

The Last-Step Minimax Algorithm

Eiji Takimoto^{1*} and Manfred K. Warmuth^{2**}

¹ Graduate School of Information Sciences, Tohoku University
Sendai, 980-8579, Japan.

t2@ecei.tohoku.ac.jp

² Computer Science Department, University of California, Santa Cruz
Santa Cruz, CA 95064, U.S.A.

manfred@cse.ucsc.edu

Abstract. We consider on-line density estimation with a parameterized density from an exponential family. In each trial t the learner predicts a parameter θ_t . Then it receives an instance \mathbf{x}_t chosen by the adversary and incurs loss $-\ln p(\mathbf{x}_t|\theta_t)$ which is the negative log-likelihood of \mathbf{x}_t w.r.t. the predicted density of the learner. The performance of the learner is measured by the regret defined as the total loss of the learner minus the total loss of the best parameter chosen off-line. We develop an algorithm called the Last-step Minimax Algorithm that predicts with the minimax optimal parameter assuming that the current trial is the last one. For one-dimensional exponential families, we give an explicit form of the prediction of the Last-step Minimax Algorithm and show that its regret is $O(\ln T)$, where T is the number of trials. In particular, for Bernoulli density estimation the Last-step Minimax Algorithm is slightly better than the standard Krichevsky-Trofimov probability estimator.

1 Introduction

Consider the following repeated game based on density estimation with a family of probability mass functions $\{p(\cdot|\theta) \mid \theta \in \Theta\}$, where Θ denotes the parameter space. The learner plays against an adversary. In each trial t the learner produces a parameter θ_t . Then the adversary provides an instance \mathbf{x}_t and the loss of the learner is $L(\mathbf{x}_t, \theta_t) := -\ln p(\mathbf{x}_t|\theta_t)$. Consider the following *regret* or *relative loss*

$$\sum_{t=1}^T L(\mathbf{x}_t, \theta_t) - \inf_{\theta_B \in \Theta} \sum_{t=1}^T L(\mathbf{x}_t, \theta_B).$$

This is the total on-line loss of the learner minus the total loss of the best parameter chosen off-line based on all T instances. The goal of the learner is to minimize the regret while the goal of the adversary is to maximize it. To get a finite regret we frequently need to restrict the adversary to choose instances from a bounded space (Otherwise the adversary could make the regret unbounded in

* This work was done while the author visited University of California, Santa Cruz.

** Supported by NSF grant CCR-9821087

just one trial). So we let \mathcal{X}_0 be the instance space from which instances are chosen. Thus the game is specified by a parametric density and the pair (Θ, \mathcal{X}_0) .

If the horizon T is fixed and known in advance, then we can use the optimal minimax algorithm. For a given history of play $(\theta_1, \mathbf{x}_1, \dots, \theta_{t-1}, \mathbf{x}_{t-1})$ of the past $t - 1$ trials, this algorithm predicts with

$$\theta_t = \operatorname{arginf}_{\theta_t \in \Theta} \sup_{\mathbf{x}_t \in \mathcal{X}_0} \inf_{\theta_{t+1} \in \Theta} \sup_{\mathbf{x}_{t+1} \in \mathcal{X}_0} \cdots \inf_{\theta_T \in \Theta} \sup_{\mathbf{x}_T \in \mathcal{X}_0} \left(\sum_{t=1}^T L(\mathbf{x}_t, \theta_t) - \inf_{\theta_B \in \Theta} \sum_{t=1}^T L(\mathbf{x}_t, \theta_B) \right).$$

The minimax algorithm achieves the best possible regret (called *minimax regret* or the *value* of the game). However this algorithm usually cannot be computed efficiently. In addition the horizon T of the game might not be known to the learner. Therefore we introduce a simple heuristic for the learner called the *Last-step Minimax Algorithm* that behaves as follows: Choose the minimax prediction assuming that the current trial is the last one (i.e. assuming that $T = t$). More precisely, the Last-step Minimax Algorithm predicts with

$$\theta_t = \operatorname{arginf}_{\theta_t \in \Theta} \sup_{\mathbf{x}_t \in \mathcal{X}_0} \left(\sum_{q=1}^t L(\mathbf{x}_q, \theta_q) - \inf_{\theta_B \in \Theta} \sum_{q=1}^t L(\mathbf{x}_q, \theta_B) \right).$$

This method for motivating learning algorithms was first used by Forster [4] for linear regression.

