No-Regret Learning in Repeated Bayesian Games

Jason Hartline Vasilis Syrgkanis Éva Tardos

Abstract

Recent price-of-anarchy analyses of games of complete information suggest that coarse correlated equilibria, which characterize outcomes resulting from no-regret learning dynamics, have near-optimal welfare. This work provides two main technical results that lift this conclusion to games of incomplete information, a.k.a., Bayesian games. First, near-optimal welfare in Bayesian games follows directly from the smoothness-based proof of near-optimal welfare in the same game when the private information is public. Second, no-regret learning dynamics converge to Bayesian coarse correlated equilibrium in these incomplete information games. These results are enabled by interpretation of a Bayesian game as a stochastic game of complete information.

1 Introduction

A recent confluence of results from game theory and learning theory gives a simple explanation for why good outcomes in large classes of games can be expected, in spite of the inherent complexity of these games and their equilibria. Many of games of interest in the computer science literature, from traffic routing to auctions, model repeated interactions. Nash equilibria of the game being repeated correspond to a stable outcome of the repeated game dynamic. However, it is well known that in most games natural game play does not lead to such stable outcomes (Nash equilibria), under any definition of “natural play”.

Blum et al. [3] suggested modeling game player as no-regret learners. Modeling game participants as no-regret learners is very appealing for a number of reasons. There are a number of simple and natural algorithm that guarantee vanishing regret for a player. Any of a number of multi-armed bandit learning algorithms can guarantee that no-regret in the limit. No-regret learning outcomes circumvent the computational difficulty of finding Nash equilibria in games, and can do this with the minimal feedback typically available in large games arising in most online scenarios, such as online auction platforms: just the player’s own payoff. When repeated interaction is between the exact same set of players, the limit points of no-regret learning form the coarse correlated equilibria of the game, an equilibrium notion that relaxes the notion of Nash equilibrium.

To measure the quality of outcomes in games Koutsoupias and Papadimitriou [8] introduced the Price of anarchy, the ratio of the quality of the worst Nash equilibrium over a socially optimal solution focusing on pure Nash equilibria of full information games. Price of anarchy results have been shown for many games from routing through simple auctions. Roughgarden [10] identifies a simple smoothness condition, exhibited by many games of interest, that implies that all Nash equilibria have near-optimal welfare. Syrgkanis and Tardos [13] extended this to mechanisms.

Over the last decade the work on understanding the quality of outcomes in games has been extended in two orthogonal directions. Bayesian games model really large games, situations when players do not know the exact identities and properties of all opponents, rather need to make decisions optimizing their expected utility based on the distribution of possible opponents. Roughgarden [11] and Syrgkanis [12] showed that price of anarchy results obtained via the smoothness framework also extend to Bayesian versions of the games. The second direction concerns full information games, and is extending the concept...
of game outcomes to learning outcomes. Blum et al. [3] introduced the price of total anarchy for the worst case welfare at a limit point of no-regret learning compared to the socially optimum outcome. Roughgarden [10] showed that price of anarchy analysis obtained via the smoothness framework extends to such learning outcomes, bounding also the price of total anarchy in full information games.

Our results. In this paper we consider learning outcomes in Bayesian games. Many important games model repeated interactions between an uncertain set of participants. Sponsored search, and more generally, online ad-auction market places, are important examples of such games. Platforms are running millions of auctions, with each individual auction slightly different, and of only very small value, but such market places have high enough volume to be the financial basis of large industries. This online auction environment is best modeled by a repeated Bayesian game: the auction game is repeated over time, with the set of participants slightly different each time, depending on many factors from budgets of the players through subtle differences in the opportunities.

Our main result is to show that the price of anarchy bounds achieved via the smoothness property of games or mechanisms extends to results on the quality of learning outcomes in Bayesian games. We achieve this by interpreting a Bayesian game as a stochastic game of complete information: there is a large population of possible participants. The game is played by randomly selecting players from the player pool, where players know only the distribution of participants, and not the actual set selected, when choosing their strategies. Concretely, we show two results. First, we show that if all players use a form of no-regret learning in the stochastic game just defined, then the outcome converges to a form of Bayesian coarse correlated equilibrium. There has been several different notions of correlated equilibrium introduced for incomplete information games, the version we use here is the analogue of the Bayesian correlated equilibrium introduced by Forges [6], and is the coarse correlated equilibrium of the corresponding stochastic full information game. Second, we show that if a game or mechanism is smooth, then the corresponding stochastic game is also smooth with the same smoothness parameters. Putting these two results together, our main contribution is to show that the price of anarchy results obtained via the smoothness technique extend to outcomes of no-regret learning in Bayesian games.

2 Preliminaries

This section describes a general game theoretic environment which includes auctions and resource allocation mechanisms. For this general environment we review the results from the literature for analyzing the social welfare that arises from no-regret learning dynamics in repeated game play. The subsequent sections of the paper will generalize this model and these results to Bayesian games, a.k.a., games of incomplete information.

General Game Form. A general game $\mathcal{M}$ is specified by a mapping from a profile $a \in \mathcal{A} \equiv \mathcal{A}_1 \times \cdots \times \mathcal{A}_n$ of allowable actions of players to an outcome. Player $i$’s utility in this game is determined by a profile of individual values $v \in V \equiv V_1 \times \cdots \times V_n$ and the (implicit) outcome of the game and is denoted $U^{\mathcal{M}}_i(a; v_i)$. In games with a social planner or principal who does not take an action in the game, the utility of the principal is $\mathcal{R}^{\mathcal{M}}(a)$. In many games of interest, such as auctions or allocation mechanisms, the utility of the principal is the revenue from payments from the players. We will use the term mechanism and game interchangeably.

