The Minimax Strategy for Gaussian Density Estimation

Eiji Takimoto[*]
t2@ecei.tohoku.ac.jp
Graduate School of Information Sciences
Tohoku University, Sendai, 980-8579 Japan

Manfred Warmuth[†]
manfred@cse.ucsc.edu
Computer Science Department
UC Santa Cruz, CA 95064

Abstract

We consider on-line density estimation with a Gaussian of unit variance. In each trial t the learner predicts a mean $\boldsymbol{\theta}_t$. Then it receives an instance $\boldsymbol{x}_t$ chosen by the adversary and incurs loss $\frac{1}{2}(\boldsymbol{\theta}_t - \boldsymbol{x}_t)^2$. The performance of the learner is measured by the regret de£ned as the total loss of the learner minus the total loss of the best mean parameter chosen off-line. We assume that the horizon T of the protocol is £xed and known to both parties. We give the optimal strategies for both the learner and the adversary. The value of the game is $\frac{1}{2}X^2(\ln T - \ln\ln T + O(\ln\ln T/\ln T))$, where X is an upper bound of the 2-norm of instances. We also consider the standard algorithm that predicts with $\boldsymbol{\theta}_t = \sum_{q=1}^{t-1} \boldsymbol{x}_q/(t-1+a)$ for a £xed a. We show that the regret of this algorithm is $\frac{1}{2}X^2(\ln T - O(1))$ regardless of the choice of a.

1 Introduction

Consider the following simple repeated game based on Gaussian density estimation. The learner plays against an adversary. In each trial t the learner produces a mean $\boldsymbol{\theta}_t$. Then the adversary provides an instance vector $\boldsymbol{x}_t$ and the loss of the learner is $\frac{1}{2}(\boldsymbol{\theta}_t - \boldsymbol{x}_t)^2$ (in other words we assume unit variance). Assume the *horizon of the game* (number of trials) is £xed to T and T is known to both parties. Consider the following *regret* or *relative loss*

$$\frac{1}{2}\sum_{t=1}^{T}(\boldsymbol{\theta}_t - \boldsymbol{x}_t)^2 - \inf_{\boldsymbol{\theta}_B}\frac{1}{2}\sum_{t=1}^{T}(\boldsymbol{\theta}_B - \boldsymbol{x}_t)^2.$$

This is the total on-line loss of the learner minus the total loss of the best mean 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.

For the analogous problem of density estimation over a discrete domain w.r.t. log loss, Shtarkov gave the minimax strategy and an implicit form of the value of the game called the *minimax regret* [8]. Freund [3] gives an explicit formula for the minimax regret for Bernoulli density estimation: $(1/2)\ln(T+1) + \ln(\pi/2) - O(1/\sqrt{T})$. The minimax strategy has also been computed for the universal portfolio problem [5]. In this case the strategy is not ef£ciently computable, but the minimax regret for the universal portfolio problem is the same as the minimax regret for Bernoulli density estimation. Our work on the minimax regret is different from a large body of work that has its roots in the Minimum Description Length community [6, 7, 12, 9, 10, 13]. In short we require the learner to choose its on-line parameters from the same model class from which the best off-line parameter is chosen. We will discuss the differences in Section 3.

In this paper we give the minimax strategy for Gaussian density estimation for both the learner and the adversary. These strategies are simple and ef£cient. At trial $1 \le t \le T$ the learner should intuitively choose the average of the past $t-1$ instances as its mean. However the optimal strategy of the learner is to choose $\boldsymbol{\theta}_t = c_t \sum_{q=1}^{t-1} \boldsymbol{x}_q$ where c_t is slightly smaller than $1/t$: $c_t = 1/(t + \ln T - \ln(t + O(\ln T)))$. (Note that c_t depends on the horizon.) We give a simple recurrence for the optimal shrinkage factor c_t. If the learner plays optimally then the regret is $\frac{1}{2}\sum_{t=1}^{T} c_t \boldsymbol{x}_t^2$. To get the minimax regret we need to restrict the adversary to choose instances of 2-norm bounded above by some constant X (Otherwise the

[*]This work was done while Eiji Takimoto was on a sabbatical leave at UC Santa Cruz.

[†]Manfred Warmuth was supported by NSF grant CCR-9821087.

adversary can make the regret unbounded in just one trial). Now the minimax regret of horizon T is

$$\frac{X^2}{2}\sum_{t=1}^{T} c_t = \frac{1}{2}X^2(\ln T - \ln\ln T + O(\ln\ln T/\ln T)). \quad (1)$$

The $-\ln\ln T$ term is surprising because many on-line games were shown to have $O(\ln T)$ upper bounds for the minimax regret [2, 5, 3, 11, 1, 13, 14].

There are some intriguing properties of the optimal strategies of both parties. First, the learner does not need to know the upper bound X on the 2-norm of the instances. Second, the strategies we give for both players are still optimal even when the opponent plays non-optimally in the past. Third, the adversary can restrict its choice of instances to two points: plus or minus X times a unit vector. Even with this restricted choice of the instances the adversary can force a regret at least as large as the game value. In other words the algorithm cannot take advantage of the restricted choice of the adversary.

Perhaps the most natural algorithm for Gaussian density estimation is to start with an initial instance $\boldsymbol{x}_0$ and predict with $\boldsymbol{\theta}_t = \frac{a\boldsymbol{x}_0 + \sum_{q=1}^{t-1}\boldsymbol{x}_q}{a+t-1}$. Here $a \geq 0$ is the multiplicity of the initial instance. The initial instance is chosen to be zero for Gaussian density estimation. This prediction algorithm is the *forward algorithm* of [1]. The same algorithm was investigated in parallel work by Gordon [4]. The forward algorithm was inspired by a similar related algorithm of Vovk for linear regression [11].

We show that the regret of the forward algorithm is larger than $\frac{1}{2}X^2(\ln T - O(1))$ regardless of the choice of a. This holds even if the constant a is allowed to depend on the horizon T. On the other hand, for the £xed choice of $a = 1$ the forward algorithm works without knowing T and the regret is at most $\frac{1}{2}X^2(1 + \ln T)$ [1]. So, for the forward algorithm there is no signi£cant gap between the cases when the horizon is known or unknown to the learner.

