Varieties of Regret in Online Prediction

Casey Marks, Amy Greenwald,
David Gondek

Department of Computer Science
Brown University
Providence, Rhode Island 02912

CS-04-09
July 2004

Varieties of Regret in Online Prediction

Casey Marks
casey@cs.brown.edu

Amy Greenwald
amygreen@cs.brown.edu

David Gondek
dcg@cs.brown.edu

Department of Computer Science
Brown University, Box 1910
Providence, RI 02912

Abstract

We present a general framework for analyzing regret in the online prediction problem. We develop this from sets of linear transformations of strategies. We establish relationships among the varieties of regret and present a class of regret-matching algorithms. Finally we consider algorithms that exhibit the asymptotic no-regret property. Our main results are an analysis of observed regret in expectation and two regret-matching algorithms that exhibit no-observed-internal-regret in expectation.

1 Introduction

The analysis of learning algorithms by calculating "regret" is of interest in both machine learning and game theory. Regret can be thought of as the difference between an algorithm's actual performance and the performance of some transformation of its history of play. In this paper we analyze regret in the context of the online prediction problem.

Much of this paper follows parts of Cesa-Bianci and Lugosi (2003). However in addition to their analysis of expected regret, we provide an analysis of observed regret in expectation (which bounds expected regret from above) inspired by unpublished work by Geoff Gordon. In Section 2 we consider sets of linear transformations of mixed strategies and the corresponding varieties of regret in both the observed and expected analyses, and then we consider the relationships among the varieties. In Section 3 we introduce a general class of regret-matching algorithms and prove a bound on their regret under certain conditions. We also consider two particular cases of regret-matching algorithms: polynomial and exponential. In Section 4 we consider the asymptotic "no regret" property.

2 Transformations and Regret

2.1 The Model

At all time steps the agent has the same finite set of actions available to it. At each iteration, the agent deterministically generates a distribution (mixed strategy) over the action set. It is then informed of the reward for each action at that time step. Let $N = \{1, 2, \ldots, n\}$ be the set of actions. Let r^t be a row vector such that r^t_j is the reward for playing action j. We assume bounded rewards; without loss of generality the rewards are in $[0, 1]$, so $r^t \in [0, 1]^n$. Let Q be the set of all distributions over N. Let $e_i \in Q$ be the distribution with all its weight on action i. Let A be the set of all such distributions (i.e., the set of all pure strategies). Let $q^t \in Q$ denote the mixed strategy played by the agent at time t, and let $a^t \in A$ be the realization of that strategy. Thus the actual payoff the agent receives at time t is $p^t = r^t \cdot a^t$, while the expected payoff is $\mathbb{E}[p^t|q^t] = r^t \cdot q^t$. Treat both a^t and q^t as row vectors.

2.2 Transformations

Consider the set of linear transformations of mixed strategies. We can represent them by the set of row-stochastic matrices, so that the transformation of $x \in Q$ is $x\phi$, where ϕ is such a matrix. Let $\Phi_{stochastic}$ be the set of all such matrices, and let Φ_{swap} be the set of all deterministic row-stochastic matrices (i.e., elements of $\Phi_{stochastic}$ whose entries are either 1 or 0).

Consider the set of mappings from the set of actions to itself, $\mathcal{F}_{swap} = \{F : N \to N\}$. Note that the operation of elements of $\mathcal{F}_{swap}$ on N is isomorphic to the operation of Φ_{swap} on A. The isomorphism between $\phi \in \Phi_{swap}$ and $F \in \mathcal{F}_{swap}$ is $(\phi)_{ij} = \delta^j_{F(i)}$, where δ is the Krondecker delta, or $e_i\phi = e_{F(i)}$.

Consider two subsets of $\mathcal{F}_{swap}$:

$$\mathcal{F}_{in} = \left\{ F_{ij} : x \mapsto \begin{cases} j & \text{if } x = i \\ i & \text{otherwise} \end{cases} \right\} \tag{1}$$

$$\mathcal{F}_{ext} = \{F_i : x \mapsto i\} \tag{2}$$

These give us corresponding subsets of Φ_{swap}, Φ_{in} and Φ_{ext}.

2.3 Regret

At time T we will evaluate regret by comparing the payoffs we obtained (or expected to obtain) to the payoffs we would have obtained by playing some transformation of our history of actions or mixed strategies.

