

RARE NASH EQUILIBRIA AND THE PRICE OF ANARCHY IN LARGE STATIC GAMES

DANIEL LACKER AND KAVITA RAMANAN

ABSTRACT. We study a static game played by a finite number of agents, in which agents are assigned independent and identically distributed random types and each agent minimizes its objective function by choosing from a set of admissible actions that depends on its type. The game is anonymous in the sense that the objective function of each agent depends on the actions of other agents only through the empirical distribution of their type-action pairs. We study the asymptotic behavior of Nash equilibria, as the number of agents tends to infinity, first by deriving laws of large numbers characterizing almost sure limit points of Nash equilibria in terms of so-called Cournot-Nash equilibria of an associated nonatomic game. Our main results are large deviation principles that characterize the probability of rare Nash equilibria and associated conditional limit theorems describing the behavior of equilibria conditioned on a rare event. The results cover situations when neither the finite-player game nor the associated nonatomic game has a unique equilibrium. In addition, we study the asymptotic behavior of the price of anarchy, complementing existing worst-case bounds with new probabilistic bounds in the context of congestion games, which are used to model traffic routing in networks.

1. INTRODUCTION

1.1. Model Description. We consider a static game played by n agents $i = 1, \dots, n$, in which the i^{th} agent is assigned a type w_i from a type space \mathcal{W} , and is allowed to choose an action x_i from a subset $\mathcal{C}(w_i)$ of the action space \mathcal{X} , so as to minimize its objective function $J_i^n(w_1, \dots, w_n, x_1, \dots, x_n)$, which can be viewed as a cost. Both \mathcal{W} and \mathcal{X} are assumed to be metric spaces, and for $w \in \mathcal{W}$, $\mathcal{C}(w)$ represents the set of admissible actions allowed for an agent of type w . We restrict our attention to anonymous games, in which an individual agent's objective function is influenced by its own type and action, but depends on the other agents' types and actions only through the empirical distribution of type-action pairs (rather than on the full configuration of types and actions of individual agents), and in addition, the form of this dependence is the same for all agents. More precisely, if we let $\delta_{(w,x)}$ denote the Dirac delta measure at the point $(w, x) \in \mathcal{W} \times \mathcal{X}$, the objective function of the i^{th} agent in the n -player game takes the form

$$J_i^n(w_1, \dots, w_n, x_1, \dots, x_n) = F\left(\frac{1}{n} \sum_{k=1}^n \delta_{(w_k, x_k)}, w_i, x_i\right),$$

for a suitable function $F : \mathcal{P}(\mathcal{W} \times \mathcal{X}) \times \mathcal{W} \times \mathcal{X} \mapsto \mathbb{R}$, where $\mathcal{P}(\mathcal{W} \times \mathcal{X})$ is the set of Borel probability measures on $\mathcal{W} \times \mathcal{X}$. We are interested in properties of Nash equilibria when the number of players is large and seek to understand the behavior of agents in terms of their *types*, not their *names*. Here, a *Nash equilibrium with type vector* $\vec{w} = (w_1, \dots, w_n) \in \mathcal{W}^n$ is any vector

2010 *Mathematics Subject Classification.* Primary: 60F10, 91A10, 91A15 Secondary: 91A13, 91A43.

Key words and phrases. nonatomic games, mean field games, Nash equilibrium, Cournot-Nash equilibrium, large deviation principle, price of anarchy, congestion games, entry games, conditional limit theorems.

The first author was supported by the National Science Foundation under Award No. DMS 1502980.

The second author was partially supported by the grant NSF CMMI 1538706.

$(x_1, \dots, x_n) \in \mathcal{X}^n$ such that for each i , x_i lies in the set $\mathcal{C}(w_i)$ of admissible actions and the objective function satisfies

$$J_i^n(w_1, \dots, w_n, x_1, \dots, x_n) = \inf_{y \in \mathcal{C}(w_i)} J_i^n(w_1, \dots, w_n, x_1, \dots, x_{i-1}, y, x_{i+1}, \dots, x_n).$$

Note that an agent's type w_i influences not only the objective function but also the set $\mathcal{C}(w_i)$ of admissible actions. For simplicity, we work exclusively with pure strategies, although we refer the interested reader to Blanchet-Carlier [7, Section 4] for extensions of a similar model setup to cover mixed strategies. For pure strategies, even existence of a Nash equilibrium in an n -player game is not always guaranteed, but we will be concerned with the large class of games for which Nash equilibria are known to exist (see Section 2.5). However, such games often admit multiple equilibria, and so we will in general not assume uniqueness of Nash equilibria.