We conjecture that if the horizon is not known then an adversary can always force any learner to have regret at least $\frac{1}{2}X^2(\ln T - O(1))$. All lower bound techniques that we know of [13, 11] are of the form $\frac{1}{2}X^2\ln T(1 - o(1))$. Thus these lower bounds do not lie above the value of the game given in (1), which can be expressed as $\frac{1}{2}X^2\ln T(1 - O(\ln\ln T/\ln T))$. This means that the known techniques are not strong enough to bring out the difference between the cases when the horizon is known or unknown.

Besides resolving the above conjecture this work raises a number of open problems. Are there other cases where the minimax strategies are simple and ef£cient? In particular, we don't know the minimax strategy for linear regression and for Gaussian density estimation with an arbitrary variance.

2 On-line density estimation with a Gaussian

We £rst give a formal framework of the on-line density estimation problem with Gaussian densities. For a vector $\boldsymbol{x}$, $\boldsymbol{x}'$ denotes the transposition of $\boldsymbol{x}$ and $\boldsymbol{x}^2$ is shorthand for the real $\boldsymbol{x}'\boldsymbol{x}$. An n dimensional Gaussian $N(\boldsymbol{\theta}, \Sigma)$ has density function

$$\frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}}\exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\theta})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\theta})\right).$$

In this paper we assume that the variance-covariance matrix Σ is £xed and known. In this case we can construct from Σ a linear transformation A that maps an instance $\boldsymbol{x}$ to $\boldsymbol{z} = A\boldsymbol{x}$ so that $\boldsymbol{z}$ is subject to the Gaussian $N(\boldsymbol{\nu}, I)$, where $\boldsymbol{\nu} = A\boldsymbol{\theta}$ and I is the unit matrix. So without loss of generality we can assume that the parameter space consists of mean vectors. Namely a mean $\boldsymbol{\theta}$ represents the density function

$$p(\boldsymbol{x}|\boldsymbol{\theta}) = \frac{1}{(2\pi)^{n/2}}\exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\theta})^2\right).$$

For $\boldsymbol{\theta}$ and an instance $\boldsymbol{x}$ we de£ne the loss as the negative log-likelihood $-\ln p(\boldsymbol{x}|\boldsymbol{\theta}) = \frac{1}{2}(\boldsymbol{x}-\boldsymbol{\theta})^2 + c$, where c is a constant independent of $\boldsymbol{x}$ and $\boldsymbol{\theta}$. Since the constant term does not matter in our analysis, we de£ne the loss for $\boldsymbol{x}$ and $\boldsymbol{\theta}$ simply as $\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\theta})^2$. We restrict the instance space $\mathcal{X}$ to the set of vectors with 2-norm at most X for some real $X > 0$. That is, we let $\mathcal{X} = \{\boldsymbol{x} \in \mathbf{R}^n \mid \|\boldsymbol{x}\| \leq X\}$, where $\|\boldsymbol{x}\| = \sqrt{\boldsymbol{x}^2}$ denotes the 2-norm of $\boldsymbol{x}$.

An on-line algorithm called the learner is a function $\hat{\boldsymbol{\theta}} : \mathcal{X}^* \to \mathbf{R}^n$ that is used to choose a parameter based on the past instance sequence. The protocol proceeds in trials. In each trial $t = 1, 2, \ldots, T$ the learner chooses a parameter $\boldsymbol{\theta}_t = \hat{\boldsymbol{\theta}}(\boldsymbol{x}^{t-1})$, where $\boldsymbol{x}^{t-1} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{t-1})$ is the instance sequence observed so far. Then the learner receives an instance $\boldsymbol{x}_t \in \mathcal{X}$ and suffers loss $\frac{1}{2}(\boldsymbol{\theta}_t - \boldsymbol{x}_t)^2$. The total loss of the learner is $\frac{1}{2}\sum_{t=1}^{T}(\boldsymbol{\theta}_t - \boldsymbol{x}_t)^2$. Let $\boldsymbol{\theta}_B$ be the best parameter in hindsight (off-line setting). Namely,

$$\boldsymbol{\theta}_B = \arg\inf_{\boldsymbol{\theta}} \frac{1}{2}\sum_{t=1}^{T}(\boldsymbol{\theta} - \boldsymbol{x}_t)^2 = \frac{\boldsymbol{x}_{1..T}}{T},$$

where $\boldsymbol{x}_{r..s}$ is shorthand for $\sum_{q=r}^{s}\boldsymbol{x}_q$. We measure the performance of the learner $\hat{\boldsymbol{\theta}}$ for a particular instance sequence $\boldsymbol{x}^T$ by the *regret*, or the *relative loss*, de£ned as

$$R_T(\hat{\boldsymbol{\theta}}, \boldsymbol{x}^T) = \frac{1}{2}\sum_{t=1}^{T}(\boldsymbol{\theta}_t - \boldsymbol{x}_t)^2 - \inf_{\boldsymbol{\theta}_B}\frac{1}{2}\sum_{t=1}^{T}(\boldsymbol{\theta}_B - \boldsymbol{x}_t)^2.$$

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 make any (probabilistic) assumption on how the instance sequence is generated. In other words, the preceding protocol can be viewed as a game between 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.

Assume that the horizon (the number of trials T) of the game is £xed and known to both the learner and the adversary. In this case the game value called the *minimax regret* is well-de£ned and given by

$$R_T = \inf_{\hat{\boldsymbol{\theta}}}\sup_{\boldsymbol{x}^T \in \mathcal{X}^T} R_T(\hat{\boldsymbol{\theta}}, \boldsymbol{x}^T).$$

Alternatively we can de£ne the minimax regret as

$$R_T = \inf_{\boldsymbol{\theta}_1}\sup_{\boldsymbol{x}_1 \in \mathcal{X}} \cdots \inf_{\boldsymbol{\theta}_T}\sup_{\boldsymbol{x}_T \in \mathcal{X}} \left(\frac{1}{2}\sum_{t=1}^{T}(\boldsymbol{\theta}_t - \boldsymbol{x}_t)^2 - \inf_{\boldsymbol{\theta}_B}\frac{1}{2}\sum_{t=1}^{T}(\boldsymbol{\theta}_B - \boldsymbol{x}_t)^2\right).$$

The minimax regret R_T is achieved when both the learner and the adversary play optimally.

