One Coin to Rule Them All

From Betting to Online Learning to Automatic Model Selection with One Algorithm

Francesco Orabona
francesco@orabona.com

David Pal
dpal@yahoo-inc.com

Yahoo Labs
New York, NY, USA

November 17, 2015

Abstract

Online Linear Optimization (OLO) is a basic building block for optimization and machine learning problems. It allows to reduce the design of learning algorithms to simple online algorithms for linear losses. In particular, Online Convex Optimization (OCO), stochastic optimization, batch optimization, and (convex) machine learning algorithms can be solved with simple instantiations of OLO algorithms.

Yet, some critical issues are currently mostly ignored by theoretical research. In particular, the adaptation to the norm of the optimal solution is not taken into account in first-order optimization algorithms and the model selection phase is often ignored in theoretical algorithms for machine learning. In fact, in theoretical works most of the time assumptions are made, for example, on the prior knowledge of the norm of the optimal solution, while in the practical world validation methods remain the only viable approach.

In this paper, we propose to use coin betting as a starting principle, instead of OLO. We will prove that an optimal betting algorithm can be used to have optimal guarantees in OLO, OCO and regularized ERM, proving novel connections between these areas.

1 Introduction

Online Convex Optimization (OCO) is a problem where an algorithm repeatedly chooses a point w_t from a convex decision set K, observes an arbitrary, or even adversarially chosen, convex loss function ℓ_t and suffers loss $\ell_t(w_t)$. The goal of the algorithm is to have a small cumulative loss. Performance of an algorithm is evaluated by the so-called regret, which is the difference of cumulative losses of the algorithm and of the (hypothetical) strategy that would choose in every round the same best point in hindsight. Typically, one tries to prove that the regret grows at most sub-linearly in time, that is, the average regret vanishes over time.

Typically, OCO is solved with a reduction of a Online Linear Optimization (OLO) problem [Cesa-Bianchi and Lugosi, 2006, Shalev-Shwartz, 2012], where the losses $\ell_t(w)$ are the linear functions $\langle w, g_t \rangle$. Indeed, many learning problems can be directly phrased as OLO, e.g., learning with expert advice [Littlestone and Warmuth, 1994, Vovk, 1998, Cesa-Bianchi et al., 1997], online combinatorial optimization Koolen et al. [2010]. A part from OCO, other problems can be also reduced to OLO, e.g. online classification and regression [Cesa-Bianchi and Lugosi, 2006, Chapters 11 and 12], multi-armed problems [Cesa-Bianchi and Lugosi, 2006, Chapter 6], and batch and stochastic optimization of convex functions Nemirovski and Yudin [1983]. Hence, a result in OLO immediately implies other results in all these domains.

However, as essential as it is, being able to solve an optimization problem is only half of the problem in machine learning. In fact, we are often interested in the generalization performance of a predictor trained from data. This depends not only on the ability to optimize a loss but also on constraining the complexity of the predictor. This is usually achieved through the use an regularized objective function, that biases the solution towards a small region of

the space. However, choosing the size of this region is usually a problem by itself, that is solved by practitioners with a variety of more or less theoretical principled tools.

In this paper, we claim that a more fundamental notion subsumes both OLO and OCO. This notion is linked to the ability of an algorithm to optimally bet on an arbitrary sequence of outcomes from a coin. We will define the notion of Master Betting Algorithm (MBA) that not only will allow us to solve OLO and OCO, but to design algorithm that do not require any explicit form of regularization nor any hyper-parameter to tune, yet are able to achieve optimal worst case guarantees. In particular, we will show that an algorithm that guarantees a reward close to the optimal sequence of bets also guarantees optimal regret in OLO/OCO and automatic model selection in regularized Empirical Risk Minimization (ERM).

We will also show connections between the optimal the betting strategy known in economics as Kelly betting [Kelly, 1956] and online learning, and hence indirectly with stochastic optimization and statistical learning.

2 Setting and Notation

These papers deals with three different settings, hence we will fist describe the common notation and then each of one them separately, trying to show the similarities between them.

Random variables will be indicated by capital italic letters, e.g. $X \sim N(0, \sigma^2)$. A vector x is a subgradient of a convex function ℓ at v iff $\ell(u) - \ell(v) \geq \langle u - v, x \rangle$ for any u in the domain of ℓ. The differential set of ℓ at v, denoted by $\partial \ell(v)$, is the set of all the subgradients of ℓ at v. We define the KL divergence between two Bernoulli distributions with parameters p and q as

$$D(p||q) := p \log \frac{p}{q} + (1-p) \log \frac{1-p}{1-q} .$$

We will denote by $\mathcal{H}$ an Hilbert space with inner product $\langle \cdot, \cdot \rangle$.

Binary and Continuous Coin Betting. The bettor starts with the amount of money ϵ. At the end of each bet in the time step t we denote the amount of money he has by $Wealth_t$ and by $Reward_t$ the amount of money gained through the t bets. Hence, $Wealth_t = Reward_t + \epsilon$. At each time step t, it has to bet a quantity of money w_t equal to a fraction of his current money, $w_t := b_t \, Wealth_{t-1}$ where $b_t \in (-1, 1)$. Note that here we consider only betting strategies that will never result in a negative quantity of money owned. Hence, $|b_t|$ must be strictly less than 1, otherwise the algorithm could lose all the money in one round. After the bet, the outcome of the coin $g_t \in \{-1, 1\}$ is revealed and the bettor wins (or loses) the quantity $w_t g_t$. Note that we will speak about a "binary coin" when the outcome of the coin is in $\{-1, 1\}$, and "continuous coin" when $g_t \in [-1, 1]$. The latter case model the situation in there are different possible prizes with unknown probabilities to be won. In both cases, we have $Wealth_t := Reward_{t-1} + w_t g_t + \epsilon = \epsilon + \sum_{i=1}^{t} w_i g_i$, so that $Reward_0 = 0$ and $Wealth_0 = \epsilon$. We will extend to above definitions also to the the case that w_t and g_t belongs to a Hilbert space $\mathcal{H}$ and $\|g_t\| \leq 1$. Hence, with a slight abuse of notation we will also define $Reward_n = \sum_{t=1}^{n} \langle w_t, g_t \rangle$, $Wealth_t = Reward_t + \epsilon$ and $w_t = \beta_t \, Wealth_{t-1}$, with $\|\beta_t\| < 1$.

OLO and OCO. Let $\mathcal{H}$ be a Hilbert space. In the OLO framework, at each round t the algorithm receives a vector $g_t \in \mathcal{X}$, picks a $w_t \in \mathcal{S} \subseteq \mathcal{H}$, and gains $\langle w_t, g_t \rangle$ (or equivalently loses -$\langle w_t, g_t \rangle$). The aim of the algorithm is to minimize the *regret*, that is the difference between the cumulative gains of the of an arbitrary and fixed competitor $u \in \mathcal{X}$, $\sum_{t=1}^{n} \langle u, g_t \rangle$, and the cumulative gains of the algorithm, $\sum_{t=1}^{n} \langle w_t, g_t \rangle$. In particular, define

$$Regret_n(u) := \sum_{t=1}^{n} \langle g_t, u - w_t \rangle .$$

In the OCO setting, instead, the algorithm receives convex loss functions ℓ_t rather than vectors. Again, the aim is to minimize the regret, that in this case is defined as

