Budapest University of Technology and Economics
Department of Computer Science and Information Theory

MACHINE LEARNING FOR INFO-COMMUNICATION SYSTEMS

by

György Ottucsák

Ph.D. Thesis

Supervisor:
Dr. László Györfi

Department of Computer Science and Information Theory
Faculty of Electrical Engineering and Informatics
Budapest University of Technology and Economics
Budapest, Hungary
August, 2007

Alulírott Ottucsák György kijelentem, hogy ezt a doktori értekezést magam készítettem, és abban csak a megadott forrásokat használtam fel. Minden olyan részt, amelyet szó szerint vagy azonos tartalomban, de átfogalmazva más forrásból átvettem, egyértelműen, a forrás megadásával megjelöltem.

Budapest, 2007. Augusztus 29.

Ottucsák György

A disszertáció bírálatai és a védésről készült jegyzőkönyv megtekinthető a Budapesti Műszaki és Gazdaságtudományi Egyetem Villamosmérnöki és Informatikai Karának Dékáni Hivatalában.

Abstract

In this thesis efficient algorithms for sequential prediction (decision) problems are studied. In general, the algorithm has to guess the next element of an unknown sequence using some knowledge about the past of the sequence and other side information. In this model the goal of the algorithm is to minimize its cumulative loss, which is accumulated from round to round (in each round one decision is made) where the loss is scored by some fixed loss function. The sequence of the outcomes is a product of some unspecified mechanism, which could be deterministic, stochastic or even adversarially adaptive to our own behavior.

As the first result of the thesis, an algorithm is given for the problem when the loss is unbounded and its performance is studied under various partial information (also called partial monitoring) settings. A wide class of partial monitoring problems are introduced: the combination of the label efficient and multi-armed bandit problems. In this setting the algorithm is only informed about the performance of its decision with probability $\varepsilon \leq 1$ and does not have access to the losses it would have suffered if it had made a different decision. It is shown that consistency can be achieved for unbounded losses, too, depending on the growth rate of the overall "worst" decision's average loss. Moreover, the above result can be applied to solve the special problem, when the loss is bounded. For bounded losses a simple modification of the previous algorithm is offered; its convergence rate coincides with that of the best "earlier algorithms", but it can be applied more easily for real life problems.

In the next part, the on-line shortest path problem is considered under various models of partial monitoring. Given a weighted directed acyclic graph whose edge weights can change in an arbitrary (adversarial) way, an algorithm (decision maker) has to choose in each round of a game a path between two distinguished vertices such that the loss of the chosen path (defined as the sum of the weights of its composing edges) be as small as possible. In a setting generalizing the multi-armed bandit problem, after choosing a path, the algorithm learns only the weights of those edges that belong to the chosen path. For this problem, an algorithm is given whose average cumulative loss in n rounds exceeds that of the best path, matched off-line to the entire sequence of the edge weights, by a quantity that is proportional to $1/\sqrt{n}$ and depends only polynomially on the number of edges of the

graph. The algorithm can be implemented with complexity that is linear in the number of rounds n (i.e., the average complexity per round is constant) and in the number of edges. An extension to the so-called label efficient setting is also given, in which the algorithm is informed about the weights of the edges corresponding to the chosen path at a total of $m \ll n$ time instances. Another extension is shown, where the algorithm competes against a time-varying path, a generalization of the problem of tracking the best expert. A version of the multi-armed bandit setting for shortest path is also discussed where the algorithm learns only the total weight of the chosen path but not the weights of the individual edges on the path. Applications to routing in packet switched networks along with simulation results are also presented.

Finally, a prediction strategy is introduced for unbounded stationary and ergodic real-valued processes and show that the average of squared errors of the algorithm converges, almost surely, to that of the optimum, given by the Bayes predictor. The algorithm is based on a combination of several simple predictors, where for this combination the methodology and results of the previous parts of the thesis are used. Furthermore an extension for the noisy setting is offered, that is when the algorithm has access only to the noisy version of the outcome sequence e.g. the "clean" process is passed through a fixed binary memoryless channel. A simple universally consistent classification scheme is provided for zero-one loss in this noisy setting.

Contents

Chapter 1

Introduction

In this chapter the framework of sequential decision problems is introduced. Section 1.1 describes the main concepts and motivation of the sequential decision problems. In Section 1.2 literature overview is given. Our contribution is described in Section 1.3, as well as a detailed overview of the thesis.

1.1 Motivation

The goal of this thesis is to design general purpose, efficient algorithms for sequential prediction (decision) problems. Prediction, as we understand it in this thesis, is concerned with guessing the short term evolution of certain phenomena. Examples include forecasting whether tomorrow will be rainy or not, or guessing the route with lowest traffic between our home and our workplace on the following working time period. These tasks look similar at an abstract level: one has to guess the next element of an unknown sequence using some knowledge about the past of the sequence and other side information available. Such problems naturally arise in real-world applications from portfolio selection in financial market through real-time optimization of websites to routing in the communication networks.

In the classical statistical theory of sequential prediction, the sequence of the elements, so called outcomes, is assumed to be a realization of a stationary stochastic process. In such a setup, the statistical property of the process based on past observations can be estimated and using this estimation efficient prediction strategies can be constructed. In that case, the performance of a prediction strategy is usually evaluated by expected value of some loss function which measures the "distance" between the predicted value and the true outcome.

However, in a large part of this thesis we use a different viewpoint. We abandon the assumption that the outcomes are generated by a well-behaved stochastic process and view the sequence of the outcomes as a product of some unspecified mechanism, which could be deterministic, stochastic or even adversarially adaptive to our own behavior. This setup

where *no probabilistic assumption* is made on how the sequence is generated is often referred to as *prediction of individual sequences*.

In this model the goal of the algorithm is to minimize its cumulative loss, which is accumulated from round to round (in each round one decision is made) where the loss is scored by some fixed loss function. At the same time, without a probabilistic model it is non-obvious how to measure the performance of the algorithm. There is no natural baseline as in the stochastic case, and for example it is easy to see that it is not possible to minimize the cumulative loss simultaneously for all possible sequences. To provide such a baseline one of the possible way is to define a set of reference forecasters (prediction rules), so called *experts*. Then the performance of the algorithm is evaluated relative to this set of experts, and the goal is to perform asymptotically as well as the best expert from the reference class matched to the observed outcome sequence off-line. The experts make their decisions available to the algorithm before the next outcome is revealed, and based on these "pieces of advices" the algorithm forms its own decision to keep close its cumulative loss to the cumulative loss of the best expert.

The difference between the cumulative loss of the algorithm to that of the best expert is called *regret*, as it measures how much the algorithm regrets, in hindsight, of not having followed the advice of the best expert.

On the one hand for "small" expert classes the regret of the algorithm converges "fast" to zero, however, the cumulative loss of the best expert may be "large". Borrowing an analogy from nonparametric statistics the first error criterion is called *estimation error* of the algorithm and the second one is the *approximation error* of the expert class. On the other hand for "large" expert classes it is vice versa, the convergence of the regret of the algorithm is slower, at the same time the cumulative loss of the best expert is smaller. In most of this thesis we focus on the minimization of the regret.

The advantage of this novel technique, prediction for individual sequences is twofold. On the one hand it is able to handle the case when the sequence of the outcomes are generated by an *adversarial mechanism*. In that case one cannot assume any stationary and probabilistic mechanism for the sequence. Indeed, that is realistic in e.g. *reactive environments* where the choice of the algorithm influences the behaviour of the environment (see below for real-world problems). On the other hand it has huge increment in the field of non-parametric statistics. Namely, one may have a probabilistic model, however, there is a need to construct prediction with good rates, i.e., to adapt the parameters of the algorithm. There are such adaptations: splitting, cross validation, complexity regularization, etc., but they work well only for memoryless sequences, which restricts seriously their applicability. Another important problem is the universally consistent prediction of ergodic sequence. The concept of individual sequence is extremely efficient such that the choices of the parameters of the algorithm are considered as experts, and the bounds on the performance of the combined (aggregated) algorithm does not depend on the properties of the actual sequence, and so these bounds result in optimal adaptation both for memoryless sequences or in universal consistency for ergodic sequences.

The concrete interpretation of the "experts" depends on the specific application. In the sequel some important problems which naturally cast as experts' advice (sequential

decision) problems are shown mostly from the framework of info-communication systems.

Let us see first an example for the above mentioned adaptation from the field of pattern recognition. We have a k-NN (Nearest Neighbor) classifier and our goal is to find the best value of the number of k. In that case each expert can run a k-NN classifier with different values of k. Another typical choice of the number of the neighbors is: $c_k n^{1/(d+2)}$, where d is the dimension of the samples. In that case each expert can use different parameters c_k.

Second, let us see some examples from info-communication systems. In these problems, the parameters of the networks and protocols are needed to be well tuned to ensure that the networks operate at the desired Quality of Service (QoS) level. For instance, the class of the experts can be some Transmission Control Protocol (TCP) variants that may use different *parameter settings* and the algorithm competes with the TCP variant which has the best parameters in hindsight. In particular, this setup is reasonable when the TCP variant has to provide good performance in a heterogeneous environment or in case of delay based TCP variants, like TCP Vegas and FAST TCP, whose performance are ultra-sensitive to the value of the parameters controlling the number of backlogged packets in the buffers of the routers on the path. These parameters are responsible for the long run performance of the flow (as throughput and fairness) and since Vegas and FAST keep these parameters constant, they cannot adapt well to the current characteristics of the network.

Another extensively studied issue is the estimation of the available bandwidth in high speed networks where the previously developed TCP variants (e.g: Reno) do not provide good utilization of the link or they may find available bandwidth too slowly. In that case each *bandwidth estimation* technique or protocol can be considered as an expert.

This approach also could be used for *modeling the bidding strategy* of participants of an auction. In Dynamic Spectrum Access networks where the allocation of the spectrum is based on an auction mechanism (e.g: English or Vickrey auction) the set of experts contains some fixed price or more complex bidding strategies and the goal of the algorithm is that its expense do not exceed too much the costs of the best bidding strategy.

Other interesting applications are in *adaptive routing*, which is of great importance in the maintenance of packet switched communication networks. A sufficiently flexible algorithm can yield increased QoS, such as reduced packet loss ratio or delay, even in case of link failures or substantially varying traffic scenarios. These algorithms require constant monitoring of the network state, and the measured information is combined to update the routing tables. Such combination can be done, for instance, with a combination of the experts' advice. More precisely, for each packet the routing algorithm has to choose an expert (path) from source to destination on which the packet is to be sent. The loss corresponding to the decision is the value of the QoS parameter we wish to optimize, such as the delay, or the number of hops on the path, or the packet loss ratio due to insufficient buffer size.

The performance of any algorithm obviously depends on how much information is available to the algorithm (the decision maker) about the experts' and its own performance. Often, only *partial information* is available to the algorithm, this is the so called *partial monitoring* setting. For example, in case of adaptive routing, it is not feasible to assume that the algorithm knows the delays of each path in the network in each moment. It is

more natural to assume that at each moment the algorithm learns information about the delay of the path its packet is sent on, and no information is available about the delay it would have suffered had it chosen a different path (e.g., this feedback is available through acknowledgments). Another example when the decision maker has the option to query the delays at a certain moments, e.g., with flooding.

Both full information (when the performance of each expert is available for the algorithm) and partial monitoring problems are well-studied in case when the experts class is "small" and the loss function is bounded. In these settings good convergence rates and also consistency results are considered. The extensions of these results to "large" expert classes or to unbounded loss functions are important open-questions, but unfortunately usually they make difficulties. For a general class of experts the computational complexity of the expert algorithms available in the literature usually grows linearly with time and with the number of the experts. This complexity may be prohibitive for large classes of experts, e.g., when an expert is a path in a network (the number of such paths is typically exponential in the size of the network). Finally, in most cases, one assumes that the loss is bounded, and such a bound is known in advance, during the design of the algorithm, which is not acceptable in case of many real-life applications. For instance, in case of adaptive routing, the algorithm has no information about the maximum value of the delay.

At this point some questions arise in connection with the above applications. Is it possible to construct an algorithm whose performance achieves asymptotically the performance of the best expert (consistency) if the bound of the loss is unknown? If yes, then is there a way to somehow extend the result for the case of partial monitoring? Furthermore, do consistent algorithms exist with low time and space complexity if the number of the experts is large (e.g., the number of the paths in a network) under partial monitoring? If we have some (stochastic) assumptions about the behavior of the outcome sequence (e.g., the delays on the links are realizations of stationary and ergodic processes in the routing problem) is it possible to improve in some sense the convergence of the algorithm? Most of the material in this thesis is devoted to provide answers to these and to related questions.

1.2 Literature overview

Research on sequential decision problems started in the 1950s, see, for example, Blackwell [15] and Hannan [43] for some of the basic results, and gained new life in the 1990s following the work of Vovk [70], Littlestone and Warmuth [53], and Cesa-Bianchi *et al.* [20]. These results show that for any bounded loss function, if the decision maker has access to the past losses of all experts, then it is possible to construct on-line algorithms that perform, for any possible behavior of the environment, almost as well as the best expert. For a good survey on prediction of individual sequences, the reader is referred to, e.g., the recent book of Cesa-Bianchi and Lugosi [21].

The theory has been extended to different directions, considering complexity issues or the amount of available information.

A representative example of the partial monitoring problem is the *multi-armed bandit*

problem where the algorithm has only information on the loss of the chosen expert. This problem was originally considered in the stochastic setting – it was assumed that the losses are randomly and independently drawn with respect to a fixed but unknown distribution – by Robbins [63] and Lai and Robbins [52] (for a recent efficient solution, see Auer *et al.* [4]). For the non-stochastic setting consistent algorithms are given in Auer *et al.* [6], [5] and Hart and Mas Colell [44]. Auer *et al.* [5] gave an algorithm whose average cumulative loss in n rounds exceeds that of the best expert by a quantity that is proportional to $\sqrt{N/n}$, where N is the number of the experts. Another example of partial monitoring problems is the *label efficient prediction problem*, where it is expensive to obtain the losses of the experts, and therefore the algorithm has the option to query this information (see Helmbold and Panizza [45] and Cesa-Bianchi *et. al* [22]). The main open problem left is to extend these results to unbounded losses.

For large classes of experts, such as the shortest path problem in graphs, the special structure of the experts allows to implement the algorithms with significantly lower complexity in the full information case, see, e.g., Helmbold and Schapire [64], Mohri [55], Auer and Warmuth [9], Helmbold and Warmuth [46], Takimoto and Warmuth [68], [69], Kalai and Vempala [49] and György *et al.* [36]. However, in case of the multi-armed bandit problem, if one applies the general bandit algorithm of Auer *et al.* [5], the resulting regret bound (on the average excess loss relative to the best expert) will be unacceptably large to be of practical use because of its square-root-type dependence on the number of expert. The most important issues here are the improvement of the algorithms in multi-armed bandit problem to achieve better regret bounds and further reduction of the computational complexity.

One may wonder whether it is possible to improve the above results if we have some probabilistic assumptions about the behavior of the outcome sequence. If the outcome sequence is a realization of a stationary and ergodic random process then one can show an algorithm (strategy) whose performance converges not only to the performance of the best expert, but in case of a carefully defined class of the experts, it also converges to the theoretical optimum that can be achieved in full knowledge of the underlying distribution generating the outcome sequence. A strategy is called *universally consistent* if it achieves asymptotically this optimum. In case of squared loss, Algoet [1] and Morvai, Yakowitz, and Györfi [57] proved that there exists a prediction strategy that can achieve this well-defined optimum. Györfi and Lugosi [32] introduced a simple universally consistent prediction strategy. We refer to Nobel [58], Singer and Feder [65], [66] and Yang [74] for closely related recent works. In case of $0-1$ loss, Ornstein [59] and Bailey [12] proved the existence of universally consistent predictors. This was later generalized by Algoet [1]. A simpler estimator with the same convergence property was introduced by Morvai, Yakowitz, and Györfi [57]. Motivated by the need for a practical estimator, Morvai, Yakowitz, and Algoet [56] introduced an even simpler algorithm. However, it is not known whether their predictor is universally consistent. Györfi, Lugosi, and Morvai [33] introduced a simple randomized universally consistent procedure with a practical appeal. Weissman and Merhav [72], [73] studied consistency in noisy environment.

1.3 Contribution and thesis overview

In this thesis we address some fundamental open questions of the sequential decision problems.

In **Chapter 2** we introduce the general model of sequential decision problems and accurately define specialized problems and algorithms of which we make extensive use later in this thesis. Moreover, this chapter also contains a more detailed literature overview.

As mentioned before, if the bound of the loss is unknown beforehand or if it can slowly grow with time, most of the existing algorithms are not applicable. In **Chapter 3** we give a new algorithm for this situation and study its performance under various partial observation settings. We introduce a wide class of partial monitoring problems: *the combination of the label efficient problem and the multi-armed bandit problem.* In the label efficient setting the algorithm is informed about the *experts' performance* only with probability $\varepsilon \leq 1$, while in the model of multi-armed bandit, only the performance of the *chosen expert* is known. In the combination of the label efficient problem and the multi-armed bandit problem the algorithm is only informed about the performance of the *chosen* expert with probability $\varepsilon \leq 1$. We show that consistency can be achieved for unbounded losses, if the growth rate of the worst expert's average square of the losses is sublinear in the number of rounds. Moreover, the above result can be applied to solve the special problem when the loss is bounded. For bounded losses a simple modification of the previous algorithm is offered; its convergence rate coincides with that of an earlier algorithm due to Auer *et al.* [5], but it can be applied more easily to practical problems.

In many applications the set of experts has a certain structure that may be exploited to construct efficient on-line decision algorithms. Construction of such algorithms has been of great interest in computational learning theory. In **Chapter 4** we study the on-line shortest path problem, a representative example of structured expert classes that has received attention in the literature for its many applications, including, among others, routing in communication networks and data compression. In this problem, a weighted directed (acyclic) graph is given whose edge weights can change in an arbitrary manner, and in each round the decision maker has to choose a path between two distinguished vertices such that the loss of the chosen path (defined as the sum of the weights of its composing edges) be as small as possible. In the multi-armed bandit setting, after choosing a path, the decision maker learns only the weights of those edges that belong to the chosen path. For this problem, an algorithm is given whose average cumulative loss in n rounds exceeds that of the best path, matched off-line to the entire sequence of the edge weights, by a quantity that is proportional to $1/\sqrt{n}$ and depends only *polynomially on the size of the graph.* The algorithm has linear complexity in the number of rounds n and in the number of edges. Motivated by Cognitive Packet Networks [28], an extension to the label efficient setting is also given, in which the decision maker is informed about the weights of the edges corresponding to the chosen path in only a fraction $m \ll n$ of the rounds. Another extension is shown where the decision maker competes against a time-varying path, a generalization of the problem of tracking the best expert. A version of the multi-armed bandit setting for shortest path is also discussed where the decision maker learns

only the total weight of the chosen path but not those of the individual edges on the path. This model is particularly important for routing minimizing the packet loss ratio.

In **Chapter 5** we provide a simple on-line procedure for the prediction of a stationary and ergodic processes. The proposed procedure does not only minimize the estimation error but also guarantees that the approximation error vanishes asymptotically. First a prediction strategy (algorithm) is given for *unbounded* stationary and ergodic real-valued processes and it is shown that the algorithm is universally consistent in case of the squared loss. Furthermore, we offer an extension for this setting, where the algorithm has access only to a noisy version of the original sequence. This setup was introduced and studied by Weissman and Merhav [72, 73]. We show a universally consistent algorithm in the noisy setting for convex loss functions (e.g., squared loss, absolute loss, etc.) and finally a simple universally consistent classification scheme is provided for $0-1$ loss both in the noiseless and in the noisy settings.

Chapter 2

Sequential Prediction

In this chapter the terminology and the introduction to the theory of sequential prediction are presented. The aim is to provide the reader with the necessary background material needed for this thesis.

2.1 Sequential prediction of individual sequences

The sequential (often referred also as on-line) decision problem considered in this thesis is described as follows. Suppose a decision maker has to make a sequence of actions. At each time instant $t = 1, 2, \ldots, n$, an action $a_t \in \mathcal{A}$ is made, where $\mathcal{A}$ denotes the action space and n is the number of rounds the algorithm is run for. Then, based on the state of the environment $y_t \in \mathcal{Y}$, where $\mathcal{Y}$ is some state space, the decision maker suffers some loss $\ell(a_t, y_t)$ with a nonnegative loss function $\ell : \mathcal{A} \times \mathcal{Y} \to \mathbb{R}$. In some special cases we take $\mathcal{A} = \mathcal{Y}$, but in general $\mathcal{A}$ may be different from $\mathcal{Y}$. The action at time t may depend on all previous actions $a_1, \ldots, a_{t-1}$, and on all the information available to the decision maker about the past behavior of the environment. This information, for example, may consist of the past environment states $y_1, \ldots, y_{t-1}$; however, the decision maker may not be able to observe the state y_i of the environment, where $i = 1, \ldots, t-1$. The goal of the decision maker is to minimize the average loss of the algorithm in the long run, that is, to minimize

$$\frac{1}{n} \sum_{t=1}^{n} \ell(a_t, y_t) \ ,$$

for large n. Since no probabilistic assumption is made on how the sequence $\{y_t\}$ is generated, it is not possible to minimize the cumulative loss of the algorithm

$$\widehat{L}_n \stackrel{\text{def}}{=} \sum_{t=1}^{n} \ell(a_t, y_t)$$

simultaneously for all $y_1, \ldots, y_n$ sequence.

For predicting individual sequences, a possible problem formulation is that we evaluate the performance of the algorithm with respect to a reference class of prediction rules, called experts such that the goal of the algorithm is to perform as well as the *best expert*. Formally, given N experts, at each time instant t, for every $i = 1, \ldots, N$, expert i chooses its action $f_{i,t} \in \mathcal{A}$ and suffers loss $\ell(f_{i,t}, y_t)$. The decision maker is allowed to make its own decision a_t using the experts' advice $f_{1,t}, \ldots, f_{N,t}$, however, without knowing the experts' loss in advance. Formally, the sequential prediction problem is given in Figure 2.1.

SEQUENTIAL PREDICTION PROBLEM

Parameters: number N of experts, state space $\mathcal{Y}$, action space $\mathcal{A}$, non-negative loss function $\ell : \mathcal{A} \times \mathcal{Y} \to \mathbb{R}$, number n of rounds (n can be ∞).

At time instants $t = 1, \ldots, n$,

(1) each expert forms its action $f_{i,t} \in \mathcal{A}$, $i = 1, \ldots, N$;

(2) the decision maker observes the actions of the experts and forms it own prediction $a_t \in \mathcal{A}$;

(3) the state of the environment $y_t \in \mathcal{Y}$ may or may not be revealed;

(4) the decision maker incurs loss $\ell(a_t, y_t)$ and each expert incurs loss $\ell(f_{i,t}, y_t)$.

Figure 2.1: Sequential prediction problem.

Denote the cumulative loss of expert i up to time n by

$$L_{i,n} = \sum_{t=1}^{n} \ell(f_{i,t}, t) \ .$$

Let us define the *normalized regret* as the difference between the average loss of the algorithm and that of the best expert, that is,

$$\frac{1}{n}\left(\widehat{L}_n - \min_{i=1,\ldots,N} L_{i,n}\right) \ .$$

The goal of the learning algorithm is to combine the experts' decisions such that the *normalized regret*, be universally small for all possible sequences of $\{y_t\}$.

If the action space is convex (in this case obviously an infinite action space is required), then the decision maker can combine the advice of the experts according to a distribution $\{p_{i,t}\}$ as follows:

$$a_t = \sum_{i=1}^{N} p_{i,t} f_{i,t} \ .$$

If the loss function $\ell(\cdot,\cdot)$ is convex in its first argument, then such deterministic algorithms can be applied (see e.g. Cesa-Bianchi and Lugosi [21]), which will be introduced in Subsection 2.2.2. For general action space, the combination of the experts' advice is formulated by randomization.

2.1.1 Randomized prediction

It can be shown that under general conditions on the loss function and on the *finite action space*, excluding such simple situations when, for example, the loss of the experts are the same, no deterministic algorithm can perform well for all possible sequence $\{y_t\}$. This is because for each deterministic algorithm one can construct a "bad" sequence on which the actual algorithm performs poorly, but the best expert does not. (At the end of this subsection a simple example is presented.)

Therefore, in case of finite action space we consider *randomized algorithms*. Without loss of generality we may assume that the decision maker always follows the advice of one of the experts. Let I_t be the (random) index of the expert was chosen by the algorithm at round t, that is, $a_t = f_{I_t,t}$ for some $I_t \in \{1,\dots,N\}$. Note that for each t, I_t is a random variable, as well as the cumulative loss of the randomized algorithm $\widehat{L}_n$. Therefore, we can assume that the decision of the decision maker is to choose an expert I_t and follow its decision $f_{I_t,t}$. Formally, the randomized prediction model is defined as follows:

RANDOMIZED PREDICTION WITH EXPERT ADVICE

Parameters: number N of experts, state space $\mathcal{Y}$, action space $\mathcal{A}$, non-negative loss function $\ell : \mathcal{A} \times \mathcal{Y} \to \mathbb{R}$, number n of rounds (n can be ∞). At time instants $t = 1,\dots,n$,

(1) each expert forms its action $f_{i,t} \in \mathcal{A}$, $i = 1,\dots,N$;

(2) the decision maker observes the actions of the experts and chooses an expert $I_t \in \{1,\dots,N\}$ randomly;

(3) the state of the environment $y_t \in \mathcal{Y}$ may or may not be revealed;

(4) the decision maker incurs loss $\ell(f_{I_t,t}, y_t)$ and each expert incurs loss $\ell(f_{i,t}, y_t)$.

Figure 2.2: Randomized prediction using expert advice.

For convenience we use the notations $\ell_{i,t}$ instead of $\ell(f_{i,t}, y_t)$ and $\ell_{I_t,t}$ instead of $\ell(f_{I_t,t}, y_t)$. Then the cumulative loss of the decision maker up to time n is

$$\widehat{L}_n = \sum_{t=1}^{n} \ell_{I_t,t},$$

and the cumulative loss of expert i is

$$L_{i,n} = \sum_{t=1}^{n} \ell_{i,t} \ .$$

The goal of the learning algorithm is the same like in non-randomized setting such that the *normalized regret*, that is the difference between the average loss of the algorithm and that of the best expert, be universally small for all possible sequences of $\{y_t\}$. More precisely, to ensure

$$\limsup_{n\to\infty} \frac{1}{n}\left(\widehat{L}_n - \min_{i=1,\dots,N} L_{i,n}\right) \leq 0$$

with probability 1 for every sequence $\{y_t\}$. Such an algorithm is called Hannan consistent [21].

In most of the cases we allow that the actions of the environment depend on the past choice of the decision maker and also on its own (independent) randomization; this is the so called *non-oblivious* (adaptive) adversaries.

As an example to show that deterministic algorithms do not work in general, consider the following example.

Example 2.1. Assume that we would like to predict a binary sequence and we have two different constant experts. The first one always predicts 0 and the second one always predicts 1. Formally, $f_{1,t} = 0$ and $f_{2,t} = 1$ for all $t = 1, 2, \dots$. Let the outcome sequence be $\{1, 0, 1, 0, 1, 0, 1, \dots\}$, that is $y_t = t \bmod 2$ for all $t = 1, 2, \dots$. Then the loss sequences of the experts are $\{1, 0, 1, 0, 1, 0, \dots\}$ and $\{0, 1, 0, 1, 0, 1, \dots\}$, respectively. Let the decision maker's strategy be that it always uses the advice of the expert that has been best so far. In case of tie it chooses randomly. This is the so called *follow-the-leader* strategy. This strategy chooses uniform randomly at time t if t is odd, and it chooses the second expert is chosen if t is even, resulting in choosing the worse expert. Then the average loss of the algorithm converges to 3/4, while the loss of both actions are asymptotically 1/2; thus the performance of the algorithm is far from optimal.

2.2 Algorithms

In this section we provide an overview of the most well-known algorithms in sequential decision problems. Mostly two types of algorithms are used: The so called "follow-the-perturbed-leader"-type algorithms employ the principle (with some additional randomization) that the so far best expert should perform well in the future, too, while weighted average algorithms choose experts randomly such that the ones with better past performance are chosen with higher probability. In what follows both types of algorithms are briefly introduced, but throughout the thesis we consider only weighted average type algorithms, as for these algorithms better regret bounds are available in case of partial monitoring scenarios. Throughout this section we show results in case when the losses are bounded with 1, that is $\ell_{i,t} \in [0, 1]$ for all i and t.

2.2.1 Follow-the-perturbed-leader algorithm

It was shown at the end of Subsection 2.1.1 that the follow-the-leader strategy is not optimal. However, a simple randomization suffices to achieve a significantly improved performance. The idea is to add small random perturbations to the cumulative losses and then follow the "perturbed leader" with best "perturbed" past performance. The first Hannan consistent algorithm which used this idea was given by Hannan [43], but here we show a recent version of this algorithm due to Kalai and Vempala [49].

FOLLOW-THE-PERTURBED-LEADER ALGORITHM

Parameters: Fix $R > 0$.
Initialization: Set $L_{i,0} = 0$ for $i = 1, \ldots, N$.
At time instants $t = 1, 2, \ldots$

(1) Select the random N-vector $\mathbf{Z}_t$ with components $Z_{i,t}$, $i = 1, \ldots, N$, uniformly from $[0, R]$.

(2) Select an expert

$$I_t = \arg\min_{i=1,\ldots,N}(L_{i,t-1} + Z_{i,t})$$

(ties are broken in favor of the smallest index).

(3) Update the loss of each expert i

$$L_{i,t} = L_{i,t-1} + \ell_{i,t}.$$

Figure 2.3: The follow-the-perturbed-leader algorithm in full information case.

The following theorem gives an upper bound on the normalized regret of the follow-the-perturbed-leader algorithm given in Figure 2.3 due to [49].

Theorem 2.1. *Assume $n, N \geq 1$, $0 < \delta < 1$, $\ell_{i,t} \in [0, 1]$ for all i and t , and let $R = \sqrt{nN}$. Then the follow-the-perturbed leader algorithm satisfies, with probability at least $1 - \delta$,*

$$\frac{1}{n}\left(\widehat{L}_n - \min_{i=1,\ldots,N} L_{i,n}\right) \leq 2\sqrt{\frac{N}{n}} + \sqrt{\frac{\ln(N/\delta)}{2n}}.$$

The weakness of this algorithm is that the upper bound has square-root-type dependence on the number N of experts. However, Kalai and Vempala [50] proposed a follow-the-perturbed-leader type algorithm which use exponential distribution instead of the uniform distribution to generate the perturbation and it obtains the "right" logarithmic dependence on N.

2.2.2 Exponentially weighted average prediction

In the "weighted average decision"-type algorithms at time instant t an expert i is chosen with probability that increases with the past performance of the expert. That is, $\mathbb{P}(I_t = i)$ is proportional to $r(L_{i,t-1})$, where r is a non-increasing function. The most popular choice of r is $r(x) = e^{-\eta x}$, leading to the exponentially weighted average prediction, where $\eta > 0$ is tuning parameter. In that case the probability that choosing action i at round $t \geq 2$

$$p_{i,t} = \frac{\exp(-\eta \sum_{s=1}^{t-1} \ell_{i,s})}{\sum_{j=1}^{N} \exp(-\eta \sum_{s=1}^{t-1} \ell_{j,s})} \qquad \text{for } i = 1, \ldots, N \ .$$

Formally, the algorithm for bounded losses is given in Figure 2.4.

EXPONENTIALLY WEIGHTED AVERAGE PREDICTOR

Parameters: Fix $\eta > 0$.
Initialization: Set $w_{i,0} = 1$ and $p_{i,1} = 1/N$ for $i = 1, \ldots, N$.
For each round $t = 1, 2, \ldots$

(1) Randomly select an expert $I_t \in \{1, \ldots, N\}$ according to the probability distribution $\mathbf{p}_t = (p_{1,t}, \ldots, p_{N,t})$.

(2) Update the weights $w_{i,t} = w_{i,t-1} e^{-\eta \ell_{i,t}}$.

(3) Calculate the updated probability distribution

$$p_{i,t+1} = \frac{w_{i,t}}{\sum_{j=1}^{N} w_{j,t}}, \qquad \text{for } i = 1, \ldots, N.$$

Figure 2.4: Exponentially weighted average algorithm.

The maximum difference between the cumulative loss of the above defined algorithm and cumulative loss of the best expert is $O(\sqrt{n \ln N})$ was proved by Littlestone and Warmuth [53]:

Theorem 2.2. *Let $n, N \geq 1$, $0 < \delta < 1$ and $\ell_{i,t} \in [0, 1]$. The exponentially weighted average algorithm with $\eta = \sqrt{8 \ln N / n}$ satisfies, with probability at least $1 - \delta$,*

$$\frac{1}{n} \left(\widehat{L}_n - \min_{i=1,\ldots,N} L_{i,n} \right) \leq \sqrt{\frac{\ln N}{2n}} + \sqrt{\frac{1}{2n} \ln \frac{1}{\delta}} \ .$$

If the action space is convex and loss is convex in its first argument, then we may use deterministic algorithm in non-adversary environment (see Section 2.1). That is, where the decisions of the algorithm is a convex combination of the expert advice according distribution $\mathbf{p}_t$ at time t.

Theorem 2.3. *Let $n, N \geq 1$, $\ell_{i,t} \in [0,1]$ and it is convex in its first argument then the non-randomized exponentially weighted average algorithm with $\eta = \sqrt{8 \ln N/n}$ satisfies,*

$$\frac{1}{n}\left(\sum_{t=1}^{n}\sum_{i=1}^{N} p_{i,t}\ell_{i,t} - \min_{i=1,\ldots,N} L_{i,n}\right) \leq \sqrt{\frac{\ln N}{2n}} \ .$$

Note that it is not a probabilistic statement, it holds for any sequence $y_1, y_2, \ldots, y_n$ for a fix n.