In the complete information game the valuation profile of the players is common knowledge. Subsequent sections will consider games of incomplete information in the independent private value model, i.e., where player $i$’s value is known only privately to player $i$. 

2
Social Welfare. We will be interested in analyzing the quality of the outcome of the game as defined by the social welfare, which is the sum of the utilities of the players and the principal. We will denote by $SW^M(a; v) = \sum_{i \in [n]} U_i^M(a; v_i) + R^M(a)$ the expected social welfare of mechanism $M$ under a randomized action profile $a$. For any valuation profile $v \in V$ we will denote the optimal social welfare, i.e., the maximum over outcomes of the game of the sum of utilities, by $\text{OPT}(v)$.

No-regret Learning and Coarse Correlated Equilibria. For complete information games, i.e., fixed valuation profile $v$, Blum et al. [3] analyzed repeated play of players using no-regret learning algorithms, and showed that this play converges to a relaxation of Nash equilibrium, namely, coarse correlated equilibrium.

Definition 1 (no-regret). A player achieves no-regret in a sequence of play $a^1, \ldots, a^T$ if his regret against any fixed strategy $a_i'$ vanishes to zero:

$$\lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} U_i^M(a_i', a_{-i}; v_i) - U_i^M(a_t; v_i) = 0.$$ (1)

Definition 2 (Coarse correlated equilibrium - CCE). A randomized action profile $a \in \Delta(A)$ is a coarse correlated equilibrium of a (static) complete information game if for every player $i$ and $a_i' \in A_i$:

$$E_a[U_i(a; v_i)] \geq E_a[U_i(a_i', a_{-i}; v_i)]$$ (2)

Theorem 3 ([3]). The empirical distribution of actions of any no-regret sequence in a repeated game converges to the set of CCE of the static game.

Price of Anarchy of CCE. Roughgarden [10] gave a unifying framework for comparing the social welfare, under various equilibrium notions including coarse correlated equilibrium, to the optimal social welfare by defining the notion of a smooth game. This framework was extended to games like auctions and allocation mechanisms by Syrgkanis [13].

Definition 4 (Smooth Mechanism [13]). A mechanism $M$ is $(\lambda, \mu)$-smooth for some $\lambda, \mu \geq 0$, if for any valuation profile $v \in V$ and for each player $i \in [n]$ there exists a randomized action $a_i^*(v)$, such that for any action profile $a \in A$:

$$\sum_{i \in [n]} U_i^M(a_i^*(v), a_{-i}; v_i) \geq \lambda \cdot \text{OPT}(v) - \mu \cdot R^M(a)$$ (3)

Many important games and mechanisms satisfy this smoothness definition for various parameters of $\lambda$ and $\mu$ (see Figure 1); the following theorem shows that the welfare of any coarse correlated equilibrium in any of these games is nearly optimal.

Theorem 5 (Efficiency of CCE [13]). If a mechanism is $(\lambda, \mu)$-smooth then the social welfare of any coarse correlated equilibrium at least $\frac{\lambda}{\max\{1, \mu\}}$ of the optimal welfare, i.e., the price of anarchy satisfies $\text{PoA} \leq \frac{\max(1, \mu)}{\lambda}$.

Price of Anarchy of No-regret Learning. Following Blum et al. [3], Theorem 3 and Theorem 5 imply that no-regret learning dynamics have near-optimal social welfare.

Corollary 6 (Efficiency of no-regret dynamics [13]). If a mechanism is $(\lambda, \mu)$-smooth then the average welfare of any no-regret dynamics of the repeated game with a fixed player set and valuation profile, achieves average social welfare at least $\frac{\lambda}{\max\{1, \mu\}}$ of the optimal welfare, i.e., the price of anarchy satisfies $\text{PoA} \leq \frac{\max(1, \mu)}{\lambda}$. 

3
Table 1: Examples of smooth games and mechanisms

| Game/Mechanism                                           | $(\lambda, \mu)$-smoothness | PoA         |
|----------------------------------------------------------|-------------------------------|-------------|
| Simultaneous First Price Auction with Submodular Bidders | $(1 - 1/e, 1)$ [13]          | $\frac{c}{e-1}$ |
| First Price Multi-Unit Auction                           | $(1 - 1/e, 1)$ [5]           | $\frac{c}{e-1}$ |
| First Price Position Auction                             | $(1/2, 1)$ [13]              | 2           |
| All-Pay Auction                                          | $(1/2, 1)$ [13]              | 2           |
| Greedy Combinatorial Auction with $d$-complements        | $(1 - 1/e, d)$ [9]           | $\frac{de}{e-1}$ |
| Proportional Bandwidth Allocation Mechanism              | $(1/4, 1)$ [13]              | 4           |
| Submodular Welfare Games                                 | $(1, 1)$-smooth game [14] [10] | 2           |
| Congestion Games with Linear Delays                      | $(5/3, 1/3)$-smooth cost game [10] | 5/2        |

Figure 1: Examples of smooth games and mechanisms

Importantly, Corollary [4] holds the valuation profile $v \in V$ fixed throughout the repeated game play. The main contribution of this paper is in extending this theory to games of incomplete information, e.g., where the values of the players are drawn at random in each round of game play.

3 Learning in Repeated Bayesian Game

We consider the following setting of a repeated Bayesian game: there exist $n$ equally sized populations of players. For each $i \in [n]$ population $P_i$ consists of a finite set of $r$ players. Each player $q \in P_i$ has some valuation $V_i(q)$. We denote with $F_i \in \Delta(V_i)$ the empirical distribution of values in population $P_i$ and with $v_i$ a random sample from $F_i$. In other words, the value of a randomly chosen player from population $P_i$ is distributed according to $F_i$.