Define ρ_ϕ to be the instantaneous ϕ-regret:

$$\begin{aligned}\rho_\phi(r,x) &= (r\cdot(x\phi) - r\cdot x) &(3)\\ &= r\cdot(x\phi - x) &(4)\\ &= r\cdot(x(\phi - I)) &(5)\end{aligned}$$

which represents the difference in payoffs between playing x and playing $x\phi$ with reward vector r.

For a (finite) sequence of T reward vectors $\{r^t\}, r^t \in [0,1]^n$ and a sequence of T strategy vectors $\{x^t\}, x^t \in Q$, define R_Φ^T to be the (cumulative) Φ-regret at time T.

$$R_\Phi^T(\{r^t\},\{x^t\}) = \max_{\phi\in\Phi}\sum_{t=0}^{T}\rho_\phi(r^t,x^t) \tag{6}$$

For specific algorithms we will consider both observed regret, $R_\Phi^T(\{r^t\},\{a^t\})$, and expected regret, $R_\Phi^T(\{r^t\},\{q^t\})$.

Proposition 1

$$R_\Phi^T(\{r^t\},\{q^t\}) \le \mathbb{E}[R_\Phi^T(\{r^t\},\{a^t\})|\{q^t\}] \tag{7}$$

Proof

$$\begin{aligned}R_\Phi^T(\{r^t\},\{q^t\}) &= \max_{\phi\in\Phi}\sum_{t=0}^{T}\rho(r^t,q^t) &(8)\\ &= \max_{\phi\in\Phi}\sum_{t=0}^{T}\mathbb{E}\left[\rho(r^t,a^t)|\{q^t\}\right] &(9)\\ &= \max_{\phi\in\Phi}\mathbb{E}\left[\sum_{t=0}^{T}\rho(r^t,a^t)|\{q^t\}\right] &(10)\\ &\le \mathbb{E}\left[\max_{\phi\in\Phi}\sum_{t=0}^{T}\rho(r^t,a^t)|\{q^t\}\right] &(11)\\ &= \mathbb{E}[R_\Phi^T(\{r^t\},\{a^t\})|\{q^t\}] &(12)\end{aligned}$$

Line (10) follows from (9) by linearity of expectation. Line (11) follows from (10) and Jensen's inequality because max is convex. ∎

2.4 Varieties of Regret

For any set of transformations, we have a corresponding version of regret. For example, Φ_{in} gives us internal regret, $R_{\Phi_{in}}^T$, which we will simply denote R_{in}^T. Similarly we can calculate stochastic, swap, and external regret.

Proposition 2 *For any* $\{r^t\}$ *and* $\{x^t\}$, $R^T_{stochastic}(\{r^t\},\{x^t\}) = R^T_{swap}(\{r^t\},\{x^t\})$.

Proof

Because $\Phi_{stochastic} \supset \Phi_{swap}$, we have

$$\max_{\phi\in\Phi_{stochastic}} \sum_{t=0}^{T} \rho_\phi(r^t, x^t) \geq \max_{\phi\in\Phi_{swap}} \sum_{t=0}^{T} \rho_\phi(r^t, x^t). \tag{13}$$

Let $\phi^* = \arg\max_{\phi\in\Phi_{stochastic}} \sum_{t=0}^{T} \rho_\phi(r^t, x^t)$. Because $\Phi_{stochastic}$ is the convex hull of Φ_{swap}, ϕ^* can be represented as the convex combination of elements of Φ_{swap}:

$$\phi^* = \sum_{\phi\in\Phi_{swap}} a_\phi \phi \tag{14}$$

so we can write

$$\begin{aligned}
R^T_{stochastic}(\{r^t\},\{x^t\}) &= \sum_{t=0}^{T} \rho_{\phi^*}(r^t, x^t) && (15)\\
&= \sum_{t=0}^{T} (r \cdot (x\phi^*) - r \cdot x) && (16)\\
&= \sum_{t=0}^{T} \left(r \cdot \left(x \sum_{\phi\in\Phi_{swap}} a_\phi \phi \right) - r \cdot x \right) && (17)\\
&= \sum_{\phi\in\Phi_{swap}} a_\phi \sum_{t=0}^{T} (r \cdot (x\phi) - r \cdot x) && (18)\\
&= \sum_{\phi\in\Phi_{swap}} a_\phi \sum_{t=0}^{T} \rho_\phi(r^t, x^t) && (19)
\end{aligned}$$

Consequently there must exist some $\phi \in \Phi_{swap}$ such that $\sum_{t=0}^{T} \rho_\phi(r^t, x^t) \geq R^T_{stochastic}(\{r^t\},\{x^t\})$, implying that $R^T_{stochastic}(\{r^t\},\{x^t\}) \leq R^T_{swap}(\{r^t\},\{x^t\})$. ∎

Because Φ_{swap} is finite, we will consider only $\Phi \subset \Phi_{swap}$. In the next section, we will show that swap regret is bounded above by n times internal regret.

