Fictitious Play in Opinion Formation Games with Random Payoffs

Dimitris Fotakis NTUA
Vardis Kandiros NTUA
Vasilis Kontonis NTUA
Stratis Skoulakis NTUA

June 15, 2018

Abstract

We study opinion formation games based on the famous model proposed by Friedkin and Johsen. In today's huge social networks the assumption that in each round all agents update their opinions by taking into account the opinions of *all* their friends could be unrealistic. Therefore, we assume that in each round each agent gets to meet with only one (uniformly random) friend of hers. Specifically, we define an opinion formation game, where at round t, agent i with intrinsic opinion $s_i \in [0,1]$ and expressed opinion $x_i(t) \in [0,1]$ meets a random neighbor j with opinion $x_j(t)$ and suffers cost that is a convex combination of $(x_i(t) - s_i)^2$ and $(x_i(t) - x_j(t))^2$.

For a dynamics in the above setting to be considered as natural it must be simple, converge to the equilibrium x^*, and perhaps most importantly, it must be a reasonable choice for selfish agents. In this work we show that *fictitious play*, is a natural dynamics for the above game. We prove that, after $O(1/\varepsilon^2)$ rounds, the opinion vector is within error ε of the equilibrium. Moreover, we show that fictitious play admits no-regret and thus, is a reasonable algorithm for agents to implement.

The classical Friedkin-Johsen dynamics converges to the equilibrium within error ε after only $O(\log(1/\varepsilon))$ rounds whereas in our setting fictitious play needs $\widetilde{O}(1/\varepsilon^2)$ rounds. A natural question is whether there exists a natural dynamics for our problem with better rate of convergence. To answer this question we show that from *no-regret* dynamics in our setting, we can construct algorithms that estimate the mean of Bernoulli random variables. By using information theoretic sample complexity lower bound techniques from statistics we show that no-regret algorithms are unlikely to converge with less than $O(1/\varepsilon)$ rounds. Interestingly, the considerably worse convergence rate is not an inherent characteristic of our setting but rather of the class of no-regret dynamics since we show that, in general, there exists a dynamics that only needs $O(\log(1/\varepsilon)^2)$ rounds in our setting, resembling the convergence rate of the original Friedkin-Johsen model.

1 Introduction

1.1 Friedkin-Johsen Model and Opinion Formation Games

In the Friedkin-Johsen model (FJ-model) an undirected graph $G(V, E)$ with n nodes, is assumed, where V denotes the agents and E the social relations between them. Each agent i poses an internal opinion $s_i \in [0, 1]$ and a self confidence coefficient $\alpha_i \in [0, 1]$. At each round $t \geq 1$, agent i updates her opinion as follows:

$$x_i(t) = (1 - \alpha_i)\frac{\sum_{j \in N_i} x_j(t-1)}{|N_i|} + \alpha_i s_i, \tag{1}$$

where N_i is the set of her neighbors. The simplicity of the update rule (1) makes the FJ-model plausible, because in real social networks it is very unlikely that agents change their opinions according to complex rules. Based on the FJ-model, in [BKO11], they propose an *opinion formation game* in which the strategy that each agent i plays, is the opinion $x_i \in [0, 1]$ that she publicly expresses incurring her a cost

$$C_i(x_i, x_{-i}) = (1 - a_i)\sum_{j \in N_i}(x_i - x_j)^2 + a_i(x_i - s_i)^2. \tag{2}$$

In [**?**] they prove that it always admits a unique Nash Equilibrium x^* and has Price of Anarchy 9/8 if G is undirected and $O(n)$ in the directed case. We denote an instance of this game as $I = (G, s, a)$. The FJ-model can be viewed as the best-response dynamics of the opinion formation game. More precisely, each agent i has an opinion $x_1(0), \ldots, x_n(0)$ and at round $t \geq 1$, suffers the cost (3) and updates her opinion so as to minimize her individual cost

$$x_i(t) = \underset{x \in [0,1]}{\operatorname{argmin}}(1 - a_i)\sum_{j \in N_i}(x - x_j(t-1))^2 + a_i(x - s_i)^2. \tag{3}$$

In [GS14] it is proved that for any instance $I = (G, s, a)$ the opinion vector $x(t) = (x_1(t), \ldots, x_n(t))$ converges to the unique equilibrium point x^*. Moreover, this is achieved by an update rule that is very simple but more importantly it is *rational*, in the sense that each agent i adopts this rule in order to minimize her individual cost.

1.2 Opinion Formation Games with Random Payoffs

In the game defined by (2), agent's i cost $C_i(x_i, x_{-i})$ is a deterministic function of the opinion vector x. Many recent works (see e.g. [ZLZ17]) study games with random payoffs, that is agent's i cost $(C_i(x_i, x_{-i}))$ is a random variable. The random payoff setting can be much more realistic, since randomness may naturally occur because of incomplete information, noise or other stochastic factors. Motivated by this line of research we introduce a random payoff variant of the opinion formation game (2).

Definition 1. *Let $I = (G, s, a)$ an instance of the opinion formation game and x the opinion vector. Each agent i,*

- *picks uniformly at random one of her neighbors $j \in N_i$*
- *suffers cost $C_i(x_i, x_{-i}) = (1 - a_i)\sum_{j \in N_i}(x_i - x_j)^2 + a_i(x_i - s_i)^2$*

This game is more compatible to a realistic setting since in real world networks (e.g. Facebook, Twitter e.t.c.), each agent may have several hundreds of friends. As a result it is far more reasonable to assume that every day, each agent meets a random small subset of her acquaintances and suffers a cost based on how much she disagrees with them. Since the expected cost of our random payoff variant equals the cost of game (2), $\mathbf{E}[C(x_i, x_{-i})] = (1 - a_i)\sum_{j \in N_i}(x_i - x_j)^2 + a_i(x_i - s_i)^2$, the results about the existence of a unique equilibrium, and the PoA bounds also hold in our variant. However, the best-response dynamics in the random payoff game does not converge to the equilibrium. In this work, we investigate whether there exists natural dynamics that leads the system to the equilibrium point x^*.

1.3 Our Results and Techniques

In this work we propose Algorithm 1 as a dynamics for the random payoff game of Definition 1. Note that the dynamics described in Algorithm 1 is the *fictitious play* in the game defined by the instance $I = (G, s, a)$. Generally speaking *fictitious play* does not guarantee convergence to the equilibrium. In Section 3 we show

---

Algorithm 1 Fictitious Play

1: Initially, each agent i has an opinion $x_1(0), \ldots, x_n(0)$.
2: At round $t \geq 1$, each agent i meets uniformly at random one of her friends $j \in N_i$
3: Suffers cost $C_i^t(x_i(t-1), x_j(t-1)) = (1-a_i)(x_i(t-1) - x_j(t-1))^2 + a_i(x_i(t-1) - s_i)^2$
4: Updates opinion

$$x_i(t) = \operatorname*{argmin}_{x \in [0,1]} \sum_{\tau=1}^{t} C_i^\tau(x, x_j(\tau-1)) \tag{4}$$

---

that, for our game with random payoffs, fictitious play converges to equilibrium, with the following rate.

Theorem 1. *Let $I = (G, s, \alpha)$ be any instance of the opinion formation game of Definition 1 with equilibrium $x^* \in [0,1]^n$. The opinion vector x^t produced by Algorithm 1 after t rounds satisfies*

$$\mathbf{E}\left[\|x^t - x^*\|_\infty\right] \leq C\sqrt{\log n}\frac{(\log t)^2}{t^{\min(1/2,\rho)}},$$

where $\rho = \min_{i \in V} a_i$ and C is a universal constant.

The update rule (4) guarantees convergence while vastly reducing the information exchange between the agents at each round. Namely, each agent learns the opinion of only one of her neighbors whereas in the classical FJ-model (1) each agent learns the opinions of all of her neighbors. In terms of total communication needed to get within distance ε of the equilibrium, the update rule (4) needs $O(|V| \log |V|)$ communication while (1) needs $O(|E|)$. Of course for this difference to be significant we need each agent to have at least $O(\log |V|)$ neighbors. A large social network like Facebook has approximately 2 billion users and each user has usually more that 100 friends which far more than $\log(2\ 10^9)$.

Apart from converging to the equilibrium, our update rule (4) is also a rational behavioral assumption since it ensures *no-regret* for the agents. Having no-regret means that the average cost for each agent i after T rounds is close to the average cost that she would suffer by expressing any fixed opinion. This is a very important feature of our update rule because even players that selfishly will to minimize their incurred cost, could choose to play according to it. In Section 4 we show the following theorem.

Theorem 2. *For every instance $I = (G, s, a)$, for every agent i*

$$\sum_{\tau=1}^{t} C_i^\tau(x_i(\tau)) \leq min_{x\in[0,1]} \sum_{\tau=1}^{t} C_i^\tau(x) + O(\log t)$$

Even though our update rule (4) has the above desired properties it only achieves convergence rate of $\widetilde{O}(1/\sqrt{t})$ for a fixed instance I with $\rho > 1/2$. For such an instance the original FJ-model outperforms our update rule since it achieves convergence rate $O(1/2^t)$. We investigate whether this gap is due to an inherent characteristic of the random payoff setting i.e. learning the opinion of just one random neighbor vs learning all of them or we could find another natural dynamics that converges exponentially fast. Can the agents select another no-regret algorithm that guarantees an asymptotically faster convergence rate? In Section 5 we investigate the following question

Question. *Is there a no-regret algorithm that the agents can choose such that for any instance $I = (G, s, a)$, $\mathbf{E}[\|x_t - x^*\|_\infty] = O(1/t)$?*

To answer this question we first show that the existence of a no-regret algorithm (that achieves this convergence rate to the equilibrium) implies the existence of an algorithm that uses i.i.d samples from a Bernoulli random variable $B(p)$ to estimate its success probability p with the same asymptotic error rate. We use sample complexity lower bound techniques from statistics to show that such a no-regret algorithm is unlikely to exist. We show that updating the opinion according to no-regret algorithms leads to update rules that are totally ignorant of the specific graph structure G. Perhaps surprisingly, in Section 6 we present an update rule that takes into account only the maximum degree of the graph and for any fixed instance I achieves convergence rate $O(1/2^{\sqrt{t}})$.

