The Follow Perturbed Leader Algorithm Protected from Unbounded One-Step Losses

Vladimir V. V'yugin

Institute for Information Transmission Problems, Russian Academy of Sciences,
Bol'shoi Karetnyi per. 19, Moscow GSP-4, 127994, Russia
vyugin@iitp.ru

Abstract. In this paper the sequential prediction problem with expert advice is considered for the case when the losses of experts suffered at each step can be unbounded. 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 an 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.

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 [6] and Vovk's [11] Aggregating Algorithm, 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 loss functions, like absolute loss, square and logarithmic loss. Arbitrary losses are less common. A survey can be found in the book of Lugosi, Cesa-Bianchi [7].

In this paper, we consider a different general approach - "Follow the Perturbed Leader FPL" algorithm, now called Hannan's algorithm [3], [5], [7]. 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 $\sqrt{2}$.

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

At the beginning of the step t *Learner*, observing cumulating losses $s_{1:t-1}^i = s_1^i + \ldots + s_{t-1}^i$ of all experts $i = 1, \ldots N$, makes a decision to follow one of these

R. Gavaldà et al. (Eds.): ALT 2009, LNAI 5809, pp. 38–52, 2009.

experts, say Expert i. At the end of step t *Learner* receives the same loss s_t^i as Expert i at step t and suffers *Learner's* cumulative loss $s_{1:t} = s_{1:t-1} + s_t^i$.

In the traditional framework, we suppose that one-step losses of all experts are bounded, for example, $0 \le s_t^i \le 1$ 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 $t = 0, 1, \ldots, 6$ 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, the "Follow 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 [3]. Kalai and Vempala [5] 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

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

where ξ^i, $i = 1, \ldots N$, $t = 1, 2, \ldots$, 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 [5] show that the expected cumulative loss of the FPL algorithm has the upper bound

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

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

Hutter and Poland [4] 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: $0 \le s_t^i \le 1$ for all i and t.

Most papers on prediction with expert advice either consider bounded losses or assume the existence of a specific loss function (see [7]). 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 [1], [2], [9].

Poland and Hutter [9] 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. [2] have considered polynomially bounded one-step losses for a modified version of the Littlestone and Warmuth algorithm [6] 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 $i = 1, 2, \ldots N$ 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.[1] 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 $0<a<1$.

In this paper, we consider also the case where the loss grows "faster than polynomial, but slower than exponential".

We present some modification of Kalai and Vempala [5] algorithm of following the perturbed leader (FPL) for the case of unrestrictedly large one-step expert losses s_t^i not bounded in advance. This algorithm uses adaptive weights depending on past cumulative losses of the experts.

We analyze the asymptotic consistency of our algorithms using nonstandard scaling. We introduce new notions of *the volume of a game* $v_t=\sum_{j=1}^{t}\max_i s_j^i$ and *the scaled fluctuation* of the game $\mathrm{fluc}(t)=\Delta v_t/v_t$, where $\Delta v_t=v_t-v_{t-1}$.

We show in Theorem 1 that the algorithm of following the perturbed leader with adaptive weights constructed in Section 2 is asymptotically consistent in the mean in the case when $v_t\to\infty$ 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 $\gamma(t)$ such that $\gamma(t)=o(1)$ as $t\to\infty$, our algorithm has the expected regret

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

where $e=2.72\ldots$ is the base of the natural logarithm.

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

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

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 2 we present some sufficient conditions under which our learning algorithm is Hannan consistent.[2]

In particular case, Corollary 1 of Theorem 1 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 1 shows that our algorithm is also Hannan consistent when $\delta>\frac{1}{2}$.

[1] Allenberg et al. [2] considered losses $-\infty<s_t^i<\infty$.

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

At the end of Section 2 we consider some applications of our algorithm for the case of standard time-scaling.

2 The Follow Perturbed Leader Algorithm with Adaptive Weights

We consider a game of prediction with expert advice with unbounded one-step losses. At each step t of the game, all N experts receive one-step losses $s_t^i \in [0, +\infty)$, $i = 1, \dots N$, and the cumulative loss of the ith expert after step t is equal to

$$s_{1:t}^i = s_{1:t-1}^i + s_t^i.$$

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 $s_{1:t-1}^i$ of the experts $i = 1, \dots N$ in hindsight.