2.5 Regret Matrix

At each time step we can calculate an $n \times n$ observed regret matrix, denoted m^t.

$$m^t_{ij} = a^t_i(r^t_j - r^t_i), \tag{20}$$

Note that this matrix only has entries in the row corresponding to the action played. We can also calculate the cumulative observed regret matrix

$$M_T = \sum_{t=1}^{T} m^t \tag{21}$$

Similarly,

$$\hat{m}_{ij}^t = \mathbb{E}[m_{ij}^t | q^t] = q_i^t (r_j^t - r_i^t), \tag{22}$$

and

$$\hat{M}^T = \sum_{t=1}^{T} \hat{m}^t \tag{23}$$

Lemma 3

$$R_\Phi^T(\{r^t\}, \{a^t\}) = \max_{F \in \mathcal{F}} \sum_{i \in N} M_{iF(i)}^T \tag{24}$$

$$R_\Phi^T(\{r^t\}, \{q^t\}) = \max_{F \in \mathcal{F}} \sum_{i \in N} \hat{M}_{iF(i)}^T \tag{25}$$

if $\mathcal{F}$ is the transformation isomorphic to $\Phi \subset \Phi_{swap}$.

Proof

Prove for $R_\Phi^T(\{r^t\}, \{q^t\})$; it is a generalization of the proof for $R^T(\{r^t\}, \{a^t\})$.

$$\begin{aligned}
R_\Phi^T(\{r^t\}, \{q^t\}) &= \max_{\phi \in \Phi} \sum_{t=0}^{T} (r^t \cdot q^t \phi - r^t \cdot q^t) && (26)\\
&= \max_{\phi \in \Phi} \sum_{t=0}^{T} \left(\left(\sum_{j \in N} r_j^t (q^t \phi)_j \right) - \sum_{i \in N} r_i^t q_i^t \right) && (27)\\
&= \max_{\phi \in \Phi} \sum_{t=0}^{T} \left(\left(\sum_{j \in N} r_j^t \sum_{i \in N} q_i^t \phi_{ij} \right) - \sum_{i \in N} r_i^t q_i^t \right) && (28)\\
&= \max_{F \in \mathcal{F}} \sum_{t=0}^{T} \left(\left(\sum_{j \in N} \sum_{i \in N} r_j^t q_i^t \delta_j^{F(i)} \right) - \sum_{i \in N} r_i^t q_i^t \right) && (29)\\
&= \max_{F \in \mathcal{F}} \sum_{t=0}^{T} \left(\left(\sum_{i \in N} r_{F(i)}^t q_i^t \right) - \sum_{i \in N} r_i^t q_i^t \right) && (30)\\
&= \max_{F \in \mathcal{F}} \sum_{t=0}^{T} \sum_{i \in N} q_i^t (r_{F(i)}^t - r_i^t) && (31)\\
&= \max_{F \in \mathcal{F}} \sum_{i \in N} \hat{M}_{iF(i)}^T && (32)
\end{aligned}$$

∎

We can rewrite external regret as

$$\begin{aligned} R^T_{ext}(\{r^t\},\{q^t\}) &= \max_{F\in\mathcal{F}_{ext}} \sum_{i\in N} \hat{M}^T_{iF(i)} && (33)\\ &= \max_{j\in N} \sum_{i\in N} \hat{M}^T_{ij} && (34) \end{aligned}$$

Similarly,

$$\begin{aligned} R^T_{in}(\{r^t\},\{q^t\}) &= \max_{F\in\mathcal{F}_{in}} \sum_{k\in N} \hat{M}^T_{kF(k)} && (35)\\ &= \max_{i,j\in N} \sum_{k\in N} \hat{M}^T_{kF_{ij}(k)} && (36)\\ &= \max_{i,j\in N} \left(\hat{M}^T_{ij} + \sum_{k\neq i} \hat{M}^T_{ii} \right) && (37)\\ &= \max_{i,j\in N} \hat{M}^T_{ij} && (38) \end{aligned}$$

Line (38) follows from line (37) because for all i, $\hat{M}^T_{ii} = 0$.