3 Difference from Rissanen's stochastic complexity model

There is a large body of work on proving regret bounds that has its roots in the Minimum Description Length community [7, 12, 9, 10, 13]. In their model, the learner is interpreted as a coding scheme for the set of sequences of length T. The coding scheme is speci£ed by a probability mass function $q(\boldsymbol{x}^T)$. (Note that q is not necessarily i.i.d.) In each trial t, the learner (coding scheme) q £rst provides the conditional $q(\cdot|\boldsymbol{x}^{t-1})$ based on the past sequence $\boldsymbol{x}^{t-1}$. The conditional de£nes the coding for the next instance. The learner then observes the instance $\boldsymbol{x}_t$ and incurs loss $-\ln q(\boldsymbol{x}_t|\boldsymbol{x}^{t-1})$ which is the code length of $\boldsymbol{x}_t$. Thus, the total loss

$$\sum_{t=1}^{T} -\ln q(\boldsymbol{x}_t|\boldsymbol{x}^{t-1}) = -\ln q(\boldsymbol{x}^T)$$

is the code length of the sequence $\boldsymbol{x}^T$. The regret of the learner q for $\boldsymbol{x}^T$ relative to a family of probability mass functions $\{p(\cdot|\boldsymbol{\theta}) \mid \boldsymbol{\theta} \in \Theta\}$ is de£ned as

$$R'_T(q, \boldsymbol{x}^T) = -\ln q(\boldsymbol{x}^T) + \ln p(\boldsymbol{x}^T|\boldsymbol{\theta}_B(\boldsymbol{x}^T)),$$

where

$$\boldsymbol{\theta}_B(\boldsymbol{x}^T) = \arg\inf_{\boldsymbol{\theta}\in\Theta} -\ln p(\boldsymbol{x}^T|\boldsymbol{\theta})$$

is the maximum likelihood estimator of $\boldsymbol{x}^T$ in Θ. That is, the regret $R'_T(q, \boldsymbol{x}^T)$ is the code length of $\boldsymbol{x}^T$ based on q minus the code length based on the ideal coding scheme for the parameter space Θ. So the regret can be thought of as the redundancy of the coding scheme q for $\boldsymbol{x}^T$ relative to Θ. Let W_T be a set of sequences of length T. Then the minimax regret for the set W_T relative to Θ is de£ned as

$$R'_T = \inf_q \sup_{\boldsymbol{x}^T\in W_T} R'_T(q, \boldsymbol{x}^T).$$

Let q^* be the optimal coding scheme that attains the minimax regret. Rissanen called the code length $-\ln q^*(\boldsymbol{x}^T)$ of $\boldsymbol{x}^T$ the *stochastic complexity* of $\boldsymbol{x}^T$ with respect to W_T and Θ. In particular, Rissanen [7] showed under some condition on Θ that if $W_T = \{\boldsymbol{x}^T \mid \boldsymbol{\theta}_B(\boldsymbol{x}^T) \in K\}$ for some compact subset $K \subseteq \Theta$, then

$$R'_T = \frac{n}{2}\ln\frac{T}{2\pi} + \ln\int_K \sqrt{|I(\boldsymbol{\theta})|}\, d\boldsymbol{\theta} + o(1), \qquad (2)$$

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 $\boldsymbol{\theta}$. In the case of Gaussian density with unit variance, the parameter space Θ is $\mathbf{R}^n$ and the Fisher information matrix is the unit matrix.

The minimax regret de£ned above is different from ours in the following two points.

1. The coding scheme q is arbitrary. In particular q does not need to be in the model class $\{p(\cdot|\boldsymbol{\theta}) \mid \boldsymbol{\theta} \in \Theta\}$. The model class is used just as the reference set to measure the performance of the learner. 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 that is used to de£ne the loss of the best off-line parameter. That is, for Gaussian density estimation, the predictions of the learner must be Gaussian as well.

2. The individual instances $\boldsymbol{x}_t$ does not need to be bounded. For Gaussian density estimation, a natural choice would be $K = \mathcal{X} = \{\boldsymbol{x} \mid \|\boldsymbol{x}\| \le X\}$. In this case, the condition $\boldsymbol{x}^T \in W_T$ means $\boldsymbol{x}_{1..T}/T \in \mathcal{X}$, which is much weaker than the condition that $\boldsymbol{x}_t \in \mathcal{X}$ for all t that we use.

In comparison with the setting in this paper, it is obvious that difference 1 gives more choices to the learner while difference 2 gives more choices to the adversary. In fact, for Gaussian density estimation, R'_T is incomparable with R_T. More precisely, for Gaussian density estimation, the minimax regret of (2) with $K = \{\boldsymbol{x} \mid \|\boldsymbol{x}\| \le X\}$ is

$$\begin{aligned} R'_T &= \frac{n}{2}\ln\frac{T}{2\pi} \\ &\quad + \ln(\text{volume of the ball of radius } X) + o(1). \\ &= \frac{n}{2}(1+\ln T) + n\ln X - \frac{n+1}{2}\ln n - \frac{1}{2}\ln\pi + o(1). \end{aligned}$$

The second term of the £rst equality comes from the fact that Fisher information matrix is the unit matrix. The second equality is derived from the fact that the volume of the ball of radius X is $X^n\pi^{n/2}/\Gamma(n/2+1)$. On the other hand, we will show that the minimax regret in our setting is

$$R_T = \frac{1}{2}X^2(\ln T - \ln\ln T) + o(1).$$

It is interesting to see that R'_T depends on the dimension n while R_T does not. Moreover, in R'_T the bound X of the instance space appears in a term independent of T while in R_T it appears in the leading term.

