# On Sequential Strategies for Loss Functions with Memory*


Neri Merhav†
Electrical Engineering Department
Technion
Haifa 32000, Israel

Erik Ordentlich‡
GlobespanVirata, Inc.
2500 Walsh Ave.
Santa Clara, CA 95051, USA

Gadiel Seroussi and Marcelo J. Weinberger
Hewlett-Packard Laboratories
1501 Page Mill Road
Palo Alto, CA 94304, USA.



**Abstract**

The problem of optimal sequential decision for individual sequences, relative to a class of competing off-line reference strategies, is studied for general loss functions with memory. This problem is motivated by applications in which actions may have "long term" effects, or there is a cost for switching from one action to another. As a first step, we consider the case in which the reference strategies are taken from a finite set of generic "experts." We then focus on finite-state reference strategies, assuming finite action and observation spaces. We show that key properties that hold for finite-state strategies in the context of memoryless loss functions, do not carry over to the case of loss functions with memory. As a result, an infinite family of randomized finite-state strategies is seen to be the most appropriate reference class for this case, and the problem is basically different from its memoryless counterpart. Based on Vovk's exponential weighting technique, infinite-horizon on-line decision schemes are devised. For an arbitrary sequence of observations of length $n$, the excess normalized loss of these schemes relative to the best expert in a corresponding reference class is shown to be upper-bounded by an $O(n^{-1/3})$ term in the case of a finite class, or an $O([(\ln n)/n]^{1/3})$ term for the class of randomized finite-state strategies. These results parallel the $O(n^{-1/2})$ bounds attained by previous schemes for memoryless loss functions. By letting the number of states in the reference class grow, the notion of finite-state predictability is also extended.



**Index Terms:** Sequential decision, on-line algorithms, general loss functions, prediction, expert advice, randomized expert.


---

*Parts of this paper were presented at the 2000 International Symposium on Information Theory, Sorrento, Italy, and at the 2001 International Symposium on Information Theory, Washington D.C., USA.

†Work partially done while this author was visiting at Hewlett-Packard Laboratories, Palo Alto, California.

‡Work partially done while this author was with Hewlett-Packard Laboratories, Palo Alto, California.

# 1 Introduction

The *sequential decision problem* [1, 2] concerns a temporal sequence of *observations* $x^n = x_1x_2\cdots x_n$ for which corresponding *actions* $b_1b_2\cdots b_n$ result in instantaneous *losses* $\ell(b_t, x_t)$ for each time instant $t$, $1 \le t \le n$, where $\ell(\cdot,\cdot)$ denotes a non-negative function. The instantaneous loss contributions from each action-observation pair yield a cumulative loss

$$\mathcal{L}_b(x^n) = \sum_{t=1}^{n} \ell(b_t, x_t)\,.$$

A sequential (or "on-line") decision algorithm (or "strategy") is a sequence of actions $b_1b_2\cdots b_n$ in which the actions $b_t$ are (possibly random) functions of the previous observations $x^{t-1}$ only. The goal in the sequential decision problem is to find an on-line strategy that approximates, in the long run and for an *arbitrary* individual sequence of observations, the performance of the best strategy in a given reference class. The best competing strategy is determined in hindsight (or "off-line"), with *full knowledge* of the given sequence of observations. The excess loss incurred by the on-line strategy over the best competing strategy is termed the *regret*. Universal lossless data compression is clearly a special case of the sequential decision problem, where $b_t$ is a conditional probability assignment $p(\cdot|x^{t-1})$ for $x_t$ and $\ell(b_t, x_t) = -\log p(x_t|x^{t-1})$. The regret corresponds to the *pointwise redundancy* of a sequential probability assignment scheme with respect to a class of models. Other instances of the sequential decision problem include gambling, prediction, and portfolio selection (see, e.g., [2]).

In the original formulation of the sequential decision problem [1], the reference class is given by all *constant* strategies, and the best constant strategy is termed the *Bayes response*. In [3], Ziv and Lempel extend this class in the context of universal compression of individual sequences, to include all *finite-state* (FS) encoders. A similar approach is used in [4] to study binary prediction, leading to the notion of FS *predictability*. This work is extended in [5] for general loss functions. In the FS setting, rather than being constant, the competing reference strategy is allowed to vary according to $b_t = g(s_t)$, where $s_t$ is a state in an FS machine (FSM) with state set $S$, driven by a next-state function $s_{t+1} = f(s_t, x_t)$. The functions $g$ and $f$, and the initial state $s_1$, are optimized off-line, with full knowledge of $x^n$. The (normalized) loss over an infinite sequence ($n \to \infty$) is considered, and a (generalized) FS predictability results from letting $|S|$ tend to infinity. Notice that, for given $f$ and $s_1$, the loss incurred with $g$ takes the form

$$\mathcal{L}_g(x^n) = \sum_{s\in S}\sum_{x\in A} n(x,s)\ell(g(s),x) \tag{1}$$

where for an observation $x$ (in a discrete alphabet $A$) and a state $s \in S$, $n(x,s)$ denotes the number of times $x_t = x$ and $s_t = s$, $t = 1, 2, \ldots, n$. Thus, the optimal $g$ consists of using, for each state, the Bayes response corresponding to the sub-sequence of observations occurring at that state, namely

$$g(s) = \arg\min_{b\in B} E_x[\ell(b,x)|s] \tag{2}$$

where $B$ is the action space and the expectation on $x$ is taken with respect to the conditional empirical distribution

$$p(x|s) = \frac{n(x,s)}{\sum_{x'\in A} n(x',s)}\,,\ x \in A,\ s \in S. \tag{3}$$

The function defined in (2) is *deterministic* and depends on $x^n$ only through the empirical distribution (3). Equation (1) also suggests a probabilistic analogy, in which the data are drawn from an FS source $\{p(x|s),\, x \in A,\, s \in S\}$, with the next-state function $f$ still dictating the evolution of the state. In that case, the expected loss for *any* strategy $b_t = \mu_t(x^{t-1}, b^{t-1})$ (not necessarily FS) is given by

$$\bar{\mathcal{L}}_{p,\mu} = \sum_{t=1}^{n}\sum_{s\in S}\sum_{b\in B} P_t(s,b) E_x[\ell(b,x)|s] \tag{4}$$

where $P_t(s,b)$ is the joint probability (with respect to $\{p(x|s)\}$ and $\{\mu_t\}$) that $(s_t, b_t) = (s,b)$. Again, the expected loss in (4) is minimized over $\mu$ by the deterministic FS strategy $\mu_t(x^{t-1}, b^{t-1}) = g(s_t)$, where $g$ is as in (2) and the expectation is with respect to the FS source.

The theory of learning with expert advice [6, 7, 8, 9] is a natural extension of the above framework, where the class of reference competing strategies is viewed as a set of *experts*. In this setting, an on-line strategy is expected to combine the (possibly random) advice of the experts, incurring a loss that approaches that of the best performing expert for each individual sequence.

The classical formulation of the sequential decision problem does not address cases in which the instantaneous losses also depend on past action-observation pairs. Such losses may capture the cost of switching from one action to another (e.g., transaction costs incurred in portfolio selection [10, 11], or energy spent in control systems), or the long term effect ("memory") of an action at a given time. For a loss function with limited memory of the form $\ell(b_{t-1}, b_t, x_t)$, the cumulative loss takes the form

$$\mathcal{L}_b(x^n) = \sum_{t=1}^{n} \ell(b_{t-1}, b_t, x_t) \tag{5}$$

where $b_0$ represents some initial action. For a randomized strategy, the average loss (with respect to the ensemble of randomly selected actions) takes the form

$$\bar{\mathcal{L}}_b(x^n) = \sum_{b',b\in B} \sum_{t=1}^{n} \Pr(b_{t-1} = b', b_t = b)\ell(b', b, x_t) \,. \tag{6}$$

The scenario in which there is a cost for switching from one action to another is akin to the *metrical task system* (MTS) problem (see, e.g., [12]), but differs fundamentally from it in that in the MTS problem it is assumed that the action $b_t$ *precedes* $x_t$. Connections between the MTS problem and the sequential decision problem are investigated in [13]. As for the case of actions with long term effects, an example is given by the following (inverse) filtering problem: Given a filter $H(z) = h_0 + h_1 z^{-1}$, and a (noisy) filter output sequence $x^n$ (the observations), the actions $b_t$ are aimed at finding the most likely input sequence $b_1 b_2 \cdots b_n$ that "explains" $x^n$ in the sense that $\mathcal{L}_b(x^n)$ is minimized, where

$$\ell(b_{t-1}, b_t, x_t) = [x_t - (h_0 b_t + h_1 b_{t-1})]^2 \,.$$

In other applications, e.g., *prefetching* in computer memory architectures, the memory may be much longer. In this example, the goal is to prefetch an address from main memory into a faster memory ("cache") ahead of time, in order to prevent stall time by the CPU when accessing this address. The sequence of observations is given by the address sequence, and the action $b_t$ consists in prefetching one (or several) address(es). While this problem has been studied as one of prediction (see, e.g., [14]), such formulation aims at predicting the next memory reference, rather than at having it already in cache *at the time it is requested*. Clearly, the loss (e.g., the CPU stall time caused if the address $x_t$ is not in cache at the time it is requested) will depend not only on $b_t$, but on the entire sequence of actions $b_1 b_2 \cdots b_t$. Based on an opportunity cost analysis, it is argued in [15] that a reasonable loss function for the prefetching application takes the form

$$\ell(b_1, b_2, \cdots, b_t, x_t) = \sum_{i=0}^{\nu} C_i(b_{t-i}, x_t) \tag{7}$$

where the contribution $C_i(b_{t-i}, x_t)$ of the action $b_{t-i}$ to the loss incurred at time $t$ is considered "in isolation" and assumed additive. This contribution does not extend beyond a given non-negative window size $\nu$ (the memory length). For this simplified loss, it is shown in [15] that an on-line algorithm that uses the parsing rule of [3] to build a decision tree, performs essentially as well as the best FS strategy determined in hindsight.