We consider the repeated game where at each iteration one player $q$ from each population is sampled uniformly and independently from other populations. The set of chosen players then participate in an instance of game $M$. We assume that each player $q_i \in P_i$, uses some no-regret learning rule to play in this repeated game. An equivalent way of viewing the game is that the set of players is fixed, but their valuations are changing and are drawn independently and identically from distributions $F_i$ at each iteration. Moreover, each player uses a no-regret algorithm for each possible valuation he can have. However, throughout the paper we will adopt the population interpretation of the model. In Definition [7] we describe the structure of the game and our notation more elaborately.

Definition 7. The repeated Bayesian game of $M$ proceeds as follows. In iteration $t$:  

1. Each player $q \in P_i$ in each population $i$ picks an action $a_{iq}^t$. For each population $i$ we denote with $\mu_i : A_i^{P_i}$ a function that takes as input a player $q \in P_i$ and outputs his action $\mu_i(q) = a_{iq}^t$.

2. From each population $i$ one player $q_i^t \in P_i$ is selected uniformly at random. Let $q^t = (q_1^t, \ldots, q_n^t)$ be the chosen profile of players and $\mu^t(q^t) = (\mu_1^t(q_1^t), \ldots, \mu_n^t(q_n^t))$ be the profile of chosen actions.

3. Each player $q_i^t$ participates in an instance of game $M$, in the role of player $i \in [n]$, with action $\mu_i^t(q_i^t)$ and experiences a utility of $U_i^M(\mu_i^t(q_i^t); V_i(q_i^t))$. The rest of the players experience zero utility.

Remark. We point out that for each player in each population to achieve no-regret he does not need to know the distribution of values in the remainder populations. There exist algorithms that can achieve the
no-regret property and simply require an oracle that returns the utility of a player at each iteration. Thus all we need to assume is that each player receives as feedback his utility at each iteration.

Remark. We also note that our results would extend to the case where at each period multiple matchings are sampled independently and players potentially participate in more than one instance of the mechanism $\mathcal{M}$ and potentially with different players from the remaining population. The only thing that the players need to observe in such a setting is their average utility that resulted from their action $\mu_t^i(q) \in A_i$ from all the instances that they participated at the given period. Such a scenario seems an appealing model in online ad auction marketplaces where players receive only average utility feedback from their bids.

Bayesian Price of Total Anarchy In this repeated game setting we want to compare the average social welfare of any sequence of play where each player uses a vanishing regret algorithm versus the average optimal welfare. Moreover, we want to quantify the worst-case such average welfare over all possible valuation distributions within each population:

$$\sup_{\mathcal{F}_1, \ldots, \mathcal{F}_n} \lim_{T \to \infty} \frac{\frac{1}{T} \sum_{t=1}^{T} \text{Opt}(V(q^t))}{\frac{1}{T} \sum_{t=1}^{T} \text{SW}(\mu^t(q^t); V(q^t))}$$

We will refer to this quantity as the Bayesian price of total anarchy. Since the valuation of the chosen player is re-drawn independently at every time step, the average optimal welfare will converge almost surely to the expected ex-post optimal welfare $E_v[\text{Opt}(v)]$ of the static incomplete information setting.

Our main theorem is that smooth mechanisms guarantee approximate efficiency even in repeated Bayesian games and not only in the fixed stationary environment.

Theorem 8 (Main Theorem). If a mechanism is $(\lambda, \mu)$-smooth then the Bayesian price of total anarchy of any vanishing regret sequence of play of the repeated Bayesian game is at most $\frac{\max\{1, \mu\}}{\lambda}$.

Roadmap of the proof. In Section 4 we analyze static Bayesian games and define Bayesian generalizations of coarse correlated equilibria. The repeated Bayesian game can be viewed as a repetition of a static game of incomplete information. In Section 5, we show that any vanishing regret sequence of play of the repeated Bayesian game, will converge almost surely to the Bayesian version of a coarse correlated equilibrium of the static incomplete information game. Therefore the Bayesian price of total anarchy will be upper bounded by the efficiency of guarantee of any Bayesian coarse correlated equilibrium. Finally, in Section 6 we show that the price of anarchy bound of smooth mechanisms directly extends to Bayesian coarse correlated equilibria, thereby providing an upper bound on the Bayesian price of total anarchy of the repeated game.

4 Bayesian Coarse Correlated Equilibria

Observe that the game that is being repeated at each time step can be viewed as Bayesian game among a set of $n$ players where the valuation of each player is private and known only by the player and is drawn from a distribution $\mathcal{F}_i$. Under this interpretation, prior to making her decision on which action to play, each player learns his own valuation $v_i$ and nothing else. A player’s strategy in this Bayesian game, is a mapping $s_i : \mathcal{A}_{V_i}$ from a valuation $v_i \in \mathcal{V}_i$ to an action $a_i \in A_i$. We will denote with $\Sigma_i = \mathcal{A}_{V_i}$ the strategy space of each player and with $\Sigma = \Sigma_1 \times \ldots \times \Sigma_n$. In an incomplete information game we want to measure welfare as compared to the expected ex-post optimal social welfare $E_v[\text{Opt}(v)]$.

When the game that is being repeated is a complete information game where no player has any private information then Theorem 3 gives a strong connection between no-regret learning and the notion of a
coarse correlated equilibrium of the static game. In this work we show an analogue of this connection for the case of Bayesian games.

