Hannan consistency in on-line learning in case of unbounded losses under partial monitoring*

Chamy Allenberg[1], Peter Auer[2], László Györfi[3], and György Ottucsák[3]**

[1] School of Computer Science
Tel Aviv University
Tel Aviv, Israel, 69978
chamy_a@netvision.net.il

[2] Chair for Information Technology
University of Leoben
Leoben, Austria, A-8700
auer@unileoben.ac.at

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

Abstract. In this paper the sequential prediction problem with expert advice is considered when the loss is unbounded under partial monitoring scenarios. We deal with a wide class of the partial monitoring problems: the combination of the label efficient and multi-armed bandit problem, that is, where the algorithm is only informed about the performance of the *chosen* expert with probability $\varepsilon \leq 1$. For bounded losses an algorithm is given whose expected regret scales with the square root of the loss of the best expert. For unbounded losses we prove that Hannan consistency can be achieved, depending on the growth rate of the average squared losses of the experts.

1 Introduction

In on-line (often referred also as sequential) prediction problems in general, an algorithm has to perform a sequence of actions. After each action, the algorithm suffers some loss, depending on the response of the environment. Its goal is to minimize its cumulative loss over a sufficiently long period of time. In the adversarial setting no probabilistic assumption is made on how the losses corresponding to different actions are generated. In particular, the losses may depend on the previous actions of the algorithm, whose goal is to perform well relative to a set of experts for any possible behavior of the environment. More precisely, the aim of the algorithm is to achieve asymptotically the same average loss (per round) as the best expert.

* The authors would like to thank Gilles Stoltz and András György for useful comments.

** György Ottucsák is eligible for the "E.M. Gold Award".

In most of the machine learning literature, one assumes that the losses are bounded, and such a bound is known in advance, when designing an algorithm. In many applications, including regression problems (Györfi and Lugosi [9]) or routing in communication networks (cf. György and Ottucsák [11]) the loss is unbounded. The main aim of this paper is to show Hannan consistency of on-line algorithms for unbounded losses under partial monitoring.

The first theoretical results concerning sequential prediction (decision) are due to Blackwell [2] and Hannan [12], but they were rediscovered by the learning community only in the 1990's, see, for example, Vovk [15], Littlestone and Warmuth [14] and Cesa-Bianchi *et al.* [3]. These results show that it is possible to construct algorithms for on-line (sequential) decision that predict almost as well as the best expert. The main idea of these algorithms is the same: after observing the past performance of the experts, in each step the decision of a randomly chosen expert is followed such that experts with superior past performance are chosen with higher probability.

However, in certain type of problems it is not possible to obtain all the losses corresponding to the decisions of the experts. Throughout the paper we use this framework in which the algorithm has a limited access to the losses. For example, in the so called multi-armed bandit problem the algorithm has only information on the loss of the chosen expert, and no information is available about the loss it would have suffered had it made a different decision (see, e.g., Auer *et al.* [1], Hart and Mas Colell [13]). Another example is label efficient prediction, where it is expensive to obtain the losses of the experts, and therefore the algorithm has the option to query this information (see Cesa-Bianchi *et. al* [5]). Finally the combination of the label efficient and the multi-armed bandit problem, where after choosing a decision, the algorithm learns its own loss if and only if it asks for it (see György and Ottucsák [11]).

Cesa-Bianchi *et. al.* [7] studied second-order bounds for exponentially weighted average forecaster and they analyzed the expected regret of the algorithm in the full monitoring case when the bound of the loss function unknown. They indicated their results in partial monitoring case.

2 Sequential prediction and partial monitoring models

The on-line decision problem considered in this paper is described as follows. Suppose an algorithm has to make a sequence of actions. At each time instant $t = 1, 2, \ldots$, an action $a_t \in \mathcal{A}$ is made, where $\mathcal{A}$ denotes the action space. Then, based on the state of the environment $y_t \in \mathcal{Y}$, where $\mathcal{Y}$ is some state space, the algorithm suffers some loss $\ell(a_t, y_t)$ with loss function $\ell : \mathcal{A} \times \mathcal{Y} \to \mathbb{R}^+$. The performance of the algorithm is evaluated relative to a set of experts, and its goal is to perform asymptotically as well as the best expert. Formally, given N experts, at each time instant t, for every $i = 1, \ldots, N$, expert i chooses an action $f_{i,t} \in \mathcal{A}$, and suffers loss $\ell(f_{i,t}, y_t)$. We assume that the action space is finite, therefore we consider algorithms that follow the advice of one of the experts, that is, $f_{I_t,t}$ for some I_t, where $I_t \in \{1, \ldots, N\}$ is a random variable.

The distribution of I_t is generated by the algorithm. I_t only depends on the past losses $\ell(f_{i,t-1}, y_t), \ldots, \ell(f_{i,1}, y_1)$ for all i and the earlier choices of the algorithm $I_{t-1}, \ldots, I_1$. For convenience we use the notations $\ell_{i,t}$ instead of $\ell(f_{i,t}, y_t)$ and $\ell_{I_t,t}$ instead of $\ell(f_{I_t,t}, y_t)$.

Formally, at each time instance $t = 1, 2, \ldots$,

1. the environment decides on the losses $\ell_{i,t} \geq 0$ of the experts $i \in \{1, \ldots, N\}$,
2. the algorithm chooses an expert $I_t \in \{1, \ldots, N\}$,
3. the algorithm suffers loss $\ell_{I_t,t}$,
4. the algorithm receives some feedback about his loss and the losses of the experts.