However the above regret bounds do not hold uniformly over sequences of any length n, since the parameter $\eta = \eta_n$ depends on n. In many applications, including parameter setting in TCP variants and routing in communication network the time horizon is not fixed and not available for the algorithm. To fix this problem the simplest idea is the *doubling trick* which appears in Cesa-Bianchi *et al.* [20]. The idea is to partition the time into periods of exponentially increasing length. At the beginning of each period, the algorithm chooses the optimal η for the length of the interval and when the periods end, reset the whole fixed-horizon algorithm, and the new value of η is selected optimally for the next period. This method give a $\sqrt{2}/(\sqrt{2}-1)$ multiplicative factor to the upper bound of the theorem. However, it is obvious that this method is not practical, because it resets its previously gathered knowledge time after time and therefore its application for a real problem is doubtful. Another more attractive method is that at each time instant t the algorithm chooses an $\eta = \eta_t$ which depends on t. It was proved by Auer *et al.* [7] that setting $\eta_t = \sqrt{8 \ln N/t}$ results in a regret bound that is only twice as much as the original (time dependent) bound.

2.2.3 Countably many experts

If the (infinite) action space is convex, then the decision maker can combine the advice of the expert according to a distribution $\{p_{i,t}\}$:

$$a_t = \sum_{i=1}^{N} p_{i,t} f_{i,t} \ .$$

Under convexity condition on the loss function, the regret of this combination is bounded by $O(1/\sqrt{n})$. It is easy to prove that this regret bound holds for countably many experts, too. The only necessary modification in the algorithm is that we have to define probability distribution over the set of positive integers $\{q_i : i = 1, 2, \ldots\}$, where $w_{i,0} = q_i$ represents the initial weight of expert i.

Theorem 2.4. *Under the assumptions on Theorem 2.3, for any countable class of experts, for $\ell_{i,t} \in [0,1]$ and for any probability distribution $\{q_i : i = 1, 2, \ldots\}$ over the set of positive integers, such that $q_i > 0$, the non-randomized exponentially weighted average prediction for all $n \geq 1$*

$$\frac{1}{n}\widehat{L}_n \leq \inf_{i \geq 1} \frac{1}{n}\left(L_{i,n} - \frac{1}{\eta}\ln\frac{1}{q_i}\right) + \frac{\eta}{8} \ .$$

2.3 Partial monitoring problems

In this section we overview expert algorithms for situations where the whole information on its own performance and on the past performance of the experts is not available to the decision maker. The algorithms presented here follow the idea of estimating the performance of the experts based on the available information, and then run the exponentially weighted average decision algorithm using the estimated losses. In general, the normalized regret of the algorithms can be bounded by $O\left(\sqrt{N \ln N/(nM)}\right)$ where M is the average number of experts whose performance are revealed to the decision maker at each time instant. We provide algorithms for the label efficient decision and multi-armed bandit problems.

To ease the notation throughout this section we also assume that the loss is upper bounded with 1.

2.3.1 Label efficient prediction

In the label efficient decision problem, after choosing its action at time t, the decision maker has the option to query the "label" y_t of the environment. The decision maker is allowed to make (average) m queries out of the n time instants, where $m \leq n$. To make the algorithm universal, the querying *has to be randomized*. In the sequel we will see that a simple biased coin does the job.

More precisely, to query a label, the decision maker uses an independent, identically distributed sequence $S_1, S_2, \ldots, S_n$ of Bernoulli random variables with $\mathbb{P}(S_t = 1) = \varepsilon$ and asks label y_t if $S_t = 1$. If y_t is known, the decision maker can calculate the losses $\ell_{i,t}$ for all $i = 1, \ldots, N$. If $\varepsilon = m/n$, then the number of the revealed labels during n rounds is approximately m for large n, and the proportion of labels queried converges to ε with probability 1 as n increases.

In order to apply the exponentially weighted average decision method in this case, the losses have to be substituted with its estimate. It is shown in Figure 2.5, estimated losses are used instead of the observed losses:

$$\widetilde{\ell}_{i,t} = \begin{cases} \frac{\ell_{i,t}}{\varepsilon}, & \text{if } S_t = 1, \\ 0, & \text{otherwise.} \end{cases}$$

Note that $\widetilde{\ell}_{i,t}$ is an unbiased estimate of the true loss $\ell_{i,t}$, as

$$\mathbb{E}\left[\widetilde{\ell}_{i,t} \middle| (S_1, I_1), \ldots, (S_{t-1}, I_{t-1})\right] = \ell_{i,t} \ .$$

The following upper bound on the normalized regret of algorithm in Figure 2.5 is due to Cesa-Bianchi *et al.* [22]. Note that this upper bound coincides with the previously proved upper bound for full information case if $m = n$.

Exponential weighting for label efficient prediction

Parameters: Fix $\eta > 0$ and $0 < \epsilon \leq 1$.
Initialization: Set $w_{i,0} = 1$ and $p_{i,1} = 1/N$ for $i = 1, \ldots, N$.
At time instants $t = 1, 2, \ldots$

(1) Select an action $I_t \in \{1, \ldots, N\}$ according to the probability distribution $\mathbf{p}_t = (p_{1,t}, \ldots, p_{N,t})$.

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

(3) if $S_t = 1$ then obtain $\ell_{i,t}$ for all i and compute the estimated loss

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

(4) Update the weights $w_{i,t} = w_{i,t-1} e^{-\eta \widetilde{\ell}_{i,t}}$.

(5) Calculate the updated probability distribution

$$p_{i,t+1} = \frac{w_{i,t}}{\sum_{j=1}^{N} w_{j,t}} \qquad i = 1, \ldots, N.$$

Figure 2.5: Exponentially weighted average decision algorithm in the label efficient setting.

Theorem 2.5. *Assume $n, N \geq 1$, $\ell_{i,t} \in [0,1]$ and $0 < \delta < 1$. If the above defined algorithm is run with parameters*

$$\varepsilon = \max\left\{0, \frac{m - \sqrt{2m\ln(4/\delta)}}{n}\right\} \quad \textit{and} \quad \eta = \sqrt{\frac{2\varepsilon \ln N}{n}},$$

then the normalized regret of the decision maker can be bounded with probability at least $1 - \delta$ as

$$\frac{1}{n}\left(\widehat{L}_n - \min_{i=1,\ldots,N} L_{i,n}\right) \leq 2\sqrt{\frac{\ln N}{m}} + 6\sqrt{\frac{\ln(4N/\delta)}{m}},$$

where m is the average number of the revealed labels.

2.3.2 The multi-armed bandit problem

In the multi-armed bandit problem, the decision maker learns its own loss $\ell_{I_t,t}$ after choosing an action (expert) I_t, but not the value $\ell_{i,t}$ of the other losses for $i \neq I_t$. Thus, the decision

maker does not have access to the losses it would have suffered if it had chosen a different action. The lack of information implies a natural strategy: namely, first the decision maker has to explore the losses of the experts (*exploration phase*) and then it may keep choosing the action with smallest estimated loss for the remaining time (*the exploitation phase*).

In the classical formulation of multi-armed bandit problems (see, e.g., Robbins [63]), it is assumed that, for each action, the losses are randomly and independently drawn with respect to a fixed but unknown distribution. This version is called the *stochastic multi-armed bandit problem* (for a recent efficient solution, see Auer *et al.* [4]). Here we consider a non-stochastic (or worst-case) version of this problem where the sequence $y_1, \ldots, y_n$, describing the state of the environment, is generated by a non-stochastic opponent (non-stochastic or adversarial multi-armed bandit problem) [6]. This non-stochastic approach is extremely useful in case of reactive environment e.g. in parameter setting of TCP variants, where the decision of the algorithm influences the losses (delays) of the other users, and vice versa.

There are some modifications relative to the full information case. First, the modified method uses *gains* instead of losses, defined as

$$g_{i,t} = 1 - \ell_{i,t} \ ,$$

where we used $0 \leq \ell_{i,t} \leq 1$ assumption.

Moreover, in contrast with the label efficient case, we use biased estimates of the gains defined as

$$\tilde{g}_{i,t} = \begin{cases} \frac{g_{i,t}+\beta}{p_{i,t}}, & \text{if } I_t = i, \\ \frac{\beta}{p_{i,t}}, & \text{otherwise} \end{cases}$$

where the role of parameter β is to control the bias (for $\beta = 0$ we obtain unbiased estimates of the true gains, since then $\mathbb{E}[\tilde{g}_{i,t}|I_1, I_2, \ldots, I_{t-1}] = g_{i,t}$) and we update the weights using $\tilde{g}_{i,t}$ in the following form

$$w_{i,t} = w_{i,t-1} e^{\eta \tilde{g}_{i,t}} \ .$$

Finally, a new parameter $0 < \gamma < 1$ is introduced that is used in the exploration phase: for I_{t+1} action i is chosen according to the probability

$$p_{i,t+1} = (1-\gamma)\frac{w_{i,t}}{\sum_{j=1}^{N} w_{j,t}} + \frac{\gamma}{N}, \quad i = 1, \ldots, N.$$

The role of γ is to ensure that $p_{i,t+1} \geq \gamma/N$ for all $i = 1, \ldots, N$. That is, instead of the pure probability distribution generated by exponential weighting, the decision maker uses a mixture of the exponentially weighted average distribution and the uniform distribution, where the latter allows the decision maker to constantly explore all possible actions. The resulting algorithm is given in Figure 2.6. The algorithm as well as the following bound on its performance is due to Auer *et al.* [6].

Exponential weighting in the multi-armed bandit setting

Parameters: Fix $\eta > 0$, $0 < \beta < 1$ and $0 < \gamma < 1$.
Initialization: Set $w_{i,0} = 1$ and $p_{i,1} = 1/N$ for $i = 1, \dots, N$.
At time instants $t = 1, 2, \dots$

(1) Select an action $I_t \in \{1, \dots, N\}$ according to the probability distribution $\mathbf{p}_t = (p_{1,t}, \dots, p_{N,t})$.

(2) Calculate the estimated gains

$$\tilde{g}_{i,t} = \begin{cases} \frac{g_{i,t}+\beta}{p_{i,t}}, & \text{if } I_t = i; \\ \frac{\beta}{p_{i,t}}, & \text{otherwise.} \end{cases}$$

(3) Update the weights $w_{i,t} = w_{i,t-1} e^{\eta \tilde{g}_{i,t}}$.

(4) Calculate the updated probability distribution

$$p_{i,t+1} = (1-\gamma)\frac{w_{i,t}}{\sum_{j=1}^{N} w_{j,t}} + \frac{\gamma}{N}, \quad i = 1, \dots, N.$$

Figure 2.6: Exponentially weighted average decision algorithm for the multi-armed bandit problem.

Theorem 2.6. *For any $0 < \delta < 1$, for any $\ell_{i,t} \in [0,1]$ and for any $n \geq 8N \ln(N/\delta)$, if algorithm in Figure 2.6 is run for the multi-armed bandit problem with parameters*

$$\beta = \sqrt{\frac{\ln(N/\delta)}{nN}}, \gamma = \frac{4N\beta}{3+\beta}, \text{ and } \eta = \frac{\gamma}{2N},$$

then, with probability at least $1 - \delta$,

$$\frac{1}{n}\left(\widehat{L}_n - \min_{i=1,\dots,N} L_{i,n}\right) \leq 5.5\sqrt{N \ln(N/\delta)/n} + \frac{\ln N}{2n} .$$

Note that the bound of the theorem, unlike to the full information case, grows with $\sqrt{N \ln N}$ instead of $\sqrt{\ln N}$. Hence, the bound is not really useful if the number of the experts N is large. The other disadvantages of this bound is that it holds only for bounded loss ($\ell_{i,t} \in [0,1]$), since the algorithm is defined via gains. In Chapter 4 below some recent results are presented to handle this problem for the special case when the class of the experts has some structure.

In Chapter 3 as well as in Chapter 4 we introduce a combination of the label efficient problem and the multi-armed bandit problems. The combination was motivated by the

routing problem in Cognitive Packet Networks described in Example 4.1 (in Section 4.4). In this combined problem, the decision maker learns its own loss only if it chooses to query it (which is allowed only for a limited number of times), and it cannot obtain information on the performance of any other action.

2.4 Sequential prediction in stationary and ergodic environment

In this section we focus on the setting when $y_1, y_2, \ldots$ are realizations of random variables $Y_1, Y_2, \ldots$. Under this assumption the performance of the decision maker (strategy) has a well-defined optimum, which can be achieved in full knowledge of the underlying distribution generating the outcome sequences. This property - that the loss of a strategy converges to the loss of the Bayes optimal predictor - is called *universal consistency* and it is going to define rigorously in the sequel.

At each time instant $t = 1, 2, \ldots$, the predictor is asked to guess the value of the next outcome y_t of a sequence of real numbers $y_1, y_2, \ldots$ with knowledge of the pasts $y_1^{t-1} = (y_1, \ldots, y_{t-1})$ (where y_1^0 denotes the empty string) and the side information vectors $x_1^t = (x_1, \ldots, x_t)$, where $x_t \in \mathbb{R}^d$. Thus, the predictor's estimate, at time t, is based on the value of x_1^t and y_1^{t-1}. A prediction strategy is a sequence $g = \{g_t\}_{t=1}^{\infty}$ of functions

$$g_t : \left(\mathbb{R}^d\right)^t \times \mathbb{R}^{t-1} \to \mathbb{R}$$

so that the prediction formed at time t is $g_t(x_1^t, y_1^{t-1})$.

In this section as well as in Chapter 5 we assume that $(x_1, y_1), (x_2, y_2), \ldots$ are realizations of the random variables $(X_1, Y_1), (X_2, Y_2), \ldots$ such that $\{(X_n, Y_n)\}_{-\infty}^{\infty}$ is a jointly stationary and ergodic process. Furthermore, in these parts of the thesis we use a little bit different notation for the cumulative loss, on the one hand to emphasize that here we have stronger assumptions on the outcome sequence on the other hand to suit the notations extensively used in the literature.

After n time instants, the *normalized cumulative prediction error* is

$$L_n(g) = \frac{1}{n} \sum_{t=1}^{n} \ell\left(g_t(X_1^t, Y_1^{t-1}), Y_t\right)$$

where $\ell(\cdot, \cdot)$ is a nonnegative loss function.

The fundamental limit for the predictability of the sequence can be determined based on a result of Algoet [2], who showed that for any prediction strategy g and stationary ergodic process $\{(X_n, Y_n)\}_{-\infty}^{\infty}$, in case of squared loss ($\ell(x, y) = (x - y)^2$)

$$\liminf_{n\to\infty} L_n(g) \geq L^* \quad \text{almost surely,} \tag{2.1}$$

where

$$L^* = \mathbb{E}\left[\ell\left(\mathbb{E}\left[Y_0 \middle| X_{-\infty}^0, Y_{-\infty}^{-1}\right], Y_0\right)\right]$$

is the minimal error of any prediction for the value of Y_0 based on the infinite past $X_{-\infty}^0, Y_{-\infty}^{-1}$. Note that it follows by stationarity and the martingale convergence theorem (see, e.g., Stout [67]) that

$$L^* = \lim_{n\to\infty} \mathbb{E}\left[\ell\left(\mathbb{E}\left[Y_n \middle| X_1^n, Y_1^{n-1}\right], Y_n\right)\right].$$

This lower bound gives sense to the following definition:

Definition 2.1. A prediction strategy g is called *universally consistent with respect to a class $\mathcal{C}$ of stationary and ergodic processes* $\{(X_n, Y_n)\}_{-\infty}^{\infty}$, if for each process in the class,

$$\lim_{n\to\infty} L_n(g) = L^* \quad \text{almost surely.}$$

Universally consistent strategies asymptotically achieve the best possible loss for all ergodic processes in the class. In the '90s Algoet [1] and Morvai, Yakowitz, and Györfi [57] proved that there exists a prediction strategy universal with respect to the class of all bounded ergodic processes. However, the prediction strategies exhibited in these papers are either very complex or have an unreasonably slow rate of convergence even for well-behaved processes. For square loss, Györfi and Lugosi [32] introduced several simple prediction strategies, which are universally consistent with respect to the class of bounded, stationary and ergodic processes.

Chapter 3

Hannan Consistency under Partial Monitoring for Unbounded Losses

In this chapter we analyze the sequential decision problem when the loss is unbounded under partial monitoring scenarios. We introduce a wide class of the partial monitoring problems: *the combination of the label efficient problem 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 this general setup a new algorithm (GREEN) is given and shown its Hannan consistency.

In Section 3.1 we introduce the combination of the label efficient and multi-armed bandit problems which was originally motivated by adaptive routing (in details see in Section 4.4). In Section 3.2 we define GREEN algorithm. In the next section (Theorem 3.1) we show that the expected regret of the algorithm scales with the square root of the loss of the best expert. The main result of the chapter is stated and proved in Section 3.4; it shows that Hannan consistency can be achieved, depending the growth rate of the worst expert's average loss. The above "unbounded" results can be utilized for the special problem when the loss is bounded. In Theorem 3.3 we offer an improvement for small losses in expected regret and a high-probability bound for the regret of a slightly modified algorithm (GREEN.SHIFT) is proved in Theorem 3.4.

3.1 Combination of the label efficient and multi-armed bandit problems

In this section we introduce a recent combination of the label efficient and the multi-armed bandit (LE+MAB) problems due to Ottucsák and György [61]. This combination was motivated by the routing problem in *Cognitive Packet Networks* (CPN) due to Gelenbe (Imperial College) *et al.* in [27, 28]. CPN model is implemented and integrated into Linux kernel 2.2.x and it is also the object of a US Patent (No. 6804201). CPN is described in details in Section 4.4 (Example 4.1).

In this combined problem, the decision maker learns its *own loss* only if it chooses to query it (which is allowed only for a limited number of times), and it cannot obtain

information on the performance of any other action. More precisely, for querying its loss the decision maker uses a binary sequence $S_1, S_2, \ldots$; If $S_t = 1$ then it queries its loss otherwise not. The following figure gives the precise definition of randomized prediction in case of the problem LE+MAB.

RANDOMIZED PREDICTION WITH EXPERT ADVICE IN PROBLEM LE+MAB

Parameters: number N of experts, state space $\mathcal{Y}$, action space $\mathcal{A}$, non-negative loss function $\ell : \mathcal{A} \times \mathcal{Y} \to \mathbb{R}$, number n of rounds (n can be ∞) and $\mu : \mathbb{N} \to \mathbb{N}$. At time instants $t = 1, \ldots, n$,

(1) each expert forms its action $f_{i,t} \in \mathcal{A}$, $i = 1, \ldots, N$;

(2) the decision maker observes the actions of the experts and chooses an expert $I_t \in \{1, \ldots, N\}$;

(3) the decision maker incurs loss $\ell(f_{I_t,t}, y_t)$ and each expert incurs loss $\ell(f_{i,t}, y_t)$;

(4) if $S_t = 1$ then the decision maker issues a new query to obtain its own loss $\ell(f_{I_t,t}, y_t)$; if no query is issued then $\ell(f_{I_t,t}, y_t)$ as well as the losses of the experts remain unknown.

Figure 3.1: Randomized prediction with expert advice in combination of the label efficient and the multi-armed bandit problems.

3.2 GREEN algorithm

In problem LE+MAB, it is easy to see (similarly to the LE case) that in order to achieve a nontrivial performance, the algorithm must use randomization.

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.* [22].

For problem LE+MAB we use GREEN algorithm with time-varying parameters introduced in Allenberg *et al.* [3]. GREEN algorithm is a variant of the exponentially weighted average algorithm of Littlestone and Warmuth [53] and it was named after the known idiom: "The grass is always greener on the other side", since GREEN assumes that the experts it did not choose had the best possible payoff (the zero loss).

Denote by $p_{i,t}$ the probability of choosing action i at time t in case of the original exponentially weighted average algorithm (predictor), 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 $\widetilde{L}_{i,t-1}$ is so called cumulative estimated loss, which will be updated later. GREEN algorithm 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 (see Step (2) of the algorithm) and $\gamma_t \geq 0$ is a time-varying threshold. Finally, the algorithm uses estimated losses which are given by

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

Therefore, the estimated loss is an unbiased estimate of the true loss with respect to its natural filtration, that is,

$$\mathbb{E}_t\Big[\widetilde{\ell}_{i,t}\Big] \stackrel{\text{def}}{=} \mathbb{E}\Big[\widetilde{\ell}_{i,t}\Big|(I_1,S_1),(I_2,S_2),\ldots,(I_{t-1},S_{t-1})\Big] = \ell_{i,t}\ .$$

The cumulative estimated loss of expert i is given by

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

The resulting algorithm is given in Figure 3.2.

GREEN algorithm for problem LE+MAB

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

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

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

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

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

Figure 3.2: GREEN algorithm for label efficient and multi-armed bandit problem.

3.3 Bounds on the expected regret

In this section an $O(1/\sqrt{n})$ bound is shown for the expected normalized regret of GREEN algorithm .

Theorem 3.1. (ALLENBERG, AUER, GYÖRFI AND OTTUCSÁK [3]). *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* GREEN *algorithm with $\gamma_t = 0$ and*

$\eta_t = 2\sqrt{\frac{\ln N}{N}} \cdot t^{-(1+\nu+\beta)/2}$ *is bounded by*

$$\mathbb{E}\Big[\widehat{L}_n\Big] - \min_{i=1,\dots,N} \mathbb{E}[L_{i,n}] \le 2\sqrt{(N \ln N)}(n+1)^{(1+\nu+\beta)/2}.$$

For the proofs we introduce the notations

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

and we split the statement into the following telescopes

$$\widehat{L}_n - \min_{i=1,\dots,N} L_{i,n} = \Big(\widehat{L}_n - \overline{L}_n\Big) + \Big(\overline{L}_n - \min_{i=1,\dots,N} \widetilde{L}_{i,n}\Big) + \Big(\min_{i=1,\dots,N} \widetilde{L}_{i,n} - \min_{i=1,\dots,N} L_{i,n}\Big). \quad (3.1)$$

Lemma 3.1. *For any sequence of losses* $\ell_{i,t} \ge 0$,

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

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

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

Thus

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

□

For bounding $\overline{L}_n - \min_{i=1,\dots,N} \widetilde{L}_{i,n}$ we use the following lemma due to Cesa-Bianchi *et al.* [23].

Lemma 3.2. *Consider any non-increasing sequence of* $\eta_1, \eta_2, \dots$ *positive learning rates and any nonnegative sequences* $\widetilde{\boldsymbol{\ell}}_1, \widetilde{\boldsymbol{\ell}}_2, \dots \in \mathbb{R}^N$ *of loss vectors, where* $\widetilde{\boldsymbol{\ell}}_t = (\widetilde{\ell}_{1,t}, \widetilde{\ell}_{2,t}, \dots, \widetilde{\ell}_{N,t})$. *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 $\mathbf{p}_t = (p_{1,t}, p_{2,t}, \dots, p_{N,t})$ *is the probability vector of the exponentially weighted average algorithm. Then, for* Green *algorithm*

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

Lemma 3.3. *With the notation of Lemma 3.2 we get for* GREEN *algorithm,*

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

Proof. With straightforward calculation we obtain

$$\begin{aligned}
\Phi(\mathbf{p}_t, \eta_t, -\widetilde{\boldsymbol{\ell}}_t) &= \sum_{i=1}^{N} p_{i,t} \widetilde{\ell}_{i,t} + \frac{1}{\eta_t} \ln \sum_{i=1}^{N} p_{i,t} e^{-\eta_t \widetilde{\ell}_{i,t}} \\
&\leq \sum_{i=1}^{N} p_{i,t} \widetilde{\ell}_{i,t} + \frac{1}{\eta_t} \ln \sum_{i=1}^{N} p_{i,t} \left(1 - \eta_t \widetilde{\ell}_{i,t} + \frac{\eta_t^2 \widetilde{\ell}_{i,t}^2}{2}\right) \qquad (3.2) \\
&\leq \sum_{i=1}^{N} p_{i,t} \widetilde{\ell}_{i,t} + \frac{1}{\eta_t} \ln \left(1 - \eta_t \sum_{i=1}^{N} p_{i,t} \widetilde{\ell}_{i,t} + \frac{\eta_t^2}{2} \sum_{i=1}^{N} p_{i,t} \widetilde{\ell}_{i,t}^2\right) \\
&\leq \frac{\eta_t}{2} \sum_{i=1}^{N} p_{i,t} \widetilde{\ell}_{i,t}^2 \;\leq\; \frac{\eta_t}{2\varepsilon_t} \sum_{i=1}^{N} \ell_{i,t} \widetilde{\ell}_{i,t} \qquad (3.3)
\end{aligned}$$

where (3.2) holds because of $e^{-x} \leq 1 - x + x^2/2$ for $x \geq 0$, and (3.3) follows from the fact that $\ln(1+x) \leq x$ for all $x > -1$, and from the definition of $\widetilde{\ell}_{i,t}$ in GREEN algorithm. □

Proof of Theorem 3.1. From (3.1) and Lemmas 3.1–3.3, we get

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

Note that

$$\mathbb{E}_t[\ell_{I_t,t}] = \sum_{i=1}^{N} \widetilde{p}_{i,t} \ell_{i,t} = \sum_{i=1}^{N} \widetilde{p}_{i,t} \mathbb{E}_t\left[\widetilde{\ell}_{i,t}\right] = \mathbb{E}_t\left[\check{\ell}_t\right]$$

and

$$\mathbb{E}\left[\min_{i=1,\dots,N} \widetilde{L}_{i,n}\right] \leq \min_{i=1,\dots,N} \mathbb{E}\left[\widetilde{L}_{i,n}\right] = \min_{i=1,\dots,N} \mathbb{E}[L_{i,n}],$$

then taking expectations we obtain

$$\mathbb{E}\left[\widehat{L}_n\right] - \min_{i=1,\dots,N} \mathbb{E}[L_{i,n}] \;\leq\; N \sum_{t=1}^{n} \gamma_t \mathbb{E}[\ell_{I_t,t}] + \frac{2 \ln N}{\eta_{n+1}} + \sum_{i=1}^{N} \sum_{t=1}^{n} \frac{\eta_t \mathbb{E}\left[\ell_{i,t} \widetilde{\ell}_{i,t}\right]}{2\varepsilon_t}. \qquad (3.4)$$

Now using $\mathbb{E}_t\left[\widetilde{\ell}_{i,t}\right] = \ell_{i,t}$ and assumptions of the theorem we have

$$\begin{aligned}\mathbb{E}\left[\widehat{L}_n\right] - \min_{i=1,\dots,N} \mathbb{E}[L_{i,n}] &\leq N\sum_{t=1}^{n}\gamma_t\mathbb{E}[\ell_{I_t,t}] + \frac{2\ln N}{\eta_{n+1}} + \sum_{i=1}^{N}\sum_{t=1}^{n}\frac{\eta_t\mathbb{E}\left[\ell_{i,t}^2\right]}{2\varepsilon_t} \\ &\leq \sqrt{N\ln N}(n+1)^{(1+\nu+\beta)/2} + \sqrt{N\ln N}\sum_{t=1}^{n} t^{(-1+\nu+\beta)/2}\end{aligned}$$

as desired. □

3.4 Hannan consistency

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

Theorem 3.2. (ALLENBERG, AUER, GYÖRFI AND OTTUCSÁK [3]). *Algorithm* GREEN *is run for the combination of the label efficient and multi-armed bandit problem. Assume that there exist universal constants $c < \infty$ and $0 \leq \nu < 1$ such that for each n*

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

For some constant $\rho > 0$ choose the parameters of the algorithm as:

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

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

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 GREEN *algorithm is Hannan consistent, that is,*

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

Remark 3.1. We derive the consequences of the theorem in special cases:

- **Full information**: With a slight modification of the proof and fixing $\beta = 0$ ($\varepsilon_t = 1$) and $\gamma_t = 0$ we get the following condition for the losses in full information case:

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

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

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

- **Label efficient setting with time-varying query rate (ε_t)**: With a modification of the proof and fixing $\gamma_t = 0$ we get the following condition for the loss function in label efficient case:

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

- **Combination of the label efficient and multi-armed bandit setting**: This is the most general case. Let $\gamma_t = t^{-1/3}$. Then the bound is

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

Remark 3.2. (Convergence rate) With an extension of Lemma 3.4 below we can retrieve the ν dependent almost sure convergence rate of the algorithm. The rate is

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

in the full information and the label efficient cases with optimal choice of the parameters and in the multi-armed bandit and "combined" cases it is

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

Remark 3.3. (Minimum amount of query rate in label efficient setting) 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 (unknown) bound, i.e., $\nu = 0$. With a slight modification of the proof of Theorem 3.2 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 Cesa-Bianchi *et al.* [22].

In order to prove Theorem 3.2, we split the proof into three lemmas by telescope:

$$\begin{aligned}\frac{1}{n}\widehat{L}_n - \frac{1}{n}\min_{i=1,\dots,N} L_{i,n} \\ = \underbrace{\frac{1}{n}\left(\widehat{L}_n - \overline{L}_n\right)}_{\text{Lemma 3.5}} + \underbrace{\frac{1}{n}\left(\overline{L}_n - \min_{i=1,\dots,N}\widetilde{L}_{i,n}\right)}_{\text{Lemma 3.6}} + \underbrace{\frac{1}{n}\left(\min_{i=1,\dots,N}\widetilde{L}_{i,n} - \min_{i=1,\dots,N} L_{i,n}\right)}_{\text{Lemma 3.7}}. \quad (3.5)\end{aligned}$$

Combining sequentially Lemma 3.5, Lemma 3.6 and Lemma 3.7 Theorem 3.2 is proved. We will show separately the almost sure convergence of the three lemmas on the right-hand side. In the sequel, we need the following lemma which is the key of the proof of Theorem 3.2:

Lemma 3.4. *Let $\{Z_t\}$ be a martingale difference sequence. Let*

$$h_t \mathbb{E}[k_t] \geq \mathbf{Var}(Z_t)$$

where

$$h_t = 1/t^a$$

for all $t = 1, 2, \dots$ and

$$K_n = \frac{1}{n}\sum_{t=1}^{n} k_t \leq Cn^b$$

and $0 \leq b < 1$ 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 [24], if $\{Z_t\}$ a martingale difference sequence with

$$\sum_{t=1}^{\infty} \frac{\mathbf{Var}(Z_t)}{t^2} < \infty \quad (3.6)$$

then

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

We have to verify (3.6). Because of $k_t = tK_t - (t-1)K_{t-1}$, and $\frac{h_t}{t} - \frac{h_{t+1}t}{(t+1)^2} \geq 0$ we have that

$$\begin{aligned}\sum_{t=1}^{n} \frac{\mathbf{Var}(Z_t)}{t^2} \leq \sum_{t=1}^{n} \frac{h_t \mathbb{E}[k_t]}{t^2} &= \sum_{t=1}^{n} \frac{h_t \mathbb{E}[(tK_t - (t-1)K_{t-1})]}{t^2} \\ &= \frac{h_n \mathbb{E}[K_n]}{n} + \sum_{t=1}^{n-1} \left(\frac{h_t}{t} - \frac{h_{t+1}t}{(t+1)^2} \right) \mathbb{E}[K_t] \\ &\leq \frac{n^{-a} C n^b}{n} + \sum_{t=1}^{n-1} \left(\frac{t^{-a}}{t} - \frac{(t+1)^{-a} t}{(t+1)^2} \right) C t^b\end{aligned}$$

which is bounded by conditions. □

Now we are ready to prove one by one the almost sure convergence of the terms in (3.5).

Lemma 3.5. *Under the conditions of the Theorem 3.2,*

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

Proof. First we use Lemma 3.1, that is

$$\widehat{L}_n - \overline{L}_n \;\leq\; \sum_{t=1}^{n} \left(\ell_{I_t,t} - \check{\ell}_t \right) + \sum_{t=1}^{n} N\gamma_t \check{\ell}_t = \sum_{t=1}^{n} Z_t + \sum_{t=1}^{n} N\gamma_t \check{\ell}_t. \tag{3.7}$$

Below we show separately, that both sums in (3.7) divided by n converge to zero almost surely. First observe that $\{Z_t\}$ is a martingale difference sequence with respect to $(I_1, S_1), \ldots, (I_{t-1}, S_{t-1})$. Observe that I_t is independent from S_t therefore we get the following bound for the variance of Z_t:

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

where $h_t = 1/\varepsilon_t$ and $k_t = \sum_{i=1}^{N} \ell_{i,t}^2$. Then applying Lemma 3.4 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 (3.7) divided by n goes to zero almost surely, that is,

$$\frac{1}{n} \sum_{t=1}^{n} N\gamma_t \check{\ell}_t = \frac{1}{n} \sum_{t=1}^{n} \frac{S_t}{\varepsilon_t} \ell_{I_t,t} N\gamma_t = \frac{1}{n} \sum_{t=1}^{n} R_t + \frac{1}{n} \sum_{t=1}^{n} \ell_{I_t,t} N\gamma_t \to 0 \quad (n \to \infty) \tag{3.8}$$

where R_t is a martingale difference sequence respect to $(I_1, S_1), \ldots, (I_{t-1}, S_{t-1})$. Bounding the variance of R_t, we obtain

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

Then using Lemma 3.4 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 (3.8) 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

$$\begin{aligned}
\frac{1}{n} \sum_{t=1}^{n} \ell_{i,t} \gamma_t &= \frac{1}{n} \sum_{t=1}^{n} (tK_{i,t} - (t-1)K_{i,t-1}) \gamma_t \\
&= K_{i,n}\gamma_n + \frac{1}{n} \sum_{t=1}^{n-1} (\gamma_t - \gamma_{t+1})\, tK_{i,t} \\
&\leq K_{i,n}\gamma_n + \frac{1}{n} \sum_{t=1}^{n-1} \gamma_t K_{i,t} \qquad (3.9) \\
&\leq \sqrt{c} \frac{1}{N} n^{\nu/2-\alpha} + \frac{1}{nN} \sum_{t=1}^{n-1} t^{\nu/2-\alpha} \sqrt{c} \to 0 \qquad (3.10)
\end{aligned}$$

where the (3.9) holds because $(\gamma_t - \gamma_{t+1})t \leq \gamma_t$ and (3.10) follows from $K_{i,n} \leq \sqrt{cn^\nu}$, the definition of the parameters and $\alpha \geq (\nu + \rho)/2$. □

Lemma 3.6 yields the relation between $\overline{L}_n$ and $\min_{i=1,\ldots,N} \widetilde{L}_{i,n}$.

