
Bound on high probability regret in loss-bandit game

Peter Auer

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

György Ottucsák*

Department of Computer Science and Information Theory
Budapest University of Technology and Economics,
Magyar Tudósok körútja 2., Budapest, Hungary, H-1117
oti@szit.bme.hu

Abstract

In this paper the multi-armed bandit problem is considered in bounded loss case, that is, when after choosing an action, the algorithm learns only the loss of the chosen action. An algorithm is given, whose average loss in n rounds exceeds that of the performance of the best fix action at most by an amount proportional to $1/\sqrt{n}$.

1 Introduction

In on-line 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 the best action for any possible behavior of the environment.

In the multi-armed bandit (MAB) setting only the loss of the chosen action is revealed for the algorithm. Auer *et al.* [2] gave an algorithm, whose average loss in n rounds exceeds that of the performance of the best fix action at most by an amount proportional to $\sqrt{N \ln N/n}$, where N is the number of the actions. This algorithm uses gain instead of the loss and during the analysis switching was need between the loss and gain.

However in some cases this "gain-loss switching" causes problem. In case of unbounded loss or in the case when the bound of the loss is unknown it is not possible to convert loss to the gain. Another case is the shortest path problem (see György *et al.* [5]) where all paths have to be same length because of the "gain-loss switching". We give an algorithm which uses only losses and it obtains same convergence rate like the algorithm proposed by [2].

2 The model

The MAB problem considered in this paper is described as follows. Suppose an algorithm has to make a sequence of actions. At each time instance $t = 1, 2, \dots$,

1. the environment decides on the losses $0 \leq \ell_{i,t} \leq 1$ of an action $i \in \{1, \dots, N\}$,
2. the algorithm chooses an action $I_t \in \{1, \dots, N\}$,
3. the algorithm suffers loss $\ell_{I_t,t}$,
4. the algorithm receives feedback on his loss.

*Computer and Automation Research Institute of the Hungarian Academy of Sciences, Machine Learning Research Group, Kende u. 11-13, Budapest, Hungary, H-1111

After n rounds the loss of the algorithm and the loss of the actions are

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

and the performance of the algorithm is measured by its regret per round, $\frac{1}{n} (\widehat{L}_n - \min_i L_{i, n})$.

3 The algorithm

In this section we show a new algorithm GREEN.SHIFT for MAB problem and we give an upper bound on its performance. The proposed algorithm is a "shifted" version of algorithm GREEN which was introduced by Allenberg *et. al.* [1]. GREEN is a variant of the weighted majority (WM) algorithm of Littlestone and Warmuth [6].

Denote by $p_{i, t}$ the probability of choosing action $i \in \{1, \dots, N\}$ at time t in case of the WM algorithm, that is,

$$p_{i, t} = \frac{e^{-\eta L'_{i, t-1}}}{\sum_{j=1}^N e^{-\eta L'_{j, t-1}}},$$

where $L'_{i, t}$ is so called cumulative biased estimated loss will be specified later. Similarly to the original GREEN algorithm GREEN.SHIFT uses also *modified probabilities* $\tilde{p}_{i, t}$ which can be determined from $p_{i, t}$,

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

where c_t is the normalizing factor and γ is a positive (time-dependent) threshold. The modified probabilities inspired the name of the algorithm. Namely it was named after the well-known idiom: "The grass is always greener on the other side of the fence", since the action it did not choose ($\tilde{p}_{i, t} = 0$) that had the best possible payoff (the zero loss).

Denote $\tilde{\ell}_{i, t}$ the conditional unbiased estimation of the true loss of each action with respect to its natural filtration, that is, $\tilde{\ell}_{i, t} = \mathbb{I}_{\{I_t=i\}} \ell_{i, t} / \tilde{p}_{i, t}$ where I_t is the choice of the algorithm at time t according to the distribution $\tilde{p}_{i, t}$ and $\mathbb{E}_t[\tilde{\ell}_{i, t}] \stackrel{\text{def}}{=} \mathbb{E}[\tilde{\ell}_{i, t} | I_1, \dots, I_{t-1}] = \ell_{i, t}$.

Instead of the unbiased estimate, a slightly smaller quantity is used by the algorithm. The (biased) estimated loss is

$$\ell'_{i, t} = \tilde{\ell}_{i, t} - \frac{\beta}{\max\{\tilde{p}_{i, t}, \gamma\}},$$

where β is a positive parameter and the maximum is necessary to avoid divide by zero. Then the cumulative estimated loss of an action is given by $L'_{i, n} = \sum_{t=1}^n \ell'_{i, t}$. The resulting algorithm is given in Figure 1.