Defining a coarse correlated equilibrium in an incomplete information game (and its sibling notion of correlated equilibrium) is not such a straightforward task and several notions have been introduced (c.f. Forges [6], Bergemann and Morris [2], Caragiannis et al. [4]). The Bayesian coarse correlated equilibrium that we define here is the analogue of the Bayes correlated equilibrium defined in Section 4.1 of [6] and corresponds to the coarse correlated equilibrium of a corresponding complete information game defined as follows: the utility of a player $i$ in the complete information game is his ex-ante expected utility from the mechanism. The strategy space of player $i$ is $\Sigma_i = A_i^{V_i}$. For a strategy profile $s \in \Sigma$, the utility of a player in the complete information game is then:

$$U^E_i(s) = \mathbb{E}_v[U_i(s(v); v_i)]$$

(5)

A Bayes-CCE is simply a CCE of this complete information game. For completeness we provide the formal definition below.

**Definition 9** (Bayesian coarse correlated equilibrium - Bayes-CCE). A randomized strategy profile $s \in \Delta(\Sigma)$ is a Bayesian coarse correlated equilibrium if for every $s'_i \in \Sigma_i$:

$$\mathbb{E}_v[\mathbb{E}_v[U_i(s(v); v_i)] \geq \mathbb{E}_v[U_i(s'_i(v_i), s_{-i}(v_{-i}); v_i)]$$

(6)

We will refer to the worst-case ratio of the expected optimal social welfare over the expected social welfare of any Bayes-CCE as Bayes-CCE-PoA.

**Remark.** In the above definitions we assumed that the space $\Sigma = A_1^{V_1} \times \ldots \times A_n^{V_n}$ admits probability distributions. This is not always obvious, since this is a space of functions. When the valuation spaces $V_i$ are finite and probability distributions on the action spaces are well defined, then this is true, and hence in the above and subsequent definitions that involve probability distributions over functions $A_i^{V_i}$ we will assume that the valuation space $V_i$ is finite. For continuous valuation spaces, the issue is a special case of the problem of defining mixed strategies in extensive form games with infinite information sets and the reader is directed to the classic work of Aumann [1] for one approach.

**Remark.** We also point out that our definition of Bayes-CCE is inherently different and more restricted than the one defined in Caragiannis et al. [4]. There, a Bayes-CCE is defined as a joint distribution $D$ over $V \times A$, such that if $(v, a) \sim D$ then for any $v_i \in V_i$ and $a'_i(v_i) \in A_i$:

$$\mathbb{E}_{(v,a)}[U_i(a; v_i)] \geq \mathbb{E}_{(v,a)}[U_i(a'_i(v_i), a_{-i}; v_i)]$$

(7)

The main difference is that the product distribution defined by a distribution in $\Delta(\Sigma)$ and the distribution of values, cannot produce any possible joint distribution over $(V, A)$, but the type of joint distributions are restricted to satisfy a conditional independence property described by Forges [6]. Namely that a player $i$’s action is conditionally independent of some other player $j$’s value, given player $i$’s type. Such a conditional independence property seems essential for the guarantees that we will present in this work to extend to a Bayes-CCE and hence do not seem to extend to the notion given in Caragiannis et al. [4].

However, as we will show in Section 5 the no-regret dynamics that we analyze, which are mathematically equivalent to the dynamics in Caragiannis et al. [4], do converge to the smallest set of Bayes-CCE that we define and for which our efficiency guarantees will extend. This extra convergence property is not needed when the mechanism satisfies the stronger semi-smoothness property defined in [4] and thereby was not needed to show efficiency bounds in their setting.
5 Convergence of Bayesian No-Regret to Bayes-CCE

For any given sequence of play of the repeated Bayesian game, which we defined in Definition 7, we define the following sequence of strategy-value pairs \((s^t, v^t)\) where \(s : \mathcal{V} \to \mathcal{A}\):

\[
s^t(v) = \begin{cases} 
\mu^t(q^t) & \text{if } V(q^t) = v \\
\text{arbitrary } a \in \mathcal{A} & \text{o.w.}
\end{cases}
\]  

(8)

and \(v^t = V(q^t)\). Then observe that all that matters to compute the average social welfare of the game for any given time step \(T\), is the empirical distribution of \((s, v)\), up till time step \(T\), denoted as \(D^T\), i.e. if \((s^T, v^T)\) is a random sample from \(D^T\):

\[
\frac{1}{T} \sum_{t=1}^{T} SW(\mu^t(q^t); V(q^t)) = \mathbb{E}_{(s,v)}[SW(s^T(v^T); v^T)]
\]  

(9)

Lemma 10 (Almost sure convergence to Bayes-CCE). Let \(D \in \Delta(\Sigma \times \mathcal{V})\) be a joint distribution of (strategy, valuation) profile pairs. Consider a sequence of play of the random matching game, where each player uses a vanishing regret algorithm and let \(D^T\) be the empirical distribution of strategy, valuation profile pairs up till time step \(T\). Suppose that there exists a subsequence of \(\{D^T\}^\infty\) that converges in distribution to \(D\). Then, almost surely, \(D\) is a product distribution, i.e. \(D = D_s \times D_v\), with \(D_s \in \Delta(\Sigma)\) and \(D_v \in \Delta(\mathcal{V})\) such that \(D_v = \mathcal{F}\) and \(D_s \in \text{Bayes-CCE}\) of the static incomplete information game with distributional beliefs \(\mathcal{F}\).