Lemma 3.6. *Under the conditions of Theorem 3.2,*

$$\limsup_{n\to\infty} \frac{1}{n} \left(\overline{L}_n - \min_{i=1,\ldots,N} \widetilde{L}_{i,n}\right) \leq 0 \qquad a.s.$$

Proof. We start by applying Lemma 3.2, that is,

$$\overline{L}_n - \min_{i=1,\ldots,N} \widetilde{L}_{i,n} \leq \frac{2 \ln N}{\eta_{n+1}} + \sum_{t=1}^{n} \Phi(\mathbf{p}_t, \eta_t, -\widetilde{\boldsymbol{\ell}}_t). \qquad (3.11)$$

To bound the quantity of $\Phi(\mathbf{p}_t, \eta_t, -\widetilde{\boldsymbol{\ell}}_t)$, our starting point is (3.3). Moreover,

$$\frac{\eta_t}{2}\sum_{i=1}^{N} p_{i,t}\widetilde{\ell}_{i,t}^2 = \frac{\eta_t}{2}\sum_{i=1}^{N} p_{i,t}\frac{\ell_{i,t}^2}{\widetilde{p}_{i,t}^2\varepsilon_t^2} S_t \mathbb{I}_{\{I_t=i\}} \leq \frac{\eta_t}{2\gamma_t\varepsilon_t}\frac{S_t}{\varepsilon_t}\ell_{I_t,t}^2 \leq \frac{\eta_t}{2\gamma_t\varepsilon_t}\frac{S_t}{\varepsilon_t}\sum_{i=1}^{N}\ell_{i,t}^2 \tag{3.12}$$

where the first inequality comes from $p_{I_t,t} \geq \gamma_t$. Combining this bound with (3.11), dividing by n and taking the limit superior we get

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

Let analyze separately the two terms on the right-hand side. The first term is zero because of the assumption of the Theorem 3.2. Concerning the second term, similarly to Lemma 3.5 we can split S_t/ε_t as follows: let us

$$\frac{S_t}{\varepsilon_t}\frac{\eta_t}{2\gamma_t\varepsilon_t}\sum_{i=1}^{N}\ell_{i,t}^2 = Z_t + \frac{\eta_t}{2\gamma_t\varepsilon_t}\sum_{i=1}^{N}\ell_{i,t}^2, \tag{3.13}$$

where Z_t is a martingale difference sequence. The variance is

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

Application of Lemma 3.4 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

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

Finally, we have to prove that the sum of the second term in (3.13) 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 3.5. Introduce $K_{i,n} = \frac{1}{n}\sum_{t=1}^{n}\ell_{i,t}^2$ then we get

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

because of $K_{i,n} \leq cn^{\nu}$ 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 3.7. *Under the conditions of Theorem 3.2,*

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

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

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

where $Z_{i,t}$ is martingale difference sequence for all i. As earlier, we use Lemma 3.4. First we bound $\mathbf{Var}(Z_{i,t})$ as follows

$$\mathbf{Var}(Z_{i,t}) = \mathbb{E}\widetilde{\ell}_{i,t}^2 \leq \frac{\mathbb{E}\left[\sum_{i=1}^{N} \ell_{i,t}^2\right]}{\varepsilon_t \gamma_t}. \tag{3.14}$$

Applying Lemma 3.4 with parameters $k_t = \sum_{i=1}^{N} \ell_{i,t}^2$ and $h_t = \frac{1}{\varepsilon_t\gamma_t}$, for each i

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

therefore

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

□

3.5 Bounded loss

If the individual losses are bounded by a constant, much stronger results can be obtained for GREEN algorithm. On the one hand, we give an improvement for small losses for expected regret. On the other hand, $O(1/\sqrt{n})$ regret bound is shown for high-probability regret.

Theorem 3.3. (Allenberg, Auer, Györfi and Ottucsák [3]). *If $\ell_{i,t} \in [0,1]$ and $\varepsilon_t = \varepsilon$ for all t, then for all n with $\min_{i=1,\dots,N} L_{i,n} \leq B$ the expected loss of* Green *algorithm 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*

$$\mathbb{E}\left[\widehat{L}_n\right] - \min_{i=1,\dots,N} \mathbb{E}[L_{i,n}] \leq 4\sqrt{\frac{B}{\varepsilon} N \ln N} + \frac{N\ln N + 2}{\varepsilon} + \frac{N\ln(\varepsilon B + 1)}{\varepsilon}.$$

Remark 3.4. The improvement in Theorem 3.3 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 3.1 is void for $\min_{i=1,\dots,N} L_{i,n} \ll \sqrt{n}$ whereas Theorem 3.3 still gives a nearly optimal bound[1].

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

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

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

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

and since $\widetilde{L}_{i,n} = \widetilde{L}_{i,T_i}$ we get that $\widetilde{L}_{i,n} \leq \widetilde{L}_{i^*,n} + \frac{\ln(1/\gamma)}{\eta}$. Plugging this bound into (3.4) and using $\ell_{i,t} \in [0,1]$ we get

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

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

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

For $\gamma = \frac{1}{N(\varepsilon B+2)}$ we have $\frac{\min_i \mathbb{E}[L_{i,n}]}{1-\gamma N} \leq \min_i \mathbb{E}[L_{i,n}] + \frac{1}{\varepsilon}$ and $\frac{1}{1-\gamma N} \leq 2$, which implies

$$\mathbb{E}\left[\widehat{L}_n\right] \leq \min_{i=1,\dots,N} \mathbb{E}[L_{i,n}] + \frac{1}{\varepsilon} + \frac{4\ln N}{\eta} + N\frac{\eta}{\varepsilon}\left(\mathbb{E}[L_{i^*,n}] + \frac{\ln N}{\eta} + \frac{\ln(\varepsilon B + 2)}{\eta}\right)$$

and, by simple calculation, the statement of the theorem. □

In the rest of this section we introduce a slightly modified version of Green algorithm for multi-armed bandit problem, so called Green.Shift. One can easily extend

[1] For $\varepsilon = 1$ optimality follows from the lower bound on the regret in [6].

the Green.Shift algorithm for problem LE+MAB based on Section 4.4. The proposed algorithm is a "shifted" version of Green algorithm .

As earlier let $\widetilde{\ell}_{i,t}$ denote the conditional unbiased estimation of the true loss of each action with respect to its natural filtration. Instead of the unbiased estimate, a slightly smaller quantity is used by the algorithm. The (biased) estimated loss is

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

where β is a positive parameter and the maximum is necessary to avoid dividing 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 3.3.

Theorem 3.4. (Auer and Ottucsák [8]). *For any* $0 < \delta < 1$ *and parameters*

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

the performance of Green.Shift *algorithm can be bounded with probability at least* $1 - \delta$ *as*

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

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

$$\frac{1}{n}\left(\widehat{L}_n - \min_{i=1,\dots,N} L_{i,n}\right) \le 7\sqrt{N\ln(N/\delta)/n} + \frac{1}{2n} N\ln(nN) .$$

Green.Shift algorithm for multi-armed bandit problem

Parameters: 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

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

(2) Calculate the probability distribution

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

(3) Calculate the modified probabilities

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

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

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

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

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

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

Figure 3.3: Green.Shift algorithm for multi-armed bandit problem.

For the proof of the theorem we need the following 2 lemmas. The first lemma is a simple modification of [21, Lemma 6.7].

Lemma 3.8. *Under the assumptions of Theorem 3.4 for any* $0 < \delta < 1$ *we have*

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

Proof. For any $u > 0$ and $c > 0$ the Chernoff bounding technique (see, e.g., [25]) implies

$$\mathbb{P}\big(L'_{i,n} > L_{i,n} + u\big) \le e^{-cu}\mathbb{E}e^{c(L'_{i,n}-L_{i,n})} \ . \tag{3.15}$$

Letting $u = \beta nN$ and $c = \beta$, therefore from (3.15):

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

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

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

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

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

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

$$\beta\left(\frac{\ell_{i,t}\mathbb{I}_{\{I_t=i\}}}{\widetilde{p}_{i,t}} - \frac{\beta}{\max\{\widetilde{p}_{i,t},\gamma_t\}} - \ell_{i,t}\right) \le \frac{\beta\ell_{I_t,t}}{\widetilde{p}_{I_t,t}} \le \frac{\beta\ell_{I_t,t}}{\gamma} \le 1$$

where the second inequality comes from $\beta \le \gamma$. Let $\mathbb{E}_t[Z_t] = \mathbb{E}[Z_t|Z_{t-1}, \ldots, Z_1]$ and using $e^x \le 1 + x + x^2$ for $x \le 1$ we have

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

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

The following lemma is a variant of Theorem 3.3.

Lemma 3.9. *Under the assumptions of Theorem 3.4 for the cumulative estimated loss we have*

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

Proof. Let $T_i = \max\{0 \le t \le n : p_{i,t} \ge \gamma\}$ be the last round where $\widetilde{p}_{i,t} > 0$. Therefore,

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

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

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

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

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

□

Proof of Theorem 3.4. For the proof of theorem the quantity of $\ln \frac{W_n}{W_0}$ is bounded, where

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

The lower bound is

$$\ln \frac{W_n}{W_0} = \ln \left(\sum_{i=1}^{N} e^{-\eta L'_{i,n}} \right) - \ln N \ge \ln \left(\max_{i=1,\ldots,N} e^{-\eta L'_{i,n}} \right) - \ln N = -\eta \min_{i=1,\ldots,N} L'_{i,n} - \ln N \ . \tag{3.16}$$

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

$$\ln \frac{W_t}{W_{t-1}} = \ln \sum_{i=1}^{N} p_{i,t} e^{-\eta \ell'_{i,t}} \le \ln \sum_{i=1}^{N} p_{i,t} \left(1 - \eta \ell'_{i,t} + \eta^2 \ell'_{i,t}\right) \le -\eta \sum_{i=1}^{N} p_{i,t} \ell'_{i,t} + \eta^2 \sum_{i=1}^{N} p_{i,t} \ell'^2_{i,t} \ . \tag{3.17}$$

Next we bound the sums in (3.17). On the one hand,

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

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

On the other hand,

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

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

$$\ln\frac{W_n}{W_0} \leq -\eta\widehat{L}_n + N\eta\gamma\widehat{L}_n + \eta\beta nN + \eta^2\textstyle\sum_{i=1}^{N} L_{i,n}' + \eta^2 2N\ . \tag{3.18}$$

Plug the results of Lemma 3.9 into (3.18) we get

$$\ln\frac{W_n}{W_0} \leq -\eta\widehat{L}_n + N\eta\gamma\widehat{L}_n + \eta\beta nN + \eta^2 N\min_{i=1,\ldots,N} L_{i,n}' + \eta N\ln(1/\gamma) + \eta^2 2N\ . \tag{3.19}$$

Combining (3.16) and (3.19) we obtain

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

By Lemma 3.8 and the union bound we have at least $1-\delta$

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

as desired. □

Chapter 4

Shortest Path Problem under Partial Monitoring

As mentioned before, the basic theoretical results of sequential decision problem were pioneered by Blackwell [15] and Hannan [43], and brought to the attention of the machine learning community in the 1990's by Vovk [70], Littlestone and Warmuth [53], and Cesa-Bianchi *et al.* [20]. These results show that for any bounded loss function, if the decision maker has access to the past losses of all experts, then it is possible to construct on-line algorithms that perform, for any possible behavior of the environment, almost as well as the best of N experts. More precisely, recalling the results are presented in Chapter 2, the per round cumulative loss of these algorithms is at most as large as that of the best expert plus a quantity proportional to $\sqrt{\ln N/n}$ for any bounded loss function, where n is the number of rounds in the decision game. The logarithmic dependence on the number of experts makes it possible to obtain meaningful bounds even if the pool of experts is very large. However, the basic prediction algorithms, such as exponentially weighted average forecasters, have a computational complexity that is proportional to the number of experts, and they are therefore practically infeasible when the number of experts is very large.

As it is described in details in Section 2.3 in certain situations the decision maker has only limited knowledge about the losses of all possible actions. For example, it is often natural to assume that the decision maker gets to know only the loss corresponding to the action it has made, and has no information about the loss it would have suffered had it made a different decision. This setup is referred to as the *multi-armed bandit problem*, and was considered, in the adversarial setting, by Auer *et al.* [5] who gave an algorithm whose normalized regret (the difference of the algorithm's average loss and that of the best expert) is upper bounded by a quantity which is proportional to $\sqrt{N \ln N/n}$. Note that, compared to the *full information* case described above where the losses of all possible actions are revealed to the decision maker, there is an extra $\sqrt{N}$ factor in the performance bound, which seriously limits the usefulness of the bound if the number of experts is large.

Another interesting example for the limited information case is the so-called *label efficient decision problem* (see Helmbold and Panizza [45]) in which it is too costly to observe the state of the environment, and so the decision maker can query the losses of all possible

actions for only a limited number of times. A recent result of Cesa-Bianchi, Lugosi, and Stoltz [22] shows that in this case, if the decision maker can query the losses m times during a period of length n, then it can achieve $O(\sqrt{\ln N/m})$ normalized regret relative to the best expert.

In many applications the set of experts has a certain structure that may be exploited to construct efficient on-line decision algorithms. The construction of such algorithms has been of great interest in computational learning theory. A partial list of works dealing with this problem includes Herbster and Warmuth [46], Vovk [71], Bousquet and Warmuth [17], Helmbold and Schapire [64], Takimoto and Warmuth [69], Kalai and Vempala [49], György *et al.* [36, 37, 38]. For a more complete survey, we refer to Cesa-Bianchi and Lugosi [21, Chapter 5].

In this chapter we study the on-line shortest path problem, a representative example of structured expert classes that has received attention in the literature for its many applications, including, among others, routing in communication networks; see, e.g., Takimoto and Warmuth [69], Awerbuch *et al.* [10], or György and Ottucsák [42], and adaptive quantizer design in zero-delay lossy source coding; see, György *et al.* [36, 37, 39]. In this problem, a weighted directed (acyclic) graph is given whose edge weights can change in an arbitrary manner, and the decision maker has to pick in each round a path between two given vertices, such that the weight of this path (the sum of the weights of its composing edges) be as small as possible.

Efficient solutions, with time and space complexity proportional to the number of edges rather than to the number of paths (the latter typically being exponential in the number of edges), have been given in the full information case, where in each round the weights of all the edges are revealed after a path has been chosen; see, for example, Mohri [55], Takimoto and Warmuth [69], Kalai and Vempala [49], and György *et al.* [38].

In the bandit setting only the weights of the edges or just the sum of the weights of the edges composing the chosen path are revealed to the decision maker. If one applies the general bandit algorithm of Auer *et al.* [5], the resulting bound will be too large to be of practical use because of its square-root-type dependence on the number of paths N. On the other hand, using the special graph structure in the problem, Awerbuch and Kleinberg [11] and McMahan and Blum [54] managed to get rid of the exponential dependence on the number of edges in the performance bound. They achieved this by extending the exponentially weighted average predictor and the follow-the-perturbed-leader algorithm of Hannan [43] to the generalization of the multi-armed bandit setting for shortest paths, when only the sum of the weights of the edges is available for the algorithm. However, the dependence of the bounds obtained in [11] and [54] on the number of rounds n is significantly worse than the $O(1/\sqrt{n})$ bound of Auer *et al.* [5]. Awerbuch and Kleinberg [11] consider the model of "non-oblivious" adversaries for shortest path (i.e., the losses assigned to the edges can depend on the previous actions of the forecaster) and prove an $O(n^{-1/3})$ bound for the expected normalized regret. McMahan and Blum [54] give a simpler algorithm than in [11] however obtain a bound of the order of $O(n^{-1/4})$ for the expected regret.

In this chapter we provide an extension of the bandit algorithm of Auer *et al.* [5] unifying

the advantages of the above approaches, with a performance bound that is polynomial in the number of edges, and converges to zero at the right $O(1/\sqrt{n})$ rate as the number of rounds increases. We achieve this bound in a model which assumes that the losses of all edges on the path chosen by the forecaster are available separately after making the decision. We also discuss the case (considered by [11] and [54]) in which only the total loss (i.e., the sum of the losses on the chosen path) is known to the decision maker. We exhibit a simple algorithm which achieves an $O(n^{-1/3})$ normalized regret *with high probability* against "non-oblivious" adversary. In this case it remains an open problem to find an algorithm whose cumulative loss is polynomial in the number of edges of the graph and decreases as $O(n^{-1/2})$ with the number of rounds. Throughout the chapter we assume that the number of rounds n in the prediction game is known in advance to the decision maker.

In Section 4.1 we formally define the on-line shortest path problem, which is extended to the multi-armed bandit setting in Section 4.2. Our new algorithm for the shortest path problem in the bandit setting is given in Section 4.3 together with its performance analysis. The algorithm is extended to solve the shortest path problem in a combined label efficient and multi-armed bandit setting in Section 4.4. Another extension, when the algorithm competes against a time-varying path is studied in Section 4.5. An algorithm for the "restricted" multi-armed bandit setting (when only the sums of the losses of the edges are available) is given in Section 4.6. Simulation results are presented in Section 4.7.

4.1 The shortest path problem

Consider a network represented by a set of vertices connected by edges, and assume that we have to send a stream of packets from a distinguished vertex, called *source*, to another distinguished vertex, called *destination*. At each time slot a packet is sent along a chosen route (path) connecting source and destination. Depending on the traffic, each edge in the network may have a different delay, and the total delay the packet suffers on the chosen path is the sum of delays of the edges composing the route. The delays may change from one time slot to the next one in an arbitrary way, and our goal is to find a way of choosing the path in each time slot such that the sum of the total delays over time is not significantly more than that of the best fixed path in the network. This adversarial version of the routing problem is most useful when the delays on the edges can change dynamically, even depending on our previous routing decisions. This is the situation in the case of ad-hoc networks, where the network topology can change rapidly, or in certain secure networks, where the algorithm has to be prepared to handle denial of service attacks, that is, situations where willingly malfunctioning vertices and links increase the delay; see, e.g., Awerbuch *et al.* [10].

This problem can be cast naturally as a sequential decision problem in which each possible path is represented by an action (expert). However, the number of paths is typically exponentially large in the number of edges, and therefore computationally efficient algorithms are called for. Two solutions of different flavor have been proposed. One of them

is based on a follow-the-perturbed-leader forecaster, see Kalai and Vempala [49], while the other is based on an efficient computation of the exponentially weighted average forecaster, see, for example, Takimoto and Warmuth [69]. Both solutions have different advantages and may be generalized in different directions.

To formalize the problem, consider a (finite) directed acyclic graph with a set of edges $E = \{e_1, \ldots, e_{|E|}\}$ and a set of vertices V. Thus, each edge $e \in E$ is an ordered pair of vertices (v_1, v_2). Let u and v be two distinguished vertices in V. A *path* from u to v is a sequence of edges $e^{(1)}, \ldots, e^{(k)}$ such that $e^{(1)} = (u, v_1)$, $e^{(j)} = (v_{j-1}, v_j)$ for all $j = 2, \ldots, k-1$, and $e^{(k)} = (v_{k-1}, v)$. Let $\mathcal{P} = \{\boldsymbol{i}_1, \ldots, \boldsymbol{i}_N\}$ denote the set of all such paths. For simplicity, we assume that every edge in E is on some path from u to v and every vertex in V is an endpoint of an edge (see Figure 4.1 for examples).

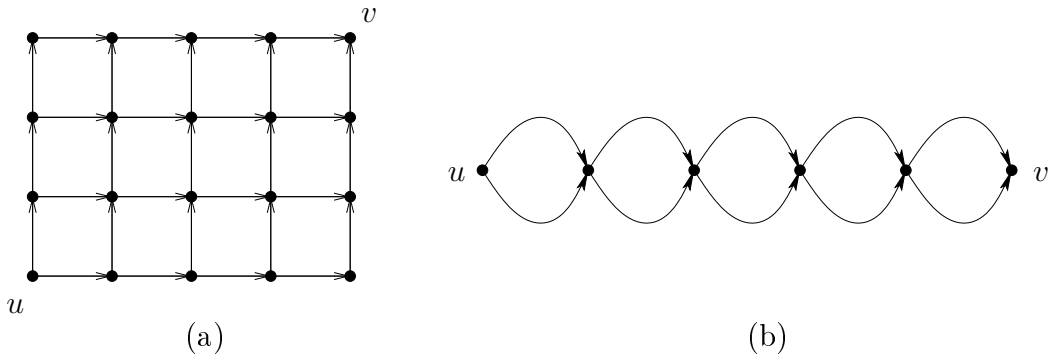

Figure 4.1: Two examples of directed acyclic graphs for the shortest path problem.

In each round $t = 1, \ldots, n$ of the decision game, the decision maker chooses a path $\boldsymbol{I}_t$ among all paths from u to v. Then a loss $\ell_{e,t} \in [0, 1]$ is assigned to each edge $e \in E$. We write $e \in \boldsymbol{i}$ if the edge $e \in E$ belongs to the path $\boldsymbol{i} \in \mathcal{P}$, and with a slight abuse of notation the loss of a path $\boldsymbol{i}$ at time slot t is also represented by $\ell_{\boldsymbol{i},t}$. Then $\ell_{\boldsymbol{i},t}$ is given as

$$\ell_{\boldsymbol{i},t} = \sum_{e \in \boldsymbol{i}} \ell_{e,t}$$

and therefore the cumulative loss up to time t of each path $\boldsymbol{i}$ takes the additive form

$$L_{\boldsymbol{i},t} = \sum_{s=1}^{t} \ell_{\boldsymbol{i},s} = \sum_{e \in \boldsymbol{i}} \sum_{s=1}^{t} \ell_{e,s}$$

where the inner sum on the right-hand side is the loss accumulated by edge e during the first t rounds of the game. The cumulative loss of the algorithm is

$$\widehat{L}_t = \sum_{s=1}^{t} \ell_{\boldsymbol{I}_s,s} = \sum_{s=1}^{t} \sum_{e \in \boldsymbol{I}_s} \ell_{e,s} \ .$$

It is well known that for a general loss sequence, the decision maker must be allowed to use randomization to be able to approximate the performance of the best expert (see, e.g., Cesa-Bianchi and Lugosi [21]). Therefore, the path $\boldsymbol{I}_t$ is chosen randomly according to some distribution $\mathbf{p}_t$ over all paths from u to v. We study the normalized regret over n rounds of the game

$$\frac{1}{n}\left(\widehat{L}_n - \min_{\boldsymbol{i}\in\mathcal{P}} L_{\boldsymbol{i},n}\right)$$

where the minimum is taken over all paths $\boldsymbol{i}$ from u to v.

In the full information case, for example, the exponentially weighted average forecaster ([70], [53], [20]), calculated over all possible paths, has regret

$$\frac{1}{n}\left(\widehat{L}_n - \min_{\boldsymbol{i}\in\mathcal{P}} L_{\boldsymbol{i},n}\right) \le K\left(\sqrt{\frac{\ln N}{2n}} + \sqrt{\frac{\ln(1/\delta)}{2n}}\right)$$

with probability at least $1-\delta$, where N is the total number of paths from u to v in the graph and K is the length of the longest path.

4.2 The multi-armed bandit setting

In this section we discuss the "bandit" version of the shortest path problem. In this setup, which is more realistic in many applications, the decision maker has only access to the losses corresponding to the paths it has chosen. For example, in the routing problem this means that information is available on the delay of the path the packet is sent on, and not on other paths in the network.

We distinguish between two types of bandit problems, both of which are natural generalizations of the simple bandit problem to the shortest path problem. In the first variant, the decision maker has access to the losses of those edges that are on the path it has chosen. That is, after choosing a path $\boldsymbol{I}_t$ at time t, the value of the loss $\ell_{e,t}$ is revealed to the decision maker if and only if $e \in \boldsymbol{I}_t$. We study this case and its extensions in Sections 4.3, 4.4, and 4.5.

The second variant is a more restricted version in which the loss of the chosen path is observed, but no information is available on the individual losses of the edges belonging to the path. That is, after choosing a path $\boldsymbol{I}_t$ at time t, only the value of the loss of the path $\ell_{\boldsymbol{I}_t,t}$ is revealed to the decision maker. Further on we call this setting as the *restricted* bandit problem for shortest path. We consider this restricted problem in Section 4.6.

Formally, the on-line shortest path problem in the multi-armed bandit setting is described as follows: at each time instance $t = 1, \ldots, n$, the decision maker picks a path $\boldsymbol{I}_t \in \mathcal{P}$ from u to v. Then the environment assigns loss $\ell_{e,t} \in [0,1]$ to each edge $e \in E$, and the decision maker suffers loss $\ell_{\boldsymbol{I}_t,t} = \sum_{e\in\boldsymbol{I}_t} \ell_{e,t}$. In the unrestricted case the losses $\ell_{e,t}$ are revealed for all $e \in \boldsymbol{I}_t$, while in the restricted case only $\ell_{\boldsymbol{I}_t,t}$ is revealed. Note that in both cases $\ell_{e,t}$ may depend on $\boldsymbol{I}_1, \ldots, \boldsymbol{I}_{t-1}$, the earlier choices of the decision maker.

For the basic multi-armed bandit problem, Auer *et al.* [5] gave an algorithm, based on exponential weighting with a biased estimate of the gains combined with uniform exploration. Applying their algorithm to the on-line shortest path problem in the bandit setting results in a performance that can be bounded, for any $0 < \delta < 1$ and fixed time horizon n, with probability at least $1 - \delta$, by

$$\frac{1}{n}\left(\widehat{L}_n - \min_{\boldsymbol{i}\in\mathcal{P}} L_{\boldsymbol{i},n}\right) \leq \frac{11K}{2}\sqrt{\frac{N\ln(N/\delta)}{n}} + \frac{K\ln N}{2n} \ .$$

(The constants follow from a slightly improved version; see Cesa-Bianchi and Lugosi [21].)

However, for the shortest path problem this bound is unacceptably large because, unlike in the full information case, here the dependence on the number of all paths N is not merely logarithmic, while N is typically exponentially large in the size of the graph (as in the two simple examples of Figure 4.1). Note that this bound also holds for the restricted setting as only the total losses on the paths are used. In order to achieve a bound that does not grow exponentially with the number of edges of the graph, it is imperative to make use of the dependence structure of the losses of the different actions (i.e., paths). Awerbuch and Kleinberg [11] and McMahan and Blum [54] do this by extending low complexity predictors, such as the follow-the-perturbed-leader forecaster [43], [49] to the restricted bandit setting. However, in both cases the price to pay for the polynomial dependence on the number of edges is a worse dependence on the length n of the game.

4.3 A bandit algorithm for shortest paths

In this section we describe a variant of the bandit algorithm of [5] which achieves the desired performance for the shortest path problem. The new algorithm uses the fact that when the losses of the edges of the chosen path are revealed, then this also provides some information about the losses of each path sharing common edges with the chosen path.

For each edge $e \in E$, and $t = 1, 2, \ldots$, introduce the *gain* $g_{e,t} = 1 - \ell_{e,t}$, and for each path $\boldsymbol{i} \in \mathcal{P}$, let the gain be the sum of the gains of the edges on the path, that is,

$$g_{\boldsymbol{i},t} = \sum_{e\in\boldsymbol{i}} g_{e,t} \ .$$

The conversion from losses to gains is done in order to facilitate the subsequent performance analysis. This has technical reasons. For the ordinary bandit problem the regret bounds of the order of $O(\sqrt{n^{-1}N\log N})$ were proved based on gains by Auer *et al.* [5] and it was only recently shown by Auer and Ottucsák [8] that it is possible to achieve the same type of bound for an algorithm based on losses. However, we do not know how to convert the latter algorithm into one that is efficiently computable for the shortest path problem.

To simplify the conversion, we assume that each path $\boldsymbol{i} \in \mathcal{P}$ is of the same length K for some $K > 0$. Note that although this assumption may seem to be restrictive at the first glance, from each acyclic directed graph (V, E) one can construct a new graph by adding at

most $(K-2)(|V|-2)+1$ vertices and edges (with constant loss zero) to the graph without modifying the losses of the paths such that each path from u to v will be of length K, where K denotes the length of the longest path of the original graph. If the number of edges is quadratic in the number of vertices, the size of the graph is not increased substantially.

A main feature of the algorithm below is that the gains are estimated for each edge and not for each path. This modification results in an improved upper bound on the performance with the number of edges in place of the number of paths. Moreover, using dynamic programming as in Takimoto and Warmuth [69], the algorithm can be computed efficiently. Another important ingredient of the algorithm is that one needs to make sure that every edge is sampled ("saw") sufficiently often. To this end, we introduce a set $\mathcal{C}$ of *covering paths* with the property that for each edge $e \in E$ there is a path $\boldsymbol{i} \in \mathcal{C}$ such that $e \in \boldsymbol{i}$. Observe that one can always find such a covering set of cardinality $|\mathcal{C}| \leq |E|$.

We note that the algorithm of [5] is a special case of the algorithm below: For any multi-armed bandit problem with N experts, one can define a graph with two vertices u and v, and N directed edges from u to v with weights corresponding to the losses of the experts. The solution of the shortest path problem in this case is equivalent to that of the original bandit problem with choosing expert $\boldsymbol{i}$ if the corresponding edge is chosen. For this graph, our algorithm reduces to the original algorithm of [5].

Note that the algorithm can be efficiently implemented using dynamic programming, similarly to Takimoto and Warmuth [28]. See the upcoming Theorem 4.1 for the formal statement.

The main result of this chapter is the following performance bound for the shortest-path bandit algorithm. It states that the normalized regret of the algorithm, after n rounds of play, is, roughly, of the order of $K\sqrt{|E| \ln N/n}$ where $|E|$ is the number of edges of the graph, K is the length of the paths, and N is the total number of paths.

A BANDIT ALGORITHM FOR SHORTEST PATHS

Parameters: real numbers $\beta > 0$, $0 < \eta, \gamma < 1$.
Initialization: Set $w_{e,0} = 1$ for each $e \in E$, $\bar{w}_{\boldsymbol{i},0} = 1$ for each $\boldsymbol{i} \in \mathcal{P}$, and $\overline{W}_0 = N$. For each round $t = 1, 2, \ldots$

(a) Choose a path $\boldsymbol{I}_t$ at random according to the distribution $\mathbf{p}_t$ on $\mathcal{P}$, defined by

$$p_{\boldsymbol{i},t} = \begin{cases} (1-\gamma)\frac{\bar{w}_{\boldsymbol{i},t-1}}{\overline{W}_{t-1}} + \frac{\gamma}{|\mathcal{C}|} & \text{if } \boldsymbol{i} \in \mathcal{C} \\ (1-\gamma)\frac{\bar{w}_{\boldsymbol{i},t-1}}{\overline{W}_{t-1}} & \text{if } \boldsymbol{i} \notin \mathcal{C}. \end{cases}$$

(b) Compute the probability of choosing each edge e as

$$q_{e,t} = \sum_{\boldsymbol{i}: e \in \boldsymbol{i}} p_{\boldsymbol{i},t} = (1-\gamma)\frac{\sum_{\boldsymbol{i}: e \in \boldsymbol{i}} \bar{w}_{\boldsymbol{i},t-1}}{\overline{W}_{t-1}} + \gamma \frac{|\{\boldsymbol{i} \in \mathcal{C} : e \in \boldsymbol{i}\}|}{|\mathcal{C}|}.$$

(c) Calculate the estimated gains

$$g'_{e,t} = \begin{cases} \frac{g_{e,t}+\beta}{q_{e,t}} & \text{if } e \in \boldsymbol{I}_t \\ \frac{\beta}{q_{e,t}} & \text{otherwise.} \end{cases}$$

(d) Compute the updated weights

$$\begin{aligned} w_{e,t} &= w_{e,t-1} e^{\eta g'_{e,t}} \\ \bar{w}_{\boldsymbol{i},t} &= \prod_{e \in \boldsymbol{i}} w_{e,t} = \bar{w}_{\boldsymbol{i},t-1} e^{\eta g'_{\boldsymbol{i},t}} \end{aligned}$$

where $g'_{\boldsymbol{i},t} = \sum_{e \in \boldsymbol{i}} g'_{e,t}$, and the sum of the total weights of the paths

$$\overline{W}_t = \sum_{\boldsymbol{i} \in \mathcal{P}} \bar{w}_{\boldsymbol{i},t}.$$

Figure 4.2: Bandit algorithm for shortest path problem.

Theorem 4.1. (GYÖRGY, LINDER AND OTTUCSÁK [41]). *For any $\delta \in (0,1)$ and parameters $0 \le \gamma < 1/2$, $0 < \beta \le 1$, and $\eta > 0$ satisfying $2\eta K|\mathcal{C}| \le \gamma$, the performance of the algorithm defined above can be bounded, with probability at least $1-\delta$, as*

$$\frac{1}{n}\left(\widehat{L}_n - \min_{\boldsymbol{i} \in \mathcal{P}} L_{\boldsymbol{i},n}\right) \le K\gamma + 2\eta K^2 |\mathcal{C}| + \frac{K}{n\beta} \ln \frac{|E|}{\delta} + \frac{\ln N}{n\eta} + |E|\beta.$$

In particular, choosing $\beta = \sqrt{\frac{K}{n|E|} \ln \frac{|E|}{\delta}}$, $\gamma = 2\eta K|\mathcal{C}|$, and $\eta = \sqrt{\frac{\ln N}{4nK^2|\mathcal{C}|}}$ yields for all

$n \geq \max\left\{\frac{K}{|E|}\ln\frac{|E|}{\delta}, 4|\mathcal{C}|\ln N\right\}$,

$$\frac{1}{n}\left(\widehat{L}_n - \min_{\boldsymbol{i}\in\mathcal{P}} L_{\boldsymbol{i},n}\right) \leq 2\sqrt{\frac{K}{n}}\left(\sqrt{4K|\mathcal{C}|\ln N} + \sqrt{|E|\ln\frac{|E|}{\delta}}\right).$$

The proof of the theorem is based on the analysis of the original algorithm of [5] with necessary modifications required to transform parts of the argument from paths to edges, and to use the connection between the gains of paths sharing common edges.