1.4 Related Work

Our work belongs to the line of work studying the seminal Friedkin-Jonhsen model [FJ90]. Bindel et al. in [BKO11] defined an opinion formation game based on the FJ-model and bounded the inefficiency of its equilibrium point with respect to the total disagreement cost. Subsequent work bounded the inefficiency of its equilibrium in variants of the latter game [BGM13, EFHS17, CKO13, BFM16]. In [GS14] they show that the convergence time depends on the spectral radius of the adjacency matrix of the graph G and provided bounds in special graph topologies. In [BFM16], a variant of the opinion formation game in which social relations depend on the expressed opinions, is studied. They prove that, the discretized version of the above game admits a potential function and thus best-response converges to the Nash equilibrium. Convergence results in other discretized variants of the FJ-model can be found in [YOA+13, FGV16].

In [FV97], [FS99], [SA] they prove that in a finite if each agent updated her mixed strategy according to a no-regret algorithm the resulting time-averaged distribution converges to Coarse Correlated Equilibrium. In the same spirit in [BEDL06] they proved that no-regret dynamics converge to NE in the case of congestion games. Later in [EMN09] they studied no regret dynamics in games with infinite strategy space. They proved that for a large class of games with concave utility function (socially concave games), the time-averaged strategy vector converges to the pure Nash equilibrium. More recent work investigate a stronger notion of convergence of no-regret dynamics. In [CHM17] they show that, in n-person finite generic games that admit unique Nash equilibrium, the strategy vector converges *locally* and exponentially fast to the PNE. They also provide conditions for *global* convergence. Our results fit in this line of research since we show that for a game with *infinite* strategy space, the strategy vector (not the time-averaged) converges to the unique Nash equilibrium.

2 Preliminaries

2.1 Notation

Let $G(V,E)$ be a graph. We denote by V the set of agents, and $N_i \subset V$ the set of neighbors of agent i. Let $n = |V|$. We denote by $x(t) \in [0,1]^n$ the vector of opinions of the agents at round t and $x_i(t)$ its i-th coordinate. We denote by $Q([0,1])$ the set of rationals in $[0,1]$.

2.2 Online Convex Optimization and No Regret Algorithms

The *Online Convex Optimization* (OCO) framework can be seen as a game played between a player and an adversary. Let $K \subseteq \mathbf{R}^n$ be a convex set and a set of functions $\mathcal{F}$ defined over K. *At round t*,

1. *the agent chooses $x_t \in K$.*
2. *the adversary observes the x_t and selects a function $f_t(x) \in \mathcal{F}$.*
3. *the player receives cost $f_t(x_t)$.*

The goal of the player is to pick x_t based on the history $(f_1, \ldots, f_{t-1})$ in a way that minimizes the total cost. We emphasize that the agent has to choose x_t before seeing f_t, otherwise the problem becomes trivial.

Definition 2. *An OCO algorithm A, at round t selects a vector $x^t \in K$ according to the history, $x^t = A_t(f_1, \dots, f_t)$.*

Definition 3. *Let the OCO algorithm A and the sequence of functions $\{f_1, \dots, f_T\} \in F^T$. The regret of A is defined by*

$$R_A(T) = \sum_{t=1}^{T} f_t(x_t) - min_{x \in K} \sum_{t=1}^{T} f_t(x)$$

If $R_A(T) = o(T)$ for any sequence $\{f_1, \dots, f_T\}$ then A is no-regret algorithm

According to the feasibility set K and the set of functions F, different no-regret algorithms with differnt regret bounds can be derived. Some seminal examples are Zinkevic's algorithm that achieves regret $R_A(T) = O\left(\sqrt{T}\right)$, when K is convex, closed and bounded and F is the set of convex functions with bounded first derivative. Hazan et al [] proposed an algorithm with $R_A(T) = O(\log T)$, when F is the set of twice diffentiable strongly convex functions.

No regret dynamics and repeated games: Agent i with s_i, a_i is the player. At round t, she adopts an opinion $x_i(t) \in [0,1]$ and then the adversary selects a function $C_i^t(x) = (1-a_i)(x-y_t)^2 + a_i(x-s_i)^2$, where $y^t \in [0,1]$. Let A a no-regret in the above OCO setting. Clearly, each agent i is willing to adopt as $x_i(t)$ the suggestion of A, ensuring that average cost is similar to the average cost of the best fixed opinion. As a result if all agents adopt the no-regret algorithm A, for a fixed instance I of the game the following dynamical process is defined.

Algorithm 2 no-regret dynamics

Let an instance $I = (G, s, a)$ of the opinion formation game and $x_1(0), \dots, x_n(0)$ the initial opinions.
At round $t \geq 1$, each agent i:
1: Adopts an opinion $x_i(t) = A_t(C_i^1, \dots, C_i^{t-1})$
2: Meets uniformly at random one of her friends $j \in N_i$
3: Suffers cost $C_i^t(x_i(t))$, where $C_i^t(x) = (1-a_i)(x - x_j(t))^2 + a_i(x-s_i)^2$

We denote as $x_A(t)$ the opinion vector when the algorithm A is selected. In case $A_t(C_i^1, \dots, C_i^{t-1}) = \text{argmin}_{x\in[0,1]} \sum_{\tau=1}^{t-1} C_i^t(x)$ we obtain the *fictitious play* defined in 1 and for simplicity we denote it as $x(t)$. As we have already menioned, in Section 3 we bound the convergence rate of the *fictitious play*, $\mathbf{E}\left[\|x(t) - x^*\|_\infty\right]$ and in Section 4, we prove that it is no-regret algorithm for the specific OCO setting. The latter does not hold in more general OCO settings. In section 4, we investigate whether there exists an algorithm A with significantly better asymptotic rate of convergence.

2.3 Estimating Bernoulli distributions

Assume that we have access to the random variables $Y_1, \dots, Y_t$, that are mutually independent and each Y_i is distributed according to a Bernoulli distribution with $p \in [0,1]$, $Y_i \sim B(p)$. Initially the parameter p is unkown. The task is to observe the realization of $y_1, \dots, y_t$ and output an estimate $\hat{p} = \theta_t(y_1, \dots, y_t)$ according to a function $\hat{\theta}_t : \{0,1\}^t \mapsto [0,1]$, that is close to the unkown parameter p.

Definition 4. *An estimator $\hat{\theta}$ is a collection of functions, $\{\hat{\theta}_t\}_{t=1}^\infty$, where $\hat{\theta}_t : \{0,1\}^t \mapsto [0,1]$.*

A well known example of such an estimator is the *sample mean*, where $\hat{\theta}_t = \frac{\sum_{i=1}^t Y_i}{t}$. Obviously for any efficient estimator, the estimate $\hat{p}$ converges to the unkown parameter p as t grows. The following definition is the standard metric for the efficiency of an estimator $\hat{\theta}$.

Definition 5. *For an estimator $\hat{\theta} = \{\hat{\theta}_t\}_{t=1}^\infty$ we define its error rate $R_p(t) = E_p[|\hat{\theta}_t(Y_1, \dots, Y_t) - p|]$*

$$\textit{where } E_p[|\hat{\theta}_t(Y_1, \dots, Y_t) - p|] = \sum_{(y_1, \dots, y^t) \in \{0,1\}^t} |\hat{\theta}_t(y^1, \dots, y^t) - p| \cdot p^{\sum_{i=1}^t y^i}(1-p)^{t - \sum_{i=1}^t y^i}$$

The quantity $E_p[|\hat{\theta}_t(Y_1,\dots,Y_t)-p|]$ is the expected distance of the estimated value $\hat{\theta}_t$ from the parameter p, when the distribution of the samples is $B(p)$. To simplify notation we also denote it as $E_p[|\hat{\theta}_t - p|]$. The error rate $R_p(t)$ quantifies the rate of convergence of the estimated value $\hat{p} = \theta_t(Y_1,\dots,Y_t)$ to the real parameter p.
Since p is unkown, any meaningful estimator $\hat{p}$ must guarantee that for all $p \in [0,1]$, $\lim_{t\to\infty} R_p(t) = 0$. For example the *sample mean* has error rate $R_p(t) \le \frac{1}{2\sqrt{t}}$ for ant $p \in [0,1]$ and clearly statisfies the above requirement. In section 5 we investigate the following question,

Question. *Is there an estimator $\hat{\theta}$ such that for all $p \in Q[0,1]$, $\lim_{t\to\infty} tR_p(t) = 0$?*

3 Fictitious Play Convergence Rate

In this section we prove that the fictitious play described as Algorithm 1 convergences to the equilibrium. We shall use the following standard concentration inequality of Hoeffding.

Lemma 1. *Let $X_1,\dots,X_t$ be independent random variables such that $0 \le X_i \le 1$ and let $X = (X_1+\dots+X_t)/t$. Then for all $t>0$, $\mathbf{P}\left[|X - \mathbf{E}\left[X\right]| \ge \lambda\right] \le 2\mathrm{e}^{-2n\lambda^2}$*

We break down the proof of Theorem **??** into two parts. We first recursively define an upper bound $e(t)$ of the error $\|x^t - x^*\|_\infty$ that holds with arbitrarily good probability $1-p$ for all $t \ge 1$. Then we work with the recursion in order to get an upper bound for $e(t)$. The following lemma provides a simple upper bound for the convergence rate of the recursion. Its technical proof can be found in Section A of the Appendix.