Probabilistic algorithm of choosing an expert

FOR $t = 1, \dots T$

Given past cumulative losses of the experts $s_{1:t-1}^i$, $i = 1, \dots N$, choose an expert i with probability $P\{I_t = i\}$.

Receive the one-step losses at step t of the expert s_t^i and suffer one-step loss $s_t = s_t^i$ of the master algorithm.

ENDFOR

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

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

where the random variable $s_{1:T}$ is the cumulative loss of the master algorithm, $s_{1:T}^i$, $i = 1, \dots N$, are the cumulative losses of the experts algorithms and E is the mathematical expectation (with respect to the probability distribution generated by probabilities $P\{I_t = i\}$, $i = 1, \dots N$, on the first T steps of the game).[3]

In the case of bounded one-step expert losses, $s_t^i \in [0, 1]$, and a convex loss function, the well-known learning algorithms have expected regret $O(\sqrt{T \log N})$ (see Lugosi, Cesa-Bianchi [7]).

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

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

A probabilistic learning algorithm is called *Hannan consistent* if

[3] For simplicity, we suppose that the experts are oblivious, i.e., they cannot use in their work random actions of the learning algorithm. The inequality (12) and the limit (13) of Theorem 1 below can be easily reformulated and proved for non-oblivious experts.

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

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 $0\le s^i_t\le 1$ all expert algorithms have total loss $\le T$ 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

$$v_t=\sum_{j=1}^{t}\max_i s^i_j.$$

Evidently, $v_{t-1}\le v_t$ for all t and $\max_i s^i_{1:t}\le v_t\le N\max_i s^i_{1:t}$ for all i and t.

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

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

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

$$\limsup_{T\to\infty}\frac{1}{v_T}\left(s_{1:T}-\min_{i=1,\dots N}s^i_{1:T}\right)\le 0 \tag{5}$$

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 $v_t\to\infty$ (or equivalently, $\max_i s^i_{1:t}\to\infty$) as $t\to\infty$.

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

$$\mathrm{fluc}(t)=\frac{\Delta v_t}{v_t}=\frac{\max_i s^i_t}{v_t}, \tag{6}$$

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

By definition $0\le\mathrm{fluc}(t)\le 1$ 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 $\mathrm{fluc}(t)\not\to 0$ as $t\to\infty$. For simplicity, we consider the case of two experts.

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

$$\mathrm{fluc}(t)\ge 1-\epsilon,$$
$$\frac{1}{v_t}E(s_{1:t}-\min_{i=1,2}s^i_{1:t})\ge\frac{1}{2}(1-\epsilon)$$

for all t.

Proof. Given a probabilistic algorithm of choosing an expert and ϵ such that $0 < \epsilon < 1$, 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 $\leq t$, let v_t be the corresponding volume, where $t = 1, 2, \ldots$.

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

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

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

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

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

for all t. By definition

$$\mathrm{fluc}(t) = \frac{M_t}{v_{t-1} + M_t} = \frac{1}{1 + \epsilon/4} \geq 1 - \epsilon$$

for all t. $\triangle$

Let $\gamma(t)$ be a computable non-increasing real function such that $0 < \gamma(t) < 1$ for all t and $\gamma(t) \to 0$ as $t \to \infty$; for example, $\gamma(t) = 1/t^\delta$, where $\delta > 0$. Define

$$\alpha_t = \frac{1}{2}\left(1 - \frac{\ln \frac{1+\ln N}{e^2-1}}{\ln \gamma(t)}\right) \text{ and} \tag{7}$$

$$\mu_t = (\gamma(t))^{\alpha_t} = \sqrt{\frac{e^2 - 1}{1 + \ln N}}(\gamma(t))^{1/2} \tag{8}$$

for all t, where $e = 2.72\ldots$ is the base of the natural logarithm.[4]

Without loss of generality we suppose that $\gamma(t) < \min\{1, (e^2 - 1)/(1 + \ln N)\}$ for all t. Then $0 < \alpha_t < 1$ for all t.

We consider an FPL algorithm with a variable learning rate

$$\epsilon_t = \frac{1}{\mu_t v_{t-1}}, \tag{9}$$

where μ_t is defined by (8) 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 $t = 1, 2, \ldots$. Also, by definition $\mu_t \to 0$ as $t \to \infty$.

[4] 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 (44).

Let $\xi_t^1,\dots\xi_t^N$, $t=1,2,\dots$, be a sequence of i.i.d random variables distributed according to 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 $\epsilon_0=\infty$.