After n rounds the loss of the algorithm and the losses of the experts are denoted by

$$\widehat{L}_n = \sum_{t=1}^{n} \ell_{I_t,t} \quad \text{and} \quad L_{i,n} = \sum_{t=1}^{n} \ell_{i,t},$$

and the performance of the algorithm is measured by its regret, $\widehat{L}_n - \min_i L_{i,n}$, or by its regret per round, $\frac{1}{n}\left(\widehat{L}_n - \min_i L_{i,n}\right)$. An algorithm is Hannan consistent [12], if

$$\limsup_{n\to\infty} \frac{1}{n}\left(\widehat{L}_n - \min_i L_{i,n}\right) \leq 0 \qquad a.s.$$

The performance of any expert algorithm obviously depends on how much information is available to the algorithm about the experts' and its own performance. Next we show the most important classes of partial monitoring according to the amount of the information available to the algorithm.

- **Full information** (FI) case: the algorithm has access to the losses $\ell_{i,t}$ of all experts.
- **Multi-armed bandit** (MAB) problem: only the loss of the chosen expert is revealed to the algorithm, i.e., only $\ell_{I_t,t}$ is known.
- **Label efficient** (LE) setting: the algorithm tosses a coin S_t whether to query for the losses.[4] If $S_t = 1$ (with probability ε_t) then the algorithm knows all $\ell_{i,t}$, $i = 1, \ldots, n$, otherwise it does not.
- **Combination of the label efficient and multi-armed bandit** (LE+MAB) setting: the algorithm queries with probability ε_t only about the loss of the chosen expert, $\ell_{I_t,t}$.

Throughout the paper we focus on problem LE+MAB because all of the other problems mentioned above are "easier", in the sense that if an algorithm is Hannan consistent for problem LE+MAB, then it is Hannan consistent for the other cases, too.

[4] It is easy to see that in order to achieve a nontrivial performance, the algorithm must use randomization in determining whether the losses should be revealed or not (cf. Cesa-Bianchi and Lugosi [4]).

3 The algorithm

In problem LE+MAB, the algorithm learns its own loss only if it chooses to query it, and it cannot obtain information on the loss of any other expert. For querying its loss the algorithm uses a sequence $S_1, S_2, \ldots$ of independent Bernoulli random variables such that

$$\mathbb{P}(S_t = 1) = \varepsilon_t,$$

and asks for the loss $\ell_{I_t,t}$ of the chosen expert I_t if $S_t = 1$, which for constant $\varepsilon_t = \varepsilon$ is identical to the label efficient algorithms in Cesa-Bianchi *et al.* [5]. We denote by LE(ε_t) the label efficient problem with time-varying parameter ε_t.

We will derive sufficient conditions for Hannan consistency for the combination of the time-varying label efficient and multi-armed bandit problem (LE(ε_t)+MAB) and then we will show that this condition can be adapted straightforwardly to the other cases.

For problem LE(ε_t)+MAB we use algorithm Green with time-varying learning rate η_t. Algorithm Green is a variant of the weighted majority (WM) algorithm of Littlestone and Warmuth [14]. Denote by $p_{i,t}$ the probability of choosing action i at time t in case of the original WM algorithm, that is,

$$p_{i,t} = \frac{e^{-\eta_t \widetilde{L}_{i,t-1}}}{\sum_{j=1}^{N} e^{-\eta_t \widetilde{L}_{j,t-1}}},$$

where $\widetilde{L}_{i,t}$ is so called cumulative estimated loss, which we will specify later. Algorithm Green uses modified probabilities $\widetilde{p}_{i,t}$ which can be calculated from $p_{i,t}$,

$$\widetilde{p}_{i,t} = \begin{cases} 0 & \text{if } p_{i,t} < \gamma_t, \\ c_t \cdot p_{i,t} & \text{if } p_{i,t} \geq \gamma_t, \end{cases}$$

where c_t is the normalizing factor and $\gamma_t \geq 0$ is a time-varying threshold. Finally, the algorithm uses estimated losses which are given by

$$\widetilde{\ell}_{i,t} = \begin{cases} \frac{\ell_{i,t}}{\widetilde{p}_{i,t}\varepsilon_t} & \text{if } I_t = i \text{ and } S_t = 1; \\ 0 & \text{otherwise,} \end{cases}$$

based on György and Ottucsák [11]. Therefore, the estimated loss is an unbiased estimate of the true loss with respect to its natural filtration, that is,

$$\mathbb{E}_t\left[\widetilde{\ell}_{i,t}\right] \stackrel{\text{def}}{=} \mathbb{E}\left[\widetilde{\ell}_{i,t} \middle| S_1^{t-1}, I_1^{t-1}\right] = \ell_{i,t}.$$

The cumulative estimated loss of an expert is given by $\widetilde{L}_{i,n} = \sum_{t=1}^{n} \widetilde{\ell}_{i,t}$. The resulting algorithm is given in Figure 1.

Algorithm Green

Let $\eta_1, \eta_2, \ldots > 0$, $\varepsilon_1, \varepsilon_2, \ldots > 0$ and $\gamma_1, \gamma_2, \ldots \geq 0$.

Initialization: $\widetilde{L}_{i,0} = 0$ for all $i = 1, \ldots, N$.

For each round $t = 1, 2, \ldots$

(1) Calculate the probability distribution

$$p_{i,t} = \frac{e^{-\eta_t \widetilde{L}_{i,t-1}}}{\sum_{i=1}^{N} e^{-\eta_t \widetilde{L}_{i,t-1}}} \qquad i = 1, \ldots, N\ .$$

(2) Calculate the modified probabilities

$$\widetilde{p}_{i,t} = \begin{cases} 0 & \text{if } p_{i,t} < \gamma_t, \\ c_t \cdot p_{i,t} & \text{if } p_{i,t} \geq \gamma_t, \end{cases}$$

where $c_t = 1/\sum_{p_{i,t} \geq \gamma_t} p_{i,t}$.

(3) Select an action $I_t \in \{1, \ldots, N\}$ according to $\widetilde{\mathbf{p}}_\mathbf{t} = (\widetilde{p}_{1,t}, \ldots, \widetilde{p}_{N,t})$.

(4) Draw a Bernoulli random variable S_t such that $\mathbb{P}(S_t = 1) = \varepsilon_t$.

