We have in (22)

$$\left|\frac{s^j_{1:t-1}-s^{j_2}_{1:t-1}}{v_{t-1}}\right|\le 2, \quad (24)$$

since $\left|\frac{s^i_{1:t-1}}{v_{t-1}}\right|\le 1$ for all t and i. Also, $\Delta v_t/v_t\le\gamma(t)$ and $\mu_t=a(\gamma(t))^{\alpha_t}$. Therefore, we obtain

$$P\{I_t=j|\xi^i=c_i, i\neq j\}\le$$
$$\le \exp\left\{\frac{4}{\mu_t}\frac{\Delta v_t}{v_t}\right\}P\{J_t=j|\xi^i=c_i, i\neq j\}\le$$
$$\le \exp\{(4/a)(\gamma(t))^{1-\alpha_t}\}P\{J_t=j|\xi^i=c_i, i\neq j\}. \quad (25)$$

Since, the inequality (25) holds for all c_i, it also holds unconditionally :

$$P\{I_t=j\}\le \exp\{(4/a)(\gamma(t))^{1-\alpha_t}\}P\{J_t=j\}. \quad (26)$$

for all $t=1,2,\dots$ and $j=1,\dots N$.

Since $s^j_t+\Delta v_t\ge 0$ for all j and t, we obtain from (26)

$$l_t+\Delta v_t=E(s^{I_t}_t+\Delta v_t)=\sum_{j=1}^N(s^j_t+\Delta v_t)P(I_t=j)\le$$
$$\le \exp\{(4/a)(\gamma(t))^{1-\alpha_t}\}\sum_{j=1}^N(s^j_t+\Delta v_t)P(J_t=j)=$$
$$=\exp\{(4/a)(\gamma(t))^{1-\alpha_t}\}(E(s^{J_t}_t)+\Delta v_t)=$$
$$=\exp\{(4/a)(\gamma(t))^{1-\alpha_t}\}(r_t+\Delta v_t)\le$$
$$\le(1+(e^{4/a}-1))(\gamma(t))^{1-\alpha_t})(r_t+\Delta v_t)=$$
$$=r_t+\Delta v_t+(e^{4/a}-1)(\gamma(t))^{1-\alpha_t}(r_t+\Delta v_t)\le$$
$$\le r_t+\Delta v_t+2(e^{4/a}-1)(\gamma(t))^{1-\alpha_t}\Delta v_t. \quad (27)$$

4. For $a\le 0$, we have $P\{\xi\ge a+b\}\le e^{|b|}P\{\xi\ge a\}$ for all b, since $P\{\xi\ge a\}=1$ and $P\{\xi\ge a+b\}\le 1$; for $a>0$, $P\{\xi\ge a+b\}\le e^{-b}P\{\xi\ge a\}$ for all b.

In the last line of (27) we have used the inequality $|r_t| \le \Delta v_t$ for all t and the inequality $e^{sr} \le 1+(e^s-1)r$ for all $0 \le r \le 1$ and $s > 0$.

Subtracting Δv_t from both sides of the inequality (27) and summing it by $t = 1, \dots T$, we obtain

$$l_{1:T} \le r_{1:T} + 2(e^{4/a}-1)\sum_{t=1}^{T}(\gamma(t))^{1-\alpha_t}\Delta v_t$$

for all T. Lemma 3 is proved. △

The following lemma, which is an analogue of the result of Kalai and Vempala (2003), gives a bound for the IFPL algorithm.

Lemma 4 *The expected cumulative loss of the* IFPL *algorithm with the learning rate (12) is bounded :*

$$r_{1:T} \le \min_i s^i_{1:T} + a(1+\ln N)\sum_{t=1}^{T}(\gamma(t))^{\alpha_t}\Delta v_t \tag{28}$$

for all T, where α_t is defined by (5).

Proof. The proof is along the line of the proof of Hutter and Poland (2004) with an exception that now the sequence ϵ'_t is not monotonic.

Let in this proof, $\mathbf{s_t} = (s^1_t, \dots s^N_t)$ be a vector of one-step losses and $\mathbf{s_{1:t}} = (s^1_{1:t}, \dots s^N_{1:t})$ be a vector of cumulative losses of the experts algorithms. Also, let $\xi = (\xi^1, \dots \xi^N)$ be a vector whose coordinates are random variables.

Recall that $\epsilon'_t = 1/(\mu_t v_t)$, $\mu_t \le \mu_{t-1}$ for all t, and $v_0 = 0$, $\epsilon'_0 = \infty$.

Define $\tilde{\mathbf{s}}_{\mathbf{1:t}} = \mathbf{s_{1:t}} - \frac{1}{\epsilon'_t}\xi$ for $t = 1, 2, \dots$. Consider the vector of one-step losses $\tilde{\mathbf{s}}_{\mathbf{t}} = \mathbf{s_t} - \xi\left(\frac{1}{\epsilon'_t} - \frac{1}{\epsilon'_{t-1}}\right)$ for the moment.

For any vector $\mathbf{s}$ and a unit vector $\mathbf{d}$ denote

$$M(\mathbf{s}) = \text{argmin}_{\mathbf{d}\in D}\{\mathbf{d}\cdot\mathbf{s}\},$$

where $D = \{(0,\dots 1), \dots, (1, \dots 0)\}$ is the set of N unit vectors of dimension N and "$\cdot$" is the inner product of two vectors.