The FPL algorithm is defined as follows:

FPL algorithm PROT

FOR $t=1,\dots T$

Choose an expert with the minimal perturbed cumulated loss on steps $<t$

$$I_t=\mathrm{argmin}_{i=1,2,\dots N}\{s_{1:t-1}^i-\frac{1}{\epsilon_t}\xi^i\}. \tag{10}$$

Receive one-step losses s_t^i for experts $i=1,\dots,N$, and receive one-step loss $s_t^{I_t}$ of the master algorithm.

ENDFOR

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

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 (9) is asymptotically consistent.

Theorem 1. *Let $\gamma(t)$ be a computable non-increasing positive real function such that $\gamma(t)\to 0$ as $t\to\infty$. Let also the game be non-degenerate and such that*

$$\mathrm{fluc}(t)\le\gamma(t) \tag{11}$$

for all t. Then the expected cumulated loss of the FPL *algorithm* PROT *with variable learning rate (9) for all t is bounded by*

$$E(s_{1:T})\le\min_i s_{1:T}^i+2\sqrt{(e^2-1)(1+\ln N)}\sum_{t=1}^{T}(\gamma(t))^{1/2}\Delta v_t. \tag{12}$$

Also, the algorithm PROT *is asymptotically consistent in the mean*

$$\limsup_{T\to\infty}\frac{1}{v_T}E(s_{1:T}-\min_{i=1,\dots N}s_{1:T}^i)\le 0. \tag{13}$$

The algorithm PROT *is Hannan consistent, i.e.,*

$$\limsup_{T\to\infty}\frac{1}{v_T}\left(s_{1:T}-\min_{i=1,\dots N}s_{1:T}^i\right)\le 0 \tag{14}$$

almost surely, if

$$\sum_{t=1}^{\infty}(\gamma(t))^2<\infty. \tag{15}$$

Proof. In the proof of this theorem we follow the proof-scheme of [4] and [5].

Let α_t be a sequence of real numbers defined by (7); 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) with the learning rate ϵ'_t defined by (16).

IFPL algorithm
FOR $t = 1, \dots T$

Define the learning rate

$$\epsilon'_t = \frac{1}{\mu_t v_t}, \text{ where } \mu_t = (\gamma(t))^{\alpha_t}, \tag{16}$$

v_t is the volume of the game at step t and α_t is defined by (7).

Choose an expert with the minimal perturbed cumulated loss on steps $\leq t$

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

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

The IFPL algorithm predicts under the knowledge of $s^i_{1:t}$, $i = 1, \dots N$ (and v_t), 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.

For any t, we have $I_t = \text{argmin}_i\{s^i_{1:t-1} - \frac{1}{\epsilon_t}\xi^i\}$ and $J_t = \text{argmin}_i\{s^i_{1:t} - \frac{1}{\epsilon'_t}\xi^i\} = \text{argmin}_i\{s^i_{1:t-1} + s^i_t - \frac{1}{\epsilon'_t}\xi^i\}$.

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

$$l_t = E(s_t^{I_t}) \text{ and } r_t = E(s_t^{J_t}),$$
$$l_{1:T} = \sum_{t=1}^{T} l_t \text{ and } r_{1:T} = \sum_{t=1}^{T} r_t,$$

respectively, where $s_t^{I_t}$ is the one-step loss of the FPL algorithm at step t and $s_t^{J_t}$ is the one-step loss of the IFPL algorithm, and E denotes the mathematical expectation.

Lemma 1. *The cumulated expected losses of the* FPL *and* IFPL *algorithms with rearning rates defined by (9) and (16) satisfy the inequality*

$$l_{1:T} \leq r_{1:T} + (e^2 - 1)\sum_{t=1}^{T}(\gamma(t))^{1-\alpha_t}\Delta v_t \tag{17}$$

for all T, where α_t is defined by (7).