We apply the Last-step Minimax Algorithm to density estimation with one-dimensional exponential families. The exponential families include many fundamental classes of distributions such as Bernoulli, Binomial, Poisson, Gaussian, Gamma and so on. In particular, we consider the game (Θ, \mathcal{X}_0) , where Θ is the exponential family that is specified by a convex¹ function F and $\mathcal{X}_0 = [A, B]$ for some $A < B$. We show that the prediction of the Last-step Minimax Algorithm is explicitly represented as

$$\theta_t = \frac{t}{B - A} \left(F(\alpha_t + B/t) - F(\alpha_t + A/t) \right),$$

where $\alpha_t = \sum_{q=1}^{t-1} \mathbf{x}_q/t$. Moreover we show that its regret is $M \ln T + O(1)$, where

$$M = \max_{A \leq \alpha \leq B} \frac{F''(\alpha)(\alpha - A)(B - \alpha)}{2}.$$

In particular, for the case of Bernoulli, we show that the regret of the Last-step Minimax Algorithm is at most

$$\frac{1}{2} \ln(T + 1) + c, \tag{1}$$

where $c = 1/2$. This is very close to the minimax regret that Shtarkov showed for the fixed horizon game [7]. The minimax regret has the same form (1) but now $c = (1/2) \ln(\pi/2) \approx .23$.

¹ The function F is the dual of the cumulant function (See next section).

Another simple and efficient algorithm for density estimation with an arbitrary exponential family is the *Forward Algorithm* of Azoury and Warmuth [2]. This algorithm predicts with $\boldsymbol{\mu}_t = (a + \sum_{q=1}^{t-1} \mathbf{x}_q)/t$ for any exponential family. Here $a \geq 0$ is a constant that is to be tuned and the mean parameter $\boldsymbol{\mu}_t$ is an alternate parameterization of the density. For a Bernoulli, the Forward Algorithm with $a = 1/2$ is the well-known Krichevsky-Trofimov probability estimator. The regret of this algorithm is again of the same form as (1) with $c = (1/2) \ln \pi \approx .57$ (See e.g. [5]). Surprisingly, the Last-step Minimax Algorithm is slightly better than the Krichevsky-Trofimov probability estimator ($c = .5$).

For general one-dimensional exponential families, the Forward Algorithm can be seen as a first-order approximation of the Last-step Minimax Algorithm. However, in the special case of Gaussian density estimation and linear regression, the Last-step Minimax Algorithm is identical to the Forward Algorithm² for some choice of a . For linear regression this was first pointed out by Forster [4].

In [2] upper bounds on the regret of the Forward Algorithm were given for specific exponential families. For all the specific families considered there, the bounds we can prove for the Last-step Minimax Algorithm are as good or better. In this paper we also give a bound of $M \ln T + O(1)$ that holds for a large class of one-dimensional exponential families. No such bound is known for the Forward Algorithm.

It is interesting to note that for Gaussian density estimation of unit variance, there exists a gap between the regret of the Last-step Minimax algorithm and the regret of the optimal minimax algorithm. Specifically, the former is $O(\ln T)$, while the latter is $O(\ln T - \ln \ln T)$ [10]. This contrasts with the case of Bernoulli, where the regret of the Last-step Minimax Algorithm is by a constant larger than the minimax regret.

Open Problems

There are a large number of open problems.

1. Is the regret of the Last-step Minimax Algorithm always of the form $O(\ln T)$ for density estimation with any member of the exponential family?
2. Does the Last-step Minimax Algorithm always have smaller regret than the Forward Algorithm?
3. For what density estimation and regression problems is the regret of the Last-step Minimax Algorithm “close to” the regret of the optimal minimax algorithm?
4. It is easy to generalize the Last-step Minimax Algorithm to the q -last-step Minimax algorithm where q is some constant larger than one. How does q affect the regret of the algorithm? How large should q be chosen so that the regret of the algorithm is essentially as good as the minimax algorithm.

² More strictly, for linear regression the Last-step Minimax Algorithm “clips” the predictions of the Forward Algorithm so that the absolute value of the predictions is bounded.

Regret Bounds from the MDL Community

There is a large body of work on proving regret bounds that has its roots in the Minimum Description Length community [6, 11, 8, 9, 12, 13]. The definition of regret used in this community is different from ours in the following two parts.

1. The learner predicts with an arbitrary probability mass function q_t . In particular q_t does not need to be in the model class $\{p(\cdot|\boldsymbol{\theta}) \mid \boldsymbol{\theta} \in \Theta\}$. On the other hand, in our setting we require the predictions of the learner to be “proper” in the sense that they must lie in the same underlying model class.
2. The individual instances \mathbf{x}_t does not need to be bounded. The adversary is instead required to choose an instance sequence $\mathbf{x}_1, \dots, \mathbf{x}_T$ so that the best off-line parameter $\boldsymbol{\theta}_B$ for the sequence belongs to a compact subset $K \subseteq \Theta$. For density estimation with an exponential family, this condition implies that $(1/T) \sum_{t=1}^T \mathbf{x}_t \in K$.

In comparison with the setting in this paper, it is obvious that part 1 gives more choices to the learner while part 2 gives more choices to the adversary. Therefore the regret bounds obtained in the MDL setting are usually incomparable with those in our setting. In particular, Rissanen [6] showed under some condition on Θ that the minimax regret is

$$\frac{n}{2} \ln \frac{T}{2\pi} + \ln \int_K \sqrt{|I(\boldsymbol{\theta})|} d\boldsymbol{\theta} + o(1), \quad (2)$$

where $\Theta \subseteq \mathbf{R}^n$ is of dimension n and

$$I(\boldsymbol{\theta}) = (E_{\boldsymbol{\theta}}(-\partial^2 \ln p(\cdot|\boldsymbol{\theta})/\partial\theta_i\partial\theta_j))_{i,j}$$

denotes the Fisher information matrix of $\boldsymbol{\theta}$. This bound is quite different from our bound $M \ln T + O(1)$.