(5) Compute the estimated loss for all $i = 1, \ldots, N$

$$\widetilde{\ell}_{i,t} = \begin{cases} \frac{\ell_{i,t}}{\widetilde{p}_{i,t}\varepsilon_t} & \text{if } I_t = i \text{ and } S_t = 1; \\ 0 & \textit{otherwise}. \end{cases}$$

(6) For all $i = 1, \ldots, N$ update the cumulative estimated loss

$$\widetilde{L}_{i,t} = \widetilde{L}_{i,t-1} + \widetilde{\ell}_{i,t}.$$

Fig. 1. Algorithm Green for LE(ε_t)+MAB

4 Bounds on the expected regret

Theorem 1. *If $\ell_{i,t}^2 \leq t^\nu$ and $\varepsilon_t \geq t^{-\beta}$ for all t, then for all n the expected loss of algorithm* Green *with $\gamma_t = 0$ and $\eta_t = 2\sqrt{\frac{\ln N}{N}} \cdot t^{-(1+\nu+\beta)/2}$ is bounded by*

$$\mathbb{E}\left[\widehat{L}_n - \min_i L_{i,n}\right] \leq 2\sqrt{(N \ln N)}(n+1)^{(1+\nu+\beta)/2}.$$

If the individual losses are bounded by a constant, a much stronger result can be obtained.

Theorem 2. *If $\ell_{i,t} \in [0,1]$ and $\varepsilon_t = \varepsilon$ for all t, then for all n with $\min_i L_{i,n} \leq B$ the expected loss of algorithm* Green *with $\gamma_t = \gamma = \frac{1}{N(B\varepsilon+2)}$ and $\eta_t = \eta = 2\sqrt{\frac{\ln N}{N}\frac{\varepsilon}{B}}$ is bounded by*

$$\mathbb{E}\left[\widehat{L}_n - \min_i L_{i,n}\right] \leq 4\sqrt{\frac{B}{\varepsilon} N \ln N} + \frac{N \ln N + 2}{\varepsilon} + \frac{\ln(\varepsilon B + 2)}{\varepsilon}.$$

Remark 1. The improvement in Theorem 2 is significant, since it bounds the regret of the algorithm in terms of the loss of the best action and not in respect to the number of rounds. For example, Theorem 1 is void for $\min_i L_{i,n} \ll \sqrt{n}$ whereas Theorem 2 still gives a nearly optimal bound[5].

Remark 2. If the magnitude of the losses is not known a-priori, the doubling trick can be used to set the parameter ν in Theorem 1 and the parameter B in Theorem 2 with no significant change in the bounds. The generalization of Theorem 2 to losses in $[a, b]$ is straightforward.

For the proofs we introduce the notations

$$\check{\ell}_t = \sum_{i=1}^{N} \widetilde{p}_{i,t}\widetilde{\ell}_{i,t}, \quad \overline{\ell}_t = \sum_{i=1}^{N} p_{i,t}\widetilde{\ell}_{i,t}, \qquad \text{and} \qquad \overline{L}_n = \sum_{t=1}^{n} \overline{\ell}_t.$$

Then

$$\widehat{L}_n - \min_i L_{i,n} = \left(\widehat{L}_n - \overline{L}_n\right) + \left(\overline{L}_n - \min_i \widetilde{L}_{i,n}\right) + \left(\min_i \widetilde{L}_{i,n} - \min_i L_{i,n}\right). \quad (1)$$

Lemma 1. *For any sequence of losses $\ell_{i,t} \geq 0$,*

$$\widehat{L}_n - \overline{L}_n \leq \sum_{t=1}^{n} \left(\ell_{I_t,t} - \check{\ell}_t\right) + \sum_{t=1}^{n} N\gamma_t\check{\ell}_t.$$

Proof. Since $p_{I_t,t}/\widetilde{p}_{I_t,t} = 1/c_t = \sum_{j:p_{j,t}\geq\gamma_t} p_{j,t} = 1 - \sum_{j:p_{j,t}<\gamma_t} p_{j,t} \geq 1 - N\gamma_t$ we have

$$\overline{\ell}_t = \sum_{i=1}^{N} p_{i,t}\widetilde{\ell}_{i,t} = p_{I_t,t}\widetilde{\ell}_{I_t,t} \geq (1 - N\gamma_t)\widetilde{p}_{I_t,t}\widetilde{\ell}_{I_t,t} = (1 - N\gamma_t)\check{\ell}_t.$$

Thus

$$\widehat{L}_n - \overline{L}_n = \sum_{t=1}^{n} \ell_{I_t,t} - \sum_{t=1}^{n} \overline{\ell}_t \leq \sum_{t=1}^{n} \left(\ell_{I_t,t} - \check{\ell}_t\right) + \sum_{t=1}^{n} N\gamma_t\check{\ell}_t.$$

□

For bounding $\overline{L}_n - \min_i \widetilde{L}_{i,n}$ we use of the following lemma.

Lemma 2 (Cesa-Bianchi *et al.* [6]). *Consider any nonincreasing sequence of $\eta_1, \eta_2, \ldots$ positive learning rates and any sequences $\widetilde{\boldsymbol{\ell}}_1, \widetilde{\boldsymbol{\ell}}_2, \ldots \in \mathbb{R}_+^N$ of loss vectors. Define the function Φ by*

$$\Phi(\mathbf{p}_t, \eta_t, -\widetilde{\boldsymbol{\ell}}_t) = \sum_{i=1}^{N} p_{i,t}\widetilde{\ell}_{i,t} + \frac{1}{\eta_t} \ln \sum_{i=1}^{N} p_{i,t} e^{-\eta_t\widetilde{\ell}_{i,t}},$$

where $\mathbf{p}_t = (p_{1,t}, p_{2,t}, \ldots, p_{N,t})$ the probability vector of the WM algorithm. Then, for Algorithm Green

$$\overline{L}_n - \min_i \widetilde{L}_{i,n} \leq \left(\frac{2}{\eta_{n+1}} - \frac{1}{\eta_1}\right) \ln N + \sum_{t=1}^{n} \Phi(\mathbf{p}_t, \eta_t, -\widetilde{\boldsymbol{\ell}}_t).$$

[5] For $\varepsilon = 1$ optimality follows from the lower bound on the regret in [1].

