Log-loss games with bounded adversaries

Jacob Andreas Dylan Hadfield-Menell

June 18, 2014

1 Introduction

The *log-loss game*, of the form

$$R_{\text{NML}} = \min_{P} \max_{\underline{x}} -\log P(\underline{x}) - \inf_{\theta} -\log P_\theta(\underline{x}) \tag{1}$$

has been a popular subject of study in the machine learning literature, due to deep connections with information theory and minimum description length learning [2], as well as the relatively simple form of its regret-minimizing estimator P_{NML}, which is given by:

$$P_{\text{NML}}(\underline{x}) = \frac{\sup_\theta P_\theta(\underline{x})}{\sum_{\underline{x}'} \sup_\theta P_\theta(\underline{x}')} \tag{2}$$

Worst-case analysis of the game assumes, by definition, that the adversary is trying to minimize the learner's regret without any restrictions on the resources it uses while doing so. In practice, however, it may not be necessary (or indeed desirable) to get bounds of this kind—real-world data are typically generated by processes of bounded computational power, memory, etc., and it would be useful to have a model in which it is possible to describe guarantees against such bounded adversaries. In this paper, we describe a generalization of the log-loss game which allows explicit penalization of "expensive" adversarial behavior. Our main result is a description of a new estimator which outperforms the NML estimator against adversaries penalized in this way, subject to only very weak assumptions about the penalty function and reference class of models.

1.1 Modeling computational constraints on adversaries

In general, it appears to be difficult to give bounds for adversaries which restricted to particular complexity classes (L, P, etc.) for the same reason that providing lower bounds on computational complexity is hard in general. Thus we need to come up with a mathematically tractable representation of a bounded adversary that doesn't require us to solve open problems in complexity theory.

One such model comes to us from the control theory literature, which suggests representing an agent's objective as a utility term plus a term for "information cost", typically chosen to be the KL divergence between the agent's

policy and some fixed distribution p_0. That is, the agent attempts to find a policy maximizing the functional

$$F[p] = \sum_x p(x)U(x) - \frac{1}{\beta}\sum_x p(x)\log\frac{p(x)}{p_0(x)} \tag{3}$$

for a scale parameter β and utility function U [4]. Here the optimal policy has a form suggestively similar to the NML estimator:

$$p^*(x) = \frac{p_0(x)e^{\beta U(x)}}{\sum'_x p_0(x')e^{\beta U(x)}} \tag{4}$$

Similar models can be found in the microeconomics literature on agency theory, which describes strategies for designing contracts between pairs of principals and agents with different utility functions [1].

Thus it seems useful in general to consider modeling bounded adversaries as trying minimize (for maxing learners) the sum of the learner's objective and some additional term encoding a penalty.

2 Results

2.1 A zero-sum game

Before we investigate the problem of minimizing regret against a penalized adversary (a non-zero-sum game, and thus structurally rather different from the normal NML case), it's instructive to consider a related zero-sum game which admits easier analysis. In particular, consider the problem of finding an optimal prediction strategy P for

$$\min_P \max_{\underline{x}} \; -\log P(\underline{x}) - \inf_\theta -\log P_\theta(\underline{x}) + K(\underline{x}) \tag{5}$$

Here, and throughout this paper, it is useful to think of $K(\underline{x})$ as acting as a kind of regularizer on the adversary's strategy. For the zero-sum game in particular, it is also possible to view $K(\underline{x})$ as a "handicap" which the learner receives when it is asked to predict particularly difficult sequences of examples.

[N.B.: for notational convenience we'll write "$+K(\underline{x})$" throughout this paper. When thinking about concrete examples involving penalties for hard sequences it may be useful to think of this term as being negative, i.e. decreasing the value of the game for the adversary on these hard sequences. As we'll see, in general it doesn't matter what form K has as long as it assigns different scores to at least two sequences.]