2 On-line Density Estimation

We first give a general framework of the on-line density estimation problem with a parametric class of distributions. Let $\mathcal{X} \subseteq \mathbf{R}^n$ denote the instance space and $\Theta \subseteq \mathbf{R}^d$ denote the parameter space. Each parameter $\boldsymbol{\theta} \in \Theta$ represents a probability distribution over \mathcal{X} . Specifically let $p(\cdot|\boldsymbol{\theta})$ denote the probability mass function that $\boldsymbol{\theta}$ represents. An on-line algorithm called the *learner* is a function $\hat{\boldsymbol{\theta}} : \mathcal{X}^* \rightarrow \Theta$ that is used to choose a parameter based on the past instance sequence. The protocol proceeds in trials. In each trial $t = 1, 2, \dots$ the learner chooses a parameter $\boldsymbol{\theta}_t = \hat{\boldsymbol{\theta}}(\mathbf{x}^{t-1})$, where $\mathbf{x}^{t-1} = (\mathbf{x}_1, \dots, \mathbf{x}_{t-1})$ is the instance sequence observed so far. Then the learner receives an instance $\mathbf{x}_t \in \mathcal{X}$ and suffers a loss defined as the negative log-likelihood of \mathbf{x}_t measured by $\boldsymbol{\theta}_t$, i.e.,

$$L(\mathbf{x}_t, \boldsymbol{\theta}_t) = -\ln p(\mathbf{x}_t|\boldsymbol{\theta}_t).$$

The total loss of the learner up to trial T is $\sum_{t=1}^T L(\mathbf{x}_t, \boldsymbol{\theta}_t)$. Let $\boldsymbol{\theta}_{B,T}$ be the best parameter in hindsight (off-line setting). Namely,

$$\boldsymbol{\theta}_{B,T} = \operatorname{arginf}_{\boldsymbol{\theta} \in \Theta} \sum_{t=1}^T L(\mathbf{x}_t, \boldsymbol{\theta}).$$

If we regard the product of the probabilities of the individual instances as the joint probability (i.e., $p(\mathbf{x}^T | \boldsymbol{\theta}) = \prod_{t=1}^T p(\mathbf{x}_t | \boldsymbol{\theta})$), then the best parameter $\boldsymbol{\theta}_{B,T}$ can be interpreted as the maximum likelihood estimator of the observed instance sequence \mathbf{x}^T . We measure the performance of the learner for a particular instance sequence $\mathbf{x}^T \in \mathcal{X}^*$ by the *regret*, or the *relative loss*, defined as

$$R(\hat{\boldsymbol{\theta}}, \mathbf{x}^T) = \sum_{t=1}^T L(\mathbf{x}_t, \boldsymbol{\theta}_t) - \sum_{t=1}^T L(\mathbf{x}_t, \boldsymbol{\theta}_{B,T}).$$

The goal of the learner is to make the regret as small as possible. In this paper we are concerned with the worst-case regret and so we do not put any (probabilistic) assumption on how the instance sequence is generated. In other words, the preceding protocol can be viewed as a game of two players, the learner and the adversary, where the regret is the payoff function. The learner tries to minimize the regret, while the adversary tries to maximize it. In most cases, to get a finite regret we need to restrict the adversary to choose instances from a bounded space (Otherwise the adversary could make the regret unbounded in just one trial). So we let $\mathcal{X}_0 \subseteq \mathcal{X}$ be the set of instances from which instances are chosen. The choice of \mathcal{X}_0 is one of the central issues for analyzing regrets in our learning model.

3 Last-step Minimax Algorithm

If the horizon T of the game is fixed and known in advance, then we can use the minimax algorithm to obtain the optimal learner in the game theoretical sense. The value of the game is the best possible regret that the learner can achieve. In most cases, the value of the game has no closed form and the minimax algorithm is computationally infeasible. Also the number of trials T might not be known to the learner. For this reasons we suggest the following simple heuristic. Assume that the current trial t is the last one (in other words, assume $T = t$) and predict as the Minimax Algorithm would under this assumption. More precisely the Last-step Minimax Algorithm predicts with

$$\begin{aligned} \boldsymbol{\theta}_t &= \operatorname{arginf}_{\boldsymbol{\theta}_t \in \Theta} \sup_{\mathbf{x}_t \in \mathcal{X}_0} \left(\sum_{q=1}^t L(\mathbf{x}_q, \boldsymbol{\theta}_q) - \sum_{q=1}^t L(\mathbf{x}_q, \boldsymbol{\theta}_{B,t}) \right) \\ &= \operatorname{arginf}_{\boldsymbol{\theta}_t \in \Theta} \sup_{\mathbf{x}_t \in \mathcal{X}_0} \left(L(\mathbf{x}_t, \boldsymbol{\theta}_t) - \sum_{q=1}^t L(\mathbf{x}_q, \boldsymbol{\theta}_{B,t}) \right). \end{aligned} \tag{3}$$

The last equality holds since the total loss up to trial $t - 1$ of the learner is constant for the inf and sup operations.