$$Regret_n^{OCO}(u) := \sum_{t=1}^{n} \left(\ell_t(w_t) - \ell_t(u) \right) .$$

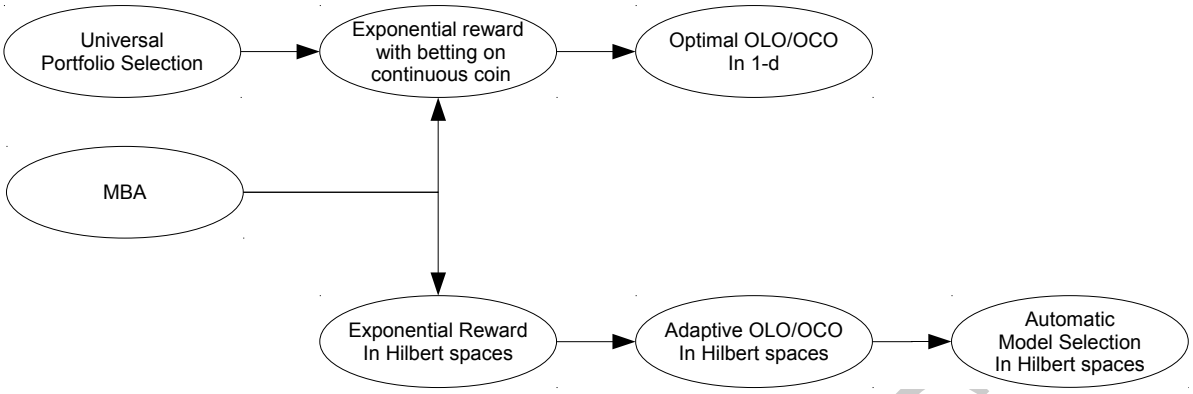

Figure 1: The links between the different areas and betting.

Regularized ERM. Let P a fixed but unknown distribution on $\mathcal{Z}$, where $\mathcal{Z}$ is an arbitrary set. We introduce a loss function $\ell : \mathcal{H} \times \mathcal{Z} \to \mathbb{R}_+$, convex and *L-Lipschitz* w.r.t. the first argument. Note that this generalizes the case where we have the composition of a predictor parametrized by a vector w and a loss function. Hence, for example, in the case of logistic regression define $\mathcal{Z} = \mathcal{H} \times [0, 1]$ and $z = (x, y)$ so that $\ell(w, z) = \ln(1 + \exp(-y\langle w, x\rangle))$. Define the *$\ell$-risk* of $w \in \mathcal{H}$, as $\mathcal{E}^\ell(w) := \mathbb{E}_{Z\sim P}[\ell(w, Z)]$. Given a training set $\{z_t\}_{t=1}^n$ of samples drawn Independent and Identically Distributed (IID) from P, the regularized ERM strategy finds a predictor $\hat{w}$ in a subset $\mathcal{S} \subseteq \mathcal{H}$, such that

$$\hat{f} = \arg\min_{w\in\mathcal{S}} \lambda R(w) + \frac{1}{n}\sum_{t=1}^{n} \ell(w, z_t),$$

where $R(w)$ is the regularizer.

3 Betting, OLO, OCO, Stochastic Optimization, and Model Selection

In this section we will show the connection between betting on a continuous coin, adaptive OLO/OCO, and automatic model selection. We will prove that, given an "optimal" betting algorithm, the same algorithm can be used to solve all the problems listed above.

We will assume that a 1-dimensional algorithm exists that satisfies the following assumption.

Assumption 1. *Assume that there exists a sequence of functions $f : \mathbb{R} \times 2^{\mathbb{R}} \to \mathbb{R}$ convex and twice differentiable in the first argument and an algorithm, that we will denote by MBA, that generates $b_t \in [-1, 1]$, using in input $x_{t-1} \in \mathbb{R}$, $z_1, \dots, z_{t-1} \in [-1, 1]$, such that*

- *f is even in the first argument*
- *$f''(x, S) > \frac{f'(x,S)}{x}$, where the derivatives are w.r.t. the first argument.*
- *$f(0, \{\}) = \epsilon > 0$;*
- $$(1 + b_t z_t) f\left(x_{t-1}, \{|z_1|, \dots, |z_{t-1}|\}\right) \geq f\left(x_{t-1} + z_t, \{|z_1|, \dots, |z_t|\}\right), \quad \forall z_t \in [-1, 1]\,. \tag{1}$$

By induction is easy to prove that such algorithm can indeed be used as a betting algorithm.

Theorem 1. *Assume that Assumption 1 holds and use the MBA with $x_{t-1} = |\sum_{i=1}^{t-1} g_i|$ and $z_t = g_t$. Then, for any sequence of outcomes $g_t \in [-1, 1]$, the following holds*

$$Wealth_n = \epsilon + \sum_{t=1}^{n} g_t w_t \geq f\left(\sum_{i=1}^{t-1} g_i, \{|z_1|, \dots, |z_n|\}\right)\,.$$

Even more importantly, the MBA can also be used to prove lower bound on the reward in Hilbert Spaces.

Theorem 2. *Assume that 1 holds. Let $g_t \in \mathcal{H}$ an arbitrary sequence of vector, such that $\|g_t\| \leq 1$ and define the vector $\theta_t = \sum_{i=1}^{t} g_i$. Use the Algorithm in 1 with the sequence $x_{t-1} = \|\theta_{t-1}\|$. Define a vectorial algorithm that at each step outputs $w_t = b_t \frac{\theta_{t-1}}{\|\theta_{t-1}\|} Wealth_{t-1}$. Then the following holds*

$$Wealth_n = \epsilon + \sum_{t=1}^{n} \langle g_t, w_t \rangle \geq f\left(\left\| \sum_{t=1}^{n} g_t \right\|, \{\|g_1\|, \dots, \|g_n\|\} \right) .$$

This theorem is useful to prove regret bounds in Hilbert spaces, as shown in the next part.

From Reward to Regret. The reward and regret view on online learning are equivalent: an algorithm guarantees low regret iff it guarantees high reward. The following Theorem makes this claim rigorous. Notice that the algorithm is exactly the same in the two setting.

Theorem 3 (McMahan and Orabona [2014])**.** *Let $\Psi : \mathcal{H} \to (-\infty, +\infty]$ be a lower semicontinuos and convex function, with $dom\Psi \neq \emptyset$. An algorithm for the player guarantees*

$$Reward_n \geq \Psi\left(\sum_{t=1}^{n} g_t \right) - \epsilon \qquad \text{for any } g_1, \dots, g_n$$

for a constant $\epsilon \in \mathbb{R}$ if and only if it guarantees

$$Regret_n(u) \leq \Psi^*(u) + \epsilon \qquad \text{for all } u \in \mathcal{H} \text{ and } g_t \in \mathcal{H}, \forall 1 \leq t \leq n . \tag{2}$$

So, a betting algorithm can be used for online learning and vice-versa. However, as it was already stressed in McMahan and Orabona [2014], the reward view has the big advantage of having one variable less, the competitor u. Moreover, as we will show in Section 5, designing and analyzing algorithms in one of the two views could be much easier than in the other one. Coupling Theorem 3 with Theorem 2, we have the following Corollary.

Corollary 1. *Assume that 1 holds. Then there exists a reduction of the MBA that produces a sequence of vectors w_t that satisfy*

$$Regret_n(u) \leq f^*\left(\|u\|, \{\|g_1\|, \dots, \|g_n\|\}\right) + \epsilon \qquad \text{for all } u \in \mathcal{H} \text{ and } g_t \in \mathcal{H}, \forall 1 \leq t \leq n,$$

where the conjugation is taken w.r.t. the first argument of f.

OLO, OCO and Stochastic Convex Optimization. From the property of the sub-gradient of a convex function we have that, for any sequence of convex functions $\ell_1, \dots, \ell_n$, vectors $w_1, \dots, w_n$, and $g_t \in \partial \ell_t(w_t)$, we have

$$Regret_n^{OCO}(u) = \sum_{t=1}^{n} \left(\ell_t(w_t) - \ell_t(u)\right) \leq \sum_{t=1}^{n} \langle w_t - u, g_t \rangle = Regret_n(u) .$$

Hence, the regret w.r.t. to arbitrary convex functions is upper bounded by the linear regret. This means that an OLO algorithm can be used to solve an OCO problem, just feeding the algorithm with loss vectors g_t equal to the subgradients of the functions ℓ_t.