4 Minimax regret for Gaussian density estimation

We £rst de£ne a sequence of shrinkage factors that will be used to de£ne the optimal predictions of the learner.

De£nition 1 *Let $\{c_t\}_{t=0,\ldots,T}$ be the sequence recursively de£ned as*

$$\begin{aligned} c_T &= 1/T, \\ c_{t-1} &= c_t + c_t^2 \quad (1 \le t \le T). \end{aligned}$$

Suppose in trial t that, based on the past sequence $\boldsymbol{x}^{t-1}$, the learner chooses a parameter $\boldsymbol{\theta}_t = \hat{\boldsymbol{\theta}}(\boldsymbol{x}^{t-1})$. Now we represent the parameter by $\Delta_t = \boldsymbol{\theta}_t - c_t\boldsymbol{x}_{1..t-1}$ and in what follows we sometimes use Δ_t to denote the choice of the learner. (Recall that $\boldsymbol{x}_{1..t-1}$ is shorthand for $\sum_{q=1}^{t-1}\boldsymbol{x}_q$.) The vector Δ_t is the offset from $c_t\boldsymbol{x}_{1..t-1}$ and the latter will be shown to be the optimal choice of the learner.

Lemma 1 *For any learner $\hat{\boldsymbol{\theta}}$ and any instance sequence $\boldsymbol{x}^T \in (\mathbf{R}^n)^T$,*

$$R_T(\hat{\boldsymbol{\theta}}, \boldsymbol{x}^T) = \frac{1}{2}\sum_{t=1}^{T} c_t \boldsymbol{x}_t^2 + \sum_{t=1}^{T}(c_t \boldsymbol{x}_{1..t-1} - \boldsymbol{x}_t)'\Delta_t + \frac{1}{2}\sum_{t=1}^{T}\Delta_t^2. \quad (3)$$

Proof. For a Gaussian the best parameter is $\boldsymbol{\theta}_B = \boldsymbol{x}_{1..T}/T$. So by de£nition the regret is

$$\begin{aligned} R_T(\hat{\boldsymbol{\theta}}, \boldsymbol{x}^T) &= \frac{1}{2}\sum_{t=1}^{T}(\boldsymbol{x}_t - \boldsymbol{\theta}_t)^2 - \frac{1}{2}\sum_{t=1}^{T}(\boldsymbol{x}_t - \boldsymbol{x}_{1..T}/T)^2 \\ &= \frac{1}{2}\sum_{t=1}^{T}\boldsymbol{\theta}_t^2 - \sum_{t=1}^{T}\boldsymbol{x}_t'\boldsymbol{\theta}_t + \frac{(\boldsymbol{x}_{1..T})^2}{2T}. \end{aligned}$$

Plugging $\boldsymbol{\theta}_t = c_t \boldsymbol{x}_{1..t-1} + \Delta_t$ into the above formula, we have

$$\begin{aligned} & R_T(\hat{\boldsymbol{\theta}}, \boldsymbol{x}^T) \\ &= \frac{1}{2}\sum_{t=1}^{T}(c_t \boldsymbol{x}_{1..t-1} + \Delta_t)^2 \\ &\quad - \sum_{t=1}^{T}\boldsymbol{x}_t'(c_t \boldsymbol{x}_{1..t-1} + \Delta_t) + \frac{1}{2}c_T(\boldsymbol{x}_{1..T})^2 \\ &= \frac{1}{2}\sum_{t=1}^{T} c_t^2(\boldsymbol{x}_{1..t-1})^2 + \frac{1}{2}\sum_{t=1}^{T}\Delta_t^2 + \sum_{t=1}^{T} c_t \boldsymbol{x}_{1..t-1}'\Delta_t \\ &\quad - \sum_{t=1}^{T}\boldsymbol{x}_t'(c_t \boldsymbol{x}_{1..t-1} + \Delta_t) + \frac{1}{2}c_T(\boldsymbol{x}_{1..T})^2. \quad (4) \end{aligned}$$

Since $c_{t-1} = c_t + c_t^2$ the £rst sum is

$$\begin{aligned} & \sum_{t=1}^{T} c_t^2(\boldsymbol{x}_{1..t-1})^2 \\ &= \sum_{t=1}^{T}(c_{t-1} - c_t)(\boldsymbol{x}_{1..t-1})^2 \\ &= \sum_{t=1}^{T}\big(c_{t-1}(\boldsymbol{x}_{1..t-1})^2 - c_t(\boldsymbol{x}_{1..t})^2 \\ &\qquad + c_t(\boldsymbol{x}_{1..t})^2 - c_t(\boldsymbol{x}_{1..t-1})^2\big) \\ &= -c_T(\boldsymbol{x}_{1..T})^2 + \sum_{t=1}^{T} c_t\left(2\boldsymbol{x}_t'\boldsymbol{x}_{1..t-1} + \boldsymbol{x}_t^2\right). \end{aligned}$$

Plugging this into (4) we have the lemma. □

Note that in the lemma we do not need to bound the instance space.

Now we show that $\Delta_t = 0$ (i.e., $\boldsymbol{\theta}_t = c_t \boldsymbol{x}_{1..t-1}$) gives the optimal choice of the learner. The next theorem follows immediately from Lemma 1.

Theorem 1 *Let $\hat{\boldsymbol{\theta}}$ be the learner that chooses $\Delta_t = 0$ for all t. Then for any instance sequence $\boldsymbol{x}^T \in \mathcal{X}^T$,*

$$R_T(\hat{\boldsymbol{\theta}}, \boldsymbol{x}^T) = \frac{1}{2}\sum_{t=1}^{T} c_t \boldsymbol{x}_t^2 \le \frac{1}{2}c_{1..T}X^2.$$

It is interesting to see that the £rst equality in the above theorem holds even when the instance space is unbounded. Moreover we can see that whenever all instances are of the same norm (i.e., $\|\boldsymbol{x}_t\| = X$) the regret is $\frac{1}{2}c_{1..T}X^2$.