Lemma 3. *With the notation of Lemma 2 we get for algorithm* Green,

$$\Phi(\mathbf{p}_t, \eta_t, -\widetilde{\boldsymbol{\ell}}_t) \le \frac{\eta_t}{2\varepsilon_t} \sum_{i=1}^N \ell_{i,t} \widetilde{\ell}_{i,t}.$$

Proof.

$$\begin{aligned}
\Phi(\mathbf{p}_t, \eta_t, -\widetilde{\boldsymbol{\ell}}_t) &= \sum_{i=1}^N p_{i,t} \widetilde{\ell}_{i,t} + \frac{1}{\eta_t} \ln \sum_{i=1}^N p_{i,t} e^{-\eta_t \widetilde{\ell}_{i,t}} \\
&\le \sum_{i=1}^N p_{i,t} \widetilde{\ell}_{i,t} + \frac{1}{\eta_t} \ln \sum_{i=1}^N p_{i,t} \left(1 - \eta_t \widetilde{\ell}_{i,t} + \frac{\eta_t^2 \widetilde{\ell}_{i,t}^2}{2}\right) && (2) \\
&\le \sum_{i=1}^N p_{i,t} \widetilde{\ell}_{i,t} + \frac{1}{\eta_t} \ln \left(1 - \eta_t \sum_{i=1}^N p_{i,t} \widetilde{\ell}_{i,t} + \frac{\eta_t^2}{2} \sum_{i=1}^N p_{i,t} \widetilde{\ell}_{i,t}^2\right) \\
&\le \frac{\eta_t}{2} \sum_{i=1}^N p_{i,t} \widetilde{\ell}_{i,t}^2 \ \le \ \frac{\eta_t}{2\varepsilon_t} \sum_{i=1}^N \ell_{i,t} \widetilde{\ell}_{i,t} && (3)
\end{aligned}$$

where (2) holds because of $e^{-x} \le 1 - x + x^2/2$ for $x \ge 0$, and (3) follows from the fact that $\ln(1+x) \le x$ for all $x > -1$, and from the definition of $\widetilde{\ell}_{i,t}$ in algorithm Green. □

Lemma 4. *For any sequence of $\ell_{i,t}$ the loss of algorithm* Green *is bounded by*

$$\begin{aligned}
\mathbb{E}\left[\widehat{L}_n - \min_i L_{i,n}\right] &\le N \sum_{t=1}^n \gamma_t \mathbb{E}[\ell_{I_t,t}] + \frac{2 \ln N}{\eta_{n+1}} + \sum_{i=1}^N \sum_{t=1}^n \frac{\eta_t \mathbb{E}\left[\ell_{i,t} \widetilde{\ell}_{i,t}\right]}{2\varepsilon_t} && (4) \\
&= N \sum_{t=1}^n \gamma_t \mathbb{E}[\ell_{I_t,t}] + \frac{2 \ln N}{\eta_{n+1}} + \sum_{i=1}^N \sum_{t=1}^n \frac{\eta_t \mathbb{E}\left[\ell_{i,t}^2\right]}{2\varepsilon_t}.
\end{aligned}$$

Proof. From (1) and Lemmas 1–3, we get

$$\begin{aligned}
\widehat{L}_n - \min_i L_{i,n} \le& \sum_{t=1}^n \left(\ell_{I_t,t} - \check{\ell}_t\right) + \sum_{t=1}^n N \gamma_t \check{\ell}_t + \left(\frac{2}{\eta_{n+1}} - \frac{1}{\eta_1}\right) \ln N \\
&+ \sum_{t=1}^n \frac{\eta_t}{2\varepsilon_t} \sum_{i=1}^N \ell_{i,t} \widetilde{\ell}_{i,t} + \left(\min_i \widetilde{L}_{i,n} - \min_i L_{i,n}\right).
\end{aligned}$$

Since $\mathbb{E}_t[\ell_{I_t,t}] = \sum_{i=1}^N \widetilde{p}_{i,t} \ell_{i,t} = \sum_{i=1}^N \widetilde{p}_{i,t} \mathbb{E}_t\left[\widetilde{\ell}_{i,t}\right] = \mathbb{E}_t\left[\check{\ell}_t\right]$ and $\mathbb{E}\left[\min_i \widetilde{L}_{i,n}\right] \le \mathbb{E}\left[\widetilde{L}_{i^*,n}\right] \le \mathbb{E}[L_{i^*,n}]$ for $i^* = \arg\min_i L_{i,n}$, taking expectations gives (4). The second line of the lemma follows from $\mathbb{E}_t\left[\widetilde{\ell}_{i,t}\right] = \ell_{i,t}$. □

Proof of Theorem 1. By simple calculation from Lemma 4. $\square$

Proof of Theorem 2. Let $T_i = \max\{0 \le t \le n : p_{i,t} \ge \gamma\}$ be the last round which contributes to $\widetilde{L}_{i,n}$. Therefore,

$$\gamma \le p_{i,T_i} = \frac{e^{-\eta \widetilde{L}_{i,T_i}}}{\sum_{j=1}^N e^{-\eta \widetilde{L}_{j,T_i}}} < \frac{e^{-\eta \widetilde{L}_{i,T_i}}}{e^{-\eta \widetilde{L}_{i^*,n}}},$$

where $i^* = \arg\min_i L_{i,n}$. After rearranging we obtain

$$\widetilde{L}_{i,T_i} \le \widetilde{L}_{i^*,n} + \frac{\ln(1/\gamma)}{\eta}$$

and since $\widetilde{L}_{i,n} = \widetilde{L}_{i,T_i}$ we get that $\widetilde{L}_{i,n} \le \widetilde{L}_{i^*,n} + \frac{\ln(1/\gamma)}{\eta}$. Plugging this bound into (4) and using $\ell_{i,t} \in [0,1]$ we get

$$\mathbb{E}\Big[\widehat{L}_n - L_{i^*,n}\Big] \le \gamma N \mathbb{E}\Big[\widehat{L}_n\Big] + \frac{2\ln N}{\eta} + N\frac{\eta}{2\varepsilon}\left(\mathbb{E}[L_{i^*,n}] + \frac{\ln(1/\gamma)}{\eta}\right).$$