In this paper, we reformulate the sequential decision problem studied in [5] to cover also loss functions with memory. We will focus on general loss functions of the form (5) and (6), for which the instantaneous losses $\ell(b', b, x)$, $b', b \in B$, $x \in A$, are bounded by a given constant; extension to longer memories is relatively straightforward. Our main result will require finite observation and action spaces $A$ and $B$, respectively; in cases in which no assumptions are required, our notation will nevertheless assume a discrete alphabet $A$.[1]

Notice that the theory of Markov decision processes (MDP's; see, e.g., [16]) is an analogous generalization (in the direction of loss functions with memory) of the probabilistic setting, in which the data are drawn from an FS source (as in Equation (4)). In an MDP, the process at time $t$ is considered to be at some state $\sigma_t$ in a countable state space, and a decision maker chooses an action $b_t = b \in B$, at a cost $L(\sigma_t, b)$. As a result, the process evolves to state $\sigma_{t+1}$ according to (action-dependent) transition probabilities $\{p_b(\sigma_{t+1}|\sigma_t)\}$. In an *average cost per stage* problem, for a given initial state $\sigma_1$, the decision maker aims at a (possibly randomized) strategy $\{\mu_t\}$ that minimizes the *normalized* expected loss over $n$ stages, $n \to \infty$, where the expectation is taken with respect to the transition probabilities and to $\{\mu_t\}$. In particular, if $\sigma_t = (b_{t-1}, s_t)$, $s_t \in S$, and $p_b((b, s_{t+1})|(b', s_t)) = p(s_{t+1}|s_t)$ for all $b' \in B$, where $\{p(s|s')\}$ are transition probabilities defined on $S$, then the normalized asymptotic expected loss takes the form

$$\bar{L}_{p,\mu} = \limsup_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \sum_{s\in S} \sum_{b',b\in B} P_t(s, b', b) L((s, b'), b) \tag{8}$$

where $P_t(s, b', b)$ is the joint probability (with respect to $\{p(s|s')\}$ and $\{\mu_t\}$) that $(s_t, b_{t-1}, b_t) = (s, b', b)$. The connection with loss functions with memory in a probabilistic setting follows, since (8) gives the normalized value of the expected loss (6) as $n \to \infty$, in case the data is drawn from an FS source with conditional probabilities $\{p(x|s),\ x \in A,\ s \in S\}$ and the state

[1] For continuous spaces, probability distributions should be appropriately replaced by density functions, and summations by integrals.

sequence evolves according to the next-state function $f$, where

$$L((s,b'),b) \triangleq L(s,b',b) = \sum_{x\in A} p(x|s)\ell(b',b,x) \tag{9}$$

and

$$p(s|s') = \sum_{x:\, s=f(s',x)} p(x|s')\,. \tag{10}$$

Sequential strategies for loss functions with memory and arbitrary observation sequences are considered in [17] in a broader setting, closely related to MDP's. However, the problem in [17] is formulated in the framework of Blackwell's so-called "approachability-excludability" theory [18], in which conditions to approach a certain performance region regardless of the observations are studied for vector loss functions. While for memoryless loss functions it is possible to define a loss vector and performance regions that measure the regret (see, e.g., [1]), this does not appear to be the case for loss functions with memory.

A sensible generalization of the notion of regret used in [5], requires the determination of a set of reference FS off-line strategies (or experts). First, we consider the case in which the experts are generic (i.e., no FSM structure is imposed), and the set of experts is finite. This case covers the deterministic FS off-line strategies that lead to the generalized predictability of [5]. Inspired by [6] in the setting of prediction with expert advice, we devise on-line strategies that perform essentially as well as any expert in a finite set, for an arbitrary sequence of observations. Specifically, we present an infinite horizon on-line scheme whose normalized excess loss vanishes at a rate $O\left([(\ln\lambda)/n]^{1/3}\right)$, where $\lambda$ denotes the cardinality of the expert set. This result parallels the case of memoryless loss functions, for which the normalized excess loss for learning schemes with expert advice (as well as for classical schemes by Hannan [1], Blackwell [18], and others) is upper-bounded by an $O(n^{-1/2})$ term. While the learning algorithm of [6] (see also [8] and [9]) suggests to (randomly) select an expert at each time $t$ based on its performance on $x^{t-1}$, here the main problem is to overcome the effect of $b_{t-1}$ in the instantaneous loss at time $t$, as this action may not agree with the expert selected at that time.

Next, we consider FSM experts. In the memoryless loss case, the "ultimate" performance yardstick for FSM experts in [5] is restricted to deterministic FSM's. This restriction is natural, since it is shown in [5] that, as $|S| \to \infty$, allowing randomization of $f$ does not improve the (generalized) predictability of an individual sequence. Moreover, for a deterministic next-state function $f$, the optimal strategy defined in (2) is deterministic. However, key properties that hold for FSM expert models in the case of memoryless loss functions break as a result of an additional dependency on the previous action, which complicates the problem. For loss functions with memory, the loss incurred by the expert model $b_t = g(s_t)$ on an individual sequence $x^n$ takes the form

$$\mathcal{L}_g(x^n) = \ell(b_0, g(s_1), x_1) + \sum_{s', s \in S} \sum_{x \in A} n(x, s's) \ell(g(s'), g(s), x) \tag{11}$$

where $n(x, s's)$ denotes the number of occurrences in $x_2, x_3, \ldots, x_n$ of an observation $x$ at state $s$, which in turn is preceded by state $s'$. Although the loss still depends on $x^n$ through occurrence counts, it turns out that the minimizing expert $g$ in (11) need not be deterministic as in (1). Thus, we cannot restrict the analysis to finite expert sets. Furthermore, we cannot rule out *randomized* next-state functions, which allow for statistical dependencies between the expert's actions. When the next-state function is randomized, without loss of generality, $g$ can be assumed deterministic. An infinite horizon on-line algorithm is derived for that case, with a regret that vanishes at a rate $O\left([(\ln n)/n]^{1/3}\right)$. While our results for the finite expert set case apply to generic experts, here we make heavy use of the FSM structure of the experts.

Next, by letting the number of states in the reference class grow, the FS predictability of [4] (and [5]) is further extended to account for the loss functions with memory considered here. Notice that [4] and [5] focus on the case $|S| \to \infty$ (for which no uniform convergence rates can be derived) as these works emphasize the use of the Lempel-Ziv parsing rule [3] to compete against an FSM of any size. For a finite sequence and a given number of states, however, competing against deterministic machines, aiming at a deterministic "FS predictability" with a regret corresponding to a reduced number of experts, may not minimize the total normalized loss. For memoryless loss functions, the limiting FS predictability (as $|S| \to \infty$) is the same as with randomized machines (for which the regret is larger due to a larger number of experts), but may be achieved at a slower rate. Thus, our results extend earlier work on FS predictability not only in the direction of loss functions with memory, but also providing a more refined regret analysis, with the above trade-off to be resolved.

The rest of this paper is organized as follows. Section 2 studies on-line strategies for finite sets of generic experts. In Section 3, we impose an FSM structure on the expert set, and discuss various alternatives for the reference class. Finally, Section 4 studies on-line strategies for randomized FSM strategies, and extends the notion of FS predictability to loss functions with memory.

## 2 On-line strategies for finite sets of experts

In this section, we study on-line strategies that compete against a finite set $\mathcal{F}$ of $\lambda$ experts for any individual sequence $x^n$. No assumptions are required on the observation and action spaces $A$ and $B$, respectively. The loss function $\ell(b', b, x)$, $b', b \in B$, $x \in A$, is assumed bounded, with a maximum loss $\ell_{\max}$. At each time instant $t$, each expert $F \in \mathcal{F}$ offers its advice, which takes the form of a probability distribution on $b_t$. If $\{b_t\}$ is randomly selected according to the strategy proposed by $F$, the expected expert loss is given by

$$\bar{\mathcal{L}}_F(x_t) = \sum_{b',b \in B} \Pr{}^{(F)}(b_{t-1} = b', b_t = b)\ell(b', b, x_t) \tag{12}$$

where $\Pr^{(F)}(\cdot)$ denotes probability with respect to the ensemble of possibly random actions of $F$. Notice that the (possibly random) information used by the expert to generate its advice is unspecified, and irrelevant at this point. An on-line strategy combines the advice of the experts, to derive a distribution on $b_t$ from which the action is drawn.

**General Approach.** In principle, it would appear that we can obtain a suitable on-line strategy for finite expert sets by directly applying the theory of learning with expert advice of [7] and [6]. A subtle complication introduced by the memory in the loss function, however, prevents such a direct application. The learning algorithms of [7] and [6] suggest to randomly select, at each time instant $t$, an element $F$ of $\mathcal{F}$ based on its expected performance on $x^{t-1}$, and draw $b_t$ according to the distribution proposed by $F$. The complication arises as these algorithms count on the fact that the average on-line loss, given the expert selection $F$, is equal to the average loss achieved by $F$ on the new observation $x_t$. This equivalence, however, may not hold for loss functions with memory, since the instantaneous losses of the on-line strategy and the expert on the new observation also depend on their respective previous actions, which may differ.

It is nevertheless possible to adapt the general approach of learning with expert advice to tracking the best strategy in $\mathcal{F}$ in the present setting. To this end, we will introduce an additional block-length parameter that determines how long we will follow the advice of a randomly selected new expert. The goal is to amortize the potential discrepancy between on-line and expert losses at the start of each block (the "cost of switching experts"), over the length of the block. Notice that "piecewise constant" strategies that follow a fixed expert during a certain block are considered in [19]. However, the piecewise constant strategies in [19] are used for off-line reference, as a generalization of competing against the best fixed expert. Our on-line strategy, inspired by [6], is first presented for the finite horizon case, in which $n$ is known in advance.

**Finite Horizon Case.** For a fixed block length $M$, let $K = \lceil n/M \rceil - 1$, and consider the parsing of $x^n$ into $K$ non-overlapping blocks of length $M$, $X_k = (x_{kM+1}, x_{kM+2}, \ldots, x_{kM+M})$, $k = 0, 1, \ldots, K-1$, followed by a possibly shorter block $X_K = (x_{KM+1}, x_{KM+2}, \ldots, x_n)$. Let $\bar{\mathcal{L}}_{F,k}$ denote the cumulative value over $k$ blocks of the loss (12) of expert $F$, with $\bar{\mathcal{L}}_{F,0} = 0$ for all $F$. The multiset of $\lambda$ cumulative losses $\{\bar{\mathcal{L}}_{F,k},\ F \in \mathcal{F}\}$, is denoted by $\mathbf{L}_{\mathcal{F},k}$. At $t = kM + 1$, $k = 0, 1, \ldots, K$, the on-line algorithm randomly selects $F \in \mathcal{F}$ according to the conditional distribution

$$P_k(F|\mathbf{L}_{\mathcal{F},k}) = \frac{\exp\{-\eta \bar{\mathcal{L}}_{F,k}\}}{\sum_{F' \in \mathcal{F}} \exp\{-\eta \bar{\mathcal{L}}_{F',k}\}} \tag{13}$$

where $\eta$ is a given positive constant. In (13), the experts are exponentially weighted based on their past performance. The actions of the selected expert are followed through $t = \min(kM + M, n)$ (i.e., during $X_k$). Notice that two different randomizations are involved: one with respect to the choice of expert, at a block level, and one with respect to the possibly random expert actions, at a sample level. Theorem 1 below gives an upper-bound on the normalized regret of the proposed scheme.