Next we give a strategy for the adversary that chooses instances: For each trial t if the learner chooses Δ_t, then the adversary chooses

$$\boldsymbol{x}_t = -X\Delta_t/\|\Delta_t\|. \quad (5)$$

Here we use a convention that $\mathbf{0}/\|\mathbf{0}\|$ denotes the unit vector $(1, 0, \ldots, 0)$. Note that $\boldsymbol{x}_t^2 = X^2$. We will use the following lemma.

Lemma 2 *For any $1 \le t \le T$ and any instance sequence $\boldsymbol{x}^{t-1} \in \mathcal{X}^{t-1}$, $\|c_t \boldsymbol{x}_{1..t-1}\| < X$.*

Proof. An easy induction shows that $c_t \le 1/t$. By the triangular inequality we have

$$\|c_t \boldsymbol{x}_{1..t-1}\| \le c_t \sum_{q=1}^{t-1}\|\boldsymbol{x}_q\| \le (t-1)X/t < X.$$

□

Theorem 2 *Let $\hat{\boldsymbol{\theta}}$ be any learner. Let $\boldsymbol{x}^T$ be the sequence in which each instance is given by (5). Then,*

$$R_T(\hat{\boldsymbol{\theta}}, \boldsymbol{x}^T) \ge \frac{1}{2}c_{1..T}X^2 + \frac{1}{2}\sum_{t=1}^{T}\Delta_t^2.$$

Proof. Lemma 1 with $\boldsymbol{x}_t = -X\Delta_t/\|\Delta_t\|$ gives

$$R_T(\hat{\boldsymbol{\theta}}, \boldsymbol{x}^T) = \frac{1}{2}c_{1..T}X^2 + \sum_{t=1}^{T}(c_t \boldsymbol{x}_{1..t-1}\Delta_t + X\|\Delta_t\|) + \frac{1}{2}\sum_{t=1}^{T}\Delta_t^2.$$

Using Lemma 2 we have

$$c_t \boldsymbol{x}_{1..t-1}\Delta_t \ge -\|c_t \boldsymbol{x}_{1..t-1}\|\|\Delta_t\| \ge -X\|\Delta_t\|.$$

This completes the theorem. □

Surprisingly the adversary can restrict its choice of instances to two points $\mathcal{X} = \{-X\boldsymbol{e}, X\boldsymbol{e}\}$, where $\boldsymbol{e}$ is an arbitrary unit vector, say, $\boldsymbol{e} = (1, 0, \ldots, 0)$. Even with this restricted choice of the instances, we claim that the adversary can force a regret at least as large as the game value. Now the adversary chooses $\boldsymbol{x}_t = -X(\boldsymbol{e}'\Delta_t)\boldsymbol{e}/|\boldsymbol{e}'\Delta_t|$. Namely,

$$\boldsymbol{x}_t = \begin{cases} -X\boldsymbol{e} & \text{if } \boldsymbol{e}'\Delta_t \ge 0, \\ X\boldsymbol{e} & \text{if } \boldsymbol{e}'\Delta_t < 0. \end{cases}$$

For this choice, it is not hard to see that the summand in the second term of (3) is still positive:

$$\begin{aligned} c_t \boldsymbol{x}_{1..t-1}'\Delta_t - \boldsymbol{x}_t'\Delta_t &\ge -\|c_t \boldsymbol{x}_{1..t-1}\||\boldsymbol{e}'\Delta_t| + X|\boldsymbol{e}'\Delta_t| \\ &> 0. \end{aligned}$$

So the claim holds.

By Theorem 1 and Theorem 2 we can conclude that the optimal strategy of the learner is to choose $\boldsymbol{\theta}_t = c_t \boldsymbol{x}_{1..t-1}$ and the minimax regret is $\frac{1}{2}c_{1..T}X^2$. It is surprising that the

optimal choice of the learner does not depend on X. This implies that the learner does not need to know the bound of the instance space. Unfortunately the coef£cients c_t (see the recurrence of De£nition 1) depend on the horizon T and do not have a closed form. However we can show tight bounds for c_t and $c_{1..T}$.

Lemma 3 *For any* $1 \le t \le T$

$$\frac{1}{t+\ln(T+1)-\ln(t+1)} \le c_t \le \frac{1}{t+\ln(T+1)-\ln(t+2+\ln(T+1))}$$

and

$$c_{1..T} = \ln T - \ln\ln T + O\left(\frac{\ln\ln T}{\ln T}\right).$$

We will give a proof in the appendix. We summarize the results of this section in the next corollary.

Corollary 1 *For Gaussian density estimation the minimax regret is*

$$R_T = \inf_{\hat{\boldsymbol{\theta}}} \sup_{\boldsymbol{x}^T} R_T(\hat{\boldsymbol{\theta}}, \boldsymbol{x}^T) = \frac{1}{2} c_{1..T} X^2 = \frac{1}{2} X^2 \left(\ln T - \ln\ln T + O\left(\frac{\ln\ln T}{\ln T}\right)\right)$$

and the in£mum is attained by the learner $\hat{\boldsymbol{\theta}}$ *given by*

$$\hat{\boldsymbol{\theta}}(\boldsymbol{x}^{t-1}) = c_t \boldsymbol{x}_{1..t-1}.$$

5 Optimal play against non-optimal player

We showed that the learner's strategy $\boldsymbol{\theta}_t = c_t \boldsymbol{x}_{1..t-1}$ is optimal in the sense that it gives the minimum regret assuming that the adversary plays optimally. However this might not be true when the adversary plays non-optimally. Surprisingly we can show that $\boldsymbol{\theta}_t = c_t \boldsymbol{x}_{1..t-1}$ is still optimal even if the adversary plays non-optimally. Similarly the adversary's choice $\boldsymbol{x}_t = -X\Delta_t/\|\Delta_t\|$ turns out to be optimal even if the learner plays non-optimally.