Solving for $\mathbb{E}\Big[\widehat{L}_n\Big]$ we find

$$\mathbb{E}\Big[\widehat{L}_n\Big] \le \frac{1}{1-\gamma N}\left[\mathbb{E}[L_{i^*,n}] + \frac{2\ln N}{\eta} + N\frac{\eta}{2\varepsilon}\left(\mathbb{E}[L_{i^*,n}] + \frac{\ln(1/\gamma)}{\eta}\right)\right].$$

For $\gamma = \frac{1}{N(\varepsilon B+2)}$ we have $\frac{L_{i^*,n}}{1-\gamma N} \le L_{i^*,n} + \frac{2}{\varepsilon}$ and $\frac{1}{1-\gamma N} \le 2$, which implies

$$\mathbb{E}\Big[\widehat{L}_n\Big] \le L_{i^*,n} + \frac{2}{\varepsilon} + \frac{4\ln N}{\eta} + N\frac{\eta}{\varepsilon}\left(L_{i^*,n} + \frac{\ln N}{\eta} + \frac{\ln(\varepsilon B+2)}{\eta}\right).$$

and, by simple calculation, the statement of the theorem. $\square$

5 Hannan consistency

In this section we derive the sufficient conditions of Hannan consistency under partial monitoring for algorithm Green using time-varying parameters in case when the bound of the loss is unknown in advance, or when the loss is unbounded.

The next result shows sufficient conditions of Hannan consistency of Algorithm Green.

Theorem 3. *Algorithm* Green *is run for the combination of the label efficient and multi armed bandit problem. Assume that for each n*

$$\max_{1\le i\le N} \frac{1}{n}\sum_{t=1}^n \ell_{i,t}^2 < cn^\nu,$$

where $c < \infty$ and $0 \le \nu < 1$. For some $\rho > 0$ choose the parameters of the algorithm as:

$$\gamma_t = t^{-\alpha}/N; \quad (\nu + \rho)/2 \leq \alpha \leq 1,$$

$$\eta_t = t^{-1+\delta}; \qquad 0 < \delta \leq 1 - \nu - \alpha - \beta - \rho$$

and

$$\varepsilon_t = \varepsilon_0 t^{-\beta}; \qquad 0 < \varepsilon_0 \leq 1 \quad \textit{and} \quad 0 \leq \beta \leq 1 - \nu - \alpha - \delta - \rho.$$

Then Algorithm Green *is Hannan consistent, that is,*

$$\limsup_{n\to\infty} \frac{1}{n}\left(\widehat{L}_n - \min_i L_{i,n}\right) \leq 0 \qquad a.s.$$

Remark 3. (Unknown ν) If ν is unknown in advance, then define a set of infinite number of experts. The experts use Algorithm 1 with different parameter ν. Since $0 \leq \nu < 1$, instead of ν we can use ν_k, a quantization of the $[0, 1)$ interval. Let $\{\nu_k\}$ is a monotonically increasing sequence which goes to 1 and let q_k be an arbitrary distribution over the set of k such that $q_k > 0$ for all k. Then using exponential weighting with time-varying learning rate in case of unbounded losses, the difference between the average loss of the (combined) algorithm and the average loss of the best expert vanishes asymptotically [10][Lemma 1]. Therefore the algorithm reaches Hannan consistency.

Remark 4. We derive the consequences of the theorem in special cases:

– **FI**: With a slight modification of the proof we get the following condition for the losses in full information case:

$$\max_{1\leq i\leq N} \frac{1}{n}\sum_{t=1}^{n} \ell_{i,t}^2 \leq O\left(n^{1-\delta-\rho}\right).$$

– **MAB**: we fix $\beta = 0$ ($\varepsilon_t = 1$). Choose $\gamma_t = t^{-1/3}$ for all t. Then the condition is for the losses

$$\max_{1\leq i\leq N} \frac{1}{n}\sum_{t=1}^{n} \ell_{i,t}^2 \leq O\left(n^{2/3-\delta-\rho}\right).$$

– **LE(ε_t)**: With a slight modification of the proof we get the following condition for the loss function in label efficient case:

$$\max_{1\leq i\leq N} \frac{1}{n}\sum_{t=1}^{n} \ell_{i,t}^2 \leq O\left(n^{1-\beta-\delta-\rho}\right).$$

– **LE(ε_t)+MAB**: This is the most general case. Let $\gamma_t = t^{-1/3}$. Then the bound is

$$\max_{1\leq i\leq N} \frac{1}{n}\sum_{t=1}^{n} \ell_{i,t}^2 \leq O\left(n^{2/3-\beta-\delta-\rho}\right).$$

Remark 5. (Convergence rate) With a slight extension of Lemma 5 we can retrieve the ν dependent almost sure convergence rate of the algorithm. The rate is

$$\frac{1}{n}\left(\widehat{L}_n - \min_i L_{i,n}\right) \leq O(n^{\nu/2-1/2}) \qquad a.s.$$

in the FI and the LE cases with optimal choice of the parameters and in the MAB and the LE+MAB cases it is

$$\frac{1}{n}\left(\widehat{L}_n - \min_i L_{i,n}\right) \leq O(n^{\nu/2-1/3}) \qquad a.s.$$

Remark 6. (Minimum amount of query rate in LE(ε_t)) Denote

$$\mu(n) = \sum_{t=1}^{n} \varepsilon_t$$

the expected query rate, that is, the expected number of queries that can be issued up to time n. Assume that the average of the loss function has a constant bound, i.e., $\nu = 0$. With a slight modification of the proof of Theorem 3 and choosing

$$\eta_t = \frac{\log\log\log t}{t} \quad \text{and} \quad \varepsilon_t = \frac{\log\log t}{t}$$

we obtain the condition for Hannan consistency, such that

$$\mu(n) = \log n \log\log n,$$

which is the same as that of to Cesa-Bianchi *et al.* [5].