Proposition 1. *Suppose K is finite. Then the learner's optimal strategy for the game described in Equation 5 is given by*

$$P_{KNML}(\underline{x}) = \frac{\sup_\theta P_\theta(\underline{x})e^{K(\underline{x})}}{\sum_{\underline{x}'}\sup_\theta P_\theta(\underline{x}')e^{K(\underline{x}')}} \tag{6}$$

Proof. Analysis of this game proceeds almost identically to that of the standard log-loss game: we can immediately see that the choice of P_{KNML} leaves the learner indifferent to the adversary's choice of $\underline{x}$, with a value of

$$V = \log \left(\sum_{\underline{x}'} \sup_{\theta} P_\theta(\underline{x}') e^{K(\underline{x}')} \right) \tag{7}$$

As already indicated, we'll call the estimator in Equation 6 P_{KNML}, short for the (rather unwieldy) *$K(\cdot)$-penalized Normalized Maximum Likelihood* estimator. Relaxing the adversary's strategy to a distribution over sequences, we have

$$V = \min_P \max_P \ \mathbb{E}_{\underline{x} \sim Q} \left[-\log P(\underline{x}) - \inf_\theta -\log P_\theta(\underline{x}) + K(\underline{x}) \right] \tag{8}$$

Sion's minimax theorem tells us that this is equivalent to the dual game:

$$= \max_Q \min_P \ \mathbb{E}_{\underline{x} \sim Q} \left[-\log P(\underline{x}) - \inf_\theta -\log P_\theta(\underline{x}) + K(\underline{x}) \right] \tag{9}$$

$$= \max_Q \ \mathbb{E}_{\underline{x} \sim Q} \left[-\log Q(\underline{x}) - \inf_\theta -\log P_\theta(\underline{x}) + K(\underline{x}) \right] \tag{10}$$

from which it can be shown without too much work that the adversary's optimal strategy is identical to Equation 6, and gives the same value for the game. It follows that the strategy specified by P_{KNML} is optimal. □

However, this game (and the corresponding analysis) is unsatisfying for a variety of reasons. First and foremost, we no longer have a bound on regret, but rather on the sum of regret and the penalty term $K(\underline{x})$, which does not have a straightforward interpretation. Moreover, we cannot show that the value of this game is even *less* than in the normal NML case (i.e., that we have gained anything by placing constraints on the adversary) unless $K(\underline{x})$ is strictly negative, when intuition suggests that arbitrary constant terms in K should not change the behavior of either player, or any measurement by which we assess the learner's performance.

In the next section, we derive standard regret bounds for the learner while maintaining a constrained adversary, addressing these issues.

2.2 A non-zero-sum game

Consider the following problem:

$$\min_P \ -\log P(\underline{X}(P)) - \inf_\theta -\log P_\theta(\underline{x})$$

where

$$\underline{X}(P) = \arg\max_{\underline{x}} \ -\log P(\underline{x}) - \inf_\theta -\log P_\theta(\underline{x}) + K(\underline{x}) \tag{11}$$

We can interpret this as a non-zero-sum game in which the learner is trying to minimize regret, while the adversary is trying to maximize regret plus a penalty term—precisely the setting described in the introduction. Note that we do not, in general, expect the same strategy to be optimal in both this setting and the original setting, and we cannot appeal to minimax theorems to prove optimality.

Nevertheless it's natural to ask what behavior we get by using the same P_{KNML} from above—as we'll see, this still allows us to improve upon regret bounds for the standard game.

We begin with the following observation.

Proposition 2. *If the learner's strategy is given by P_{KNML}, the adversary has a value of*

$$V_A = \log\left(\sum_{\underline{x}'} \sup_\theta P_\theta(\underline{x}') e^{K(\underline{x}')}\right) \tag{12}$$

Suppose the adversary plays $\underline{x}^*$. Then the learner achieves a value (regret) of

$$R_{KNML}(\underline{x}^*) = \log\left(\sum_{\underline{x}'} \sup_\theta P_\theta(\underline{x}') e^{K(\underline{x}') - K(\underline{x}^*)}\right) \tag{13}$$