Lemma 2. *Let $e(t)$ be a function satisfying the recursion*

$$e(t) = \delta(t) + (1-\alpha)\frac{\sum_{\tau=0}^{t-1} e(\tau)}{t} \text{ and } e(0) = \|x^0 - x^*\|_\infty,$$

where $\delta(t) = \sqrt{\frac{\ln(Dt^2)}{t}}$, $\delta(0)=0$, and $D > \mathrm{e}^2$ is a positive constant. Then $e(t) \le \sqrt{2\ln(D)}\frac{(\ln t)^{3/2}}{t^{\min(\rho,1/2)}}$.

We are now ready to fully prove Theorem **??**. We restate it here for the sake of completeness.

Theorem 3. *Let $I = (G,s,\alpha)$ be any instance of the opinion formation game of Definition 1 with equilibrium $x^* \in [0,1]^n$. The opinion vector x^t produced by Algorithm 1 after t rounds satisfies*

$$\mathbf{E}\left[\|x^t - x^*\|_\infty\right] \le C\sqrt{\log n}\frac{(\log t)^2}{t^{\min(1/2,\rho)}},$$

where $\rho = \min_{i\in V} a_i$ and C is a universal constant.

Proof. Let W_i^t be the random variable corresponding to the selected neighbor of agent i, at round t of the dynamics 1. Therefore, each agent $i \in V$ updates her opinion as follows

$$x_i(t) = \operatorname*{argmin}_{x\in[0,1]} \sum_{\tau=1}^{t} C_i^\tau(x, x_{W_i^\tau}(\tau-1)) = (1-\alpha_i)\frac{\sum_{\tau=1}^t x_{W_i^\tau}(\tau-1)}{t} + \alpha_i s_i$$

We want to find a bound $e(t)$ for the random variable of the error $\|x^t - x^*\|_\infty$ such that with probability $1-p$ it holds for all $t\ge 1$, $\|x^t - x^*\|_\infty \le e(t)$. We derive a recursion for the error $e(t)$ as follows. As we have already mentioned for any instance I there exists a unique equilibrium vector x^*. Since $W_i^\tau \sim \mathrm{U}(N_i)$ we have that $\mathbf{E}\left[x^*_{W_i^\tau}\right] = \frac{\sum_{j\in N_i} x_j^*}{|N_i|}$. Since W_i^τ are independent random variables, we can use Hoeffding's inequality (Lemma 1) to get $\mathbf{P}\left[|(\sum_{\tau=1}^t x^*_{W_i^\tau})/t - (\sum_{j\in N_i} x_j^*)/|N_i| > \delta(t)\right] < 6p/(\pi^2 n t^2)$, where $\delta(t) = \sqrt{\frac{\log(\pi^2 n t^2/(6p))}{t}}$.

To bound the probability of error for all rounds $t = 1$ to ∞ and all agents in V, we use the union bound to get

$$\sum_{i=1}^{\infty} \mathbf{P}\left[\max_{i\in V}\left|\frac{\sum_{\tau=1}^{t} x^*_{W_i^\tau}}{t} - \frac{\sum_{j\in N_i} x^*_j}{|N_i|}\right| > \delta(t)\right] \le \sum_{i=1}^{\infty}\frac{6}{\pi^2}\frac{1}{t^2}\sum_{i=1}^{n}\frac{p}{n} = p$$

As a result with probability $1-p$ we have that for all t and all $i \in V$

$$\left|\frac{\sum_{\tau=1}^{t} x^*_{W_i^\tau}}{t} - \frac{\sum_{j\in N_i} x^*_j}{|N_i|}\right| \le \delta(t) \tag{5}$$

We will prove our claim by induction. We assume that $\|x^\tau - x^*\|_\infty \le e(\tau)$ for all $\tau \le t-1$. Then

$$\begin{aligned}
x_i(t) &= (1-\alpha_i)\frac{\sum_{\tau=1}^{t} x_{W_i^\tau}(\tau-1)}{t} + \alpha_i s_i \\
&\le (1-\alpha_i)\frac{\sum_{\tau=1}^{t} x^*_{W_i^\tau} + \sum_{\tau=1}^{t} e(\tau-1)}{t} + \alpha_i s_i && (6)\\
&\le (1-\alpha_i)\left(\frac{\sum_{\tau=1}^{t} x^*_{W_i^\tau}}{t} + \frac{\sum_{\tau=0}^{t-1} e(\tau)}{t}\right) + \alpha_i s_i \\
&\le (1-\alpha_i)\left(\frac{\sum_{j\in N_i} x^*_j}{|N_i|} + \delta(t) + \frac{\sum_{\tau=0}^{t-1} e(\tau)}{t}\right) + \alpha_i s_i && (7)\\
&\le x^*_i + \delta(t) + (1-\alpha)\left(\frac{\sum_{\tau=0}^{t-1} e(\tau)}{t}\right)
\end{aligned}$$

We get (6) from the induction step and (7) from inequality (5). Similarly, we can prove that $x_i(t) \ge x^*_i - \delta(t) - (1-\alpha)\frac{\sum_{\tau=1}^{t} e(\tau)}{t}$. As a result $\|x_i(t) - x^*\|_\infty \le e(t)$ and the induction is complete. Therefore, we have that with probability at least $1-p$, $\|x^t - x^*\|_\infty \le e(t)$, where $e(t)$ satisfies the following recursive relation

$$e(t) = \delta(t) + (1-\alpha)\frac{\sum_{\tau=0}^{t-1} e(\tau)}{t} \text{ and } e(0) = \|x^0 - x^*\|_\infty,$$

where $\delta(t) = \sqrt{\frac{\log(\pi^2 n t^2/(6p))}{t}}$ for $t \ge 1$ and $\delta(0) = 0$. By setting $D = \pi^2 n/6p > \mathrm{e}^2$ for sufficiently small p in Lemma 2 we get that with probability at least $1 - 1/\sqrt{t}$

$$e(t) \le \sqrt{\ln D}\,\frac{(\ln t)^{3/2}}{t^{\min(\rho,1/2)}}.$$

Therefore, by the law of total expectation the expected error is upper bounded by

$$\mathbf{E}\left[\|x_t - x^*\|_\infty\right] \le \frac{1}{\sqrt{t}} + e(t)\left(1 - \frac{1}{\sqrt{t}}\right) \le C\sqrt{\ln n}\,\frac{(\ln t)^2}{t^{\min(\rho,1/2)}},$$

where C is a sufficiently large universal constant.

■

4 Fictitious Play is no-regret

In this section we consider the Online Convex Optimization problem that we defined in Subsection 2.2. For every agent i with a_i, s_i, the following OCO problem is defined. For $b \in [0,1]$ we define $C_b(z) =$

$\alpha_i(z-s_i)^2+(1-\alpha_i)(z-b)^2$. The feasibility set is $K=[0,1]$ and the set of functions that the adversary chooses from is $\mathcal{F}_i=\{C_b(z):b\in[0,1]\}$. Since the functions of $\mathcal{F}_i$ are uniquely determined by b the adversary simply chooses a sequence b^t. We now show that fictitious play admits no-regret for this OCO problem. To simplify notation, since we have fixed an agent i, we drop the subscript i from the following, i.e. we denote α_i by α. Let z^t be the choice that A makes at time t.

Theorem 4. *Let* $z^t=\operatorname{argmin}_{z\in[0,1]}\sum_{\tau=0}^{t-1}C^\tau(z)$ *then*

$$\sum_{t=0}^{T}C^t(z^t)\le\min_{z\in[0,1]}\sum_{t=0}^{T}C^t(z)+O(\log T).$$

In order to prove this claim, we will first define a similar rule y^t that takes into account the function C^t for the "prediction" at time t. Intuitively, this guarantees that the rule admits no regret.

Lemma 3. *Let* $y^t=\operatorname{argmin}_{z\in[0,1]}\sum_{\tau=0}^{t}C^\tau(z)$ *then*

$$\sum_{t=0}^{T}C^t(y^t)\le\min_{z\in[0,1]}\sum_{t=0}^{T}C^t(z)$$

Proof. By definition of y^t, $\min_{z\in[0,1]}\sum_{t=0}^{T}C^t(z)=\sum_{t=0}^{T}C^t(y^T)$, so

$$\begin{aligned}\sum_{t=0}^{T}C^t(y^t)-\min_{z\in[0,1]}\sum_{t=0}^{T}C^t(z)&=\sum_{t=0}^{T}C^t(y^t)-\sum_{t=0}^{T}C^t(y^T)\\&=\sum_{t=0}^{T-1}C^t(y^t)-\sum_{t=0}^{T-1}C^t(y^T)\\&\le\sum_{t=0}^{T-1}C^t(y^t)-\sum_{t=0}^{T-1}C^t(y^{T-1})\\&=\sum_{t=0}^{T-2}C^t(y^t)-\sum_{t=0}^{T-2}C^t(y^{T-1})\end{aligned}$$

Continuing in the same way, we get $\sum_{t=0}^{T}C^t(y^t)\le min_{z\in[0,1]}\sum_{t=0}^{T}C^t(z)$. ■

Now we can derive some intuition for the reason that *fictitious play* admits no regret. Since the cost incurred by the sequence y^t is at most that of the best fixed strategy, we can compare the cost incurred by z^t with that of y^t. However, the functions in $\mathcal{F}_i$ are Lipschitz-continuous and more specifically quadratic. These functions are all "similar" to each other, so the extra function C^t that y^t takes as input doesn't change dramatically the minimum point of the sum. Thus, for each t the numbers z^t and y^t are quite close and as a result the difference in their cost must be quite small.

Lemma 4. *For all* t, $C^t(z^t)\le C^t(y^t)+2\frac{1-\alpha}{t+1}+\frac{(1-\alpha)^2}{(t+1)^2}$.