We first show that

$$\sum_{t=1}^{T} M(\tilde{\mathbf{s}}_{\mathbf{1:t}})\cdot\tilde{\mathbf{s}}_{\mathbf{t}} \le M(\tilde{\mathbf{s}}_{\mathbf{1:T}})\cdot\tilde{\mathbf{s}}_{\mathbf{1:T}}. \tag{29}$$

For $T = 1$ this is obvious. For the induction step from $T-1$ to T we need to show that

$$M(\tilde{\mathbf{s}}_{\mathbf{1:T}})\cdot\tilde{\mathbf{s}}_{\mathbf{T}} \le M(\tilde{\mathbf{s}}_{\mathbf{1:T}})\cdot\tilde{\mathbf{s}}_{\mathbf{1:T}} - M(\tilde{\mathbf{s}}_{\mathbf{1:T-1}})\cdot\tilde{\mathbf{s}}_{\mathbf{1:T-1}}.$$

This follows from $\tilde{\mathbf{s}}_{\mathbf{1:T}} = \tilde{\mathbf{s}}_{\mathbf{1:T-1}} + \tilde{\mathbf{s}}_{\mathbf{T}}$ and

$$M(\tilde{\mathbf{s}}_{\mathbf{1:T}})\cdot\tilde{\mathbf{s}}_{\mathbf{1:T-1}} \ge M(\tilde{\mathbf{s}}_{\mathbf{1:T-1}})\cdot\tilde{\mathbf{s}}_{\mathbf{1:T-1}}.$$

We rewrite (29) as follows

$$\sum_{t=1}^{T} M(\tilde{\mathbf{s}}_{\mathbf{1:t}})\cdot\mathbf{s_t} \le M(\tilde{\mathbf{s}}_{\mathbf{1:T}})\cdot\tilde{\mathbf{s}}_{\mathbf{1:T}} + \sum_{t=1}^{T} M(\tilde{\mathbf{s}}_{\mathbf{1:t}})\cdot\xi\left(\frac{1}{\epsilon'_t} - \frac{1}{\epsilon'_{t-1}}\right). \tag{30}$$

By definition of M we have

$$M(\tilde{\mathbf{s}}_{\mathbf{1:T}})\cdot\tilde{\mathbf{s}}_{\mathbf{1:T}} \le M(\mathbf{s}_{\mathbf{1:T}})\cdot\left(\mathbf{s}_{\mathbf{1:T}} - \frac{\xi}{\varepsilon'_T}\right) = \\ = \min_{\mathbf{d}\in D}\{\mathbf{d}\cdot\mathbf{s}_{\mathbf{1:T}}\} - M(\mathbf{s}_{\mathbf{1:T}})\cdot\frac{\xi}{\varepsilon'_T} \ . \quad (31)$$

The expectation of the last term in (31) is equal to $\frac{1}{\varepsilon'_T} = \mu_T v_T$.

The second term of (30) can be rewritten

$$\sum_{t=1}^{T} M(\tilde{\mathbf{s}}_{\mathbf{1:t}})\cdot\xi\left(\frac{1}{\varepsilon'_t} - \frac{1}{\varepsilon'_{t-1}}\right) = \\ = \sum_{t=1}^{T}(\mu_t v_t - \mu_{t-1}v_{t-1})M(\tilde{\mathbf{s}}_{\mathbf{1:t}})\cdot\xi \ . \quad (32)$$

We will use the standard inequality for the mathematical expectation E

$$0 \le E(M(\tilde{\mathbf{s}}_{\mathbf{1:t}})\cdot\xi) \le E(\max_i \xi^i) \le 1 + \ln N. \quad (33)$$

The proof of this inequality uses ideas from Kalai and Vempala (2003) and Hutter and Poland (2004) (Lemma 1).

We have for the exponentially distributed random variables ξ^i, $i = 1,\dots N$,

$$P\{\max_i \xi^i \ge a\} = P\{\exists i(\xi^i \ge a)\} \le \sum_{i=1}^{N} P\{\xi^i \ge a\} = N\exp\{-a\}. \quad (34)$$

Since for any non-negative random variable η, $E(\eta) = \int\limits_0^\infty P\{\eta \ge y\}dy$, by (34) we have

$$E(\max_i \xi^i - \ln N) = \\ = \int_0^\infty P\{\max_i \xi^i - \ln N \ge y\}dy \le \\ \le \int_0^\infty N\exp\{-y - \ln N\}dy = 1.$$

Therefore, $E(\max_i \xi^i) \le 1 + \ln N$.

By (33) the expectation of (32) has the upper bound

$$\sum_{t=1}^{T} E(M(\tilde{\mathbf{s}}_{\mathbf{1:t}})\cdot\xi)(\mu_t v_t - \mu_{t-1}v_{t-1}) \le (1 + \ln N)\sum_{t=1}^{T}\mu_t\Delta v_t.$$

Here we have used the inequality $\mu_t \le \mu_{t-1}$ for all t,

Since $E(\xi^i) = 1$ for all i, the expectation of the last term in (31) is equal to

$$E\left(M(\mathbf{s}_{\mathbf{1:T}})\cdot\frac{\xi}{\varepsilon'_T}\right) = \frac{1}{\varepsilon'_T} = \mu_T v_T. \quad (35)$$

Combining the bounds (30)-(32) and (35), we obtain

$$r_{1:T} = E\left(\sum_{t=1}^{T} M(\tilde{\mathbf{s}}_{\mathbf{1:t}})\cdot \mathbf{s_t}\right) \le$$
$$\le \min_i s^i_{1:T} - \mu_T v_T + (1+\ln N)\sum_{t=1}^{T}\mu_t\Delta v_t \le$$
$$\le \min_i s^i_{1:T} + (1+\ln N)\sum_{t=1}^{T}\mu_t\Delta v_t.$$

Lemma is proved. $\triangle$.

We finish now the proof of the theorem.

The inequality (13) of Lemma 3 and the inequality (28) of Lemma 4 imply the inequality

$$E(s_{1:T}) \le \min_i s^i_{1:T} +$$
$$+\sum_{t=1}^{T}(2(e^{4/a}-1)(\gamma(t))^{1-\alpha_t} + a(1+\ln N)(\gamma(t))^{\alpha_t})\Delta v_t. \tag{36}$$

for all T.

The optimal value (5) of α_t can be easily obtained by minimization of each member of the sum (36) by α_t. In this case μ_t is equal to (6) and (36) is equivalent to

$$E(s_{1:T}) \le \min_i s^i_{1:T} + 2\sqrt{2a(e^{4/a}-1)(1+\ln N)}\sum_{t=1}^{T}(\gamma(t))^{1/2}\Delta v_t, \tag{37}$$

where a is a parameter of the algorithm PROT.

Also, for each $\varepsilon > 0$ an a exists such that $2a(e^{4/a}-1) < 8+\varepsilon$. Therefore, we obtain (9).