6 Proof

In order to prove Theorem 3, we split the proof into three lemmas by telescope as before:

$$\begin{aligned}
&\frac{1}{n}\widehat{L}_n - \frac{1}{n}\min_i L_{i,n} \\
&= \underbrace{\frac{1}{n}\left(\widehat{L}_n - \overline{L}_n\right)}_{\text{Lemma 6}} + \underbrace{\frac{1}{n}\left(\overline{L}_n - \min_i \widetilde{L}_{i,n}\right)}_{\text{Lemma 7}} + \underbrace{\frac{1}{n}\left(\min_i \widetilde{L}_{i,n} - \min_i L_{i,n}\right)}_{\text{Lemma 8}}. \qquad (5)
\end{aligned}$$

Combine sequentially Lemma 6, Lemma 7 and Lemma 8 to prove Theorem 3. We will show separately the almost sure convergence of the three terms on the right-hand side. In the sequel, we need the following lemma which is the key of the proof of Theorem 3:

Lemma 5. *Let $\{Z_t\}$ a martingale difference sequence. Let*

$$h_t k_t \geq \mathbf{Var}(Z_t)$$

where

$$h_t = 1/t^a$$

for all $t = 1, 2, \ldots$ *and*

$$K_n = \frac{1}{n}\sum_{t=1}^{n} k_t \le C n^b$$

and $0 \le b < 1$ *and* $b - a < 1$. *Then*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} Z_t = 0 \qquad a.s.$$

Proof. By the strong law of large numbers for martingale differences due to Chow [8], if $\{Z_t\}$ a martingale difference sequence with

$$\sum_{t=1}^{\infty} \frac{\mathbf{Var}(Z_t)}{t^2} < \infty \tag{6}$$

then

$$\lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} Z_t = 0 \qquad a.s.$$

We have to verify (6). Because of $k_t = tK_t - (t-1)K_{t-1}$, and $\frac{h_t}{t} - \frac{h_{t+1}t}{(t+1)^2} \ge 0$ we have that

$$\begin{aligned}
\sum_{t=1}^{n} \frac{\mathbf{Var}(Z_t)}{t^2} \le \sum_{t=1}^{n} \frac{h_t k_t}{t^2} &= \sum_{t=1}^{n} \frac{h_t\left(tK_t - (t-1)K_{t-1}\right)}{t^2} \\
&= \frac{h_n K_n}{n} + \sum_{t=1}^{n-1}\left(\frac{h_t}{t} - \frac{h_{t+1}t}{(t+1)^2}\right)K_t. \\
&\le \frac{n^{-a}Cn^b}{n} + \sum_{t=1}^{n-1}\left(\frac{t^{-a}}{t} - \frac{(t+1)^{-a}t}{(t+1)^2}\right)Ct^b
\end{aligned}$$

which is bounded by conditions. □

Now we are ready to prove one by one the almost sure convergence of the terms in (5).

Lemma 6. *Under the conditions of the Theorem 3,*

$$\lim_{n\to\infty} \frac{1}{n}\left(\widehat{L}_n - \overline{L}_n\right) = 0 \qquad a.s.$$

Proof. First we use Lemma 1, that is

$$\widehat{L}_n - \overline{L}_n \le \sum_{t=1}^{n}\left(\ell_{I_t,t} - \check{\ell}_t\right) + \sum_{t=1}^{n} N\gamma_t\check{\ell}_t = \sum_{t=1}^{n} Z_t + \sum_{t=1}^{n} N\gamma_t\check{\ell}_t. \tag{7}$$

Below we show separately, that both sums in (7) divided by n converge to zero almost surely. First observe that $\{Z_t\}$ is a martingale difference sequence with respect to I^{t-1} and S^{t-1}. Observe that I_t is independent from S_t therefore we get the following bound for the variance of Z_t:

$$\mathbf{Var}(Z_t) = \mathbb{E}\big[Z_t^2\big] = \mathbb{E}\big[(\ell_{I_t,t} - \check{\ell}_t)^2\big] \le \frac{1}{\varepsilon_t}\sum_{i=1}^{N} \ell_{i,t}^2 \stackrel{\text{def}}{=} h_t k_t,$$

where $h_t = 1/\varepsilon_t$ and $k_t = \sum_{i=1}^{N} \ell_{i,t}^2$. Then applying Lemma 5 we obtain

$$\lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} Z_t = 0 \quad a.s.$$

Next we show that the second sum in (7) divided by n goes to zero almost surely, that is,

$$\frac{1}{n}\sum_{t=1}^{n} N\gamma_t\check{\ell}_t = \frac{1}{n}\sum_{t=1}^{n}\frac{S_t}{\varepsilon_t}\ell_{I_t,t}N\gamma_t = \frac{1}{n}\sum_{t=1}^{n} R_t + \frac{1}{n}\sum_{t=1}^{n}\ell_{I_t,t}N\gamma_t \to 0 \quad (n\to\infty) \tag{8}$$

where R_t is a martingale difference sequence respect to S_1^{t-1} and I_1^t. Bounding the variance of R_t, we obtain

$$\mathbf{Var}(R_t) \le N^2\frac{\gamma_t^2}{\varepsilon_t}\sum_{i=1}^{N}\ell_{i,t}^2.$$