Proof. We first prove that for all t,

$$|z^t-y^t|\le\frac{1-\alpha}{t+1}. \tag{8}$$

By definition $z^t=\alpha s+(1-\alpha)\frac{\sum_{\tau=0}^{t-1}b_\tau}{t}$ and $y^t=\alpha s+(1-\alpha)\frac{\sum_{\tau=0}^{t}b_\tau}{t+1}$.

$$\begin{aligned}|z^t-y^t|&=(1-\alpha)\left|\frac{\sum_{\tau=0}^{t-1}b_\tau}{t}-\frac{\sum_{\tau=0}^{t}b_\tau}{t+1}\right|\\&=(1-\alpha)\left|\frac{\sum_{\tau=0}^{t-1}b_\tau-tb^t}{t(t+1)}\right|\\&\le\frac{1-\alpha}{t+1}\end{aligned}$$

The last inequality follows from the fact that $b_\tau \in [0,1]$. We now use inequality (8) to bound the difference $C^t(z^t) - C^t(y^t)$.

$$\begin{aligned}
C^t(z^t) &= \alpha(z^t - s)^2 + (1-\alpha)(z^t - y_t)^2 \\
&\le \alpha(y^t - s)^2 + 2\alpha \left|y^t - s\right| \left|z^t - y^t\right| + \alpha \left|z^t - y^t\right|^2 \\
&\quad + (1-\alpha)(y_t - y_t)^2 + 2(1-\alpha)\left|y^t - y_t\right| \left|z^t - y^t\right| + (1-\alpha)\left|z^t - y^t\right|^2 \\
&\le C^t(y^t) + 2\left|z^t - y^t\right| + \left|y^t - z^t\right|^2 \\
&\le C^t(y^t) + 2\frac{1-\alpha}{t+1} + \frac{(1-\alpha)^2}{(t+1)^2}
\end{aligned}$$

■

Theorem 4 easily follows since

$$\begin{aligned}
\sum_{t=0}^{T} C^t(z^t) \le \sum_{t=0}^{T} C^t(y^t) + \sum_{t=0}^{T} 2\frac{1-\alpha}{t+1} + \sum_{t=0}^{T} \frac{(1-\alpha)^2}{(t+1)^2} \\
\le \min_{z\in[0,1]} \sum_{t=0}^{T} C^t(z) + 2(1-\alpha)(\log T + 1) + (1-\alpha)\frac{\pi^2}{6} \\
\le \min_{z\in[0,1]} \sum_{t=0}^{T} C^t(z) + O(\log T)
\end{aligned}$$

5 Lower Bound for no-regret Dynamics

As we have already discussed for any fixed instance I with $\rho \ge 1/2$, *fictitious play* acheives convergence rate $\mathbf{E}_I\left[\|x(t) - x^*\|_\infty\right] = O\left(1/\sqrt{t}\right)$, this rate is outperformed by the rate of the original *FJ model* convergence rate $\mathbf{E}_I\left[\|x(t) - x^*\|_\infty\right] = O\left(1/2^t\right)$. An interesting question is whether this gap is due to the limited information exchange between the agents in the random payoff variant or can be reduced with another no-regret algorithm A.

Claim 1. *Let A a no-regret algorithm and let $x_A(t)$ the opinion vector defined in* **??**. *For any $c > 0$, there exists an instance I_A such that $\mathbf{E}_I\left[\|x_A(t) - x^*\|_\infty\right] = \Omega(1/t^{1+c})$.*

The above claim states that rationality in selfish agents comes with the price of slow convergence to the equilibrium point. Although we were not able to provide a full proof of Claim 1, our results indicate that the existence of an algorithm A, not satisfying it, is highly unlike. Based on lower bound techniques develloped in the statistics literature, we first prove the following lower bound for any estimator $\hat{\theta}$ of Bernoulli distributions.

Theorem 5. *Let a Bernoulli estimator $\hat{\theta}$ with error rate $R_p(t)$. Then, for all $[a,b] \subseteq [0,1]$,*

$$\lim_{t\to\infty} t^{1+c} \int_a^b R_p(t)dp = +\infty$$

Now suppose that there exists a no-regret algorithm A not statisfying Claim 1 i.e. for all instances I,

$$\lim_{t\to\infty} t^{1+c} \mathbf{E}_I\left[\|x_A(t) - x^*\|_\infty\right]$$

In Theorem 6 we show that A can be used as an estimator $\hat{\theta_A}$ for Bernoulli distributions with risk asymptotically the same.

Theorem 6. *Let the no-regret algorithm A such that for all instances I, $\lim_{t\to\infty} t^{1+c}\mathbf{E}_I\left[\|x_A(t)-x^*\|_\infty\right]=0$. Let $\hat{\theta_A}$ the estimator constructed from A. Then for all $p\in Q[0,1]$,*

$$\lim_{t\to\infty} t^{1+c}R_p(t)=0$$

As a result, the existence of such an algorithm A implies the existence of an estimator $\hat{\theta_A}$ whose risk $R_p(t)$ satisfy the following statements:

- for all $p\in Q([0,1])$, $\lim_{t\to\infty} t^{1+c}R_p(t)=0$
- for all $[a,b]\subseteq[0,1]$, $\lim_{t\to\infty} t^{1+c}\int_a^b R_p(t)dp=+\infty$

The above two statements are controversial for any function $R_p(t)$ that satisfies minimal technical assumptions (e.g. $R_p(t)=f(p)g(t)$), meaning that the first statement is violated (the second is ensured by Theorem 5). Unfortunately one can construct degenerate functions $R_p(t)$ that simultaneously satisfy the above two statements. For example $R_p(t)=\Pi_{i=1}^t(p-q_i)^2$, where the sequence $\{q_i\}_{i=1}^\infty$ is an enumeration of the rationals in $[0,1]$. Although the existence of such function means that we cannot derive a full proof of Claim 1, their exremely degenerate form indicates that such a risk rate $R_p(t)$ cannot be obtained by any *reasonable* estimator $\hat{\theta_A}$. Implying that Claim 1 must be satisfied by A.

5.1 Proof of Theorem 6

In the following Lemma we show how we can use algorithm A to construct an estimator $\hat{\theta_A}$ for Bernoulli distributions.

Lemma 5. *For any algorithm A, we can construct a Bernoulli estimator $\hat{\theta_A}$ such that for all $p\in Q([0,1])$, there exists an instance I_p such that*

$$R_p(t)\le 2E_{I_p}[||x_A(t)-x^*||_\infty]$$

Proof. At first we remind that an estimator $\hat{\theta}$ is a sequence of fuctions $\{\hat{\theta}_t\}_{t=1}^\infty$, where $\theta_t:\ \{0,1\}^t\mapsto[0,1]$. We construct such a sequence using the algorithm A. We also remind that when an agent i runs algorithm A, she selects $x_i(t)$ according to the cost functions $\{C_i^\tau\}$ that she has already reveived

$$x_i(t)=A_t(C_i^1,\dots,C_i^{t-1})$$

Consider an agent i with $a_i=1$ and $s_i=0$ that runs A. Then $C_i^t(x)=x^2$ for all t and $x_i(t)=A_t(x^2,\dots,x^2)$. The latter means that $x_i(t)$ only depends on t, $x_i(t)=h_0(t)$. Equivalently, if $a_i=1$ and $s_i=1$ then $x_i(t)=A_t((1-x)^2,\dots,(1-x)^2)$ and $x_i(t)=h_1(t)$. Finally, consider an agent i with $a_i=1/2$ and $s_i=0$. In this case $C_i^t=\frac{1}{2}x^2+\frac{1}{2}(x-y_t)^2$, where $y_t\in[0,1]$ is the opinion of the neighbor $j\in N_i$ that i met at round t. As a result, $x_i(t)=A_t(\frac{1}{2}x^2+\frac{1}{2}(x-y_1)^2,\dots,\frac{1}{2}x^2+\frac{1}{2}(x-y_{t-1})^2)=f_t(y_1,\dots,y_{t-1})$. The estimator $\hat{\theta_A}$ is the following sequences $\{\hat{\theta}_t\}_{t=1}^\infty$

$$\hat{\theta}_t(Y_1,\dots,Y_t)=\frac{1}{2}f_{t+1}(h_{Y_1}(1),\dots,h_{Y_t}(t))$$

Observe that $\hat{\theta}_t:\{0,1\}^t\mapsto[0,1]$ meaning that $\hat{\theta_A}$ is a valid estimator for Bernoulli distributions.

Now for any $p\in Q([0,1])$, we construct an appropriate instance I_p s.t. $R_p(t)=\mathbf{E}_p\left[|\hat{\theta}_t-p|\right]\le 2\mathbf{E}_{I_p}\left[\|x^t-x^*\|_\infty\right]$. For $p=\frac{k}{n}$ consider the following instance I_p with $n+1$ agents:

- A central agent with $s_c=0$ and $a_c=1/2$.
- Directed edges from the central agent to all the other agents.
- k agents with $s_i=0$ and $a_i=1$

- $n-k$ agents with $s_i = 1$ and $a_i = 1$

We just need to prove that in I_p, $\mathbf{E}_p\left[|\hat{\theta}_t - p|\right] \leq 2\mathbf{E}_{I_p}\left[\|x^t - x^*\|_\infty\right]$. Notice that $x_c^* = \frac{p}{2}$ and $x_i^* = s_i$ if $i \neq c$. . At round t, if the oracle returns to the center agent the value $h_1(t)$ of a 1-agent, then $Y_t = 1$ otherwise $Y_t = 0$. As a result, $\mathbf{P}\left[Y_t = 1\right] = p$ and

$$\begin{aligned}\mathbf{E}_{I_p}\left[\|x^t - x^*\|_\infty\right] &\geq \mathbf{E}_{I_p}\left[||x_c^t - x_c^*||\right] \\ &= \mathbf{E}_p\left[|\frac{\widehat{\theta}_t}{2} - \frac{p}{2}\right]|] = R_p(t)\end{aligned}$$

■

Theorem 7. *Let the no-regret algorithm A such that for all instances I, $\lim_{t\to\infty} t^{1+c}\mathbf{E}_I\left[\|x_A(t) - x^*\|_\infty\right] = 0$. Let $\hat{\theta_A}$ the estimator constructed from A. Then for all $p \in Q[0,1]$,*

$$\lim_{t\to\infty} t^{1+c} R_p(t) = 0$$

The proof follows by direct application of the Lemma 5

5.2 Proof of Theorem 5

Our proof builts on a standard technique in the statistics literature, for proving lower bounds on the risk of estimators. This technique reduces the estimation problem to the canonical hypothesis testing problem for which information theoretic lower bounds exist. Due to lack of space we are not able to explain this very interesting reduction. The interested reader can find a detailed explanation in [][Duchi]. We present a major Definition and Lemma of this technique in the case of Bernoulli distribution i.e. Definition 6 and Lemma 6 that are the starting point of our proof.