Theorem 1 For any $0 < \delta < 1$ and parameters

$$\sqrt{\frac{\ln(N/\delta)}{nN}} \leq \beta \leq \frac{1}{N}, \quad \beta \leq \gamma \leq \frac{1}{N} \quad \text{and} \quad 0 < \eta \leq \sqrt{\frac{\ln N}{nN}},$$

the performance of Algorithm 1 can be bounded with probability at least $1 - \delta$ as

$$\widehat{L}_n \leq N\gamma\widehat{L}_n + 2\beta nN + (1 + \eta N) \min_i L_{i, n} + \eta\beta nN^2 + N \ln(1/\gamma) + 2N\eta + \frac{\ln N}{\eta}.$$

In particular, choosing $\beta = \sqrt{\frac{\ln(N/\delta)}{nN}}$, $\gamma = \beta$, $\eta = \sqrt{\frac{\ln N}{nN}}$ and if $n \geq N \ln(N/\delta)$ then we have

$$\frac{1}{n} (\widehat{L}_n - \min_i L_{i, n}) \leq 7\sqrt{N \ln(N/\delta)/n} + \frac{1}{2n} N \ln(nN).$$

For the proof of the theorem we need the following 2 lemmas. The first lemma is a simple modification of [3, Lemma 6.7].

Algorithm GREEN.SHIFT

Let $\eta > 0$, $\beta > 0$ and $\gamma > 0$.

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

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

(1) Calculate the weights of the actions

$$w_{i,t} = e^{-\eta L'_{i,t-1}} \quad i = 1, \dots, N \quad \text{and} \quad W_t = \sum_{i=1}^N w_{i,t}.$$

(2) Calculate the probability distribution

$$p_{i,t} = \frac{w_{i,t}}{W_t} \quad i = 1, \dots, N.$$

(3) Calculate the modified probabilities

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

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

(4) Compute the estimated loss for all $i = 1, \dots, N$

$$\ell'_{i,t} = \tilde{\ell}_{i,t} - \frac{\beta}{\max\{\tilde{p}_{i,t}, \gamma\}} = \begin{cases} \frac{\ell_{i,t}}{\tilde{p}_{i,t}} - \frac{\beta}{\max\{\tilde{p}_{i,t}, \gamma\}} & \text{if } I_t = i; \\ -\frac{\beta}{\max\{\tilde{p}_{i,t}, \gamma\}} & \text{otherwise.} \end{cases}$$

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

$$L'_{i,t} = L'_{i,t-1} + \ell'_{i,t}.$$

Figure 1: Algorithm GREEN.SHIFT for MAB

Lemma 1 Under the assumptions of Theorem 1 for any $0 < \delta < 1$ we have

$$\mathbb{P}(L'_{i,n} > L_{i,n} + \beta n N) \leq \frac{\delta}{N}, \quad i \in \{1, \dots, N\}.$$

The following lemma is a variant of [1, Theorem 2].

Lemma 2 Under the assumptions of Theorem 1 for the cumulative estimated loss we have

$$L'_{i,n} \leq \min_{j=1, \dots, N} L'_{j,n} + \frac{\ln(1/\gamma)}{\eta}.$$

Proof of Theorem 1. For the proof of theorem the quantity of $\ln \frac{W_n}{W_0}$ is bounded, where

$$W_t = \sum_{i=1}^N w_{i,t}, \quad t \geq 1 \quad \text{and} \quad W_0 = N.$$

The lower bound is

$$\ln \frac{W_n}{W_0} = \ln \left(\sum_{i=1}^N e^{-\eta L'_{i,n}} \right) - \ln N \geq \ln \left(\max_{i=1, \dots, N} e^{-\eta L'_{i,n}} \right) - \ln N = -\eta \min_{i=1, \dots, N} L'_{i,n} - \ln N. \quad (1)$$

For the upper bound note that $-\eta \ell'_{i,t} \leq 1$ for all i and t , therefore

$$\ln \frac{W_t}{W_{t-1}} = \ln \sum_{i=1}^N p_{i,t} e^{-\eta \ell'_{i,t}} \leq \ln \sum_{i=1}^N p_{i,t} (1 - \eta \ell'_{i,t} + \eta^2 \ell'_{i,t}^2) \leq -\eta \sum_{i=1}^N p_{i,t} \ell'_{i,t} + \eta^2 \sum_{i=1}^N p_{i,t} \ell'_{i,t}^2. \quad (2)$$

Next we bound the sums in (2). On the one hand,

$$\sum_{i=1}^N p_{i,t} \ell'_{i,t} = \frac{p_{I_t,t}}{\tilde{p}_{I_t,t}} \ell_{I_t,t} - \beta \sum_{i=1}^N \frac{p_{i,t}}{\max\{\tilde{p}_{i,t}, \gamma\}} \geq \frac{p_{I_t,t}}{\tilde{p}_{I_t,t}} \ell_{I_t,t} - \beta N \geq (1 - N\gamma) \ell_{I_t,t} - \beta N,$$

since $p_{I_t,t}/\tilde{p}_{I_t,t} = 1/c_t = \sum_{j:p_{j,t} \geq \gamma} p_{j,t} = 1 - \sum_{j:p_{j,t} < \gamma} p_{j,t} \geq 1 - N\gamma$.