For all i, j, $\sum_{i\in N} \hat{M}^T_{ij} \leq (n-1)\max_i \hat{M}^T_{ij}$, so

$$R^T_{ext}(\{r^t\},\{q^t\}) \leq (n-1) R^T_{in}(\{r^t\},\{q^t\}). \quad (39)$$

Finally,

$$\begin{aligned} R^T_{swap}(\{r^t\},\{q^t\}) &= \max_{F\in\mathcal{F}_{swap}} \sum_{i\in N} (\hat{M}^T)_{iF(i)} && (40)\\ &= \max_{\vec{h}\in N^n} \sum_{i\in N} (\hat{M}^T)_{ih_i} && (41)\\ &= \max_{h_1\in N} \max_{h_2\in N} \cdots \max_{h_n\ \in N} \sum_{i\in N} (\hat{M}^T)_{ih_i} && (42)\\ &= \sum_{i\in N} \max_{h_i\in N} \hat{M}_{ih_i} && (43)\\ &= \sum_{i\in N} \max_{j\in N} \hat{M}_{ij} && (44) \end{aligned}$$

And because for all i, j, $\sum_i \max_{i\in N} \hat{M}^T_{ij} \leq n \max_{i,j\in N} \hat{M}^T_{ij}$, we get

$$R^T_{swap}(\{r^t\},\{q^t\}) \leq n R^T_{in}(\{r^t\},\{q^t\}). \quad (45)$$

Similar calculations yeild the same relationships between M^T and $R^T_\Phi(\{r^t\},\{a^t\})$.

In summary, we have the following relationships among our varieties of regret in both the expected and observed cases:

$$\begin{aligned} R^T_{swap} &= R^T_{stochastic} && (46)\\ R^T_{in} &\leq R^T_{swap} && (47)\\ R^T_{ext} &\leq R^T_{swap} && (48)\\ R^T_{ext} &\leq (n-1)R^T_{in} && (49)\\ R^T_{swap} &\leq nR^T_{in} && (50) \end{aligned}$$

3 Regret Matching

3.1 Regret Matching Algorithms

We generalize Theorem 3 in Greenwald and Jafari (2003) to dot products with arbitrary vectors.

Theorem 4 (Fixed Point Strategy) *Let Y be a $|\Phi|$-dimensional vector of non-negative real numbers such that $\sum_{\phi\in\Phi} Y_\phi > 0$, Define an $n \times n$ dimensional matrix*

$$A(Y,\Phi) = \frac{\sum_{\phi\in\Phi} Y_\phi \phi}{\sum_{\phi\in\Phi} Y_\phi}. \quad (51)$$

Let $q(Y,\Phi)$ be a fixed point distribution of A, so $q(Y,\Phi) = q(Y,\Phi)A(Y,\Phi)$ Then $q \in Q$ and

$$\rho(r, q(Y,\Phi)) \cdot Y = 0 \quad (52)$$

Proof

Note that each ϕ is a stochastic matrix, and A is the convex combination of ϕs, so A is a stochastic matrix. Therefore it has a fixed point that is a distribution (a valid mixed strategy).

$$\begin{aligned} \rho(r, q(T,\Phi)) \cdot Y &= \sum_\phi \rho_\phi(r, q(T,\Phi))Y_\phi && (53)\\ &= \sum_\phi (r \cdot (q(T,\Phi)\phi - q(T,\Phi)))Y_\phi && (54)\\ &= \sum_\phi r \cdot (Y_\phi q(T,\Phi)\phi - Y_\phi q(T,\Phi)) && (55)\\ &= r \cdot \left(\sum_\phi Y_\phi q(T,\Phi)\phi - \sum_\phi Y_\phi q(T,\Phi) \right) && (56) \end{aligned}$$

$$= \left(\sum_{\phi} Y_{\phi}\right) r \cdot \left(q(T,\Phi)\frac{\sum_{\phi} Y_{\phi}\phi}{\sum_{\phi} Y_{\phi}} - q(T,\Phi)\right) \quad (57)$$

$$= \left(\sum_{\phi} Y_{\phi}\right) r \cdot (q(T,\Phi) - q(T,\Phi)) \quad (58)$$

$$= 0 \quad (59)$$

where line (58) follows from (57) because $q(T,\Phi)$ is the fixed point of $A(Y,\Phi)$. ∎