Proof. We will denote with

\[
r_i(a^*_i, a; v_i) = U_i(a^*_i, a; v_i) - U_i(a; v_i),
\]

the regret of a player for action \(a^*_i\) at action profile \(a\), when he has value \(v_i\). For a \(q \in P_i\) let \(x^i_t(q) = 1_{q^t=q}\).

Since the sequence has vanishing regret for each player in population \(P_i\), it must be that for any \(q_i \in P_i\) and any \(s^t_i : \mathcal{V} \to \mathcal{A}\):

\[
\sum_{t=1}^{T} x^i_t(q_i) \cdot r_i(s^*_i(V_i(q_i)); \mu^t(v_i; q^t); V_i(q_i)) \leq o(T)
\]

Summing over all players in \(P_i\) we get:

\[
\sum_{t=1}^{T} r_i(s^*_i(V_i(q^t)); \mu^t(q^t); V_i(q^t)) \leq o(T)
\]

Using the definition of \(s^t\) from Equation 8 and \(v^t = V_i(q^t)\), we can rewrite the above as:

\[
\sum_{t=1}^{T} r_i(s^*_i(v^t); s^t(v^t); v^t) \leq o(T)
\]

For any fixed \(T\), let \(D^T_s \in \Delta(\Sigma)\) denote the empirical distribution of \(s^t\) and let \(s^t\) be a random sample from \(D^T_s\). For each \(s \in \Sigma\), let \(T_s \subset [T]\) denote the time steps such that \(s^t = s\) for each \(t \in T_s\). Then we get:

\[
\mathbb{E}_{s} \left[ \frac{1}{|T_s|} \sum_{t \in T_s} r_i(s^*_i(v^t), s(v^t); v^t) \right] \leq \frac{o(T)}{T}
\]
For any $s \in \Sigma$, let $T_{s,v} = \{ t \in T_s : v^t = v \}$. Then we can re-write:

$$E_s \left[ \sum_{v \in V} \frac{|T_{s,v}|}{|T_s|} r_i \left( s_i^*(v_i), s(v); v_i \right) \right] \leq \frac{o(T)}{T} \tag{10}$$

Now we observe that $\frac{|T_{s,v}|}{|T_s|}$ is the empirical frequency of the valuation vector $v \in V$, when filtered at time steps where the strategy vector was $s$. Since at each time step $t$ the valuation vector $v^t$ is picked independently from the distribution of valuation profiles $F$, this is the empirical frequency of $T_s$ independent samples from $F$.

By standard arguments from empirical processes theory, if $T_s \to \infty$ then this empirical distribution converges almost surely to the distribution $F$. On the other hand if $T_s$ doesn’t go to $\infty$, then the empirical frequency of strategy $s$ vanishes to 0 as $T \to \infty$ and therefore has measure zero in the above expectation as $T \to \infty$. Thus for any convergent subsequence of $\{D^T\}$, if $D$ is the limit distribution, then if $s$ is in the support of $D$, then almost surely the distribution of $v$ conditional on strategy $s$ is $F$. Thus we can write $D$ as a product distribution $D_s \times F$.

Moreover, if we denote with $v$ the random variable that follows distribution $F$, then the limit of Inequality (10) for any convergent subsequence, will give that:

$$a.s.: E_s E_v \left[ r_i \left( s_i^*(v_i), s(v); v_i \right) \right] \leq 0$$

Thus $D_s^T$ is in the set of Bayes-CCE of the static incomplete incomplete information game among $n$ players, where the valuation profile is drawn from $F$.

**Theorem 11.** The Bayesian price of total anarchy is upper bounded by the Bayesian price of anarchy of Bayesian coarse correlated equilibria.

6 **Efficiency of Smooth Mechanisms at Bayes Coarse Correlated Equilibria**

In this section we show that smoothness of a mechanism $\mathcal{M}$ implies that any Bayes-CCE of the incomplete information setting achieves at least $\frac{\lambda}{\max(1, \mu)}$ of the expected optimal welfare. To show this we will adopt the interpretation of Bayes-CCE that we used in the previous section, as coarse correlated equilibria of a more complex normal form game. We can interpret this complex normal form game as the game that arises from a complete information mechanism $\mathcal{M}^{EX}$. The main theorem that we will show is that whenever mechanism $\mathcal{M}$ is $(\lambda, \mu)$-smooth, then also mechanism $\mathcal{M}^{EX}$ is $(\lambda, \mu)$-smooth. Then we will invoke a theorem of [13] [10], which shows that any coarse correlated equilibria of a mechanism in the complete information setting achieve at least $\frac{\lambda}{\max(1, \mu)}$ of the optimal welfare.

More formally, we define the mechanism $\mathcal{M}^{EX}$ where the action space of a player is the set of functions from a type to an action of the complete information setting, $\mathcal{A}^{EX}_i \equiv (V_i \to A_i)$. The utility of a player in this mechanism is his ex-ante expected utility, $U_i^{EX}(s) = E_v \left[ U_i^M(s(v)) \right]$. Observe that the utility $U_i^{EX}$ does not depend on any private information of the player. It only depends on the underlying utility function $U_i^M$ and on the distributions $\mathcal{F}_i$ which are all common knowledge. The set of possible outcomes in this ex-ante game corresponds to the set of mappings from an allocation profile to an outcome in the underlying mechanism $\mathcal{M}$. The optimal welfare of this game, is then the expected ex-post optimal welfare $Opt^{EX} = E_v \left[ Opt(v) \right]$.

The main result of this section is to show that if a mechanism $\mathcal{M}$ is smooth according to definition [1] then $\mathcal{M}^{EX}$ is also a smooth mechanism and therefore by [13] [10] every coarse correlated equilibria of $\mathcal{M}^{EX}$,
which corresponds to a Bayes-CCE of the incomplete information game, will achieve at least \( \frac{\lambda}{\max\{1, \mu\}} \) of the expected optimal welfare.