3.1 Last-step minimax algorithm for exponential families

For a vector $\boldsymbol{\theta}$, $\boldsymbol{\theta}'$ denotes the transposition of $\boldsymbol{\theta}$. A class Θ_G of distributions is said to be an exponential family if parameter $\boldsymbol{\theta} \in \Theta_G$ has density function

$$p(\mathbf{x}|\boldsymbol{\theta}) = p_0(\mathbf{x}) \exp(\boldsymbol{\theta}'\mathbf{x} - G(\boldsymbol{\theta})),$$

where $p_0(\mathbf{x})$ represents any factor of density which does not depend on $\boldsymbol{\theta}$. The parameter $\boldsymbol{\theta}$ is called the *natural* parameter. The function $G(\boldsymbol{\theta})$ is a normalization factor so that $\int_{\mathbf{x} \in \mathcal{X}} p(\mathbf{x}|\boldsymbol{\theta}) d\mathbf{x} = 1$ holds, and it is called the *cumulant function* that characterizes the family Θ_G . We first review some basic properties of the family. For further details, see [3, 1]. Let $g(\boldsymbol{\theta})$ denote the gradient vector $\nabla_{\boldsymbol{\theta}} G(\boldsymbol{\theta})$. It is well known that G is a strictly convex function and $g(\boldsymbol{\theta})$ equals the mean of \mathbf{x} , i.e. $g(\boldsymbol{\theta}) = \int_{\mathbf{x} \in \mathcal{X}} \mathbf{x} p(\mathbf{x}|\boldsymbol{\theta}) d\mathbf{x}$. We let $g(\boldsymbol{\theta}) = \boldsymbol{\mu}$ and call $\boldsymbol{\mu}$ the expectation parameter. Since G is strictly convex, the map $g(\boldsymbol{\theta}) = \boldsymbol{\mu}$ has an inverse: Let $f := g^{-1}$. Sometimes it is more convenient to use the expectation parameter $\boldsymbol{\mu}$ instead of its natural parameter $\boldsymbol{\theta}$. Define the second function F over the set of expectation parameters as

$$F(\boldsymbol{\mu}) = \boldsymbol{\theta}'\boldsymbol{\mu} - G(\boldsymbol{\theta}). \quad (4)$$

The function F is called the *dual* of G and strictly convex as well. It is easy to check that $f(\boldsymbol{\mu}) = \nabla_{\boldsymbol{\mu}} F(\boldsymbol{\mu})$. Thus the two parameters $\boldsymbol{\theta}$ and $\boldsymbol{\mu}$ are related by

$$\boldsymbol{\mu} = g(\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}} G(\boldsymbol{\theta}), \quad (5)$$

$$\boldsymbol{\theta} = f(\boldsymbol{\mu}) = \nabla_{\boldsymbol{\mu}} F(\boldsymbol{\mu}). \quad (6)$$

For parameter $\boldsymbol{\theta}$, the negative log-likelihood of \mathbf{x} is $G(\boldsymbol{\theta}) - \boldsymbol{\theta}'\mathbf{x} + \ln P_0(\mathbf{x})$. Since the last term is independent of $\boldsymbol{\theta}$ and thus does not affect the regret, we define the loss function simply as

$$L(\mathbf{x}, \boldsymbol{\theta}) := G(\boldsymbol{\theta}) - \boldsymbol{\theta}'\mathbf{x}. \quad (7)$$

It is easy to see that, for an instance sequence \mathbf{x}^t up to trial t , the best off-line parameter $\boldsymbol{\mu}_{B,t}$ is given by $\boldsymbol{\mu}_{B,t} = \mathbf{x}_{1..t}/t$ (thus, $\boldsymbol{\theta}_{B,t} = f(\mathbf{x}_{1..t}/t)$), where $\mathbf{x}_{1..t}$ is shorthand for $\sum_{q=1}^t \mathbf{x}_q$. Moreover the total loss of $\boldsymbol{\theta}_{B,t}$ is

$$\sum_{q=1}^t L(\mathbf{x}_q, \boldsymbol{\theta}_{B,t}) = -tF(\mathbf{x}_{1..t}/t). \quad (8)$$

From (3), (7) and (8), it immediately follows that the Last-step Minimax Algorithm for the family Θ_G predicts with

$$\boldsymbol{\theta}_t = \operatorname{arginf}_{\boldsymbol{\theta} \in \Theta_G} \sup_{\mathbf{x}_t \in \mathcal{X}_0} (G(\boldsymbol{\theta}) - \boldsymbol{\theta}'\mathbf{x}_t + tF(\mathbf{x}_{1..t}/t)). \quad (9)$$