Definition 6. *A finite family of Bernoulli distribution $\mathcal{P} = \{B(p_1), \ldots, B(p_n)\}$ is called a 2δ-packing if $|p_i - p_j| > 2\delta$ for all $i \neq j$.*

Lemma 6 (Fano's Inequality)**.** *Let $\hat{\theta}$ an estimator for the parameter p of a Bernoulli distribution and $\mathcal{P} = \{B(p_1), \ldots, B(p_n)\}$ a 2δ-packing. Then*

$$\frac{1}{n}\sum_{i=1}^{n} R_{p_i}(t) \geq \delta\left(1 - \frac{\log 2}{\log n} - \frac{t}{n^2 \log n}\sum_{i,j} D_{kl}\left(B(p_i), B(p_j)\right)\right)$$

At first for a fixed interval $[a,b]$ we select an appropriate 2δ-packing $\mathcal{P}$ and apply Lemma 6. The most challenging part is to appropriately select $\mathcal{P}$ such that the $\sum_{i,j} D_{kl}\left(B(p_i), B(p_j)\right)$ can be efficiently upper bounded. In our case $\mathcal{P}$ consists of t different distributions and $\delta = \frac{b-a}{2t}$. For further details one can see the proof of the Lemma 13.

Lemma 7. *For the interval $[a,b] \subseteq [0,1]$. Let the 2δ-packing $\mathcal{P} = \{B(p_1), \ldots, B(p_t)\}$ with $p_i = a + \frac{b-a}{2t} + i\frac{b-a}{t}$. Then,*

$$\frac{1}{t}\sum_{i=1}^{t} R_{p_i}(t) \geq \frac{b-a}{2t}\left(c_1 - \frac{c_2}{t}\right)$$

where $c_1, c_2 > 0$ and $c_1 > \frac{1}{2}$.

Now our goal is to upper bound the quantity $\frac{1}{t}\sum_{i=1}^{t} R_{p_i}(t)$ by the quantity $\int_a^b R_p(t)dp$. Our proof if based in the following simple idea. Assume that we draw t samples from the $B(p)$ and t samples from the $B(q)$, where $|p-q| \leq \frac{1}{t}$. These two product distribution are approximately the same since we only t samples are drawn and their parameters differ only by $\frac{1}{t}$. At the same time we have that $|\hat{\theta}_t - p| \simeq |\hat{\theta}_t - q|$ meaning that $R_p(t) \simeq R_q(t)$ ($R_p(t) = \mathbf{E}_p\left[|\hat{\theta}_t - p|\right]$). The above intuition is formallized and proved in Lemma 14.

Lemma 8. *For every $p \in [p_i - \frac{b-a}{2t}, p_i + \frac{b-a}{2t}]$,*

$$R_{p_i}(t) \leq C_a^b R_p(t) + |p - p_i|$$

where C_a^b is a constant that depends on a, b.

Using Lemma 14, we can easily upper bound $\frac{1}{t}\sum_{i=1}^{t} R_{p_i}(t)$ by $\int_a^b R_p(t)dp$.

Lemma 9. *Let an interval $[a, b] \subseteq [0, 1]$ and the parameters $p_i = a + \frac{b-a}{2t} + i\frac{b-a}{t}$. Then,*

$$\frac{1}{t}\sum_{i=1}^{t} R_{p_i}(t) \leq \frac{C_a^b}{b-a}\int_a^b R_p(t)dp + \frac{b-a}{4t}$$

Now we are ready to prove Theorem 5.

Theorem 8. *Let a Bernoulli estimator $\hat{\theta}$ with error rate $R_p(t)$. Then, for all $[a, b] \subseteq [0, 1]$ and $c > 0$,*

$$\lim_{t\to\infty} t^{1+c}\int_a^b R_p(t)dp = +\infty$$

Proof. By combining Lemmas (13), (15) and multiplying by t^{1+c} we get

$$\frac{c_a^b}{b-a}t^{1+c}\int_a^b \mathbf{E}_p\left[\left|\hat{\theta}^t - p\right|\right]dp \geq \frac{b-a}{2}t^c\left(\left(c_1 - \frac{1}{2}\right) - \frac{c_2}{t}\right)$$

The coefficient of t^c in the right hand side of (5.2) is positive, so $\lim_{t\to\infty} t^{1+c}\int_a^b R_p(t)dp = +\infty$ ■

6 A Graph Aware Update Rule

In Section 5 we saw that in our *imperfect information* for any no-regret algorithm $\mathcal{A}$ there exists a graph G such that the corresponding dynamics needs poly$(1/\varepsilon)$ rounds to achieve error ε. However, in the *perfect information* FJ-model there exists a simple update rule that requires only $\log(1/\varepsilon)$ rounds. Notice that our lower bound crucially depends on the fact that $\mathcal{A}$ is no-regret. At this point, a natural question is whether, this exponential gap is a generic restriction of our imperfect information model. We answer this question in the negative. More precisely the reason that no-regret dynamics suffer slow convergence is that the update rule depends only on the expressed opinions. In this section we exhibit an update rule that depends on the graph G and achieves exponentially fast convergence. Precisely, our update rule depends on the expressed opinions, the number of neighbors of each agent, and the number of agents n. The latter is interesting because information about the graph G combined with agents that do not act selfishly, can restore the exponential convergence rate.

The main idea of the protocol is straightforward: to counterbalance the imperfect information, the agents can spend some rounds to simulate one round of the original FJ-model. To do this, they agree to stop updating their expressed opinion for a large enough window of rounds so that everybody learns, with high probability, *exactly* the average of the opinions of their neighbors. Following the ideas of Section 3 an agent could just average the opinions that she gets in this window. Unfortunately this would again result in a poly$(1/\varepsilon)$-round protocol. However this can be fixed by using the additional knowledge of the number of agents and the number of neighbors. Precisely, each agent i keeps an array with the frequencies of the different opinions that she observes. The catch is that at the end of the window, she rounds each frequency to the closest multiple of $1/d_i$, where d_i is the number of neighbors of agent i. This rounding step is crucial to ensure the exponential convergence rate. To see that this works, first notice that if all agents stop updating their opinions for a number of rounds, each agents just needs to specify exactly how many of her neighbors share a specific observed opinion. If the length of the window is large the frequency of a specific opinion at the end of the window will be sufficiently close to the true frequency. Since the frequencies of the opinions that agent i observes can only be multiples of $1/d_i$, we can round the estimated frequencies to the closest multiple of $1/d_i$ to recover the true frequencies and use them to get the exact average of the opinions of the neighbors. To bound the length of the window we use a VC-dimension argument and show that with $n^2 \log n$ rounds each agent knows the frequencies within error smaller than $1/n$ with constant probability, which then can be trivially amplified by repeating the procedure.

We next state a version of the standard VC-Inequality that we will use in our argument. Let P be a discrete distribution over $[n]$, and let $S_1, \ldots, S_t$ be t i.i.d samples drawn from P, i.e. $(S_1, \ldots, S_t) \sim P^t$. The empirical distribution $\hat{P}_t$ is the following estimator of the density of P.

$$\hat{P}_t(A) = \frac{\sum_{i=1}^{t} \mathbb{1}_{[S_i \in A]}}{t}, \tag{9}$$

where $A \subseteq [n]$. In words, $\hat{P}_t$ simply counts how many times the value i appeared in the samples $S_1, \ldots S_t$. We will use the following version of the classical result of Vapnik and Chervonenkis.

Lemma 10. *Let $\mathcal{A}$ be a collection of subsets of $\{1, \ldots, n\}$ and let $S_{\mathcal{A}}(t)$ be the Vapnik-Chervonenkis shatter coefficient, defined by*

$$S_{\mathcal{A}(t)} = \max_{x_1, \ldots, x_t \in [n]} \left| \{\{x_1, \ldots, x_t\} \cap A : A \in \mathcal{A}\} \right|.$$

Then

$$\mathbf{E}_{P^t}\left[\max_{A \in \mathcal{A}} \left|\hat{P}_t(A) - P(A)\right|\right] \le 2\sqrt{\frac{\log 2S_{\mathcal{A}}(t)}{t}}$$

Algorithm 3 Graph Aware Update Rule

$x_i(0) \leftarrow s_i$.
$M_1 = O(\ln(n/\varepsilon))$, $M_2 = O(d^2 \ln(d))$
for $l = 1, \ldots, \ln(1/\varepsilon)$ **do**
  Keep a set A of tuples $(x, \text{freq}(x))$ of and an array B of length M_1.
  for $j = 1, \ldots, M_1$ **do**
    for $k = 1, \ldots, M_2$ **do**
      Get the opinion X_k of a random neighbor.
      if X_j is not in A **then**
        Insert $(X_k, 1)$ to A.
      else
        $(X_k, \text{freq}(X_k)) \leftarrow (X_k, \text{freq}(X_k) + 1)$.
    Divide all frequencies of A by M_2.
    Round all frequencies of to the closest multiple of $1/d_i$.
    $B(j) \leftarrow \alpha_i \sum_{x \in A} \text{freq}(x) + (1 - \alpha_i) s_i$.
  $x_i(t) \leftarrow \text{majority}_j B(j)$.