We have $\sum_{t=1}^{T}\Delta v_t = v_T$ for all T, $v_t \to \infty$ and $\gamma(t) \to 0$ as $t \to \infty$. Then by Toeplitz lemma (see Lemma 9 of Section A)

$$\frac{1}{v_T}\left(2\sqrt{(8+\varepsilon)(1+\ln N)}\sum_{t=1}^{T}(\gamma(t))^{1/2}\Delta v_t\right) \to 0$$

as $T \to \infty$. Therefore, the FPL algorithm PROT is asymptotically consistent in the mean, that is, the relation (11) of Theorem 2 is proved.

In case where all losses are nonnegative: $s^i_t \in [0,+\infty)$, the inequality (24) can be replaced on

$$\left|\frac{s^j_{1:t-1} - s^{j_2}_{1:t-1}}{v_{t-1}}\right| \le 1$$

for all t and i. We need not to add the term Δv_t to the right-hand side of the inequality (14). Also, we need not to add Δv_t to both parts of inequality (27).

In this case an analysis of the proof of Lemma 3 shows that the bound (37) can be replaced on

$$E(s_{1:T}) \le \min_i s^i_{1:T} + 2\sqrt{a(e^{2/a}-1)(1+\ln N)}\sum_{t=1}^{T}(\gamma(t))^{1/2}\Delta v_t,$$

where a is a parameter of the algorithm PROT.

For each $\varepsilon > 0$ an a exists such that $a(e^{2/a}-1) < 2+\varepsilon$. Using this parameter a, we obtain a version of (9) for nonnegative losses—the inequality (10). △

We study now the Hannan consistency of our algorithm.

Theorem 5 *Assume that all conditions of Theorem 2 hold and*

$$\sum_{t=1}^{\infty}(\gamma(t))^2 < \infty. \tag{38}$$

Then the algorithm PROT *is Hannan consistent:*

$$\limsup_{T\to\infty}\frac{1}{v_T}\left(s_{1:T}-\min_{i=1,\ldots N}s^i_{1:T}\right)\le 0 \tag{39}$$

almost surely.

Proof. So far we assumed that perturbations $\xi^1,\ldots,\xi^N$ are sampled only once at time $t=0$. This choice was favorable for the analysis. As it easily seen, under expectation this is equivalent to generating new perturbations $\xi^1_t,\ldots,\xi^N_t$ at each time step t; also, we assume that all these perturbations are i.i.d for $i=1,\ldots,N$ and $t=1,2,\ldots$. Lemmas 3, 4 and Theorem 2 remain valid for this case. This method of perturbation is needed to prove the Hannan consistency of the algorithm PROT.

We use some version of the strong law of large numbers to prove the Hannan consistency of the algorithm PROT.

Proposition 6 *Let $g(x)$ be a positive nondecreasing real function such that $x/g(x)$, $g(x)/x^2$ are non-increasing for $x>0$ and $g(x)=g(-x)$ for all x.*

Let the assumptions of Theorem 2 hold and

$$\sum_{t=1}^{\infty}\frac{g(\Delta v_t)}{g(v_t)} < \infty. \tag{40}$$

Then the FPL *algorithm* PROT *is Hannan consistent, that is, (4) holds as $T\to\infty$ almost surely.*

Proof. The proof is based on the following lemma.

Lemma 7 *Let a_t be a nondecreasing sequence of real numbers such that $a_t\to\infty$ as $t\to\infty$ and X_t be a sequence of independent random variables such that $E(X_t)=0$, for $t=1,2,\ldots$. Let also, $g(x)$ satisfies assumptions of Proposition 6. Then the inequality*

$$\sum_{t=1}^{\infty}\frac{E(g(X_t))}{g(a_t)} < \infty \tag{41}$$

implies

$$\frac{1}{a_T}\sum_{t=1}^{T}X_t \to 0 \tag{42}$$

as $T\to\infty$ almost surely.

The proof of this lemma is given in Section A.

Put $X_t = (s_t - E(s_t))/2$, where s_t is the loss of the FPL algorithm PROT at step t, and $a_t = v_t$ for all t. By definition $|X_t| \le \Delta v_t$ for all t. Then (41) is valid, and by (42)

$$\frac{1}{v_T}(s_{1:T} - E(s_{1:T})) = \frac{1}{v_T}\sum_{t=1}^{T}(s_t - E(s_t)) \to 0$$

as $T \to \infty$ almost surely. This limit and the limit (11) imply (39). △

By Lemma 6 the algorithm PROT is Hannan consistent, since (38) implies (40) for $g(x) = x^2$. Theorem 5 is proved. △

Non-asymptotic version of Theorem 5 can be obtained but this requires more heavy technics from probability theory (see Petrov 1975).

4. Specializations of Theorems 2 and 5

In this section we discuss some special cases of Theorems 2 and 5.

In case of bounded experts losses $s_t^i \in [0,1]$, assume that an auxiliary "bad" expert i_0 exists for which $s_t^{i_0} = 1$ for all t. Then $\Delta v_t = 1$ and the volume becomes time: $v_t = t$ for all t (we put $v_0 = 0$). So, we can take $\gamma(t) = t^{-1}$ for all t. In this case the regret (10) of Theorem 2 is equal to $4\sqrt{(2+\varepsilon)(1+\ln N)T}$ that is very close to classical bounds from Hutter and Poland (2004), Kalai and Vempala (2003) and Lugosi and Cesa-Bianchi (2006).

Allenberg et al. (2006) and Poland and Hutter (2005) considered polynomially bounded one-step losses. We consider a specific example of the bound (9) for polynomial case.

Corollary 8 *Assume that $|s_t^i| \le t^{\alpha}$ for all t and $i = 1,\dots N$, and $v_t \ge t^{\alpha+\delta}$ for all t, where α and δ are positive real numbers. Let also, in the algorithm* PROT, *$\gamma(t) = t^{-\delta}$ and μ_t is defined by (6). Then*

- *(i) the algorithm* PROT *is asymptotically consistent in the mean for any $\alpha > 0$ and $\delta > 0$;*
- *(ii) this algorithm is Hannan consistent for any $\alpha > 0$ and $\delta > \frac{1}{2}$;*
- *(iii) the expected loss of this algorithm is bounded :*

$$E(s_{1:T}) \le \min_i s_{1:T}^i + 2\sqrt{(8+\varepsilon)(1+\ln N)}T^{1-\frac{1}{2}\delta+\alpha} \tag{43}$$

  as $T \to \infty$, where $\varepsilon > 0$ is a parameter of the algorithm.[5]

This corollary follows directly from Theorem 2, where condition (38) of Theorem 2 holds for $\delta > \frac{1}{2}$.