**Theorem 12** (From complete information to Bayesian smoothness). If a mechanism \( \mathcal{M} \) is \((\lambda, \mu)\)-smooth, then for any vector of independent valuation distributions \( \mathcal{F} = (\mathcal{F}_1, \ldots, \mathcal{F}_n) \), the ex-ante complete information mechanism \( \mathcal{M}^{\text{EX}} \) defined by the incomplete information setting is also \((\lambda, \mu)\)-smooth.

**Proof.** Consider the following randomized deviation for each player \( i \) that depends only on the information that he has which is his own value \( v_i \): he random samples a valuation profile \( w \sim \mathcal{F} \). Then he plays according to the randomized action \( s_i^*(v_i, w_{-i}) \), i.e., the player deviates using the randomized action guaranteed by the smoothness property for his true type \( v_i \) and the random sample of the types of the others \( w_{-i} \). Since this is not a profitable deviation for player \( i \):

\[
E_{v, w}[U_i(s_i^*(v_i, w_{-i}), s_{-i}(v_{-i}); v_i)] = E_{v, w}[U_i(s_i^*(w_i, w_{-i}), s_{-i}(v_{-i}); w_i)]
\]

\[
= E_{v, w}[U_i(s_i^*(w), s_{-i}(v_{-i}); w_i)],
\]

where the first equation is an exchange of variable names and regrouping using independence. Summing over players and using smoothness of \( \mathcal{M} \):

\[
\sum_{i \in [n]} U_i^{\text{EX}}(s_i^*, s_{-i}) = E_{v, w} \left[ \sum_{i \in [n]} U_i(s_i^*(w), s_{-i}(v_{-i}); w_i) \right]
\]

\[
\geq E_{v, w} \left[ \lambda \text{OPT}(w) - \mu R^{\mathcal{M}}(s(v)) \right] = \lambda E_{w} \left[ \text{OPT}(w) \right] - \mu R^{\text{EX}}(s)
\]

**Corollary 13.** Every Bayes-CCE of the incomplete information setting of a smooth mechanism \( \mathcal{M} \), achieves welfare at least \( \frac{\lambda}{\max\{1, \mu\}} \) of the expected optimal welfare.

### 7 Finite Time Analysis and Convergence Rates

In this section we analyze the convergence rate to Bayes-CCE and we show approximate efficiency results even for finite time, when players are allowed to have some \( \epsilon \)-regret.

**Theorem 14.** Consider the repeated matching game with a \((\lambda, \mu)\)-smooth mechanism. Suppose that for any \( T \geq T^0 \), each player in each of the \( n \) populations has regret at most \( \frac{\lambda}{\max\{1, \mu\}} \). Then for every \( \delta \) and \( \rho \), there exists a \( T^*(\delta, \rho) \), such that for any \( T \geq \min\{T^0, T^*\} \), with probability \( 1 - \rho \):

\[
\frac{1}{T} \sum_{t=1}^{T} SW(s^t(v^t); v^t) \geq \frac{\lambda}{\max\{1, \mu\}} E_{\nu} \left[ \text{OPT}(v) \right] - \delta - \mu \cdot \epsilon
\]

(11)

Moreover, \( T^*(\delta, \rho) \leq \frac{54n^3|\Sigma||V|^2H^3}{\delta^2}\). \( \log \left( \frac{2}{\rho} \right) \).

**References**

[1] Robert J. Aumann. Mixed and behavior strategies in infinite extensive games. *Advances in Game Theory, Annals of Mathematics Study 52*, pages 627–650, 1964.
[2] Dirk Bergemann and Stephen Morris. Correlated Equilibrium in Games with Incomplete Information. Cowles Foundation Discussion Papers 1822, Cowles Foundation for Research in Economics, Yale University, October 2011.

[3] Avrim Blum, MohammadTaghi Hajiaghayi, Katrina Ligett, and Aaron Roth. Regret minimization and the price of total anarchy. In Proceedings of the Fortyeth Annual ACM Symposium on Theory of Computing, STOC ’08, pages 373–382, New York, NY, USA, 2008. ACM.

[4] Ioannis Caragiannis, Christos Kaklamanis, Panagiotis Kanellopoulos, Maria Kyropoulou, Brendan Lucier, Renato Paes Leme, and ´Eva Tardos. Bounding the inefficiency of outcomes in generalized second price auctions. Journal of Economic Theory, (0):–, 2014.

[5] Bart de Keijzer, Evangelos Markakis, Guido Schfer, and Orestis Telelis. Inefficiency of standard multi-unit auctions. In HansL. Bodlaender and GiuseppeF. Italiano, editors, Algorithms ESA 2013, volume 8125 of Lecture Notes in Computer Science, pages 385–396. Springer Berlin Heidelberg, 2013.

[6] Franoise Forges. Five legitimate definitions of correlated equilibrium in games with incomplete information. Theory and Decision, 35(3):277–310, 1993.

[7] Ramesh Johari and John N. Tsitsiklis. Efficiency loss in a network resource allocation game. Mathematics of Operations Research, 29(3):pp. 407–435, 2004.

[8] Elias Koutsoupias and Christos Papadimitriou. Worst-case equilibria. In Proceedings of the 16th annual conference on Theoretical aspects of computer science, STACS’99, pages 404–413, Berlin, Heidelberg, 1999. Springer-Verlag.