**Theorem 1** *Let*

$$M^* = 2\left(\frac{n}{\ln \lambda}\right)^{\frac{1}{3}} \tag{14}$$

*and*

$$\eta^* = \frac{2}{\ell_{\max}}\left(\frac{\ln \lambda}{n}\right)^{\frac{2}{3}}. \tag{15}$$

*Let the expected loss of the proposed on-line strategy with parameters $\eta = \eta^*$ and $M = \lceil M^* \rceil$, over $x^n$, be denoted by $\bar{\mathcal{L}}_{\text{fes/fh}}(x^n)$ (finite expert set/finite horizon). Let $\bar{\mathcal{L}}_{\min}(x^n)$ denote the* *loss of the expert in $\mathcal{F}$ that performs best on $x^n$. Then,*

$$\frac{1}{n}\left(\bar{\mathcal{L}}_{\text{fes/fh}}(x^n) - \bar{\mathcal{L}}_{\min}(x^n)\right) \leq \frac{3}{2}\,\ell_{\max}\left(\frac{\ln\lambda}{n}\right)^{\frac{1}{3}} + \frac{\ell_{\max}(1+(\ln\lambda)/8)}{n} + O\left(n^{-\frac{4}{3}}\right). \tag{16}$$

*Proof:* The bounding technique (similar to the one used for proving [20, Theorem 1], which builds on [7]) is based on the sums $W_k$, $0 \leq k \leq K+1$, defined by

$$W_k \triangleq \sum_{F\in\mathcal{F}} \exp\{-\eta\bar{\mathcal{L}}_{F,k}\}\,. \tag{17}$$

It consists in comparing an upper bound and a lower bound to $\ln(W_{K+1}/W_0)$. For the lower bound, we have:

$$\begin{aligned}\ln\frac{W_{K+1}}{W_0} &= \ln\sum_{F\in\mathcal{F}}\exp\{-\eta\bar{\mathcal{L}}_{F,K+1}\} - \ln\lambda \\ &\geq \ln\max_{F\in\mathcal{F}}\exp\{-\eta\bar{\mathcal{L}}_{F,K+1}\} - \ln\lambda \\ &= -\eta\bar{\mathcal{L}}_{\min}(x^n) - \ln\lambda \end{aligned} \tag{18}$$

where the first equality follows from $\bar{\mathcal{L}}_{F,0} = 0$, so that $W_0 = \lambda$, and the last equality follows from the definition of $\bar{\mathcal{L}}_{F,K+1}$ as the total loss accumulated by $F$ over $K+1$ blocks, namely over the entire sequence $x^n$.

To get an upper bound, we first define $\bar{\mathcal{L}}_F(X_k)$ as the average loss accumulated by an expert $F \in \mathcal{F}$ over block $X_k$, namely

$$\bar{\mathcal{L}}_F(X_k) = \sum_{b',b\in B}\sum_{t=kM+1}^{m_k}\Pr{}^{(F)}(b_{t-1}=b', b_t=b)\,\ell(b',b,x_t) \tag{19}$$

where $m_k = kM + M$, $0 \leq k < K$, and $m_K = n$. With this definition, for $k = 0, 1, \ldots, K$,

$$\bar{\mathcal{L}}_{F,k+1} = \sum_{i=0}^{k}\bar{\mathcal{L}}_F(X_i) = \bar{\mathcal{L}}_{F,k} + \bar{\mathcal{L}}_F(X_k)\,. \tag{20}$$

Thus, by (17), (20), and (13),

$$\ln\frac{W_{k+1}}{W_k} = \ln\sum_{F\in\mathcal{F}} P_k(F|\mathbf{L}_{\mathcal{F},k})\exp\{-\eta\bar{\mathcal{L}}_F(X_k)\}\,. \tag{21}$$

Next, recall that Hoeffding's inequality [21] asserts that for any zero-mean random variable $Z$ taking values in a bounded interval of size $R$, $\ln E[e^Z] \leq R^2/8$. A trivial extension that allows for a general mean $\mu$ of $Z$, gives

$$\ln E[e^Z] \leq \mu + \frac{R^2}{8}\,. \tag{22}$$

Applying (22) to the right-hand side of (21), we get

$$\ln \sum_{F\in\mathcal{F}} P_k(F|\mathbf{L}_{\mathcal{F},k}) \exp\{-\eta \bar{\mathcal{L}}_F(X_k)\} \le -\eta E_F\{\bar{\mathcal{L}}_F(X_k)|\mathbf{L}_{\mathcal{F},k}\} + \frac{\eta^2 M^2 \ell_{\max}^2}{8} \tag{23}$$

where the expectation on $F$ is with respect to $P_k(\cdot|\mathbf{L}_{\mathcal{F},k})$. Combining equations (21) and (23) and taking the summation over $k = 0, 1, \ldots, K$, we get

$$\begin{aligned}\ln \frac{W_{K+1}}{W_0} &= \sum_{k=0}^{K} \ln \frac{W_{k+1}}{W_k} \\ &\le -\eta \sum_{k=0}^{K} E_F\{\bar{\mathcal{L}}_F(X_k)|\mathbf{L}_{\mathcal{F},k}\} + \frac{\eta^2 M^2 \ell_{\max}^2 (K+1)}{8}\,. \end{aligned} \tag{24}$$

The summation on the right-most side of (24) differs from $\bar{\mathcal{L}}_{\text{fes/fh}}(x^n)$ in that it does not account for the possible change in the expert selection from one block to another, which only affects the term $\Pr^{(F)}(b_{kM} = b', b_{kM+1} = b)$ in (19), as $b_{kM}$ follows the expert used in the previous block. This correction cannot cost more than an additional loss of $\ell_{\max}$ for each block $X_k$, $k > 0$, and therefore

$$\bar{\mathcal{L}}_{\text{fes/fh}}(x^n) \le \sum_{k=0}^{K} E_F\{\bar{\mathcal{L}}_F(X_k)|\mathbf{L}_{\mathcal{F},k}\} + K\ell_{\max}\,.$$

Thus, by (24),

$$\begin{aligned}\ln \frac{W_{K+1}}{W_0} &\le -\eta \bar{\mathcal{L}}_{\text{fes/fh}}(x^n) + \frac{\eta^2 M^2 \ell_{\max}^2 (K+1)}{8} + \eta K \ell_{\max} \\ &\le -\eta \bar{\mathcal{L}}_{\text{fes/fh}}(x^n) + \frac{\eta^2 \ell_{\max}^2 M(n+M-1)}{8} + \eta \ell_{\max} \frac{n-1}{M}\,. \end{aligned} \tag{25}$$

Finally, by comparing the upper and the lower bounds on $\ln(W_{K+1}/W_0)$, (18) and (25), we get

$$\bar{\mathcal{L}}_{\text{fes/fh}}(x^n) \le \bar{\mathcal{L}}_{\min}(x^n) + \frac{\ln \lambda}{\eta} + \frac{\eta \ell_{\max}^2 M(n+M-1)}{8} + \ell_{\max}\frac{n-1}{M}\,. \tag{26}$$

For a given horizon $n$, the design parameters $M^*$ and $\eta^*$ given in (14) and (15), respectively, approximately minimize the regret. Using $M = \lceil M^* \rceil$ and $\eta = \eta^*$ yields the upper-bound (16), where the last two terms in the right-hand side account for the discrepancy between $M$ and $M^*$, and for the possibly incomplete last block $X_K$. □

**Infinite Horizon Case.** For infinite horizon, we use the "doubling technique" (see, e.g., [9]), dividing time into non-overlapping contiguous super-segments of lengths $N_j = \lceil n_0 a^j \rceil$, $j = 0, 1, \cdots$, for given positive constants $n_0$ and $a$, $a > 1$. In the $(j+1)$-st super-segment, the above finite horizon algorithm is used with design parameters $M = \lceil M_j \rceil$ and $\eta = \eta_j$, where $M_j$ and $\eta_j$ are specified by (14) and (15), respectively, but with $N_j$ replacing $n$. The expected loss of this infinite horizon scheme over a prefix $x^n$ of a sequence of observations is denoted $\bar{\mathcal{L}}_{\text{fes/ih}}(x^n)$ (finite expert set/infinite horizon). Theorem 2 below states that for *all* $n$, the normalized regret is bounded as in Theorem 1, but with a larger constant.

**Theorem 2** *For any prefix $x^n$ of an infinite sequence of observations,*

$$\frac{1}{n}\left(\bar{\mathcal{L}}_{\text{fes/ih}}(x^n) - \bar{\mathcal{L}}_{\min}(x^n)\right) \le \frac{3}{2}\, C(a)\ell_{\max}\left(\frac{\ln \lambda}{n}\right)^{\frac{1}{3}} + O\left(\frac{\ln n}{n}\right) \tag{27}$$

*where*

$$C(a) = \frac{[a(a-1)]^{\frac{2}{3}}}{a^{\frac{2}{3}} - 1}\,.$$

*Proof:* Let $J(n)$ denote the index of the super-segment in which $x_n$ falls, and let $Y_j$, $0 \le j \le J(n)$, denote the sequence of observations that fall in the $(j+1)$-st super-segment. We have $Y_j = (x_{T_j+1}, x_{T_j+2}, \ldots, x_{T_{j+1}})$, where $T_0 \stackrel{\Delta}{=} 0$, $T_{j+1} - T_j = N_j$ for $0 \le j < J(n)$, and $T_{J(n)+1} \stackrel{\Delta}{=} n \le T_{J(n)} + N_{J(n)}$. By definition of the infinite horizon scheme, we have

$$\left|\bar{\mathcal{L}}_{\text{fes/ih}}(x^n) - \sum_{j=0}^{J(n)} \bar{\mathcal{L}}_{\text{fes/fh}}(Y_j)\right| \le (J(n)+1)\ell_{\max} \tag{28}$$

where the difference is due to the possible discrepancy between $b_0$ and the last action of each super-segment. We can upper-bound each loss $\bar{\mathcal{L}}_{\text{fes/fh}}(Y_j)$, $0 \le j < J(n)$, using Theorem 1, and the bound is still valid if we replace the loss of the best performing expert in each super-segment with that of the expert $F$ that performs best on $x^n$. On the other hand, the design parameters used in the last super-segment $Y_{J(n)}$, of length $n - T_{J(n)}$, are optimized for a possibly longer horizon $N_{J(n)}$. However, notice that for given design parameters, the regret terms in (26) grow with the sequence length, and therefore the upper bound still holds if we replace the sequence length with the (possibly longer) horizon for which the parameters are optimized. Thus, the (unnormalized) bound of Theorem 1 is still valid if we replace the sequence length with a longer horizon in the regret term. As a result, for $0 \le j \le J(n)$ we have

$$\bar{\mathcal{L}}_{\text{fes/fh}}(Y_j) \le \bar{\mathcal{L}}_F(Y_j) + \frac{3}{2}\,\ell_{\max}[(\ln \lambda)N_j^2]^{\frac{1}{3}} + O(1)$$

where $\bar{\mathcal{L}}_F(Y_j)$ denotes the expected loss of $F$ over $Y_j$. Therefore, (28) implies

$$\frac{1}{n}\bar{\mathcal{L}}_{\text{fes/ih}}(x^n) \le \frac{1}{n}\bar{\mathcal{L}}_{\min}(x^n) + \frac{3\ell_{\max}}{2n}\sum_{j=0}^{J(n)}[(\ln \lambda)N_j^2]^{\frac{1}{3}} + O\left(\frac{J(n)}{n}\right)\,. \tag{29}$$