Definition 5 *For some Φ, let $m = |\Phi|$. If $f : \mathbb{R}^m \to (\mathbb{R}^+)^m$ (where $\mathbb{R}^+ = \{x \in \mathbb{R} : x \geq 0\}$), then an algorithm that plays*

$$q^T = q\left(f\left(\sum_{t=0}^{T-1} \rho(r^t, x^t)\right), \Phi, T\right) \quad \text{if } \left\| f\left(\sum_{t=0}^{T-1} \rho(r^t, x^t)\right)\right\|_1 > 0 \quad (60)$$

is an f,Φ,$\{x^t\}$ regret matching algorithm. *(If the 1-norm is equal to zero, the algorithm may play an arbitrary strategy.) If $\{x^t\} = \{q^t\}$ we will call it an expected regret matching algorithm; if $\{x^t\} = \{a^t\}$ we call it an observed regret matching algorithm.*

3.2 Regret Matching Bounds

We can generalize the Blackwell-style Theorem 1 in Cesa-Bianchi and Lugosi (2003) to a statement about expectations.

Theorem 6 *Let $\Psi : \mathbb{R}^m \to \mathbb{R}^+$ such that Ψ is convex and twice differentiable and can be written*

$$\Psi(u) = \sum_{i=1}^{m} \psi(u_i) \quad (61)$$

for some $\psi : \mathbb{R} \to \mathbb{R}^+$. Let $z^0, z^1, \ldots \in R^N$ be a sequence of random variables and let $Z^T = \sum_{t=0}^{T} z^t$. Let $g : \mathbb{R}^+ \to \mathbb{R}^+$ be a non-decreasing, concave, and twice differentiable function such that

$$\frac{1}{2} \sup_{u \in \mathbb{R}^m} g'(\Psi(u)) \sum_{i=1}^{m} \psi''(u_i)(z_i^t)^2 \leq C(t) \quad (62)$$

for some $C : \mathbb{N} \to \mathbb{R}$. If, for all t, $\mathbb{E}[z^t | Z^{t-1}] \cdot \nabla\Psi(Z^{t-1}) \leq B(t)$ for some $B : \mathbb{N} \to \mathbb{R}$, then for all t

$$\mathbb{E}[g(\Psi(Z^T))] - g(\Psi(0)) \leq \sum_{t=0}^{T} (B(t) + C(t)) \quad (63)$$

The proof is an adaptation of the proof of Cesa-Bianchi and Lugosi's theorem using the linearity of expectation and the law of iterated expectations.

Proposition 7 (Regret Matching Bound) *Let Ψ,ψ,g satisfy the conditions of Theorem 6. Let $f = \nabla\Psi$. If*

$$\frac{1}{2}\sup_{u\in\mathbb{R}^m} g'(\Psi(u))\sum_{\phi}\psi''(u_\phi)(\rho_\phi(r^t,x^t))^2 \leq C(t) \tag{64}$$

an f,Φ-expected regret matching algorithm has the property

$$g\left(\Psi\left(\sum_{t=0}^{T}\vec{\rho}(r^t,q^t)\right)\right) \leq g(\Psi(0)) + \sum_{t=0}^{T} C(t) \tag{65}$$

and an f,Φ-observed regret matching algorithm has the property

$$\mathbb{E}\left[g\left(\Psi\left(\sum_{t=0}^{T}\vec{\rho}(r^t,a^t)\right)\right)\right] \leq g(\Psi(0)) + \sum_{t=0}^{T} C(t). \tag{66}$$

Proof

Apply Theorem 6 with $z^t = \rho(r^t,x^t)$. We must satisfy $\mathbb{E}[z^t|Z^{t-1}]\cdot\nabla\Psi(Z^{t-1}) \leq B(t)$ for some $B:\mathbb{N}\to\mathbb{R}$. In the case of expected regret matching $\mathbb{E}[\rho(r^t,x^t)|Z^{t-1}] = \mathbb{E}[\rho(r^t,q^t)|q^t] = \rho(r^t,q^t)$. In the case of observed regret matching $\mathbb{E}[\rho(r^t,x^t)|Z^{t-1}] = \mathbb{E}[\rho(r^t,q^t)|q^t] = \rho(r^t,q^t)$. If

$$\left\| f\left(\sum_{t=0}^{T-1}\rho(r^t,x^t)\right)\right\|_1 > 0 \tag{67}$$

then the definition of regret matching and Theorem 4 give us that

$$\begin{aligned} 0 &= \rho(r^t,q^t)\cdot f\left(\sum_{t-0}^{T-1}\rho(r^t,x^t)\right) && (68)\\ &= \mathbb{E}[\rho(r^t,x^t)|Z^{t-1}]\cdot f\left(\sum_{t-0}^{T-1}\rho(r^t,x^t)\right) && (69)\\ &= \mathbb{E}[z^t|Z^{t-1}]\cdot\nabla\Psi(Z^{t-1}). && (70)\end{aligned}$$