To be more precise we extend the notion of regret in the situation where initial choices of the both players are given. Let $h^{t-1} = (\boldsymbol{\theta}_1, \boldsymbol{x}_1, \ldots, \boldsymbol{\theta}_{t-1}, \boldsymbol{x}_{t-1})$ be any history of play up to trial $t-1$. Note that the choices in the history are not necessarily optimal. Now we de£ne the minimax regret given h^{t-1} as

$$R_T|_{h^{t-1}} = \inf_{\boldsymbol{\theta}_t} \sup_{\boldsymbol{x}_t \in \mathcal{X}} \inf_{\boldsymbol{\theta}_{t+1}} \sup_{\boldsymbol{x}_{t+1} \in \mathcal{X}} \cdots \inf_{\boldsymbol{\theta}_T} \sup_{\boldsymbol{x}_T \in \mathcal{X}} \left(\frac{1}{2}\sum_{q=1}^{T} (\boldsymbol{x}_q - \boldsymbol{\theta}_q)^2 - \frac{1}{2}\sum_{q=1}^{T} (\boldsymbol{x}_q - \boldsymbol{x}_{1..T}/T)^2\right).$$

Similarly for history $h^{t-1} \circ \boldsymbol{\theta}_t$, i.e., h^{t-1} followed by some $\boldsymbol{\theta}_t$, we de£ne the minimax regret given $h^{t-1} \circ \boldsymbol{\theta}_t$ as

$$R_T|_{h^{t-1}\circ\boldsymbol{\theta}_t} = \sup_{\boldsymbol{x}_t \in \mathcal{X}} \inf_{\boldsymbol{\theta}_{t+1}} \sup_{\boldsymbol{x}_{t+1} \in \mathcal{X}} \cdots \inf_{\boldsymbol{\theta}_T} \sup_{\boldsymbol{x}_T \in \mathcal{X}} \left(\frac{1}{2}\sum_{q=1}^{T} (\boldsymbol{x}_q - \boldsymbol{\theta}_q)^2 - \frac{1}{2}\sum_{q=1}^{T} (\boldsymbol{x}_q - \boldsymbol{x}_{1..T}/T)^2\right).$$

Clearly, for the empty history $h = \epsilon$, $R_T|_h$ gives the minimax regret which was shown to be $R_T = \frac{1}{2} c_{1..T} X^2$ in Corollary 1.

The arguments of the previous section combined with an easy induction can be used to show the next theorem.

Theorem 3 *Let* $h^{t-1} = (\boldsymbol{\theta}_1, \boldsymbol{x}_1, \ldots \boldsymbol{\theta}_{t-1}, \boldsymbol{x}_{t-1})$ *be any history of play. Then*

$$R_T|_{h^{t-1}} = \inf_{\boldsymbol{\theta}_t} R_T|_{h^{t-1}\circ\boldsymbol{\theta}_t} = \frac{1}{2}\sum_{q=1}^{t-1} c_q \boldsymbol{x}_q^2 + \sum_{q=1}^{t-1} (c_q \boldsymbol{x}_{1..q-1} - \boldsymbol{x}_q)' \Delta_q + \frac{1}{2}\sum_{q=1}^{t-1} \Delta_q^2 + \frac{1}{2} c_{t..T} X^2$$

and the in£mum over $\boldsymbol{\theta}_t$ *is attained at* $c_t \boldsymbol{x}_{1..t-1}$. *Moreover for any* $\boldsymbol{\theta}_t$

$$R_T|_{h^{t-1}\circ\boldsymbol{\theta}_t} = \sup_{\boldsymbol{x}_t \in \mathcal{X}} R_T|_{h^{t-1}\circ\boldsymbol{\theta}_t\circ\boldsymbol{x}_t} \ge R_T|_{h^{t-1}} + \Delta_t^2$$

and the supremum over $\boldsymbol{x}_t$ *is attained at* $-X\Delta_t/\|\Delta_t\|$.

6 A Lower bound for the forward algorithm

By Lemma 3 the optimal shrinkage factor c_t is roughly $1/(t + \ln T - \ln t)$. A good approximation to c_t might be to use shrinkage factors of the form $1/(t-1+a)$, for some universal constant $a \ge 0$. This learner is called the *forward* algorithm. The constant a parameterizes a prior [1, 4]. In particular, Azoury and Warmuth [1] showed that the forward algorithm with $a = 1$ has the worst case regret of $\frac{1}{2} X^2 (1 + \ln T)$.

More precisely the forward algorithm is the Bayes-optimal algorithm that minimizes the expected regret under the following probabilistic setup: The adversary £rst chooses $p \in [0, 1]$ according to $(a/2, a/2)$-beta prior and then in each trial generates $\boldsymbol{x}_t = X$ with probability p and $\boldsymbol{x}_t = -X$ with $1 - p$. (Here $n = 1$.) In this probabilistic setup the expected regret of the forward algorithm is shown to be $\frac{a}{2(a+1)} X^2 \ln T + O(1)$ [11]. When a is large then the optimal algorithm has expected regret at least $\frac{1-\epsilon}{2} X^2 \ln T$. Thus this probabilistic argument [11] gives a lower bound of $\frac{1-\epsilon}{2} X^2 \ln T$ for the (worst-case) regret of any algorithm. Note that this lower bound lies below the minimax regret of $\frac{1}{2} X^2 (\ln T - \ln\ln T + o(1))$ proven in this paper.

In this section we show that a particular adversary can force the forward algorithm to have regret at least $\frac{1}{2} X^2 (\ln T - O(1))$. The sequence produced by the adversary is decidedly not i.i.d. For the sake of simplicity we assume $X = 1$, $n = 1$ and we shift a by one. Thus the forward algorithm predicts with $\boldsymbol{\theta}_t = \boldsymbol{x}_{1..t-1}/(t + a)$. In other words, $\Delta_t = (1/(t + a) - c_t)\boldsymbol{x}_{1..t-1}$. In the appendix we show that c_t is of the form $1/(t + d_t)$ where d_t is a monotonically decreasing sequence ending with $d_T = 0$. So there exists a $t_0 \in \{0, \ldots, T\}$ such that for $1 \le t \le t_0$, Δ_t has the same sign as $\boldsymbol{x}_{1..t-1}$, and for $t_0 + 1 \le t \le T$, Δ_t has the opposite sign as $\boldsymbol{x}_{1..t-1}$. So the adversary's strategy of Theorem 3 chooses the instance $\boldsymbol{x}_t = -\mathrm{sgn}(\boldsymbol{x}_{1..t-1})$, for

