Second-order quantile bounds in online learning

Jaouad Mourtada

Journal club, Thursday, March 16rd

Abstract

These notes reflect the contents of the oral presentation of the paper Koolen and van Erven (2015) given in the journal club. After introducing the Hedge setting and providing some context, including a minimax regret bound, we discuss two kinds of adaptivity to "easy" data: second-order bounds and quantile bounds. We then describe the Squint algorithm proposed by Koolen and van Erven (2015) and show how this strategy can combine those two kinds of adaptivity (something that previous methods could not achieve), while at the same time leading to an efficiently implementable closed-form algorithm.

Contents

1 Introduction: the Hedge setting

Before introducing the so-called *Hedge setting* (Freund and Schapire, 1997), we first describe the related setting of *prediction with expert advice* (see Cesa-Bianchi and Lugosi (2006)) as a motivation.

1.1 Prediction with expert advice

Most approaches in traditional statistics and statistical learning theory proceed under the assumption that the observations are generated by an (unknown) probability distribution, and provide estimators or learning algorithms that perform well — in expectation and/or with high probability — under this assumption. While this stochasticity assumption is often innocuous, legitimate and useful, it is sometimes harder to justify; in such contexts, the traditional statistical results provide virtually no guarantee on the performance of the corresponding methods.

A radically different approach is to make no assumption at all on the signal, and to aim at methods that perform well even in the worst case, often called "adversarial" since they may be modelled as a *game* between the learner (that makes forecasts) and the environment (that generates the true values of the signal). Since it is clearly impossible for the learner to predict well the signal in the worst case without any further assumptions, a more reasonable objective is to predict almost as well as the best of a given set of forecasters called *experts*, which may be black-box prediction mechanisms, learning algorithms or even human experts.

This is formalised in the setting of *prediction with expert advice*:

Problem 1 (Prediction with expert advice)**.** Assume we are given a convex *prediction space* $\mathscr{X}$, a *signal space* $\mathscr{Y}$, a loss function $\ell : \mathscr{X} \times \mathscr{Y} \to \mathbf{R}$ and a number $K \geqslant 2$ of *experts* $k = 1, \dots, K$. At each time step $t \geqslant 1$, the following actions occur sequentially:

1. The experts $k = 1, \dots, K$ output their predictions $x_t^k \in \mathscr{X}$.
2. The learner combines them to form his own forecast $f_t \in \mathscr{X}$.
3. The environment then decides the true value of the signal $y_t \in \mathscr{Y}$.

The goal of the learner is to maintain a small *cumulative regret*

$$R_T^k := \sum_{t=1}^{T} \ell(f_t, y_t) - \sum_{t=1}^{T} \ell(x_t^k, y_t) \tag{1}$$

for each time horizon $T \geqslant 1$, and with respect to each expert $k = 1, \dots, K$, no matter what the experts' predictions $\boldsymbol{x}_t = (x_t^k)_{1 \leqslant k \leqslant K}$ and the values of the signal y_t are.

Specifically, one wants to find strategies that guarantee sub-linear regret bounds: $\max_{1 \leqslant k \leqslant K} R_T^k = o(T)$. It turns out that this is possible, modulo one small additional assumption on ℓ beyond just convexity. As is the case in statistical learning, the curvature of the loss function enables to obtain fast rates: namely, if the loss function is η-*exp-concave* for some $\eta > 0$, in the sense that $\exp(-\eta\, \ell(\cdot, y))$ is concave on $\mathscr{X}$ for each $y \in \mathscr{Y}$, then the *exponential weights* algorithm, which forecasts

$$f_t = \frac{\sum_{k=1}^{K} e^{-\eta L_{t-1}^k} x_t^k}{\sum_{k=1}^{K} e^{-\eta L_{t-1}^k}} \tag{2}$$

with $L_t^k := \sum_{s=1}^{t} \ell(x_s^k, y_s)$ denotes the *cumulated loss*, achieves the regret bound

$$\max_{1 \leqslant k \leqslant K} R_T^k \leqslant \frac{1}{\eta} \log K \tag{3}$$

for each $T \geqslant 1$. This means that the per-round regret R_T/T tends to 0 at the speed $O((\log K)/T)$.