Now, assuming $N_0 > 1$,

$$\begin{aligned}\sum_{j=0}^{J(n)} N_j^{\frac{2}{3}} &= \sum_{j=0}^{J(n)} (N_j-1)^{\frac{2}{3}}\left(1+\frac{1}{N_j-1}\right)^{\frac{2}{3}} \\ &< \sum_{j=0}^{J(n)} (N_j-1)^{\frac{2}{3}}\left(1+\frac{2}{3(N_j-1)}\right) \\ &< \sum_{j=0}^{J(n)} (n_0a^j)^{\frac{2}{3}} + \frac{2(J(n)+1)}{3}\,. \end{aligned} \tag{30}$$

In addition,

$$n > T_{J(n)} = \sum_{i=0}^{J(n)-1} N_i \geq n_0 \cdot \frac{a^{J(n)}-1}{a-1}\,.$$

Thus,

$$J(n) < \frac{\ln[1+n(a-1)/n_0]}{\ln a} \tag{31}$$

which also yields

$$\sum_{j=0}^{J(n)} a^{\frac{2j}{3}} < \frac{\left(\frac{na(a-1)}{n_0}+a\right)^{\frac{2}{3}}-1}{a^{\frac{2}{3}}-1} < \frac{\left(\frac{na(a-1)}{n_0}\right)^{\frac{2}{3}}}{a^{\frac{2}{3}}-1} \tag{32}$$

where the last inequality holds for all $n > 8n_0a^2/[27(a-1)]$. The theorem follows from equations (29)–(32). □

Notice that the optimal choice of $a$ and $n_0$ depends on $n$. As $n \to \infty$, the optimal $a$ is the value that minimizes $C(a)$, which is close to $a = 1.64$, and the bound on the regret is about 2.64 times higher than in the finite horizon case. However, any choice $a > 1$ guarantees a regret that vanishes at the same rate as in the finite horizon case.

**Almost-Sure Convergence.** Next, we establish convergence of the on-line loss to $\bar{\mathcal{L}}_{\min}(x^n)$ in the almost-sure sense for the random choice of experts, while still maintaining, at this point, the average performance measure with respect to the possibly random actions of each expert. Almost-sure convergence with respect to the latter requires additional assumptions on the experts, so as to guarantee that the loss obeys a "law of large numbers." Such assumptions will be made in Section 4. Here, a slightly different choice of the block length parameter is needed. Let $\mathcal{L}_{\text{fes/ih}}(x^n)$ denote the (random) average loss incurred over $x^n$ by a specific realization of the randomly selected experts according to the infinite horizon scheme (where $N_j$ is as in Theorem 2 but $M_j$ and $\eta_j$ are still unspecified).

**Theorem 3** *Let* $M_j = K_1 N_j^{\frac{1}{3}-\varphi_1}$ *and* $\eta_j = K_2 N_j^{-\frac{1}{3}-\varphi_2}$, *where* $0 < \varphi_1 < 1/3$, $0 < \varphi_2 < 2/3$, *and* $K_1$, $K_2$ *are positive constants. Then,*

$$\limsup_{n\to\infty} \frac{\mathcal{L}_{\text{fes/ih}}(x^n) - \bar{\mathcal{L}}_{\min}(x^n)}{n} \leq 0 \quad a.s.$$

*Proof:* Let $J(n)$ denote the index of the super-segment in which a time instant $n$ falls, and let $m(n)$ denote the total number of blocks up to and including the block in which $x_n$ is located. Recalling the definition of $T_j$ as the cumulative length of the first $j$ super-segments, we have

$$m(n) = \sum_{j=0}^{J(n)-1} \left\lceil \frac{N_j}{M_j} \right\rceil + \left\lceil \frac{n - T_{J(n)}}{M_{J(n)}} \right\rceil .$$

Let $F_0, F_1, \ldots, F_{m(n)-1}$ denote the sequence of expert selections in the specific realization leading to the loss $\mathcal{L}_{\text{fes/ih}}(x^n)$. Notice that these selections are independent (given $x^n$), and so are the corresponding random losses $\bar{\mathcal{L}}_{F_k}(X_k)$ accumulated by $F_k$ over block $X_k$, $k = 0, 1, \ldots, m(n)-1$. For a sequence of zero-mean independent random variables $Z_k$ taking values in bounded intervals of sizes $R_k$, respectively, $k = 1, 2, \ldots, m$, the Chernoff bounding technique used in conjunction with Hoeffding's inequality (see (22)) yields, for every $\epsilon > 0$,

$$\Pr\{\sum_{k=1}^{m} Z_k \geq m\epsilon\} \leq \exp\{-\frac{2m^2\epsilon^2}{\sum_{k=1}^{m} R_k^2}\} . \tag{33}$$

Thus, by letting $m = m(n)$, $Z_k = \bar{\mathcal{L}}_{F_k}(X_k) - E_F\{\bar{\mathcal{L}}_{F_k}(X_k)|\mathbf{L}_{\mathcal{F},k}\}$, and $R_k = \ell_{\max} M_{j(k)}$, $k = 0, 1, \ldots, m(n)-1$, where $j(k)$ denotes the index of the super-segment to which block $k$ belongs, we can apply the bound in (33) to obtain

$$\Pr\left\{ \sum_{k=0}^{m(n)-1} [\bar{\mathcal{L}}_{F_k}(X_k) - E_F\{\bar{\mathcal{L}}_{F_k}(X_k)|\mathbf{L}_{\mathcal{F},k}\}] \geq m(n)\epsilon \right\} \leq \exp\{-\frac{2m(n)^2\epsilon^2}{\ell_{\max}^2 \sum_{k=1}^{m(n)} M_{j(k)}^2}\} . \tag{34}$$

Now, proceeding as in the proof of Theorem 2, but with the design parameters $\eta_j$ and $M_j$ specified for Theorem 3, we have

$$\sum_{k=0}^{m(n)-1} E_F\{\bar{\mathcal{L}}_{F_k}(X_k)|\mathbf{L}_{\mathcal{F},k}\} \leq \bar{\mathcal{L}}_{\min}(x^n) + O(n^{\varphi_2+1/3}) + O(n^{1-\varphi_1-\varphi_2}) . \tag{35}$$

In addition,

$$\mathcal{L}_{\text{fes/ih}}(x^n) \leq \sum_{k=0}^{m(n)-1} \bar{\mathcal{L}}_{F_k}(X_k) + \ell_{\max} m(n) . \tag{36}$$

It is easy to see that $m(n) = \Theta(n^{\varphi_1+2/3})$ and $\sum_{k=0}^{m(n)-1} M^2_{j(k)} = O(n^{4/3-\varphi_1})$, so that (34), (35), and (36) imply

$$\Pr\left\{\frac{\mathcal{L}_{\mathrm{fes/ih}}(x^n) - \bar{\mathcal{L}}_{\min}(x^n)}{n} \geq \delta(n)\right\} \leq \exp\{-K_3 n^{3\varphi_1}\} \tag{37}$$

for some positive constant $K_3$, where $\delta(n) = O(n^{\varphi_2-2/3}) + O(n^{-\varphi_1-\varphi_2}) + O(n^{\varphi_1-1/3})$. Since $\delta(n) \to 0$ and the right-hand side of (37) forms a summable series, the result follows by the Borel-Cantelli Lemma. □

For a given $\varphi_1$, which determines the decay rate at the right-hand side of (37), the order of $\delta(n)$ at the left-hand side is minimized by choosing $\varphi_2$ such that $\max\{0, \frac{1}{3}-2\varphi_1\} < \varphi_2 \leq \frac{1}{3}+\varphi_1$. In that case we have $\delta(n) = O(n^{\varphi_1-1/3})$.

**Discussion.** An example of the finite expert set considered in this section is the set of deterministic FS strategies of the form $b_t = g(s_t)$, as studied for memoryless loss functions, for *finite* action space $B$ and state space $S$. In this case, $\log\lambda \leq |S|\log|B|$. The MDP probabilistic analogy, in which the data are drawn from an FS source $\{p(x|s),\ x \in A,\ s \in S\}$ and the expected loss takes the form (8), suggests an alternative deterministic FS model. Assuming that $\{p(x|s)\}$ yields an irreducible Markov chain, the expected loss in (8) is minimized over the strategies $\mu$ by the *deterministic* FS strategy

$$b_t = \mu_t(x^{t-1}, b^{t-1}) = g(s_t, b_{t-1})$$

independent of $s_1$ and $b_0$ [16, Vol. 2, Ch. 4]. This strategy is termed *stationary*, and is obtained by solving the following linear program:

$$\begin{aligned}
\text{minimize:} \quad & \sum_{s,b',b} q(s,b',b)L(s,b',b) \text{ over } q \\
\text{subject to:} \quad & \sum_{b} q(s,b',b) = \sum_{s',b''} q(s',b'',b')p(s|s')\,, \\
& \sum_{s,b',b} q(s,b',b) = 1\,, \text{ and} \\
& q(s,b',b) \geq 0
\end{aligned} \tag{38}$$

where $p(s|s')$ is given by (10). At most one action $b$ satisfies $q(s,b',b) \neq 0$, and we set $g(s,b') = b$. Thus, the MDP setting suggests the use of a *finite* set of *deterministic* reference FS off-line strategies of the form $b_t = g(s_t, b_{t-1})$, that are obtained as optimal stationary strategies for some FS sources $\{p(x|s)\}$. The resulting number of experts satisfies $\log\lambda \leq |S|\cdot|B|\log|B|$.