Proof. Let $c_1, \dots c_N$ be nonnegative real numbers and

$$m_j = \min_{i \neq j}\{s^i_{1:t-1} - \frac{1}{\epsilon_t}c_i\},$$
$$m'_j = \min_{i \neq j}\{s^i_{1:t} - \frac{1}{\epsilon'_t}c_i\} = \min_{i \neq j}\{s^i_{1:t-1} + s^i_t - \frac{1}{\epsilon'_t}c_i\}.$$

Let $m_j = s^{j_1}_{1:t-1} - \frac{1}{\epsilon_t} c_{j^1}$ and $m'_j = s^{j_2}_{1:t} - \frac{1}{\epsilon'_t} c_{j_2} = s^{j_2}_{1:t-1} + s^{j_2}_t - \frac{1}{\epsilon'_t} c_{j_2}$. By definition and since $j_2 \neq j$ we have

$$m_j = s^{j_1}_{1:t-1} - \frac{1}{\epsilon_t} c_{j_1} \leq s^{j_2}_{1:t-1} - \frac{1}{\epsilon_t} c_{j^2} \leq s^{j_2}_{1:t-1} + s^{j_2}_t - \frac{1}{\epsilon_t} c_{j_2} = \tag{18}$$

$$s^{j_2}_{1:t} - \frac{1}{\epsilon'_t} c_{j_2} + \left(\frac{1}{\epsilon'_t} - \frac{1}{\epsilon_t} \right) c_{j_2} = m'_j + \left(\frac{1}{\epsilon'_t} - \frac{1}{\epsilon_t} \right) c_{j_2}. \tag{19}$$

We compare conditional probabilities $P\{I_t = j | \xi^i = c_i, i \neq j\}$ and $P\{J_t = j | \xi^i = c_i, i \neq j\}$.

The following chain of equalities and inequalities is valid:

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

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

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

$$P\{\xi^j \geq \epsilon'_t (s^j_{1:t-1} - m_j) + (\epsilon_t - \epsilon'_t)(s^j_{1:t-1} - m_j) | \xi^i = c_i, i \neq j\} \leq \tag{20}$$

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

$$(\epsilon_t - \epsilon'_t)(s^j_{1:t-1} - s^{j_2}_{1:t-1} + \frac{1}{\epsilon_t} c_{j_2}) | \xi^i = c_i, i \neq j\} = \tag{21}$$

$$\exp\{-(\epsilon_t - \epsilon'_t)(s^j_{1:t-1} - s^{j_2}_{1:t-1})\} \times \tag{22}$$

$$P\{\xi^j \geq \epsilon'_t (s^j_{1:t-1} - m_j) + (\epsilon_t - \epsilon'_t) \frac{1}{\epsilon_t} c_{j_2} | \xi^i = c_i, i \neq j\} \leq \tag{23}$$

$$\exp\{-(\epsilon_t - \epsilon'_t)(s^j_{1:t-1} - s^{j_2}_{1:t-1})\} \times$$

$$P\{\xi^j \geq \epsilon'_t (s^j_{1:t} - s^j_t - m'_j - \left(\frac{1}{\epsilon'_t} - \frac{1}{\epsilon_t} \right) c_{j_2}) + \tag{24}$$

$$(\epsilon_t - \epsilon'_t) \frac{1}{\epsilon_t} c_{j_2} | \xi^i = c_i, i \neq j\} = \tag{25}$$

$$\exp\{-(\epsilon_t - \epsilon'_t)(s^j_{1:t-1} - s^{j_2}_{1:t-1}) + \epsilon'_t s^j_t\} \times \tag{26}$$

$$P\{\xi^j \geq \epsilon'_t (s^j_{1:t} - m'_j) | \xi^i = c_i, i \neq j\} =$$

$$\exp\left\{ - \left(\frac{1}{\mu_t v_{t-1}} - \frac{1}{\mu_t v_t} \right) (s^j_{1:t-1} - s^{j_2}_{1:t-1}) + \frac{s^j_t}{\mu_t v_t} \right\} \times \tag{27}$$

$$P\{\xi^j > \frac{1}{\mu_t v_t} (s^j_{1:t} - m'_j) | \xi^i = c_i, i \neq j\} \leq$$

$$\exp\left\{ - \frac{\Delta v_t}{\mu_t v_t} \frac{(s^j_{1:t-1} - s^{j_2}_{1:t-1})}{v_{t-1}} + \frac{\Delta v_t}{\mu_t v_t} \right\} \times \tag{28}$$

$$P\{\xi^j > \frac{1}{\mu_t v_t} (s^j_{1:t} - m'_j) | \xi^i = c_i, i \neq j\} =$$

$$\exp\left\{ \frac{\Delta v_t}{\mu_t v_t} \left(1 - \frac{s^j_{1:t-1} - s^{j_2}_{1:t-1}}{v_{t-1}} \right) \right\} P\{J_t = 1 | \xi^i = c_i, i \neq j\}. \tag{29}$$

Here the inequality (20)-(21) follows from (18) and $\epsilon_t \geq \epsilon'_t$. We have used twice, in change from (21) to (22) and in change from (25) to (26), the equality $P\{\xi > a+b\} = e^{-b}P\{\xi > a\}$ for any random variable ξ distributed according to the exponential law. The equality (23)-(24) follows from (19). We have used in change from (27) to (28) the equality $v_t - v_{t-1} = \Delta v_t$ and the inequality $s_t^j \leq \Delta v_t$ for all j and t.