Otherwise, $f\left(\sum_{t=0}^{T-1}\rho(r^t,x^t)\right) = 0$, and so the dot product must also be zero. Therefore we can set $B(t) = 0$, giving the desired result. If $\{x^t\} = \{q^t\}$ then we have

$$\mathbb{E}\left[g\left(\Psi\left(\sum_{t=0}^{T}\vec{\rho}(r^t,q^t)\right)\right)\right] = g\left(\Psi\left(\sum_{t=0}^{T}\vec{\rho}(r^t,q^t)\right)\right) \quad \blacksquare \tag{71}$$

3.3 Specific Regret Matching Algorithms

Let a^+ denote $\max\{a,0\}$. Following Corollary 1 in Cesa-Bianchi and Lugosi, we have

Corollary 8 (Polynomial Regret Matching) *Let $f(u) = p(u_i^+)^{p-1}$ for $p > 2$. Then an f,Φ-expected regret matching algorithm has the property*

$$R_\Phi^T(\{r^t\}, \{q^t\}) \leq \sqrt{(p-1)\sum_{t=0}^{T} \|\rho^t(r^t, q^t)\|_p^2} \tag{72}$$

In particular, for both Φ_{in} and Φ_{ext}

$$R_\Phi^T(\{r^t\}, \{q^t\}) \leq \sqrt{(p-1)nT}. \tag{73}$$

An f,Φ-observed regret matching algorithm has the property

$$\mathbb{E}\left[R_\Phi^T(\{r^t\}, \{q^t\})\right] \leq \sqrt{(p-1)\sum_{t=0}^{T} \|\rho(r^t, q^t)\|_p^2} \tag{74}$$

In particular, for both Φ_{in} and Φ_{ext}

$$\mathbb{E}\left[R_\Phi^T(\{r^t\}, \{a^t\})\right] \leq \sqrt{(p-1)nT}. \tag{75}$$

Proof

The proof for expected regret matching is similar to the proof of Corollary 1 in Cesa-Bianchi and Lugosi. To prove the observed case, use Jensen's inequality to get $\left(\mathbb{E}\left[R_\Phi^T(\{r^t\}, \{a^t\})\right]\right)^2 \leq \mathbb{E}\left[\left(R_\Phi^T(\{r^t\}, \{a^t\})\right)^2\right]$. ∎

Following Corollary 2 in Cesa-Bianchi and Lugosi, we have

Corollary 9 (Exponential Regret Matching) *Let $f(u) = \eta e^{\eta u_i}$ where $\eta > 0$. For some Φ, let $m = |\Phi|$. Then an f,Φ-expected regret matching algorithm has the property*

$$R_\Phi^T(\{r^t\}, \{q^t\}) \leq \frac{\ln m}{\eta} + \frac{\eta t}{2} \tag{76}$$

An f,Φ-observed regret matching algorithm has the property

$$\mathbb{E}[R_\Phi^T(\{r^t\}, \{a^t\})] \leq \frac{\ln m}{\eta} + \frac{\eta t}{2} \tag{77}$$

The proof is similar to Cesa-Bianchi and Lugosi's proof.

3.4 Implementation of Internal and External Regret Matching

Note that for Φ_{in},

$$A_{ij}(Y, \Phi_{in})) = \frac{\sum_{k,l \in N} Y_{kl} \delta_j^{F_{kl}(i)}}{\sum_{k,l \in N} Y_{kl}} \tag{78}$$

$$= \frac{\sum_{l \in N} \left(Y_{il} \delta_j^l + \sum_{k \neq i} Y_{kl} \delta_j^l \right)}{\sum_{k,l \in N} Y_{kl}} \tag{79}$$

$$= \frac{Y_{ij} + \delta_j^i \left(\sum_{l \in N} \sum_{k \neq i} Y_{kl} \right)}{\sum_{k,l \in N} Y_{kl}} \tag{80}$$