Similarly to what happens in supervised learning, this fast rate cannot be guaranteed for a general loss function. From now on, we only assume that the loss function ℓ is convex and bounded; in this case, the minimax regret is of order $\sqrt{T \log K}$, meaning that the per-round regret is of order $\sqrt{(\log K)/T}$.

1.2 The Hedge setting

Strategies for prediction with expert advice often take the form of a convex combination of the predictions of the experts, with weights w_t^k that depend on the past performance of the experts:

$$f_t = \sum_{k=1}^{K} w_t^k x_t^k \tag{4}$$

where $\boldsymbol{w}_t = (w_t^k)_{1 \leqslant k \leqslant K}$ is a probability distribution on the experts. When the loss function ℓ is convex and bounded (say $0 \leqslant \ell \leqslant 1$), such strategies can be analysed in a simpler setting. Indeed, the convexity of ℓ implies that, when f_t is defined as in (4), we have whatever the value of $y_t \in \mathscr{Y}$ is:

$$\ell(f_t, y_t) \leqslant \sum_{k=1}^{K} w_t^k \, \ell(x_t^k, y_t) = \boldsymbol{w}_t \cdot \boldsymbol{\ell}_t \tag{5}$$

with $\boldsymbol{\ell}_t = (\ell_t^k)_{1 \leqslant k \leqslant K} = (\ell(x_t^k, y_t))_{1 \leqslant k \leqslant K} \in [0,1]^K$ the loss vector of the experts. To bound the regret with respect to expert k, it thus suffices to control $\sum_{t=1}^{T} (\boldsymbol{w}_t \cdot \boldsymbol{x}_t - \ell_t^k)$ for each k; this formulation depends neither on the spaces $\mathscr{X}$ and $\mathscr{Y}$, nor on the loss function ℓ, but only on the loss vectors $\boldsymbol{\ell}_t$. This leads to the following simpler formulation:

Problem 2 (Hedge)**.** Assume we are given a finite set $\{1, \dots, K\}$ of experts. At each time step $t \geqslant 1$, the following actions occur sequentially:

1. The learner chooses a probability distribution $\boldsymbol{w}_t = (w_t^k)_{1 \leqslant k \leqslant K}$ on the set of experts.
2. The environment then chooses the loss vector $\boldsymbol{\ell}_t \in [0,1]^K$.

The goal of the learner is to maintain a small *cumulative regret*

$$R_T^k := \sum_{t=1}^{T} r_t^k \; ; \quad r_t^k := \boldsymbol{w}_t \cdot \boldsymbol{\ell}_t - \ell_t^k = (\boldsymbol{w}_t - \boldsymbol{e}_k) \cdot \boldsymbol{\ell}_t \tag{6}$$

for each time horizon $T \geqslant 1$ and with respect to each expert $k = 1, \dots, K$, no matter what the losses $\boldsymbol{\ell}_t \in [0,1]^K$ are.

1.3 Minimax regret for the Hedge setting

We now briefly study the *minimax regret*, *i.e.* the best upper bound we can guarantee on $R_T := \max_{1 \leqslant k \leqslant K} R_T^k$ in the worst case scenario, and how to achieve it. As we announced before, this bound is of order $\sqrt{T \log K}$.

If T is known, we can attain this bound by using a properly tuned exponential weights algorithm, defined by (up to renormalisation) $w_t^k \propto \exp(-\eta L_{t-1}^k)$. The proof relies on the following inequality: for each k,

$$\boldsymbol{w}_t \cdot \boldsymbol{\ell}_t \leqslant -\frac{1}{\eta} \log \left(\sum_{k=1}^{K} w_t^k e^{-\eta \ell_t^k} \right) + \frac{\eta}{8} = \ell_t^k + \frac{1}{\eta} \log \frac{w_{t+1}^k}{w_t^k} + \frac{\eta}{8} \tag{7}$$

where the first inequality follows from a Hoeffding upper bound[1] on $\mathbb{E}\, e^{-\eta X}$, where the random variable X takes the value ℓ_t^k with probability w_t^k, and the equality is a consequence of the form of the exponential weights. Subtracting ℓ_t^k on both sides of (7), summing on $t = 1, \dots, T$, and noting that $w_1^k = \frac{1}{K}$ and $w_{T+1}^k \leqslant 1$, we get

$$R_T^k \leqslant \frac{1}{\eta} \log K + \frac{\eta T}{8} \tag{8}$$

which, knowing T, is minimized by taking $\eta = \sqrt{8(\log K)/T}$; this yields the regret bound $R_T \leqslant \sqrt{(T/2) \log K}$.