If $\delta = 1$ the regret from (43) is asymptotically equivalent to the regret from Allenberg et al. (2006) (see Section 1).

For $\alpha = 0$ we have the case of bounded loss function ($|s_t^i| \le 1$ for all i and t). The FPL algorithm PROT is asymptotically consistent in the mean if $v_t \ge \beta(t)$ for all t, where $\beta(t)$ is an arbitrary positive unbounded non-decreasing computable function (we can get $\gamma(t) = 1/\beta(t)$ in this case). This algorithm is Hannan consistent if (38) holds, that is,

$$\sum_{t=1}^{\infty}(\beta(t))^{-2} < \infty.$$

5. Recall that given ε we tune the parameter a of the algorithm PROT.

For example, this condition be satisfied for $\beta(t) = t^{1/2}\ln t$.

Let us show that the bound (9) of Theorem 2 that holds against oblivious experts also holds against non-oblivious (adaptive) ones.

In non-oblivious case, it is natural to generate at each time step t of the algorithm PROT a new vector of perturbations $\bar{\xi}_{\mathbf{t}} = (\xi_t^1, \ldots, \xi_t^N)$, $\bar{\xi}_0$ is empty set. Also, it is assumed that all these perturbations are i.i.d according to the exponential distribution P, where $i = 1, \ldots, N$ and $t = 1, 2, \ldots$. Denote $\bar{\xi}_{\mathbf{1:t}} = (\bar{\xi}_1, \ldots, \bar{\xi}_t)$.

Non-oblivious experts can react at each time step t on past decisions $s_1, s_2, \ldots s_{t-1}$ of the FPL algorithm and on values of $\bar{\xi}_1, \ldots, \bar{\xi}_{t-1}$.

Therefore, losses of experts and regret depend now from random perturbations:

$$
s_t^i = s_t^i(\bar{\xi}_{\mathbf{1:t-1}}),\ i = 1, \ldots, N,
$$
$$
\Delta v_t = \Delta v_t(\bar{\xi}_{\mathbf{1:t-1}}),
$$

where $t = 1, 2, \ldots$.

In non-oblivious case, condition (8) is a random event. We assume in Theorem 2 that in the game of prediction with expert advice regulated by the FPL-protocol the event

$$
\mathrm{fluc}(t) \leq \gamma(t) \text{ for all } t
$$

holds almost surely.

An analysis of the proof of Theorem 2 shows that in non-oblivious case, the bound (9) is an inequality for the random variable

$$
\sum_{t=1}^{T} E(s_t) - \min_i s_{1:T}^i -
$$
$$
-2\sqrt{(8+\varepsilon)(1+\ln N)} \sum_{t=1}^{T} (\gamma(t))^{1/2} \Delta v_t \leq 0,
$$

which holds almost surely with respect to the product distribution P^{t-1}, where the loss of the FPL algorithm s_t depend on a random perturbation ξ_t at step t and on losses of all experts on steps $< t$. Also, E is the expectation with respect to P.

Taking expectation $E_{1:T-1}$ with respect to the product distribution P^{t-1} we obtain a version of (9) for non-oblivious case

$$
E_{1:T}\left(s_{1:T} - \min_i s_{1:T}^i - 2\sqrt{(8+\varepsilon)(1+\ln N)} \sum_{t=1}^{T} (\gamma(t))^{1/2} \Delta v_t\right) \leq 0
$$

for all T.

5. An Example: Zero-sum Experts

In this section we present an example of a game, where losses of experts cannot be bounded in advance.[6]

6. This example is a modified version of an example from V'yugin (2009a).

Let $S = S(t)$ be a function representing evolution of a stock price. Two experts will represent two concurrent methods of buying and selling shares of this stock.

Let M and T be positive integer numbers and let the time interval $[0,T]$ be divided on a large number M of subintervals. Define a discrete time series of stock prices

$$S_0 = S(0), S_1 = S(T/(M)), S_2 = S(2T/(M)) \dots, S_M = S(T). \tag{44}$$

In this paper, volatility is an informal notion. We say that the difference $(S_T - S_0)^2$ represents the macro volatility and the sum $\sum\limits_{i=0}^{T-1}(\Delta S_i)^2$, where $\Delta S_i = S_{i+1} - S_i$, $i = 1, \dots T-1$, represents the micro volatility of the time series (44).

The game between an investor and the market looks as follows: the investor can use the long and short selling. At beginning of time step t Investor purchases the number C_t of shares of the stock by S_{t-1} each. At the end of trading period the market discloses the price S_{t+1} of the stock, and the investor incur his current income or loss $s_t = C_t \Delta S_t$ at the period t. We have the following equality

$$(S_T - S_0)^2 = (\sum_{t=0}^{T-1} \Delta S_t)^2 =$$
$$= \sum_{t=0}^{T-1} 2(S_t - S_0)\Delta S_t + \sum_{t=0}^{T-1}(\Delta S_t)^2. \tag{45}$$

The equality (45) leads to the two strategies for investor which are represented by two experts. At the beginning of step t Experts 1 and 2 hold the number of shares

$$C_t^1 = 2C(S_t - S_0), \tag{46}$$
$$C_t^2 = -C_t^1, \tag{47}$$

where C is an arbitrary positive constant.

These strategies at step t earn the incomes $s_t^1 = 2C(S_t - S_0)\Delta S_t$ and $s_t^2 = -s_t^1$. The strategy (46) earns in first T steps of the game the income

$$s_{1:T}^1 = \sum_{t=1}^{T} s_t^1 = 2C((S_T - S_0)^2 - \sum_{t=1}^{T-1}(\Delta S_t)^2).$$

The strategy (47) earns in first T steps the income $s_{1:T}^2 = -s_{1:T}^1$.