For the analysis we introduce some notation:

$$G_{\boldsymbol{i},n} = \sum_{t=1}^{n} g_{\boldsymbol{i},t} \qquad \text{and} \qquad G'_{\boldsymbol{i},n} = \sum_{t=1}^{n} g'_{\boldsymbol{i},t}$$

for each $\boldsymbol{i} \in \mathcal{P}$ and

$$G_{e,n} = \sum_{t=1}^{n} g_{e,t} \qquad \text{and} \qquad G'_{e,n} = \sum_{t=1}^{n} g'_{e,t}$$

for each $e \in E$, and

$$\widehat{G}_n = \sum_{t=1}^{n} g_{\boldsymbol{I}_t,t}.$$

Note that $g'_{e,t}$, $g'_{\boldsymbol{i},t}$, $G'_{e,n}$, and $G'_{\boldsymbol{i},n}$ are random variables that depend on $\boldsymbol{I}_t$.

The following lemma, shows that the deviation of the true cumulative gain from the estimated cumulative gain is of the order of $\sqrt{n}$. The proof is a modification of [21, Lemma 6.7].

Lemma 4.1. *For any $\delta \in (0,1)$, $0 \leq \beta < 1$ and $e \in E$ we have*

$$\mathbb{P}\left[G_{e,n} > G'_{e,n} + \frac{1}{\beta}\ln\frac{|E|}{\delta}\right] \leq \frac{\delta}{|E|}\ .$$

Proof. Fix $e \in E$. For any $u > 0$ and $c > 0$, by the Chernoff bound we have

$$\mathbb{P}[G_{e,n} > G'_{e,n} + u] \leq e^{-cu}\mathbb{E}e^{c(G_{e,n}-G'_{e,n})}\ . \tag{4.1}$$

Letting $u = \ln(|E|/\delta)/\beta$ and $c = \beta$, we get

$$e^{-cu}\mathbb{E}e^{c(G_{e,n}-G'_{e,n})} = e^{-\ln(|E|/\delta)}\mathbb{E}e^{\beta(G_{e,n}-G'_{e,n})} = \frac{\delta}{|E|}\mathbb{E}e^{\beta(G_{e,n}-G'_{e,n})}\ ,$$

so it suffices to prove that $\mathbb{E}e^{\beta(G_{e,n}-G'_{e,n})} \leq 1$ for all n. To this end, introduce

$$Z_t = e^{\beta(G_{e,t}-G'_{e,t})}\ .$$

Below we show that $\mathbb{E}_t[Z_t] \le Z_{t-1}$ for $t \ge 2$ where $\mathbb{E}_t$ denotes the conditional expectation $\mathbb{E}[\cdot|\boldsymbol{I}_1, \ldots, \boldsymbol{I}_{t-1}]$. Clearly,

$$Z_t = Z_{t-1} \exp\left(\beta\left(g_{e,t} - \frac{\mathbb{I}_{\{e\in\boldsymbol{I}_t\}} g_{e,t} + \beta}{q_{e,t}}\right)\right) .$$

Taking conditional expectations, we obtain

$$\begin{aligned}
&\mathbb{E}_t[Z_t] \\
&= Z_{t-1}\mathbb{E}_t\left[\exp\left(\beta\left(g_{e,t} - \frac{\mathbb{I}_{\{e\in\boldsymbol{I}_t\}} g_{e,t} + \beta}{q_{e,t}}\right)\right)\right] \\
&= Z_{t-1} e^{-\frac{\beta^2}{q_{e,t}}} \mathbb{E}_t\left[\exp\left(\beta\left(g_{e,t} - \frac{\mathbb{I}_{\{e\in\boldsymbol{I}_t\}} g_{e,t}}{q_{e,t}}\right)\right)\right] \\
&\le Z_{t-1} e^{-\frac{\beta^2}{q_{e,t}}} \mathbb{E}_t\left[1 + \beta\left(g_{e,t} - \frac{\mathbb{I}_{\{e\in\boldsymbol{I}_t\}} g_{e,t}}{q_{e,t}}\right) + \beta^2\left(g_{e,t} - \frac{\mathbb{I}_{\{e\in\boldsymbol{I}_t\}} g_{e,t}}{q_{e,t}}\right)^2\right] \qquad (4.2) \\
&= Z_{t-1} e^{-\frac{\beta^2}{q_{e,t}}} \mathbb{E}_t\left[1 + \beta^2\left(g_{e,t} - \frac{\mathbb{I}_{\{e\in\boldsymbol{I}_t\}} g_{e,t}}{q_{e,t}}\right)^2\right] \qquad (4.3) \\
&\le Z_{t-1} e^{-\frac{\beta^2}{q_{e,t}}} \mathbb{E}_t\left[1 + \beta^2\left(\frac{\mathbb{I}_{\{e\in\boldsymbol{I}_t\}} g_{e,t}}{q_{e,t}}\right)^2\right] \\
&\le Z_{t-1} e^{-\frac{\beta^2}{q_{e,t}}} \left(1 + \frac{\beta^2}{q_{e,t}}\right) \\
&\le Z_{t-1} . \qquad (4.4)
\end{aligned}$$

Here (4.2) holds since $\beta \le 1$, $g_{e,t} - \frac{\mathbb{I}_{\{e\in\boldsymbol{I}_t\}} g_{e,t}}{q_{e,t}} \le 1$ and $e^x \le 1 + x + x^2$ for $x \le 1$. (4.3) follows from $\mathbb{E}_t\left[\frac{\mathbb{I}_{\{e\in\boldsymbol{I}_t\}} g_{e,t}}{q_{e,t}}\right] = g_{e,t}$. Finally, (4.4) holds by the inequality $1 + x \le e^x$. Taking expectations on both sides proves $\mathbb{E}[Z_t] \le \mathbb{E}[Z_{t-1}]$. A similar argument shows that $\mathbb{E}[Z_1] \le 1$, implying $\mathbb{E}[Z_n] \le 1$ as desired. □

Proof of Theorem 4.1. As usual in the analysis of exponentially weighted average forecasters, we start with bounding the quantity $\ln \frac{\overline{W}_n}{\overline{W}_0}$. On the one hand, we have the lower bound

$$\ln \frac{\overline{W}_n}{\overline{W}_0} = \ln \sum_{\boldsymbol{i}\in\mathcal{P}} e^{\eta G'_{\boldsymbol{i},n}} - \ln N \ge \eta \max_{\boldsymbol{i}\in\mathcal{P}} G'_{\boldsymbol{i},n} - \ln N . \qquad (4.5)$$

To derive a suitable upper bound, first notice that the condition $\eta \le \frac{\gamma}{2K|\mathcal{C}|}$ implies $\eta g'_{\boldsymbol{i},t} \le 1$ for all $\boldsymbol{i}$ and t, since

$$\eta g'_{\boldsymbol{i},t} = \eta \sum_{e\in\boldsymbol{i}} g'_{e,t} \le \eta \sum_{e\in\boldsymbol{i}} \frac{1+\beta}{q_{e,t}} \le \frac{\eta K(1+\beta)|\mathcal{C}|}{\gamma} \le 1$$

where the second inequality follows because $q_{e,t} \geq \gamma/|\mathcal{C}|$ for each $e \in E$.
Therefore, using the fact that $e^x \leq 1 + x + x^2$ for all $x \leq 1$, for all $t = 1, 2, \ldots$ we have

$$\begin{aligned}
\ln \frac{\overline{W}_t}{\overline{W}_{t-1}} &= \ln \sum_{\boldsymbol{i} \in \mathcal{P}} \frac{\bar{w}_{\boldsymbol{i},t-1}}{\overline{W}_{t-1}} e^{\eta g'_{i,t}} \\
&= \ln \left(\sum_{\boldsymbol{i} \in \mathcal{P}} \frac{p_{\boldsymbol{i},t} - \frac{\gamma}{|\mathcal{C}|} \mathbb{I}_{\{\boldsymbol{i} \in \mathcal{C}\}}}{1-\gamma} e^{\eta g'_{i,t}} \right) \qquad (4.6) \\
&\leq \ln \left(\sum_{\boldsymbol{i} \in \mathcal{P}} \frac{p_{\boldsymbol{i},t} - \frac{\gamma}{|\mathcal{C}|} \mathbb{I}_{\{\boldsymbol{i} \in \mathcal{C}\}}}{1-\gamma} \left(1 + \eta g'_{\boldsymbol{i},t} + \eta^2 g'^2_{\boldsymbol{i},t} \right) \right) \\
&\leq \ln \left(1 + \sum_{\boldsymbol{i} \in \mathcal{P}} \frac{p_{\boldsymbol{i},t}}{1-\gamma} \left(\eta g'_{\boldsymbol{i},t} + \eta^2 g'^2_{\boldsymbol{i},t} \right) \right) \\
&\leq \frac{\eta}{1-\gamma} \sum_{\boldsymbol{i} \in \mathcal{P}} p_{\boldsymbol{i},t} g'_{\boldsymbol{i},t} + \frac{\eta^2}{1-\gamma} \sum_{\boldsymbol{i} \in \mathcal{P}} p_{\boldsymbol{i},t} g'^2_{\boldsymbol{i},t} \qquad (4.7)
\end{aligned}$$

where (4.6) follows form the definition of $p_{\boldsymbol{i},t}$, and (4.7) holds by the inequality $\ln(1+x) \leq x$ for all $x > -1$.

Next we bound the sums in (4.7). On the one hand,

$$\begin{aligned}
\sum_{\boldsymbol{i} \in \mathcal{P}} p_{\boldsymbol{i},t} g'_{\boldsymbol{i},t} &= \sum_{\boldsymbol{i} \in \mathcal{P}} p_{\boldsymbol{i},t} \sum_{e \in \boldsymbol{i}} g'_{e,t} = \sum_{e \in E} g'_{e,t} \sum_{\boldsymbol{i} \in \mathcal{P}: e \in \boldsymbol{i}} p_{\boldsymbol{i},t} \\
&= \sum_{e \in E} g'_{e,t} q_{e,t} = g_{\boldsymbol{I}_t,t} + |E|\beta.
\end{aligned}$$

On the other hand,

$$\begin{aligned}
\sum_{\boldsymbol{i} \in \mathcal{P}} p_{\boldsymbol{i},t} g'^2_{\boldsymbol{i},t} &= \sum_{\boldsymbol{i} \in \mathcal{P}} p_{\boldsymbol{i},t} \left(\sum_{e \in \boldsymbol{i}} g'_{e,t} \right)^2 \\
&\leq \sum_{\boldsymbol{i} \in \mathcal{P}} p_{\boldsymbol{i},t} K \sum_{e \in \boldsymbol{i}} g'^2_{e,t} \\
&= K \sum_{e \in E} g'^2_{e,t} \sum_{\boldsymbol{i} \in \mathcal{P}: e \in \boldsymbol{i}} p_{\boldsymbol{i},t} \\
&= K \sum_{e \in E} g'^2_{e,t} q_{e,t} \\
&= K \sum_{e \in E} q_{e,t} g'_{e,t} \frac{\beta + \mathbb{I}_{\{e \in \boldsymbol{I}_t\}} g_{e,t}}{q_{e,t}} \\
&\leq K(1+\beta) \sum_{e \in E} g'_{e,t}
\end{aligned}$$

where the first inequality is due to the inequality between the arithmetic and quadratic mean, and the second one holds because $g_{e,t} \leq 1$. Therefore,

$$\ln \frac{\overline{W}_t}{\overline{W}_{t-1}} \leq \frac{\eta}{1-\gamma} \left(g_{\boldsymbol{I}_t,t} + |E|\beta \right) + \frac{\eta^2 K(1+\beta)}{1-\gamma} \sum_{e \in E} g'_{e,t} \ .$$

Summing for $t = 1, \ldots, n$, we obtain

$$\begin{aligned} \ln \frac{\overline{W}_n}{\overline{W}_0} &\leq \frac{\eta}{1-\gamma} \left(\widehat{G}_n + n|E|\beta \right) + \frac{\eta^2 K(1+\beta)}{1-\gamma} \sum_{e \in E} G'_{e,n} \\ &\leq \frac{\eta}{1-\gamma} \left(\widehat{G}_n + n|E|\beta \right) + \frac{\eta^2 K(1+\beta)}{1-\gamma} |\mathcal{C}| \max_{\boldsymbol{i} \in \mathcal{P}} G'_{\boldsymbol{i},n} \end{aligned} \tag{4.8}$$

where the second inequality follows since $\sum_{e \in E} G'_{e,n} \leq \sum_{\boldsymbol{i} \in \mathcal{C}} G'_{\boldsymbol{i},n}$. Combining the upper bound with the lower bound (4.5), we obtain

$$\widehat{G}_n \geq (1 - \gamma - \eta K(1+\beta)|\mathcal{C}|) \max_{\boldsymbol{i} \in \mathcal{P}} G'_{\boldsymbol{i},n} - \frac{1-\gamma}{\eta} \ln N - n|E|\beta. \tag{4.9}$$

Now using Lemma 4.1 and applying the union bound, for any $\delta \in (0,1)$ we have that, with probability at least $1 - \delta$,

$$\widehat{G}_n \geq (1 - \gamma - \eta K(1+\beta)|\mathcal{C}|) \left(\max_{\boldsymbol{i} \in \mathcal{P}} G_{\boldsymbol{i},n} - \frac{K}{\beta} \ln \frac{|E|}{\delta} \right) - \frac{1-\gamma}{\eta} \ln N - n|E|\beta \ ,$$

where we used $1 - \gamma - \eta K(1+\beta)|\mathcal{C}| \geq 0$ which follows from the assumptions of the theorem. Since $\widehat{G}_n = Kn - \widehat{L}_n$ and $G_{\boldsymbol{i},n} = Kn - L_{\boldsymbol{i},n}$ for all $\boldsymbol{i} \in \mathcal{P}$, we have

$$\begin{aligned} \widehat{L}_n &\leq Kn \left(\gamma + \eta(1+\beta)K|\mathcal{C}| \right) + (1 - \gamma - \eta(1+\beta)K|\mathcal{C}|) \min_{\boldsymbol{i} \in \mathcal{P}} L_{\boldsymbol{i},n} \\ &\quad + (1 - \gamma - \eta(1+\beta)K|\mathcal{C}|) \frac{K}{\beta} \ln \frac{|E|}{\delta} + \frac{1-\gamma}{\eta} \ln N + n|E|\beta \end{aligned}$$

with probability at least $1 - \delta$. This implies

$$\begin{aligned} \widehat{L}_n - \min_{\boldsymbol{i} \in \mathcal{P}} L_{\boldsymbol{i},n} &\leq Kn\gamma + \eta(1+\beta)nK^2|\mathcal{C}| + \frac{K}{\beta} \ln \frac{|E|}{\delta} + \frac{1-\gamma}{\eta} \ln N + n|E|\beta \\ &\leq Kn\gamma + 2\eta nK^2|\mathcal{C}| + \frac{K}{\beta} \ln \frac{|E|}{\delta} + \frac{\ln N}{\eta} + n|E|\beta \end{aligned}$$

with probability at least $1 - \delta$, which is the first statement of the theorem. Setting

$$\beta = \sqrt{\frac{K}{n|E|} \ln \frac{|E|}{\delta}} \qquad \text{and} \qquad \gamma = 2\eta K|\mathcal{C}|$$

results in the inequality

$$\widehat{L}_n - \min_{\boldsymbol{i}\in\mathcal{P}} L_{\boldsymbol{i},n} \le 4\eta n K^2 |\mathcal{C}| + \frac{\ln N}{\eta} + 2\sqrt{nK|E| \ln \frac{|E|}{\delta}}$$

which holds with probability at least $1-\delta$ if $n \ge (K/|E|)\ln(|E|/\delta)$ (to ensure $\beta \le 1$). Finally, setting

$$\eta = \sqrt{\frac{\ln N}{4nK^2|\mathcal{C}|}}$$

yields the last statement of the theorem ($n \ge 4\ln N|\mathcal{C}|$ is required to ensure $\gamma \le 1/2$). □

Next we analyze the computational complexity of the algorithm. The next result shows that the algorithm is feasible as its complexity is linear in the size (number of edges) of the graph.

Theorem 4.2. (GYÖRGY, LINDER AND OTTUCSÁK [41]). *The proposed algorithm can be implemented efficiently with time complexity $O(n|E|)$ and space complexity $O(|E|)$.*

Proof. The two complex steps of the algorithm are steps (a) and (b), both of which can be computed, similarly to Takimoto and Warmuth [69], using dynamic programming. To perform these steps efficiently, first we order the vertices of the graph. Since we have an acyclic directed graph, its vertices can be labeled (in $O(|E|)$ time) from 1 to $|V|$ such that $u = 1$, $v = |V|$, and if $(v_1, v_2) \in E$, then $v_1 < v_2$. For any pair of vertices $u_1 < v_1$ let $\mathcal{P}_{u_1,v_1}$ denote the set of paths from u_1 to v_1, and for any vertex $s \in V$, let

$$H_t(s) = \sum_{\boldsymbol{i}\in\mathcal{P}_{s,v}} \prod_{e\in\boldsymbol{i}} w_{e,t}$$

and

$$\widehat{H}_t(s) = \sum_{\boldsymbol{i}\in\mathcal{P}_{u,s}} \prod_{e\in\boldsymbol{i}} w_{e,t} \ .$$

Given the edge weights $\{w_{e,t}\}$, $H_t(s)$ can be computed recursively for $s = |V|-1, \dots, 1$, and $\widehat{H}_t(s)$ can be computed recursively for $s = 2, \dots, |V|$ in $O(|E|)$ time (letting $H_t(v) = \widehat{H}_t(u) = 1$ by definition). In step (a), first one has to decide with probability γ whether $\boldsymbol{I}_t$ is generated according to the graph weights, or it is chosen uniformly from $\mathcal{C}$. If $\boldsymbol{I}_t$ is to be drawn according to the graph weights, it can be shown that its vertices can be chosen one by one such that if the first k vertices of $\boldsymbol{I}_t$ are $v_0 = u, v_1, \dots, v_{k-1}$, then the next vertex of $\boldsymbol{I}_t$ can be chosen to be any $v_k > v_{k-1}$, satisfying $(v_{k-1}, v_k) \in E$, with probability $w_{(v_{k-1},v_k),t-1} H_{t-1}(v_k)/H_{t-1}(v_{k-1})$. The other computationally demanding step, namely step (b), can be performed easily by noting that for any edge (v_1, v_2),

$$q_{(v_1,v_2),t} \;=\; (1-\gamma)\frac{\widehat{H}_{t-1}(v_1) w_{(v_1,v_2),t-1} H_{t-1}(v_2)}{H_{t-1}(u)} + \;\gamma\frac{|\{\boldsymbol{i}\in\mathcal{C} : (v_1,v_2)\in\boldsymbol{i}\}|}{|\mathcal{C}|}$$

as desired. □

4.4 A combination of the label efficient and bandit settings

In this section we investigate a combination of the multi-armed bandit and the label efficient problems. This means that the decision maker only has access to the loss of all the edges on the chosen path upon request and the total number of requests must be bounded by a constant m. This combination is motivated by some applications, in which feedback information is costly to obtain.

In the general label efficient decision problem, after taking an action, the decision maker has the option to query the losses of all possible actions. For this problem, Cesa-Bianchi *et al.* [22] proved an upper bound on the normalized regret of order $O(K\sqrt{\ln(4N/\delta)/(m)})$ which holds with probability at least $1-\delta$, where K is the length of the longest path in the graph.

Our model of the label-efficient bandit problem for shortest paths is motivated by an application to a particular packet switched network model. This model, called the Cognitive Packet Network (CPN), was introduced by Gelenbe *et al.* [27, 28].

Example 4.1. (Cognitive Packet Network) CPN is a specific autonomic technique that offers adaptive routing as a way to better QoS to users and it is oriented toward to use of self-awareness in the network and it is based on strictly automatic defence without human intervention.

In these networks a particular type of packets, called smart packets, are used to explore the network (e.g., the delay of the chosen path). These packets do not carry any useful data; they are merely used for exploring the network. The other type of packets are the data packets, which do not collect any information about their paths. The task of the decision maker is to send packets from the source to the destination over paths with minimum average transmission delay (or packet loss rate). In this scenario, smart packets are used to query the delay (or loss) of the chosen path. However, as these packets do not transport information, there is a trade-off between the number of queries and the usage of the network. If data packets are on the average α times larger than smart packets (note that typically $\alpha \gg 1$) and ε is the proportion of time instances when smart packets are used to explore the network, then $\varepsilon/(\varepsilon+\alpha(1-\varepsilon))$ is the proportion of the bandwidth sacrificed for well informed routing decisions.

The CPN model is implemented and integrated into *Linux kernel 2.2.x* and it is the object of the *US Patent No. 6804201*. The performance of the CPN is extensively studied experimentally in a test-bed (with 80 nodes) [26] in Imperial College. These experimental measurements are focused on the techniques using genetic algorithm [29] and neural network [27] to choose the next path. However, these papers do not touch on the theoretical optimality of the proposed methods.

We study a combined algorithm which, at each time slot t, queries the loss of the chosen path with probability ε (as in the solution of the label efficient problem proposed in [22]), and, similarly to the multi-armed bandit case, computes biased estimates $g'_{i,t}$ of the true

gains $g_{\boldsymbol{i},t}$. Just as in the previous section, it is assumed that each path of the graph is of the same length K.

The algorithm differs from our bandit algorithm of the previous section only in step (c), which is modified in the spirit of [22]. The modified step is given below:

MODIFIED STEP FOR THE LABEL EFFICIENT BANDIT ALGORITHM FOR SHORTEST PATHS

(c') Draw a Bernoulli random variable S_t with $\mathbb{P}((S_t = 1)) = \varepsilon$, and compute the estimated gains

$$g'_{e,t} = \begin{cases} \frac{g_{e,t}+\beta}{\varepsilon q_{e,t}} S_t & \text{if } e \in \boldsymbol{I}_t \\ \frac{\beta}{\varepsilon q_{e,t}} S_t & \text{if } e \notin \boldsymbol{I}_t \end{cases} .$$

Figure 4.3: The modified step for the LE+MAB problem for shortest path.

The performance of the algorithm is analyzed in the next theorem, which can be viewed as a combination of Theorem 4.1 in the preceding section and Theorem 2 of [22].

Theorem 4.3. (GYÖRGY, LINDER AND OTTUCSÁK [41]). *For any $\delta \in (0,1)$, $\varepsilon \in (0,1]$, parameters $\eta = \sqrt{\frac{\varepsilon \ln N}{4nK^2|\mathcal{C}|}}$, $\gamma = \frac{2\eta K|\mathcal{C}|}{\varepsilon} \leq 1/2$, and $\beta = \sqrt{\frac{K}{n|E|\varepsilon} \ln \frac{2|E|}{\delta}} \leq 1$, and for all*

$$n \geq \frac{1}{\varepsilon} \max \left\{ \frac{K^2 \ln^2(2|E|/\delta)}{|E| \ln N}, \frac{|E| \ln(2|E|/\delta)}{K}, 4|\mathcal{C}| \ln N \right\}$$

the performance of the algorithm defined above can be bounded, with probability at least $1-\delta$, as

$$\begin{aligned} &\frac{1}{n}\left(\widehat{L}_n - \min_{\boldsymbol{i}\in\mathcal{P}} L_{\boldsymbol{i},t}\right) \\ &\quad \leq \sqrt{\frac{K}{n\varepsilon}}\left(4\sqrt{K|\mathcal{C}|\ln N} + 5\sqrt{|E|\ln\frac{2|E|}{\delta}} + \sqrt{8K\ln\frac{2}{\delta}}\right) + \frac{4K}{3n\varepsilon}\ln\frac{2N}{\delta} \\ &\quad \leq \frac{27K}{2}\sqrt{\frac{|E|\ln\frac{2N}{\delta}}{n\varepsilon}} . \end{aligned}$$

If ε is chosen as $(m - \sqrt{2m\ln(1/\delta)})/n$ then, with probability at least $1-\delta$, the total number of queries is bounded by m (see [21, Lemma 6.1]) and the performance bound above is of the order of $K\sqrt{|E|\ln(N/\delta)/m}$.

For the proof we need the following two lemmas. The first is the Bernstein's inequality for martingales differences [13].

Lemma 4.2. *Let $X_1, \ldots, X_n$ be a martingale difference sequence such that $X_t \in [a, b]$ with probability one ($t = 1, \ldots, n$). Assume that, for all t,*

$$\mathbb{E}\left[X_t^2 | X_{t-1}, \ldots, X_1\right] \leq \sigma^2 \ a.s.$$

Then, for all $\varepsilon > 0$,

$$\mathbb{P}\left\{\sum_{t=1}^{n} X_t > \varepsilon\right\} \leq e^{\frac{-\varepsilon^2}{2n\sigma^2 + 2\varepsilon(b-a)/3}}$$

and therefore

$$\mathbb{P}\left\{\sum_{t=1}^{n} X_t > \sqrt{2n\sigma^2 \ln \delta^{-1}} + 2\ln \delta^{-1}(b-a)/3\right\} \leq \delta.$$

Similarly to Theorem 4.1, we need a lemma which reveals the connection between the true and the estimated cumulative losses. However, here we need a more careful analysis because the "shifting term" $\frac{\beta}{\varepsilon q_{e,t}} S_t$, is a random variable.

Lemma 4.3. *For any $0 < \delta < 1$, $0 < \varepsilon \leq 1$, for any*

$$n \geq \frac{1}{\varepsilon} \max\left\{\frac{K^2 \ln^2(2|E|/\delta)}{|E| \ln N}, \frac{K \ln(2|E|/\delta)}{|E|}\right\} \ ,$$

parameters

$$\frac{2\eta K|\mathcal{C}|}{\varepsilon} \leq \gamma, \qquad \eta = \sqrt{\frac{\varepsilon \ln N}{4nK^2|\mathcal{C}|}} \quad \text{and} \quad \beta = \sqrt{\frac{K}{n|E|\varepsilon} \ln \frac{2|E|}{\delta}} \leq 1 \ ,$$

and $e \in E$, we have

$$\mathbb{P}\left[G_{e,n} > G'_{e,n} + \frac{4}{\beta\varepsilon} \ln \frac{2|E|}{\delta}\right] \leq \frac{\delta}{2|E|} \ .$$

Proof. Fix $e \in E$. Using (4.1) with $u = \frac{4}{\beta\varepsilon} \ln \frac{2|E|}{\delta}$ and $c = \frac{\beta\varepsilon}{4}$, it suffices to prove for all n that

$$\mathbb{E}\left[e^{c(G_{e,n} - G'_{e,n})}\right] \leq 1 \ .$$

Similarly to Lemma 4.1 we introduce $Z_t = e^{c(G_{e,t} - G'_{e,t})}$ and we show that $Z_1, \ldots, Z_n$ is a supermartingale, that is $\mathbb{E}_t[Z_t] \leq Z_{t-1}$ for $t \geq 2$ where $\mathbb{E}_t$ denotes $\mathbb{E}[\cdot|(\boldsymbol{I}_1, S_1), \ldots, (\boldsymbol{I}_{t-1}, S_{t-1})]$. Taking conditional expectations, we obtain

$$\begin{aligned}
\mathbb{E}_t[Z_t] &= Z_{t-1}\mathbb{E}_t\left[e^{c\left(g_{e,t} - \frac{\mathbb{I}_{\{e \in I_t\}} S_t g_{e,t} + S_t\beta}{q_{e,t}\varepsilon}\right)}\right] \\
&\leq Z_{t-1}\mathbb{E}_t\left[1 + c\left(g_{e,t} - \frac{\mathbb{I}_{\{e \in I_t\}} S_t g_{e,t} + S_t\beta}{q_{e,t}\varepsilon}\right)\right. \\
&\quad \left. + c^2\left(g_{e,t} - \frac{\mathbb{I}_{\{e \in I_t\}} S_t g_{e,t} + S_t\beta}{q_{e,t}\varepsilon}\right)^2\right]. \qquad (4.10)
\end{aligned}$$

Since

$$\mathbb{E}_t\left[g_{e,t} - \frac{\mathbb{I}_{\{e\in I_t\}}S_t g_{e,t} + S_t\beta}{q_{e,t}\varepsilon}\right] = -\frac{\beta}{q_{e,t}}$$

and

$$\mathbb{E}_t\left[\left(g_{e,t} - \frac{\mathbb{I}_{\{e\in I_t\}}S_t g_{e,t}}{q_{e,t}\varepsilon}\right)^2\right] \le \mathbb{E}_t\left[\left(\frac{\mathbb{I}_{\{e\in I_t\}}S_t g_{e,t}}{q_{e,t}\varepsilon}\right)^2\right] \le \frac{1}{q_{e,t}\varepsilon}$$

we get from (4.10) that

$$\begin{aligned}
&\mathbb{E}_t[Z_t] \\
&\quad\le\; Z_{t-1}\mathbb{E}_t\left[1 - \frac{c\beta}{q_{e,t}} + \frac{c^2}{q_{e,t}\varepsilon} + c^2\left(\frac{2\mathbb{I}_{\{e\in I_t\}}S_t g_{e,t}\beta}{q_{e,t}^2\varepsilon^2} - \frac{2g_{e,t}S_t\beta}{q_{e,t}\varepsilon} + \frac{S_t\beta^2}{q_{e,t}^2\varepsilon^2}\right)\right] \\
&\quad\le\; Z_{t-1}\left(1 + \frac{c}{q_{e,t}}\left(-\beta + \frac{c}{\varepsilon} + c\beta\left(\frac{2}{\varepsilon} + \frac{\beta}{q_{e,t}\varepsilon}\right)\right)\right). \qquad (4.11)
\end{aligned}$$

Since $c = \beta\varepsilon/4$ we have

$$\begin{aligned}
-\beta + \frac{c}{\varepsilon} + c\beta\left(\frac{2}{\varepsilon} + \frac{\beta}{q_{e,t}\varepsilon}\right) &= -\frac{3\beta}{4} + \frac{\beta^2\varepsilon}{4}\left(\frac{2}{\varepsilon} + \frac{\beta}{q_{e,t}\varepsilon}\right) \\
&= -\frac{3\beta}{4} + \frac{\beta^2}{2} + \frac{\beta^3}{4q_{e,t}} \\
&\le -\frac{\beta}{4} + \frac{\beta^3}{4q_{e,t}} \\
&\le -\frac{\beta}{4} + \frac{\beta^3|\mathcal{C}|}{4\gamma} \qquad (4.12) \\
&\le 0, \qquad (4.13)
\end{aligned}$$

where (4.12) follows from $q_{e,t} \ge \frac{\gamma}{|\mathcal{C}|}$ and (4.13) holds since $\beta \le 1$ and by

$$\frac{\beta^2|\mathcal{C}|}{\gamma} \le \frac{\beta^2\varepsilon}{2\eta K} \le 1\ ,$$

and the last inequality is ensured by $n \ge \frac{K^2\ln^2(2|E|/\delta)}{\varepsilon|E|\ln N}$, the assumption of the lemma.

Combining (4.11) and (4.13) we get that $\mathbb{E}_t[Z_t] \le Z_{t-1}$. Taking expectations on both sides of the inequality, we get $\mathbb{E}[Z_t] \le \mathbb{E}[Z_{t-1}]$ and since $\mathbb{E}[Z_1] \le 1$, we obtain $\mathbb{E}[Z_n] \le 1$ as desired. □

Proof of Theorem 4.3. The proof of the theorem is a generalization of that of Theorem 4.1, and follows the same lines with some extra technicalities to handle the effects of

the modified step (c'). Therefore, in the following we emphasize only the differences. First note that (4.5) and (4.7) also hold in this case. Bounding the sums in (4.7), one obtains

$$\sum_{\boldsymbol{i}\in\mathcal{P}} p_{\boldsymbol{i},t} g'_{\boldsymbol{i},t} = \frac{S_t}{\varepsilon}\left(g_{\boldsymbol{I}_t,t} + |E|\beta\right)$$

and

$$\sum_{\boldsymbol{i}\in\mathcal{P}} p_{\boldsymbol{i},t} g'^2_{\boldsymbol{i},t} \le \frac{1}{\varepsilon} K(1+\beta) \sum_{e\in E} g'_{e,t} \ .$$

Plugging these bounds into (4.7) and summing for $t = 1, \ldots, n$, we obtain

$$\ln \frac{\overline{W}_n}{\overline{W}_0} \ \le \ \frac{\eta}{1-\gamma} \sum_{t=1}^{n} \frac{S_t}{\varepsilon}\left(g_{I_t,t} + |E|\beta\ \right) + \frac{\eta^2 K(1+\beta)}{(1-\gamma)\varepsilon} |\mathcal{C}| \max_{\boldsymbol{i}\in\mathcal{P}} G'_{\boldsymbol{i},n} \ .$$

Combining the upper bound with the lower bound (4.5), we obtain

$$\sum_{t=1}^{n} \frac{S_t}{\varepsilon}\left(g_{\boldsymbol{I}_t,t} + |E|\beta\ \right) \ \ge \ \left(1-\gamma-\frac{\eta K(1+\beta)|\mathcal{C}|}{\varepsilon}\right) \max_{\boldsymbol{i}\in\mathcal{P}} G'_{\boldsymbol{i},n} - \frac{\ln N}{\eta} \ . \tag{4.14}$$

To relate the left-hand side of the above inequality to the real gain $\sum_{t=1}^{n} g_{\boldsymbol{I}_t,t}$, notice that

$$X_t = \frac{S_t}{\varepsilon}\left(g_{\boldsymbol{I}_t,t} + |E|\beta\right) - \left(g_{\boldsymbol{I}_t,t} + |E|\beta\right)$$

is a martingale difference sequence with respect to $(\boldsymbol{I}_1, S_1), (\boldsymbol{I}_2, S_2), \ldots$. Now for all $t = 1, \ldots, n$, we have the bound

$$\begin{aligned}
\mathbb{E}\left[X_t^2 | (\boldsymbol{I}_1, S_1), \ldots, (\boldsymbol{I}_{t-1}, S_{t-1})\right] \ &\le \ \mathbb{E}\left[\frac{S_t}{\varepsilon^2}(g_{\boldsymbol{I}_t,t} + |E|\beta)^2 \middle| (\boldsymbol{I}_1, S_1), \ldots, (\boldsymbol{I}_{t-1}, S_{t-1})\right] \\
&\le \ \frac{(K + |E|\beta)^2}{\varepsilon} \\
&\le \ \frac{4K^2}{\varepsilon} \stackrel{\text{def}}{=} \sigma^2,
\end{aligned} \tag{4.15}$$

where (4.15) holds by $n \ge \frac{|E|\ln(2|E|/\delta)}{K\varepsilon}$ (to ensure $\beta|E| \le K$). We know that

$$X_t \in \left[-2K, \left(\frac{1}{\varepsilon} - 1\right) 2K\right]$$

for all t. Now apply Bernstein's inequality for martingale differences (see Lemma 4.2 in the Appendix) to obtain

$$\mathbb{P}\left[\sum_{t=1}^{n} X_t > u\right] \le \frac{\delta}{2} \ , \tag{4.16}$$

where

$$u = \sqrt{2n\frac{4K^2}{\varepsilon}\ln\left(\frac{2}{\delta}\right)} + \frac{4K}{3\varepsilon}\ln\left(\frac{2}{\delta}\right).$$

From (4.16) we get

$$\mathbb{P}\left[\sum_{t=1}^{n}\frac{S_t}{\varepsilon}\left(g_{\boldsymbol{I}_t,t} + |E|\beta\right) \geq \widehat{G}_n + \beta n|E| + u\right] \leq \frac{\delta}{2}\,. \tag{4.17}$$

Now Lemma 4.3, the union bound, and (4.17) combined with (4.14) yield, with probability at least $1-\delta$,

$$\begin{aligned}\widehat{G}_n \;\geq\; & \left(1-\gamma-\frac{\eta K(1+\beta)|\mathcal{C}|}{\varepsilon}\right)\left(\max_{\boldsymbol{i}\in\mathcal{P}} G_{\boldsymbol{i},n} - \frac{4K}{\beta\varepsilon}\ln\frac{2|E|}{\delta}\right)\\ & -\frac{\ln N}{\eta} - \beta n|E| - u\end{aligned}$$

since the coefficient of $G'_{i,n}$ is greater than zero by the assumptions of the theorem.