Theorem 9. *Let $I = (G(V, E), s, a)$ be an instance of the opinion formation game of Definition 1 with $a > 1/2$. Let d be the maximum degree of the graph G and $n = |V|$. There exists an update rule after $O(d^2 \log^2 n \log^2(1/\varepsilon))$ rounds achieves expected error $\mathbf{E}[\|x_t - x^*\|_\infty] \le \varepsilon$.*

Proof. According to the update rule 3 all agents fix their opinions $x_i(t)$ for $M_1 \times M_2$ rounds. To estimate the sum of the opinions each agent estimates the frequencies k_j/d_i. Since the neighbors have at most d_i different opinions we can map the opinions to natural numbers in $[d_i]$. At each round the agent gets the opinion of a random neighbor and therefore the samples X_i that she observes are drawn from a discrete distribution P supported on $[d_i]$. If k_j be the (absolute) frequency of the opinion j namely the number of neighbors that express j as their opinion, then the probability $P(j)$ of opinion j is k_j/d_i. To learn the probabilities $P(j)$ using samples from P, we let $\mathcal{A} = \{\{1\}, \{2\}, \ldots, \{d_i\}\}$ and use Lemma 10 to get that

$$\mathbf{E}_{P^m}\left[\max_{j\in[d_i]} \left|\hat{P}_m(j) - P(j)\right|\right] \leq 2\sqrt{\frac{\log 2d_i}{m}},$$

since $S_{\mathcal{A}} \leq n$. Therefore, an agent can draw $m = 100n^2 \log(2n)$ to learn the frequencies k_j/d_i within expected error $1/(5d)$. Notice now that the array A after line 12 corresponds to the empirical distribution of equation (9). Notice that if the agents have estimations of the frequencies k_j/d_i with error smaller than $1/d$, then by rounding them to the closest multiple of $1/d_i$ they learn the frequencies exactly. By Markov's inequality we have that with probability at least $4/5$ the rounded frequencies are exactly correct. By standard Chernoff bounds we have that if the agents repeat the above procedure $\ln(1/\delta)$ times and keep the most frequent of the answers $B(j)$, then they will obtain the correct answer with probability at least $1-\delta$. We know that, having computed the *exact* average of the opinions of the neighbors $(\log(1/\varepsilon))$ rounds are enough to achieve error ε. Since we need all nodes to succeed at computing the exact averages for $(\log(1/\varepsilon))$ rounds, from the union bound we get that for $\delta < \frac{\varepsilon}{n\ln(1/\varepsilon)}$, with probability at least $1-\varepsilon$ the error is at most ε. Finally, from the law of total expectation, after $T = O(d^2 \log d \log(\varepsilon)(\log(n/\varepsilon) + \log\log(1/\varepsilon))$ rounds the expected error is $\mathbf{E}\left[\|x_T - x^*\|_\infty\right] = (1-\varepsilon)\varepsilon + \varepsilon \leq 2\varepsilon$. ∎

7 Rational Lower Bound

Theorem 10. *Let $\theta_t : \{0,1\}^t \to [0,1]$ be a sequence of estimators for the success probability p of a Bernoulli random variable with distribution P. There exists $p \in [0,1]$ such that*

$$\lim_{t\to\infty} t^2\, \mathbf{E}_{X\sim P^t}\left[|\theta_t(X) - p|\right] > 0.$$

Proof. Observe that $\theta_t(\{0,1\}^t)$ has cardinality at most 2^t. Since

$$\sum_{0\leq i\leq t} \sum_{\|x\|_1 = i} |\theta_t(x) - p|\, p^i(1-p)^{t-i} \geq \sum_{0\leq i\leq t} \binom{t}{i} \left|\frac{\sum_{\|x\|_1=i}\theta_t(x)}{\binom{t}{i}} - p\right| p^i(1-p)^{t-i}.$$

Thus, without loss of generality we assume that $\theta_t(\{0,1\}^t)$ contains at most $t+1$ discrete points.

In the following, we work in the measure space $(\mathbf{R}, \mathcal{M}, \mu)$, where μ is the Lebesgue measure, and $\mathcal{M}$ is the σ-algebra of the Lebesgue measurable sets. Suppose that there exists no such $p \in [0,1]$. Let

$$A = \{p \in [0,1] \ :\ \lim_{t\to\infty} t^2\, \mathbf{E}_{X\sim P^t}\left[|\theta_t(X) - p|\right] = 0\}.$$

Then $A = [0,1]$ and A is measurable as an interval. Notice that,

$$A \subseteq \bigcup_{t=1}^{\infty} \bigcap_{k=t}^{\infty} A_k,$$

where $A_k = \{p \in [0,1] \ :\ R_k(p) < 1/2\}$, and $R_k(p) = k^2\, \mathbf{E}_{X\sim P^k}\left[|\theta_k(X) - p|\right]$. We have that $R_k : [0,1] \to [0,+\infty)$ is polynomial of degree t in p and therefore it is a measurable function. Thus, A_k is measurable. We now show that

$$A_k \subseteq B_k := \{p \in [0,1] \ :\ k^2 \min_{0\leq i\leq k} |\theta_k(i) - p| < 1\}.$$

We prove this by contradiction. Suppose that $p \in A_k$ but $p \notin B_k$. Since $p \in A_k$ we have that

$$R_k(p) = k^2 \sum_{i=0}^{k} \binom{k}{i} |\theta_k(i) - p|\, p^i(1-p)^{k-i} \geq k^2 \min_{0\leq i\leq k} |\theta_k(i) - p| \sum_{i=0}^{k} \binom{k}{i} p^i(1-p)^{k-i} \geq 1.$$

Since the functions $p \mapsto k^2 |\theta_k(i) - p|$ are measurable, their pointwise minimum is measurable and therefore the sets B_k are also measurable. We next proceed to bound $\mu(B_k)$. Since θ_k can only take k different values we have that there exist $k+1$ intervals (a_{k_i}, b_{k_i}) of length at most $2/k^2$ such that $B_k = \bigcup_{i=0}^{k} \left(a_{k_i}, b_{k_i}\right)$. Since μ is subadditive we have

$$\mu(B_k) \leq \sum_{i=0}^{k} \frac{2}{k^2} = \frac{2(k+1)}{k^2}.$$

Now observe that

$$\mu(A) \leq \mu\left(\bigcup_{t=1}^{\infty} \bigcap_{k=t}^{\infty} A_k\right) \leq \mu\left(\bigcup_{t=1}^{\infty} \bigcap_{k=t}^{\infty} B_k\right) \leq \sum_{t=1}^{\infty} \mu\left(\bigcap_{k=t}^{\infty} B_k\right) \leq \sum_{t=1}^{\infty} \lim_{k\to\infty} \mu(B_k) = 0$$

Which is a contradiction since we assumed $A = [0,1]$. ■

References

[BEDL06] Avrim Blum, Eyal Even-Dar, and Katrina Ligett. Routing without regret: On convergence to nash equilibria of regret-minimizing algorithms in routing games. In *Proceedings of the Twenty-fifth Annual ACM Symposium on Principles of Distributed Computing*, PODC '06, pages 45–52, New York, NY, USA, 2006. ACM.

[BFM16] Vittorio Bilò, Angelo Fanelli, and Luca Moscardelli. Opinion formation games with dynamic social influences. In Yang Cai and Adrian Vetta, editors, *Web and Internet Economics*, pages 444–458, Berlin, Heidelberg, 2016. Springer Berlin Heidelberg.

[BGM13] Kshipra Bhawalkar, Sreenivas Gollapudi, and Kamesh Munagala. Coevolutionary opinion formation games. In *Symposium on Theory of Computing Conference, STOC'13, Palo Alto, CA, USA, June 1-4, 2013*, pages 41–50, 2013.

[BKO11] David Bindel, Jon M. Kleinberg, and Sigal Oren. How bad is forming your own opinion? In *IEEE 52nd Annual Symposium on Foundations of Computer Science, FOCS 2011, Palm Springs, CA, USA, October 22-25, 2011*, pages 57–66, 2011.

[CHM17] Johanne Cohen, Amélie Héliou, and Panayotis Mertikopoulos. Hedging under uncertainty: Regret minimization meets exponentially fast convergence. In *Algorithmic Game Theory - 10th International Symposium, SAGT 2017, L'Aquila, Italy, September 12-14, 2017, Proceedings*, pages 252–263, 2017.

[CKO13] Flavio Chierichetti, Jon M. Kleinberg, and Sigal Oren. On discrete preferences and coordination. In *ACM Conference on Electronic Commerce, EC '13, Philadelphia, PA, USA, June 16-20, 2013*, pages 233–250, 2013.

[EFHS17] Markos Epitropou, Dimitris Fotakis, Martin Hoefer, and Stratis Skoulakis. Opinion formation games with aggregation and negative influence. In *Algorithmic Game Theory - 10th International Symposium, SAGT 2017, L'Aquila, Italy, September 12-14, 2017, Proceedings*, pages 173–185, 2017.

[EMN09] Eyal Even-Dar, Yishay Mansour, and Uri Nadav. On the convergence of regret minimization dynamics in concave games. In *Proceedings of the 41st Annual ACM Symposium on Theory of Computing, STOC 2009, Bethesda, MD, USA, May 31 - June 2, 2009*, pages 523–532, 2009.