The number of shares C_t^1 in the strategy (46) or number of shares $C_t^2 = -C_t^1$ in the strategy (47) can be positive or negative. The one-step gains s_t^1 and $s_t^2 = -s_t^1$ are unbounded and can be positive or negative: $s_t^i \in (-\infty, +\infty)$.

Informally speaking, the first strategy will show a large return if

$$(S_T - S_0)^2 \gg \sum_{i=0}^{T-1}(\Delta S_i)^2;$$

the second one will show a large return when

$$(S_T - S_0)^2 \ll \sum_{i=0}^{T-1}(\Delta S_i)^2.$$

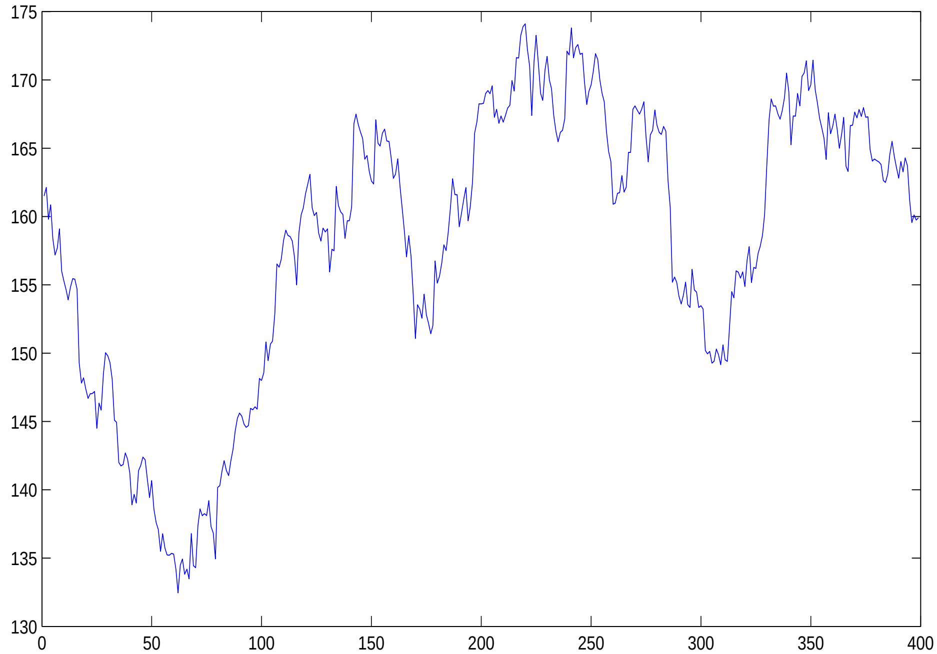

Figure 4: Evolution of Gazprom stock price

There is an uncertainty domain for these strategies, that is, a case when both $\gg$ and $\ll$ do not hold. The idea of these strategies is based on the paper of Cheredito (2003) (see also Rogers 1997 and Delbaen and Schachermayer 1994) who have constructed arbitrage strategies for a financial market that consists of money market account and a stock whose price follows a fractional Brownian motion with drift or an exponential fractional Brownian motion with drift. Vovk (2003) has reformulated these strategies for discrete time. We use these strategies to define a mixed strategy which incur gain when macro and micro volatilities of time series differ. There is no uncertainty domain for continuous time.

We analyze this game in the decision theoretic online learning (DTOL) framework (Freund and Schapire, 1997). We introduce *Learner* that can choose between two strategies (46) and (47). To change from follow the leader framework to DTOL we derandomize the FPL algorithm PROT.[7] We interpret the expected one-step gain $E(s_t)$ gain as the weighted average of one-step gains of experts strategies. In more detail, at each step t, *Learner* divide his investment in proportion to the probabilities of expert strategies (46) and (47) computed by the FPL algorithm and suffers the gain

$$G_t = 2C(S_t - S_0)(P\{I_t = 1\} - P\{I_t = 2\})\Delta S_t$$

at any step t, where C is an arbitrary positive constant; $G_{1:T} = \sum_{t=1}^{T} G_t = E(s_{1:T})$ is the *Learner's* cumulative gain.

7. To apply Theorem 2 we interpreted gain as a negative loss.

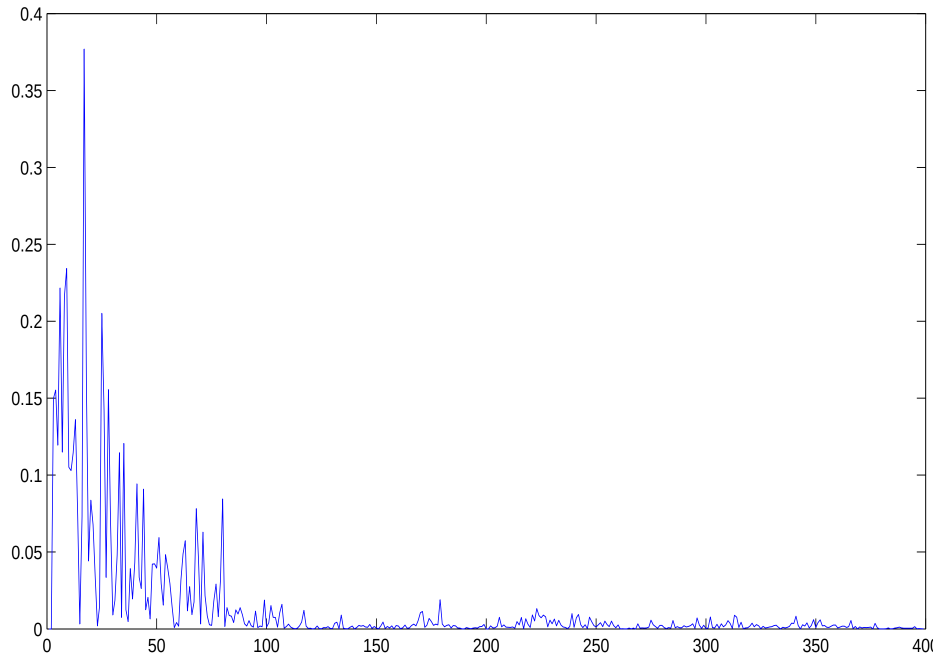

Figure 5: Fluctuation of the game

Assume that $|s_t^1| = o(\sum_{i=1}^t |s_i^1|)$ as $t \to \infty$. Let $\gamma(t) = \mu$ for all t, where μ is arbitrary small positive number. Then for any $\varepsilon > 0$

$$G_{1:T} \geq \left| \sum_{t=1}^{T} s_t^1 \right| - 2\mu^{1/2}\sqrt{(8+\varepsilon)(1+\ln N)} \left(\sum_{t=1}^{T} |s_t^1| + v_0 \right)$$

for all sufficiently large T, and for some $v_0 \geq 0$.