Also, a regret bound can be transformed into a convergence guarantee for optimization of convex functions. In particular, we have the following Theorem from Cesa-Bianchi et al. [2004].

Theorem 4. *Let $w_1, \cdots, w_n \in \mathcal{H}$ the vector produced by an OLO algorithm with a regret guarantee $Regret_n(u)$. Let $\ell : \mathcal{H} \to \mathbb{R}$ a convex function, and fix the vectors g_t to be unbiased estimate of the gradient of ℓ in w_t. Then, the following holds*

$$\mathbb{E}\left[\ell\left(\frac{1}{n} \sum_{t=1}^{n} w_t \right) \right] \leq \min_u \ \ell(u) + \frac{\mathbb{E}[Regret_n(u)]}{n} .$$

High probability bounds can be also easily obtained, assuming more on the function ℓ [Cesa-Bianchi et al., 2004]. The above theorem says that, if the regret grows, for example, as $\mathcal{O}(\sqrt{n})$, the OLO algorithm can be used as a stochastic optimization algorithm with convergence in expectation $\mathcal{O}(\frac{1}{\sqrt{n}})$.

Adaptive Algorithms for OLO and Self-tuning Model Selection. In learning theory a key concept is the one of regularization. If the concept class we are learning is too rich, we need to constrain the complexity of the trained predictor. A regularizer is achieving this biasing the classifier towards a small region of the space. However, it is known that the optimal amount of regularization is completely problem-dependent. Hence, the regularizer becomes another parameter to be learned.

Most, if not all, the machine learning algorithms uses a two-stages process to find the optimal amount of regularization. First, the algorithm is trained with a fixed regularization parameter. Second, its generalization performance is estimated together with a change in the regularization parameter. These two steps are repeated till convergence.

Surprisingly enough, [Orabona, 2014] proved that, when the regularizer is the squared norm of the Hilbert space, the above lengthy procedure can be avoided with a stochastic learner. In particular, instead of solving a series of regularized ERM problems, with different amounts of regularization, one can use a simple parameter-free stochastic gradient descent procedure over the training samples and achieve the same performance. More rigorously, the following theorem holds.

Theorem 5. *Assume that there is an online algorithm whose $Regret_n(u)$ is $\mathcal{O}\left(\|u\| \sqrt{n}(\ln(n))^{\alpha}\right)$. Then the following holds*

$$\mathbb{E}\left[\mathcal{E}^{\ell}\left(\frac{1}{n}\sum_{t=1}^{n} w_t\right)\right] \leq \mathcal{O}\left(\inf_{u} \min_{\lambda>0} \mathcal{E}^{\ell}(u) + \lambda \|u\|^2 + \frac{(\ln(n))^{2\alpha}}{\lambda n}\right) . \tag{3}$$

To better appreciate this bound, it is useful to compare it with a generalization bound we get from a regularized ERM solution, e.g. the one in Sridharan et al. [2009]. There, they prove that, with probability at least $1-\delta$, we have

$$\mathcal{E}^{\ell}(\hat{w}) \leq \mathcal{O}\left(\inf_{u} \mathcal{E}^{\ell}(u) + \lambda \|u\|^2 + \frac{\ln \frac{1}{\delta}}{\lambda n}\right), \tag{4}$$

where

$$\hat{w} = \arg\min_{w} \frac{\lambda}{2}\|w\|^2 + \sum_{t=1}^{n} \ell(w_t, z_t) .$$

Ignoring the fact that the bound in Theorem 5 is only in expectation, the bound in (4) is missing the minimum over λ, that is present in (3). Hence, to obtain the optimal performance in the regularized ERM setting we have to use different values of λ, while in Theorem 5 the tuning is automatic. This is particular important in infinite dimensional Hilbert spaces, where the infimum could not be attained by a vector in the space Orabona [2014].

The only missing piece is to prove that the existence of the MBA guarantees to have an algorithm that satisfies the regret guarantee in the hypothesis of Theorem 5. Indeed, we have the following Theorem.

Theorem 6. *Assume that 1 holds and there exist constants $\alpha, \beta, \gamma, \delta \geq 0$ independent from t such that the function $f(x, \{z_1, \dots, z_t\}) \geq \beta t^{-\gamma} \exp(\frac{x^2}{\alpha t}) - \delta$. Then there exists an algorithm, that uses the MBA as a subroutine, that guarantees $Regret_n(u) = \mathcal{O}\left(\|u\| \sqrt{n \ln(n\|u\| + 1)}\right)$.*

To summarize, the existence of a simple 1-d algorithm guarantees, though a straightforward reduction, the possibility of learning in Hilbert spaces, in the online setting and stochastic setting. Moreover, if the algorithm guarantees an exponential growth of the wealth, a reduction allows us to use it to obtain optimal rates of convergence to the best risk in a function class, without having to choose the regularization parameter. In the next section we will explain what are the difficulties in designing such 1-d algorithm and in Section 5 we will present our solution.

4 Kelly Betting, Mixture Forecasters and Portfolio Selection

Before trying to design the MBA, we will explore what are the theoretical limits that such an algorithm should obey. These limits will guide us in the design of the algorithm.

Betting on a binary coin. Given the link with betting and MBA, exploring the limits of the reward obtainable with betting will imply limits on the MBA. We will consider the case that the outcomes are binary, i.e. $g_t \in \{-1, 1\}$. First, we will analyze the class of betting strategy that bets a fixed amount of the current reward for the entire game. The following theorem is well-known and it has been shown, for example, in McMahan and Abernethy [2013].

Theorem 7. *Let an algorithm bet a fixed fraction of his current reward, i.e.* $w_t = b\, Wealth_{t-1}$*, where* $0 \leq b \leq 1$*. Then, for any sequence* $g_1, \ldots, g_n \in \{-1, 1\}$*, we have*

$$\max_b \; Wealth_n = \epsilon \exp\left(n\, D\left(\frac{1}{2} + \frac{\sum_{t=1}^n g_t}{2n} \middle\| \frac{1}{2}\right)\right) . \tag{5}$$

Moreover the optimal b *is* $\frac{\sum_{t=1}^n g_t}{n}$.

The above theorem tells us that any algorithms that keeps the fraction of money to bet at each round fixed cannot gain more that the quantity on the r.h.s. of (5). Moreover, the optimal fraction is proportional to the empirical estimate of the probability that $g_t = 1$.

One might ask if having a fraction of money that changes over time might help and the answer is substantially negative. The next Theorem shows that we can only hope for logarithmic gain in the exponent w.r.t. (5).

Theorem 8. *For* $n \in \mathbb{N}$ *even and any betting strategy* w_t *with an initial amount of money equal to* ϵ*, there exists a sequence of* $g_t \in \{-1, 1\}$ *such that*

$$Wealth_n \leq \epsilon \min\left\{\left(\exp\left(\frac{1}{6}\right)\sqrt{2\pi}\frac{|\sum_{t=1}^n g_t|}{\sqrt{n}} + 2\exp\left(\frac{1}{6}\right) - 1\right)\exp\left(n\, D\left(\frac{1}{2} + \frac{|\sum_{t=1}^n g_t|}{2n} \middle\| \frac{1}{2}\right)\right), 2^n\right\}.$$

These last two theorems suggest that we should aim at obtaining an exponential gain, up to logarithmic terms in the exponent.

Notice that, if we knew that the setting is stochastic, the optimal fraction of money in Theorem 5 is nothing else than the Kelly criterion, that is derived with the objective of maximizing the expectation of the logarithm of the reward, rather than the expected profit from each bet [Kelly, 1956]. In the latter case, one would be led to gamble all he had when presented with a favorable bet, and, in the case of a lost, one would have no capital with which to place subsequent bets. In most gambling scenarios, the Kelly strategy will do better than any essentially different strategy in the long run.