$1 \le t \le t_0$, and $\boldsymbol{x}_t = \mathrm{sgn}(\boldsymbol{x}_{1..t-1})$, for $t_0 + 1 \le t \le T$. This strategy produces the following instance sequence: $\boldsymbol{x}^T = (1, -1, 1, -1, \ldots, 1, -1, 1, 1, \ldots, 1)$. Namely,

$$\boldsymbol{x}_t = \begin{cases} 1 \text{ if } (1 \le t \le t_0 \text{ and } t \text{ is odd}) \text{ or } (t \ge t_0 + 1), \\ -1 \text{ if } (1 \le t \le t_0 \text{ and } t \text{ is even}). \end{cases} \tag{6}$$

Note that this would not be the optimal choice for an adversary playing against the forward algorithm because the forward algorithm plays non-optimally in the future. However the adversary given in Theorem 3 assumes that the learner plays optimally in the future. We will show that the above sequence with an appropriate choice of t_0 makes the regret large enough for obtaining the lower bound.

Theorem 4 *Let $\hat{\boldsymbol{\theta}}$ be the forward algorithm that predicts with $\hat{\boldsymbol{\theta}}(\boldsymbol{x}^{t-1}) = \boldsymbol{x}_{1..t-1}/(t+a)$ for a £xed $a \in \mathbf{R}$. Then there exists a $t_0 \in \{0, \ldots, T\}$ such that the instance sequence $\boldsymbol{x}^T$ de£ned in (6) gives*

$$R_T(\hat{\boldsymbol{\theta}}, \boldsymbol{x}^T) \ge \frac{1}{2} \ln T - c - o(1),$$

where $c \le 0.55$.

Proof. By the de£nition of the instance sequence, it is obvious that

$$\boldsymbol{\theta}_t = \begin{cases} 0 & \text{if } 1 \le t \le t_0 \text{ and } t \text{ is odd}, \\ \frac{1}{t+a} & \text{if } 1 \le t \le t_0 \text{ and } t \text{ is even}, \\ \frac{t-t_0-1}{t+a} & \text{if } t \ge t_0 + 1. \end{cases}$$

It is straightforward to show that the regret becomes

$$\begin{aligned} R_T(\hat{\boldsymbol{\theta}}, \boldsymbol{x}^T) &= \frac{1}{2} \sum_{t=1}^{T} (\boldsymbol{\theta}_t - \boldsymbol{x}_t)^2 - \frac{1}{2} \sum_{t=1}^{T} \left(\boldsymbol{x}_t - \frac{\boldsymbol{x}_{1..T}}{T} \right)^2 \\ &= \frac{1}{2} \sum_{\substack{1 \le t \le t_0 \\ t:\text{even}}} \left(\frac{1}{(t+a)^2} + \frac{2}{t+a} \right) \\ &\quad + \frac{1}{2} (t_0 + a + 1)^2 \sum_{t=t_0+1}^{T} \frac{1}{(t+a)^2} \\ &\quad - \frac{t_0}{2} + \frac{t_0^2}{2T}. \end{aligned}$$

The £rst sum is lower-bounded by

$$\begin{aligned} &\frac{1}{2} \sum_{1 \le q \le t_0/2} \left(\frac{1}{(2q+a)^2} + \frac{2}{2q+a} \right) \\ &\ge \frac{1}{2} \int_1^{t_0/2+1} \left(\frac{1}{(2x+a)^2} + \frac{2}{2x+a} \right) dx \\ &= \frac{1}{4(a+2)} - \frac{1}{4(t_0+a+2)} \\ &\quad + \frac{1}{2} \ln(t_0 + a + 2) - \frac{1}{2} \ln(a+2) \end{aligned}$$

and similarly the second sum by

$$\begin{aligned} &\frac{1}{2} (t_0 + a + 1)^2 \left(\frac{1}{t_0+a+1} - \frac{1}{T+a+1} \right) \\ &= \frac{t_0 + a + 1}{2} - \frac{(t_0+a+1)^2}{2(T+a+1)}. \end{aligned}$$

So the regret is lower-bounded by

$$\begin{aligned} R_T(\hat{\boldsymbol{\theta}}, \boldsymbol{x}^T) \ge\ & \frac{1}{4(a+2)} - \frac{1}{4(t_0+a+2)} \\ & + \frac{1}{2} \ln(t_0 + a + 2) - \frac{1}{2} \ln(a+2) \\ & + \frac{a+1}{2} - \frac{(t_0+a+1)^2}{2(T+a+1)} + \frac{t_0^2}{2T}. \end{aligned} \tag{7}$$

Plugging $t_0 = 0$ and $t_0 = T$ into the above formula, we have

$$R_T(\hat{\boldsymbol{\theta}}, \boldsymbol{x}^T) \ge \frac{a+1}{2} - \frac{(a+1)^2}{2(T+a+1)}$$

and

$$\begin{aligned} R_T(\hat{\boldsymbol{\theta}}, \boldsymbol{x}^T) \ge\ & \frac{1}{2} \ln(T + a + 2) - \frac{1}{2} \ln(a+2) \\ & + \frac{1}{4(a+2)} - \frac{1}{4(T+a+2)}, \end{aligned}$$

respectively. From these we can show that if $a \ge \ln T$ or $a \le 1.45$ then $R_T(\hat{\boldsymbol{\theta}}, \boldsymbol{x}^T) \ge \frac{1}{2} \ln T - c - o(1)$. So in the following we assume $1.45 \le a \le \ln T$. In this case we choose $t_0 = T/(2a+1)$. Then (7) becomes

$$\begin{aligned} R_T(\hat{\boldsymbol{\theta}}, \boldsymbol{x}^T) \ge\ & \frac{1}{2} \ln T - \frac{1}{2} \ln(2a+1) + \frac{1}{4(a+2)} \\ & - \frac{1}{2} \ln(a+2) + \frac{a+1}{2} \left(\frac{2a}{2a+1} \right)^2 - o(1). \end{aligned}$$

The r.h.s. of the above formula is monotonically increasing in a when $a \ge 1$ and so is minimized at $a = 1.45$. A simple calculation shows that $R_T(\hat{\boldsymbol{\theta}}, \boldsymbol{x}^T) \ge \frac{1}{2} \ln T - c - o(1)$. □

Acknowledgments

The authors are grateful to Jürgen Forster and Kenji Yamanishi for useful discussions and to Jun'ichi Takeuchi for helping understand the stochastic complexity result quoted in Section 3.