When T is unknown and we want a bound valid for every $T \geqslant 1$, we can use a variant of the exponential weights: $w_t^k \propto \exp(-\eta_t L_{t-1}^k)$ with time-varying rate $\eta_t = \sqrt{8(\log K)/t}$; this yields a similar bound $R_T = O(\sqrt{T \log K})$. Hence, we can achieve of regret bound of order $\sqrt{T \log K}$; it turns out that this bound cannot be improved (Cesa-Bianchi and Lugosi, 2006), in the sense that whatever the learner's strategy is, the environment can always ensure a regret of at least $C\sqrt{T \log K}$ for some universal $C > 0$.

2 Beyond minimax regret bounds: two kinds of adaptivity

We saw in the previous section that the minimax regret was of order $\sqrt{T \log K}$, and that a properly-tuned exponential weights algorithm (also called *Hedge algorithm*) enables to achieve this bound. At first sight, it would seem that the Hedge problem is essentially solved.

However, the Hedge algorithm turns out to be underwhelming in practice: often, we would be better off by using the simple *Follow-the-Leader* (FTL) algorithm, that just predicts at each step the same as the expert with the smallest cumulative loss (although it can be seen that the worst-case regret of this latter strategy is much worse, in fact *linear* in T). Indeed, the Hedge strategy tends to be overly conservative, since it tries to protect itself against a worst-case signal that may be generated by an adversary. As it turns out, the Hedge algorithm often *attains* a regret of order $\sqrt{T}$, even in "easy" cases when more naive strategies would perform better.

With this in mind, it is natural to seek strategies that, while retaining the worst-case guarantees of the Hedge algorithm, can adapt to the difficulty of the problem and learn more "aggressively" when the data permits it. Two notions of adaptivity, largely orthogonal, were considered in the literature. In this section, we review and motivate these notions.

[1] The Hoeffding inequality states that, for every random variable X with $0 \leqslant X \leqslant 1$ and every $\lambda \in \mathbf{R}$, $\log \mathbb{E} \exp(\lambda(X - \mathbb{E}\, X)) \leqslant \lambda^2/8$. Applying this with $\lambda = -\eta$ yields $\mathbb{E}\, X \leqslant -\frac{1}{\eta} \log \mathbb{E}\, e^{-\eta X} + \frac{\eta}{8}$ for $\eta > 0$.

2.1 Second-order regret bounds

A first line of improvement is to replace the term T in the bound $\sqrt{T \log K}$ by a data-dependent quantity that is always smaller than T, and can potentially be much smaller when the problem is "easier".

First-order bounds. A first improvement takes advantage of the fact that the best expert may predict the data well, and use higher learning rates if this occurs. An exponential weights algorithm with the data-dependent learning rate $\eta_t \approx \sqrt{(\log K)/L^\star_{t-1}}$, where $L^\star_t = \min_{1 \leqslant k \leqslant K} L^k_t$ is the loss of the best expert, leads to the regret bound (Cesa-Bianchi and Lugosi, 2006)

$$R_T = L_T - L^\star_T \lesssim \sqrt{L^\star_T \log K} + \log K \,. \tag{9}$$

Since $L^\star_T \leqslant T$, this bound is always lower than the minimax regret bound $\sqrt{T \log K}$; when one expert predicts the data well, *i.e.* $L^\star_T \ll T$, this *first-order bound* (that depends on the losses) becomes much tighter. In particular, in the extreme case when one experts predicts the data perfectly (*i.e.* $L^\star_T = 0$) or really well (in the sense that its cumulative loss $L^\star_T$ remains bounded), the first-order bound (9) ensures a constant regret, a significant improvement over the $\sqrt{T}$ minimax bound.

Second-order bounds. While first-order bounds are appealing when the best expert performs well, there are many other scenarios when the problem is in some sense "easy" but when the first-order methods fail to adapt to this. A typical case is when one expert consistently outperforms the others, but still has a non-trivial loss (e.g. linear in T); this occurs for example when the losses ℓ^k_t of each expert k are drawn from an i.i.d. distribution with mean $\overline{\ell}^k \in [0, 1]$.