Practically speaking, w.l.o.g. assume that the probability of tail is p, $p > \frac{1}{2}$. The Kelly criterion applied to this coin and with equal wins, corresponds in betting on that tail on each round a fraction of the current money equal to $2p - 1$. This simple rule assures that in expectation the growth rate will be exponential.

However, the Kelly criterion can be used only if the coin is stochastic and if we knew the bias of the coin exactly. One might argue that this is equally unlikely as knowing beforehand the empirical frequencies of 1 and -1 in the sequence. One might ask what is the optimal strategy in the case that the sequence of outcomes of the coin is non-stochastic. A little known result by Krichevsky and Trofimov Krichevsky and Trofimov [1981], says that above procedure can still be used, substituting the unknown probability with a slightly biased running estimate. In particular, the fraction to bet is set to

$$\hat{b}_t = \frac{|\sum_{i=1}^{t-1} g_t|}{t},$$

and we bet on the faces that appeared more in the past. In particular the following Theorem holds for the Krichevsky-Trofimov (KT) algorithm.

Theorem 9. *Bet at each round a quantity equal to* $\hat{b}_t\, Wealth_{t-1}$ *on the face that appeared more in the past. Then*

$$Wealth_n \geq \epsilon \exp\left(n\, D\left(\frac{1}{2} + \frac{\sum_{t=1}^n g_t}{2n} \middle\| \frac{1}{2}\right) - \frac{1}{2}\ln n - \ln 2\right) .$$

That is, this simple procedure guarantees an exponential reward as well, only a factor $\sqrt{n}$ less than having known in advance the total number of heads in the sequence of n rounds. It is also known that this factor cannot be improved.

Betting on a continuous coin. Hence, the problem of betting on a binary coin is solved. However, the following problem cannot be solved with the Krichevsky-Trofimov forecaster. Consider the same betting scenario as before, with the only difference that the outcome of the coin is now a real number between $+1$ and -1. We can interpret this as a betting scenario in which the maximum amount of money that can be won or lost in each bet is fixed, but revealed to the bettor only after the bet is done. The formalism is still the same, because $w_t g_t$ is still the amount of money won. However, the simple change makes this problem much harder than before. Indeed, to solve it we now have to use a Portfolio Selection algorithm. In fact, it is possible to consider the following equivalent problem. On each time step, we have to divide our wealth between two stocks. The gains given by the market are coded in the vector m_t that are equal to $[1+g_t, 1-g_t]$. The algorithm will return a division of the wealth of the form $[a_t, 1-a_t]$. Then this can be used to bet on the continuous coin using $\beta_t = 2a_t - 1$. It is easy to see that with this reduction the wealth of the portfolio selection is equal to the reward on the continuous coin, that is

$$Wealth_{t-1}\left((1+g_t)a_t + (1-g_t)(1-a_t)\right) = Wealth_{t-1} + Wealth_{t-1}\beta_t g_t \ .$$

However, this reduction will not solve our problem. The reasons are that, even using the optimal algorithm for portfolio selection Cover and Ordentlich [1996], we do not have an explicit lower bound on the reward. In fact, the Universal Portfolio algorithm Cover and Ordentlich [1996] only assures us that the wealth is close to the optimal one, but there is no closed formula for the optimal wealth. Also, the Universal Portfolio algorithm strategy itself cannot be computed in a closed formula and it has to be approximated Kalai and Vempala [2003]. Note that $|g_t|$ can be equal to 1, so we cannot use the efficient algorithm in Hazan et al. [2007]. Finally, it is not known if the Universal Portfolio algorithm satisfies the Assumption 1, hence we cannot use it to solve OLO/OCO problems in Hilbert spaces.

To overcome these problems, in the next section, we will present a simple algorithm that satisfies Assumption 1 and it has $\mathcal{O}(1)$ complexity per prediction.

5 COCOB: A Master Betting Algorithm for Continuous Coins

Algorithm 1 COCOB

Parameters: $a > 2, \delta > 1$
Initialize: $Wealth_0 = \epsilon, G_0 = \delta$
for $t = 1, 2, \dots$ **do**
  Receive θ_{t-1} and G_{t-1}
  Calculate fraction and direction to bet: $\beta_t = 2\,S\left(\frac{4\theta_{t-1}}{a(G_{t-1}+1)}\right) - 1$
  Bet $b_t = \beta_t Wealth_{t-1}$
  Win (or lose) $b_t z_t$
  Update your money: $Wealth_t = Wealth_{t-1} + b_t z_t$
end for

As said in Section 3, we need an efficient MBA algorithm for continuous coin betting. Hence, in this section we show how a very simple algorithm that satisfies Assumption 1.

The first observation is that the wealth of the optimal strategy, as proved in Theorems 7 and 8, is expected to grow exponential in time depending on $D\left(\frac{1}{2} + \frac{\sum_{t=1}^n g_t}{2n} \middle\| \frac{1}{2}\right)$. If we try to achieve exactly this term, we are (probably) doomed to use the universal portfolio selection algorithm or to assume a binary coin. However, we might try to achieve something that is very close to that quantity. The idea is to sacrifice a bit of the guaranteed growth of the wealth in order to gain generality. In particular we aim at designing a MBA.

It is easy to show that

$$D\left(\frac{1}{2} + \frac{\sum_{t=1}^n g_t}{2n} \middle\| \frac{1}{2}\right) \geq \frac{(\sum_{t=1}^n g_t)^2}{2n^2},$$

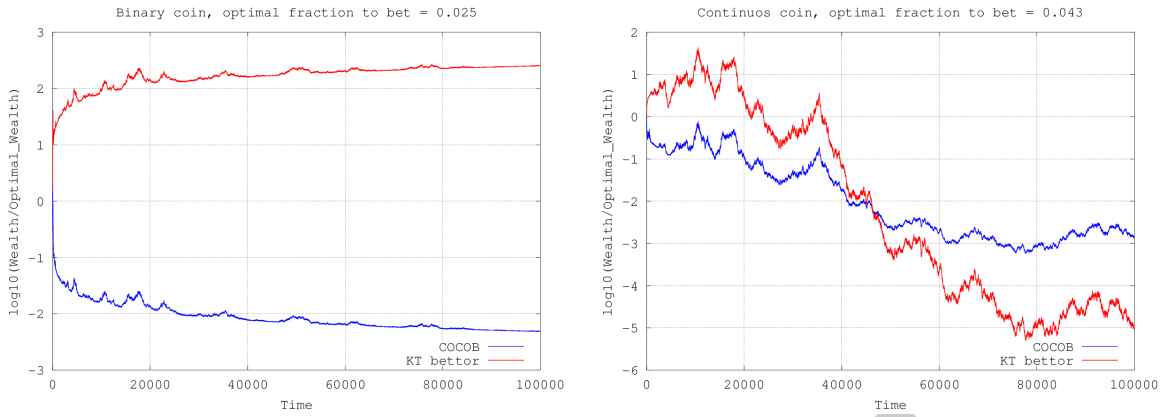

Figure 2: Log ratio between the algorithms and the optimal a posteriori fixed betting strategy, binary (left) and continuous (right) coin.

where the approximation is very good around $\sum_{t=1}^{n} g_t = 0$ and it is of the order of $\mathcal{O}(\frac{(\sum_{t=1}^{n} g_t)^4}{n^4})$. It was first proposed (but not solved) by McMahan and Abernethy [2013] to use the above approximation as a target wealth. Indeed, we show below that it is possible to design a simple algorithm that has an exponential reward that depends on this quantity. Also, our guarantee will be data-dependent and the resulting algorithm will satisfy Assumption 1.

The Continous Coin Betting (COCOB) algorithm is shown in Algorithm 1 and we can prove it is a MBA.