[9] B. Lucier and A. Borodin. Price of anarchy for greedy auctions. In Proceedings of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms, SODA ’10, pages 537–553, Philadelphia, PA, USA, 2010. Society for Industrial and Applied Mathematics.

[10] T. Roughgarden. Intrinsic robustness of the price of anarchy. In Proceedings of the 41st annual ACM symposium on Theory of computing, STOC ’09, pages 513–522, New York, NY, USA, 2009. ACM.

[11] Tim Roughgarden. The price of anarchy in games of incomplete information. SIGecom Exch., 11(1):18–20, June 2012.

[12] Vasilis Syrgkanis. Bayesian games and the smoothness framework. CoRR, abs/1203.5155, 2012.

[13] Vasilis Syrgkanis and ´Eva Tardos. Composable and efficient mechanisms. In Proceedings of the Forty-fifth Annual ACM Symposium on Theory of Computing, STOC ’13, pages 211–220, New York, NY, USA, 2013. ACM.

[14] A. Vetta. Nash equilibria in competitive societies, with applications to facility location, traffic routing and auctions. In Foundations of Computer Science, 2002. Proceedings. The 43rd Annual IEEE Symposium on, pages 416–425, 2002.
A Proof of Theorem 11

Let \( D \in \Delta(\Sigma \times V) \) be a joint distribution, such that there is a subsequence of \( \{D^T\}_T \), converging in distribution to \( D \). Then by Lemma 10 almost surely, \( D \) is a product distribution, i.e. \( D \in \Delta(\Sigma) \times \Delta(V) \) and that the marginal on \( V \) is equal to \( F \) and the marginal on \( \Sigma \) is a Bayes-CCE of the static incomplete information game with distributional beliefs \( F \).

Therefore, if \( \rho \) is the Bayes-CCE - PoA of the mechanism, and if \( (s,v) \) is a random sample from \( D \), then almost surely:

\[
E_{s,v}[SW(s;v)] \geq \frac{1}{\rho}E_v[OPT(v)]
\]  

Thus the limit average social welfare of any convergent subsequence will be at least \( \frac{1}{\rho}E_v[OPT(v)] \), which then implies that almost surely:

\[
\lim_{T \to \infty} \inf \frac{1}{T} \sum_{t=1}^{T} SW(\mu_t^i(q^t);V(q^t)) \geq \frac{1}{\rho}E_v[OPT(v)] = \frac{1}{\rho} \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} OPT(V(q^t))
\]

Thus for any non-measure zero event, for any \( \epsilon \), there exists a \( f(\epsilon) \) such that for any \( T \geq f(\epsilon) \):

\[
\frac{1}{T} \sum_{t=1}^{T} SW(\mu_t^i(q^t);V(q^t)) \geq \frac{1}{\rho} \sum_{t=1}^{T} OPT(V(q^t)) - \epsilon
\]

With no loss of generality we can assume that \( E_v[OPT(v)] > 0 \) (o.w. valuations are all zero and theorem holds trivially). The average optimal welfare converges almost surely to \( E_v[OPT(v)] \), we get that for any non-measure zero event, there exists a \( g(\delta) \) such that for \( T \geq g(\delta) \), \( \frac{1}{T} \sum_{t=1}^{T} OPT(V(q^t)) \) is bounded away from zero. Thereby, we can turn the additive error into a multiplicative one, i.e. for any non-measure zero event and for any \( \epsilon' \) there exists \( w(\epsilon') \) such that for any \( T \geq w(\epsilon') \):

\[
\frac{1}{T} \sum_{t=1}^{T} SW(\mu_t^i(q^t);V(q^t)) \geq \frac{1}{\rho} (1 + \epsilon') \frac{1}{T} \sum_{t=1}^{T} OPT(V(q^t))
\]

This implies that almost surely:

\[
\lim_{T \to \infty} \sup \frac{1}{T} \sum_{t=1}^{T} OPT(V(q^t)) \leq \rho = \text{Bayes-CCE-PoA}
\]

B Proof of Theorem 14

Fix a population \( i \) and a bayesian strategy \( s^*_i \in \Sigma_i \), as well as a bayesian strategy profile \( s \in \Sigma \). For shorter notation we will denote:

\[
\pi_i(s^*_i, s, v) = U_i(s^*_i(v_i), s_{-i}(v_{-i}); v_i).
\]

For a time step \( T \), let \( p^T(s) = \frac{|T|}{T} \) be the empirical distribution of a bayesian strategy \( s \) and with \( p^T(v|s) = \frac{|T|}{T} \) be the empirical distribution of values conditional on a bayesian strategy \( s \). The average utility of a population \( i \) up till time step \( T \), when switching to a fixed bayesian strategy \( s^*_i \), can be written as:

\[
\frac{1}{T} \sum_{t=1}^{T} \pi(s^*_i, s^t, v^t) = \sum_{s \in \Sigma} p^T(s) \sum_{v \in V} p^T(v|s) \cdot \pi_i(s^*_i, s, v)
\]  

(13)
We will show that for any \( s^*_i \), there exists a \( T^*(\delta, \rho) \) such that for any \( T \geq T^*(\delta, \rho) \), with probability \( 1 - \rho \):

\[
\sum_{s \in \Sigma} p^T(s) \sum_{v \in \mathcal{V}} p^T(v|s) \cdot \pi_i(s^*_i, s, v) \geq \sum_{s \in \Sigma} p^T(s) \mathbb{E}_v [\pi_i(s^*_i, s, v)] - \frac{\epsilon}{n} \tag{14}
\]

where \( v \) is a random variable drawn from the distribution of valuation profiles \( \mathcal{F} \). We will denote with \( p(v) \) the density function implied by distribution \( \mathcal{F} \).