Under condition of Theorem 2 we show that strategy of algorithm PROT is "defensive" in some weak sense :

$$G_{1:T} - \left| \sum_{t=1}^{T} s_t^1 \right| \geq -o\left(\sum_{t=1}^{T} |s_t^1| + v_0 \right) \tag{48}$$

as $T \to \infty$.

Some experimental results are shown on Figures 4–6. The strategies (46) and (47) were applied to the Russian Gazprom stock (ticker symbol—GAZP) downloaded from FINAM site.[8] We get $C = 600$. We have used the stock closing price time series on period from 02 July to 02 September 2009 with periodicity 60 minutes between two neighboring time-points; the size of time series is 400 points. The stock price was volatile during the playing period, its value changed slightly during this period from 163.45 Rub to 159.90 Rub (see Figure 4).

8. FINAM is at `http://www.finam.ru/analysis/export/default.asp`.

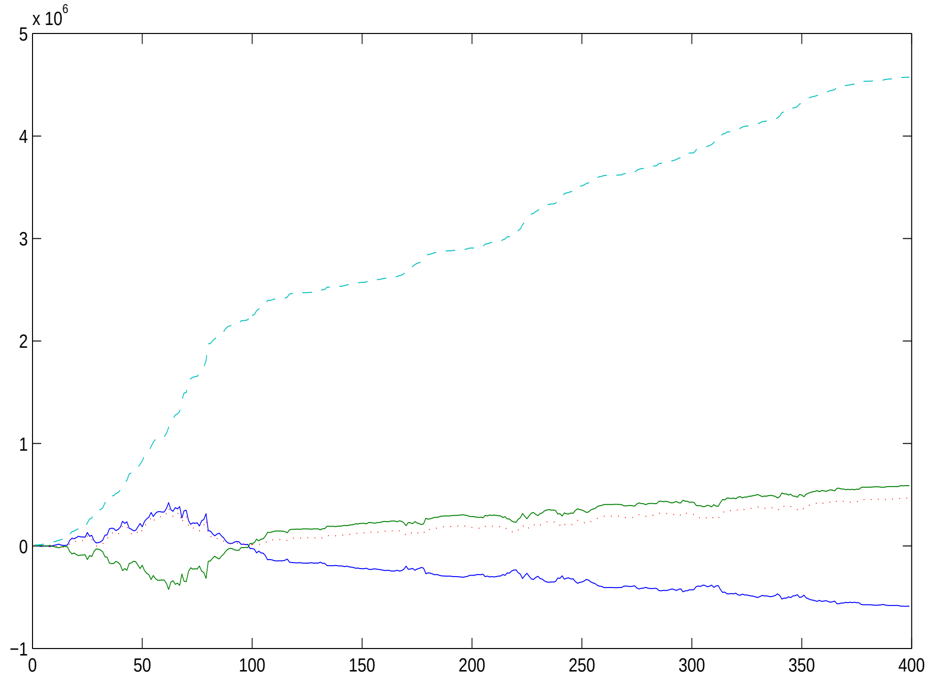

Figure 6: Two symmetric solid lines—gains of two zero sums strategies, dotted line—expected gain of the algorithm PROT (without transaction costs), dashed line—volume of the game

Two symmetric solid lines on Figure 6 are gains of two zero sums strategies (46) and (47), dotted line—expected gain of the algorithm PROT, dashed line—volume of the game. The scaled fluctuation of the game is presented on Figure 5. We get $\gamma(t) = t^{-1/2}$. The first strategy (46) was favorite at about 100 first steps, the second strategy (47) was favorite at the rest of the playing period. The algorithm PROT suffered sufficiently large income—456970 Rub (without transaction costs) (see Figure 6) and 230099 Rub when transaction costs were subtracted.

6. Conclusion

In this paper we try to extend methods of the theory of prediction with expert advice for the case when experts one-step gains cannot be bounded in advance. The traditional measures of performance are invalid for general unbounded case.

To measure the asymptotic performance of our algorithm, we replace the traditional time-scale on a volume-scale. New notion of volume of a game and scaled fluctuation of a game are introduced in this paper. In case of two zero-sum experts, the volume equals to the sum of all transactions between experts.

Using the notion of the scaled fluctuation of a game, we can define very broad classes of games (experts) for which our algorithm PROT is asymptotically consistent in the modified sense. The restrictions imposed on such games are formulated in a relative form: "the logarithmic derivative" of the volume of the game must be $o(t)$ as $t \to \infty$.

Our work supplements results of Cesa-Bianchi et al. (2007), where the bounds for a regret were obtained under the very general assumptions. Authors of this paper do not study asymptotic consistency of their algorithm. Our bounds for the regret are defined in terms of a volume of the game and our learning algorithm is asymptotically consistent in the mean and almost surely.

Algorithms for unbounded losses have appeared in the literature, but none of the papers deal with FPL and "fast-growing" losses. Looking closely at the requirements of this paper, the quantity $\mathrm{fluc}(t)$ has to decrease to 0, which to imply that the rate of growth of the losses has to be slower than exponential. Given the results of Allenberg et al. (2006), who can deal with polynomial growth of loss, this paper is more general in the regime "faster than polynomial, but slower than exponential".