Proof. For Equation 12, just substitute P_{KNML} into the second part of Equation 11. For Equation 13, we perform the corresponding substitution and get

$$R_{\text{KNML}}(\underline{x}^*) = \log\left(\sum_{\underline{x}'} \sup_\theta P_\theta(\underline{x}') e^{K(\underline{x}')}\right) - K(\underline{x}^*) \tag{14}$$

from which we only need to bring $-K(\underline{x}^*)$ inside the logarithm and distribute across the summation to obtain the desired result. □

Note in particular that the adversary remains indifferent to all of its options upon observing P_{KNML}, but the player's regret actually depends on the adversary's choice. This curious situation suggests that the learner should be able to perturb P_{KNML} slightly in order to "nudge" the adversary towards the $\underline{x}$ which maximizes $K(\underline{x})$ (thus locally minimizing the learner's regret over all possible choices of $\underline{x}^*$ in Equation 13). In particular, we claim the following:

Proposition 3. *For every $\varepsilon > 0$, consider the strategy*

$$P_{\epsilon KNML}(\underline{x}) = \frac{\sup_\theta P_\theta(\underline{x}) e^{K(\underline{x}) - \varepsilon \mathbb{1}[\underline{x} = \underline{x}^*]}}{\sum_{\underline{x}'} \sup_\theta P_\theta(\underline{x}') e^{K(\underline{x}') - \varepsilon \mathbb{1}[\underline{x} = \underline{x}^*]}}$$

where

$$\underline{x}^* \in \arg\max_{\underline{x}} K(\underline{x}) \tag{15}$$

(if the max is not unique, we can choose one arbitrarily). $P_{\epsilon KNML}$ achieves a regret of

$$R_{\epsilon KNML} = \log\left(\sum_{\underline{x}'} \sup_{\theta} P_\theta(\underline{x}') e^{K(\underline{x}')-K(\underline{x}^*)}\right) + \varepsilon = R_{KNML}(\underline{x}^*) + \varepsilon \tag{16}$$

Proof. Again, this is mostly crank-turning. Substitute $P_{\epsilon\text{KNML}}$ into Equation 11 and obtain

$$\underline{X}(P_{\epsilon\text{KNML}}) = \arg\max_{\underline{x}} \log\left(\sum_{\underline{x}'} \sup_{\theta} P_\theta(\underline{x}') e^{K(\underline{x}')-K(\underline{x}^*)-\varepsilon\mathbb{1}[\underline{x}'=\underline{x}^*]}\right) + \varepsilon\mathbb{1}[\underline{x}=\underline{x}^*]$$

The only term which depends on $\underline{x}$ is $\varepsilon\mathbb{1}[\underline{x}=\underline{x}^*]$, which is trivially maximized by choosing $\underline{x}=\underline{x}^*$. Then, substituting back into the learner's side of Equation 11, we have a regret of

$$\begin{aligned} R_{\epsilon\text{KNML}} &= \log\left(\sum_{\underline{x}'} \sup_{\theta} P_\theta(\underline{x}') e^{K(\underline{x}')-K(\underline{x}^*)-\varepsilon\mathbb{1}[\underline{x}'=\underline{x}^*]}\right) + \varepsilon\mathbb{1}[\underline{x}^*=\underline{x}^*] \\ &\leq \log\left(\sum_{\underline{x}'} \sup_{\theta} P_\theta(\underline{x}') e^{K(\underline{x}')-K(\underline{x}^*)}\right) + \varepsilon\mathbb{1}[\underline{x}^*=\underline{x}^*] \\ &= R_{\text{KNML}}(\underline{x}^*) + \varepsilon \end{aligned} \tag{17}$$