Then using Lemma 5 with parameters $h_t = \gamma_t^2/\varepsilon_t$ and $k_t = \sum_{i=1}^{N}\ell_{i,t}^2$ we get

$$\lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n} R_t = 0 \quad a.s.$$

The proof is finished by showing, that the second sum in (8) goes to zero. i.e.,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\ell_{I_t,t}N\gamma_t = \lim_{n\to\infty} N\sum_{i=1}^{N}\frac{1}{n}\sum_{t=1}^{n}\ell_{i,t}\gamma_t = 0.$$

Introduce $K_{i,n} = \frac{1}{n}\sum_{t=1}^{n}\ell_{i,t}$ then for all i

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n}\ell_{i,t}\gamma_t &= \frac{1}{n}\sum_{t=1}^{n}(tK_{i,t} - (t-1)K_{i,t-1})\gamma_t \\ &= K_{i,n}\gamma_n + \frac{1}{n}\sum_{t=1}^{n-1}(\gamma_t - \gamma_{t+1})\, tK_{i,t} \\ &\le K_{i,n}\gamma_n + \frac{1}{n}\sum_{t=1}^{n-1}\gamma_t K_{i,t} & (9) \\ &\le \sqrt{c}\frac{1}{N}n^{\nu/2-\alpha} + \frac{1}{nN}\sum_{t=1}^{n-1}t^{\nu/2-\alpha}\sqrt{c} \to 0 & (10)\end{aligned}$$

where the (9) holds because $(\gamma_t - \gamma_{t+1})t \le \gamma_t$ and (10) follows from $K_{i,n} \le \sqrt{cn^\nu}$, the definition of the parameters and $\alpha \ge (\nu + \rho)/2$. $\square$

Lemma 7 yields the relation between $\overline{L}_n$ and $\min_i \widetilde{L}_{i,n}$.

Lemma 7. *Under the conditions of Theorem 3,*

$$\limsup_{n\to\infty} \frac{1}{n}\left(\overline{L}_n - \min_i \widetilde{L}_{i,n}\right) \le 0 \qquad a.s.$$

Proof. We start by applying Lemma 2, that is,

$$\overline{L}_n - \min_i \widetilde{L}_{i,n} \le \frac{2\ln N}{\eta_{n+1}} + \sum_{t=1}^{n} \Phi(\mathbf{p}_t, \eta_t, -\widetilde{\boldsymbol{\ell}}_t). \tag{11}$$

To bound the quantity of $\Phi(\mathbf{p}_t, \eta_t, -\widetilde{\boldsymbol{\ell}}_t)$, our starting point is (3). Moreover,

$$\frac{\eta_t}{2}\sum_{i=1}^{N} p_{i,t}\widetilde{\ell}_{i,t}^2 = \frac{\eta_t}{2}\sum_{i=1}^{N} p_{i,t}\frac{\ell_{i,t}^2}{\widetilde{p}_{i,t}^2\varepsilon_t^2} S_t \mathbb{I}_{\{I_t=i\}} \le \frac{\eta_t}{2\gamma_t\varepsilon_t}\frac{S_t}{\varepsilon_t}\ell_{I_t,t}^2 \le \frac{\eta_t}{2\gamma_t\varepsilon_t}\frac{S_t}{\varepsilon_t}\sum_{i=1}^{N}\ell_{i,t}^2 \tag{12}$$

where the first inequality comes from $p_{I_t,t} \ge \gamma_t$. Combining this bound with (11), dividing by n and taking the limit we get

$$\limsup_{n\to\infty}\frac{1}{n}\left(\overline{L}_n - \min_i \widetilde{L}_{i,n}\right) \le \limsup_{n\to\infty}\frac{2\ln N}{n\eta_{n+1}} + \limsup_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\frac{\eta_t}{2\gamma_t\varepsilon_t}\frac{S_t}{\varepsilon_t}\sum_{i=1}^{N}\ell_{i,t}^2.$$

Let analyze separately the two terms on the right-hand side. The first term is zero because of the assumption of the Theorem 3. Concerning the second term, similarly to Lemma 6 we can split S_t/ε_t as follows: let us

$$\frac{S_t}{\varepsilon_t}\frac{\eta_t}{2\gamma_t\varepsilon_t}\sum_{i=1}^{N}\ell_{i,t}^2 = Z_t + \frac{\eta_t}{2\gamma_t\varepsilon_t}\sum_{i=1}^{N}\ell_{i,t}^2, \tag{13}$$

where Z_t is a martingale difference sequence. The variance is

$$\mathbf{Var}(Z_t) = \mathbb{E}\left[\frac{\eta_t^2 S_t}{\gamma_t^2\varepsilon_t^2}\left(\sum_{i=1}^{N}\ell_{i,t}^2\right)^2\right] = \frac{\eta_t^2}{\varepsilon_t\gamma_t^2}\left(\sum_{i=1}^{N}\ell_{i,t}^2\right)^2.$$

Application of Lemma 5 with $h_t = \frac{\eta_t^2}{\varepsilon_t\gamma_t^2}$ and $k_t = \left(\sum_{i=1}^{N}\ell_{i,t}^2\right)^2$ yields

$$\lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n} Z_t = 0 \qquad a.s.$$

where we used that

$$\frac{1}{n}\sum_{t=1}^{n} k_t \le \frac{1}{n}\left(\sum_{t=1}^{n}\sqrt{k_t}\right)^2 \le N^2c^2n^{1+2\nu}.$$

Finally, we have to prove that the sum of the second term in (13) goes to zero, that is,