Second-order bounds (Cesa-Bianchi et al., 2007; de Rooij et al., 2014; Gaillard et al., 2014; Wintenberger, 2017) are a further refinement of first-order bounds, that typically take the following form:

$$R_T \lesssim \sqrt{V_T \log K} + \log K \tag{10}$$

where V_T denotes some second-order quantity that depends on a "variance" of the losses. Different kinds of variance have been considered in the literature; of particular interest is the quantity $V_T = \sum_{t=1}^T v_t$, where v_t is the variance of the losses ℓ^k_t under the probability distribution $\boldsymbol{w}_t$ on the experts. Choosing $\eta_t \approx \sqrt{(\log K)/V_{t-1}}$ yields the regret bound (10); although the interpretation of this bound is less straightforward than the previous ones (note that V_T depends on the weights $\boldsymbol{w}_t$, and hence on the algorithm), it can be remarkably tight. Indeed, we have $V_T \lesssim L^\star_T \leqslant T$, so this bounds implies both the minimax guarantee and the first-order bound, but V_T can be much smaller than $L^\star_T$, especially when one expert consistently outperforms the others, so that the weights concentrate on this expert.

For instance, Gaillard et al. (2014) showed that second-order regret bounds implied a *constant* cumulative regret (the optimal rate), both in expectation and with high probability, when the losses of each expert over time are i.i.d.

2.2 Quantile regret bounds

A different kind of adaptivity to easy data is formalised by *quantile bounds*: while second-order bounds refine the T term in the minimax bound $\sqrt{T \log K}$, quantile bounds are concerned with the dependence on the total number of experts K. Intuitively, when multiple experts are performing well, we should be able to derive better regret bounds. Such a scenario may naturally occur, e.g. when the experts themselves correspond to learning algorithms that improve with more data, or when the finite set of experts corresponds to a discretization of a continuous model (in which case the number of good experts may be large if the model is over-discretized).

Quantile bounds. Quantile bounds (Chaudhuri et al., 2009; Chernov and Vovk, 2010; Luo and Schapire, 2015) deal with the aforementioned issue, by expressing the regret in terms of the *share* of good experts instead of the total number of experts. More precisely, let π be a probability measure on the set $\{1, \dots, K\}$ of experts, which we call a *prior*. A *quantile bound* is a regret bound of the form

$$\min_{k \in \mathcal{K}} R_T^k \lesssim \sqrt{T \log \frac{1}{\pi(\mathcal{K})}} \tag{11}$$

for *every* subset $\mathcal{K} \subset \{1, \dots, K\}$. Such a bound can be interpreted in the following way: for every $\varepsilon \in (0, 1)$, the regret with respect to the best ε-quantile of experts (ranked by their losses) is a $O(\sqrt{T \log \varepsilon^{-1}})$. Note that, when π is the uniform prior on the experts, the bound (11) implies $R_T \lesssim \sqrt{T \log K}$ by taking $\mathcal{K} = \{k\}$ for $k = 1, \dots, K$.

Relative entropy bounds. Let us also mention relative entropy bounds, a generalization of quantile bounds[2]. Such a bound takes the form

$$R_T^\rho \lesssim \sqrt{T \log \Delta(\rho \,\|\, \pi)} \tag{12}$$

for *every* distribution ρ on the experts, where we defined $R_T^\rho := \sum_{k=1}^K \rho(k) R_T^k$ and where $\Delta(\rho \,\|\, \pi) := \sum_{k=1}^K \rho(k) \log \frac{\rho(k)}{\pi(k)}$ denotes the *relative entropy* (also called *Kullback-Leibler divergence*) between ρ and π. The relative entropy bound (12) implies the quantile bound (11): to see this, consider for any subset $\mathcal{K}$ of experts the distribution $\rho(k) = \pi(k \,|\, \mathcal{K})$; the implication follows from the fact that $\Delta(\rho \,\|\, \pi) = -\log \pi(\mathcal{K})$ and the inequality $\min_{k \in \mathcal{K}} R_T^k \leqslant R_T^\rho$.

Quantile and relative entropy bounds are particularly appealing when the number of experts K is large; they can even be formulated for continuous sets of experts.

3 The Squint algorithm

3.1 Objective

We saw in the previous section two kinds of adaptivity; before Koolen and van Erven (2015), those two were studied separately, using very different and incompatible tech-

[2] Although in their paper Koolen and van Erven (2015) only considered quantile bounds, Wouter Koolen later issued a note http://blog.wouterkoolen.info/Squint_PAC/post.html in which he showed that the bounds directly extended to relative entropies.

niques. The purpose of the Squint algorithm is to combine the two through a unified analysis.