Since $\widehat{G}_n = Kn - \widehat{L}_n$ and $G_{\boldsymbol{i},n} = Kn - L_{\boldsymbol{i},n}$, we have

$$\begin{aligned}\widehat{L}_n \;\leq\; & \left(1-\gamma-\frac{K(1+\beta)\eta|\mathcal{C}|}{\varepsilon}\right)\min_{\boldsymbol{i}\in\mathcal{P}} L_{\boldsymbol{i},n} + Kn\left(\gamma + \frac{K(1+\beta)\eta|\mathcal{C}|}{\varepsilon}\right)\\ & + \left(1-\gamma-\frac{K(1+\beta)\eta|\mathcal{C}|}{\varepsilon}\right)\frac{4K}{\beta\varepsilon}\ln\frac{2|E|}{\delta} + \beta n|E| + \frac{\ln N}{\eta} + u\\ \leq\; & \min_{\boldsymbol{i}\in\mathcal{P}} L_{\boldsymbol{i},n} + Kn\left(\gamma + \frac{K(1+\beta)\eta|\mathcal{C}|}{\varepsilon}\right) + 5\beta n|E| + \frac{\ln N}{\eta} + u\,,\end{aligned}$$

where we used the fact that $\frac{K}{\beta\varepsilon}\ln\frac{2|E|}{\delta} = \beta n|E|$.

Substituting the value of β, η and γ, we have

$$\begin{aligned}\widehat{L}_n - \min_{\boldsymbol{i}\in\mathcal{P}} L_{\boldsymbol{i},n} \leq & Kn\frac{2K\eta|\mathcal{C}|}{\varepsilon} + Kn\frac{2K\eta|\mathcal{C}|}{\varepsilon} + \frac{\ln N}{\eta} + 5\beta n|E| + u\\ \leq & 4K\sqrt{\frac{n|\mathcal{C}|\ln N}{\varepsilon}} + 5\sqrt{\frac{n|E|K\ln(2|E|/\delta)}{\varepsilon}} + u\\ \leq & \sqrt{\frac{nK}{\varepsilon}}\left(4\sqrt{K|\mathcal{C}|\ln N} + 5\sqrt{|E|\ln(2|E|/\delta)} + \sqrt{8K\ln(2/\delta)}\right)\\ & + \frac{4K}{3\varepsilon}\ln(2/\delta)\end{aligned}$$

as desired. □

4.5 A bandit algorithm for tracking the shortest path

Our goal in this section is to extend the bandit algorithm so that it is able to compete with time-varying paths under small computational complexity. This is a variant of the problem known as *tracking the best expert*; see, for example, Herbster and Warmuth [46], Vovk [71], Auer and Warmuth [9], Bousquet and Warmuth [17], Herbster and Warmuth [47].

To describe the loss the decision maker is compared to, consider the following "m-partition" prediction scheme: the sequence of paths is partitioned into $m+1$ contiguous segments, and on each segment the scheme assigns exactly one of the N paths. Formally, an m-partition $\texttt{Part}(n, m, \mathbf{t}, \underline{\boldsymbol{i}})$ of the n paths is given by an m-tuple $\mathbf{t} = (t_1, \ldots, t_m)$ such that $t_0 = 1 < t_1 < \cdots < t_m < n+1 = t_{m+1}$, and an $(m+1)$-vector $\underline{\boldsymbol{i}} = (\boldsymbol{i}_0, \ldots, \boldsymbol{i}_m)$ where $\boldsymbol{i}_j \in \mathcal{P}$. At each time instant t, $t_j \le t < t_{j+1}$, path $\boldsymbol{i}_j$ is used to predict the best path. The cumulative loss of a partition $\texttt{Part}(n, m, \mathbf{t}, \underline{\boldsymbol{i}})$ is

$$L(\texttt{Part}(n, m, \mathbf{t}, \underline{\boldsymbol{i}})) = \sum_{j=0}^{m} \sum_{t=t_j}^{t_{j+1}-1} \ell_{\boldsymbol{i}_j, t} = \sum_{j=0}^{m} \sum_{t=t_j}^{t_{j+1}-1} \sum_{e \in \boldsymbol{i}_j} \ell_{e,t}.$$

The goal of the decision maker is to perform as well as the best time-varying path (partition), that is, to keep the normalized regret

$$\frac{1}{n}\left(\widehat{L}_n - \min_{\mathbf{t}, \underline{\boldsymbol{i}}} L(\texttt{Part}(n, m, \mathbf{t}, \underline{\boldsymbol{i}}))\right)$$

as small as possible (with high probability) for all possible outcome sequences.

In the "classical" tracking problem there is a relatively small number of "base" experts and the goal of the decision maker is to predict as well as the best "compound" expert (i.e., time-varying expert). However in our case, base experts correspond to all paths of the graph between source and destination whose number is typically exponentially large in the number of edges, and therefore we cannot directly apply the computationally efficient methods for tracking the best expert. György, Linder, and Lugosi [38] develop efficient algorithms for tracking the best expert for certain large and structured classes of base experts, including the shortest path problem. The purpose of the following algorithm is to extend the methods of [38] to the bandit setting when the forecaster only observes the losses of the edges on the chosen path.

The following performance bounds shows that the normalized regret with respect to the best time-varying path which is allowed to switch paths m times is roughly of the order of $K\sqrt{(m/n)|\mathcal{C}| \ln N}$.

A BANDIT ALGORITHM FOR TRACKING SHORTEST PATHS

Parameters: real numbers $\beta > 0$, $0 < \eta, \gamma < 1$, $0 \leq \alpha \leq 1$.
Initialization: Set $w_{e,0} = 1$ for each $e \in E$, $\bar{w}_{\boldsymbol{i},0} = 1$ for each $\boldsymbol{i} \in \mathcal{P}$, and $\overline{W}_0 = N$. For each round $t = 1, 2, \ldots$

(a) Choose a path $\boldsymbol{I}_t$ according to the distribution $\mathbf{p}_t$ defined by

$$p_{\boldsymbol{i},t} = \begin{cases} (1-\gamma)\frac{\bar{w}_{\boldsymbol{i},t-1}}{\overline{W}_{t-1}} + \frac{\gamma}{|\mathcal{C}|} & \text{if } \boldsymbol{i} \in \mathcal{C}; \\ (1-\gamma)\frac{\bar{w}_{\boldsymbol{i},t-1}}{\overline{W}_{t-1}} & \text{if } \boldsymbol{i} \notin \mathcal{C}. \end{cases}$$

(b) Compute the probability of choosing each edge e as

$$q_{e,t} = \sum_{\boldsymbol{i}: e \in \boldsymbol{i}} p_{\boldsymbol{i},t} = (1-\gamma)\frac{\sum_{\boldsymbol{i}: e \in \boldsymbol{i}} \bar{w}_{\boldsymbol{i},t-1}}{\overline{W}_{t-1}} + \gamma \frac{|\{\boldsymbol{i} \in \mathcal{C} : e \in \boldsymbol{i}\}|}{|\mathcal{C}|}.$$

(c) Calculate the estimated gains

$$g'_{e,t} = \begin{cases} \frac{g_{e,t}+\beta}{q_{e,t}} & \text{if } e \in \boldsymbol{I}_t; \\ \frac{\beta}{q_{e,t}} & \text{otherwise.} \end{cases}$$

(d) Compute the updated weights

$$\begin{aligned} \bar{v}_{\boldsymbol{i},t} &= \bar{w}_{\boldsymbol{i},t-1} e^{\eta g'_{\boldsymbol{i},t}} \\ \bar{w}_{\boldsymbol{i},t} &= (1-\alpha)\bar{v}_{\boldsymbol{i},t} + \frac{\alpha}{N}\overline{W}_t \end{aligned}$$

where $g'_{\boldsymbol{i},t} = \sum_{e \in \boldsymbol{i}} g'_{e,t}$ and $\overline{W}_t$ is the sum of the total weights of the paths, that is,

$$\overline{W}_t = \sum_{\boldsymbol{i} \in \mathcal{P}} \bar{v}_{\boldsymbol{i},t} = \sum_{\boldsymbol{i} \in \mathcal{P}} \bar{w}_{\boldsymbol{i},t}.$$

Figure 4.4: Bandit algorithm for tracking the shortest path.

Theorem 4.4. (GYÖRGY, LINDER, LUGOSI AND OTTUCSÁK [40]). *For any $\delta \in (0,1)$ and parameters $0 \le \gamma < 1/2$, $\alpha, \beta \in [0,1]$, and $\eta > 0$ satisfying $2\eta K|\mathcal{C}| \le \gamma$, the performance of the algorithm defined above can be bounded, with probability at least $1-\delta$, as*

$$\begin{aligned}\frac{1}{n}\left(\widehat{L}_n - \min_{\mathbf{t},\underline{\boldsymbol{i}}} L(\mathtt{Part}(n,m,\mathbf{t},\underline{\boldsymbol{i}}))\right) \\ \le\ & K\left(\gamma + \eta(1+\beta)K|\mathcal{C}|\right) + \frac{K(m+1)}{n\beta}\ln\frac{|E|(m+1)}{\delta} \\ & + \beta|E| + \frac{1}{n\eta}\ln\left(\frac{N^{m+1}}{\alpha^m(1-\alpha)^{n-m-1}}\right)\ .\end{aligned}$$

In particular, choosing

$$\beta = \sqrt{\frac{K(m+1)}{n|E|}\ln\frac{|E|(m+1)}{\delta}}, \qquad \gamma = 2\eta K|\mathcal{C}|, \qquad \alpha = \frac{m}{n-1},$$

and

$$\eta = \sqrt{\frac{(m+1)\ln N + m\ln\frac{e(n-1)}{m}}{4nK^2|\mathcal{C}|}}$$

we have, for all $n \ge \max\left\{\frac{K(m+1)}{|E|}\ln\frac{|E|(m+1)}{\delta}, 4|\mathcal{C}|D\right\}$,

$$\widehat{L}_n - \min_{\mathbf{t},\underline{\boldsymbol{i}}} L(\mathtt{Part}(n,m,\mathbf{t},\underline{\boldsymbol{i}})) \le 2\sqrt{\frac{K}{n}}\left(\sqrt{4K|\mathcal{C}|D} + \sqrt{|E|(m+1)\ln\frac{|E|(m+1)}{\delta}}\right),$$

where

$$D = (m+1)\ln N + m\left(1 + \ln\frac{n-1}{m}\right).$$

The proof of the theorem is a combination of that of our Theorem 4.1 and Theorem 1 of [38]. We will need the following three lemmas.

Lemma 4.4. *For any $1 \le t \le t' \le n$ and any $\boldsymbol{i} \in \mathcal{P}$,*

$$\frac{\bar{v}_{\boldsymbol{i},t'}}{\bar{w}_{\boldsymbol{i},t-1}} \ge e^{\eta G'_{\boldsymbol{i},[t,t']}}(1-\alpha)^{t'-t}$$

where $G'_{\boldsymbol{i},[t,t']} = \sum_{\tau=t}^{t'} g'_{\boldsymbol{i},\tau}$.

Proof. The proof is a straightforward modification of the one in Herbster and Warmuth [46]. From the definitions of $v_{\boldsymbol{i},t}$ and $w_{\boldsymbol{i},t}$ (see step (d) of the algorithm) it is clear that for any $\tau \ge 1$,

$$\bar{w}_{\boldsymbol{i},\tau} = (1-\alpha)\bar{v}_{\boldsymbol{i},\tau} + \frac{\alpha}{N}\overline{W}_\tau \ge (1-\alpha)e^{\eta g'_{\boldsymbol{i},\tau}}\bar{w}_{\boldsymbol{i},\tau-1}\ .$$

Applying this equation iteratively for $\tau = t, t+1, \ldots, t'-1$, and the definition of $\bar{v}_{\boldsymbol{i},t}$ (step (d)) for $\tau = t'$, we obtain

$$\begin{aligned}\bar{v}_{\boldsymbol{i},t'} = \bar{w}_{\boldsymbol{i},t'-1} e^{\eta g'_{\boldsymbol{i},t'}} &\geq e^{\eta g'_{\boldsymbol{i},t'}} \prod_{\tau=t}^{t'-1} \left((1-\alpha)e^{\eta g'_{\boldsymbol{i},\tau}}\right) \bar{w}_{\boldsymbol{i},t-1} \\ &= e^{\eta G'_{\boldsymbol{i},[t,t']}} (1-\alpha)^{t'-t} \bar{w}_{\boldsymbol{i},t-1}\end{aligned}$$

which implies the statement of the lemma. □

Lemma 4.5. *For any $t \geq 1$ and $\boldsymbol{i}, \boldsymbol{j} \in \mathcal{P}$, we have*

$$\frac{\bar{w}_{\boldsymbol{i},t}}{\bar{v}_{\boldsymbol{j},t}} \geq \frac{\alpha}{N}$$

Proof. By the definition of $\bar{w}_{\boldsymbol{i},t}$ we have

$$\bar{w}_{\boldsymbol{i},t} = (1-\alpha)\bar{v}_{\boldsymbol{i},t} + \frac{\alpha}{N}\overline{W}_t \geq \frac{\alpha}{N}\overline{W}_t \geq \frac{\alpha}{N}\bar{v}_{\boldsymbol{j},t} \ .$$

This completes the proof of the lemma. □

The next lemma is a simple corollary of Lemma 4.1.

Lemma 4.6. *For any $\delta \in (0,1)$, $0 \leq \beta \leq 1$, $t \geq 1$ and $e \in E$ we have*

$$\mathbb{P}\left[G_{e,t} > G'_{e,t} + \frac{1}{\beta}\ln\frac{|E|(m+1)}{\delta}\right] \leq \frac{\delta}{|E|(m+1)} \ .$$

Proof of Theorem 4.4. Again, we upper bound $\ln \overline{W}_n/\overline{W}_0$ the same way as in Theorem 4.1. Then we get

$$\ln\frac{\overline{W}_n}{\overline{W}_0} \ \leq \ \frac{\eta}{1-\gamma}\left(\widehat{G}_n + n|E|\beta\right) + \frac{\eta^2 K(1+\beta)}{1-\gamma}|\mathcal{C}| \max_{\boldsymbol{i}\in\mathcal{P}} G'_{\boldsymbol{i},n} \ . \tag{4.18}$$

Let $\texttt{Part}(n, m, \mathbf{t}, \underline{\boldsymbol{i}})$ be an arbitrary partition. Then the lower bound is obtained as

$$\ln\frac{\overline{W}_n}{\overline{W}_0} = \ln\sum_{\boldsymbol{j}\in\mathcal{P}}\frac{\bar{w}_{\boldsymbol{j},n}}{N} = \ln\sum_{\boldsymbol{j}\in\mathcal{P}}\frac{\bar{v}_{\boldsymbol{j},n}}{N} \geq \ln\frac{\bar{v}_{\boldsymbol{i}_m,n}}{N} \tag{4.19}$$

(recall that $\boldsymbol{i}_m$ denotes the path used in the last segment of the partition). Now $v_{\boldsymbol{i}_m,n}$ can be rewritten in the form of the following telescoping product

$$\bar{v}_{\boldsymbol{i}_m,n} = \bar{w}_{\boldsymbol{i}_0,t_0-1}\frac{\bar{v}_{\boldsymbol{i}_0,t_1-1}}{\bar{w}_{\boldsymbol{i}_0,t_0-1}}\prod_{j=1}^{m}\left(\frac{\bar{w}_{\boldsymbol{i}_j,t_j-1}}{\bar{v}_{\boldsymbol{i}_{j-1},t_j-1}}\frac{\bar{v}_{\boldsymbol{i}_j,t_{j+1}-1}}{\bar{w}_{\boldsymbol{i}_j,t_j-1}}\right).$$

Therefore, applying Lemmas 4.4 and 4.5, we have

$$\begin{aligned}\bar{v}_{\boldsymbol{i}_m,n} &\geq \bar{w}_{\boldsymbol{i}_0,t_0-1}\left(\frac{\alpha}{N}\right)^m \prod_{j=0}^{m}\left((1-\alpha)^{t_{j+1}-1-t_j} e^{\eta G'_{\boldsymbol{i}_j,[t_j,t_{j+1}-1]}}\right)\\ &= \left(\frac{\alpha}{N}\right)^m e^{\eta G'(\mathtt{Part}(n,m,\mathbf{t},\underline{\boldsymbol{i}}))}(1-\alpha)^{n-m-1}.\end{aligned}$$

Combining the lower bound with the upper bound (4.18), we have

$$\begin{aligned}&\ln\left(\frac{\alpha^m(1-\alpha)^{n-m-1}}{N^{m+1}}\right) + \max_{\mathbf{t},\underline{\boldsymbol{i}}} \eta G'(\mathtt{Part}(n,m,\mathbf{t},\underline{\boldsymbol{i}}))\\ &\quad\leq \tfrac{\eta}{1-\gamma}\left(\widehat{G}_n + n|E|\beta\right) + \tfrac{\eta^2 K(1+\beta)}{1-\gamma}|\mathcal{C}|\max_{\boldsymbol{i}\in\mathcal{P}} G'_{\boldsymbol{i},n}\ ,\end{aligned}$$

where we used the fact that $\mathtt{Part}(n,m,\mathbf{t},\underline{\boldsymbol{i}})$ is an arbitrary partition. After rearranging and using $\max_{\boldsymbol{i}\in\mathcal{P}} G'_{\boldsymbol{i},n} \leq \max_{\mathbf{t},\underline{\boldsymbol{i}}} G'(\mathtt{Part}(n,m,\mathbf{t},\underline{\boldsymbol{i}}))$ we get

$$\begin{aligned}\widehat{G}_n &\geq (1-\gamma-\eta K(1+\beta)|\mathcal{C}|)\max_{\mathbf{t},\underline{\boldsymbol{i}}} G'(\mathtt{Part}(n,m,\mathbf{t},\underline{\boldsymbol{i}}))\\ &\quad -n|E|\beta - \frac{1-\gamma}{\eta}\ln\left(\frac{N^{m+1}}{\alpha^m(1-\alpha)^{n-m-1}}\right).\end{aligned}$$

Now since $1-\gamma-\eta K(1+\beta)|\mathcal{C}| \geq 0$, by the assumptions of the theorem and from Lemma 4.6 with an application of the union bound we obtain that, with probability at least $1-\delta$,

$$\begin{aligned}\widehat{G}_n \geq{}& (1-\gamma-\eta K(1+\beta)|\mathcal{C}|)\left(\max_{\mathbf{t},\underline{\boldsymbol{i}}} G(\mathtt{Part}(n,m,\mathbf{t},\underline{\boldsymbol{i}})) - \frac{K(m+1)}{\beta}\ln\frac{|E|(m+1)}{\delta}\right)\\ &- n|E|\beta - \frac{1-\gamma}{\eta}\ln\left(\frac{N^{m+1}}{\alpha^m(1-\alpha)^{n-m-1}}\right).\end{aligned}$$

Since $\widehat{G}_n = Kn - \widehat{L}_n$ and $G(\mathtt{Part}(n,m,\mathbf{t},\underline{\boldsymbol{i}})) = Kn - L(\mathtt{Part}(n,m,\mathbf{t},\underline{\boldsymbol{i}}))$, we have

$$\begin{aligned}\widehat{L}_n &\leq (1-\gamma-\eta K(1+\beta)|\mathcal{C}|)\min_{\mathbf{t},\boldsymbol{i}} L(\mathtt{Part}(n,m,\mathbf{t},\underline{\boldsymbol{i}})) + Kn\left(\gamma+\eta(1+\beta)K|\mathcal{C}|\right)\\ &\quad + (1-\gamma-\eta(1+\beta)K|\mathcal{C}|)\frac{K(m+1)}{\beta}\ln\frac{|E|(m+1)}{\delta} + n|E|\beta\\ &\quad + \frac{1}{\eta}\ln\left(\frac{N^{m+1}}{\alpha^m(1-\alpha)^{n-m-1}}\right).\end{aligned}$$

This implies that, with probability at least $1-\delta$,

$$\begin{aligned}&\widehat{L}_n - \min_{\mathbf{t},\underline{\boldsymbol{i}}} L(\mathtt{Part}(n,m,\mathbf{t},\underline{\boldsymbol{i}}))\\ &\quad\leq Kn\left(\gamma+\eta(1+\beta)K|\mathcal{C}|\right) + \frac{K(m+1)}{\beta}\ln\frac{|E|(m+1)}{\delta}\\ &\qquad + n|E|\beta + \frac{1}{\eta}\ln\left(\frac{N^{m+1}}{\alpha^m(1-\alpha)^{n-m-1}}\right).\end{aligned} \tag{4.20}$$

To prove the second statement, let $H(p) = -p\ln p - (1-p)\ln(1-p)$ and $D(p \parallel q) = p\ln\frac{p}{q} + (1-p)\ln\frac{1-p}{1-q}$. Optimizing the value of α in the last term of (4.20) gives

$$\begin{aligned}
&\frac{1}{\eta}\ln\left(\frac{N^{m+1}}{\alpha^m(1-\alpha)^{n-m-1}}\right)\\
&\quad= \frac{1}{\eta}\left((m+1)\ln(N) + m\ln\frac{1}{\alpha} + (n-m-1)\ln\frac{1}{1-\alpha}\right)\\
&\quad= \frac{1}{\eta}\big((m+1)\ln(N) + (n-1)(D_b(\alpha^* \parallel \alpha) + H_b(\alpha^*))\big)
\end{aligned}$$

where $\alpha^* = \frac{m}{n-1}$. For $\alpha = \alpha^*$ we obtain

$$\begin{aligned}
&\frac{1}{\eta}\ln\left(\frac{N^{m+1}}{\alpha^m(1-\alpha)^{n-m-1}}\right)\\
&\quad= \frac{1}{\eta}\left((m+1)\ln(N) + (n-1)(H_b(\alpha^*))\right)\\
&\quad= \frac{1}{\eta}\left((m+1)\ln(N) + m\ln((n-1)/m)\right.\\
&\qquad\left.+(n-m-1)\ln(1+m/(n-m-1))\right)\\
&\quad\le \frac{1}{\eta}\left((m+1)\ln(N) + m\ln((n-1)/m) + m\right)\\
&\quad= \frac{1}{\eta}\left((m+1)\ln(N) + m\ln\frac{e(n-1)}{m}\right) \stackrel{\text{def}}{=} \frac{1}{\eta}D
\end{aligned}$$

where the inequality follows since $\ln(1+x) \le x$ for $x > 0$. Therefore

$$\begin{aligned}
&\widehat{L}_n - \min_{\mathbf{t},\underline{\boldsymbol{i}}} L(\mathtt{Part}(n,m,\mathbf{t},\underline{\boldsymbol{i}}))\\
&\quad\le Kn\left(\gamma + \eta(1+\beta)K|\mathcal{C}|\right) + \frac{K(m+1)}{\beta}\ln\frac{|E|(m+1)}{\delta} + n|E|\beta + \frac{1}{\eta}D\ .
\end{aligned}$$

which is the first statement of the theorem. Setting

$$\beta = \sqrt{\frac{K(m+1)}{n|E|}\ln\frac{|E|(m+1)}{\delta}},\ \gamma = 2\eta K|\mathcal{C}|, \text{ and } \eta = \sqrt{\frac{D}{4nK^2|\mathcal{C}|}}$$

results in the second statement of the theorem, that is,

$$\begin{aligned}
&\widehat{L}_n - \min_{\mathbf{t},\underline{\boldsymbol{i}}} L(\mathtt{Part}(n,m,\mathbf{t},\underline{\boldsymbol{i}}))\\
&\qquad\le 2\sqrt{nK}\left(\sqrt{4K|\mathcal{C}|D} + \sqrt{|E|(m+1)\ln\frac{|E|(m+1)}{\delta}}\right). \quad \square
\end{aligned}$$

Similarly to [38], the proposed algorithm has an alternative version, which is efficiently computable:

AN ALTERNATIVE BANDIT ALGORITHM FOR TRACKING SHORTEST PATHS

For $t = 1$, choose $\boldsymbol{I}_1$ uniformly from the set $\mathcal{P}$. For $t \geq 2$,

(a) Draw a Bernoulli random variable Γ_t with $\mathbb{P}(\Gamma_t = 1) = \gamma$.

(b) If $\Gamma_t = 1$, then choose $\boldsymbol{I}_t$ uniformly from $\mathcal{C}$.

(c) If $\Gamma_t = 0$,

(c1) choose τ_t randomly according to the distribution

$$\mathbb{P}\{\tau_t = t'\} = \begin{cases} \frac{(1-\alpha)^{t-1} Z_{1,t-1}}{W_{t-1}} & \text{for } t' = 1 \\ \frac{\alpha(1-\alpha)^{t-t'} W_{t'} Z_{t',t-1}}{N W_t} & \text{for } t' = 2, \ldots, t \end{cases}$$

where $Z_{t',t-1} = \sum_{\boldsymbol{i} \in \mathcal{P}} e^{\eta G'_{\boldsymbol{i},[t',t-1]}}$ for $t' = 1, \ldots, t-1$, and $Z_{t,t-1} = N$;

(c2) given $\tau_t = t'$, choose $\boldsymbol{I}_t$ randomly according to the probabilities

$$\mathbb{P}\{I_t = \boldsymbol{i} | \tau_t = t'\} = \begin{cases} \frac{e^{\eta G'_{\boldsymbol{i},[t',t-1]}}}{Z_{t',t-1}} & \text{for } t' = 1, \ldots, t-1 \\ \frac{1}{N} & \text{for } t' = t. \end{cases}$$

Figure 4.5: An alternative version of the bandit algorithm for tracking shortest path.

With a slight modification of the proof of Theorem 2 in [38], it can be shown that the alternative and the original algorithms are equivalent. Moreover, in a way completely analogous to [38], in this alternative formulation of the algorithm one can compute the probabilities the normalization factors $Z_{t',t-1}$ efficiently, as the baseline bandit algorithm for shortest paths has an $O(n|E|)$ time complexity by Theorem 4.2. Therefore the factors $\overline{W}_t$ and hence the probabilities $\mathbb{P}\{\boldsymbol{I}_t = \boldsymbol{i} | \tau_t = t'\}$ can also be computed efficiently as in [38]. In particular, it follows from Theorem 3 of [38] that the time complexity of the alternative bandit algorithm for tracking the shortest path is $O(n^2|E|)$.

4.6 An algorithm for the restricted multi-armed bandit problem

In this section we consider the situation where the decision maker receives information only about the performance of the whole chosen path, but the individual edge losses are not available. That is, the forecaster has access to the sum $\ell_{\boldsymbol{I}_t,t}$ of losses over the chosen path $\boldsymbol{I}_t$ but not to the losses $\{\ell_{e,t}\}_{e \in \boldsymbol{I}_t}$ of the edges belonging to $\boldsymbol{I}_t$.

This is the problem formulation considered by McMahan and Blum [54] and Awerbuch and Kleinberg [11]. McMahan and Blum provided a relatively simple algorithm whose regret is at most of the order of $n^{-1/4}$, while Awerbuch and Kleinberg gave a more complex algorithm to achieve $O(n^{-1/3})$ regret. In this section we combine the strengths of these papers, and propose a simple algorithm with regret at most of the order of $n^{-1/3}$. Moreover, our bound holds with high probability, while the above-mentioned papers prove bounds for the expected regret only. The proposed algorithm uses ideas very similar to those of McMahan and Blum [54]. The algorithm alternates between choosing a path from a "basis" $\boldsymbol{B}$ to obtain unbiased estimates of the loss (exploration step), and choosing a path according to exponential weighting based on these estimates.

A simple way to describe a path $\boldsymbol{i} \in \mathcal{P}$ is a binary row vector with $|E|$ components which are indexed by the edges of the graph such that, for each $e \in E$, the eth entry of the vector is 1 if $e \in \boldsymbol{i}$ and 0 otherwise. With a slight abuse of notation we will also denote by $\boldsymbol{i}$ the binary row vector representing path $\boldsymbol{i}$. In the previous sections, where the loss of each edge along the chosen path is available to the decision maker, the complexity stemming from the large number of paths was reduced by representing all information in terms of the edges, as the set of edges spans the set of paths. That is, the vector corresponding to a given path can be expressed as the linear combination of the unit vectors associated with the edges (the eth component of the unit vector representing edge e is 1, while the other components are 0). While the losses corresponding to such a spanning set are not observable in the restricted setting of this section, one can choose a subset of $\mathcal{P}$ that forms a *basis*, that is, a collection of b paths which are linearly independent and each path in $\mathcal{P}$ can be expressed as a linear combination of the paths in the basis. We denote by $\boldsymbol{B}$ the $b \times |E|$ matrix whose rows $\boldsymbol{b}^1, \ldots, \boldsymbol{b}^b$ represent the paths in the basis. Note that b is equal to the maximum number of linearly independent vectors in $\{\boldsymbol{i} : \boldsymbol{i} \in \mathcal{P}\}$, so $b \leq |E|$.

Let $\boldsymbol{\ell}_t^{(E)}$ denote the (column) vector of the edge losses $\{\ell_{e,t}\}_{e\in E}$ at time t, and let $\boldsymbol{\ell}_t^{(\boldsymbol{B})} = (\ell_{\boldsymbol{b}^1,t}, \ldots, \ell_{\boldsymbol{b}^b,t})^T$ be a b-dimensional column vector whose components are the losses of the paths in the basis $\boldsymbol{B}$ at time t. If $\alpha_{\boldsymbol{b}^1}^{(\boldsymbol{i},\boldsymbol{B})}, \ldots, \alpha_{\boldsymbol{b}^b}^{(\boldsymbol{i},\boldsymbol{B})}$ are the coefficients in the linear combination of the basis paths expressing path $\boldsymbol{i} \in \mathcal{P}$, that is, $\boldsymbol{i} = \sum_{j=1}^b \alpha_{\boldsymbol{b}^j}^{(\boldsymbol{i},\boldsymbol{B})} \boldsymbol{b}^j$, then the loss of path $\boldsymbol{i} \in \mathcal{P}$ at time t is given by

$$\ell_{\boldsymbol{i},t} = \langle \boldsymbol{i}, \boldsymbol{\ell}_t^{(E)} \rangle = \sum_{j=1}^b \alpha_{\boldsymbol{b}^j}^{(\boldsymbol{i},\boldsymbol{B})} \langle \boldsymbol{b}^j, \boldsymbol{\ell}_t^{(E)} \rangle = \sum_{j=1}^b \alpha_{\boldsymbol{b}^j}^{(\boldsymbol{i},\boldsymbol{B})} \ell_{\boldsymbol{b}^j,t} \tag{4.21}$$

where $\langle \cdot, \cdot \rangle$ denotes the standard inner product in $\mathbb{R}^{|E|}$. In the algorithm we obtain estimates $\tilde{\ell}_{\boldsymbol{b}^j,t}$ of the losses of the basis paths and use (4.21) to estimate the loss of any $\boldsymbol{i} \in \mathcal{P}$ as

$$\tilde{\ell}_{\boldsymbol{i},t} = \sum_{j=1}^b \alpha_{\boldsymbol{b}^j}^{(\boldsymbol{i},\boldsymbol{B})} \tilde{\ell}_{\boldsymbol{b}^j,t} \ . \tag{4.22}$$

It is algorithmically advantageous to calculate the estimated path losses $\tilde{\ell}_{\boldsymbol{i},t}$ from an intermediate estimate of the individual edge losses. Let $\boldsymbol{B}^+$ denote the Moore-Penrose inverse of $\boldsymbol{B}$ defined by $\boldsymbol{B}^+ = \boldsymbol{B}^T(\boldsymbol{B}\boldsymbol{B}^T)^{-1}$, where $\boldsymbol{B}^T$ denotes the transpose of $\boldsymbol{B}$ and $\boldsymbol{B}\boldsymbol{B}^T$ is invertible since the rows of $\boldsymbol{B}$ are linearly independent. (Note that $\boldsymbol{B}\boldsymbol{B}^+ = I_b$, the $b \times b$ identity matrix, and $\boldsymbol{B}^+ = \boldsymbol{B}^{-1}$ if $b = |E|$.). Then letting $\tilde{\boldsymbol{\ell}}_t^{(\boldsymbol{B})} = (\tilde{\ell}_{\boldsymbol{b}^1,t}, \ldots, \tilde{\ell}_{\boldsymbol{b}^b,t})^T$ and

$$\tilde{\boldsymbol{\ell}}_t^{(E)} = \boldsymbol{B}^+ \tilde{\boldsymbol{\ell}}_t^{(\boldsymbol{B})}$$

it is easy to see that $\tilde{\ell}_{\boldsymbol{i},t}$ in (4.22) can be obtained as $\tilde{\ell}_{\boldsymbol{i},t} = \langle \boldsymbol{i}, \tilde{\boldsymbol{\ell}}_t^{(E)} \rangle$, or equivalently

$$\tilde{\ell}_{\boldsymbol{i},t} = \sum_{e\in\boldsymbol{i}} \tilde{\ell}_{e,t}.$$

This form of the path losses allows for an efficient implementation of exponential weighting via dynamic programming [69].