Theorem 10. *Let $a \geq 2$. Then the function*

$$f(x, \{z_1, \dots, z_t\}) := \epsilon \exp\left(\frac{x^2}{a(1+\delta+\sum_{i=1}^{t-1} z_i)} - \sum_{i=1}^{n} \frac{z_i}{a(1+\delta+\sum_{j=1}^{i-1} z_j)}\right)$$

and the betting

$$b_t := Reward_{t-1}\left(2\,S\left(\frac{4x}{a(1+\delta+\sum_{i=1}^{t-1} z_i)}\right) - 1\right),$$

satisfy Assumption 1.

Notice that the a regulates the trade-off between the first and second term. When $a = 2$, we maximize the contribute of the first term.

Corollary 2. *Set $a = 2$ in Algorithm 1. Then we have that the following hold*

- *Feed Algorithm 1 with $G_{t-1} = \sum_{i=1}^{t-1} |g_t| + \delta + 1$ and $\theta_{t-1} = |\sum_{i=1}^{t-1} g_i|$. Algorithm 1 guarantees an exponential reward in $\frac{(\sum_{t=1}^{n} g_t)^2}{2\sum_{t=1}^{n} |g_t|}$ and a regret w.r.t. any real number u of $\tilde{\mathcal{O}}(|u|\sqrt{\sum_{t=1}^{n} |g_t|})$.*
- *Feed Algorithm 1 with $G_{t-1} = \sum_{i=1}^{t-1} \|g_t\| + \delta + 1$ and $\theta_{t-1} = \|\theta_{t-1}\|$. Predicting with $w_t = b_t \frac{\sum_{i=1}^{t} g_t}{\|\sum_{i=1}^{t} g_t\|}$ guarantees a regret w.r.t. any $u \in \mathcal{H}$ of $\tilde{\mathcal{O}}(\|u\| \sqrt{\sum_{t=1}^{n} \|g_t\|})$ and an exponential reward in $\frac{(\sum_{t=1}^{n} g_t)^2}{2\sum_{t=1}^{n} \|g_t\|}$.*

6 Experiments

In this section we show some empirical results.

First, we tested the binary and continuous betting scenario. We compared the performance of COCOB and the KT algorithm, compared to the optimal a posteriori fixed betting strategy. This latter one is simply computed as

$$\beta^* = \arg\max_{\beta} \prod_{t=1}^{n} (1 + g_t \beta) \; .$$

In Figure 2, we show the log ratio between the algorithms and the optimal a posteriori fixed betting strategy. We see that, according to the theory, the performance of the KT strategy on binary coin is better than the one of COCOB. Also, the KT algorithm is actually better than the a posteriori fixed optimal strategy, because it adapts to small asymmetries at the beginning of the sequence of outcomes, resulting in a small fixed gain w.r.t. the fixed optimal strategy.

The situation is reversed in the more interesting setting of the continuous coin. Here, we see that COCOB shows and advantage of KT. Even more importantly, the difference between KT and the optimal strategy sharply increases over time, due to the suboptimal nature of the KT algorithm in this setting.

References

H. H. Bauschke and P. L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer Publishing Company, Incorporated, 1st edition, 2011.

A.V. Boyd. Inequalities for Mills' ratio. *Reports of Statistical Application Research (Union of Japanese Scientists and Engineers)*, 6:44–46, 1959.

N. Cesa-Bianchi and G. Lugosi. *Prediction, learning, and games*. Cambridge University Press, 2006.

N. Cesa-Bianchi, Y. Freund, D. Haussler, D. P. Helmbold, R. E. Schapire, and M. K. Warmuth. How to use expert advice. *J. ACM*, 44(3):427–485, 1997.

N. Cesa-Bianchi, A. Conconi, and C. Gentile. On the generalization ability of on-line learning algorithms. *IEEE Trans. on Information Theory*, 50(9):2050–2057, 2004.

T. M. Cover and E. Ordentlich. Universal portfolios with side information. *Information Theory, IEEE Transactions on*, 42(2):348–363, 1996.

F. Orabona. Dimension-free Exponentiated Gradient. In C.J.C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pages 1806–1814. Curran Associates, Inc., 2013.

E. Hazan, A. Agarwal, and S. Kale. Logarithmic regret algorithms for online convex optimization. *Machine Learning*, 69(2-3):169–192, 2007.

A. Hoorfar and M. Hassani. Inequalities on the Lambert W function and hyperpower function. *J. Inequal. Pure and Appl. Math*, 9(2), 2008.

A. Kalai and S. Vempala. Efficient algorithms for universal portfolios. *J. Mach. Learn. Res.*, 3:423–440, March 2003.

J. L. Kelly. A new interpretation of information rate. *Information Theory, IRE Transactions on*, 2(3):185–189, September 1956.

W. M. Koolen, M. K. Warmuth, and J. Kivinen. Hedging structured concepts. In *Proc. of COLT*, pages 93–105, 2010.

R. E. Krichevsky and V. K. Trofimov. The performance of universal encoding. *IEEE Transactions on Information Theory*, 27(2):199–206, 1981.

N. Littlestone and M. K. Warmuth. The weighted majority algorithm. *Information and Computation*, 108(2):212–261, 1994.

B. D. McKay. On Littlewoods estimate for the binomial distribution. *Advanced Applied Probability*, 21:475–478, 1989.

B. McMahan and J. Abernethy. Minimax optimal algorithms for unconstrained linear optimization. In C.J.C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pages 2724–2732. Curran Associates, Inc., 2013.

H. B. McMahan and F. Orabona. Unconstrained online linear learning in Hilbert spaces: Minimax algorithms and normal approximations. In *COLT*, 2014.

A. Nemirovski and D. B. Yudin. *Problem complexity and method efficiency in optimization*. Wiley, 1983.

F. Orabona. Simultaneous model selection and optimization through parameter-free stochastic learning. In *Advances in Neural Information Processing Systems 27*, 2014.

S. Shalev-Shwartz. Online learning and online convex optimization. *Foundations and Trends in Machine Learning*, 4 (2), 2012.

K. Sridharan, S. Shalev-Shwartz, and N. Srebro. Fast rates for regularized objectives. In D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou, editors, *Advances in Neural Information Processing Systems 21*, pages 1545–1552. Curran Associates, Inc., 2009.

V. Vovk. A game of prediction with expert advice. *Journal of Computer and System Sciences*, 56:153–173, 1998.

A Proofs

First we state some technical lemmas that will be used in the following proofs.

Define the Lambert function $W(x) : \mathbb{R} \to \mathbb{R}$ as the one that satisfies the equality[1]

$$x = W(x) \exp\left(W(x)\right), \ \forall x \geq 0. \tag{6}$$

It satisfies the following properties.

Lemma 1. *The Lambert function satisfies* $0.6321 \log(x+1) \leq W(x) \leq \log(x+1), \forall x \geq 0$.

Proof. We first prove the lower bound. From (6) we have

$$W(x) = \log\left(\frac{x}{W(x)}\right) \tag{7}$$

$$= \log\left(\frac{x}{\log(x/W(x))}\right). \tag{8}$$

From the first equality, for any $a > 0$, we get

$$W(x) \leq \frac{1}{a\,e}\left(\frac{x}{W(x)}\right)^a$$

that is

$$W(x) \leq \left(\frac{1}{a\,e}\right)^{\frac{1}{1+a}} x^{\frac{a}{1+a}}. \tag{9}$$

[1]For $x < 0$ the Lambert function is multivalued. Hence, to avoid complication and because we only need positive arguments, we will define it only for positive values of x.