[FGV16] Diodato Ferraioli, Paul W. Goldberg, and Carmine Ventre. Decentralized dynamics for finite opinion games. *Theor. Comput. Sci.*, 648(C):96–115, October 2016.

[FJ90] Noah E. Friedkin and Eugene C. Johnsen. Social influence and opinions. *The Journal of Mathematical Sociology*, 15(3-4):193–206, 1990.

[FS99] Yoav Freund and Robert E. Schapire. Adaptive game playing using multiplicative weights. *Games and Economic Behavior*, 29(1):79 – 103, 1999.

[FV97] Dean P. Foster and Rakesh V. Vohra. Calibrated learning and correlated equilibrium. *Games and Economic Behavior*, 21(1):40 – 55, 1997.

[GS14] Javad Ghaderi and R. Srikant. Opinion dynamics in social networks with stubborn agents: Equilibrium and convergence rate. *Automatica*, 50(12):3209–3215, 2014.

[SA] Hart Sergiu and MasâĂŘColell Andreu. A simple adaptive procedure leading to correlated equilibrium. *Econometrica*, 68(5):1127–1150.

[YOA+13] Mehmet Ercan Yildiz, Asuman E. Ozdaglar, Daron Acemoglu, Amin Saberi, and Anna Scaglione. Binary opinion dynamics with stubborn agents. *ACM Trans. Economics and Comput.*, 1(4):19:1–19:30, 2013.

[ZLZ17] Yichi Zhou, Jialian Li, and Jun Zhu. Identify the nash equilibrium in static games with random payoffs. In *Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017*, pages 4160–4169, 2017.

A Fictitious Play Donvergence Rate

We give here the proof of the following technical lemma that we used to derive an upper bound on the rate of convergence of the fictitious play dynamics in Section 3. We restate Lemma **??** for completeness.

Lemma 11. *Let $e(t)$ be a function satisfying the recursion*

$$e(t) = \delta(t) + (1-\alpha)\frac{\sum_{\tau=0}^{t-1} e(\tau)}{t} \text{ and } e(0) = \|x^0 - x^*\|_\infty,$$

where $\delta(t) = \sqrt{\frac{\ln(Dt^2)}{t}}$, $\delta(0) = 0$, and $D > \mathrm{e}^2$ is a positive constant. Then $e(t) \leq \sqrt{2\ln(D)}\frac{(\ln t)^{3/2}}{t^{\min(\rho,1/2)}}$.

Proof. Observe that for all $t \geq 0$ the function $e(t)$ the following recursive relation

$$e(t+1) = e(t)\left(1 - \frac{\rho}{t+1}\right) + \delta(t+1) - \delta(t) + \frac{\delta(t)}{t+1} \tag{10}$$

For $t = 0$ we have that

$$e(1) = (1-\rho)e(0) + \delta(1) = (1-\rho)e(0) + \sqrt{\ln D} \tag{11}$$

Observe that for $D > \mathrm{e}^2$, $\delta(t)$ is decreasing for all $t \geq 1$. Therefore, $\delta(t+1) - \delta(t) + \frac{\delta(t)}{t+1} \leq \frac{\delta(t)}{t+1}$ and from equations (10) and (11) we get that for all $t \geq 0$

$$e(t+1) \leq e(t)\left(1 - \frac{\rho}{t+1}\right) + \frac{\sqrt{\ln(D(t+1)^2)}}{(t+1)^{3/2}} \leq e(t)\left(1 - \frac{\rho}{t+1}\right) + \frac{\sqrt{2\ln(D(t+1))}}{(t+1)^{3/2}}$$

Now let $g(t) = \frac{\sqrt{2\ln(Dt)}}{t^{3/2}}$ to obtain for all $t \geq 1$

$$\begin{aligned}
e(t) &\leq (1 - \frac{\rho}{t})e(t-1) + g(t) \\
&\leq (1 - \frac{\rho}{t})(1 - \frac{\rho}{t-1})e(t-2) + (1 - \frac{\rho}{t})g(t-1) + g(t) \\
&\leq (1 - \frac{\rho}{t})\cdots(1-\rho)e(0) + \sum_{\tau=1}^{t} g(\tau)\prod_{i=\tau+1}^{t}(1 - \frac{\rho}{i}) \\
&\leq \frac{e(0)}{t^\rho} + \sum_{\tau=1}^{t} g(\tau)e^{-\rho\sum_{i=\tau+1}^{t}\frac{1}{i}} \\
&\leq \frac{e(0)}{t^\rho} + \sum_{\tau=1}^{t} g(\tau)e^{-\rho(H_t - H_\tau)} \\
&\leq \frac{e(0)}{t^\rho} + e^{-\rho H_t}\sum_{\tau=1}^{t} g(\tau)e^{\rho H_\tau} \\
&\leq \frac{e(0)}{t^\rho} + \frac{\sqrt{2}}{t^\rho}\sum_{\tau=1}^{t} \tau^\rho\frac{\sqrt{\ln(D\tau)}}{\tau^{3/2}} \\
&\leq \frac{e(0)}{t^\rho} + \frac{\sqrt{2\ln D}}{t^\rho}\sum_{\tau=1}^{t} \frac{\sqrt{\ln\tau}}{\tau^{3/2-\rho}}
\end{aligned}$$

We observe that

$$\sum_{\tau=1}^{t} \frac{\sqrt{\ln\tau}}{\tau^{3/2-\rho}} \leq \int_{\tau=1}^{t} \frac{\sqrt{\ln\tau}}{\tau^{3/2-\rho}}\mathrm{d}\tau \tag{12}$$

since, $\tau \mapsto \frac{\sqrt{\ln\tau}}{\tau^{3/2-\rho}}$ is a decreasing function of τ for all $\rho \in [0,1]$.

- If $\rho \leq 1/2$ then

$$\int_{\tau=1}^{t} \tau^{\rho} \frac{\sqrt{\ln \tau}}{\tau^{3/2}} \mathrm{d}\tau \leq \sqrt{\ln t} \int_{\tau=1}^{t} \frac{1}{\tau} \mathrm{d}\tau = (\ln t)^{3/2}$$

- If $\rho > 1/2$ then

$$\begin{aligned}\int_{\tau=1}^{t} \tau^{\rho} \frac{\sqrt{\ln \tau}}{\tau^{3/2}} \mathrm{d}\tau &= \int_{\tau=1}^{t} \tau^{\rho-1/2} \frac{\sqrt{\ln \tau}}{\tau} d\tau \\ &= \frac{2}{3} \int_{\tau=1}^{t} \tau^{\rho-1/2} ((\ln \tau)^{3/2})' \mathrm{d}\tau \\ &= \frac{2}{3} (\ln t)^{3/2} - (\rho - 1/2) \frac{2}{3} \int_{\tau=1}^{t} \tau^{\rho-3/2} (\ln \tau)^{3/2} \mathrm{d}\tau \\ &\leq \frac{2}{3} (\ln t)^{3/2}\end{aligned}$$

■

B Lower Bound

Lemma 12. *Let $P = B(p)$ and $Q = B(q)$ be two Bernoulli distributions then*

$$D_{\mathrm{kl}}(P \| Q) \leq \left(\frac{1}{p(1-p)} + \frac{1}{q(1-q)} \right) \frac{(p-q)^2}{2}$$

Proof. We have that $D_{\mathrm{kl}}(P \| Q) = p \log(p/q) + (1-p) \log((1-p)/(1-q))$. Let

$$f(x, y) = x \log \left(\frac{x}{y} \right) + (1-x) \log \left(\frac{1-x}{1-y} \right)$$

To simplify notation fix y and let $g(x) = f(x, y)$. We have

$$g'(x) = -\log \left(\frac{1-x}{1-y} \right) + \log \left(\frac{x}{y} \right), \qquad g''(x) = \frac{1}{(1-x)x}$$

If $x < y$ using Taylor in the interval $[x, y]$ we have for a $\xi \in [x, y]$

$$g(x) = g(y) + g'(y)(y-x) + g''(\xi)(y-x)^2/2 = g''(\xi)(x-y)^2/2.$$

If $x > y$ the using Taylor in the interval $[y, x]$ we get the same expression as above. Since $g''(x)$ is convex and is minimized at $x_0 = 1/2$ we have that if $|x - 1/2| < |y - 1/2|$ then $g''(\xi) \leq g''(y)$ else $g''(\xi) \leq g''(x)$. Therefore,

$$g(x) = f(x, y) \leq \max(g''(x), g''(y))(x-y)^2/2 \leq (g''(x) + g''(y))(x-y)^2/2$$

■

We are now going to use Lemma (6) to prove the following lemma, which provides a lower bound for the mean of the expectation of errors of the estimator.

Lemma 13. *Suppose we choose t Bernoulli distributions $\{P_i\}_{i=0}^{t-1}$ with the following probabilities:*

$$p_i = a + \frac{b-a}{2t} + i \frac{b-a}{t}, i = 0, ..., t-1$$

. Then:

$$\frac{1}{t} \sum_{i=1}^{t} \mathbf{E} \left[|\hat{\theta}_t - p_i| \right] \geq \frac{b-a}{2t} \left(c_1 - \frac{c_2}{t} \right)$$

where $c_1, c_2 > 0$ and $c_1 > \frac{1}{2}$.