In the case $|S| = 1$, the same model naturally arises without recourse to a probabilistic setting. In that case, consider randomized reference strategies, not necessarily FS, in which $b_t$ is any *stationary* random process, with $b_0$ randomly selected according to the stationary marginal distribution (rather than being a fixed initial action). Let $q(b_{t-1} = b', b_t = b)$, $b', b \in B$, denote the marginal distribution on pairs. Then, the average loss (with respect to the ensemble of randomly selected actions) takes the form

$$\mathcal{L}_q(x^n) = \sum_{b,b' \in B} q(b', b) \sum_{x \in A} np(x)\ell(b', b, x) \triangleq n \sum_{b,b' \in B} q(b', b) L(b', b) \tag{39}$$

where $p(x)$ denotes the empirical probability of $x \in A$ for the sequence $x^n$. Thus, the performance is controlled by the transition probabilities from one action to the next induced by the marginals, namely $q(b|b') = q(b', b)/\sum_b q(b', b)$, which are subject to off-line optimization. Here, as in the memoryless loss case, the optimal reference strategy depends on $x^n$ through its empirical distribution $p(x)$. The linear program defined by the minimization of the right-hand side of (39) over $q$, is a particular case of (38) (obtained in the MDP case), and admits a deterministic solution $b_t = g(b_{t-1})$. Clearly, each possible solution corresponds to a simple cycle in the fully connected directed graph whose nodes represent the elements of $B$, with $b_0$ randomly selected among the actions in the cycle with a uniform distribution. If we associate a weight $L(b, b')$ to an edge from action $b$ to action $b'$, then the best strategy corresponds to the simple cycle with smallest average weight among its edges, which can be found with a Viterbi-like procedure. Due to the random choice of $b_0$, these experts are not deterministic. However, if we now allow optimization of the choice of $b_0$ within the actions in the cycle, we obtain a deterministic expert with a smaller or equal loss. Clearly, in this case,

$$\log \lambda \leq \log \sum_{i=1}^{|B|} \binom{|B|}{i} \cdot i! \leq |B| \log |B| \, .$$

The above graph theoretic considerations can be generalized to the MDP case in which $|S| \geq 1$, provided that the state transitions are independent of the data, i.e., $f(s, x) = f(s)$ for all $x \in A$ (e.g., in the case $|S| = 1$). The off-line strategies can be found by identifying simple cycles with minimum average weight in a graph with nodes in $S \times B$. The edges in the graph connect pairs $(s, b')$ to corresponding pairs $(f(s), b)$, with associated weights $L(s, b', b)$. An edge from $(s, b')$ to $(f(s), b)$ in the selected simple cycle indicates that $b = g(s, b')$.

For given $S$ and $f$, the reference set $b_t = g(s_t, b_{t-1})$ is richer than the traditional reference set $b_t = g(s_t)$ used for memoryless loss functions. However, it is in turn a subset of the traditional model, if the latter is defined for a state space $S \times B$, and an appropriate next-state function. This fact follows from observing that any strategy of the form $b_t = g(s_t, b_{t-1})$ can be represented as $b_t = g(\sigma_t)$, $\sigma_t \in S \times B$, by re-defining the state transitions as $\sigma_{t+1} = f'(\sigma_t, x_t)$, where for $\sigma = (s, b) \in S \times B$, $f'(\sigma, x) = (f(s, x), g(s, b))$. Notice that the next-state function $f'$ also depends on $g$. In particular, if $g$ is randomized, so is $f'$.

## 3 FSM reference models for loss functions with memory

We will now impose an FSM structure on the expert set, for a finite action space $B$. As discussed in Section 1, for memoryless loss functions the analysis can be confined to deterministic reference strategies. In this section, we show that key properties that hold for FSM expert models in that case, do not carry over to the case of loss functions with memory, which complicates the problem. In particular, it is evidenced that the deterministic model is too narrow for this case, in the sense of offering a less ambitious target performance than randomized FSM expert models. As a result, in Section 4 we analyze on-line strategies for a randomized FSM expert model.

We start by showing that, contrary to the memoryless loss case, the minimizing function $g$ in the loss expression (11) for an FSM model $b_t = g(s_t)$ (and a given next-state function) need not be deterministic. Intuitively, the reason is that with a deterministic $g$, the mapping remains the same for the first two arguments of $\ell(g(s'), g(s), x)$ in (11), whereas randomization allows for selection of two "different" mappings, thus accounting for a convex combination of terms not all of which are covered by the deterministic mappings. In contrast, in the memoryless loss case, taking expectation over a randomized mapping $g$ would generate a mixture only of terms that can be implemented deterministically, one of which is the best. The following example demonstrates a randomized optimal mapping $g$.

**Example 1.** Consider a binary prediction problem in which

$$\ell(b_{t-1}, b_t, x_t) = \begin{cases} 0 & \text{if } x_t = b_t \\ \frac{1}{2} & \text{if } x_t = b_{t-1} \text{ and } x_t \neq b_t \\ 1 & \text{otherwise.} \end{cases}$$

Let the underlying FSM for the set of experts be Markov of order 1, namely, $s_{t+1} = x_t$, and for a (possibly randomized) strategy $g(s)$ denote $q_s = \Pr\{g(s) = 1\}$, $s \in \{0, 1\}$. Straightforward computations show that for a sequence $x^n$, $n \to \infty$ (so that we can neglect edge effects), the loss in (11) takes the form

$$\mathcal{L}_g(x^n) = n(1) + \frac{n}{2}\left[(\rho - \nu - \varphi)q_0 + (3\rho + 3\nu + \varphi - 3)q_1 + \rho q_0^2 + \nu q_0 q_1 + (1 - \rho - \nu)q_1^2\right] \quad (40)$$

where $\rho \triangleq n(00)/n$, $\nu \triangleq 2n(01)/n$, $\varphi \triangleq [n(001) + n(011)]/n$, and $n(y_1 y_2 \cdots y_m)$ denotes the number of occurrences of a string $y_1 y_2 \cdots y_m$ in $x^n$. Now, for a constant $p$, $1/4 < p < 3/4$, there exists a non-empty Markov type such that for any sequence $x^n$ in the type, $n(yz0) = p\, n(yz) + O(1/n)$, $y, z \in \{0, 1\}$ (see [22]). Neglecting the $O(1/n)$ term (which accounts for edge effects and non-integer values), we have $\rho = p^2$, $\nu = 2p(1 - p)$, and $\varphi = p(1 - p)$, so that, after normalization, the loss in (40) is given by

$$\frac{\mathcal{L}_g(x^n)}{n} = 1 - p + \frac{[pq_0 + (1 - p)q_1]^2}{2} + (4p - 3)\ \frac{pq_0 + (1 - p)q_1}{2}$$

which achieves its minimum value $1-p-(3-4p)^2/8$ for any strategy such that $pq_0+(1-p)q_1 = (3-4p)/2$. The four possible deterministic strategies, in turn, yield losses $1-p$, $1-5p(1-p)/2$, $3p(1-p)/2$, and $p$, respectively. Clearly, a randomized strategy prevails for any $p \neq \frac{1}{2}$.

Notice that in the above example, the loss of the best single-state deterministic machine is $n \min\{p, 1 - p\}$, which for $p$ in the range $(1/3, 2/3)$ is larger than the loss $3np(1 - p)/2$ that can be obtained with the best deterministic machine with a Markov order of 1. Thus, it may be advantageous to increase the number of states even if, as in this case, the state refinement process does not provide additional information in the sense of further refining the conditional empirical distributions. In contrast, for memoryless loss functions, state refinements may translate into better performance only if they result in a better "discrimination" of the conditional empirical probabilities, as the optimal $g(s)$, $s \in S$, depends solely on the occurrence counts given $s$. The advantage of state refinement even without a change in the empirical conditional probabilities, is further illustrated by the following example.

**Example 2.** Consider the "counting" sequence [3] that lists, for example, in lexicographical order, all the binary words of length $k$, $k = 1, 2, \cdots$. For this sequence, the empirical conditional probability of 0 given any past string tends to $\frac{1}{2}$ as $n$ grows, regardless of the length of the conditioning string. Clearly, the (normalized) loss of a deterministic single-state machine (with the loss function of Example 1) is $\frac{3}{8}$. However, it can be shown that a Markov machine of order $k$ can achieve the normalized loss $\frac{1}{4} + \delta(k)$, where $\delta(k) \to 0$ as $k \to \infty$. This loss can be achieved by stepping through a Hamiltonian circuit in the de Brujin graph of order $k$ (see [23]), alternatively assigning actions $g(s) = 0$ and $g(s) = 1$ as all the states $s \in S$ are visited. Notice that a normalized loss of $\frac{1}{4}$ can be achieved with a (non-Markovian) two-state periodic FSM that generates actions $b_{2t} = 0$ and $b_{2t+1} = 1$, $t = 0, 1, \cdots$.

As the Markov order $k$ grows, the performance in the counting sequence example approaches the best achievable performance for any FSM. This property conforms to the case of memoryless loss functions [5]. However, this is not always the case for loss functions with memory, as demonstrated by the following example.

**Example 3.** Consider the loss function $\ell(b', b, x) = 1$ if $b = b'$, and $\ell(b', b, x) = 0$ otherwise (here, the loss does not depend on $x \in A$). For the all-zero sequence $x^n = 000\cdots$, the state sequence is constant in a Markov model, regardless of the Markov order. Therefore, a Markov strategy will be constant, with a normalized loss of 1. In contrast, the two-state FSM strategy that generates the action sequence $b^n = 010101\cdots$ has a zero loss.

The above examples show that while for the expert model $b_t = g(s_t)$ the loss in (11) depends on the data only through its type with respect to an FSM with state $(s', s) \in S^2$, the model does not inherit some of the properties of the memoryless loss case. Furthermore, the following example provides evidence to the need to deal with randomized experts also if, as suggested by the MDP analogy, we let the expert's advice $b_t$ depend also on the previous advice $b_{t-1}$. Notice that in the case $b_t = g(s_t, b_{t-1})$, the type of $x^n$ will not in general determine the loss in (6).[2]

**Example 4.** Let $x^n$ $(n \to \infty)$ be a binary periodic sequence with period $x^{18}$ =

[2]An exception is given by the case $|S| = 1$, with the initial action $b_0$ chosen according to the steady-state distribution of the first-order Markov chain $\{b_t\}$ (see Equation (39)).