A motivating example of a game with two zero-sum experts from Section 5 shows some practical significance of these problems. The FPL algorithm with variable learning rates is simple to implement and it is bringing satisfactory experimental results when prices follow fractional Brownian motion. The drawback of this playing strategy is that the defense condition (48) is too weak—it has only an asymptotic form. In cases, where regimes of high and low volatilities quickly changing the algorithm PROT may suffer a large loss. This is an open problems for further research: how to construct a defensive strategy for *Learner* in sense of Shafer and Vovk (2001)? This means that *Learner* starting with some initial capital never goes to debt and suffer a gain when macro and micro volatilities differ.

There are other open problems for further research. Can we incorporate our results obtained in fluctuation-volume setting into the framework presented in Cesa-Bianchi et al. (2007), where a powerful technics for the Weighed Majority algorithm based on second order quantity—variance of losses, was developed?

We have used the FPL algorithm, since its analysis remains easy for an adaptive learning rate, in contrast to WM-derivatives. It would be useful to analyze the performance of the well known algorithms from DTOL framework (like "Hedge" of Freund and Schapire 1997 or "Normal Hedge" of Chaudhuri et al. 2009) for the case of unbounded losses in terms of the volume and scaled fluctuation of a game.

Some improvement of the regret (9) can be achieved using in (27) a more tight bound of the exponent $e^r \leq 1 + r + (e-2)r^2$ (for $|r| \leq 1$) in place of the linear bound used in the proof of Lemma 3.

There is a gap between Proposition 1 and Theorem 2, since we assume in this theorem that the game satisfies $\mathrm{fluc}(t) \leq \gamma(t) \to 0$, where $\gamma(t)$ is computable. Also, the function $\gamma(t)$ is a parameter of our algorithm PROT. Does there exists an asymptotically consistent learning algorithm in case where $\limsup_{t\to\infty} \mathrm{fluc}(t) = 0$ and where the function $\gamma(t)$ is not a parameter of this algorithm?

Can we apply "double trick" method for the sequence $\mathrm{fluc}(t)$, $t = 1, 2, \ldots$, to avoid parameter $\gamma(t)$ from the learning algorithm is an open question. A problem is that $\mathrm{fluc}(t)$ is not monotone though $\limsup_{t\to\infty} \mathrm{fluc}(t) = 0$.

Let $\gamma_i(t)$ be any computable (by i and t) sequence of non-increasing (by t) functions such that for any i, $0 < \gamma_i(t) \leq 1$ for all t and $\gamma_i(t) \to 0$ as $t \to \infty$.[9] We can construct a version of the algorithm

9. A case $\sup_i \gamma_i(t) = 1$ for all t is possible for these functions.

PROT which is asymptotically consistent in the mean for any game satisfying

$$\limsup_{t\to\infty} \frac{\mathrm{fluc}(t)}{\gamma_i(t)} < \infty \tag{49}$$

for some i. To solve this problem define a computable non-increasing function $\gamma(t)$ such that

- $0 < \gamma(t) \le 1$,
- $\gamma(t) \to 0$ as $t \to \infty$,
- for any i there exists an t_i such that $\gamma(t) \ge \gamma_i(t)$ for all $t \ge t_i$.

Evidently, the algorithm PROT with the parameter $\gamma(t)$ is asymptotically consistent in the mean for any game such that (49) holds for some i.

We consider in this paper only the full information case. An analysis of these problems under partial monitoring is a subject for a further research.

Acknowledgments

This research was partially supported by Russian foundation for fundamental research: 09-07-00180-a and 09-01-00709a.

Appendix A. Proof of Lemma 7

The proof of Lemma 7 is based on Kolmogorov's theorem on three series and its corollaries. For completeness of presentation we reconstruct the proof from Petrov (1975) (Chapter IX, Section 2).

For any random variable X and a positive number c denote

$$X^c = \begin{cases} X \text{ if } |X| \le c \\ 0 \text{ otherwise.} \end{cases}$$

The Kolmogorov's theorem on three series says:

For any sequence of independent random variables X_t, $t = 1, 2, \ldots$, the following implications hold

- If the series $\sum_{t=1}^{\infty} X_t$ is convergent almost surely then the series $\sum_{t=1}^{\infty} EX_t^c$, $\sum_{t=1}^{\infty} DX_t^c$ and $\sum_{t=1}^{\infty} P\{|X_t| \ge c\}$ are convergent for each $c > 0$, where E is the mathematical expectation and D is the variation.
- The series $\sum_{t=1}^{\infty} X_t$ is convergent almost surely if all these series are convergent for some $c > 0$.

See Shiryaev (1980) for the proof.

Assume conditions of Lemma 7 hold. We will prove that

$$\sum_{t=1}^{\infty} \frac{Eg(X_t)}{g(a_t)} < \infty \tag{50}$$

implies

$$\sum_{t=1}^{\infty} \frac{X_t}{a_t} < \infty$$

almost surely. From this, by Kroneker's lemma 10 (see below), the series

$$\frac{1}{a_t} \sum_{t=1}^{\infty} X_t \tag{51}$$

is convergent almost surely.

Let V_t be a distribution function of the random variable X_t. Since g non-increases,

$$P\{|X_t| > a_t\} \leq \int_{|x| \geq a_t} \frac{g(x)}{g(a_t)} dV_t(x) \leq \frac{Eg(X_t)}{g(a_t)}.$$

Then by (50)

$$\sum_{t=1}^{\infty} P\left\{ \left| \frac{X_t}{a_t} \right| \geq 1 \right\} < \infty \tag{52}$$

almost surely. Denote

$$Z_t = \begin{cases} X_t \text{ if } |X_t| \leq a_t \\ 0 \text{ otherwise.} \end{cases}$$

By definition $x^2/g(x)) \leq a_t/g(a_t)$ for $|x| < a_t$. Rearranging, we obtain $x^2/a_t \leq g(x)/g(a_t)$ for these x. Therefore,

$$EZ_t^2 = \int\limits_{|x|<a_t} x^2 dV_t(x) \leq \frac{a_t^2}{g(a_t)} \int\limits_{|x|<a_t} g(x) dV_t(x) \leq \frac{a_t^2}{g(a_t)} Eg(X_t).$$

By (50) we obtain

$$\sum_{t=1}^{\infty} E\left(\frac{Z_t}{a_t}\right)^2 < \infty. \tag{53}$$

Since $EX_t = \int\limits_{-\infty}^{\infty} x dV_t(x) = 0$,

$$|EZ_t| = \left| \int\limits_{|x|>a_t} x dV_t(x) \right| \leq \frac{a_t}{g(a_t)} \int\limits_{|x|>a_t} g(x) dV_t(x) \leq \frac{a_t}{g(a_t)} Eg(X_t). \tag{54}$$

By (50)

$$\sum_{t=1}^{\infty} E\left(\frac{X_t}{a_t}\right)^1 \leq \sum_{t=1}^{\infty} \left| E\left(\frac{Z_t}{a_t}\right) \right| < \infty.$$

From (52)–(54) and the theorem on three series we obtain (51).