1.2. Discussion of Results and Related Work. It is in general hard to explicitly identify the set of Nash equilibria, especially in large games. Thus, we instead study the behavior of Nash equilibria in the limit as the number of agents goes to infinity. Specifically, under the assumption that the types of agents in the n -player game are sampled independently from a common type distribution $\lambda_0 \in \mathcal{P}(\mathcal{W})$, where $\mathcal{P}(\mathcal{W})$ denotes the space of probability measures on \mathcal{W} , the goal of this work is to study the asymptotic behavior of corresponding Nash equilibria as the number of agents goes to infinity. To emphasize that they are random, we will use capital letters $\{W_i\}$ to denote the sequence of i.i.d. types and the array $\{X^n = (X_1^n, \dots, X_n^n)\}$ of associated agent actions in the sequence of n -player games. Under fairly general conditions (see the standing assumption in Section 2 below), we first state a strong law of large numbers (Theorem 2.2) that shows that almost sure limit points of sequences of (random) empirical type-action distributions $\frac{1}{n} \sum_{i=1}^n \delta_{(W_i, X_i^n)}$ associated with Nash equilibria can be characterized as Cournot-Nash equilibria of a certain nonatomic game associated with the type distribution λ_0 . When there is a unique Cournot-Nash equilibrium for the nonatomic game, this implies almost-sure convergence, as $n \rightarrow \infty$, of the empirical type-action distributions of Nash equilibria of n -player games to the corresponding Cournot-Nash equilibrium.

Our precise framework is related to several existing models in the literature. In particular, the nonatomic game is similar to the model considered by Blanchet and Carlier [7], which is itself a reparametrization of the seminal framework of Mas-Colell [30]. The particular definition that we use (see Section 2.1) is a slight generalization that has two new features. First, it allows for the incorporation of a constraint map \mathcal{C} that specifies the admissible set of actions associated with each agent type. This extension is necessary to cover interesting examples such as the class of congestion games described in Section 1.3. Theorem 2.2 is one of many related (and largely equivalent) laws of large numbers in the literature on large games, notably [21, 23, 13, 24, 7]. Second, our model involves *unknown* or *random* types, whereas all of these papers work with *known* or *deterministic* sequences of type vectors (w_1^n, \dots, w_n^n) satisfying $\frac{1}{n} \sum_{i=1}^n \delta_{w_i^n} \rightarrow \lambda_0$. Although limited to complete information and homogeneous beliefs, our model setup is nonetheless also reminiscent of Harsanyi's formalism of *Bayesian games* [22].

The primary focus of this work is the estimation of the probability that a Nash equilibrium of an n -player game makes a large deviation from the law of large numbers limit when n is large. Our first main set of results, stated in Section 2.2, concern the large deviations behavior of any sequence of (random) empirical type-action distributions associated with Nash equilibria, under the assumption that there is a unique Cournot-Nash equilibrium for the corresponding nonatomic game. Specifically, Theorem 2.4 establishes a large deviations principle (LDP) for such a sequence, which provides precise asymptotic estimates of the exponential rate of decay of probabilities of the occurrence of rare Nash equilibria (i.e., those that are far from the Cournot-Nash equilibrium), and the exponential decay rate is expressed in terms of quantities that are

derived from the more tractable nonatomic game. Establishing an LDP (as opposed to just obtaining large deviation bounds) sheds light on the behavior of Nash equilibria conditioned on a rare event, as exemplified by the conditional limit result in Theorem 2.5. **In a sense, the conditional limit theorem tells us precisely *how* large deviations occur, when they do occur, which marks an important practical implication of the LDP.**

Uniqueness of the Cournot-Nash equilibrium holds for many nonatomic games, including the important class of potential games with strictly convex potential. This covers many congestion games, discussed in more detail in Section 2.5. The foundational work on finite potential games is [32], but we refer to [9] for interesting developments on potential games for general (possibly uncountable) type and action spaces. In more recent work [8, 10], Blanchet et al. exploit a connection with optimal transport to develop methods for computing Cournot-Nash equilibria even for non-potential games.

However, there are also cases of interest for which the nonatomic game admits multiple equilibria. To address this situation, in Section 2.3 we also consider the large deviation behavior of the set of empirical distributions induced by all the Nash equilibria of an n -player game. We first state an analogous law of large numbers result in Theorem 2.7 that shows convergence of the sequence of sets of Nash equilibria to the corresponding set of Cournot-Nash equilibria for the nonatomic game, and then establish a corresponding LDP in Theorem 3.6. The choice of topology on the space of sets of distributions for this LDP is rather subtle. One needs to identify a topology that is weak enough for the LDP to hold, but strong enough that the LDP can provide useful information. We show in Corollary 2.9 that, indeed, our LDP provides interesting information on the probability of outliers or rare equilibria even in the non-unique setting. Additionally, as elaborated below, in Section 2.4, we also show that the LDP is useful for obtaining interesting asymptotic results on the price of anarchy.

Our results appear to be the first LDPs for any kind for large games. Philosophically, the paper that is closest to ours is that of Menzel [31], which adopts a similar statistical perspective to large- n asymptotics in order to derive a central limit theorem in addition to a law of large numbers like Theorem 3.5. Although the model specification in [31] is very different from our own, Menzel interprets his results as “expansions” of the n -player games around the nonatomic game “limit”, which is useful because the latter is typically more tractable. Likewise, our results provide asymptotics for n -player quantities in terms of quantities derived from the associated nonatomic game, namely, the rate functions in Theorems 2.4 and 2.8). However, rather than addressing econometric questions as in Menzel, we focus on the probabilistic nature of equilibria arising from a large number of random types.