3.2 For one dimensional exponential families

In what follows we only consider one dimensional exponential families. Let the instance space be $\mathcal{X}_0 = [A, B]$ for some reals $A < B$. Since F is convex, the supremum over \mathbf{x}_t of (9) is attained at a boundary of \mathcal{X}_0 , i.e., $\mathbf{x}_t = A$ or $\mathbf{x}_t = B$. So

$$\boldsymbol{\theta}_t = \operatorname{arginf}_{\boldsymbol{\theta} \in \Theta_G} \max \left\{ G(\boldsymbol{\theta}) - A\boldsymbol{\theta} + tF(\alpha_t + A/t), G(\boldsymbol{\theta}) - B\boldsymbol{\theta} + tF(\alpha_t + B/t) \right\}, \quad (10)$$

where $\alpha_t = \mathbf{x}_{1..t-1}/t$. It is not hard to see that the minimax parameter $\boldsymbol{\theta}_t$ must satisfy $\boldsymbol{\mu}_t = g(\boldsymbol{\theta}_t) \in [A, B]$. So we can restrict the parameter space to

$$\Theta_{G, \mathcal{X}_0} = \{ \boldsymbol{\theta} \in \Theta_G \mid g(\boldsymbol{\theta}) \in [A, B] \}.$$

Since for any $\boldsymbol{\theta} \in \Theta_{G, \mathcal{X}_0}$

$$\frac{\partial}{\partial \boldsymbol{\theta}} \left(G(\boldsymbol{\theta}) - A\boldsymbol{\theta} + tF(\alpha_t + A/t) \right) = g(\boldsymbol{\theta}) - A \geq 0,$$

the first term in the maximum of (10) is monotonically increasing in $\boldsymbol{\theta}$. Similarly the second term is monotonically decreasing. So the minimax parameter $\boldsymbol{\theta}_t$ must be the solution to the equation

$$G(\boldsymbol{\theta}) - A\boldsymbol{\theta} + tF(\alpha_t + A/t) = G(\boldsymbol{\theta}) - B\boldsymbol{\theta} + tF(\alpha_t + B/t).$$

Solving this, we have

$$\boldsymbol{\theta}_t = \frac{t}{B - A} (F(\alpha_t + B/t) - F(\alpha_t + A/t)). \quad (11)$$

Let us confirm that $\boldsymbol{\mu}_t = g(\boldsymbol{\theta}_t) \in [A, B]$. Since F is convex,

$$\begin{aligned} F(\alpha_t + B/t) &= F(\alpha_t + A/t + (B - A)/t) \\ &\geq F(\alpha_t + A/t) + f(\alpha_t + A/t)(B - A)/t. \end{aligned}$$

Plugging this into (11), we have $\boldsymbol{\theta}_t \geq f(\alpha_t + A/t)$. Since g is monotonically increasing and $f = g^{-1}$,

$$\boldsymbol{\mu}_t = g(\boldsymbol{\theta}_t) \geq g(f(\alpha_t + A/t)) = \alpha_t + A/t \geq A. \quad (12)$$

Similarly we can show that

$$F(\alpha_t + A/t) \geq F(\alpha_t + B/t) - f(\alpha_t + B/t)(B - A)/t,$$

which implies

$$\boldsymbol{\mu}_t = g(\boldsymbol{\theta}_t) \leq g(f(\alpha_t + B/t)) = \alpha_t + B/t \leq B.$$

Hence we proved that $\boldsymbol{\mu}_t \in [A, B]$. Note that this argument also shows that

$$\alpha_t + A/t \leq \boldsymbol{\mu}_t \leq \alpha_t + B/t.$$

Therefore, the prediction $\boldsymbol{\mu}_t$ of the Last-step Minimax Algorithm (for the expectation parameter) converges $\alpha_t = \mathbf{x}_{1..t-1}/t$, which is the prediction of the Forward Algorithm.

3.3 Analysis on the regret

Let

$$\delta_t = L(\mathbf{x}_t, \boldsymbol{\theta}_t) - \sum_{q=1}^t L(\mathbf{x}_q, \boldsymbol{\theta}_{B,t}) + \sum_{q=1}^{t-1} L(\mathbf{x}_q, \boldsymbol{\theta}_{B,t-1}).$$