For every $k \in \{1, \dots, K\}$ and $T \geqslant 1$, define $V_T^k := \sum_{t=1}^T (r_t^k)^2 = \sum_{t=1}^T (\boldsymbol{w}_t \cdot \boldsymbol{\ell}_t - \ell_t^k)^2$; the aim is to achieve a *second-order relative entropy bound* of the form

$$R_T^\rho \lesssim \sqrt{V_T^\rho (C_{\mathrm{lr}} + \Delta(\rho \,\|\, \pi))} \tag{13}$$

for *every* distribution ρ, where we denote $V_T^\rho := \sum_{k=1}^K \rho(k) V_T^k$, and where C_{lr} is an overhead we incur that should be kept small.

3.2 The Squint potential

A general way to design online learning strategies and to control their regret is to use a *potential* $\Phi_T(\boldsymbol{r}_{1:T})$, an increasing function of the regrets that the weights are built to control. For example, the proof of the regret of the exponential weights algorithm sketched in section 1.3 amounts to show that the following potential

$$\Phi_T(\boldsymbol{r}_{1:T}) := \mathop{\mathbb{E}}_{\pi(k)} e^{\eta R_T^k - \eta^2 T/8} \tag{14}$$

is decreasing when we choose the weights $w_t^k \propto \pi(k) e^{-\eta L_{t-1}^k}$.

In order to obtain a second-order bound, the potential has to depend not only on the regrets R_T^k, but also on the cumulated variances V_T^k; moreover, in order to adapt to the quantities V_T^ρ and $\Delta(\rho \,\|\, \pi)$ for every ρ simultaneously, one has to find a way to automatically tune the learning rate η. With this in mind, the Squint potential is defined as

$$\Phi_T = \Phi_T(\boldsymbol{r}_{1:T}) := \mathop{\mathbb{E}}_{\pi(k)\gamma(\eta)} \left[e^{\eta R_T^k - \eta^2 V_T^k} - 1 \right] \tag{15}$$

where $\gamma(\eta) d\eta$ is a *prior* distribution put on learning rates $\eta \in [0, \frac{1}{2}]$ to be specified later (see section 4). The weights $\boldsymbol{w}_T$ chosen by the Squint algorithm are designed to control the potential and taken to be

$$\boldsymbol{w}_{T+1} := \frac{\mathbb{E}_{\pi(k)\gamma(\eta)} \left[e^{\eta R_T^k - \eta^2 V_T^k} \eta \, \boldsymbol{e}_k \right]}{\mathbb{E}_{\pi(k)\gamma(\eta)} \left[e^{\eta R_T^k - \eta^2 V_T^k} \eta \right]} \quad i.e. \quad w_{T+1}^k \propto \pi(k) \int_0^{1/2} e^{\eta R_T^k - \eta^2 V_T^k} \eta \, \gamma(\eta) d\eta \,. \tag{16}$$

Lemma 1. *For choice of weights of equation (16), we have* $0 = \Phi_0 \geqslant \ldots \geqslant \Phi_{T-1} \geqslant \Phi_T$.

Proof. We use the following elementary inequality: for every $x \geqslant -\frac{1}{2}$, $e^{x - x^2} \leqslant 1 + x$. Applying this to $\eta \, r_{T+1}^k = \eta \, (\boldsymbol{w}_{T+1} - \boldsymbol{e}_k) \cdot \boldsymbol{\ell}_{T+1}$ for $\eta \leqslant \frac{1}{2}$, we get

$$\begin{aligned}
\Phi_{T+1} - \Phi_T &= \mathop{\mathbb{E}}_{\pi(k)\gamma(\eta)} \left[e^{\eta R_T^k - \eta^2 V_T^k} \left(e^{\eta r_{T+1}^k - \eta^2 (r_{T+1}^k)^2} - 1 \right) \right] \\
&\leqslant \mathop{\mathbb{E}}_{\pi(k)\gamma(\eta)} \left[e^{\eta R_T^k - \eta^2 V_T^k} \eta \, (\boldsymbol{w}_{T+1} - \boldsymbol{e}_k) \cdot \boldsymbol{\ell}_{T+1} \right] \\
&= 0 \,,
\end{aligned}$$

the last inequality being a consequence of the definition (16) of $\boldsymbol{w}_{T+1}$. □

To see intuitively how the inequality $\Phi_T \leqslant 0$ implies a second-order relative entropy bound, assume for simplicity that the prior γ puts some positive mass $\gamma(\eta) > 0$ on a learning $\eta > 0$ to be specified later.