For $\phi \sim F_{ij} \in \mathcal{F}_{in}$, $\rho_\phi(r^t, a^t) = m_{ij}^t$ and $\rho_\phi(r^t, q^t) = \hat{m}_{ij}^t$, so $\sum_{t=0}^T \rho_\phi(r^t, a^t) = M_{ij}^T$ and $\sum_{t=0}^T \rho_\phi(r^t, q^t) = \hat{M}_{ij}^T$. Therefore an f,internal-expected regret matching algorithm plays the fixed point of

$$A_{ij}(f(\hat{M}), \Phi_{in})) = \frac{f_{ij}(\hat{M}) + \delta_j^i \left(\sum_{l \in N} \sum_{k \neq i} f_{kl}(\hat{M}) \right)}{\sum_{k,l \in N} f_{kl}(\hat{M})} \tag{81}$$

and an f,internal-observed regret matching algorithm plays the fixed point of

$$A_{ij}(f(M), \Phi_{in})) = \frac{f_{ij}(M) + \delta_j^i \left(\sum_{l \in N} \sum_{k \neq i} f_{kl}(M) \right)}{\sum_{k,l \in N} f_{kl}(M)} \tag{82}$$

when the matrix is well-defined.

Also,

$$A_{ij}(Y, \Phi_{ext})) = \frac{\sum_{k \in N} Y_k \delta_j^k}{\sum_{k \in N} Y_k} \tag{83}$$

$$= \frac{Y_j}{\sum_{k \in N} Y_k} \tag{84}$$

and calculating the fixed point equation:

$$\sum_{k \in N} q_k(Y, \Phi_{ext}) A_{kj}(Y, \Phi_{ext}) = q_j(Y, \Phi_{ext}) \tag{85}$$

$$\sum_{k \in N} q_k(Y, \Phi_{ext}) \frac{Y_j}{\sum_{l \in N} Y_l} = q_j(Y, \Phi_{ext}) \tag{86}$$

$$\frac{Y_j}{\sum_{l \in N} Y_l} = q_j(Y, \Phi_{ext}) \tag{87}$$

For $\phi \sim F_i \in \mathcal{F}_{ext}$, $\rho_\phi(r, x) = r_i - r \cdot x$. In particular $\rho_\phi(r^t, a^t) = \sum_{j \in N} m_{ij}^t$ and $\rho_\phi(r^t, q^t) = \sum_{j \in N} \hat{m}_{ij}^t$. Therefore an f,external-expected regret matching algorithm plays

$$q_i^t = \frac{\sum_{j \in N} f_{ij}(\hat{M}^t)}{\sum_{i,j \in N} f_{ij}(\hat{M}^t)} \tag{88}$$

and an f,external-observed regret matching algorithm plays

$$q_i^t = \frac{\sum_{j \in N} f_{ij}(M^t)}{\sum_{i,j \in N} f_{ij}(M^t)} \tag{89}$$

when the vector is well-defined.

4 No Regret

4.1 Definitions

Definition 10 *An algorithm is said to be* no-expected-Φ-regret *if for all* $\{r^t\}$, $R_\Phi^T(\{r^t\}, \{q^t\}) \in o(T)$, *or equivalently if* $\limsup_{T\to\infty} \frac{1}{T} R_\Phi^T(\{r^t\}, \{q^t\}) \leq 0$.

Cessa-Bianchi and Lugosi call an algorithm that is no-expected-external-regret *Hannan consistent.*

Definition 11 *An algorithm is said to be* no-obseved-Φ-regret in expectation *if for all* $\{r^t\}$, $\mathbb{E}[R_\Phi^T(\{r^t\}, \{a^t\})] \in o(T)$, *or equivalently if* $\limsup_{T\to\infty} \mathbb{E}[\frac{1}{T} R_\Phi^T(\{r^t\}, \{a^t\})] \leq 0$.

Proposition 12 *If an algorithm is no-observed-*Φ*-regret in expectation, then it is no-expected-*Φ*-regret.*

Proof

This is a direct consequence of Proposition 7. ∎

Because swap regret bounds any Φ-regret from above, no-expected-swap-regret implies any other no-expected-regret. Because internal regret bounds swap regret by a constant factor, no-expected-interal-regret implies no-expected-swap-regret. Therefore no-expected-internal-regret is the strongest version of no-expected-regret, and it implies no-expected-external regret. The analogous statements also hold for no-observed-Φ-regret in expectation.

4.2 Examples

Polynomial regret matching gives a bound that is $O(\sqrt{nT})$, and therefore such algorithms will have no-regret properties. Exponential regret matching can also give a sub-linear bound with the proper setting of the parameter η. For example setting $\eta = \sqrt{2 \ln m / T}$, where

$m = |\Phi|$, yields a bound of $\sqrt{2T \ln m}$. However, this method requires foreknowledge of the horizon T.