The expression in the exponent (29) is bounded

$$\left| \frac{s^j_{1:t-1} - s^{j_2}_{1:t-1}}{v_{t-1}} \right| \leq 1, \tag{30}$$

since $\frac{s^i_{1:t-1}}{v_{t-1}} \leq 1$ and $s^i_{1:t-1} \geq 0$ for all t and i.

Therefore, we obtain

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

$$\exp\left\{\frac{2}{\mu_t}\frac{\Delta v_t}{v_t}\right\} P\{J_t = j | \xi^i = c_i, i \neq j\} \leq \tag{31}$$

$$\exp\{2(\gamma(t))^{1-\alpha_t}\} P\{J_t = j | \xi^i = c_i, i \neq j\}. \tag{32}$$

Since, the inequality (32) holds for all c_i, it also holds unconditionally

$$P\{I_t = j\} \leq \exp\{2(\gamma(t))^{1-\alpha_t}\} P\{J_t = j\}. \tag{33}$$

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

Using inequality $\exp\{2x\} \leq 1 + (e^2 - 1)x$ for all x such that $0 \leq x \leq 1$, we obtain from (33) the lower bound

$$l_t = E(s_t^{I_t}) = \sum_{j=1}^{N} s_t^j P(I_t = j) \leq$$

$$\exp\{2(\gamma(t))^{1-\alpha_t}\} \sum_{j=1}^{N} s_t^j P(J_t = j) = \exp\{2(\gamma(t))^{1-\alpha_t}\} E(s_t^{J_t}) =$$

$$\exp\{2(\gamma(t))^{1-\alpha_t}\} r_t \leq (1 + (e^2 - 1)(\gamma(t))^{1-\alpha_t}) r_t. \tag{34}$$

Since $r_t \leq \Delta v_t$ for all t, the inequality (34) implies

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

for all T. Lemma 1 is proved. $\triangle$

The following lemma, which is an analogue of the result from [5], gives a bound for the IFPL algorithm.

Lemma 2. *The expected cumulative loss of the* IFPL *algorithm with the learning rate (16) is bounded by*

$$r_{1:T} \leq \min_i s^i_{1:T} + (1 + \ln N) \sum_{t=1}^{T} (\gamma(t))^{\alpha_t} \Delta v_t \tag{35}$$

for all T, where α_t is defined by (7).

Proof. The proof is along the line of the proof from Hutter and Poland [4] with an exception that now the sequence ϵ'_t is not monotonic.

Let in this proof, $s_t = (s_t^1, \dots s_t^N)$ be a vector of one-step losses and $s_{1:t} = (s_{1:t}^1, \dots s_{1:t}^N)$ 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 $\epsilon'_t = 1/(\mu_t v_t)$, $\mu_t \leq \mu_{t-1}$ for all t, and $v_0 = 0$, $\epsilon'_0 = \infty$.

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

For any vector s and a unit vector d denote

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

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

We first show that

$$\sum_{t=1}^{T} M(\tilde{s}_{1:t}) \cdot \tilde{s}_t \leq M(\tilde{s}_{1:T}) \cdot \tilde{s}_{1:T}. \tag{36}$$

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

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

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

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

We rewrite (36) as follows

$$\sum_{t=1}^{T} M(\tilde{s}_{1:t}) \cdot s_t \leq M(\tilde{s}_{1:T}) \cdot \tilde{s}_{1:T} + \sum_{t=1}^{T} M(\tilde{s}_{1:t}) \cdot \xi \left(\frac{1}{\epsilon'_t} - \frac{1}{\epsilon'_{t-1}}\right). \tag{37}$$

By definition of M we have

$$M(\tilde{s}_{1:T}) \cdot \tilde{s}_{1:T} \leq M(s_{1:T}) \cdot \left(s_{1:T} - \frac{\xi}{\epsilon'_T}\right) = \min_{d \in D}\{d \cdot s_{1:T}\} - M(s_{1:T}) \cdot \frac{\xi}{\epsilon'_T}. \tag{38}$$

The expectation of the last term in (38) is equal to $\frac{1}{\epsilon'_T} = \mu_T v_T$.