On the other hand,

$$\begin{aligned} \sum_{i=1}^N p_{i,t} \ell'_{i,t} &= \sum_{i=1}^N p_{i,t} \left(\tilde{\ell}_{i,t} - \frac{\beta}{\max\{\tilde{p}_{i,t}, \gamma\}} \right) \ell'_{i,t} \leq \ell_{I_t,t} \ell'_{I_t,t} - \beta \sum_{i=1}^N \frac{p_{i,t} \ell'_{i,t}}{\max\{\tilde{p}_{i,t}, \gamma\}} \\ &\leq \ell_{I_t,t} \ell'_{I_t,t} + \beta^2 \sum_{i=1}^N \frac{1}{\max\{\tilde{p}_{i,t}, \gamma\}} \\ &\leq \ell_{I_t,t} \ell'_{I_t,t} + \frac{\beta^2 N}{\gamma} \\ &\leq \sum_{i=1}^N \ell'_{i,t} + \frac{\beta N}{\gamma} + \frac{\beta^2 N}{\gamma} \\ &\leq \sum_{i=1}^N \ell'_{i,t} + N + \beta N, \end{aligned}$$

where the last inequality follows from $\beta \leq \gamma$. Summing over $t = 1, \dots, n$, we have that

$$\ln \frac{W_n}{W_0} \leq -\eta \hat{L}_n + N\eta\gamma \hat{L}_n + \eta\beta nN + \eta^2 \sum_{i=1}^N L'_{i,n} + \eta^2 2N. \quad (3)$$

Plug the results of Lemma 2 into (3) we get

$$\ln \frac{W_n}{W_0} \leq -\eta \hat{L}_n + N\eta\gamma \hat{L}_n + \eta\beta nN + \eta^2 N \min_{i=1, \dots, N} L'_{i,n} + \eta N \ln(1/\gamma) + \eta^2 2N. \quad (4)$$

Combining (1) and (4) we obtain

$$\hat{L}_n \leq N\gamma \hat{L}_n + \beta nN + (1 + \eta N) \min_{i=1, \dots, N} L'_{i,n} + N \ln(1/\gamma) + 2\eta N + \frac{\ln N}{\eta}.$$

By Lemma 1 and the union bound we have at least $1 - \delta$

$$\hat{L}_n \leq N\gamma \hat{L}_n + 2\beta nN + (1 + \eta N) \min_{i=1, \dots, N} L_{i,n} + \eta\beta nN^2 + N \ln(1/\gamma) + 2\eta N + \frac{\ln N}{\eta}$$

as desired. \square

References

- [1] C. Allenberg, P. Auer, L. Györfi, and Gy. Ottucsák. Hannan consistency in on-line learning in case of unbounded losses under partial monitoring. In *Proceedings of 17th International Conference on Algorithmic Learning Theory, ALT 2006, LNAI 4264*, pages 229–243, Barcelona, Spain, Oct. 2006.
- [2] P. Auer, N. Cesa-Bianchi, Y. Freund, and R. Schapire. The non-stochastic multi-armed bandit problem. *SIAM Journal on Computing*, 32(1):48–77, 2002.
- [3] N. Cesa-Bianchi and G. Lugosi. *Prediction, Learning, and Games*. Cambridge University Press, Cambridge, 2006.
- [4] L. Devroye, L. Györfi, and G. Lugosi. *A Probabilistic Theory of Pattern Recognition*. Springer-Verlag, New York, 1996.
- [5] A. Györfi, T. Linder, and Gy. Ottucsák. The shortest path problem under partial monitoring. In *Proc. of 19th Annual Conference on Learning Theory, COLT 2006, LNAI 4005*, pages 468–482, Pittsburgh, USA, June 2006.
- [6] N. Littlestone and M. K. Warmuth. The weighted majority algorithm. *Information and Computation*, 108:212–261, 1994.