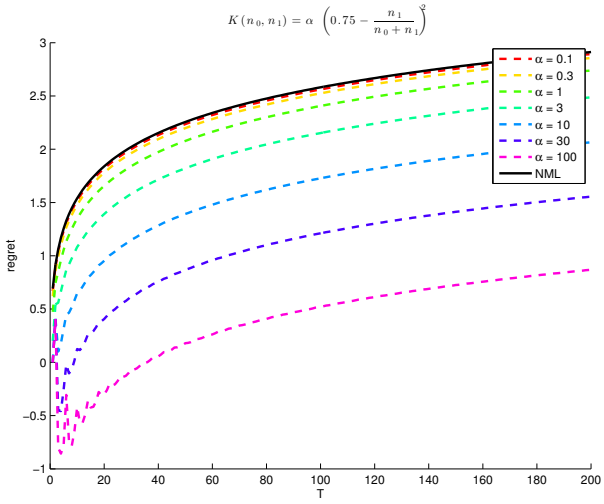
□

Because $\underline{x}^*$ maximizes K, every term $e^{K(\underline{x}')-K(\underline{x}^*)}$ is at most one; moreover, if K assigns a different penalty to any two $\underline{x}_1$ and $\underline{x}_2$, then we can always choose ε small enough so that $R_{\epsilon\text{KNML}}$ is strictly better than the regret for the standard game—as desired, penalizing the adversary has led to an improved regret bound for the learner.

3 Observations and examples

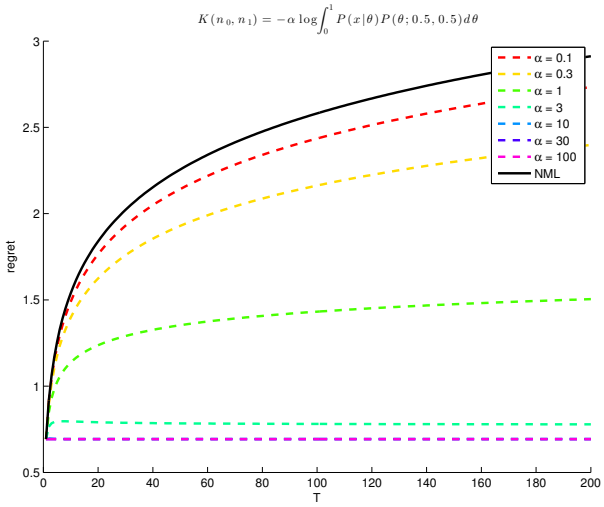
Basic limiting behaviors As a sanity check, observe a few simple properties of this bound: First, desired in Section 2.1, regret is invariant under translations of K: $(K(\underline{x}_1)+C)-(K(\underline{x}_2)+C) = K(\underline{x}_1)-K(\underline{x}_2)$. If we consider scaling K by a multiplicative factor D, we find that as $D\to 0$ we recover the original regret, and as $|D|\to\infty$, all of the mass is placed on maximizers of K; if the max is unique, the learner suffers only ε regret (the adversary is almost completely predictable).

Behaviors and pathologies for the binomial game It's also possible to gain intuition about the behavior of the by plotting $R_{\epsilon\text{KNML}}$ regret curves for various penalties and game lengths. We include a few here. All examples are for binary example sequences, with P_θ the binomial family.
Squared deviation from a target 3/4 ratio of heads to tails:



Observe in particular that for large enough penalties and short enough games, the difficulty of achieving a fraction close to $\frac{3}{4}$ results in oscillation of regret.

Posterior probability of the data under a symmetric Beta(0.5, 0.5) prior.



Note that instead choosing the noninformative prior Beta(1,1) would have resulted in a K constant for all choices of $\underline{x}$; in this case we recover the regular log-loss regret.

4 Learning with unknown penalties

What if the exact form of the adversary's penalty is unknown? Certainly in the case where it can be an arbitrary function we don't expect to be able to prove anything. But suppose we know that the penalty function is some member of a (possibly uncountable) class $\{K_1, K_2, \dots\}$. We claim:

Proposition 4. *If the adversary's penalty is unknown to the learner, but known to be drawn from some fixed class $\mathcal{K}$, then for any $\varepsilon' > 0$ it is still possible to achieve a regret bound of*

$$\log\left(\sum_{\underline{x}'} \sup_{\theta} P_\theta(\underline{x}') e^{K_L(\underline{x})' - K_L(\underline{x}^*) + D}\right) + \varepsilon' \tag{18}$$

where

$$K_L = \arg\min_{K \in \mathcal{K}} \max_{\underline{x}, K' \in \mathcal{K}} K'(\underline{x}) - K(\underline{x}) \tag{19}$$

and

$$D = \max_{\underline{x}, K' \in \mathcal{K}} K'(\underline{x}) - K_L(\underline{x}) \tag{20}$$