101010000000111111, let $B = \{0,1\}$, and consider the loss function

$$\ell(b_{t-1}, b_t, x_t) = \begin{cases} 0 & \text{if } x_t = b_t \cdot b_{t-1} \\ 1 & \text{otherwise.} \end{cases}$$

Let the underlying FSM for the set of experts be Markov of order 1, so that $s_{t+1} = x_t$, $t > 0$, and assume $s_1 = 1$. Notice that each period in $x^n$ is composed of (three) pieces with very different Markov types. While it appears that no deterministic machine can perform well on all the pieces, the idea is that a randomized machine may be more able to compromise. An exhaustive search shows that the best deterministic expert $b_t = g^{(d)}(s_t, b_{t-1})$ is given by $b_0 = 1$, and $b_t = s_t$, $t > 0$ (independent of $b_{t-1}$), with a loss $\ell_{g^{(d)}} = 5$ per period over any integer number of periods (thus, the normalized loss tends to $5/18$ as $n \to \infty$). It is easy to verify that a randomized expert $g^{(r)}(s_t, b_{t-1})$ that chooses $b_t = 1$ with probability $q$ in case $s_t = b_{t-1} = 0$ and otherwise behaves as the above deterministic expert, achieves an expected loss per period (over any integer number of periods)

$$\bar{\ell}_{g^{(r)}} = 5 - \frac{q - q^7}{1 + q}\,.$$

Since $\bar{\ell}_{g^{(r)}} < 5$ for $0 < q < 1$, it follows that the best expert of the form $b_t = g(s_t, b_{t-1})$ for $x^n$ is randomized.

Instead of considering randomized experts $b_t = g(s_t, b_{t-1})$, we can "hide" the dependency on $b_{t-1}$ through a transformation of the state space, as discussed at the end of Section 2. Notice that the resulting next-state function is randomized if $g$ is randomized. A randomized next-state function introduces statistical dependencies between the expert's actions, without an explicit dependency of $g$ on $b_{t-1}$. Moreover, it results in a richer expert model.

We conclude that the most appropriate FSM expert model takes the form $b_t = g(s_t)$, where the next-state function is randomized, with transition probabilities $p_f(s_{t+1}|s_t, x_t)$. We can assume $g$ to be deterministic, since for a randomized rule based on a conditional distribution $q_g(b_t|s_t)$, an equivalent expert with at most $|S| \times |B|$ states and for which $g$ is deterministic, can be built as follows: Split each state into $|B|$ states, and for an observation $x \in A$, transition from the composite state $(s', b')$ to $(s, b)$ with probability $p_f(s|s', x)q_g(b|s)$; then, at state $(s, b)$, choose action $b$ deterministically.

It is interesting to notice that, while randomization of the next-state function results in a richer expert model, for a given mapping $g$ this model can be viewed as one in which $S$ is a set of *actions* and the FSM (driven by $x^n$) is *Markov of order* 1. The random action $s_t$ is selected (conditioned on the previous action and the state) according to $p_f(s_t|s_{t-1}, x_{t-1})$, and incurs a loss $\ell_g(s_{t-1}, s_t, x_t) \stackrel{\Delta}{=} \ell(g(s_{t-1}), g(s_t), x_t)$. Thus, under this interpretation, the next-state function is deterministic and the (randomized) choice of action depends on the previous action.

## 4 On-line strategies for randomized FSM experts

In this section, for a given state space $S$ with cardinality $\gamma$, we consider on-line strategies that perform essentially as well as the best randomized FSM expert, which is given by a mapping $b_t = g(s_t)$, a set of transition probabilities $p_f(s_{t+1}|s_t, x_t)$, and an initial state $s_1$, for any individual sequence $x^n$. The observation space $A$ and the action space $B$ are assumed finite, with respective cardinalities $\alpha$ and $\beta$. Again, the result is first presented for a given finite horizon $n$. As in Section 2, a block-length parameter will determine how long the advice of an expert is followed.

Since the set $\mathcal{F}$ of experts is infinite, the idea is to construct a finite grid in the space of probability matrices $\{p_f(s|s', x),\ s', s \in S,\ x \in A\}$ in order to approximate the best performing expert on $x^n$ by a point in the grid (together with a deterministic mapping $g$ and an initial state). By (6), the expert loss $\bar{\mathcal{L}}_{p,g}(x^n)$ is given by

$$\bar{\mathcal{L}}_{p,g}(x^n) = \ell(b_0, g(s_1), x_1) + \sum_{t=2}^{n} \sum_{s' \in S} \Pr\{s_{t-1} = s'|x^{t-2}\} \sum_{s \in S} p_f(s|s', x_{t-1}) \ell(g(s'), g(s), x_t) \quad (41)$$

where $\Pr\{s_1 = s'|x^0\} \stackrel{\Delta}{=} 1$ when $s'$ is a given initial state, and 0 otherwise. Thus, the main difficulty is in bounding the error accumulated in the computation of $\Pr\{s_{t-1} = s'|x^{t-2}\}$ as a result of the approximation. This probability is obtained from the product of transition probability matrices $Q(x_1) \cdot Q(x_2) \cdot \ldots \cdot Q(x_{t-2})$, $Q(x)$ being the stochastic matrix $\{p_f(s|s', x),\ s, s' \in S,\ x \in A\}$, where the optimum $Q$ in each factor of this product is approximated by a grid point, and so, the overall approximation error grows with $t$. Since the regret of an on-line strategy that combines expert predictions grows with the number of experts, we are faced with two conflicting goals which need to be compromised. On the one hand, a dense grid would provide a good approximation of the best randomized strategy in the continuum; however, on the other hand, the grid cannot be too dense as it would otherwise include a larger number of experts with the risk of non-vanishing normalized regret. Nonetheless, thanks to the fact that the regret depends on the number of experts logarithmically, it turns out that a polynomially growing grid satisfies both requirements.

Specifically, consider a grid $G_{\epsilon_n}$ in which all the transition probabilities $p_f(s|s', x)$ are integer multiples of $\epsilon_n$, for a vanishing parameter $\epsilon_n$ to be specified later. Let $\mathcal{F}_n$ denote the resulting set of experts, where an expert $F \in \mathcal{F}_n$ is given by a point in the grid, an initial state $s_1^{(F)}$, and a mapping $g$ (a fixed initial action $b_0$ is chosen for all experts, since its effect does not propagate). Recalling $\alpha = |A|$, $\gamma = |S|$, and $\beta = |B|$, the cardinality $\lambda_n$ of $\mathcal{F}_n$ satisfies

$$\lambda_n \leq \gamma \cdot \beta^\gamma \cdot \left(\frac{1}{\epsilon_n}\right)^{\gamma^2\alpha} \triangleq \lambda_n^* . \tag{42}$$

For a fixed block length $M$, let a parsing of $x^n$ into $K+1$ non-overlapping blocks be defined as in Theorem 1. At $t = kM+1$, $k = 0, 1, \ldots, K$, the on-line algorithm randomly selects $F \in \mathcal{F}_n$ according to the conditional distribution (13), where $\mathcal{F}$ is replaced by (the finite set) $\mathcal{F}_n$, the cumulative expected loss $\bar{\mathcal{L}}_{F,k}$ of expert $F$ over $k$ blocks is computed as in (41), and $\eta$ is a given positive constant. The initial state $s_{Mk+1}$ of that block is selected according to the marginal distribution on $s_{Mk+1}$ that would have resulted from applying $F$ on $x^{Mk}$, independent of previous state occurrences. Thus, the corresponding distribution on $b_{Mk+1}$ differs from the one prescribed by Theorem 1, as it follows the marginal, rather than the conditional distribution. This choice guarantees statistical independence between actions corresponding to different blocks, a property that will be used later to show also almost-sure convergence. The marginal is computed and maintained recursively, starting from the initial state $s_1^{(F)}$. The actions of the selected expert are followed through $t = \min(kM+M, n)$ (i.e., during the $(k+1)$-st block). Theorem 4 below gives an upper bound on the normalized expected regret of the proposed scheme.

**Theorem 4** *Let $\epsilon_n = 1/(\lceil n^{1+d}\rceil \gamma^2)$, $d > 0$, and let $M^*$ and $\eta^*$ be specified as in equations (14) and (15), respectively, but with $\lambda$ replaced by $\lambda_n^*$, given by (42). Let the expected loss of the proposed on-line strategy with parameters $\eta = \eta^*$ and $M = \lceil M^* \rceil$, over $x^n$, be denoted by*

$\bar{\mathcal{L}}_{\text{rfs/fh}}(x^n)$ *(randomized finite-state expert/finite horizon). Let* $\bar{\mathcal{L}}^{\gamma}_{\min}(x^n)$ *denote the expected loss of the (randomized) expert in* $\mathcal{F}$ *that performs best on* $x^n$*. Then,*

$$\frac{1}{n}\left(\bar{\mathcal{L}}_{\text{rfs/fh}}(x^n) - \bar{\mathcal{L}}^{\gamma}_{\min}(x^n)\right) \leq \frac{3}{2}\,\ell_{\max}\left(\frac{(1+d)\gamma^2\alpha\ln n}{n}\right)^{\frac{1}{3}} + O(n^{-\frac{1}{3}}) + O(n^{-d})\,. \tag{43}$$

Theorem 4 implies that for $d > 1/3$, the normalized regret is $O\left([(\ln n)/n]^{1/3}\right)$.