To analyze the algorithm we need an upper bound on the magnitude of the coefficients $\alpha_{\boldsymbol{b}^j}^{(\boldsymbol{i},\boldsymbol{B})}$. For this, we invoke the definition of a barycentric spanner from [11]: the basis $\boldsymbol{B}$ is called a *C-barycentric spanner* if $|\alpha_{\boldsymbol{b}^j}^{(\boldsymbol{i},\boldsymbol{B})}| \le C$ for all $\boldsymbol{i} \in \mathcal{P}$ and $j = 1, \ldots, b$. Awerbuch and Kleinberg [11] show that a 1-barycentric spanner exists if $\boldsymbol{B}$ is a square matrix (i.e., $b = |E|$) and give a low-complexity algorithm which finds a C-barycentric spanner for $C > 1$. We use their technique to show that a 1-barycentric spanner also exists in case of a non-square $\boldsymbol{B}$, when the basis is chosen to maximize the absolute value of the determinant of $\boldsymbol{B}\boldsymbol{B}^T$. As before, b denotes the maximum number of linearly independent vectors (paths) in $\mathcal{P}$.

Lemma 4.7. *For a directed acyclic graph, the set of paths $\mathcal{P}$ between two dedicated nodes has a 1-barycentric spanner. Moreover, let $\boldsymbol{B}$ be a $b \times |E|$ matrix with rows from $\mathcal{P}$ such*

that $\det[\boldsymbol{B}\boldsymbol{B}^T] \neq 0$. *If* $\boldsymbol{B}_{-j,\boldsymbol{i}}$ *is the matrix obtained from* $\boldsymbol{B}$ *by replacing its* j*th row by* $\boldsymbol{i} \in \mathcal{P}$ *and*

$$\left|\det\left[\boldsymbol{B}_{-j,\boldsymbol{i}}\boldsymbol{B}_{-j,\boldsymbol{i}}^T\right]\right| \leq C^2 \left|\det\left[\boldsymbol{B}\boldsymbol{B}^T\right]\right| \tag{4.23}$$

for all $j = 1, \ldots, b$ *and* $\boldsymbol{i} \in \mathcal{P}$*, then* $\boldsymbol{B}$ *is a* C*-barycentric spanner.*

Proof. Let $\boldsymbol{B}$ be a basis of $\mathcal{P}$ with rows $\boldsymbol{b}^1, \ldots, \boldsymbol{b}^b \in \mathcal{P}$ that maximizes $|\det[\boldsymbol{B}\boldsymbol{B}^T]|$. Then, for any path $\boldsymbol{i} \in \mathcal{P}$, we have $\boldsymbol{i} = \sum_{j=1}^b \alpha_{\boldsymbol{b}^j}^{(\boldsymbol{i},\boldsymbol{B})} \boldsymbol{b}^j$ for some coefficients $\{\alpha_{\boldsymbol{b}^j}^{(\boldsymbol{i},\boldsymbol{B})}\}$. Now for the matrix $\boldsymbol{B}_{-1,\boldsymbol{i}} = [\boldsymbol{i}^T, (\boldsymbol{b}^2)^T, \ldots, (\boldsymbol{b}^b)^T]^T$ we have

$$\begin{aligned}
&\left|\det\left[\boldsymbol{B}_{-1,\boldsymbol{i}}\boldsymbol{B}_{-1,\boldsymbol{i}}^T\right]\right| \\
&= \left|\det\left[\boldsymbol{B}_{-1,\boldsymbol{i}}\boldsymbol{i}^T, \boldsymbol{B}_{-1,\boldsymbol{i}}(\boldsymbol{b}^2)^T, \boldsymbol{B}_{-1,\boldsymbol{i}}(\boldsymbol{b}^3)^T, \ldots, \boldsymbol{B}_{-1,\boldsymbol{i}}(\boldsymbol{b}^b)^T\right]\right| \\
&= \left|\det\left[\left(\sum_{j=1}^b \alpha_{\boldsymbol{b}^j}^{(\boldsymbol{i},\boldsymbol{B})} \boldsymbol{B}_{-1,\boldsymbol{i}}\boldsymbol{b}^j\right)^T, \boldsymbol{B}_{-1,\boldsymbol{i}}(\boldsymbol{b}^2)^T, \boldsymbol{B}_{-1,\boldsymbol{i}}(\boldsymbol{b}^3)^T, \ldots, \boldsymbol{B}_{-1,\boldsymbol{i}}(\boldsymbol{b}^b)^T\right]\right| \\
&= \left|\sum_{j=1}^b \alpha_{\boldsymbol{b}^j}^{(\boldsymbol{i},\boldsymbol{B})} \det\left[\boldsymbol{B}_{-1,\boldsymbol{i}}(\boldsymbol{b}^j)^T, \boldsymbol{B}_{-1,\boldsymbol{i}}(\boldsymbol{b}^2)^T, \boldsymbol{B}_{-1,\boldsymbol{i}}(\boldsymbol{b}^3)^T, \ldots, \boldsymbol{B}_{-1,\boldsymbol{i}}(\boldsymbol{b}^b)^T\right]\right| \\
&= |\alpha_{\boldsymbol{b}^1}^{(\boldsymbol{i},\boldsymbol{B})}| \left|\det\left[\boldsymbol{B}_{-1,\boldsymbol{i}}\boldsymbol{B}^T\right]\right| \\
&= \left(\alpha_{\boldsymbol{b}^1}^{(\boldsymbol{i},\boldsymbol{B})}\right)^2 \left|\det\left[\boldsymbol{B}\boldsymbol{B}^T\right]\right|
\end{aligned}$$

where last equality follows by the same argument the penultimate equality was obtained. Repeating the same argument for $\boldsymbol{B}_{-j,\boldsymbol{i}}$, $j = 2, \ldots, b$ we obtain

$$\left|\det\left[\boldsymbol{B}_{-j,\boldsymbol{i}}\boldsymbol{B}_{-j,\boldsymbol{i}}^T\right]\right| = \left(\alpha_{\boldsymbol{b}^j}^{(\boldsymbol{i},\boldsymbol{B})}\right)^2 \left|\det\left[\boldsymbol{B}\boldsymbol{B}^T\right]\right|. \tag{4.24}$$

Thus the maximal property of $|\det[\boldsymbol{B}\boldsymbol{B}^T]|$ implies $|\alpha_{\boldsymbol{b}^j}^{(\boldsymbol{i},\boldsymbol{B})}| \leq 1$ for all $j = 1, \ldots, b$. The second statement follows trivially from (4.23) and (4.24). □

Awerbuch and Kleinberg [11, Proposition 2.4] also present an iterative algorithm to find a C-barycentric spanner if $\boldsymbol{B}$ is a square matrix. Their algorithm has two parts. First, starting from the identity matrix, the algorithm replaces sequentially the rows of the matrix in each step by maximizing the determinant with respect to the given row. This is done by calling b times an optimization oracle, to compute $\arg\max_{\boldsymbol{i}\in\mathcal{P}} |\det[\boldsymbol{B}_{-j,\boldsymbol{i}}]|$ for $j = 1, 2, \ldots, b$. In the second part the algorithm replaces an arbitrarily row j of the matrix in each iteration with some $\boldsymbol{i} \in \mathcal{P}$ if $|\det[\boldsymbol{B}_{-j,\boldsymbol{i}}]| > C|\det[\boldsymbol{B}]|$. It is shown that the oracle is called in the second part $O(b \log_C b)$ times for $C > 1$. In case $\boldsymbol{B}$ is not a square matrix, the algorithm carries over if we have access to an alternative optimization oracle that can compute $\arg\max_{\boldsymbol{i}\in\mathcal{P}} |\det[\boldsymbol{B}_{-j,\boldsymbol{i}}\boldsymbol{B}_{-j,\boldsymbol{i}}^T]|$: In the first b steps, all the rows of the matrix are replaced (first part), then we can iteratively replace one row in each step, using the oracle, to maximize the determinant $|\det[\boldsymbol{B}_{-j,\boldsymbol{i}}\boldsymbol{B}_{-j,\boldsymbol{i}}^T]|$ in $\boldsymbol{i}$ until (4.23) is satisfied for all j and $\boldsymbol{i}$. By Lemma 4.7, this results is a C-barycentric spanner. Similarly to [11, Lemma 2.5], it can be shown that the alternative optimization oracle is called $O(b \log_C b)$ times for $C > 1$.

For simplicity (to avoid carrying the constant C), assume that we have a 2-barycentric spanner $\boldsymbol{B}$. Based on the ideas of label efficient prediction, the next algorithm gives a simple solution to the restricted shortest path problem. The algorithm is very similar to that of the algorithm in the label efficient case, but here we cannot estimate the edge losses directly. Therefore, we query the loss of a (random) basis vector from time to time, and create unbiased estimates $\tilde{\ell}_{\boldsymbol{b}^j,t}$ of the losses of basis paths $\ell_{\boldsymbol{b}^j,t}$, which are then transformed into edge-loss estimates.

The performance of the algorithm is analyzed in the next theorem. The proof follows the argument of Cesa-Bianchi *et al.* [22], but we also have to deal with some additional technical difficulties. Note that in the theorem we do not assume that all paths between u and v have equal length.

Theorem 4.5. (GYÖRGY, LINDER, LUGOSI AND OTTUCSÁK [40]). *Let K denote the length of the longest path in the graph. For any $\delta \in (0,1)$, parameters $0 < \varepsilon \le \frac{1}{K}$ and $\eta > 0$ satisfying $\eta \le \varepsilon^2$, and $n \ge \frac{8b}{\varepsilon^2} \ln \frac{4bN}{\delta}$, the performance of the algorithm defined above can be bounded, with probability at least $1-\delta$, as*

$$\widehat{L}_n - \min_{\boldsymbol{i}\in\mathcal{P}} L_{\boldsymbol{i},n} \le K\left(\frac{\eta b}{\varepsilon}Kn + \sqrt{\frac{n}{2}\ln\frac{4}{\delta}} + n\varepsilon + \frac{\sqrt{2n\varepsilon\ln\frac{4}{\delta}}}{K} + \frac{16}{3}b\sqrt{2n\frac{b}{\varepsilon}\ln\frac{4bN}{\delta}}\right) + \frac{\ln N}{\eta}$$

In particular, choosing

$$\varepsilon = \left(\frac{Kb}{n}\ln\frac{4bN}{\delta}\right)^{1/3} \qquad \text{and} \qquad \eta = \varepsilon^2$$

we obtain

$$\widehat{L}_n - \min_{\boldsymbol{i}\in\mathcal{P}} L_{\boldsymbol{i},n} \le 9.1K^2b\left(Kb\ln(4bN/\delta)\right)^{1/3} n^{2/3} \ .$$

The theorem is proved using the following two lemmas. The first one is an easy consequence of Bernstein's inequality:

Lemma 4.8. *Under the assumptions of Theorem 4.5, the probability that the algorithm queries the basis more than $n\varepsilon + \sqrt{2n\varepsilon\ln\frac{4}{\delta}}$ times is at most $\delta/4$.*

Using the estimated loss of a path $\boldsymbol{i} \in \mathcal{P}$ given in (4.22), we can estimate the cumulative loss of $\boldsymbol{i}$ up to time n as

$$\tilde{L}_{\boldsymbol{i},n} = \sum_{t=1}^{n} \widetilde{\ell}_{\boldsymbol{i},t} \ .$$

The next lemma demonstrates the quality of these estimates.

A BANDIT ALGORITHM FOR THE RESTRICTED SHORTEST PATH PROBLEM

Parameters: $0 < \varepsilon, \eta \leq 1$.
Initialization: Set $w_{e,0} = 1$ for each $e \in E$, $\bar{w}_{\boldsymbol{i},0} = 1$ for each $\boldsymbol{i} \in \mathcal{P}$, $\overline{W}_0 = N$. Fix a basis $\boldsymbol{B}$, which is a 2-barycentric spanner. For each round $t = 1, 2, \ldots$

(a) Draw a Bernoulli random variable S_t such that $\mathbb{P}(S_t = 1) = \varepsilon$;

(b) If $S_t = 1$, then choose the path $\boldsymbol{I}_t$ uniformly from the basis $\boldsymbol{B}$. If $S_t = 0$, then choose $\boldsymbol{I}_t$ according to the distribution $\{p_{\boldsymbol{i},t}\}$, defined by

$$p_{\boldsymbol{i},t} = \frac{\bar{w}_{\boldsymbol{i},t-1}}{\overline{W}_{t-1}} \ .$$

(c) Calculate the estimated loss of all edges according to

$$\tilde{\boldsymbol{\ell}}_t^{(E)} = \boldsymbol{B}^{+} \tilde{\boldsymbol{\ell}}_t^{(\boldsymbol{B})} \ ,$$

where $\tilde{\boldsymbol{\ell}}_t^{(E)} = \{\tilde{\ell}_{e,t}^{(E)}\}_{e \in E}$, and $\tilde{\boldsymbol{\ell}}_t^{(\boldsymbol{B})} = (\tilde{\ell}_{\boldsymbol{b}^1,t}^{(\boldsymbol{B})}, \ldots, \tilde{\ell}_{\boldsymbol{b}^b,t}^{(\boldsymbol{B})})$ is the vector of the estimated losses

$$\tilde{\ell}_{\boldsymbol{b}^j,t} = \frac{S_t}{\varepsilon} \ell_{\boldsymbol{b}^j,t} \mathbb{I}_{\{\boldsymbol{I}_t = \boldsymbol{b}^j\}} b$$

for $j = 1, \ldots, b$.

(d) Compute the updated weights

$$\begin{aligned} w_{e,t} &= w_{e,t-1} e^{-\eta \tilde{\ell}_{e,t}}, \\ \bar{w}_{\boldsymbol{i},t} &= \prod_{e \in \boldsymbol{i}} w_{e,t} = \bar{w}_{\boldsymbol{i},t-1} e^{-\eta \sum_{e \in \boldsymbol{i}} \tilde{\ell}_{e,t}}, \end{aligned}$$

and the sum of the total weights of the paths

$$\overline{W}_t = \sum_{\boldsymbol{i} \in \mathcal{P}} \bar{w}_{\boldsymbol{i},t} \ .$$

Figure 4.6: Bandit algorithm for the restricted shortest path problem.

Lemma 4.9. *Let $0 < \delta < 1$ and assume $n \geq \frac{8b}{\varepsilon} \ln \frac{4bN}{\delta}$. For any $\boldsymbol{i} \in \mathcal{P}$, with probability at least $1 - \delta/4$,*

$$\sum_{t=1}^{n} \sum_{\boldsymbol{i} \in \mathcal{P}} p_{\boldsymbol{i},t} \ell_{\boldsymbol{i},t} - \sum_{t=1}^{n} \sum_{\boldsymbol{i} \in \mathcal{P}} p_{\boldsymbol{i},t} \widetilde{\ell}_{\boldsymbol{i},t} \leq \frac{8}{3} b \sqrt{2n \frac{bK^2}{\varepsilon} \ln \frac{4b}{\delta}} \ .$$

Furthermore, with probability at least $1 - \delta/(4N)$,

$$\tilde{L}_{\boldsymbol{i},n} - L_{\boldsymbol{i},n} \leq \frac{8}{3} b \sqrt{2n \frac{bK^2}{\varepsilon} \ln \frac{4bN}{\delta}} \ .$$

Proof. We may write

$$\begin{aligned}
\sum_{t=1}^{n} \sum_{\boldsymbol{i} \in \mathcal{P}} p_{\boldsymbol{i},t} \ell_{\boldsymbol{i},t} - \sum_{t=1}^{n} \sum_{\boldsymbol{i} \in \mathcal{P}} p_{\boldsymbol{i},t} \widetilde{\ell}_{\boldsymbol{i},t} &= \sum_{t=1}^{n} \sum_{\boldsymbol{i} \in \mathcal{P}} p_{\boldsymbol{i},t} \sum_{j=1}^{b} \alpha_{\boldsymbol{b}^j}^{(\boldsymbol{i},\boldsymbol{B})} \left(\ell_{\boldsymbol{b}^j,t} - \widetilde{\ell}_{\boldsymbol{b}^j,t} \right) \\
&= \sum_{j=1}^{b} \sum_{t=1}^{n} \left[\sum_{\boldsymbol{i} \in \mathcal{P}} p_{\boldsymbol{i},t} \alpha_{\boldsymbol{b}^j}^{(\boldsymbol{i},\boldsymbol{B})} \left(\ell_{\boldsymbol{b}^j,t} - \widetilde{\ell}_{\boldsymbol{b}^j,t} \right) \right] \\
&\stackrel{\text{def}}{=} \sum_{j=1}^{b} \sum_{t=1}^{n} X_{\boldsymbol{b}^j,t} \ . \qquad (4.25)
\end{aligned}$$

Note that for any $\boldsymbol{b}^j$, $X_{\boldsymbol{b}^j,t}$, $t = 1, 2, \ldots$ is a martingale difference sequence with respect to $(\boldsymbol{I}_t, S_t)$, $t = 1, 2, \ldots$ as $\mathbb{E}_t \widetilde{\ell}_{\boldsymbol{b},t} = \ell_{\boldsymbol{b},t}$. Also,

$$\mathbb{E}_t[X_{\boldsymbol{b}^j,t}^2] \leq \left(\sum_{\boldsymbol{i} \in \mathcal{P}} p_{\boldsymbol{i},t} \alpha_{\boldsymbol{b}^j}^{(\boldsymbol{i},\boldsymbol{B})} \right)^2 \mathbb{E}_t \left[\left(\widetilde{\ell}_{\boldsymbol{b}^j,t} \right)^2 \right] \leq \sum_{\boldsymbol{i} \in \mathcal{P}} p_{\boldsymbol{i},t} \left(\alpha_{\boldsymbol{b}^j}^{(\boldsymbol{i},\boldsymbol{B})} \right)^2 \frac{K^2 b}{\varepsilon} \leq 4 \frac{K^2 b}{\varepsilon} \qquad (4.26)$$

and

$$|X_{\boldsymbol{b}^j,t}| \leq \left| \sum_{\boldsymbol{i} \in \mathcal{P}} p_{\boldsymbol{i},t} \alpha_{\boldsymbol{b}^j}^{(\boldsymbol{i},\boldsymbol{B})} \right| \left| \ell_{\boldsymbol{b}^j,t} - \widetilde{\ell}_{\boldsymbol{b}^j,t} \right| \leq \sum_{\boldsymbol{i} \in \mathcal{P}} p_{\boldsymbol{i},t} \left| \alpha_{\boldsymbol{b}^j}^{(\boldsymbol{i},\boldsymbol{B})} \right| \frac{Kb}{\varepsilon} \leq 2 \frac{Kb}{\varepsilon} \qquad (4.27)$$

where the last inequalities in both cases follow from the fact that $\boldsymbol{B}$ is a 2-barycentric spanner. Then, using Bernstein's inequality for martingale differences (Lemma 4.2), we have, for any fixed $\boldsymbol{b}^j$,

$$\mathbb{P} \left[\sum_{t=1}^{n} X_{\boldsymbol{b}^j,t} \geq \frac{8}{3} \sqrt{2n \frac{bK^2}{\varepsilon} \ln \frac{4b}{\delta}} \right] \leq \frac{\delta}{4b} \qquad (4.28)$$

where we used (4.26), (4.27) and the assumption of the lemma on n. The proof of the first statement is finished with an application of the union bound and its combination with (4.25).

For the second statement we use a similar argument, that is,

$$\begin{aligned}\sum_{t=1}^{n}(\widetilde{\ell}_{\boldsymbol{i},t}-\ell_{\boldsymbol{i},t})=\sum_{j=1}^{b}\alpha_{\boldsymbol{b}^j}^{(\boldsymbol{i},\boldsymbol{B})}\sum_{t=1}^{n}(\widetilde{\ell}_{\boldsymbol{b}^j,t}-\ell_{\boldsymbol{b}^j,t}) &\leq \sum_{j=1}^{b}\left|\alpha_{\boldsymbol{b}^j}^{(\boldsymbol{i},\boldsymbol{B})}\right|\left|\sum_{t=1}^{n}(\widetilde{\ell}_{\boldsymbol{b}^j,t}-\ell_{\boldsymbol{b}^j,t})\right| \\ &\leq 2\sum_{j=1}^{b}\left|\sum_{t=1}^{n}(\widetilde{\ell}_{\boldsymbol{b}^j,t}-\ell_{\boldsymbol{b}^j,t})\right| . \qquad (4.29)\end{aligned}$$

Now applying Lemma 4.2 for a fixed $\boldsymbol{b}^j$ we get

$$\mathbb{P}\left[\sum_{t=1}^{n}(\widetilde{\ell}_{\boldsymbol{b}^j,t}-\ell_{\boldsymbol{b}^j,t})\geq\frac{4}{3}\sqrt{2n\frac{K^2b}{\varepsilon}\ln\frac{4bN}{\delta}}\right]\leq\frac{\delta}{4bN} \qquad (4.30)$$

because of $\mathbb{E}_t[(\widetilde{\ell}_{\boldsymbol{b}^j,t}-\ell_{\boldsymbol{b}^j,t})^2]\leq\frac{K^2b}{\varepsilon}$ and $-K\leq\widetilde{\ell}_{\boldsymbol{b}^j,t}-\ell_{\boldsymbol{b}^j,t}\leq K\left(\frac{b}{\varepsilon}-1\right)$. The proof is completed by applying the union bound to (4.30) and combining the result with (4.29). □

Proof of Theorem 4.5. Similarly to earlier proofs, we follow the evolution of the term $\ln\frac{\overline{W}_n}{\overline{W}_0}$. In the same way as we obtained (4.5) and (4.7), we have

$$\ln\frac{\overline{W}_n}{\overline{W}_0}\geq-\eta\min_{\boldsymbol{i}\in\mathcal{P}}\widetilde{L}_{\boldsymbol{i},n}-\ln N$$

and

$$\ln\frac{\overline{W}_n}{\overline{W}_0}\leq\sum_{t=1}^{n}\left(-\eta\sum_{\boldsymbol{i}\in\mathcal{P}}p_{\boldsymbol{i},t}\widetilde{\ell}_{\boldsymbol{i},t}+\frac{\eta^2}{2}\sum_{\boldsymbol{i}\in\mathcal{P}}p_{\boldsymbol{i},t}\widetilde{\ell}_{\boldsymbol{i},t}^2\right) .$$

Combining these bounds, we obtain

$$\begin{aligned}-\min_{\boldsymbol{i}\in\mathcal{P}}\widetilde{L}_{\boldsymbol{i},n}-\frac{\ln N}{\eta} &\leq \sum_{t=1}^{n}\left(-\sum_{\boldsymbol{i}\in\mathcal{P}}p_{\boldsymbol{i},t}\widetilde{\ell}_{\boldsymbol{i},t}+\frac{\eta}{2}\sum_{\boldsymbol{i}\in\mathcal{P}}p_{\boldsymbol{i},t}\widetilde{\ell}_{\boldsymbol{i},t}^2\right) \\ &\leq \left(-1+\frac{\eta Kb}{\varepsilon}\right)\sum_{t=1}^{n}\sum_{\boldsymbol{i}\in\mathcal{P}}p_{\boldsymbol{i},t}\widetilde{\ell}_{\boldsymbol{i},t} ,\end{aligned}$$

because $0\leq\widetilde{\ell}_{\boldsymbol{i},t}\leq\frac{2Kb}{\varepsilon}$. Applying the results of Lemma 4.9 and the union bound, we have, with probability $1-\delta/2$,

$$\begin{aligned}&-\min_{\boldsymbol{i}\in\mathcal{P}}L_{\boldsymbol{i},n}-\frac{8}{3}b\sqrt{2n\frac{K^2b}{\varepsilon}\ln\frac{4bN}{\delta}} \\ &\leq \left(-1+\frac{\eta Kb}{\varepsilon}\right)\left(\sum_{t=1}^{n}\sum_{\boldsymbol{i}\in\mathcal{P}}p_{\boldsymbol{i},t}\ell_{\boldsymbol{i},t}-\frac{8}{3}b\sqrt{2n\frac{K^2b}{\varepsilon}\ln\frac{4b}{\delta}}\right)+\frac{\ln N}{\eta} \\ &\leq \left(-1+\frac{\eta Kb}{\varepsilon}\right)\sum_{t=1}^{n}\sum_{\boldsymbol{i}\in\mathcal{P}}p_{\boldsymbol{i},t}\ell_{\boldsymbol{i},t}+\frac{8}{3}b\sqrt{2n\frac{K^2b}{\varepsilon}\ln\frac{4b}{\delta}}+\frac{\ln N}{\eta} . \qquad (4.31)\end{aligned}$$

Introduce the sets

$$\mathcal{T}_n \stackrel{\text{def}}{=} \{t : 1 \le t \le n \text{ and } S_t = 0\} \qquad \text{and} \qquad \overline{\mathcal{T}}_n \stackrel{\text{def}}{=} \{t : 1 \le t \le n \text{ and } S_t = 1\}$$

of "exploitation" and "exploration" steps, respectively. Then, by the Hoeffding-Azuma inequality [48] we obtain that, with probability at least $1 - \delta/4$,

$$\sum_{t \in \mathcal{T}_n} \sum_{\boldsymbol{i} \in \mathcal{P}} p_{\boldsymbol{i},t} \ell_{\boldsymbol{i},t} \ge \sum_{t \in \mathcal{T}_n} \ell_{\boldsymbol{I}_t,t} - \sqrt{\frac{|\mathcal{T}_n| K^2}{2} \ln \frac{4}{\delta}} \; .$$

Note that for the exploration steps $t \in \overline{\mathcal{T}}_n$, as the algorithm plays according to a uniform distribution instead of $p_{\boldsymbol{i},t}$, we can only use the trivial lower bound zero on the losses, that is,

$$\sum_{t \in \overline{\mathcal{T}}_n} \sum_{\boldsymbol{i} \in \mathcal{P}} p_{\boldsymbol{i},t} \ell_{\boldsymbol{i},t} \ge \sum_{t \in \overline{\mathcal{T}}_n} \ell_{\boldsymbol{I}_t,t} - K |\overline{\mathcal{T}}_n| \; .$$

The last two inequalities imply

$$\sum_{t=1}^{n} \sum_{\boldsymbol{i} \in \mathcal{P}} p_{\boldsymbol{i},t} \ell_{\boldsymbol{i},t} \ge \widehat{L}_n - \sqrt{\frac{|\mathcal{T}_n| K^2}{2} \ln \frac{4}{\delta}} - K |\overline{\mathcal{T}}_n| \; . \tag{4.32}$$

Then, by (4.31), (4.32) and Lemma 4.8 we obtain, with probability at least $1 - \delta$,

$$\begin{aligned} &\widehat{L}_n - \min_{\boldsymbol{i} \in \mathcal{P}} L_{\boldsymbol{i},n} \\ &\quad \le \quad K \left(\frac{\eta b}{\varepsilon} K n + \sqrt{\frac{n}{2} \ln \frac{4}{\delta}} + n\varepsilon + \frac{\sqrt{2n\varepsilon \ln \frac{4}{\delta}}}{K} + \frac{16}{3} b \sqrt{2n \frac{b}{\varepsilon} \ln \frac{4bN}{\delta}} \right) + \frac{\ln N}{\eta} \end{aligned}$$

where we used $\widehat{L}_n \le Kn$ and $|\mathcal{T}_n| \le n$. Substituting the values of ε and η gives

$$\begin{aligned} \widehat{L}_n - \min_{\boldsymbol{i} \in \mathcal{P}} L_{\boldsymbol{i},n} \quad &\le \quad K^2 b n \varepsilon + \frac{1}{4} K n \varepsilon + K n \varepsilon + \frac{1}{2} n \varepsilon + \frac{16}{3} b \sqrt{K} n \varepsilon + n\varepsilon \\ &\le \quad 9.1 K^2 b n \varepsilon \end{aligned}$$

where we used $\sqrt{\frac{n}{2} \ln \frac{4}{\delta}} \le \frac{1}{4} n\varepsilon$, $\sqrt{2n\varepsilon \ln \frac{4}{\delta}} \le \frac{1}{2} n\varepsilon$, $\sqrt{n \frac{bK}{\varepsilon} \ln \frac{4N}{\delta}} = n\varepsilon$, and $\frac{\ln N}{\eta} \le n\varepsilon$ (from the assumptions of the theorem). □

4.7 Simulation results

To further investigate our new algorithms, we have conducted some simple simulations. As the main motivation of this work is to improve earlier algorithms in case the number of paths is exponentially large in the number of edges, we tested the algorithms on the small

graph shown in Figure 4.1 (b), which has one of the simplest structures with exponentially many (namely $2^{|E|/2}$) paths.

The losses on the edges were generated by a sequence of independent and uniform random variables, with values from $[0, 1]$ on the upper edges, and from $[0.32, 1]$ on the lower edges, resulting in a (long-term) optimal path consisting of the upper edges. We ran the tests for $n = 10000$ steps, with confidence value $\delta = 0.001$. To establish baseline performance, we also tested the EXP3 algorithm of Auer *et al.* [5] (note that this algorithm does not need edge losses, only the loss of the chosen path). For the version of our bandit algorithm that is informed of the individual edge losses (edge-bandit), we used the simple 2-element cover set of the paths consisting of the upper and lower edges, respectively (other 2-element cover sets give similar performance). For our restricted shortest path algorithm (path-bandit) the basis $\{uuuuu, uuuul, uuull, uulll, ullll, lllll\}$ was used, where u (resp. l) in the kth position denotes the upper (resp. lower) edge connecting v_{k-1} and v_k. In this example the performance of the algorithm appeared to be independent of the actual choice of the basis; however, in general we do not expect this behavior. Two versions of the algorithm of Awerbuch and Kleinberg [11] were also simulated. With its original parameter setting (AwKl), the algorithm did not perform well. However, after optimizing its parameters off-line (AwKl tuned), substantially better performance was achieved. The normalized regret of the above algorithms, averaged over 30 runs, as well as the regret of the fixed paths in the graph are shown in Figure 4.7.

Although all algorithms showed better performance than the bound for the edge-bandit algorithm, the latter showed the expected superior performance in the simulations. Furthermore, our algorithm for the restricted shortest path problem outperformed Awerbuch and Kleinberg's (AwKl) algorithm, while being inferior to its off-line tuned version (AwKl tuned). It must be noted that similar parameter optimization did not improve the performance of our path-bandit algorithm, which showed robust behavior with respect to parameter tuning.

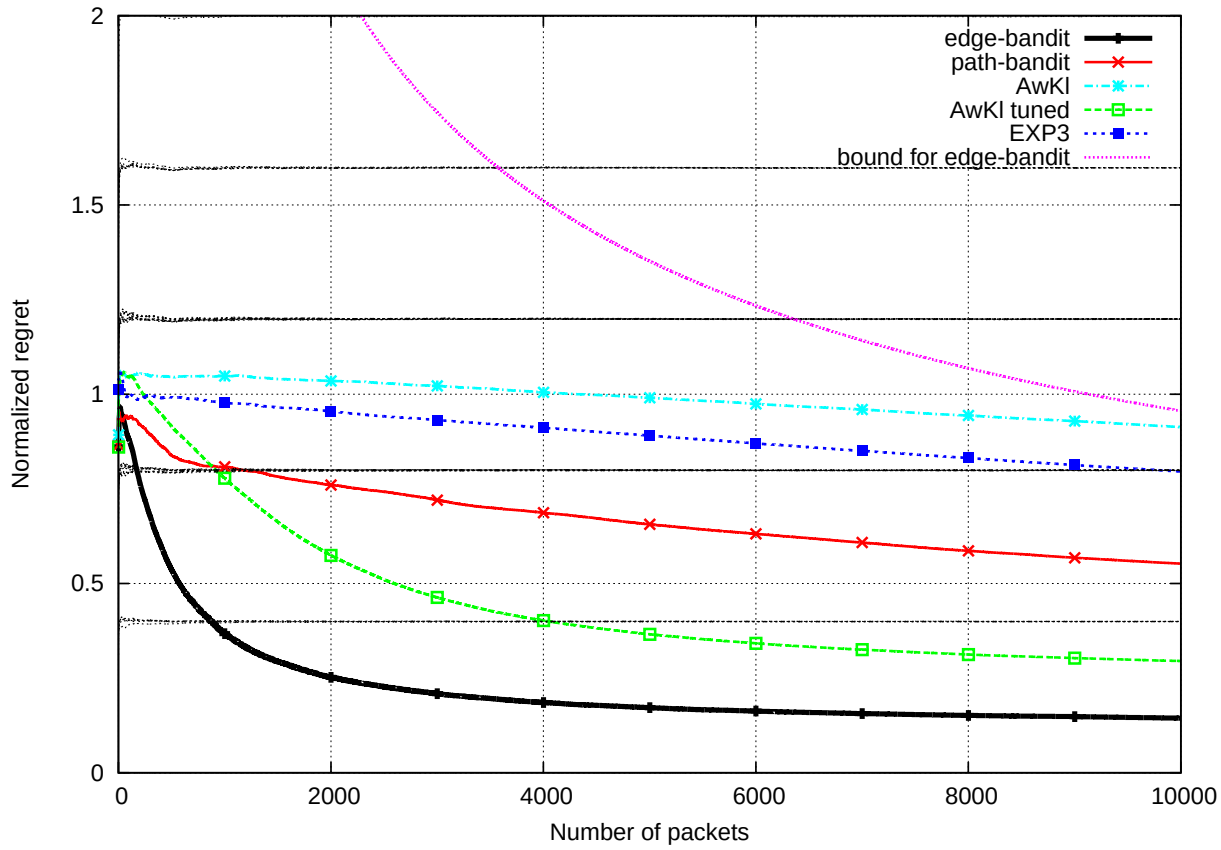

Figure 4.7: Normalized regret of several algorithms for the shortest path problem. The gray dotted lines show the normalized regret of fixed paths in the graph.