Proof. Suppose the adversary is penalized with K_A, and the learner's strategy is given by $P_{\epsilon\text{KNML}}$ with $K = K_L$. Then the K terms no longer cancel in the adversary's objective, and the adversary now tries to optimize

$$\arg\max \; K_A(\underline{x}) - K_L(\underline{x}) + \varepsilon \mathbb{1}[\underline{x} = \underline{x}^*]$$

where $\underline{x}^*$ maximizes K_L. To ensure that the adversary still chooses $\underline{x}^*$, we can no longer take ε arbitrarily small, but it suffices (for all K_A) to set it using the maximum difference $K_A(\underline{x}) - K_L(\underline{x})$. Thus, more generally, let the learner choose K_L as in Equation 19, and set $\varepsilon = D + \varepsilon'$ (with $\varepsilon' > 0$ and D as in Equation 20). Then for all K_A the adversary will choose $\underline{x}^*$, and the learner suffers a regret of

$$\log\left(\sum_{\underline{x}'} \sup_{\theta} P_\theta(\underline{x}') e^{K_L(\underline{x}')}\right) - K_L(\underline{x}^*) + D + \varepsilon'$$

which is Equation 18. □

Simple algebra shows that this strategy remains an improvement over the standard regret bound where

$$D \leq R_{\text{NML}} - R_{\epsilon\text{KNML}} \tag{21}$$

Note that when D does not satisfy this inequality it is always possible to achieve R_{NML} by playing P_{NML} instead.

One simple consequence of this result is that there exists some $\underline{x}^*$ which is optimal for all K in $\mathcal{K}$ (e.g. if all members of the class are the same except for some scaling parameter) it is possible to achieve the $\varepsilon KNML$ regret bound exactly.

5 Relation to existing work

Because computation of the denominator for the NML estimator is often intractable, various other related estimators are considered in the literature. A particularly well-behaved estimator is the Sequetial NML estimator

$$P_{\text{SNML}}(x_t|\underline{x}_{1:T-1}) = \frac{\sup_\theta P_\theta(x_t)}{\sum_{x'} \sup_\theta P_\theta(x')}$$

In fact, it can be easily shown that by choosing K equal to regret from the first $t-1$ rounds, this is exactly the optimal estimator for the zero-sum game initially considered [3].

Other NML-like estimators which look approximately similar to this one include the result of letting K penalize θ inside the regret term rather than $\underline{x}$; this has the form of a (possibly-unnormalized) prior on θ, and can be used to ensure convergence of the denominator when it is otherwise badly behaved [2].

6 Conclusion

We have introduced a non-zero-sum variant of the log-loss game, in which the adversary pays a penalty term $K(\underline{x})$ for each sequence $\underline{x}$ of examples it reveals, and demonstrated that with simple and fairly general conditions on this penalty term, the learner can achieve regret better than for the standard log-loss game.

Various continuations of this work remain possible. The most pressing issue is whether the $P_{\epsilon\text{KNML}}$ estimator is optimal (in the limit where $\varepsilon \to 0$); so far we have neither been able to prove this nor to construct a counterexample. It's also interesting to consider whether particular choices of K allow easy-to-compute predictors (not necessarily $P_{\epsilon\text{KNML}}$) which nonetheless achieve sub-NML regrets against a penalized adversary. For the formulation of the problem with uncertainty about K, it seems likely that it is possible to achieve better guarantees using more sophisticated tools from the analysis of non-perfect-information games, and more generally to achieve a sharper characterization of D in terms of the properties of $\mathcal{K}$.

References

[1] Robert Gibbons. Lecture notes on agency theory, 2010.

[2] Peter Grünwald. A tutorial introduction to the minimum description length principle, 2004.

[3] Fares Hedayati. Minimax optimality in online learning under logarithmic loss with parametric constraint experts, 2013.

[4] Pedro A Ortega and Daniel D Lee. An adversarial interpretation of information-theoretic bounded rationality.