Since

$$\sum_{t=1}^T \delta_t = \sum_{t=1}^T L(\mathbf{x}_t, \boldsymbol{\theta}_t) - \sum_{t=1}^T L(\mathbf{x}_t, \boldsymbol{\theta}_{B,T}) = R(\hat{\boldsymbol{\theta}}, \mathbf{x}^T),$$

bounding δ_t for all individual t 's is a way to obtain an upper bound of the regret $R(\hat{\boldsymbol{\theta}}, \mathbf{x}^T)$. By (12) and (8), the prediction $\boldsymbol{\theta}_t$ of the Last-step Minimax Algorithm (given by (11)) satisfies

$$L(\mathbf{x}_t, \boldsymbol{\theta}_t) - \sum_{q=1}^t L(\mathbf{x}_q, \boldsymbol{\theta}_{B,t}) \leq G(\boldsymbol{\theta}_t) - A\boldsymbol{\theta}_t + tF(\alpha_t + A/t)$$

for any \mathbf{x}_t . Moreover, applying (8) with t replaced by $t - 1$, we have

$$\sum_{q=1}^{t-1} L(\mathbf{x}_q, \boldsymbol{\theta}_{B,t-1}) = -(t-1)F(\mathbf{x}_{1..t-1}/(t-1)) = -(t-1)F\left(\frac{t}{t-1}\alpha_t\right).$$

Hence we have

$$\delta_t \leq G(\boldsymbol{\theta}_t) - A\boldsymbol{\theta}_t + tF(\alpha_t + A/t) - (t-1)F\left(\frac{t}{t-1}\alpha_t\right). \tag{13}$$

In the subsequent sections we will give an upper bound of the regret by bounding the right-hand-side of the above formula.

4 Density estimation with a Bernoulli

For a Bernoulli, an expectation parameter $\boldsymbol{\mu} = g(\boldsymbol{\theta})$ represents the probability distribution over $\mathcal{X} = \{0, 1\}$ given by $p(0|\boldsymbol{\mu}) = 1 - \boldsymbol{\mu}$ and $p(1|\boldsymbol{\mu}) = \boldsymbol{\mu}$. In this case we have $\Theta_G = \mathbf{R}$, $\mathcal{X} = \mathcal{X}_0 = \{0, 1\}$, $G(\boldsymbol{\theta}) = \ln(1 + e^\boldsymbol{\theta})$ and $F(\boldsymbol{\mu}) = \boldsymbol{\mu} \ln \boldsymbol{\mu} + (1 - \boldsymbol{\mu}) \ln(1 - \boldsymbol{\mu})$. From (11) it follows that in each trial t the Last-step Minimax Algorithm predicts with

$$\boldsymbol{\theta}_t = t \ln \left[\frac{(\alpha_t + 1/t)^{\alpha_t + 1/t} (1 - \alpha_t - 1/t)^{1 - \alpha_t - 1/t}}{\alpha_t^{\alpha_t} (1 - \alpha_t)^{1 - \alpha_t}} \right], \tag{14}$$

where $\alpha_t = \mathbf{x}_{1..t-1}/t$. In other words, the prediction for the expectation parameter is

$$\boldsymbol{\mu}_t = \frac{(k+1)^{k+1} (t-k-1)^{t-k-1}}{k^k (t-k)^{t-k} + (k+1)^{k+1} (t-k-1)^{t-k-1}},$$

where $k = \mathbf{x}_{1..t-1}$. This is different from the Krichevsky-Trofimov probability estimator (the Forward Algorithm with $a = \frac{1}{2}$) [5, 2] that predicts with $\boldsymbol{\mu}_t =$

$(k + 1/2)/t$. The worst case regret of the standard algorithm was shown³ to be $(1/2) \ln(T + 1) + (1/2) \ln \pi$. Surprisingly, the regret of the Last-step Minimax Algorithm is slightly better.

Theorem 1. *Let $\hat{\theta}$ be the Last-step Minimax Algorithm that makes predictions according to (14). Then for any instance sequence $\mathbf{x}^T \in \{0, 1\}^*$,*

$$R(\hat{\theta}, \mathbf{x}^T) \leq \frac{1}{2} \ln(T + 1) + \frac{1}{2}.$$

Proof. Recall that the regret is $R(\hat{\theta}, \mathbf{x}^T) = \sum_{t=1}^T \delta_t$ and δ_t is upper-bounded by (13), i.e.,

$$\delta_t \leq G(\theta_t) + tF(\alpha_t) - (t - 1)F\left(\frac{t}{t - 1}\alpha_t\right).$$

(Note that for the case of Bernoulli the above inequality is an equality.) We can show that the r.h.s. of the above formula is concave in α_t and maximized at $\alpha_t = (t - 1)/(2t)$. Plugging this into (14) we have $\theta_t = 0$. So

$$\begin{aligned} \delta_t &\leq G(0) + tF\left(\frac{t - 1}{2t}\right) - (t - 1)F(1/2) \\ &= \ln 2 + \frac{t - 1}{2} \ln \frac{t - 1}{2t} + \frac{t + 1}{2} \ln \frac{t + 1}{2t} - (t - 1) \ln(1/2) \\ &= \frac{t - 1}{2} \ln(t - 1) + \frac{t + 1}{2} \ln(t + 1) - t \ln t \\ &= \left(\frac{t + 1}{2} \ln(t + 1) - \frac{t}{2} \ln t\right) - \left(\frac{t}{2} \ln t - \frac{t - 1}{2} \ln(t - 1)\right). \end{aligned}$$

Therefore

$$\begin{aligned} R(\hat{\mu}, \mathbf{x}^T) &= \sum_{t=1}^T \delta_t \leq \frac{T + 1}{2} \ln(T + 1) - \frac{T}{2} \ln T \\ &= \frac{1}{2} \ln((T + 1)(1 + 1/T)^T) \\ &\leq \frac{1}{2} \ln(T + 1) + \frac{1}{2}. \end{aligned}$$

This completes the theorem.