4 Appendix

4.1 Proof of Lemma 1

For any $u > 0$ and $c > 0$ the Chernoff bounding technique (see, e.g., [4]) implies

$$\mathbb{P}(L'_{i,n} > L_{i,n} + u) \leq e^{-cu} \mathbb{E} e^{c(L'_{i,n} - L_{i,n})}. \quad (5)$$

Letting $u = \beta n N$ and $c = \beta$, therefore from (5):

$$e^{-cu} \mathbb{E} e^{c(L'_{i,n} - L_{i,n})} = e^{-\beta^2 n N} \mathbb{E} e^{\beta(L'_{i,n} - L_{i,n})} \leq \frac{\delta}{N} \mathbb{E} e^{\beta(L'_{i,n} - L_{i,n})},$$

where the inequality comes from $\sqrt{\frac{\ln(N/\delta)}{nN}} \leq \beta$. Thus it suffices to prove that

$$\mathbb{E} e^{\beta(L'_{i,n} - L_{i,n})} \leq 1.$$

For $t = 1, \dots, n$, introducing, a random variable $Z_t = e^{\beta(L'_{i,t} - L_{i,t})}$ we clearly have

$$Z_t = e^{\beta(\ell'_{i,t} - \ell_{i,t})} Z_{t-1}.$$

Note that $\beta(\ell'_{i,t} - \ell_{i,t}) \leq 1$ because

$$\beta \left(\frac{\ell_{i,t} \mathbb{1}_{\{I_t=i\}}}{\tilde{p}_{i,t}} - \frac{\beta}{\max\{\tilde{p}_{i,t}, \gamma\}} - \ell_{i,t} \right) \leq \frac{\beta \ell_{I_t,t}}{\tilde{p}_{I_t,t}} \leq \frac{\beta \ell_{I_t,t}}{\gamma} \leq 1$$

where the second inequality comes from $\beta \leq \gamma$. Moreover, $e^x \leq 1 + x + x^2$ for $x \leq 1$ therefore

$$\begin{aligned} \mathbb{E}_t[Z_t] &= Z_{t-1} \mathbb{E}_t \left[e^{\beta(\tilde{\ell}_{i,t} - \frac{\beta}{\max\{\tilde{p}_{i,t}, \gamma\}} - \ell_{i,t})} \right] \\ &= Z_{t-1} e^{-\frac{\beta^2}{\max\{\tilde{p}_{i,t}, \gamma\}}} \mathbb{E}_t \left[e^{\beta(\tilde{\ell}_{i,t} - \ell_{i,t})} \right] \\ &\leq Z_{t-1} e^{-\frac{\beta^2}{\max\{\tilde{p}_{i,t}, \gamma\}}} \mathbb{E}_t \left[1 + \beta(\tilde{\ell}_{i,t} - \ell_{i,t}) + \beta^2(\tilde{\ell}_{i,t} - \ell_{i,t})^2 \right] \\ &= Z_{t-1} e^{-\frac{\beta^2}{\max\{\tilde{p}_{i,t}, \gamma\}}} \mathbb{E}_t \left[1 + \beta^2(\tilde{\ell}_{i,t} - \ell_{i,t})^2 \right] \\ &\leq Z_{t-1} e^{-\frac{\beta^2}{\max\{\tilde{p}_{i,t}, \gamma\}}} \left(1 + \frac{\beta^2 \ell_{i,t}^2}{\max\{\tilde{p}_{i,t}, \gamma\}} \right) \\ &\leq Z_{t-1}, \end{aligned}$$

where we used $\mathbb{E}_t[\tilde{\ell}_{i,t} - \ell_{i,t}] = 0$ and $1 + x \leq e^x$. Taking expected values of both sides of the inequality we have $\mathbb{E}_t Z_t \leq \mathbb{E}_t Z_{t-1}$ and since $\mathbb{E}_t Z_1$ the proof is concluded. \square

4.2 Proof of Lemma 2

Let $T_i = \max\{0 \leq t \leq n : p_{i,t} \geq \gamma\}$ be the last round where $\tilde{p}_{i,t} > 0$. Therefore,

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

where $i^* = \arg \min_{i=1, \dots, N} L'_{i,n}$. After rearranging we obtain

$$L'_{i,T_i} \leq L'_{i^*,T_i} + \frac{\ln(1/\gamma)}{\eta}.$$

Since $L'_{i,T_i} = L'_{i,n} + \frac{\beta(n-T_i-1)}{\gamma}$ and $L'_{i^*,T_i} \leq L'_{i^*,n} + \sum_{t=T_i+1}^n \frac{\beta}{\max\{\tilde{p}_{i^*,t}, \gamma\}}$ we get that

$$L'_{i,n} \leq L'_{i^*,n} + \beta \sum_{t=T_i+1}^n \left(\frac{1}{\max\{\tilde{p}_{i,t}, \gamma\}} - \frac{1}{\gamma} \right) + \frac{\ln(1/\gamma)}{\eta} \leq L'_{i^*,n} + \frac{\ln(1/\gamma)}{\eta}.$$

\square