Proof. Without loss of generality, we can assume that t is a multiple of 5. For simplicity, we set $R_{p_i}(t) = \mathbf{E}\left[\left|\hat{\theta}_t - p_i\right|\right]$. By the choice of p_i, we have that

$$\left|p_i - p_j\right| \geq \frac{b-a}{t}$$

, for every distinct i, j in the family. This means that for the family of t distributions that we picked, the property of Lemma (6) is satisfied for $\delta = \frac{b-a}{2t}$. Thus, by dividing the t distributions into groups of 5 and applying Lemma (6) to each one, we obtain:

$$\begin{aligned}
\frac{\sum_{k=i}^{i+4} R_{p_i}(t)}{5} &\geq \frac{b-a}{2t}\left(1 - \frac{I(X;V) + \log 2}{\log n}\right) \\
&= \frac{b-a}{2t}\left(1 - \frac{\log 2}{\log 5} - \frac{I(X;V)}{\log 5}\right) \\
&= \frac{b-a}{2t}\left(c_1 - \frac{I(X;V)}{\log 5}\right) \qquad (13)
\end{aligned}$$

where $c_1 = 1 - \frac{\log 2}{\log 5} > \frac{1}{2}$. We are now going to upper bound the mutual information $I(X;V)$ for every group of 5 distributions. We denote by P_i^t the distribution on vectors with t coordinates, where each coordinate is sampled independently from P_i. Let $U = \{i, i+1, i+2, i+3, i+4\}$ be a family of 5 distributions. Using a well known inequality REFERENCE NEEDED HERE:

$$\begin{aligned}
I(X;V) &\leq \frac{1}{5^2}\sum_{i,j \in U} D_{\mathrm{kl}}\left(P_i^t \| P_j^t\right) \\
&\leq D_{\mathrm{kl}}\left(P_i^t \| P_{i+4}^t\right) \\
&= tD_{\mathrm{kl}}\left(P_i \| P_{i+4}\right)
\end{aligned}$$

Using Lemma 12 we obtain

$$\begin{aligned}
tD_{\mathrm{kl}}\left(P_i \| P_{i+4}\right) &\leq t\frac{(p_i - p_{i+4})^2}{2}\left(\frac{1}{p_i(1-p_i)} + \frac{1}{p_{i+4}(1-p_{i+4})}\right) \\
&\leq t\left(\frac{1}{a(1-a)} + \frac{1}{b(1-b)}\right)(p_i - p_{i+4})^2 \\
&= t\left(\frac{1}{a(1-a)} + \frac{1}{b(1-b)}\right)\left(\frac{b-a}{2t} + i\frac{b-a}{t} - \frac{b-a}{t} - (i+4)\frac{b-a}{t}\right)^2 \\
&= t\left(\frac{1}{a(1-a)} + \frac{1}{b(1-b)}\right)\frac{16(b-a)^2}{t^2} \\
&= \frac{16(b-a)^2}{t}\left(\frac{1}{a(1-a)} + \frac{1}{b(1-b)}\right)
\end{aligned}$$

So, if we set

$$c_2 = 16\frac{(b-a)^2}{\log 5}\left(\frac{1}{a(1-a)} + \frac{1}{b(1-b)}\right)$$

then by equation (13) we obtain:

$$\sum_{k=i}^{i+4} R_{p_i}(t) \geq \frac{5(b-a)}{2t}\left(c_1 - \frac{c_2}{t}\right) \qquad (14)$$

Inequality (14) holds for every group of 5 distributions. By summing up for all groups:

$$\sum_{i=0}^{t-1} R_{p_i}(t) \geq t\frac{b-a}{2t}\left(c_1 - \frac{c_2}{t}\right) = \frac{b-a}{2}\left(c_1 - \frac{c_2}{t}\right)$$

which is what we wanted to prove. ■

Next, we state a lemma regarding an upper bound on the expectation of the error in each p_i.

Lemma 14. *Let $p, q \in [0,1]$ and Let $P = B(p)$, $Q = B(q)$ be two Bernoulli distributions with corresponding product distributions P^t, Q^t. Moreover, let $\hat{\theta}_t : \{0,1\}^t \to [0,1]$ be an estimator. Then*

$$\left| \mathbf{E}_{P^t}\left[|\hat{\theta}_t - p|\right] - \mathbf{E}_{Q^t}\left[|\hat{\theta}_t - q|\right] \right| \leq \sqrt{\mathbf{E}_{P^t}\left[(\hat{\theta}_t - p)^2\right]}\sqrt{\left(\frac{q^2}{p} + \frac{(1-q)^2}{1-p}\right)^t - 1} + |p-q|,$$

Proof. By the triangle inequality we have

$$\left| \mathbf{E}_{P^t}\left[|\hat{\theta}_t - p|\right] - \mathbf{E}_{Q^t}\left[|\hat{\theta}_t - q|\right] \right| \leq \left| \mathbf{E}_{P^t}\left[|\hat{\theta}_t - p|\right] - \mathbf{E}_{Q^t}\left[|\hat{\theta}_t - p|\right] \right| + |p-q| \tag{15}$$

To simplify notation let $X = (X_1, \ldots, X_t)$ be the sample and let $r(X) = |\hat{\theta}_t(X) - p|$. Moreover, denote by $p(x)$ resp. $q(x)$ the density functions of $B(p)$ resp. $B(q)$. $p_t : \{0,1\}^t \to [0,1]$ resp. q_t be the density functions of the product distribution P^t, i.e. namely $p_t(x) = \prod_{i=1}^t p(x_i)$. We have

$$\begin{aligned}
\mathbf{E}_{P_t}[r(X)] - \mathbf{E}_{Q^t}[r(X)] &= \sum_{x\in\{0,1\}^t} r(x)p_t(x) - \sum_{x\in\{0,1\}^t} r(x)q_t(x) \\
&= \sum_{x\in\{0,1\}^t} r(x)(p_t(x) - q_t(x)) \\
&= \sum_{x\in\{0,1\}^t} r(x)\sqrt{p_t(x)}\,\frac{p_t(x) - q_t(x)}{\sqrt{p_t(x)}}
\end{aligned}$$

From Cauchy-Schwarz inequality we have

$$\left(\mathbf{E}_{P^t}[r(X)] - \mathbf{E}_{Q^t}[r(X)]\right)^2 \leq \left(\sum_{x\in\{0,1\}^t} r^2(x)p_t(x)\right)\left(\sum_{x\in\{0,1\}^t} \frac{(p_t(x) - q_t(x))^2}{p_t(x)}\right) \tag{16}$$

Notice that

$$\begin{aligned}
\sum_{x\in\{0,1\}^t} \frac{(p_t(x) - q_t(x))^2}{p_t(x)} &= \sum_{x\in\{0,1\}^t} p_t(x) - 2q_t(x) + \frac{q_t^2(x)}{p_t(x)} \\
&= -1 + \sum_{x\in\{0,1\}^t} \frac{q_t^2(x)}{p_t^2(x)} p_t(x) \\
&= -1 + \mathbf{E}_{P^t}\left[\frac{q_t^2(X)}{p_t^2(X)}\right]
\end{aligned}$$

Since X_i are independent we have

$$\mathbf{E}_{P^t}\left[\prod_{i=1}^t \frac{q^2(X_i)}{p^2(X_i)}\right] = \left(\mathbf{E}_P\left[\frac{q^2(X_1)}{p^2(X_1)}\right]\right)^t = \left(\frac{q^2}{p} + \frac{(1-q)^2}{1-p}\right)^t \tag{17}$$

Combining equations (15), (16), (17) yields the result. ■

Now, we find an upper bound for the quantity $\sum_{i=0}^{t-1} E_i$. Our goal is to find an upper bound where the coefficient of $(b-a)/t$ is less than $1/4$.

Lemma 15.

$$\frac{1}{t}\sum_{i=0}^{t-1} E_i \leq \frac{c_a^b}{b-a}\int_a^b \mathbf{E}\left[|\hat{\theta}_t - p|\right] dp + \frac{b-a}{4t}$$

Proof. By Lemma (14) we get that for every $p \in [-(b-a)/2t + p_i, p_i + (b-a)/2t]$:

$$E_i \leq c_a^b \mathbf{E}\left[|\hat{\theta}_t - p|\right]_p + |p_i - p|$$

We integrate both sides of the equation with respect to p and get:

$$\int_{p_i-\frac{b-a}{2t}}^{p_i+\frac{b-a}{2t}} E_i dp \leq c_a^b \int_{p_i-\frac{b-a}{2t}}^{p_i+\frac{b-a}{2t}} \mathbf{E}\left[|\hat{\theta}_t - p|\right]_p dp + \int_{p_i-\frac{b-a}{2t}}^{p_i+\frac{b-a}{2t}} |p_i - p|\, dp$$

Hence:

$$\begin{aligned}\frac{b-a}{t}E_i &\leq c_a^b \int_{p_i-\frac{b-a}{2t}}^{p_i+\frac{b-a}{2t}} \mathbf{E}\left[|\hat{\theta}_t - p|\right]_p dp + 2\int_{p_i-\frac{b-a}{2t}}^{p_i} |p_i - p|\, dp \\ &= c_a^b \int_{p_i-\frac{b-a}{2t}}^{p_i+\frac{b-a}{2t}} \mathbf{E}\left[|\hat{\theta}_t - p|\right]_p dp + \frac{1}{4}\frac{(b-a)^2}{t^2}\end{aligned}$$

Therefore:

$$\frac{E_i}{t} \leq \frac{c_a^b}{b-a}\int_{p_i-\frac{b-a}{2t}}^{p_i+\frac{b-a}{2t}} \mathbf{E}\left[|\hat{\theta}_t - p|\right]_p dp + \frac{b-a}{4t^2} \tag{18}$$

By summing up Equation (18) for all t intervals, we get:

$$\begin{aligned}\frac{\sum_{i=0}^{t-1} E_i}{t} &\leq \frac{c_a^b}{b-a}\sum_{i=0}^{t-1}\int_{p_i-\frac{b-a}{2t}}^{p_i+\frac{b-a}{2t}} \mathbf{E}\left[|\hat{\theta}_t - p|\right]_p dp + \frac{b-a}{4t} \\ &= \frac{c_a^b}{b-a}\int_a^b \mathbf{E}\left[|\hat{\theta}_t - p|\right]_p dp + \frac{b-a}{4t}\end{aligned}$$

■