5 Density Estimation with a General Exponential Family

In this section we give an upper bound of the regret of the Last-step Minimax Algorithm for a general exponential family, provided that the second and the third derivative of $F(\mu)$ is bounded for any $\mu \in [A, B]$. Note that the Bernoulli family do not satisfy this condition because the second derivative $F''(\mu) = 1/\mu + 1/(1 - \mu)$ is unbounded when $\mu = 0$ and $\mu = 1$.

³ This regret is achieved in the case where the sequence consists of all 0s or all 1s.

Theorem 2. Assume that $|F''(\boldsymbol{\mu})|$ and $|F'''(\boldsymbol{\mu})|$ is upper-bounded by a constant for any $\boldsymbol{\mu} \in [A, B]$. Then for any instance sequence $\mathbf{x}^T \in [A, B]^T$, the regret of the Last-step Minimax Algorithm is upper-bounded by

$$R(\hat{\boldsymbol{\theta}}, \mathbf{x}^T) \leq M \ln T + O(1),$$

where

$$M = \max_{A \leq \alpha \leq B} \frac{F''(\alpha)(\alpha - A)(B - \alpha)}{2}.$$

Proof. As in the case of the Bernoulli, we will bound

$$\delta_t \leq G(\boldsymbol{\theta}_t) - A\boldsymbol{\theta}_t + tF(\alpha_t + A/t) - (t - 1)F(t\alpha_t/(t - 1))$$

for each t to obtain an upper bound of the regret $R(\hat{\boldsymbol{\theta}}, \mathbf{x}^T) = \sum_{t=1}^T \delta_t$. The prediction $\boldsymbol{\theta}_t$ of the Last-step Minimax Algorithm is given by (11), i.e.,

$$\boldsymbol{\theta}_t = \frac{t}{B - A} \left(F(\alpha_t + B/t) - F(\alpha_t + A/t) \right),$$

where $\alpha_t = \mathbf{x}_{1..t-1}/t$. Applying Taylor's expansion of F up to the third degree, we have

$$\begin{aligned} F(\alpha_t + B/t) &= F(\alpha_t + A/t) + f(\alpha_t + A/t) \frac{B - A}{t} + \frac{f'(\alpha_t + A/t)}{2} \left(\frac{B - A}{t} \right)^2 \\ &\quad + O(1/t^3) \\ &= F(\alpha_t + A/t) + f(\alpha_t + A/t) \frac{B - A}{t} + \frac{f'(\alpha_t)}{2} \left(\frac{B - A}{t} \right)^2 \\ &\quad + O(1/t^3). \end{aligned}$$

Note that the last term $O(1/t^3)$ contains the hidden factors $f''(\alpha_t)$ and $f''(\alpha_t + A/t)$, which are assumed to be bounded by a constant. So the Last-step Minimax prediction is rewritten as

$$\boldsymbol{\theta}_t = f(\alpha_t + A/t) + \frac{f'(\alpha_t)}{2} \frac{B - A}{t} + O(1/t^2).$$

The Taylor's expansion of G gives

$$\begin{aligned} G(\boldsymbol{\theta}_t) &= G(f(\alpha_t + A/t)) + g(f(\alpha_t + A/t)) \frac{f'(\alpha_t)(B - A)}{2t} + O(1/t^2) \\ &= (\alpha_t + A/t)f(\alpha_t + A/t) - F(\alpha_t + A/t) \\ &\quad + \frac{\alpha_t f'(\alpha_t)(B - A)}{2t} + O(1/t^2). \end{aligned} \tag{15}$$

Here we used the relations $f = g^{-1}$ and $G(f(\boldsymbol{\mu})) = f(\boldsymbol{\mu})\boldsymbol{\mu} - F(\boldsymbol{\mu})$ (See (4) and (6)). Similarly

$$(t - 1)F \left(\frac{t}{t - 1} \alpha_t \right) = (t - 1)F \left((\alpha_t + A/t) + (\alpha_t/(t - 1) - A/t) \right)$$

$$\begin{aligned}
 &= (t-1) \left[F(\alpha_t + A/t) + f(\alpha_t + A/t)(\alpha_t/(t-1) - A/t) \right. \\
 &\quad \left. + \frac{1}{2} f'(\alpha_t + A/t)(\alpha_t/(t-1) - A/t)^2 + O(1/t^3) \right] \\
 &= (t-1)F(\alpha_t + A/t) + (\alpha_t + A/t)f(\alpha_t + A/t) - Af(\alpha_t + A/t) \\
 &\quad + \frac{1}{2t} f'(\alpha_t)(\alpha_t - A)^2 + O(1/t^2). \tag{16}
 \end{aligned}$$

Thus, (15) $- A\theta_t + tF(\alpha_t + A/t) - (16)$ gives

$$\begin{aligned}
 \delta_t &\leq \frac{f'(\alpha_t)(\alpha_t - A)(B - \alpha_t)}{2t} + O(1/t^2) \\
 &\leq \frac{M}{t} + O(1/t^2).
 \end{aligned}$$

This establishes the theorem.