References

[1] K. Azoury and M. K. Warmuth. Relative loss bounds for on-line density estimation with the exponential family of distributions. *Journal of Machine Learning*, 43(3):211–246, June 2001. Special issue on *Theoretical Advances in On-line Learning, Game Theory and Boosting*, edited by Yoram Singer.

[2] T. M. Cover. Universal portfolios. *Mathematical Finance*, 1(1):1–29, 1991.

[3] 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.

[4] G. J. Gordon. Approximate solutions to Markov decision processes. Ph. D. thesis, Department of Computer Science, Carnegie Mellon University, Pittsburgh. Technical report CMU-CS-99-143, June 1999.

[5] E. Ordentlich and T. Cover. The cost of achieving the best portfolio in hindsight. *Mathematics of Operations Research*, 23(4):960–982, 1998.

[6] J. Rissanen. Stochastic complexity in learning. In *Computational Learning Theory: Eurocolt '95*, pages 196–210. Springer-Verlag, 1995.
[7] J. Rissanen. Fisher information and stochastic complexity. *IEEE Transactions on Information Theory*, 42(1):40–47, 1996.
[8] Y. M. Shtarkov. Universal sequential coding of single messages. *Prob. Pered. Inf.*, 23:175–186, 1987.
[9] J. Takeuchi and A. Barron. Asymptotically minimax regret for exponential families. In *SITA '97*, pages 665–668, 1997.
[10] J. Takeuchi and A. Barron. Asymptotically minimax regret by bayes mixtures. In *IEEE ISIT '98*, 1998.
[11] V. Vovk. Competitive on-line linear regression. Technical Report CSD-TR-97-13, Department of Computer Science, Royal Holloway, University of London, 1997.
[12] Q. Xie and A. Barron. Asymptotic minimax regret for data compression, gambling, and prediction. *IEEE Trans. on Information Theory*, 46(2):431–445, 2000.
[13] 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.
[14] K. Yamanshi. Extended stochastic complexity and minimax relative loss analysis,. In *Proc. 10th International Conference on Algorithmic Learning Theory - ALT'99*, pages 26–38. Sringer Verlag, 1999. volume 1720 of *Lecture Notes in Arti£cial Intelligence*.

A Proof of Lemma 3

We want to estimate c_t and $c_{1..T} = \sum_{t=1}^{T} c_t$, where the sequence $\{c_t\}$ is de£ned by the recurrence

$$c_{t-1} = c_t(1 + c_t) \tag{8}$$

for $1 \le t \le T$ and $c_T = 1/T$. Taking logarithm for both sides of (8) we have

$$\ln c_{t-1} - \ln c_t = \ln(1 + c_t).$$

Since the inequalities $x - x^2/2 \le \ln(1 + x) \le x$ hold for any $x \ge 0$,

$$c_t - c_t^2/2 \le \ln c_{t-1} - \ln c_t \le c_t.$$

By (8) we can replace c_t^2 by $c_{t-1} - c_t$ and get

$$c_t - (c_{t-1} - c_t)/2 \le \ln c_{t-1} - \ln c_t \le c_t.$$

Summing the above inequalities for t over $\{1, \ldots, T\}$ we have

$$c_{1..T} - (c_0 - c_T)/2 \le \ln c_0 - \ln c_T \le c_{1..T}.$$

Since $c_T = 1/T$ we get both upper and lower bounds for $c_{1..T}$:

$$\ln T + \ln c_0 \le c_{1..T} \le \ln T + \ln c_0 + (c_0 - 1/T)/2. \tag{9}$$

We need to estimate c_0 as a function of T to express the bound in a closed form. Now we de£ne another sequence $\{d_t\}$ so that

$$c_t = \frac{1}{t + d_t}$$

holds for any t. Plugging the above equality into (8) we have a recurrence with respect to $\{d_t\}$:

$$d_{t-1} = d_t + \frac{1}{d_t + t + 1}. \tag{10}$$

Note that $d_T = 0$ and $c_0 = 1/d_0$. It is clear from (10) that

$$d_{t-1} \le d_t + \frac{1}{t+1} \le \frac{1}{t+1} + \frac{1}{t+2} + \cdots + \frac{1}{T+1}.$$

The r.h.s. is a harmonic sum and so we have

$$d_t \le \ln(T + 1) - \ln(t + 1). \tag{11}$$

This gives a lower bound of c_t:

$$c_t = \frac{1}{t + d_t} \ge \frac{1}{t + \ln(T + 1) - \ln(t + 1)}. \tag{12}$$

Similarly, plugging (11) into the second term of (10) we have

$$\begin{aligned} d_{t-1} &\ge d_t + \frac{1}{t + 1 + \ln(T + 1) - \ln(t + 1)} \\ &\ge d_t + \frac{1}{t + 1 + \ln(T + 1)} \end{aligned}$$

and so

$$\begin{aligned} d_t &\ge \ln(T + 2 + \ln(T + 1)) - \ln(t + 2 + \ln(T + 1)) \\ &\ge \ln(T + 1) - \ln(t + 2 + \ln(T + 1)), \end{aligned}$$

which gives an upper bound of c_t:

$$c_t \le \frac{1}{t + \ln(T + 1) - \ln(t + 2 + \ln(T + 1))}. \tag{13}$$

From (12) and (13) it follows that

$$\frac{1}{\ln(T + 1)} \le c_0 \le \frac{1}{\ln(T + 1) - \ln(2 + \ln(T + 1))}.$$

Plugging this into (9) we can easily get

$$c_{1..T} = \ln T - \ln \ln T + O\left(\frac{\ln \ln T}{\ln T}\right).$$