*Proof:* Let $\{p^*_f(s|s',x),\ s',s \in S,\ x \in A\}$ denote the transition probabilities of the expert $F \in \mathcal{F}$ that performs best on $x^n$. It can readily be verified that there exists a set of transition probabilities $\{\hat{p}_f(s|s',x),\ s',s \in S,\ x \in A\}$ in $G_{\epsilon_n}$ such that

$$|\hat{p}_f(s|s',x) - p^*_f(s|s',x)| \leq \frac{\gamma-1}{\gamma}\cdot\epsilon_n \tag{44}$$

for all $s',s \in S$, $x \in A$. To bound the error accumulated in the computation of $\Pr\{s_{t-1} = s'|x^{t-2}\}$ in (41), $t > 2$, as a result of approximating $\{p^*_f(s|s',x)\}$ with the grid point $\{\hat{p}_f(s|s',x)\}$, we write the probabilities corresponding to each set of distributions as

$$\Pr{}^*\{s_{t-1} = s'|x^{t-2}\} = \mathbf{1}^T_{s_1^{(F)}}\cdot Q^*(x_1)\cdot Q^*(x_2)\cdot\ldots\cdot Q^*(x_{t-2})\cdot\mathbf{1}_{s'}$$

and

$$\hat{\Pr}\{s_{t-1} = s'|x^{t-2}\} = \mathbf{1}^T_{s_1^{(F)}}\cdot\hat{Q}(x_1)\cdot\hat{Q}(x_2)\cdot\ldots\cdot\hat{Q}(x_{t-2})\cdot\mathbf{1}_{s'} \tag{45}$$

respectively, where $\mathbf{1}_s$ denotes a (column) $\gamma$-vector with a single non-zero entry, which corresponds to $s \in S$ and equals 1, and $Q^*(x_i)$ and $\hat{Q}(x_i)$, $x_i \in A$, $1 \leq i \leq t-2$, denote, respectively, the stochastic matrices $\{p^*_f(s|s',x_i),\ s,s' \in S\}$ and $\{\hat{p}_f(s|s',x_i),\ s,s' \in S\}$. Equation (45) can be written as

$$\hat{\Pr}\{s_{t-1} = s'|x^{t-2}\} = \mathbf{1}^T_{s_1^{(F)}}\cdot[Q^*(x_1)+\Delta(x_1)]\cdot[Q^*(x_2)+\Delta(x_2)]\cdot\ldots\cdot[Q^*(x_{t-2})+\Delta(x_{t-2})]\cdot\mathbf{1}_{s'}$$

where $\Delta(x_i)$, $1{\leq}i{\leq}t{-}2$, denote error matrices whose entries, by the approximation bound (44), are all upper-bounded by $(\gamma-1)\epsilon_n/\gamma$ in absolute value. Thus,

$$\hat{\Pr}\{s_{t-1} = s'|x^{t-2}\} = \Pr{}^*\{s_{t-1} = s'|x^{t-2}\} + \sum_{j=1}^{2^{t-2}-1}\mathbf{1}^T_{s_1^{(F)}}\cdot\boldsymbol{Z}_j\cdot\mathbf{1}_{s'} \tag{46}$$

where $\boldsymbol{Z}_j$, $1 \leq j < 2^{t-2}$, denote products of $t-2$ matrices each of which is either $Q^*(x_i)$ or $\Delta(x_i)$ for some $i$, $1 \leq i \leq t-2$, and at least one of the factors is $\Delta(x_i)$. For a matrix $\boldsymbol{Z}_j$ formed by $\kappa$ factors of type $\Delta(x_i)$ $(1 \leq \kappa \leq t-2)$, we have

$$\boldsymbol{Z}_j = R_1\cdot\Delta(x_{i_1})\cdot R_2\cdot\Delta(x_{i_2})\ldots R_\kappa\cdot\Delta(x_{i_\kappa})\cdot R_{\kappa+1}$$

for some indexes $0 \triangleq i_0 < i_1 < i_2 < \ldots < i_\kappa < i_{\kappa+1} \triangleq t-1$, where $R_h$, $1 \le h \le \kappa+1$, are *stochastic matrices* of the form

$$R_h \triangleq Q^*(x_{i_{h-1}+1}) \cdot \ldots \cdot Q^*(x_{i_h-1})$$

provided $i_{h-1}+1 \ne i_h$, and $R_h = I$ (the $\gamma \times \gamma$ identity matrix) otherwise.

Now, since $R_h$ is stochastic, every entry in $R_h \cdot \Delta(x_{i_h})$ is upper-bounded by $(\gamma-1)\epsilon_n/\gamma$ in absolute value. Therefore, the absolute value of every entry in $\prod_h R_h \cdot \Delta(x_{i_h})$ is upper-bounded by $[(\gamma-1)\epsilon_n]^\kappa/\gamma$, and after further post-multiplication by $R_{\kappa+1}$, the absolute value of every entry of $\boldsymbol{Z}_j$ is upper-bounded by $[(\gamma-1)\epsilon_n]^\kappa$. Thus, we can upper-bound the right-hand side of (46) to obtain

$$\hat{\Pr}\{s_{t-1}=s'|x^{t-2}\} \le \Pr{}^*\{s_{t-1}=s'|x^{t-2}\} + \sum_{\kappa=1}^{t-2} \binom{t-2}{\kappa} \cdot [(\gamma-1)\epsilon_n]^\kappa = \Pr{}^*\{s_{t-1}=s'|x^{t-2}\} + \theta_t \tag{47}$$

where $\theta_t = (1+(\gamma-1)\epsilon_n)^{t-2} - 1$. It then follows from (41) that, for any initial state $s_1^{(F)}$ and mapping $g$,

$$\begin{aligned}
\bar{\mathcal{L}}_{\hat{p},g}(x^n) &\le \ell(b_0, g(s_1^{(F)}), x_1) \\
&+ \sum_{t=2}^{n} \sum_{s' \in S} [\Pr{}^*\{s_{t-1}=s'|x^{t-2}\} + \theta_t] \sum_{s \in S} [p_f^*(s|s', x_{t-1}) + \frac{\gamma-1}{\gamma}\epsilon_n] \ell(g(s'), g(s), x_t) \\
&\le \bar{\mathcal{L}}_{p^*,g}(x^n) + \ell_{\max}\gamma \left( n\epsilon_n + (1+\gamma\epsilon_n)\sum_{t=2}^{n} \theta_t \right) \\
&\le \bar{\mathcal{L}}_{p^*,g}(x^n) + n\ell_{\max}\gamma[\epsilon_n + (1+\gamma\epsilon_n)\theta_n] \triangleq \bar{\mathcal{L}}_{p^*,g}(x^n) + n\delta_n
\end{aligned} \tag{48}$$

where the last inequality follows from $\theta_t$ being an increasing function of $t$. Thus, the loss achieved by any expert in $\mathcal{F}$ is approximated by an expert in $\mathcal{F}_n$ up to an excess (normalized) loss $\delta_n = O(n\gamma^2\epsilon_n)$. In particular, the loss $\bar{\mathcal{L}}_{\text{grid}}(x^n)$ of the expert in $\mathcal{F}_n$ that performs best on $x^n$ satisfies

$$\bar{\mathcal{L}}_{\text{grid}}(x^n) \le \bar{\mathcal{L}}^{\gamma}_{\min}(x^n) + n\delta_n \tag{49}$$

where $\delta_n = O(n^{-d})$ for $\epsilon_n = 1/(\lceil n^{1+d} \rceil \gamma^2)$.

Next, we use Theorem 1 to upper-bound the regret $\bar{\mathcal{L}}_{\text{rfs/fh}}(x^n) - \bar{\mathcal{L}}^{\gamma}_{\text{grid}}(x^n)$ of the proposed on-line strategy with respect to the finite set of experts $\mathcal{F}_n$. The choice of the initial state $s_{Mk+1}$ of block $k$ according to the marginal (rather than conditional) distribution corresponding to the evolution of the selected expert $F$ along $x^{Mk}$, prevents from direct application of the theorem. However, the use of the correct marginal implies that the joint distributions of pairs of consecutive actions, which determine the expected loss, will not differ beyond the first sample in the block from those dictated by $F$. Now, since $b_{Mk}$ corresponds to a possibly different expert selection, this discrepancy was already accounted for in the proof of Theorem 1. Therefore, we can still use the theorem to upper-bound the regret. Finally, since

$$\bar{\mathcal{L}}_{\mathrm{rfs/fh}}(x^n) - \bar{\mathcal{L}}^{\gamma}_{\min}(x^n) = (\bar{\mathcal{L}}_{\mathrm{rfs/fh}}(x^n) - \bar{\mathcal{L}}_{\mathrm{grid}}(x^n)) + (\bar{\mathcal{L}}_{\mathrm{grid}}(x^n) - \bar{\mathcal{L}}^{\gamma}_{\min}(x^n))$$

the result follows from (49), Theorem 1, and (42). □

For infinite horizon, again, time is divided into non-overlapping contiguous super-segments of lengths $N_j = \lceil n_0 a^j \rceil$, $j = 0, 1, \cdots$, for given positive constants $n_0$ and $a$, $a > 1$. In the $(j+1)$-st super-segment, the above finite horizon algorithm is used with an expert set $\mathcal{F}_{N_j}$ (given by a grid $G_{\epsilon_{N_j}}$, with $\epsilon_n$ as in Theorem 4), and design parameters $M = \lceil M_j \rceil$ and $\eta = \eta_j$, where $M_j$ and $\eta_j$ differ from $M^*$ and $\eta^*$, respectively, in that $N_j$ replaces $n$ in their definition. The expected loss of this infinite horizon scheme over a prefix $x^n$ of a sequence of observations is denoted $\bar{\mathcal{L}}_{\mathrm{rfs/ih}}(x^n)$ (randomized finite-state expert/infinite horizon). Theorem 5 below states that for *all* $n$, the normalized regret is essentially bounded as in Theorem 4, but with a larger constant. The proof is omitted, since it is analogous to that of Theorem 2.

**Theorem 5** *For any prefix $x^n$ of an infinite sequence of observations,*

$$\frac{1}{n}\left(\bar{\mathcal{L}}_{\mathrm{rfs/ih}}(x^n) - \bar{\mathcal{L}}^{\gamma}_{\min}(x^n)\right) \le \frac{3}{2}\, C(a)\ell_{\max}\left(\frac{(1+d)\gamma^2\alpha\ln n}{n}\right)^{\frac{1}{3}} + O(n^{-\frac{1}{3}}) + O(n^{-d})\,.$$

As for almost-sure convergence, it can be established with respect to both randomization levels (the choice of experts and the expert actions), as shown in Theorem 6 below. This pure almost-sure property is stronger than the one presented for generic experts in Theorem 3, which was only with respect to the random selection of experts at each block (with expectation still taken for the inner randomization on the expert actions). Let $\mathcal{L}_{\mathrm{rfs/ih}}(x^n)$ denote the (random) loss incurred over $x^n$ by a specific realization of the randomly selected experts and their random actions, according to the infinite horizon scheme, but with $M_j$ and $\eta_j$ as in Theorem 3.