Using (9) in (7), we have

$$W(x) \geq \log\left(\frac{x}{\left(\frac{1}{a\,e}\right)^{\frac{1}{1+a}}\, x^{\frac{a}{1+a}}}\right) = \frac{1}{1+a}\log\left(a\,e\,x\right)\ .$$

Consider now the function $g(x) = \frac{x}{x+1} - \frac{b}{\log(1+b)(b+1)}\log(x+1), x \geq b$. This function has a maximum in $x^* = (1+\frac{1}{b})\log(1+b) - 1$, the derivative is positive in $[0, x^*]$ and negative in $[x^*, b]$. Hence the minimum is in $x = 0$ and in $x = b$, where it is equal to 0. Using the property just proved on g, we have that for $x \leq b$, setting $a = \frac{1}{x}$, we have

$$W(x) \geq \frac{x}{x+1} \geq \frac{b}{\log(1+b)(b+1)}\log(x+1)\ .$$

For $x > b$, setting $a = \frac{x+1}{ex}$, we have

$$W(x) \geq \frac{e\,x}{(e+1)x+1}\log(x+1) \geq \frac{e\,b}{(e+1)b+1}\log(x+1) \tag{10}$$

Hence, we set b such that

$$\frac{e\,b}{(e+1)b+1} = \frac{b}{\log(1+b)(b+1)}$$

Numberically, $b = 1.71825...$, so

$$W(x) \geq 0.6321\log(x+1)\ .$$ □

For the upper bound, we use Theorem 2.3 in Hoorfar and Hassani [2008], that says that

$$W(x) \leq \log\frac{x+C}{1+\log(C)}, \quad \forall x > -\frac{1}{e},\ C > \frac{1}{e}.$$

Setting $C = 1$, we obtain the stated bound.

Lemma 2. *Define* $f(\theta) = \beta\exp\frac{x^2}{2\alpha}$, *for* $\alpha, \beta > 0$, $x \geq 0$. *Then*

$$f^*(y) = y\sqrt{\alpha W\left(\frac{\alpha y^2}{\beta^2}\right)} - \beta\exp\left(\frac{W\left(\frac{\alpha y^2}{\beta^2}\right)}{2}\right).$$

Moreover

$$f^*(y) \leq y\sqrt{\alpha\log\left(\frac{\alpha y^2}{\beta^2}+1\right)} - \beta.$$

Proof. From the definition of Fenchel dual, we have

$$f^*(y) = \max_x\ x\,y - f(x) = \max_x\ x\,y - \beta\exp\frac{x^2}{2\alpha} \leq x^*\,y - \beta$$

where $x^* = \arg\max_x x\,y - f(x)$. We now use the fact that x^* satisfies $y = f'(x^*)$, to have

$$x^* = \sqrt{\alpha W\left(\frac{\alpha y^2}{\beta^2}\right)},$$

where the function $W : \mathbb{R}_+ \to \mathbb{R}$ is the Lambert function that satisfies

$$x = W(x)\exp\left(W(x)\right).$$

Hence, to obtain an upper bound we need an upper bound to the Lambert function. We use Theorem 2.3 in Hoorfar and Hassani [2008], that says that

$$W(x) \leq \log \frac{x+C}{1+\log(C)}, \quad \forall x > -\frac{1}{e},\ C > \frac{1}{e}.$$

Setting $C = 1$, we obtain the stated bound. □

Lemma 3 ([Bauschke and Combettes, 2011, Example 13.7])**.** *Let $\phi : \mathbb{R} \to (-\infty, +\infty]$ be even. Then $(\phi * \|\cdot\|)^* = \phi^* * \|\cdot\|$.*

Corollary 3. *Define $f(\theta) = \beta \exp \frac{\|\theta\|^2}{2\alpha}$, for $\alpha, \beta > 0$. Then*

$$f^*(y) \leq \|\theta\| \sqrt{\alpha \log \left(\frac{\alpha \|\theta\|^2}{\beta^2} + 1 \right)} - \beta.$$

A.1 Proof of Theorem 2

Proof. For simplicity denote by $f_t(\cdot) = f\left(\cdot, \{\|g_1\|, \ldots, \|g_t\|\}\right)$. We will prove the thesis by induction. The base case is verified from the first point of Assumption 1. The, we assume that

$$\epsilon + \sum_{t=1}^{n-1} \langle g_t, w_t \rangle \geq f_{n-1}\left(\left\| \sum_{t=1}^{n-1} g_t \right\| \right),$$

and we want to prove that

$$\epsilon + \sum_{t=1}^{n} \langle g_t, w_t \rangle \geq f_n\left(\left\| \sum_{t=1}^{n} g_t \right\| \right).$$

We have that

$$\begin{aligned}
&\epsilon + \sum_{t=1}^{n} \langle g_t, w_t \rangle - f_n\left(\left\| \sum_{t=1}^{n} g_t \right\| \right) \\
&= \langle g_n, w_n \rangle + \epsilon + \sum_{t=1}^{n-1} \langle g_t, w_t \rangle - f_n\left(\left\| \sum_{t=1}^{n} g_t \right\| \right) \\
&= \left(1 + \frac{b_n}{\|\theta_{n-1}\|} \langle \theta_{n-1}, g_n \rangle \right) \left(\sum_{t=1}^{n-1} \langle g_t, w_t \rangle + \epsilon \right) - f_n\left(\left\| \sum_{t=1}^{n} g_t \right\| \right) \\
&\geq \left(1 + \frac{b_n}{\|\theta_{n-1}\|} \langle \theta_{n-1}, g_n \rangle \right) f_{n-1}\left(\left\| \sum_{t=1}^{n-1} g_t \right\| \right) - f_n\left(\left\| \sum_{t=1}^{n} g_t \right\| \right) \\
&= \left(1 + \frac{b_n}{\|\theta_{n-1}\|} \langle \theta_{n-1}, g_n \rangle \right) f_{n-1}\left(\left\| \sum_{t=1}^{n-1} g_t \right\| \right) - f_n\left(\left\| g_n + \sum_{t=1}^{n-1} g_t \right\| \right) \\
&\geq \min_{r \in \{-1,1\}} \left(1 + r\, b_n \|g_n\|\right) f_{n-1}\left(\|\theta_{n-1}\|\right) - f_n\left(\left\|\|\theta_{n-1}\| + r\, \|g_n\|\right\|\right) \\
&= \min_{r \in \{-1,1\}} \left(1 + r\, b_n \|g_n\|\right) f_{n-1}\left(\|\theta_{n-1}\|\right) - f_n\left(\|\theta_{n-1}\| + r\, \|g_n\|\right) \\
&\geq 0,
\end{aligned}$$

where the first inequality comes from the induction hypothesis, the second one using Lemma 8 in McMahan and Orabona [2014] and the last one by the hypothesis on the MBA. □

A.2 Proof of Theorem 5

Proof. The statement is readily proved using the inequality $a\,b \leq \frac{a^2}{\lambda} + \lambda b^2,\ \forall a, b \geq 0, \lambda > 0$ and Theorem 4. □

A.3 Proof of Theorem 6

Proof. From Theorem 2 we have that

$$Reward_n = Wealth_n - \epsilon \geq B n^{-C} \exp\left(\frac{\left\|\sum_{t=1}^n g_t\right\|^2}{An}\right) - D - \epsilon\,.$$

We now apply Theorem 3 and Corollary 3 to have that

$$Regret_n(u) \leq \|u\| \sqrt{\frac{AT}{2} \log\left(\frac{\frac{A}{2} T^{2C+1} \|u\|^2}{B^2} + 1\right)} - \frac{B}{T^C} + \epsilon + D\,.$$ □

A.4 Proofs of Theorem 7 and Theorem 8

We first state a couple of useful indentities. For any $0 \leq p < 1$

$$D\left(\frac{1}{2} + \frac{p}{2} \middle\| \frac{1}{2}\right) = D\left(\frac{1}{2} - \frac{p}{2} \middle\| \frac{1}{2}\right) = \frac{1+p}{2}\log(1+p) + \frac{1-p}{2}\log(1-p).$$