Chapter 5

On-line Prediction in case of Stationary and Ergodic Processes

One may wonder whether it is possible to improve the statements for individual sequences if we have some assumptions about the behavior of the outcome sequences $y_1, y_2, \ldots$. To spot a possible way of the improvement we recall our analogy with *pattern recognition*, which was mentioned in Section 1.1. In Chapter 3 and in Chapter 4 we have dealt with the minimization of the *estimation error*, that is, that measures the difference between the normalized regret of the best expert from a fixed expert class and the normalized regret of our algorithm. However, if we have e.g. stationary and ergodic assumption on the outcome sequence we can say also something about the *approximation error*, which describes how far is the performance of the best expert from the performance of the *Bayes-optimal* predictor, which can be achieved only in full knowledge of the underlying distribution of the outcome process.

In this chapter we provide simple on-line procedures for the prediction of a sequences in stationary and ergodic environment, which not only minimize the *estimation error* but also guarantee that the *approximation error* vanishes asymptotically. The proposed algorithms are based on a combination of several simple predictors (experts). One of the huge increment using this "model-less" expert advice approach that it provides "adaptation" also in case of *dependent* outcome sequence, where the classical methods (splitting and cross-validation) is not applicable.

In Section 5.1 we introduce a prediction strategy (algorithm) for *unbounded* stationary and ergodic real-valued processes and show that the average of squared errors of the algorithm converges, almost surely, to that of the optimum, given by the Bayes predictor. This property – that the loss of a strategy converges to the loss of the theoretical optimum – is called *universal consistency*. In Section 5.2 we offer an extension for the noisy setting, that is when the algorithm has access only to the noisy version of the original sequence. The "clean" process is passed through a fixed binary memoryless channel (e.g. Binary Symmetric Channel). This setup was introduced and studied by Weissman and Merhav [72, 73]. Theorem 5.2 proves the universal consistency of an algorithm in the noisy setting for the loss function which is convex in its first argument (e.g.: squared loss, absolute loss, etc.).

Finally, in Section 5.3 we provide a simple universally consistent classification scheme for zero-one loss in the noisy setting.

5.1 Universal prediction of unbounded time series: squared loss

The problem of time series analysis and prediction has a long and rich history, probably dating back to the pioneering work of Yule in 1927 [75]. The application scope is vast, as time series modeling is routinely employed across the entire and diverse range of applied statistics, including problems in genetics, in info-communications systems, machine condition monitoring, financial investments, marketing and econometrics. Most of the research activity until the 1970s was concerned with parametric approaches to the problem whereby a simple, usually linear model is fitted to the data or it was assumed that the process is the sum of a sequence from a restricted class or a Gaussian process (for a comprehensive account we refer the reader to the monograph of Brockwell and Davies [19]). While many appealing mathematical properties of the parametric paradigm have been established, it has become clear over the years that the limitations of the approach may be rather severe, essentially due to overly rigid constraints which are imposed on the processes. For example, it turned out that financial processes cannot be modeled by linear processes. One of the more promising solutions to overcome this problem has been the extension of classic nonparametric methods to the time series framework (see for example Györfi, Härdle, Sarda and Vieu [30] and Bosq [16] for a review and references).

The present section is devoted to the nonparametric problem of sequential prediction of *unbounded*, ergodic real valued sequences which we do not require to necessarily satisfy the classical statistical assumptions for bounded, autoregressive or Markovian processes. Indeed, our goal is to show consistency results under a strict minimum of conditions. Consistency for ergodic sequence can be proved using the powerful machine learning bounds derived for individual sequences.

To fix the context, we suppose that at each time instant $t = 1, 2, \ldots$, the predictor is asked to guess the value of the next outcome y_t of a sequence of real numbers $y_1, y_2, \ldots$ with knowledge of the past $y_1^{t-1} = (y_1, \ldots, y_{t-1})$ (where y_1^0 denotes the empty string) and the side information vectors $x_1^t = (x_1, \ldots, x_t)$, where $x_t \in \mathbb{R}^d$. Thus, the predictor's estimate, at time t, is based on the value of x_1^t and y_1^{t-1}. A prediction strategy is a sequence $g = \{g_t\}_{t=1}^{\infty}$ of functions

$$g_t : (\mathbb{R}^d)^t \times \mathbb{R}^{t-1} \to \mathbb{R}$$

so that the prediction formed at time t is $g_t(x_1^t, y_1^{t-1})$.

Throughout the chapter we assume that $(x_1, y_1), (x_2, y_2), \ldots$ are realizations of the random variables $(X_1, Y_1), (X_2, Y_2), \ldots$ such that $\{(X_n, Y_n)\}_{-\infty}^{\infty}$ is a jointly stationary and ergodic process.

After n time instants, the *normalized cumulative prediction error* is

$$L_n(g) = \frac{1}{n}\sum_{t=1}^{n} \ell(g_t(X_1^t, Y_1^{t-1}), Y_t) = \frac{1}{n}\sum_{t=1}^{n} (g_t(X_1^t, Y_1^{t-1}) - Y_t)^2,$$

where $\ell : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ now denotes the squared loss.

The results of this chapter are given in an autoregressive (on-line learning) framework, that is, the value Y_t is predicted based on the past observations (X_1^t and Y_1^{t-1}). The fundamental limit for the predictability of the sequence can be determined based on a result of Algoet [2], who showed that for any prediction strategy g and stationary ergodic process $\{(X_n, Y_n)\}_{-\infty}^{\infty}$,

$$\liminf_{n\to\infty} L_n(g) \geq L^* \quad \text{almost surely,} \tag{5.1}$$

where

$$L^* = \mathbb{E}\Big\{ \big(Y_0 - \mathbb{E}\big[Y_0 \big| X_{-\infty}^0, Y_{-\infty}^{-1}\big]\big)^2 \Big\}$$

is the minimal mean squared error of any prediction for the value of Y_0 based on the infinite past $X_{-\infty}^0, Y_{-\infty}^{-1}$. Note that it follows by stationarity and the martingale convergence theorem (see, e.g., Stout [67]) that

$$L^* = \lim_{n\to\infty} \mathbb{E}\Big\{ \big(Y_n - \mathbb{E}\big[Y_n \big| X_1^n, Y_1^{n-1}\big]\big)^2 \Big\}.$$

This lower bound gives sense to the following definition:

Definition 5.1. A prediction strategy g is called *universally consistent with respect to a class $\mathcal{C}$ of stationary and ergodic processes* $\{(X_n, Y_n)\}_{-\infty}^{\infty}$, if for each process in the class,

$$\lim_{n\to\infty} L_n(g) = L^* \quad \text{almost surely.}$$

Universally consistent strategies asymptotically achieve the best possible loss for all ergodic processes in the class.

In case of squared loss Algoet [1] proved that there exists a prediction strategy that can achieve this well-defined optimum. Using machine learning principles, Györfi and Lugosi [32] introduced several simple prediction strategies, which are universally consistent with respect to the class of bounded, stationary and ergodic processes. In this section we extend the results of [32] to *unbounded processes*. We refer to Nobel [58], Singer and Feder [65], [66] and Yang [74] for closely related recent works.

The prediction strategy g is defined, at each time instant, as a convex combination of *elementary predictors*, where the weighting coefficients depend on the past performance of each elementary predictor.

The goal of each simple predictor is to estimate the regression function $\mathbb{E}\big[Y_n \big| X_1^n, Y_1^{n-1}\big]$ at time instance n. We define an infinite array of elementary predictors $h^{(k,l)}$, $k, l = 1, 2, \ldots$ as follows. Let $\mathcal{P}_l = \{A_{l,j}, j = 1, 2, \ldots, m_l\}$ be a sequence of finite partitions of $\mathbb{R}$, and let

$\mathcal{Q}_l = \{B_{l,j}, j = 1, 2, \ldots, m_l'\}$ be a sequence of finite partitions of $\mathbb{R}^d$. Introduce the corresponding quantizers:

$$F_l(y) = j, \text{ if } y \in A_{l,j}$$

and

$$G_l(x) = j, \text{ if } x \in B_{l,j} \ .$$

With some abuse of notation, for any n and $y_1^n \in \mathbb{R}^n$, we write $F_l(y_1^n)$ for the sequence $F_l(y_1), \ldots, F_l(y_n)$, and similarly, for $x_1^n \in (\mathbb{R}^d)^n$, we write $G_l(x_1^n)$ for the sequence $G_l(x_1), \ldots, G_l(x_n)$.

Fix positive integers k, l, and for each $(k+1)$-long string z of positive integers, and for each k-long string s of positive integers, define the partitioning regression function estimate

$$\widehat{E}_n^{(k,l)}(x_1^n, y_1^{n-1}, z, s) = \frac{\sum_{\{k<t<n: G_l(x_{t-k}^t)=z,\, F_l(y_{t-k}^{t-1})=s\}} y_t}{\left|\{k < t < n : G_l(x_{t-k}^t) = z,\, F_l(y_{t-k}^{t-1}) = s\}\right|},$$

for all $n > k+1$ where $0/0$ is defined to be 0. Because of the original sequence is unbounded we have to control (bound) the predicted value of each expert. Therefore we introduce a truncation function to prevent from that the experts' prediction have "too big" values, that is,

$$T_n(z) = \begin{cases} n^\delta & \text{if } z > n^\delta; \\ z & \text{if } |z| \le n^\delta; \\ -n^\delta & \text{if } z < -n^\delta, \end{cases}$$

where

$$0 < \delta < 1/8.$$

Now we are ready to define the elementary predictor $h^{(k,l)}$ by

$$h_n^{(k,l)}(x_1^n, y_1^{n-1}) = T_n\left(\widehat{E}_n^{(k,l)}(x_1^n, y_1^{n-1}, G_l(x_{n-k}^n), F_l(y_{n-k}^{n-1}))\right),$$

for $n = 1, 2, \ldots$. That is, $h_n^{(k,l)}$ quantizes the sequence x_1^n, y_1^{n-1} according to the partitions $\mathcal{Q}_l$ and $\mathcal{P}_l$, and looks for all appearances of the last seen quantized strings $G_l(x_{n-k}^n)$ of length $k+1$ and $F_l(y_{n-k}^{n-1})$ of length k in the past. Then it predicts according to the truncation of the average of the y_t's following the string.

The proposed prediction algorithm proceeds based on exponential weighting average algorithm. Formally, let $\{q_{k,l}\}$ be a probability distribution on the set of all pairs (k, l) of positive integers such that for all k, l, $q_{k,l} > 0$. For $\eta_t > 0$, and define the weights

$$w_{k,l,t} = q_{k,l} e^{-\eta_t (t-1) L_{t-1}(h^{(k,l)})}$$

and their normalized values

$$p_{k,l,t} = \frac{w_{k,l,t}}{W_t} \ ,$$

where

$$W_t = \sum_{i,j=1}^{\infty} w_{i,j,t} \ .$$

The prediction strategy g is defined by

$$g_t(x_1^t, y_1^{t-1}) = \sum_{k,l=1}^{\infty} p_{k,l,t} h_t^{(k,l)}(x_1^t, y_1^{t-1}) \ , \qquad t = 1, 2, \ldots \tag{5.2}$$

Theorem 5.1. (GYÖRFI AND OTTUCSÁK [35]). *Assume that*
(a) the sequences of partition $\mathcal{P}_l$ is nested, that is, any cell of $\mathcal{P}_{l+1}$ is a subset of a cell of $\mathcal{P}_l$, $l = 1, 2, \ldots$;
(b) the sequences of partition $\mathcal{Q}_l$ is nested;
(c) the sequences of partition $\mathcal{P}_l$ is asymptotically fine, i.e., if

$$\mathrm{diam}(A) = \sup_{x,y \in A} \|x - y\|$$

denotes the diameter of a set, then for each sphere S centered at the origin

$$\lim_{l \to \infty} \max_{j: A_{l,j} \cap S \neq \emptyset} \mathrm{diam}(A_{l,j}) = 0 \ ;$$

(d) the sequences of partition $\mathcal{Q}_l$ is asymptotically fine.
Choose the parameter η_t of the algorithm as

$$\eta_t = \frac{1}{\sqrt{t}} \ .$$

Then the prediction scheme g defined above is universally consistent with respect to the class of all ergodic processes such that

$$\mathbb{E}\{Y_1^4\} < \infty.$$

Here we describe two results, which are used in the analysis. The first lemma is a modification of the analysis of Auer *et al.* [7], which allows of the handling the case when the parameter of the algorithm (η_t) is time-dependent and the number of the elementary predictors is infinite.

Lemma 5.1. (GYÖRFI AND OTTUCSÁK [35]). *Let $h^{(1)}, h^{(2)}, \ldots$ be a sequence of prediction strategies (experts). Let $\{q_k\}$ be a probability distribution on the set of positive integers. Denote the normalized loss of the expert $h = (h_1, h_2, \ldots)$ by*

$$L_n(h) = \frac{1}{n} \sum_{t=1}^{n} \ell_t(h),$$

where

$$\ell_t(h) = \ell(h_t, Y_t)$$

and the loss function ℓ is convex in its first argument h. Define

$$w_{k,t} = q_k e^{-\eta_t (t-1) L_{t-1}(h^{(k)})}$$

where $\eta_t > 0$ is monotonically decreasing, and

$$p_{k,t} = \frac{w_{k,t}}{W_t}$$

where

$$W_t = \sum_{k=1}^{\infty} w_{k,t} \ .$$

If the prediction strategy $g = (g_1, g_2, \dots)$ is defined by

$$g_t = \sum_{k=1}^{\infty} p_{k,t} h_t^{(k)} \qquad t = 1, 2, \dots$$

then for every $n \geq 1$,

$$L_n(g) \leq \inf_k \left(L_n(h^{(k)}) - \frac{\ln q_k}{n\eta_{n+1}} \right) + \frac{1}{2n} \sum_{t=1}^{n} \eta_t \sum_{k=1}^{\infty} p_{k,t} \ell_t^2(h^{(k)}).$$

Proof. Introduce some notations:

$$w'_{k,t} = q_k e^{-\eta_{t-1}(t-1)L_{t-1}(h^{(k)})},$$

which is the weight $w_{k,t}$, where η_t is replaced by η_{t-1} and the sum of these are

$$W'_t = \sum_{k=1}^{\infty} w'_{k,t}.$$

We start the proof with the following chain of bounds:

$$\begin{aligned}
\frac{1}{\eta_t} \ln \frac{W'_{t+1}}{W_t} &= \frac{1}{\eta_t} \ln \frac{\sum_{k=1}^{\infty} w_{k,t} e^{-\eta_t \ell_t(h^{(k)})}}{W_t} \\
&= \frac{1}{\eta_t} \ln \sum_{k=1}^{\infty} p_{k,t} e^{-\eta_t \ell_t(h^{(k)})} \\
&\leq \frac{1}{\eta_t} \ln \sum_{k=1}^{\infty} p_{k,t} \left(1 - \eta_t \ell_t(h^{(k)}) + \frac{\eta_t^2}{2} \ell_t^2(h^{(k)}) \right)
\end{aligned}$$

because of $e^{-x} \le 1 - x + x^2/2$ for $x \ge 0$. Moreover,

$$\begin{aligned}
\frac{1}{\eta_t} \ln \frac{W'_{t+1}}{W_t} &\le \frac{1}{\eta_t} \ln \left(1 - \eta_t \sum_{k=1}^{\infty} p_{k,t} \ell_t(h^{(k)}) + \frac{\eta_t^2}{2} \sum_{k=1}^{\infty} p_{k,t} \ell_t^2(h^{(k)}) \right) \\
&\le -\sum_{k=1}^{\infty} p_{k,t} \ell_t(h^{(k)}) + \frac{\eta_t}{2} \sum_{k=1}^{\infty} p_{k,t} \ell_t^2(h^{(k)}) \qquad (5.3) \\
&= -\sum_{k=1}^{\infty} p_{k,t} \ell(h_t^{(k)}, Y_t) + \frac{\eta_t}{2} \sum_{k=1}^{\infty} p_{k,t} \ell_t^2(h^{(k)}) \\
&\le -\ell \left(\sum_{k=1}^{\infty} p_{k,t} h_t^{(k)}, Y_t \right) + \frac{\eta_t}{2} \sum_{k=1}^{\infty} p_{k,t} \ell_t^2(h^{(k)}) \qquad (5.4) \\
&= -\ell_t(g) + \frac{\eta_t}{2} \sum_{k=1}^{\infty} p_{k,t} \ell_t^2(h^{(k)}) \qquad (5.5)
\end{aligned}$$

where (5.3) follows from the fact that $\ln(1+x) \le x$ for all $x > -1$ and in (5.4) we used the convexity of the loss $\ell(h, y)$ in its first argument h. From (5.5) after rearranging we obtain

$$\ell_t(g) \le -\frac{1}{\eta_t} \ln \frac{W'_{t+1}}{W_t} + \frac{\eta_t}{2} \sum_{k=1}^{\infty} p_{k,t} \ell_t^2(h^{(k)}) \ .$$

Then write a telescope formula:

$$\begin{aligned}
\frac{1}{\eta_t} \ln W_t - \frac{1}{\eta_t} \ln W'_{t+1} &= \left(\frac{1}{\eta_t} \ln W_t - \frac{1}{\eta_{t+1}} \ln W_{t+1} \right) \\
&\quad + \left(\frac{1}{\eta_{t+1}} \ln W_{t+1} - \frac{1}{\eta_t} \ln W'_{t+1} \right) \\
&= (A_t) + (B_t).
\end{aligned}$$

We have that

$$\begin{aligned}
\sum_{t=1}^{n} A_t &= \sum_{t=1}^{n} \left(\frac{1}{\eta_t} \ln W_t - \frac{1}{\eta_{t+1}} \ln W_{t+1} \right) \\
&= \frac{1}{\eta_1} \ln W_1 - \frac{1}{\eta_{n+1}} \ln W_{n+1} \\
&= -\frac{1}{\eta_{n+1}} \ln \sum_{k=1}^{\infty} q_k e^{-\eta_{n+1} n L_n(h^{(k)})} \\
&\le -\frac{1}{\eta_{n+1}} \ln \sup_k q_k e^{-\eta_{n+1} n L_n(h^{(k)})} \\
&= -\frac{1}{\eta_{n+1}} \sup_k \left(\ln q_k - \eta_{n+1} n L_n(h^{(k)}) \right) \\
&= \inf_k \left(n L_n(h^{(k)}) - \frac{\ln q_k}{\eta_{n+1}} \right).
\end{aligned}$$

$\frac{\eta_{t+1}}{\eta_t} \le 1$, therefore applying Jensen's inequality for concave function, we get that

$$\begin{aligned} W_{t+1} &= \sum_{i=1}^{\infty} q_i e^{-\eta_{t+1} t L_t(h^{(i)})} \\ &= \sum_{i=1}^{\infty} q_i \left(e^{-\eta_t t L_t(h^{(i)})}\right)^{\frac{\eta_{t+1}}{\eta_t}} \\ &\le \left(\sum_{i=1}^{\infty} q_i e^{-\eta_t t L_t(h^{(i)})}\right)^{\frac{\eta_{t+1}}{\eta_t}} \\ &= \left(W'_{t+1}\right)^{\frac{\eta_{t+1}}{\eta_t}} . \end{aligned}$$

Thus,

$$\begin{aligned} B_t &= \frac{1}{\eta_{t+1}} \ln W_{t+1} - \frac{1}{\eta_t} \ln W'_{t+1} \\ &\le \frac{1}{\eta_{t+1}} \frac{\eta_{t+1}}{\eta_t} \ln W'_{t+1} - \frac{1}{\eta_t} \ln W'_{t+1} \\ &= 0. \end{aligned}$$

We can summarize the bounds:

$$L_n(g) \le \inf_k \left(L_n(h^{(k)}) - \frac{\ln q_k}{n\eta_{n+1}} \right) + \frac{1}{2n} \sum_{t=1}^{n} \eta_t \sum_{k=1}^{\infty} p_{k,t} \ell_t^2(h^{(k)}) .$$

□

The next lemma is due to Breiman [18], and its proof may also be found in Györfi *et al.* [31].

Lemma 5.2. *Let $Z = \{Z_i\}_{-\infty}^{\infty}$ be a stationary and ergodic time series. Let T denote the left shift operator. Let f_i be a sequence of real-valued functions such that for some function f, $f_i(Z) \to f(Z)$ almost surely. Assume that $\mathbb{E} \sup_i |f_i(Z)| < \infty$. Then*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} f_i(T^i Z) = \mathbb{E} f(Z)$$

almost surely.

Proof of Theorem 5.1. Because of (5.1), it is enough to show that

$$\limsup_{n\to\infty} L_n(g) \le L^* \qquad \text{a.s.} \tag{5.6}$$

By a double application of the ergodic theorem, as $n \to \infty$, a.s.,

$$\begin{aligned}\widehat{E}_n^{(k,l)}(X_1^n, Y_1^{n-1}, z, s) &= \frac{\frac{1}{n}\sum_{\{k<t<n:\, G_l(X_{t-k}^t)=z,\, F_l(Y_{t-k}^{t-1})=s\}} Y_t}{\frac{1}{n}\left|\{k<t<n:\, G_l(X_{t-k}^t)=z,\, F_l(Y_{t-k}^{t-1})=s\}\right|} \\ &\to \frac{\mathbb{E}\{Y_0 I_{\{G_l(X_{-k}^0)=z,\, F_l(Y_{-k}^{-1})=s\}}\}}{\mathbb{P}\{G_l(X_{-k}^0)=z,\, F_l(Y_{-k}^{-1})=s\}} \\ &= \mathbb{E}\{Y_0 \mid G_l(X_{-k}^0)=z,\, F_l(Y_{-k}^{-1})=s\},\end{aligned}$$

and therefore for all z and s

$$T_n\left(\widehat{E}_n^{(k,l)}(X_1^n, Y_1^{n-1}, z, s)\right) \to \mathbb{E}\{Y_0 \mid G_l(X_{-k}^0)=z,\, F_l(Y_{-k}^{-1})=s\}.$$

Now we can write

$$\begin{aligned}L_n(h^{(k,l)}) &= \frac{1}{n}\sum_{t=1}^n (h^{(k,l)}(X_1^t, Y_1^{t-1}) - Y_t)^2 \\ &= \frac{1}{n}\sum_{t=1}^n \left(T_t\left(\widehat{E}_t^{(k,l)}(X_1^t, Y_1^{t-1}, G_l(X_{t-k}^t), F_l(Y_{t-k}^{t-1}))\right) - Y_t\right)^2 . \qquad (5.7)\end{aligned}$$

To use Lemma 5.2 we have to verify $\mathbb{E}\sup_i |f_i(Y_{-\infty}^{\infty}, X_{-\infty}^{\infty})| < \infty$, where

$$f_i(X_{-\infty}^{\infty}, Y_{-\infty}^{\infty}) = (h^{(k,l)}(X_{1-i}^0, Y_{1-i}^{-1}) - Y_0)^2.$$

One can show that is enough to verify only the numerator of $\hat{E}_n^{(k,l)}(X_{1-k}^0, Y_{1-k}^{-1}, z, s)$ divided by n is finite for each individual z and s. For this we can apply maximal ergodic theorem (see Krengel [51] Theorem 6.3 with parameter $p = 2$). Now using Lemma 5.2, as $n \to \infty$, almost surely, we get from (5.7)

$$\begin{aligned}L_n(h^{(k,l)}) &\to \mathbb{E}\{(Y_0 - \mathbb{E}\{Y_0 \mid G_l(X_{-k}^0),\, F_l(Y_{-k}^{-1})\})^2\} \\ &\stackrel{\text{def}}{=} \epsilon_{k,l}.\end{aligned}$$

$\mathbb{E}\{Y_0 \mid G_l(X_{-k}^0),\, F_l(Y_{-k}^{-1})\}$ is a martingale indexed by the pair (k, l), since the partitions $\mathcal{P}_l$ and $\mathcal{Q}_l$ are nested. Thus, the martingale convergence theorem (see, e.g., Stout [67]) and assumptions (c) and (d) for the sequences of partitions implies that

$$\inf_{k,l} \epsilon_{k,l} = \lim_{k,l\to\infty} \epsilon_{k,l} = \mathbb{E}\left\{\left(Y_0 - \mathbb{E}\{Y_0 | X_{-\infty}^0, Y_{-\infty}^{-1}\}\right)^2\right\} = L^*$$

(cf. Györfi and Lugosi [32]).

To prove (5.6) apply Lemma 5.1 with choice $\eta_t = \frac{1}{\sqrt{t}}$ and for the squared loss $\ell_t(h) = (h_t - Y_t)^2$, then the squared loss is convex in its first argument h, so

$$L_n(g) \le \inf_{k,l}\left(L_n(h^{(k,l)}) - \frac{2\ln q_{k,l}}{\sqrt{n}}\right) + \frac{1}{2n}\sum_{t=1}^n \frac{1}{\sqrt{t}} \sum_{k,l=1}^{\infty} p_{k,l,t}\left(h^{(k,l)}(X_1^t, Y_1^{t-1}) - Y_t\right)^4. \qquad (5.8)$$

On the one hand, almost surely,

$$\begin{aligned}\limsup_{n\to\infty}\inf_{k,l}\left(L_n(h^{(k,l)})-\frac{2\ln q_{k,l}}{\sqrt{n}}\right) &\le \inf_{k,l}\limsup_{n\to\infty}\left(L_n(h^{(k,l)})-\frac{2\ln q_{k,l}}{\sqrt{n}}\right)\\ &= \inf_{k,l}\limsup_{n\to\infty}L_n(h^{(k,l)})\\ &= \inf_{k,l}\epsilon_{k,l}\\ &= \lim_{k,l\to\infty}\epsilon_{k,l}\\ &= L^*.\end{aligned}$$

On the other hand,

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n}\frac{1}{\sqrt{t}}\sum_{k,l}p_{k,l,t}(h^{(k,l)}(X_1^t,Y_1^{t-1})-Y_t)^4 &\le \frac{8}{n}\sum_{t=1}^{n}\frac{1}{\sqrt{t}}\sum_{k,l}p_{k,l,t}\left(h^{(k,l)}(X_1^t,Y_1^{t-1})^4+Y_t^4\right)\\ &\le \frac{8}{n}\sum_{t=1}^{n}\frac{1}{\sqrt{t}}\sum_{k,l}p_{k,l,t}\left(t^{4\delta}+Y_t^4\right)\\ &= \frac{8}{n}\sum_{t=1}^{n}\frac{t^{4\delta}+Y_t^4}{\sqrt{t}},\end{aligned}$$

therefore, almost surely,

$$\limsup_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\frac{1}{\sqrt{t}}\sum_{k,l}p_{k,l,t}(h^{(k,l)}(X_1^t,Y_1^{t-1})-Y_t)^4\le\limsup_{n\to\infty}\frac{8}{n}\sum_{t=1}^{n}\frac{Y_t^4}{\sqrt{t}}=0,$$

where we applied that $\mathbb{E}\{Y_1^4\}<\infty$ and $0<\delta<\frac{1}{8}$. Summarizing these bounds, we get that, almost surely,

$$\limsup_{n\to\infty}L_n(g)\le L^*$$

and the proof of the theorem is finished. □

Remark 5.1. (CHOICE OF $q_{k,l}$) Theorem 5.1 is true independently of the choice of the $q_{k,l}$'s as long as these values are strictly positive for all k and l. In practice, however, the choice of $q_{k,l}$ may have an impact on the performance of the predictor. For example, if the distribution $\{q_{k,l}\}$ has a very rapidly decreasing tail, then the term $-\ln q_{k,l}/\sqrt{n}$ will be large for moderately large values of k and l, and the performance of g will be determined by the best of just a few of the elementary predictors $h^{(k,l)}$. Thus, it may be advantageous to choose $\{q_{k,l}\}$ to be a large-tailed distribution. For example, $q_{k,l}=c_0k^{-2}l^{-2}$ is a safe choice, where c_0 is an appropriate normalizing constant.

Remark 5.2. (GENERAL LOSSES) It is easy to extend Theorem 5.1 to the loss function

$$\ell(x,y)=|x-y|^r\ ,$$

where $r\ge 1$.

Remark 5.3. (IMPLEMENTATION) The proposed algorithm in not computationally feasible to implement it because of the infinite number of simple predictors. However, in practical scenarios e.g. in regression problem Biau, Bleakley, Györfi and Ottucsák [14] or in portfolio selection problems cf. Györfi, Lugosi and Udina [34] and Ottucsák and Vajda [62] it seems that a relatively small proportion of experts ($k = 1, \dots, 5$ and $l = 1, \dots, 10$) provides good experimental results. Moreover, for the higher values of k and l the achieved performance is from bad to worse.

5.2 Universal prediction for binary memoryless channel: general convex loss

In this section we investigate the case when the predictor has only incomplete information. Here $\{(X_n, Y_n)\}_{-\infty}^{\infty}$ is a jointly stationary and ergodic process and both X_t and Y_t are binary valued. The predictor's estimate, at time t, is based on the value of X_1^{t-1} and a prediction strategy is a sequence $\overline{g} = \{\overline{g}_t\}_{t=1}^{\infty}$ of functions

$$\overline{g}_t : \{0,1\}^{t-1} \to \mathbb{R}$$

so that the prediction formed at time t is $\overline{g}_t(X_1^{t-1})$.

Obviously, on the one hand this model is a special case of the previous setup (because the outcome is a binary value sequence), on the other hand it handles a more general class of loss functions (convex losses) and takes less assumption on the amount of the information (uses only past side information).

After n time instants, the *normalized cumulative loss* is

$$L_n(\overline{g}) \stackrel{\text{def}}{=} \frac{1}{n}\sum_{t=1}^{n} \ell(\overline{g}_t(X_1^{t-1}), Y_t)$$

where $\ell : \mathbb{R} \times \{0,1\} \to [0, K]$ is a bounded loss function, which is convex in its first argument. This model was introduced and studied in Weissman and Merhav [72, 73].

The key property of the loss function, which allows to obtain universal consistency in the case noisy environment, is that the loss function can be "linearized" in Y_t, that is,

$$L_n(\overline{g}) = \frac{1}{n}\sum_{t=1}^{n}\left[(1-Y_t)\ell(\overline{g}_t(X_1^{t-1}), 0) + Y_t\ell(\overline{g}_t(X_1^{t-1}), 1)\right]$$

because Y_t is binary. This form allows us to estimate Y_t much easier (directly) irrespectively of the loss function.

The prediction with side information only is a delicate problem, because Y_t neither in the learning, nor in the prediction is available. In that case the fundamental limit for the predictability of the sequence can be determined as follows. Let

$$g_t^*(X_1^{t-1}) = \mathbb{E}(Y_t | X_1^{t-1})$$

be the Bayes-optimal predictor and its normalized cumulative loss is

$$L_n(g_t^*) = \frac{1}{n}\sum_{t=1}^{n} \ell(g^*(X_1^{t-1}), Y_t) \ .$$

Now define

$$\delta_t = \ell(\overline{g}_t(X_1^{t-1}), Y_t) - \mathbb{E}\left(\ell(\overline{g}_t(X_1^{t-1}), Y_t)|X_1^{t-1}\right)$$

then we can write

$$\begin{aligned} L_n(\overline{g}) &= \frac{1}{n}\sum_{t=1}^{n}\delta_t + \frac{1}{n}\sum_{t=1}^{n}\mathbb{E}\left(\ell(\overline{g}_t(X_1^{t-1}), Y_t)|X_1^{t-1}\right) \\ &\geq \frac{1}{n}\sum_{t=1}^{n}\delta_t + \frac{1}{n}\sum_{t=1}^{n}\mathbb{E}\left(\ell(g_t^*(X_1^{t-1}), Y_t)|X_1^{t-1}\right) \ . \end{aligned}$$

Weissman and Merhav [73, Lemma 1] proved

$$\frac{1}{n}\sum_{t=1}^{n}\delta_t \to 0 \qquad \text{a.s.}$$

under the condition that $\{(X_n, Y_n)\}_{n=-\infty}^{\infty}$ is conditionally mixing in the sense that

$$\sum_{s=1}^{\infty}\sup_{t\geq 1}\mathbb{E}\left|\mathbb{P}\{Y_{t+s}=a|Y_t=a, X_1^{t+s-1}\} - \mathbb{P}\{Y_{t+s}=a|X_1^{t+s-1}\}\right| < \infty, \tag{5.9}$$

where $a \in \{0,1\}$. Therefore, we get

$$\liminf_{n\to\infty} L_n(\overline{g}) \geq \liminf_{n\to\infty} L_n(g^*) = R^* \ , \tag{5.10}$$

with

$$R^* = \mathbb{E}\left\{(1-Y_0)\ell(\mathbb{E}\{Y_0 \mid X_{-\infty}^{-1}\}, 0) + Y_0\ell(\mathbb{E}\{Y_0 \mid X_{-\infty}^{-1}\}, 1)\right\} \ . \tag{5.11}$$

Similarly to Definition 5.1 we call a prediction strategy $\overline{g}$ *universally consistent* with respect to a class $\mathcal{C}$ of stationary and ergodic processes $\{(X_n, Y_n)\}_{-\infty}^{\infty}$ if for each process in the class,

$$\lim_{n\to\infty} L_n(\overline{g}) = R^* \quad \text{almost surely.}$$

Henceforth, we assume that the connection between Y_t and X_t are characterized by an *binary memoryless channel* as, e.g., binary symmetric channel or binary erasure channel. It means that Y_t is the input of the channel and X_t is the output of the channel, and based on the past outputs X_1^{t-1} we want to estimate the input Y_t. We suppose also that the crossover probabilities of the channel are *known* for the algorithm. This assumption is indeed a realistic one in many applications, where noisy medium is well-characterized statistically.