Let ρ be an arbitrary probability distribution on the experts. We will use the following *change of measure inequality*: for every function $h : \{1, \dots, K\} \to \mathbf{R}$, we have

$$\log \mathop{\mathbb{E}}_{\pi(k)} e^{h(k)} \geqslant -\Delta(\rho \,\|\, \pi) + \mathop{\mathbb{E}}_{\rho(k)} h(k)\,. \tag{17}$$

Indeed, writing $\pi(k) \geqslant \rho(k)(\frac{\rho(k)}{\pi(k)})^{-1}$ (with equality when $\rho(k) > 0$) and using the concavity of the log function, we have

$$\log \mathop{\mathbb{E}}_{\pi(k)} e^{h(k)} \geqslant \log \mathop{\mathbb{E}}_{\rho(k)} \left[\left(\frac{\rho(k)}{\pi(k)}\right)^{-1} e^{h(k)}\right] \geqslant \mathop{\mathbb{E}}_{\rho(k)} \left[-\log \frac{\rho(k)}{\pi(k)} + h(k)\right] = -\Delta(\rho \,\|\, \pi) + \mathop{\mathbb{E}}_{\rho(k)} h(k)\,.$$

Now write, using the inequality $0 \geqslant \Phi_T$ and applying the change of measure inequality (17) to the function $h(k) = \eta R_T^k - \eta^2 V_T^k$,

$$\begin{aligned}
1 &\geqslant \mathop{\mathbb{E}}_{\pi(k)\gamma(\eta)} \left[e^{\eta R_T^k - \eta^2 V_T^k}\right] \\
&\geqslant \gamma(\eta) \mathop{\mathbb{E}}_{\pi(k)} \left[e^{\eta R_T^k - \eta^2 V_T^k}\right] \\
&\geqslant \gamma(\eta) \exp\left(-\Delta(\rho \,\|\, \pi) + \mathop{\mathbb{E}}_{\rho(k)} \eta R_T^k - \eta^2 V_T^k\right) \\
&= \gamma(\eta) \exp\left(-\Delta(\rho \,\|\, \pi) + \eta R_T^\rho - \eta^2 V_T^\rho\right)
\end{aligned}$$

since we assumed η was arbitrary (*i.e.* that the γ puts mass on the optimal η), we can take it to be the value $\widehat{\eta} = \frac{R_T^\rho}{2V_T^\rho}$ that maximizes the quantity $\eta R_T^\rho - \eta^2 V_T^\rho$. The above bound then becomes

$$1 \geqslant \gamma(\widehat{\eta}) \exp\left(-\Delta(\rho \,\|\, \pi) + \frac{(R_T^\rho)^2}{2V_T^\rho}\right) \tag{18}$$

which can be rearranged to give:

$$R_T^\rho \leqslant \sqrt{2V_T^\rho \left(\log \frac{1}{\gamma(\widehat{\eta})} + \Delta(\rho \,\|\, \pi)\right)}\,. \tag{19}$$

Equation (19) is a second-order relative entropy regret bound, with an overhead C_{lr} that depends on the mass put on the optimal learning rate. Of course, the imprecise step in our derivation was when we assumed that the optimal η had a positive weight. In practice, the priors we will consider (see section 4) have a density, and to get a rigorous bound we write (similarly to what we did before)

$$1 \geqslant \exp(-\Delta(\rho \,\|\, \pi)) \int_0^{1/2} e^{\eta R_T^\rho - \eta^2 V_T^\rho} \gamma(\eta) d\eta\,; \tag{20}$$

then, one has to show that the prior $\gamma(\eta)d\eta$ puts a sufficient mass *around* the optimal $\widehat{\eta}$.

4 Choice of the prior

In section 3, we have described the Squint potential, algorithm and given a sketch of proof of its regret bound. In order to determine the algorithm, it remains to specify the prior $\gamma(\eta)d\eta$ on $\eta \in [0, \frac{1}{2}]$ so that:

1. The weights given by equation (16) have a closed-form expression, so that they can be computed explicitly.
2. The prior puts enough weight around each η, so that the overhead C_{lr} is small.