The extension for continuity of $D(\frac{1}{2} + \frac{p}{2}||\frac{1}{2})$ in $p = 1$ is $\log(2)$. Also,

$$\left(\frac{n}{n-q}\right)^{n-q}\left(\frac{n}{q}\right)^{q} = 2^n \exp\left(-nD\left(\frac{q}{n}\middle\|\frac{1}{2}\right)\right),$$

and

$$(1+x)^{\frac{1+x}{2}}(1-x)^{\frac{1-x}{2}} = \exp\left(D\left(\frac{1}{2} + \frac{x}{2}\middle\|\frac{1}{2}\right)\right) \tag{11}$$

Also, for any $-\frac{1}{2} \leq x \leq \frac{1}{2}$ we have

$$\frac{x^2}{2} + \frac{x^4}{12} \leq D\left(\frac{1}{2} + \frac{x}{2}\middle\|\frac{1}{2}\right) \leq \frac{x^2}{2} + \frac{x^4}{5}.$$

Proof of Theorem 7. From the betting strategy we have

$$Wealth_t = Wealth_{t-1} + w_t\, g_t = Wealth_{t-1} + \beta\, Wealth_{t-1}\, g_t = Wealth_{t-1}(1 + \beta\, g_t)\,.$$

Hence

$$Wealth_n = \epsilon \prod_{t=1}^n (1 + \beta g_t) = \epsilon (1+\beta)^{\frac{n+Z}{2}} (1-\beta)^{\frac{n-Z}{2}},$$

where $G = \sum_{t=1}^n g_t$. It is easy to show that the maximum value of $Wealth_n$ w.r.t. β is in $\beta = \frac{G}{n}$. Hence, we have

$$Wealth_n = \epsilon\left(1 + \frac{G}{n}\right)^{\frac{n+G}{2}}\left(1 - \frac{G}{n}\right)^{\frac{n-G}{2}} = \epsilon\left[\left(1 + \frac{G}{n}\right)^{\frac{1+\frac{G}{n}}{2}}\left(1 - \frac{G}{n}\right)^{\frac{1-\frac{G}{n}}{2}}\right]^n = \exp\left(D\left(\frac{1}{2} + \frac{G}{2n}\middle\|\frac{1}{2}\right)\right),$$

where in the last equality we used (11). □

The following tail bound for Binomial variables is from F. Orabona [2013] and we report here the proof for completeness.

Theorem 11. *Let $n \geq 2$ an even number of Bernoulli random variables b_i. Then for any $k \in \mathbb{N}_0$ such that $k \leq \frac{1}{2}n-1$, we have*

$$P\left(\sum_{i=1}^{n} b_i \geq \frac{1}{2}n + k\right) \geq \frac{\exp\left(-n\, D\left(\frac{1}{2}+\frac{k}{n}\middle\|\frac{1}{2}\right)\right)}{2\exp\left(\frac{1}{6}\right)} \frac{\sqrt{2\pi}}{(\pi-1)y+\sqrt{y^2+2\pi}},$$

where $y = \frac{2k}{\sqrt{n}}$.

Proof. We use Theorem 2 in McKay [1989], that specialized to our case says that

$$P\left(\sum_{i=1}^{n} b_i \geq \frac{1}{2}n + k\right) \geq \sqrt{n}\binom{n-1}{\frac{1}{2}n+k-1} 2^{-n} \frac{Q(y)}{\phi(y)}, \tag{12}$$

where $\phi(x)$ is the unit variance, zero mean Gaussian, $\frac{1}{\sqrt{2\pi}}\exp(-\frac{x^2}{2})$ and $Q(x)$ is its CDF, $\int_x^{+\infty}\phi(u)du$.

We start lower bounding the ratio $\frac{Q(y)}{\phi(y)}$. Using the inequality in Boyd [1959], that says

$$\frac{Q(y)}{\phi(y)} = \exp\left(\frac{x^2}{2}\right)\int_x^{+\infty}\exp\left(-\frac{t^2}{2}\right)dt \geq \frac{\pi}{(\pi-1)x+\sqrt{x^2+2\pi}}.$$

To bound the binomial coefficient we make use of the following Stirling approximation, for any $n \geq 1$,

$$\sqrt{2\pi n}n^n\exp(-n) < n! < \exp\left(\frac{1}{12}\right)\sqrt{2\pi n}n^n\exp(-n)\,.$$

Hence, for any $n \geq 2$ and $1 \leq q \leq n-1$, after some algebra we obtain

$$\begin{aligned}\binom{n}{q} &\geq \frac{1}{\exp\left(\frac{1}{6}\right)\sqrt{2\pi}}\left(\frac{n}{n-q}\right)^{n-q}\left(\frac{n}{q}\right)^{q}\sqrt{\frac{n}{q(n-q)}} \\ &\geq \frac{1}{\exp\left(\frac{1}{6}\right)\sqrt{2\pi}}2^n\exp\left(-nD\left(\frac{q}{n}\middle\|\frac{1}{2}\right)\right)\sqrt{\frac{n}{q(n-q)}}.\end{aligned}$$

where in the equality we used the definition of $D\left(\cdot\|\cdot\right)$. Also, we have

$$\binom{n-1}{\frac{1}{2}n+k-1} = \binom{n}{\frac{1}{2}n+k}\left(\frac{1}{2}+\frac{k}{n}\right). \tag{13}$$

Putting together (12)-(13), and using the definition of y we have

$$\begin{aligned}P\left(\sum_{i=1}^{n} b_i \geq \frac{1}{2}n + k\right) &\geq \frac{1}{\exp\left(\frac{1}{6}\right)\sqrt{2\pi}}\exp\left(-n\,D\left(\frac{1}{2}+\frac{k}{n}\middle\|\frac{1}{2}\right)\right)\sqrt{\frac{\frac{1}{2}+\frac{k}{n}}{\frac{1}{2}-\frac{k}{n}}}\frac{Q(y)}{\phi(y)} \\ &\geq \frac{1}{\exp\left(\frac{1}{6}\right)\sqrt{2\pi}}\exp\left(-n\,D\left(\frac{1}{2}+\frac{k}{n}\middle\|\frac{1}{2}\right)\right)\frac{\pi}{(\pi-1)y+\sqrt{y^2+2\pi}}.\end{aligned}$$

□

We can now prove Theorem 8.

Proof of Theorem 8. First observe that, even knowing all the outcomes of g_t, we cannot gain more than $\epsilon 2^n$, simply betting on each round all the money on the correct outcome.

Then, for a specific function $g(\cdot,\cdot)$ that grows in the first argument, we will show that

$$\min_{|\sum_{t=1}^{n} g_t| \geq 2k}\sum_{t=1}^{n} w_t g_t \leq \epsilon\, g(k,n),\ \forall 0 \leq k \leq n-1.$$

When $k = n/2$, we cannot gain more than $\epsilon 2^n$, this is why we can safely consider only $0 \leq k \leq n/2 - 1$. From this we will infer that

$$\min_{g_t} \sum_{t=1}^{n} w_t g_t - \epsilon \min\left(g\left(\frac{|\sum_{t=1}^{n} g_t|}{2}, n\right), 2^n\right) \leq 0,$$

that implies the stated inequality.

Set r_t as independent random variable that assumes the value of 1 with probability 0.5 and -1 with probability 0.5. Hence, we have that $\mathbb{E}[\sum_{t=1}^{n} w_t r_t] = 0$, and also $\sum_{t=1}^{n} w_t r_t \geq -\epsilon$ because we never lose more than the initial amount of money.