In what follows we will denote with \( H = \max_{i \in [n], v_i \in \mathcal{V}_i, x_i \in \mathcal{X}_i} v_i(x_i) \) the maximum possible value of any player. Thus observe that the utility of any player is upper bounded by \( H \) and that the revenue collected by any player at equilibrium is upper bounded by \( H \).

For a time period \( T \), let \( G = \{ s \in \Sigma : p^T(s) \geq \zeta \} \). Then observe that:

\[
\sum_{s \in \Sigma} p^T(s) \sum_{v \in \mathcal{V}} (p^T(v|s) - p(v)) \cdot \pi_i(s^*_i, s, v) \geq \sum_{s \in G} p^T(s) \sum_{v \in \mathcal{V}} (p^T(v|s) - p(v)) \cdot \pi_i(s^*_i, s, v) - \zeta \cdot |\Sigma| \cdot H
\]

Observe that for any \( s \in G \), \( |T_s| \geq \zeta \cdot T \). Thus \( p^T(v|s) \) is the empirical mean of at least \( \zeta \cdot T \) independent random samples of a Bernoulli trial with success probability \( p(v) \). Hence, by Hoeffding bounds, we have that \( |p^T(v|s) - p(v)| \leq t \) with probability at least \( 1 - 2 \exp(-2 \cdot \zeta \cdot T \cdot t^2) \). Thus with that much probability we get:

\[
\sum_{s \in \Sigma} p^T(s) \sum_{v \in \mathcal{V}} (p^T(v|s) - p(v)) \cdot \pi_i(s^*_i, s, v) \geq -t \cdot |\mathcal{V}| \cdot H - \zeta \cdot |\Sigma| \cdot H
\]

By setting \( t = \frac{\epsilon}{2 |\mathcal{V}| \cdot |\Sigma|^2} \) and \( \zeta = \frac{16 |\Sigma| \cdot |\mathcal{V}|^2 \cdot H^2}{n^2} \log \left( \frac{2}{\rho} \right) \), we get the claimed property in Equation (14).

Now suppose that after time step \( T^0 \) each player in a population has regret \( \epsilon/n \). Thus the average utility of the population is at least the utility from switching to any fixed bayesian strategy \( s^*_i \), minus an error term of \( \epsilon/n \):

\[
\sum_{s \in \Sigma} p^T(s) \sum_{v \in \mathcal{V}} p^T(v|s) \pi_i(s_i, s, v) \geq \sum_{s \in \Sigma} p^T(s) \sum_{v \in \mathcal{V}} p^T(v|s) \pi_i(s^*_i, s, v) - \frac{\epsilon}{n} \tag{15}
\]

From the previous analysis, for any \( T \geq \min\{T^0, T^*(\frac{2\delta}{\rho}, \rho)\} \), we get that with probability \( 1 - \rho \):

\[
\sum_{s \in \Sigma} p^T(s) \sum_{v \in \mathcal{V}} p^T(v|s) \pi_i(s_i, s, v) \geq \sum_{s \in \Sigma} p^T(s) \mathbb{E}_v [\pi_i(s^*_i, s, v)] - \frac{2\delta}{3n} - \frac{\epsilon}{n} \tag{16}
\]

Summing over all populations and using the bayesian smoothness property of the mechanism from Theorem 12 we have that with probability \( 1 - \rho \):

\[
\sum_{s \in \Sigma} p^T(s) \sum_{v \in \mathcal{V}} \sum_i \pi_i(s_i, s, v) \geq \sum_{s \in \Sigma} p^T(s) (\lambda \mathbb{E}_v [\text{OPT}(v)] - \mu \mathcal{R}^\text{EX}(s)) - \frac{2\delta}{3} - \epsilon
\]

\[
\geq \lambda \mathbb{E}_v [\text{OPT}(v)] - \mu \sum_{s \in \Sigma} p^T(s) \mathcal{R}^\text{EX}(s) - \frac{2\delta}{3} - \epsilon
\]

To conclude the theorem we observe that since for any \( s \in \Sigma \), \( |p^T(v|s) - p(v)| \leq \frac{\delta}{3 n |\mathcal{V}| |\Sigma|^2} \), we get that:

\[
\mathcal{R}^\text{EX}(s) = \sum_{v \in \mathcal{V}} p(v) \mathcal{R}(s(v)) \leq \sum_{v \in \mathcal{V}} p^T(v|s) \mathcal{R}(s(v)) + \frac{\delta}{3} \tag{17}
\]
Since, the revenue collected by a player at any action in the support of an equilibrium is at most $H$. By the latter we can combine the revenue on the right hand side with the utility on the left hand side. We can also bound the remaining $(\mu - 1)$ of the revenue, by $(\mu - 1)$ of the average welfare minus $\epsilon$, since each player in each population can always drop out of the auction and therefore his average utility at an $\frac{\epsilon}{n}$-regret sequence must be at least $-\frac{\epsilon}{n}$.

Hence, we get that:

$$\sum_{s \in \Sigma} \mathbb{P}(s) \sum_{v \in V} \mathbb{P}(v|s) SW(s(v); v) \geq \frac{\lambda}{\max\{1, \mu\}} \mathbb{E}_v[\text{Opt}(v)] - \delta - \mu \cdot \epsilon$$

(18)

Thus choosing $T^*(\rho, \frac{2H}{3n}) = \frac{54 \cdot n^2 \cdot |\Sigma| \cdot |V|^2 \cdot H^3}{\delta^3} \log \left( \frac{2}{\rho} \right)$, we get the conditions of the theorem.