The case where the horizon T is not known in advance may be addressed by a scheme partitioning the time sequence into S "epochs" where the sth epoch is of length T_s. Moreover, at the beginning of the sth epoch, η_s is reset appropriately and the algorithm is restarted. This is similar to the scheme proposed in Freund and Schapire (1999), only we consider a different partitioning of the sequence. In particular, the sequence may be divided into $\lceil \log T \rceil$ epochs where the sth epoch, $T_s = 2^{s-1}$, consists of actions $2^{s-1}, \ldots, (2^s - 1)$ for $s > 0$ and $T_0 = 0$. Setting $\eta_s = \sqrt{2 \ln m / T_s}$ then the regret for epoch s is $R \leq \sqrt{2^s \ln m}$. The regret over all epochs is:

$$
\begin{aligned}
R &\leq \sum_{s=0}^{\lceil \log T \rceil} \sqrt{2 T_s \ln m} && (90)\\
&= \sum_{s=0}^{\lceil \log T \rceil} \sqrt{2^s \ln m} && (91)\\
&= \sqrt{\ln m} \sum_{s=0}^{\lceil \log T \rceil} 2^{s/2} && (92)\\
&= \sqrt{\ln m} \left(\sqrt{2^{\lceil \log T \rceil + 2}} + \sqrt{2^{\lceil \log T \rceil + 1}} - 1 - \sqrt{2} \right) && (93)\\
&\leq \sqrt{\ln m} \left(\sqrt{2^{\log T + 3}} + \sqrt{2^{\log T + 2}} - 1 - \sqrt{2} \right) && (94)\\
&= \sqrt{\ln m} \left(\sqrt{8T} + \sqrt{4T} - 1 - \sqrt{2} \right) && (95)\\
&= O(\sqrt{T \ln m}) && (96)
\end{aligned}
$$

where (92) uses the closed form solution for $S_K = \sum_{s=0}^{K} 2^{s/2}$ obtained as follows:

$$
\begin{aligned}
S_K + 2^{\frac{K+1}{2}} &= 2^0 + \sum_{i=0}^{K} 2^{\frac{i+1}{2}} && (97)\\
S_K &= -2^{\frac{K+1}{2}} + 1 + \sum_{i=0}^{K} 2^{\frac{i+1}{2}} && (98)\\
S_K &= -2^{\frac{K+1}{2}} + 1 + \sum_{i=0}^{K} 2^{\frac{i+1}{2}} && (99)\\
S_K &= -2^{\frac{K+1}{2}} + 1 + \sqrt{2} \sum_{i=0}^{K} 2^{\frac{i}{2}} && (100)\\
S_K &= -2^{\frac{K+1}{2}} + 1 + \sqrt{2} S_K && (101)\\
\left(1 - \sqrt{2}\right) S_K &= 1 - 2^{\frac{K+1}{2}} && (102)\\
S_K &= \frac{1 - 2^{\frac{K+1}{2}}}{\left(1 - \sqrt{2}\right)} && (103)
\end{aligned}
$$

$$S_K = \frac{\left(1 - \sqrt{2^{K+1}}\right)\left(1 + \sqrt{2}\right)}{-1} \quad (104)$$

$$S_K = \sqrt{2^{K+2}} + \sqrt{2^{K+1}} - 1 - \sqrt{2} \quad (105)$$

If $|\Phi|$ is polynomial in n, this bound is superior to the polynomial matching bound for large n.

5 Conclusion

We have developed an analysis of the expectation of observed regret, which is more powerful than expected regret. This has allowed us to derive two algorithms which achieve no-observed-internal-regret, which is the strongest version of no-regret that can be derived from time-invariant linear transformations. However, this property is only achieved in expectation; we would like to extend the analysis to be able to bound the probability of observed regret exceeding a particular value.

Bibliography

Nicolo Cesa-Bianchi and Gabor Lugosi (2003). Potential-Based Algorithms in On-Line Prediction and Game Theory. *Machine Learning*, 51, 239-261.

Yoav Freund and Robert E. Schapire (1999). Adaptive Game Playing Using Multiplicative Weights. *Games and Economic Behavior*, 29, 79-103.

Amy Greenwald and Amir Jafari (2003). A General Class of No-Regret Learning Algorithms and Game-Theoretic Equilibria. In *Proceedings of the 16th Annual Conference on Computational Learning Theory.*