**Theorem 6** *If* $0 < \varphi_1 < 1/3$ *and* $0 < \varphi_2 < 2/3$*, then*

$$\limsup_{n\to\infty} \frac{\mathcal{L}_{\mathrm{rfs/ih}}(x^n) - \bar{\mathcal{L}}^{\gamma}_{\min}(x^n)}{n} \leq 0 \quad a.s.$$

*Proof:* We proceed as in the proof of Theorem 3, except that here the independent random losses of interest are $\mathcal{L}_{\mathrm{rfs/ih}}(X_k)$, $k = 0, 1, \ldots, m(n)-1$, which correspond to *specific realizations* of the on-line actions over block $X_k$. The independence of these losses follows from the choice of initial state for each block. Thus,

$$\Pr\left\{\sum_{k=0}^{m(n)-1} [\mathcal{L}_{\mathrm{rfs/ih}}(X_k) - E_F\{\bar{\mathcal{L}}_{\mathrm{rfs/ih}}(X_k)|\mathbf{L}_{\mathcal{F},k}\}] \geq m(n)\epsilon\right\} \leq \exp\{-\frac{2m(n)^2\epsilon^2}{\ell^2_{\max}\sum_{k=1}^{m(n)} M^2_{j(k)}}\}$$

implying

$$\Pr\left\{\mathcal{L}_{\mathrm{rfs/ih}}(x^n) - \bar{\mathcal{L}}_{\mathrm{rfs/ih}}(x^n) \geq m(n)\epsilon\right\} \leq \exp\{-\frac{2m(n)^2\epsilon^2}{\ell^2_{\max}\sum_{k=1}^{m(n)} M^2_{j(k)}}\}\,. \tag{50}$$

A result analogous to Theorem 5, but for the re-specified design parameters $\eta_j$ and $M_j$, takes the form

$$\bar{\mathcal{L}}_{\mathrm{rfs/ih}}(x^n) \leq \bar{\mathcal{L}}^{\gamma}_{\min}(x^n) + O(n^{\varphi_2+1/3}\ln n) + O(n^{1-\varphi_1-\varphi_2}) + O(n^{-d}) + \ell_{\max}m(n)\,. \tag{51}$$

Since the asymptotic behavior of $m(n)$ and $M_{j(k)}$ is the same as in the proof of Theorem 3, by (50) and (51), the proof is completed similarly. □

**FSM experts with an unbounded number of states.** Next, we consider an infinite sequence $x^\infty$, and analyze an on-line strategy that competes against randomized FSM experts of *any* size ($\gamma \to \infty$). The corresponding reference performance for randomized FSM experts is given by

$$\bar{\ell}^{\infty}_{\min}(x^\infty) = \lim_{\gamma\to\infty}\limsup_{n\to\infty}\frac{\bar{\mathcal{L}}^{\gamma}_{\min}(x^n)}{n}$$

where the limit on $\gamma$ exists since $\bar{\mathcal{L}}^{\gamma}_{\min}(x^n)$ is non-negative and non-increasing on $\gamma$. To achieve a vanishing normalized regret, we use the algorithm of Theorem 5, but with a value $\gamma_j$ of $\gamma$ that gradually increases with the super-segment index $j$. Let $\bar{\mathcal{L}}_{\mathrm{seq}}(x^n)$ denote the resulting loss over $x^n$.

**Corollary 1** *If the size* $\gamma_j$ *of the state space used for super-segment* $j$ *satisfies* $\lim_{j\to\infty}\gamma_j = \infty$ *and* $\gamma_j = o(\sqrt{N_j}/\log^2 N_j)$*, and* $d > 1/3$*, then*

$$\limsup_{n\to\infty}\frac{\bar{\mathcal{L}}_{\mathrm{seq}}(x^n)}{n} \leq \bar{\ell}^{\infty}_{\min}(x^\infty)\,.$$

*Proof:* Let $Y_j$ denote the sequence of observations that fall in the $(j+1)$-st super-segment. By Theorem 4 we have

$$\bar{\mathcal{L}}_{\text{seq}}(Y_j) - \bar{\mathcal{L}}^{\gamma_j}_{\text{min}}(Y_j) = O\left([\gamma_j^2 N_j^2 \log N_j]^{\frac{1}{3}}\right) = O\left(\frac{N_j \epsilon(N_j)}{\log N_j}\right) \tag{52}$$

where $\lim_{j\to\infty} \epsilon(N_j) = 0$. For any positive integer $\gamma$, we have $\gamma_j > \gamma$ for sufficiently large $j$, implying

$$\bar{\mathcal{L}}^{\gamma_j}_{\text{min}}(Y_j) \le \bar{\mathcal{L}}^{\gamma}_{\text{min}}(Y_j)\,.$$

Therefore, by (52), we have

$$\bar{\mathcal{L}}_{\text{seq}}(x^n) - \bar{\mathcal{L}}^{\gamma}_{\text{min}}(x^n) = O\left(\sum_{j=0}^{J(n)} \frac{N_j \epsilon(N_j)}{\log N_j}\right) \tag{53}$$

where $J(n)$ is the index of the super-segment in which $n$ falls. Now, if $\epsilon(N_j) = O((\log N_j)/N_j)$, then the right-hand side of (53) is $O(J(n))$. Otherwise, the right-hand side of (53) is $O(n\epsilon(n)J(n)/\log n)$. In any case, since $J(n) = O(\log n)$ (see proof of Theorem 2), we have

$$\limsup_{n\to\infty} \frac{\bar{\mathcal{L}}_{\text{seq}}(x^n)}{n} \le \limsup_{n\to\infty} \frac{\bar{\mathcal{L}}^{\gamma}_{\text{min}}(x^n)}{n}\,. \tag{54}$$

Since (54) holds for any $\gamma > 0$, the proof is complete. □

**Discussion.** Corollary 1 parallels similar results for memoryless loss functions [5]. In the memoryless loss case, the "ultimate" performance yardstick is restricted to deterministic FSM's, which is justified by the asymptotic equivalence of the deterministic and the randomized classes. For a finite sequence and a given number of states, however, we can still ask which approach is better in terms of total normalized loss:

a. To compete against deterministic machines only, as in Section 2 (where an action may also depend on the previous action), aiming at a deterministic "FS predictability," with a regret corresponding to a reduced number of experts. For memoryless loss functions, the limiting FS predictability (as $|S| \to \infty$) is the same as with randomized machines, but may be achieved at a slower rate.

b. To compete against randomized machines, for which the regret is larger due to a larger number of experts, but the limiting value may be smaller for loss functions with memory.

Since the best approach may depend on the data sequence, it cannot be decided ahead of time. However, we can regard each approach as a randomized "meta-expert" and, again, combine the advice of these two meta-experts as in Section 2. By Theorem 2 (in the infinite horizon case), the main term in the (normalized) excess loss (with respect to the best of the above two strategies) is $1.5C(a)\ell_{\max}((\ln 2)/n)^{1/3}$. Thus, this mixed strategy is equivalent to picking the best approach, but with an additional regret corresponding to only two experts.

## Acknowledgments

Many thanks to Tom Cover for useful discussions in the early stages of this investigation, and to Tsachy Weissman for insightful comments on the manuscript.

## References


[1] J. F. Hannan, "Approximation to Bayes risk in repeated plays," in *Contributions to the Theory of Games, Volume III, Annals of Mathematics Studies*, pp. 97–139, Princeton, NJ, 1957.

[2] N. Merhav and M. Feder, "Universal prediction," *IEEE Trans. Inform. Theory*, vol. IT-44, pp. 2124–2147, Oct. 1998.

[3] J. Ziv and A. Lempel, "Compression of individual sequences via variable-rate coding," *IEEE Trans. Inform. Theory*, vol. IT-24, pp. 530–536, Sept. 1978.

[4] M. Feder, N. Merhav, and M. Gutman, "Universal prediction of individual sequences," *IEEE Trans. Inform. Theory*, vol. IT-38, pp. 1258–1270, July 1992.

[5] N. Merhav and M. Feder, "Universal schemes for sequential decision from individual data sequences," *IEEE Trans. Inform. Theory*, vol. IT-39, pp. 1280–1292, July 1993.

[6] V. G. Vovk, "Aggregating strategies," in *Proc. of the 3rd Annual Workshop on Computational Learning Theory*, (San Mateo, California), pp. 372–383, 1990.

[7] N. Littlestone and M. K. Warmuth, "The weighted majority algorithm," *Information and Computation*, vol. 108, pp. 212–261, 1994.

[8] D. Haussler, J. Kivinen, and M. K. Warmuth, "Sequential prediction of individual sequences under general loss functions," *IEEE Trans. Inform. Theory*, vol. IT-44, pp. 1906–1925, Sept. 1998.

[9] N. Cesa-Bianchi, Y. Freund, D. P. Helmbold, D. Haussler, R. E. Schapire, and M. K. Warmuth, "How to use expert advice," *Journal of the ACM*, vol. 44, no. 3, pp. 427–485, 1997.

[10] A. Blum and A. Kalai, "Universal portfolios with and without transaction costs," *Machine Learning*, vol. 35, no. 3, pp. 193–205, 1999.

[11] G. Iyengar and T. Cover, "Growth optimal investment in horse race markets with costs," *IEEE Trans. Inform. Theory*, vol. IT-46, pp. 2675–2683, Nov. 2000.

[12] A. Borodin, N. Linial, and M. Saks, "An optimal online algorithm for metrical task systems," *Journal of the ACM*, vol. 39, pp. 745–763, 1992.

[13] A. Blum and C. Burch, "On-line learning and the metrical task system problem," *Machine Learning*, vol. 39, no. 1, pp. 35–58, 2000.

[14] J. S. Vitter and P. Krishnan, "Optimal prefetching via data compression," in *Proc. 1991 Conf. Foundation of Computer Sci. (FOCS)*, pp. 121–130, 1991.

[15] M. J. Weinberger and E. Ordentlich, "On-line decision making for a class of loss functions via Lempel-Ziv parsing," in *Proc. 2000 Data Compression Conference*, (Snowbird, Utah, USA), pp. 163–172, Mar. 2000.

[16] D. P. Bertsekas, *Dynamic programming and optimal control.* Belmont, Massachusetts: Athena Scientific, 1995.

[17] N. Shimkin and A. Shwartz, "Guaranteed performance regions in Markovian systems with competing decision makers," *IEEE Trans. Autom. Control*, vol. 38, pp. 84–95, Jan. 1993.

[18] D. Blackwell, "An analog for the minimax theorem for vector payoffs," *Pacific J. Math.*, vol. 6, pp. 1–8, 1956.

[19] M. Herbster and M. K. Warmuth, "Tracking the best expert," *Machine Learning*, vol. 32, no. 2, pp. 286–294, 1998.

[20] N. Cesa-Bianchi and G. Lugosi, "On prediction of individual sequences," *Annals of Statistics*, vol. 27, no. 6, 1999.

[21] W. Hoeffding, "Probability inequalities for sums of bounded random variables," *J. Amer. Statist. Assoc.,* vol. 58, pp. 13–30, 1963.

[22] P. Whittle, "Some distributions and moment formulae for the Markov chain," *J. Roy. Statist. Soc.,* Ser. B, 17, pp. 235–242, 1955.

[23] D. E. Knuth, *The Art of Computer Programming.* Reading, Massachusetts: Addison Wesley Longman, Inc., 1997.