Then the algorithm is able to construct a random variable $\tilde{r}(X_t, \mathbf{C})$ which is an efficient estimate of original bit Y_t where $\mathbf{C}$ is the channel matrix:

$$\mathbf{C} = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix} ,$$

and $0 \leq p, q < \frac{1}{2}$ are the crossover probabilities of the channel. More precisely, let

$$\tilde{r}(X_t, \mathbf{C}) = \frac{X_t - p}{1-p-q}$$

which is a conditionally unbiased estimate of Y_t respect to X_1^{t-1}. Namely,

$$\begin{aligned} \mathbb{E}\{X_t|Y_t\} &= I_{\{Y_t=0\}}[(1-p)Y_t + p(1-Y_t)] + I_{\{Y_t=1\}}[(1-q)Y_t + q(1-Y_t)] \\ &= I_{\{Y_t=0\}}[p(1-Y_t)] + I_{\{Y_t=1\}}[(1-q)Y_t] \\ &= p + Y_t(1-p-q) \end{aligned}$$

and therefore

$$\begin{aligned} \mathbb{E}\{\tilde{r}(X_t, \mathbf{C})|X_1^{t-1}\} &= \mathbb{E}\left\{ \frac{X_t - p}{1-p-q} \middle| X_1^{t-1} \right\} \\ &= \mathbb{E}\left\{ \frac{\mathbb{E}\left\{X_t|Y_t, X_1^{t-1}\right\} - p}{1-p-q} \middle| X_1^{t-1} \right\} \\ &= \mathbb{E}\left\{ \frac{\mathbb{E}\left\{X_t|Y_t\right\} - p}{1-p-q} \middle| X_1^{t-1} \right\} \\ &= \mathbb{E}\{Y_t|X_1^{t-1}\} , \end{aligned}$$

where the third equation follows from the memoryless property of the channel.

The algorithm is defined, at each time instant, as a combination of *simple predictors*, where the weighting coefficients depend on the past performance of each simple predictor.

We define an infinite array of elementary predictors $h^{(k)}$, $k = 1, 2, \ldots$ as follows. Let $J_n^{(k)}$ be the locations of the matches of the last seen binary string x_{n-k}^{n-1} of length k in the past:

$$J_n^{(k)} = \{k < t < n : x_{t-k}^{t-1} = x_{n-k}^{n-1}\} .$$

Now define the elementary predictor $h^{(k)}$ by

$$h^{(k)}(x_1^{n-1}) = \tilde{r}\left(\frac{\sum_{\{t \in J_n^{(k)}\}} x_t}{|J_n^{(k)}|}, \mathbf{C} \right)$$

$n > k+1$, where $0/0$ is defined to be 0. Note that $h^{(k)}(x_1^{n-1}) \in \left[\frac{-p}{1-p-q}; \frac{1-p}{1-p-q}\right]$.

Since, the predictor has no access to the "clean" sequence Y_t thus to measure its own performance (loss) it must use another type of the loss function based on X_t only. Define

the following loss function introduced by Weissman and Merhav [72]: let $\widetilde{\ell} : \mathbb{R} \times \{0,1\} \to [\frac{-pK}{1-2p}, \frac{(1-p)K}{1-2p}]$ be the *estimated loss*, where K is the upper bound of $\ell(\cdot,\cdot)$. More precisely, let

$$\widetilde{\ell}(\overline{g}_t(X_1^{t-1}), X_t) \stackrel{\text{def}}{=} \tilde{r}(1 - X_t, \mathbf{C})\ell(\overline{g}_t(X_1^{t-1}), 0) + \tilde{r}(X_t, \mathbf{C})\ell(\overline{g}_t(X_1^{t-1}), 1) \ ,$$

which is an (conditionally) unbiased estimate of the k-th expert's true loss. The cumulative estimated loss of the k-th expert is given by

$$\tilde{L}_n(h^{(k)}) = \frac{1}{n}\sum_{t=1}^{n} \widetilde{\ell}(h^{(k)}(X_1^{t-1}), X_t) \ .$$

The proposed prediction algorithm proceeds as follows: let $\{q_k\}$ be a probability distribution on the set of all k of positive integers such that for all k, $q_k > 0$. For $\eta_t > 0$, define the weights

$$w_{k,t} = q_k e^{-\eta_t(t-1)\tilde{L}_{t-1}(h^{(k)})}$$

and their normalized values

$$p_{k,t} = \frac{w_{k,t}}{\sum_{i=1}^{\infty} w_{i,t}} \ .$$

The prediction strategy $\overline{g}$ is defined by

$$\overline{g}_t(x_1^{t-1}) = \sum_{k=1}^{\infty} p_{k,t} h^{(k)}(x_1^{t-1}) \ , \qquad t = 1, 2, \dots \ . \tag{5.12}$$

Theorem 5.2. (OTTUCSÁK AND GYÖRFI [60]). *Assume that $\{Y_t\}$ is stationary ergodic, and $\{X_t\}$ is the output sequence of a binary memoryless channel if $\{Y_t\}$ is the input sequence. The prediction scheme $\overline{g}$ defined above is universally consistent with respect to the class of all ergodic processes satisfying (5.9).*

For the proof of the theorem we use the next lemma is due to Weissman and Merhav [72] (Lemma 2).

Lemma 5.3. *If $\ell(\cdot,\cdot) \in [0, B]$ then for any predictor g*

$$\limsup_{n\to\infty} \frac{\sqrt{n}|L_n(\overline{g}) - \tilde{L}_n(\overline{g})|}{\sqrt{\log\log n}} \le C(\mathbf{C}) \quad \text{a.s.},$$

where $C(\mathbf{C})$ is a deterministic constant depending on the channel matrix.

Proof of Theorem 5.2. Because of (5.9) we have (5.10), therefore it is enough to show that

$$\limsup_{n\to\infty} L_n(g) \le R^* \qquad \text{a.s.}$$

Now we can write

$$\begin{aligned}
\limsup_{n\to\infty} L_n(\overline{g}) - R^* &\leq \limsup_{n\to\infty} |L_n(\overline{g}) - \tilde{L}_n(\overline{g})| && (5.13)\\
&\quad + \limsup_{n\to\infty} \tilde{L}_n(\overline{g}) - \inf_k \limsup_{n\to\infty} \tilde{L}_n(h^{(k)}) && (5.14)\\
&\quad + \inf_k \limsup_{n\to\infty} \tilde{L}_n(h^{(k)}) - \inf_k \limsup_{n\to\infty} L_n(h^{(k)}) && (5.15)\\
&\quad + \inf_k \limsup_{n\to\infty} L_n(h^{(k)}) - R^*. && (5.16)
\end{aligned}$$

(5.13) and (5.15) goes to zero because of Lemma 5.3. For (5.14), we can apply Lemma 5.1 with $\overline{\ell}(\cdot,\cdot) = \widetilde{\ell}(\cdot,\cdot) + \frac{pK}{1-p-q}$, where the last additive term ensures that $\overline{\ell}(\cdot,\cdot) \geq 0$. Then $\overline{\ell}(\cdot,\cdot) \in [0, B]$, where $B = \frac{K}{1-p-q}$ and we have

$$\begin{aligned}
\limsup_{n\to\infty} \tilde{L}_n(\overline{g}) &\leq \limsup_{n\to\infty} \inf_k \left(\tilde{L}_n(h^{(k)}) - \frac{2B \ln q_k}{\sqrt{n}} \right)\\
&\leq \inf_k \limsup_{n\to\infty} \left(\tilde{L}_n(h^{(k)}) - \frac{2B \ln q_k}{\sqrt{n}} \right)\\
&\leq \inf_k \limsup_{n\to\infty} \tilde{L}_n(h^{(k)}) \ .
\end{aligned}$$

Thus it remains to show that (5.16) is smaller than zero:

$$\inf_k \limsup_{n\to\infty} L_n(h^{(k)}) - R^* \leq 0 \ .$$

By an application of the ergodic theorem, as $n \to \infty$, a.s.,

$$\begin{aligned}
h_n^{(k)}(X_1^{n-1}) &= \tilde{r}\left(\frac{\sum_{\{t \in J_n^{(k)}\}} X_t}{|J_n^{(k)}|}, \mathbf{C} \right)\\
&\to \tilde{r}\left(\mathbb{E}\{X_0 | X_{-k}^{-1}\}, \mathbf{C} \right)\\
&= \mathbb{E}\{\tilde{r}(X_0, \mathbf{C}) | X_{-k}^{-1}\}\\
&= \mathbb{E}\{Y_0 | X_{-k}^{-1}\} \ .
\end{aligned}$$

By Lemma 5.2, as $n \to \infty$, almost surely,

$$\begin{aligned}
L_n(h^{(k)}) &= \frac{1}{n} \sum_{t=1}^{n} \ell(h^{(k)}(X_1^{t-1}), Y_t)\\
&\to \mathbb{E}\{\ell(\mathbb{E}\{Y_0 \mid X_{-k}^{-1}\}, Y_0)\\
&= \mathbb{E}\{(1-Y_0)\ell(\mathbb{E}\{Y_0 \mid X_{-k}^{-1}\}, 0) + Y_0 \ell(\mathbb{E}\{Y_0 \mid X_{-k}^{-1}\}, 1)\}\\
&\stackrel{\text{def}}{=} \epsilon_k.
\end{aligned}$$

Thus, the martingale convergence theorem (see, e.g., Stout [67, Theorem 2.8.6.]) implies that

$$\inf_k \epsilon_k = \lim_{k\to\infty} \epsilon_k = \mathbb{E}\left\{(1-Y_0)\ell(\mathbb{E}\{Y_0 \mid X_{-\infty}^{-1}\}, 0) + Y_0\ell(\mathbb{E}\{Y_0 \mid X_{-\infty}^{-1}\}, 1)\right\} = R^*$$

as desired. □

Remark 5.4. (PREDICTION UNDER CHANNEL UNCERTAINTY) If we assume that sometimes the algorithm has access to the original bit Y_t, then we may construct a universal consistent prediction scheme even if p and q are unknown for the algorithm. However in a number of cases there are expensive to obtain Y_t, therefore the forecaster has the option to query this information. For query it used i.i.d. sequence $S_1, S_2, \ldots, S_n$ of Bernoulli random variables such that $\mathbb{P}\{S_t = 1\} = \epsilon$ and asks label Y_t if $S_t = 1$. Then the algorithm can construct an efficient estimate of the crossover probabilities:

$$\tilde{p}_n = \frac{\sum_{t=1}^n I_{\{X_t=1, Y_t=0\}} S_t}{\sum_{t=1}^n I_{\{Y_t=0\}} S_t}$$

and

$$\tilde{q}_n = \frac{\sum_{t=1}^n I_{\{X_t=0, Y_t=1\}} S_t}{\sum_{t=1}^n I_{\{Y_t=1\}} S_t} ,$$

where $\tilde{p}_n \to p$ and $\tilde{q}_n \to q$. Now using these estimates in $\widetilde{\ell}(\cdot,\cdot)$ and $\tilde{r}(\cdot,\cdot)$ we obtain a universal prediction scheme. The above described situation appears when the algorithm is supported by a human expert or we have a second no noisy-channel. For example, in case of natural language processing (e.g. 8 bits represent a character), the human observer select the best possible reconstruction, which e.g, can be found in the "dictionary" and fits in with the context.

5.3 Universal prediction for binary memoryless channel: zero-one loss

In this section we apply the same ideas to the seemingly more difficult classification (or pattern recognition) problem. The strategy of the classifier is a sequence $\overline{f} = \{\overline{f}_t\}_{t=1}^{\infty}$ of decision functions

$$\overline{f}_t : \{0,1\}^{t-1} \to \{0,1\}$$

so that the classification formed at time t is $f_t(X_1^{t-1})$. The *normalized cumulative* $0-1$ *loss* for any fixed pair of sequences X_1^n, Y_1^n is now

$$R_n(\overline{f}) = \frac{1}{n}\sum_{t=1}^n I_{\{\overline{f}_t(X_1^{t-1}) \neq Y_t\}}.$$

(5.9) implies (5.10) such that

$$\liminf_{n\to\infty} R_n(\overline{f}) \geq R^* \tag{5.17}$$

where

$$R^* = \mathbb{E}\left\{ \min\left(\mathbb{P}\{Y_0 = 1|X_{-\infty}^{-1}\}, \mathbb{P}\{Y_0 = 0|X_{-\infty}^{-1}\}\right) \right\}.$$

Consider the prediction scheme $\overline{g}_t(X_1^{t-1})$ with squared loss $\ell(x,y) = (x-y)^2$, introduced in the previous section, and then introduce the corresponding classification scheme:

$$\overline{f}_t(X_1^{t-1}) = \begin{cases} 1 & \text{if } \overline{g}_t(X_1^{t-1}) > 1/2; \\ 0 & \text{otherwise.} \end{cases}$$

The main result of this section is the universal consistency of this simple classification scheme:

Theorem 5.3. (OTTUCSÁK AND GYÖRFI [60]). *Assume that $\{Y_t\}$ is stationary ergodic, and $\{X_t\}$ is the output sequence of a binary memoryless channel if $\{Y_t\}$ is the input sequence. The classification scheme $\overline{f}$ defined above satisfies*

$$\lim_{n\to\infty} R_n(\overline{f}) = R^* \quad \textit{almost surely}$$

for any stationary and ergodic process $\{(X_n, Y_n)\}_{n=-\infty}^{\infty}$ satisfying (5.9).

For the proof we need the following corollary of Theorem 5.2.

Corollary 5.1. *Under the conditions of Theorem 5.2,*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \left(\mathbb{E}\{Y_t \mid X_{-\infty}^{t-1}\} - \overline{g}_t(X_1^{t-1})\right)^2 = 0 \quad \text{a.s.} \tag{5.18}$$

where $\overline{g}_t$ is the predictor for squared loss $\ell(x,y) = (x-y)^2$ in noisy setting.

Proof. The ergodic theorem implies that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \mathbb{E}\left\{ \left(Y_t - \mathbb{E}\{Y_t \mid X_{-\infty}^{t-1}\}\right)^2 \Big| X_{-\infty}^{t-1} \right\} = L^* \qquad \text{a.s.}$$

and note that

$$\begin{aligned} \mathbb{E}\{ \left(Y_t - g_t(X_1^{t-1})\right)^2 | X_{-\infty}^{t-1}\} &= \mathbb{E}\{\left(Y_t - \mathbb{E}\{Y_t \mid X_{-\infty}^{t-1}\}\right)^2 \mid X_{-\infty}^{t-1}\} \\ &\quad + \left(\mathbb{E}\{Y_t \mid X_{-\infty}^{t-1}\} - g_t(X_1^{t-1})\right)^2, \end{aligned}$$

therefore in order to finish the proof it suffices to show

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \mathbb{E}\left\{ \left(Y_t - g_t(X_1^{t-1})\right)^2 \Big| X_{-\infty}^{t-1} \right\} = L^* \qquad \text{a.s.} \tag{5.19}$$

By Theorem 5.2 with squared loss, we have

$$\lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\left(Y_t-g_t(X_1^{t-1})\right)^2=L^* \qquad \text{a.s.}$$

Thus, for (5.19), we have to prove that

$$\begin{aligned}&\frac{1}{n}\sum_{t=1}^{n}\left(\left(Y_t-g_t(X_1^{t-1})\right)^2-\mathbb{E}\{\left(Y_t-g_t(X_1^{t-1})\right)^2\mid X_{-\infty}^{t-1}\}\right)\\&\quad=\frac{1}{n}\sum_{t=1}^{n}\left(Y_t^2-\mathbb{E}\{Y_t^2\mid X_{-\infty}^{t-1}\}\right)\\&\qquad-2\frac{1}{n}\sum_{t=1}^{n}g_t(X_1^{t-1})(Y_t-\mathbb{E}\{Y_t\mid X_{-\infty}^{t-1}\})\to 0 \qquad \text{a.s.}\end{aligned}$$

By the ergodic theorem and the assumption (5.9) we have

$$\frac{1}{n}\sum_{t=1}^{n}\left(Y_t^2-\mathbb{E}\{Y_t^2\mid X_{-\infty}^{t-1}\}\right)\to 0 \qquad \text{a.s.}$$

and

$$\frac{1}{n}\sum_{t=1}^{n}(Y_t-\mathbb{E}\{Y_t\mid X_{-\infty}^{t-1}\})\to 0 \qquad \text{a.s.}$$

which imply the assertion. □

Proof of Theorem 5.3. Because of (5.17) we have to show that

$$\limsup_{n\to\infty}R_n(\overline{f})\le R^* \quad \text{a.s.}$$

Introduce the Bayes classification scheme using the infinite past:

$$f_t^*(X_{-\infty}^{t-1})=\begin{cases}1 & \text{if } \mathbb{P}\{Y_t=1\mid X_{-\infty}^{t-1}\}>1/2;\\ 0 & \text{otherwise,}\end{cases}$$

and its normalized cumulative $0-1$ loss:

$$R_n(f^*)=\frac{1}{n}\sum_{t=1}^{n}I_{\{f_t^*(X_{-\infty}^{t-1})\neq Y_t\}}.$$

Put

$$\bar{R}_n(\overline{f})=\frac{1}{n}\sum_{t=1}^{n}\mathbb{P}\{\overline{f}_t(X_1^{t-1})\neq Y_t\mid X_{-\infty}^{t-1}\}$$

and

$$\bar{R}_n(f^*) = \frac{1}{n}\sum_{t=1}^{n} \mathbb{P}\{f_t^*(X_{-\infty}^{t-1}) \neq Y_t \mid X_{-\infty}^{t-1}\}.$$

Because of assumption (5.9) we have

$$R_n(\overline{f}) - \bar{R}_n(\overline{f}) \to 0 \qquad \text{a.s.}$$

and

$$R_n(f^*) - \bar{R}_n(f^*) \to 0 \qquad \text{a.s.},$$

moreover, by the Breiman ergodic theorem

$$\bar{R}_n(f^*) \to R^* \qquad \text{a.s.}$$

so we have to show that

$$\limsup_{n\to\infty}(\bar{R}_n(\overline{f}) - \bar{R}_n(f^*)) \leq 0 \qquad \text{a.s.}$$

Theorem 2.2 in Devroye, Györfi and Lugosi [25] implies that

$$\begin{aligned}
\bar{R}_n(\overline{f}) - \bar{R}_n(f^*) &= \frac{1}{n}\sum_{t=1}^{n}\Big(\mathbb{P}\{\overline{f}_t(X_1^{t-1}) \neq Y_t \mid X_{-\infty}^{t-1}\} \\
&\qquad -\mathbb{P}\{f_t^*(X_{-\infty}^{t-1}) \neq Y_t \mid X_{-\infty}^{t-1}\}\Big) \\
&\leq 2\frac{1}{n}\sum_{t=1}^{n}\left|\mathbb{E}\{Y_t \mid X_{-\infty}^{t-1}\} - \overline{g}_t(X_1^{t-1})\right| \\
&\leq 2\sqrt{\frac{1}{n}\sum_{t=1}^{n}\left|\mathbb{E}\{Y_t \mid X_{-\infty}^{t-1}\} - \overline{g}_t(X_1^{t-1})\right|^2} \\
&\to 0 \qquad \text{a.s.},
\end{aligned}$$

where in the last step we applied the result of Corollary 5.1. □

Acknowledgments

First of all I would like to thank my supervisor László Györfi. Laci constantly supported my research and his kind guidance and encouragement during these three years are highly appreciated. I am also grateful to András György. Andris, who was my coauthor in almost half of the papers concerning in this thesis, was excellent guide in exploring different topics in machine learning and his support was an enormous help during the development of this work. I have got very much from both of them both professionally and personally, and I am extremely grateful for that.

I would also like to thank Gábor Lugosi for initiating me into the subject of prediction of individual sequences and for the enjoyable common research in Barcelona and in Budapest. I am grateful to Tamás Linder, who is the coauthor of the results in Chapter 4; His exactitude and thoroughness inspired me. I am also thankful to Gilles Stoltz for the pleasurable common research in Paris and in Budapest.

I would also like to thank Peter Auer and István Vajda junior, who are coauthors of some results treated in this dissertation.

My colleagues from the Computer and Automation Research Institute of the Hungarian Academy of Sciences supported me in my research work. I want to thank András Antos, Levente Kocsis and Csaba Szepesvári for all their help, support, interest and valuable hints.

I would also like to thank my roommates, my ex-roommates and colleagues at V2 building of Budapest University of Technology and Economics for providing such a pleasant working atmosphere.

The financial support of High Speed Networks Laboratory, Department of Telecommunications and Media Informatics, Budapest University of Technology and Economics is gratefully acknowledge.

Finally, I would like to express my debt of gratitude to my family for their warm, loving support and understanding. I am especially grateful for the love and the support of my wife, Gabi.

To everyone who has aided my efforts, both those named here and those not, let me simply say: thank you.

Bibliography

[1] P. Algoet. Universal schemes for prediction, gambling, and portfolio selection. *Annals of Probability*, 20:901–941, 1992.

[2] P. Algoet. The strong law of large numbers for sequential decisions under uncertainity. *IEEE Transactions on Information Theory*, 40:609–634, 1994.

[3] 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, Lecture Notes in Computer Science 4264*, pages 229–243, Barcelona, Spain, Oct. 2006.

[4] P. Auer, N. Cesa-Bianchi, and P. Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine Learning*, 47:235–256, May 2002.

[5] 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.

[6] 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.

[7] P. Auer, N. Cesa-Bianchi, and C. Gentile. Adaptive and self-confident on-line learning algorithms. *Journal of Computer and System Sciences*, 64(1):48–75, 2002. A preliminary version has appeared in *Proc. 13th Ann. Conf. Computational Learning Theory*.

[8] P. Auer and Gy. Ottucsák. Bound on high-probability regret in loss-bandit game. Preprint, 2006. http://www.szit.bme.hu/˜oti/green.pdf.

[9] P. Auer and M. K. Warmuth. Tracking the best disjunction. *Machine Learning*, 32(2):127–150, 1998.

[10] B. Awerbuch, D. Holmer, H. Rubens, and R. Kleinberg. Provably competitive adaptive routing. In *Proceedings of IEEE INFOCOM 2005*, volume 1, pages 631–641, March 2005.

[11] B. Awerbuch and R. D. Kleinberg. Adaptive routing with end-to-end feedback: distributed learning and geometric approaches. In *Proceedings of the 36th Annual ACM Symposium on the Theory of Computing, STOC 2004*, pages 45–53, Chicago, IL, USA, Jun. 2004. ACM Press.

[12] D. H. Bailey. *Sequential schemes for classifying and predicting ergodic processes.* PhD thesis, Stanford University, 1976.

[13] S. Bernstein. *The Theory of Probabilities.* Gastehizdat Publishing House, Moscow, 1946.

[14] G. Biau, K. Bleakely, L. Györfi, and Gy. Ottucsák. Nonparametric sequential prediction of time series, 2007. (submitted to JRSSB.).

[15] D. Blackwell. An analog of the minimax theorem for vector payoffs. *Pacific Journal of Mathematics*, 6:1–8, 1956.

[16] D. Bosq. *Nonparametric Statistics for Stochastic Processes. Estimation and Prediction.* Lecture Notes in Statistics, 110. Springer-Verlag, New York, 1996.

[17] O. Bousquet and M. K. Warmuth. Tracking a small set of experts by mixing past posteriors. *Journal of Machine Learning Research*, 3:363–396, Nov. 2002.

[18] L. Breiman. The individual ergodic theorem of information theory. *Annals of Mathematical Statistics*, 28:809–811, 1957. Correction. *Annals of Mathematical Statistics*, 31:809–810, 1960.

[19] P. Brockwell and R. A. Davis. *Time Series: Theory and Methods.* Springer-Verlag, New York, Second edition, 1991.

[20] 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.

[21] N. Cesa-Bianchi and G. Lugosi. *Prediction, Learning, and Games.* Cambridge University Press, Cambridge, 2006.

[22] N. Cesa-Bianchi, G. Lugosi, and G. Stoltz. Minimizing regret with label efficient prediction. *IEEE Trans. Inform. Theory*, IT-51:2152–2162, June 2005.

[23] N. Cesa-Bianchi, Y. Mansour, and G. Stoltz. Improved second-order bounds for prediction with expert advice. In *COLT 2005, Lecture Notes in Computer Science 3559*, pages 217–232, 2005.

[24] Y. S. Chow. Local convergence of martingales and the law of large numbers. *Annals of Mathematical Statistics*, 36:552–558, 1965.

[25] L. Devroye, L. Györfi, and G. Lugosi. *A Probabilistic Theory of Pattern Recognition.* Springer-Verlag, New York, 1996.

[26] E. Gelenbe. Self–aware network homepage, `http://san.ee.ic.ac.uk/`.

[27] E. Gelenbe, M. Gellman, R. Lent, P. Liu, and P. Su. Autonomous smart routing for network QoS. In *Proceedings of First International Conference on Autonomic Computing*, pages 232–239, New York, May 2004. IEEE Computer Society.

[28] E. Gelenbe, R. Lent, and Z. Xhu. Measurement and performance of a cognitive packet network. *Journal of Computer Networks*, 37:691–701, 2001.

[29] E. Gelenbe, P. Liu, and J. Laine. Genetic algorithms for autonomic route discovery. In *Distributed Intelligent Systems: Collective Intelligence and Its Applications, 2006. DIS 2006. IEEE Workshop on*, pages 371–376, June 2006.

[30] L. Györfi, W. Härdle, P. Sarda, and P. Vieu. *Nonparametric Curve Estimation from Time Series.* Lecture Notes in Statistics, 60. Springer-Verlag, Berlin, 1989.

[31] L. Györfi, M. Kohler, A. Krzyzak, and H. Walk. *A Distribution-Free Theory of Nonparametric Regression.* Springer, New York, 2002.

[32] 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.

[33] L. Györfi, G. Lugosi, and G. Morvai. A simple randomized algorithm for consistent sequential prediction of ergodic time series. *IEEE Transactions on Information Theory*, 45:2642–2650, 1999.

[34] L. Györfi, G. Lugosi, and F. Udina. Nonparametric kernel-based sequential investment strategies. *Mathematical Finance*, 16:337–357, 2006.

[35] L. Györfi and Gy. Ottucsák. Sequential prediction of unbounded stationary time series. *IEEE Transactions on Information Theory*, 53:1866–1872, 2007.

[36] A. György, T. Linder, and G. Lugosi. Efficient algorithms and minimax bounds for zero-delay lossy source coding. *IEEE Transactions on Signal Processing*, 52:2337–2347, Aug. 2004.

[37] A. György, T. Linder, and G. Lugosi. A "follow the perturbed leader"-type algorithm for zero-delay quantization of individual sequences. In *Proc. Data Compression Conference*, pages 342–351, Snowbird, UT, USA, Mar. 2004.

[38] A. György, T. Linder, and G. Lugosi. Tracking the best of many experts. In *Proceedings of the 18th Annual Conference on Learning Theory, COLT 2005, Lecture Notes in Computer Science 3559*, pages 204–216, Bertinoro, Italy, Jun. 2005. Springer.

[39] A. György, T. Linder, and G. Lugosi. Tracking the best quantizer. In *Proceedings of the IEEE International Symposium on Information Theory*, pages 1163–1167, Adelaide, Australia, June-July 2005.

[40] A. György, T. Linder, G. Lugosi, and Gy. Ottucsák. The on-line shortest path problem under partial monitoring. *Journal of Machnine Learning Research (accepted)*, 2007.

[41] A. György, T. Linder, and Gy. Ottucsák. The shortest path problem under partial monitoring. In *Proc. of 19th Annual Conference on Learning Theory, COLT 2006, Lecture Notes in Computer Science 4005*, pages 468–482, Pittsburgh, USA, June 2006.

[42] A. György and Gy. Ottucsák. Adaptive routing using expert advice. *The Computer Journal*, 49(2):180–189, 2006.

[43] 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.

[44] S. Hart and A. Mas-Colell. A simple adaptive procedure leading to correlated equilibrium. *Econometria*, 68(5):181–200, 2002.

[45] D. P. Helmbold and S. Panizza. Some label efficient learning results. In *Proceedings of the 10th Annual Conference on Computational Learning Theory*, pages 218–230. ACM press, 1997.

[46] M. Herbster and M. K. Warmuth. Tracking the best expert. *Machine Learning*, 32(2):151–178, 1998.

[47] M. Herbster and M. K. Warmuth. Tracking the best linear predictor. *Journal of Machine Learning Research*, 1:281–309, 2001.

[48] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58:13–30, 1963.

[49] A. Kalai and S Vempala. Efficient algorithms for the online decision problem. In B. Schölkopf and M. Warmuth, editors, *Proceedings of the 16th Annual Conference on Learning Theory and the 7th Kernel Workshop, COLT-Kernel 2003, Lecture Notes in Computer Science 2777*, pages 26–40, New York, USA, Aug. 2003. Springer.

[50] A. Kalai and S Vempala. Efficient algorithms for on-line optimization. *Journal of Computer and System Sciences*, 71(3):291–307, 2005.

[51] U. Krengel. *Ergodic Theorems*. Walter de Gruyter, Berlin-New York, 1985.

[52] T.L. Lai and H. Robbins. Asymptotically efficient adaptive allocation rules. *Advances in Applied Mathematics*, 6:4–22, 1985.

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

[54] H. B. McMahan and A. Blum. Online geometric optimization in the bandit setting against an adaptive adversary. In *Proceedings of the 17th Annual Conference on Learning Theory, COLT 2004, Lecture Notes in Computer Science 3120*, pages 109–123, Banff, Canada, Jul. 2004. Springer.

[55] M. Mohri. General algebraic frameworks and algorithms for shortest distance problems. Technical Report 981219-10TM, AT&T Labs Research, 1998.

[56] G. Morvai, S. Yakowitz, and P. Algoet. Weakly convergent stationary time series. *IEEE Transactions on Information Theory*, 43:483–498, 1997.

[57] G. Morvai, S. Yakowitz, and L. Györfi. Nonparametric inference for ergodic, stationary time series. *Annals of Statistics*, 24:370–379, 1996.

[58] A. Nobel. On optimal sequential prediction for general processes. *IEEE Transactions on Information Theory*, 49:83–98, 2003.

[59] D.S. Ornstein. Guessing the next output of a stationary process. *Israel Journal of Mathematics*, 30:292–296, 1978.

[60] Gy. Ottucsák and L. Györfi. Sequential prediction of binary sequences with side information only. In *IEEE International Symposium on Information Theory 2007*, pages 2351–2355, Nice, France, June 2007.

[61] Gy. Ottucsák and A. György. A combination of the label efficient and the multi-armed bandit problems in adversarial setting. Preprint, 2005. http://www.szit.bme.hu/~oti/preprints/comb.pdf.

[62] Gy. Ottucsák and I. Vajda. An asymptotic analysis of the mean-variance portfolio selection. *Statistics & Decisions*, 25:63–88, 2007.

[63] H. Robbins. Some aspects of the sequential design of experiments. *Bullettin of the American Mathematical Society*, 55:527–535, 1952.

[64] R. E. Schapire and D. P. Helmbold. Predicting nearly as well as the best pruning of a decision tree. *Machine Learning*, 27:51–68, 1997.

[65] A. C. Singer and M. Feder. Universal linear prediction by model order weighting. *IEEE Transactions on Signal Processing*, 47:2685–2699, 1999.

[66] A. C. Singer and M. Feder. Universal linear least-squares prediction. In *Proceedings of the IEEE International Symposium on Information Theory*, 2000.

[67] W. F. Stout. *Almost sure convergence*. Academic Press, New York, 1974.

[68] E. Takimoto and M. K. Warmuth. Path kernels and multiplicative updates. In J. Kivinen and R. H. Sloan, editors, *Proceedings of the 15th Annual Conference on Computational Learning Theory, COLT 2002, Lecture Notes in Computer Science 2375*, Lecture Notes in Computer Science 2375, pages 74–89, Berlin–Heidelberg, Jul. 2002. Springer-Verlag.

[69] E. Takimoto and M. K. Warmuth. Path kernels and multiplicative updates. *Journal of Machine Learning Research*, 4:773–818, 2003.

[70] V. Vovk. Aggregating strategies. In *Proceedings of the Third Annual Workshop on Computational Learning Theory*, pages 372–383, Rochester, NY, Aug. 1990. Morgan Kaufmann.

[71] V. Vovk. Derandomizing stochastic prediction strategies. *Machine Learning*, 35(3):247–282, Jun. 1999.

[72] T. Weissman and N. Merhav. Universal prediction of binary individual sequences in the presence of noise. *IEEE Trans. Inform. Theory*, 47(6):2151–2173, July 2001.

[73] T. Weissman and N. Merhav. Universal prediction of random binary sequences in a noisy environment. *Annals of Applied Probability*, 14(1):54–89, Feb. 2004.

[74] Y. Yang. Combining different procedures for adaptive regression. *Journal of Multivariate Analysis*, 74:135–161, 2000.

[75] U. Yule. On a method of investigating periodicities in disturbed series, with special reference to wölfer's sunspot numbers. *Philos. Trans. Roy. Soc.*, 226:267–298, 1927.