4.1 Uniform prior

A simple choice of prior $\gamma(\eta)d\eta$ is the uniform distribution $\gamma(\eta) = \frac{1}{2}$. The weights (16) then have a closed-form expression in terms of the erf function; moreover, the overhead we incur in the regret bound is $C_{\mathrm{lr}} = \log V_T^\rho$. While this may seem acceptable, for large V_T^ρ this term is of order $\log T$, which as it turns out can be improved.

4.2 A near-optimal prior

To get a tighter regret bound, one has to put more mass to the small learning rates η. A nearly optimal prior that gives more mass around 0 while still being integrable is $\gamma(\eta) \propto \frac{1}{\eta(\log \eta)^2}$. For this prior, the incurred overhead is $C_{\mathrm{lr}} = \log\log V_T^\rho$ which is almost constant; unfortunately, the weights of equation (16) no longer have a closed form expression.

4.3 An improper prior

It turns out that one can combine the closed-form expression for the coefficients $\boldsymbol{w}_T$ with a small, doubly-logarithmic overhead C_{lr}, by choosing the improper prior $\gamma(\eta) = \frac{1}{\eta}$. For this prior, one has to be much more cautious in the derivations, since for example the function $\frac{1}{\eta}e^{\eta R_T^k - \eta^2 V_T^k}$ is not integrable.

However, the potential (15) is still defined provided we keep the "-1" term *inside* the integral. Additionally, the weights of equation (16) are not only well-defined (since $\eta\,\gamma(\eta) = 1$ cancel out in the integral), but they also admit a closed form expression: by recognizing the integral of a Gaussian density, we see that

$$w_{T+1}^k \propto \pi(k)\int_0^{1/2} e^{\eta R_T^k - \eta^2 V_T^k} d\eta = \pi(k)\frac{\sqrt{\pi}e^{\frac{(R_T^k)^2}{4V_T^k}}\left(\operatorname{erf}\left(\frac{R_T^k}{2\sqrt{V_T^k}}\right) - \operatorname{erf}\left(\frac{R_T^k - V_T^k}{2\sqrt{V_T^k}}\right)\right)}{2\sqrt{V_T^k}} \tag{21}$$

Moreover, this choice leads to the small overhead $C_{\mathrm{lr}} = \log\log T$.

References

Cesa-Bianchi, N. and Lugosi, G. (2006). *Prediction, Learning, and Games*. Cambridge University Press, Cambridge, New York, USA.

Cesa-Bianchi, N., Mansour, Y., and Stoltz, G. (2007). Improved second-order bounds for prediction with expert advice. *Machine Learning*, 66:321–352.

Chaudhuri, K., Freund, Y., and Hsu, D. J. (2009). A parameter-free hedging algorithm. In Bengio, Y., Schuurmans, D., Lafferty, J. D., Williams, C. K. I., and Culotta, A., editors, *Advances in Neural Information Processing Systems 22*, pages 297–305. Curran Associates, Inc.

Chernov, A. and Vovk, V. (2010). Prediction with advice of unknown number of experts. In *Proceedings of the Twenty-Sixth Annual Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 117–125, Corvallis, Oregon. AUAI Press.

de Rooij, S., van Erven, T., Grünwald, P., and Koolen, W. M. (2014). Follow the leader if you can, hedge if you must. *Journal of Machine Learning Research*, 15:1281–1316.

Freund, Y. and Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119–139.

Gaillard, P., Stoltz, G., and van Erven, T. (2014). A second-order bound with excess losses. In *Proceedings of the 27th Annual Conference on Learning Theory (COLT)*, pages 176–196.

Koolen, W. M. and van Erven, T. (2015). Second-order quantile methods for experts and combinatorial games. In *Proceedings of the 28th Annual Conference on Learning Theory (COLT)*, pages 1155–75.

Luo, H. and Schapire, R. E. (2015). Achieving all with no parameters: AdaNormalHedge. In *Proceedings of the 28th Annual Conference on Learning Theory (COLT)*, pages 1286–1304.

Wintenberger, O. (2017). Optimal learning with Bernstein online aggregation. *Machine Learning*, 106(1):119–141.