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \sum_{i=1}^{N} \frac{\eta_t}{2\gamma_t \varepsilon_t} \ell_{i,t}^2 = 0$$

for which we use same argument as in Lemma 6. Introduce $K_{i,n} = \frac{1}{n}\sum_{t=1}^{n} \ell_{i,t}^2$ then we get

$$\begin{aligned}
\frac{1}{n}\sum_{t=1}^{n} \ell_{i,t}^2 \frac{\eta_t}{2\gamma_t\varepsilon_t} &= K_{i,n}\frac{\eta_n}{2\gamma_n\varepsilon_n} + \frac{1}{n}\sum_{t=1}^{n-1}\left(\frac{\eta_t}{2\gamma_t\varepsilon_t} - \frac{\eta_{t+1}}{2\gamma_{t+1}\varepsilon_{t+1}}\right) tK_{i,t} \\
&\le K_{i,n}\frac{\eta_n}{2\gamma_n\varepsilon_n} + \frac{1}{n}\sum_{t=1}^{n-1} \frac{\eta_t}{2\gamma_t\varepsilon_t} K_{i,t} \\
&\le Ncn^{\nu-1+\alpha+\beta+\delta} + \frac{1}{n}\sum_{t=1}^{n-1} Nct^{\nu-1+\alpha+\beta+\delta} \to 0
\end{aligned}$$

because of $K_{i,n} \le cn^{\nu}$ and $\nu < 1 - \alpha - \beta - \delta - \rho$. □

Finally, the last step is to analyze the difference between the estimated loss and the true loss.

Lemma 8. *Under the conditions of Theorem 3,*

$$\lim_{n\to\infty} \frac{1}{n}\left(\min_i \widetilde{L}_{i,n} - \min_i L_{i,n}\right) = 0 \qquad a.s.$$

Proof. First, bound the difference of the minimum of the true and the estimated loss. Obviously,

$$\begin{aligned}
\frac{1}{n}\left(\min_i \widetilde{L}_{i,n} - \min_j L_{j,n}\right) \le \sum_{i=1}^{N}\left|\frac{1}{n}\left(\widetilde{L}_{i,n} - L_{i,n}\right)\right| &= \sum_{i=1}^{N}\left|\frac{1}{n}\sum_{t=1}^{n}(\widetilde{\ell}_{i,t} - \ell_{i,t})\right| \\
&= \sum_{i=1}^{N}\left|\frac{1}{n}\sum_{t=1}^{n} Z_{i,t}\right|,
\end{aligned}$$

where $Z_{i,t}$ is martingale difference sequence for all i. As earlier, we use Lemma 5. First we bound $\mathbf{Var}(Z_{i,t})$ as follows

$$\mathbf{Var}(Z_{i,t}) = \mathbb{E}\widetilde{\ell}_{i,t}^2 \le \frac{\sum_{i=1}^{N} \ell_{i,t}^2}{\varepsilon_t\gamma_t}. \tag{14}$$

Applying Lemma 5 with parameters $k_t = \ell_{i,t}^2$ and $h_t = \frac{1}{\varepsilon_t\gamma_t}$, for each fixed i

$$\lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} Z_{i,t} = 0 \qquad a.s.$$

therefore

$$\lim_{n\to\infty}\sum_{i=1}^{N}\left|\frac{1}{n}\sum_{t=1}^{n}Z_{i,t}\right| = 0 \qquad a.s.$$

□

References

1. P. Auer, N. Cesa-Bianchi, Y. Freund, and R. E. Schapire. Gambling in a rigged casino: the adversial multi-armed bandit problem. In *Proceedings of the 36th Annual Symposium on Foundations of Computer Science, FOCS 1995*, pages 322–331, Washington, DC, USA, Oct. 1995. IEEE Computer Society Press, Los Alamitos, CA.
2. D. Blackwell. An analog of the minimax theorem for vector payoffs. *Pacific Journal of Mathematics*, 6:1–8, 1956.
3. N. Cesa-Bianchi, Y. Freund, D. P. Helmbold, D. Haussler, R. Schapire, and M. K. Warmuth. How to use expert advice. *Journal of the ACM*, 44(3):427–485, 1997.
4. N. Cesa-Bianchi and G. Lugosi. *Prediction, Learning, and Games.* Cambridge University Press, Cambridge, 2006.
5. N. Cesa-Bianchi, G. Lugosi, and G. Stoltz. Minimizing regret with label efficient prediction. *IEEE Trans. Inform. Theory*, IT-51:2152–2162, June 2005.
6. N. Cesa-Bianchi, Y. Mansour, and G. Stoltz. Improved second-order bounds for prediction with expert advice. In *COLT 2005*, pages 217–232, 2005.
7. N. Cesa-Bianchi, Y. Mansour, and G. Stoltz. Improved second-order bounds for prediction with expert advice, 2006. (submitted).
8. Y. S. Chow. Local convergence of martingales and the law of large numbers. *Annals of Mathematical Statistics*, 36:552–558, 1965.
9. L. Györfi and G. Lugosi. Strategies for sequential prediction of stationary time series. In M. Dror, P. L'Ecuyer, and F. Szidarovszky, editors, *Modelling Uncertainty: An Examination of its Theory, Methods and Applications*, pages 225–248. Kluwer Academic Publishers, 2001.
10. L. Györfi and Gy. Ottucsák. Sequential prediction of unbounded stationary time series, 2006.
11. A. György and Gy. Ottucsák. Adaptive routing using expert advice. *The Computer Journal*, 49(2):180–189, 2006.
12. J. Hannan. Approximation to bayes risk in repeated plays. In M. Dresher, A. Tucker, and P. Wolfe, editors, *Contributions to the Theory of Games*, volume 3, pages 97–139. Princeton University Press, 1957.
13. S. Hart and A. Mas-Colell. A simple adaptive procedure leading to correlated equilibrium. *Econometria*, 68(5):181–200, 2002.
14. N. Littlestone and M. K. Warmuth. The weighted majority algorithm. *Information and Computation*, 108:212–261, 1994.
15. V. Vovk. Aggregating strategies. In *Proceedings of the Third Annual Workshop on Computational Learning Theory*, pages 372–383, Rochester, NY, Aug. 1990. Morgan Kaufmann.