The second term of (37) can be rewritten

$$\sum_{t=1}^{T} M(\tilde{s}_{1:t}) \cdot \xi \left(\frac{1}{\epsilon'_t} - \frac{1}{\epsilon'_{t-1}}\right) = \sum_{t=1}^{T} (\mu_t v_t - \mu_{t-1} v_{t-1}) M(\tilde{s}_{1:t}) \cdot \xi. \tag{39}$$

We will use the inequality for mathematical expectation E

$$0 \leq E(M(\tilde{s}_{1:t}) \cdot \xi) \leq E(M(\xi) \cdot \xi) = E(\max_i \xi^i) \leq 1 + \ln N. \tag{40}$$

The proof of this inequality uses ideas of Lemma 1 from [4].

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

$$P\{\max_i \xi^i \geq a\} = P\{\exists i(\xi^i \geq a)\} \leq \sum_{i=1}^{N} P\{\xi^i \geq a\} = N \exp\{-a\}. \tag{41}$$

Since for any non-negative random variable η, $E(\eta) = \int\limits_0^\infty P\{\eta \geq y\}dy$, by (41) we have

$$E(\max_i \xi^i - \ln N) = \int\limits_0^\infty P\{\max_i \xi^i - \ln N \geq y\}dy \leq$$
$$\int\limits_0^\infty N \exp\{-y - \ln N\}dy = 1.$$

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

By (40) the expectation of (39) has the upper bound

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

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

Since $E(\xi^i) = 1$ for all i, the expectation of the last term in (38) is equal to

$$E\left(M(s_{1:T}) \cdot \frac{\xi}{\epsilon'_T}\right) = \frac{1}{\epsilon'_T} = \mu_T v_T. \tag{42}$$

Combining the bounds (37)-(39) and (42), we obtain

$$r_{1:T} = E\left(\sum_{t=1}^{T} M(\tilde{s}_{1:t}) \cdot s_t\right) \leq$$
$$\min_i s^i_{1:T} - \mu_T v_T + (1 + \ln N) \sum_{t=1}^{T} \mu_t \Delta v_t \leq$$
$$\min_i s^i_{1:T} + (1 + \ln N) \sum_{t=1}^{T} \mu_t \Delta v_t. \tag{43}$$

Lemma is proved. $\triangle$.

We finish now the proof of the theorem.

The inequality (17) of Lemma 1 and the inequality (35) of Lemma 2 imply the inequality

$$E(s_{1:T}) \leq \min_i s^i_{1:T} + \sum_{t=1}^{T}((e^2-1)(\gamma(t))^{1-\alpha_t} + (1+\ln N)(\gamma(t))^{\alpha_t})\Delta v_t. \quad (44)$$

for all T.

The optimal value (7) of α_t can be easily obtained by minimization of each member of the sum (44) by α_t. In this case μ_t is equal to (8) and (44) is equivalent to (12).

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 [10]

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

as $T \to \infty$. Therefore, the FPL algorithm PROT is asymptotically consistent in the mean, i.e., the relation (13) of Theorem 1 is proved.

We use some version of the strong law of large numbers to formulate a sufficient condition for Hannan consistency of the algorithm PROT.

Lemma 3. *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 1 hold and

$$\sum_{t=1}^{\infty} \frac{g(\Delta v_t)}{g(v_t)} < \infty. \quad (45)$$

Then the FPL *algorithm* PROT *is Hannan consistent, i.e., (5) holds as $T \to \infty$ almost surely.*

Proof. We use Theorem 11 from Petrov [8] (Chapter IX, Section 2) which gives sufficient conditions in order that the strong law of large numbers holds for a sequence of independent unbounded random variables:

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 Lemma 3. By Theorem 11 from Petrov [8] the inequality

$$\sum_{t=1}^{\infty} \frac{E(g(X_t))}{g(a_t)} < \infty \quad (46)$$

implies

$$\frac{1}{a_T}\sum_{t=1}^{T} X_t \to 0 \quad (47)$$

as $T \to \infty$ almost surely.