5.1 Density estimation with a Gaussian of unit variance

For a Gaussian of unit variance, an expectation parameter μ represents the density

$$p(x|\mu) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x - \mu)^2\right).$$

Thus we have $\Theta_G = \mathbf{R}$, $\mathcal{X} = \mathbf{R}$, $\mathcal{X}_0 = [A, B]$, $G(\theta) = \frac{1}{2}\theta^2$ and $F(\mu) = \frac{1}{2}\mu^2$. In this case, the Last-step Minimax Algorithm predicts with

$$\theta_t = \mu_t = \alpha_t + \frac{A + B}{2t}.$$

Since $F''(\mu) = 1$ for all μ , Theorem 2 says that the regret of the Last-step Minimax Algorithm is

$$R(\hat{\theta}, x^T) \leq \frac{(B - A)^2}{8} \ln T + O(1).$$

Note that for Gaussian density estimation the Last-step Minimax Algorithm predicts with the same value as the Forward Algorithm. So here we just have alternate proofs for previously published bounds [2].

5.2 Density estimation with a Gamma of unit shape parameter

For a Gamma of unit shape parameter, an expectation parameter μ represents the density

$$p(x|\mu) = \mu e^{-x\mu}.$$

In this case we have $\Theta_G = (-\infty, 0)$, $\mathcal{X} = (0, \infty)$, $\mathcal{X}_0 = [A, B]$, $G(\theta) = -\ln(-\theta)$ and $F(\mu) = -1 - \ln \mu$. The Last-step Minimax Algorithm predicts with

$$\theta_t = -1/\mu = -\frac{t}{B - A} (\ln(\alpha_t + B/t) - \ln(\alpha_t + A/t)).$$

Since $F''(\boldsymbol{\mu}) = 1/\boldsymbol{\mu}^2$, Theorem 2 says that the regret of the Last-step Minimax Algorithm is

$$R(\hat{\boldsymbol{\theta}}_t, \mathbf{x}^T) \leq \frac{(B-A)^2}{8AB} \ln T + O(1).$$

Previously, the $O(\ln T)$ regret bound is also shown for the Forward Algorithm [2]. However, the hidden constant in the order notation has not been explicitly specified.

Acknowledgments

The authors are grateful to Jun'ichi Takeuchi for useful discussions.

References

1. S. Amari. *Differential Geometrical Methods in Statistics*. Springer Verlag, Berlin, 1985.
2. K. Azoury and M. K. Warmuth. Relative loss bounds for on-line density estimation with the exponential family of distributions. In *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*, pages 31–40, San Francisco, CA, 1999. Morgan Kaufmann. To appear in Machine Learning.
3. O. Barndorff-Nielsen. *Information and Exponential Families in Statistical Theory*. Wiley, Chichester, 1978.
4. J. Forster. On relative loss bounds in generalized linear regression. In *12th International Symposium on Fundamentals of Computation Theory*, pages 269–280, 1999.
5. Y. Freund. Predicting a binary sequence almost as well as the optimal biased coin. In *Proc. 9th Annu. Conf. on Comput. Learning Theory*, pages 89–98. ACM Press, New York, NY, 1996.
6. J. Rissanen. Fisher information and stochastic complexity. *IEEE Transactions on Information Theory*, 42(1):40–47, 1996.
7. Y. M. Shtarkov. Universal sequential coding of single messages. *Prob. Pered. Inf.*, 23:175–186, 1987.
8. J. Takeuchi and A. Barron. Asymptotically minimax regret for exponential families. In *SITA '97*, pages 665–668, 1997.
9. J. Takeuchi and A. Barron. Asymptotically minimax regret by bayes mixtures. In *IEEE ISIT '98*, 1998.
10. E. Takimoto and M. Warmuth. The minimax strategy for Gaussian density estimation. In *To appear in COLT2000*, 2000.
11. Q. Xie and A. Barron. Asymptotic minimax regret for data compression, gambling, and prediction. *IEEE Trans. on Information Theory*, 46(2):431–445, 2000.
12. K. Yamanishi. A decision-theoretic extension of stochastic complexity and its applications to learning. *IEEE Transaction on Information Theory*, 44(4):1424–39, July 1998.
13. K. Yamanishi. Extended stochastic complexity and minimax relative loss analysis. In *Proc. 10th International Conference on Algorithmic Learning Theory - ALT'99*, volume 1720 of *Lecture Notes in Artificial Intelligence*, pages 26–38. Springer-Verlag, 1999.