Hence we have

$$\min_{|\sum_{t=1}^{n} g_t| \geq 2k} \sum_{t=1}^{n} w_t g_t \leq \mathbb{E}\left[\sum_{t=1}^{n} w_t r_t \middle| \left|\sum_{t=1}^{n} r_t\right| \geq 2k\right], \ \forall 0 \leq k \leq n-1.$$

For any $k \geq 0$, it follows that

$$\begin{aligned}
0 = \mathbb{E}\left[\sum_{t=1}^{n} w_t r_t\right] &= \mathbb{E}\left[\sum_{t=1}^{n} w_t r_t \middle| \left|\sum_{t=1}^{n} r_t\right| < 2k\right] P\left(\left|\sum_{t=1}^{n} r_t\right| < 2k\right) \\
&\quad + \mathbb{E}\left[\sum_{t=1}^{n} w_t r_t \middle| \left|\sum_{t=1}^{n} r_t\right| \geq 2k\right] P\left(\left|\sum_{t=1}^{n} r_t\right| \geq 2k\right) \\
&= \mathbb{E}\left[\sum_{t=1}^{n} w_t r_t \middle| \left|\sum_{t=1}^{n} r_t\right| < 2k\right] \left(1 - P\left(\left|\sum_{t=1}^{n} r_t\right| \geq 2k\right)\right) \\
&\quad + \mathbb{E}\left[\sum_{t=1}^{n} w_t r_t \middle| \left|\sum_{t=1}^{n} r_t\right| \geq 2k\right] P\left(\left|\sum_{t=1}^{n} r_t\right| \geq 2k\right) \\
&\geq -\epsilon\left(1 - P\left(\left|\sum_{t=1}^{n} r_t\right| \geq 2k\right)\right) + \mathbb{E}\left[\sum_{t=1}^{n} w_t r_t \middle| \left|\sum_{t=1}^{n} r_t\right| \geq 2k\right] P\left(\left|\sum_{t=1}^{n} r_t\right| \geq 2k\right),
\end{aligned}$$

hence

$$\begin{aligned}
\mathbb{E}\left[\sum_{t=1}^{n} w_t r_t \middle| \left|\sum_{t=1}^{n} r_t\right| \geq 2k\right] &\leq \frac{\epsilon}{P\left(|\sum_{t=1}^{n} r_t| \geq 2k\right)} - \epsilon = \frac{\epsilon}{2P\left(\sum_{t=1}^{n} r_t \geq 2k\right)} - \epsilon \\
&= \frac{\epsilon}{2P\left(\sum_{t=1}^{n} \frac{r_t+1}{2} \geq \frac{1}{2}n + k\right)} - \epsilon.
\end{aligned}$$

Notice that $\frac{r_t+1}{2}$ are Bernoulli random variables. Using Theorem 11, we have

$$\mathbb{E}\left[\sum_{t=1}^{n} w_t r_t \middle| \left|\sum_{t=1}^{n} r_t\right| \geq 2k\right] \leq \epsilon\left(\exp\left(\frac{1}{6}\right)\sqrt{2\pi}\frac{2k}{\sqrt{n}} + 2\exp\left(\frac{1}{6}\right) - 1\right)\exp\left(n\, D\left(\frac{1}{2} + \frac{k}{n}\middle\|\frac{1}{2}\right)\right). \qquad \square$$

A.5 Proof of Theorem 10

Proof. We want to prove that, for any t

$$(1 + b_t g_t)\epsilon \exp\left(\frac{x^2}{a(\delta + \sum_{i=1}^{t-1} z_i)} - \sum_{i=1}^{t-1} \frac{z_i}{a(\delta + \sum_{j=1}^{i-1} z_j)}\right) \leq \epsilon \exp\left(\frac{(x+g_t)^2}{a(\delta + \sum_{i=1}^{t} z_i)} - \sum_{i=1}^{t} \frac{z_i}{a(\delta + \sum_{j=1}^{i-1} z_j)}\right).$$

The above is equivalent to prove

$$\ln(1 + b_t g_t) + \frac{x^2}{a(\delta + \sum_{i=1}^{t-1} z_i)} - \sum_{i=1}^{t-1} \frac{z_i}{a(\delta + \sum_{j=1}^{i-1} z_j)} \leq \frac{(x+g_t)^2}{a(\delta + \sum_{i=1}^{t} z_i)} - \sum_{i=1}^{t} \frac{z_i}{a(\delta + \sum_{j=1}^{i-1} z_j)}.$$

Denote by $G_{t-1} = \sum_{i=1}^{t-1} z_i$ and consider the function

$$\phi(g) = -\log(1+\beta_t g) + \frac{(x+g)^2}{a(G_{t-1}+|g|)} \, .$$

We have that $\phi(g)$ is piece-wise convex on $[-\infty, 0]$ and $[0, \infty]$. Hence, we have that

$$\begin{aligned}
\phi(g) &\leq \phi(0) + g(\phi(1) - \phi(0)), \forall 0 \leq g \leq 1 \\
\phi(g) &\leq \phi(0) + g(\phi(0) - \phi(-1)), \forall -1 \leq g \leq 0 \, .
\end{aligned}$$

Also, we set β_t such that $\phi(1) = \phi(-1)$, that is

$$\beta_t = \frac{A_{t-1} - 1}{A_{t-1} + 1} = 2\, S\left(\frac{4x}{a(G_{t-1}+1)}\right) - 1$$

where $A_{t-1} = \exp\left(\frac{4x}{a(G_{t-1}+1)}\right)$ and $S(x) = \frac{1}{1+\exp(-x)}$. Hence we have

$$\phi(g) \leq \phi(0) + |g|(\phi(1) - \phi(0)), \forall -1 \leq g \leq 1,$$

that is

$$\begin{aligned}
\frac{x^2}{aG_{t-1}} - \frac{(x+g)^2}{a(G_{t-1}+|g|)} + \log(1+\beta_t g_t) &= \phi(0) - \phi(g) \\
&\geq |g|\left(\phi(0) - \phi(1)\right) \\
&= |g|\left(\frac{x^2}{aG_{t-1}} - \frac{(x+1)^2}{a(G_{t-1}+1)} + \log(1+\beta_t)\right), \forall -1 \leq g \leq 1 \, .
\end{aligned}$$

Using this relation we have that

$$\begin{aligned}
&- \frac{(x+g_t)^2}{a(G_{t-1}+|g_t|)} + \frac{x^2}{aG_{t-1}} + \log(1+\beta_t g_t) - \sum_{i=1}^{t-1} \frac{|g_i|}{a(G_{i-1}+1)} \\
&\geq |g_t|\left(\frac{x^2}{aG_{t-1}} - \frac{(x+1)^2}{a(G_{t-1}+1)} + \log(1+\beta_t G_t)\right) - \sum_{i=1}^{t-1} \frac{|g_i|}{a(G_{i-1}+1)} \\
&= |g_t|\left(\frac{ax^2 - 2xG_{t-1}}{a^2 G_{t-1}(G_{t-1}+1)} + \log(1+\beta_t G_t)\right) - \sum_{i=1}^{t} \frac{|g_i|}{a(G_{i-1}+1)} \\
&\geq |g_t|\left(\frac{ax^2}{a^2 G_{t-1}(G_{t-1}+1)} - \frac{2x}{a(G_{t-1}+1)} + \log(1+\beta_t)\right) - \sum_{i=1}^{t} \frac{|g_i|}{a(G_{i-1}+1)} \, .
\end{aligned}$$

We now use the Taylor expansion, to obtain

$$\log\left(1 + \frac{\exp(x)-1}{\exp(x)+1}\right) \geq \frac{x}{2} - \frac{x^2}{8} \qquad \forall x \in \mathbb{R}$$

and, using the expression of β_t, have

$$\log\left(1+\beta_t\right) = \log\left(1 + \frac{\exp\left(\frac{4x}{a(G_{t-1}+1)}\right) - 1}{\exp\left(\frac{4x}{a(G_{t-1}+1)}\right) + 1}\right) \geq \frac{2x}{a(G_{t-1}+1)} - \frac{2x^2}{a^2(G_{t-1}+1)^2}.$$

Hence the expression

$$\frac{ax^2}{a^2 G_{t-1}(G_{t-1}+1)} - \frac{2x}{a(G_{t-1}+1)} + \log(1+\beta_t)$$

is greater than zero if $a \geq 2$, that is true by definition of a. □