We have used Toeplitz and Kroneker's lemmas.

Lemma 9 *(Toeplitz) Let x_t be a sequence of real numbers and b_t be a sequence of nonnegative real numbers such that $a_t = \sum\limits_{i=1}^{t} b_i \to \infty$, $x_t \to x$ and $|x| < \infty$. Then*

$$\frac{1}{a_t}\sum_{i=1}^{t} b_i x_i \to x. \tag{55}$$

Proof. For any $\varepsilon > 0$ an t_ε exists such that $|x_t - x| < \varepsilon$ for all $t \geq t_\varepsilon$. Then

$$\left|\frac{1}{a_t}\sum_{I=1}^{t} b_i(x_i - x)\right| \leq \frac{1}{a_t}\sum_{i<t_\varepsilon}|b_i(x_i - x)| + \varepsilon$$

for all $t \geq t_\varepsilon$. Since $a_t \to \infty$, we obtain (55).

Lemma 10 *(Kroneker) Assume $\sum\limits_{t=1}^{\infty} x_t < \infty$ and $a_t \to \infty$ Then*

$$\frac{1}{a_t}\sum_{i=1}^{t} a_i x_i \to 0.$$

The proof is the straightforward corollary of Toeplitz lemma.

References

C. Allenberg, P. Auer, L. Gyorfi and Gy. Ottucsak. Hannan consistency in on-Line learning in case of unbounded losses under partial monitoring. In Jose L. Balcazar, Philip M. Long, Frank Stephan, editors, *Proceedings of the 17th International Conference ALT 2006, Lecture Notes in Computer Science 4264*, pages 229–243, Springer-Verlag, Berlin, Heidelberg, 2006.

K. Chaudhuri, Y. Freund and D. Hsu. A parameter-free hedging algorithm. *Technical Report arXiv:0903.2851v1 [cs.LG]*, March 2009.

P. Cheredito. Arbitrage in fractional Brownian motion. *Finance and Statistics*, 7:533–553, 2003.

N. Cesa-Bianchi, Y. Mansour and G. Stoltz. Improved second-order bounds for prediction with expert advice. *Machine Learning*, 66(2–3):321–352, 2007.

F. Delbaen and W. Schachermayer. A general version of the fundamental theorem of asset pricing. *Mathematische Annalen*, 300:463–520, 1994.

Y. Freund and R.E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55:119–139, 1997.

J. Hannan. Approximation to Bayes risk in repeated plays. In M. Dresher, A.W. Tucker, and P. Wolfe, editors, *Contributions to the Theory of Games 3*, pages 97-139, Princeton University Press, 1957.

M. Hutter and J. Poland. Prediction with expert advice by following the perturbed leader for general weights. In S.Ben-Dawid, J.Case, A.Maruoka, editors *Proceedings of the 15th International Conference ALT 2004, Lecture Notes in Artificial Intelligence 3244*, pages 279–293, Springer-Verlag, Berlin, Heidelberg, 2004.

M. Hutter and J. Poland. Adaptive online prediction by following the perturbed leader. *Journal of Machine Learning Research*, 6:639–660, 2005.

A. Kalai and S. Vempala. Efficient algorithms for online decisions. In Bernhard Scholkopf, Manfred K. Warmuth, editors, *Proceedings of the 16th Annual Conference on Learning Theory COLT 2003, Lecture Notes in Computer Science 2777*, pages 506–521, Springer-Verlag, Berlin, 2003. Extended version in *Journal of Computer and System Sciences*, 71:291–307, 2005.

N. Littlestone and M.K. Warmuth. The weighted majority algorithm. *Information and Computation*, 108:212–261, 1994.

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

V.V. Petrov. Sums of independent random variables. *Ergebnisse der Mathematik und Ihrer Grenzgebiete*, Band 82, Springer-Verlag, New York, Heidelberg, Berlin, 1975.

J. Poland and M. Hutter. Defensive universal learning with experts for general weights. In S.Jain, H.U.Simon and E.Tomita, editors, *Proceedings of the 16th International Conference ALT 2005, Lecture Notes in Artificial Intelligence 3734*, pages 356–370, Springer-Verlag, Berlin, Heidelberg, 2005.

C. Rogers. Arbitrage with fractional Brownian motion. *Mathematical Finance*, 7:95–105, 1997.

G. Shafer and V. Vovk. *Probability and Finance. It's Only a Game!*, Wiley, New York, 2001.

A.N. Shiryaev. *Probability*. Springer-Verlag, Berlin, 1980.

V. Vovk. Aggregating strategies. In M. Fulk and J. Case, editors, *Proceedings of the 3rd Annual Workshop on Computational Learning Theory*, pages 371–383, Morgan Kaufmann, San Mateo, CA, 1993.

V. Vovk. A game-theoretic explanation of the $\sqrt{dt}$ effect. Working paper #5, Available online at `http://www.probabilityandfinance.com`, 2003.

V.V. V'yugin. The follow perturbed leader algorithm protected from unbounded one-step losses. In Ricard Gavalda, Gabor Lugosi, Thomas Zeugmann, and Sandra Zilles, editors, *Proceedings of the 20th International Conference ALT 2009, Lecture Notes in Artificial Intelligence 5809*, pages. 38-52, Springer-Verlag, Berlin Heidelberg, 2009.

V.V. V'yugin. Learning volatility of discrete time series using prediction with expert advice. In O. Watanabe and T. Zeugmann, editors, *SAGA 2009, Lecture Notes in Computer Science 5792*, pages 16-30, Springer-Verlag, Berlin, Heidelberg, 2009.