Put $X_t = s_t - E(s_t)$, 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 (46) is valid, and by (47)

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

as $T \to \infty$ almost surely.

This limit and the limit (13) imply (14). $\triangle$

By Lemma 3 the algorithm PROT is Hannan consistent, since (15) implies (45) for $g(x) = x^2$. Theorem 1 is proved. $\triangle$

Authors of [2] and [9] considered polynomially bounded one-step losses. We consider a specific example of the bound (44) for polynomial case.

Corollary 1. *Assume that $s_t^i \le t^a$ for all t and $i = 1, \dots N$, and*

$$\liminf_{t\to\infty} \frac{v_t}{t^{a+\delta}} > 0,$$

where a and δ are positive real numbers. Let also in the algorithm PROT, *$\gamma(t) = t^{-\delta}$ and $\mu_t = (\gamma(t))^{\alpha_t}$, where α_t is defined by (7). Then*

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

$$E(s_{1:T}) \le \min_i s_{1:T}^i + 2\sqrt{(e^2-1)(1+\ln N)}T^{1-\frac{1}{2}\delta+a} \tag{48}$$

as $T \to \infty$.

This corollary follows directly from Theorem 1, where condition (15) of Theorem 1 holds for $\delta > \frac{1}{2}$.

If $\delta = 1$ the regret from (48) is asymptotically equivalent to the regret from Allenberg et al. [2] (see Section 1).

For $a = 0$ we have the case of bounded loss function ($0 \le s_t^i \le 1$ 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 (15) holds, i.e.

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

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

Theorem 1 is also valid for the standard time scaling, i.e., when $v_T = T$ for all T, and when losses of experts are bounded, i.e., $a = 0$. Then the expected regret has the upper bound

$$2\sqrt{(e^2-1)(1+\ln N)}\sum_{t=1}^{T}(\gamma(t))^{1/2} \le 4\sqrt{(e^2-1)(1+\ln N)T}$$

which is similar to bounds from [4] and [5].

Acknowledgments

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

References

1. Cesa-Bianchi, N., Mansour, Y., Stoltz, G.: Improved second-order bounds for prediction with expert advice. Machine Learning 66(2-3), 321–352 (2007)
2. Allenberg, C., Auer, P., Gyorfi, L., Ottucsak, G.: Hannan consistency in on-Line learning in case of unbounded losses under partial monitoring. In: Balcázar, J.L., Long, P.M., Stephan, F. (eds.) ALT 2006. LNCS (LNAI), vol. 4264, pp. 229–243. Springer, Heidelberg (2006)
3. Hannan, J.: Approximation to Bayes risk in repeated plays. In: Dresher, M., Tucker, A.W., Wolfe, P. (eds.) Contributions to the Theory of Games, vol. 3, pp. 97–139. Princeton University Press, Princeton (1957)
4. Hutter, M., Poland, J.: Prediction with expert advice by following the perturbed leader for general weights. In: Ben-David, S., Case, J., Maruoka, A. (eds.) ALT 2004. LNCS (LNAI), vol. 3244, pp. 279–293. Springer, Heidelberg (2004)
5. Kalai, A., Vempala, S.: Efficient algorithms for online decisions. In: Schölkopf, B., Warmuth, M.K. (eds.) COLT/Kernel 2003. LNCS (LNAI), vol. 2777, pp. 26–40. Springer, Heidelberg (2003); Extended version in Journal of Computer and System Sciences 71, 291–307 (2005)
6. Littlestone, N., Warmuth, M.K.: The weighted majority algorithm. Information and Computation 108, 212–261 (1994)
7. Lugosi, G., Cesa-Bianchi, N.: Prediction, Learning and Games. Cambridge University Press, New York (2006)
8. Petrov, V.V.: Sums of independent random variables. Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 82. Springer, Heidelberg (1975)
9. Poland, J., Hutter, M.: Defensive universal learning with experts. For general weight. In: Jain, S., Simon, H.U., Tomita, E. (eds.) ALT 2005. LNCS (LNAI), vol. 3734, pp. 356–370. Springer, Heidelberg (2005)
10. Shiryaev, A.N.: Probability. Springer, Berlin (1980)
11. Vovk, V.G.: Aggregating strategies. In: Fulk, M., Case, J. (eds.) Proceedings of the 3rd Annual Workshop on Computational Learning Theory, San Mateo, CA, pp. 371–383. Morgan Kaufmann, San Francisco (1990)