Finally, we apply our large deviation analysis to derive high-probability bounds of the so-called *price of anarchy* as the number of agents grows. The *price of anarchy*, a term first introduced by Koutsoupias and Papadimitriou [26], is a measure of the degradation of efficiency in a system due to the selfish behavior of its agents, and it is defined roughly as follows. Given a type vector $\vec{w} \in \mathcal{W}^n$, the socially optimal cost is the least average cost of all players over all associated admissible type-action pairs, and the price of anarchy of the n -player game is the ratio of the worst-case (or highest) average cost induced by any Nash equilibrium to the corresponding socially optimum cost (see Section 2.4 for a precise definition). The price of anarchy measures the degradation of efficiency in a system due to the selfish behavior of its agents.

1.3. Motivation. The motivation for this study is twofold. Firstly, our results apply to a class of games introduced by Rosenthal in [34] called *congestion games*, which have found widespread application in modeling traffic routing, in both physical and communications networks, particularly in the field of algorithmic game theory [33]. In the context of traffic modeling, the congestion game is played on a network, represented by a finite graph, and the type of an agent

is associated with a certain source-destination pair, represented by a pair of vertices in the graph. The distribution of types, λ_0 , could be assumed to be known from historical data. Given a realization of these types, the agents of the game can be viewed as drivers who competitively choose their routes (between their associated source and destination) to minimize travel time, leading to a corresponding traffic outcome determined by the Nash equilibrium (or the set of Nash equilibria, when multiple exist). Depending on what types of drivers use the network on a given day, a variety of different equilibrium outcomes may occur, with probabilities determined by the type distribution. Typically, when the number n of drivers is large, the equilibrium is well approximated by the Cournot-Nash equilibrium (of the model with a continuum of agents). But when n is finite, sampling n independent types can lead to exceptional cases, resulting in exceptional equilibrium outcomes; for instance, a certain route (A, B) in the network may by chance face unusually high demand one day, leading to a highly congested equilibrium. Our large deviations results quantitatively characterize such rare equilibrium outcomes, identifying asymptotic decay rates for their probabilities and, perhaps more importantly, revealing *how* rare events are most likely to occur. For example, our results could describe (asymptotically) how likely it is for a given route (A, B) to exceed some prescribed level of congestion and, moreover, how the rest of the network tends to look when route (A, B) is exceptionally congested.

In managing network traffic, an important quantity is the average travel time (latency) faced by the agents. A central planner managing the network might prefer to assign to each agent a *socially optimal* route, which minimizes the average travel time, but this is rarely feasible. When agents choose routes selfishly, to minimize their own travel times, the resulting social cost or average travel time is typically socially suboptimal, and the price of anarchy is a popular measure of this suboptimality [35]. In Section 2.5, we describe the class of congestion games, and illustrate how our main results can be used to provide new probabilistic bounds on the price of anarchy for such games. In particular, Corollary 2.13 shows how to translate a bound on the price of anarchy in a nonatomic game into a high probability bound on the price of anarchy in the corresponding finite (but large) game. In particular, our results complement existing worst-case bounds such as those of Christodoulou and Koutsoupias [15] on the price of anarchy for n -player congestion games determined by a class of linear cost functions by providing with-high-probability bounds for the price of anarchy arising from a fixed cost function in that class.

While we focus on congestion games as a motivating example, our framework encompasses many different types of large static games appearing in applications, with notable examples including *entry games* [5, 11, 4] and *auctions* [27, 25]. In both of these examples, it is more natural to interpret agents as *maximizing a payoff*, which we identify with $-F$, the negative of the cost function. A prototypical entry game, borrowed from [5], has $\mathcal{X} = \{0, 1\}$, an arbitrary type space \mathcal{W} , and payoff $-F(m, w, x) = x[f(m^x\{1\}) + g(w)]$, for some functions f and g , where f is decreasing. The action $x = 1$ means the agent “enters the market.” An agent that does not enter receives no payoff, while an agent that enters receives a payoff which is decreasing in the fraction $m^x\{1\}$ of agents that enter. All of our main results apply to entry games, as long as f and g are continuous. We discuss entry games in somewhat more detail in Section 3.5, as an illustration of our conditional limit theorem.

On the other hand, our results do not apply to many models of auctions, for which the payoff function is discontinuous. More specifically, a typical auction model has $\mathcal{W} = \mathcal{X} \subset [0, \infty)$ and a payoff function $F(m, w, x)$ with discontinuities at points where $x = \max \text{supp}(m^x)$, where $\text{supp}(m)$ represents the support of the distribution m . For instance, in an auction of a single unit of a single good, the classical first-price auction has payoff $-F(m, w, x) = (w - x)1_{\{x \geq \max \text{supp}(m^x)\}}$, with the type w representing the intrinsic value of the good. That is, when the bid x of a given agent becomes the maximum bid, the payoff of the agent jumps from zero to

$w - x$. It is not clear if our main results should still hold in the presence of such discontinuities. Extending our results to include these other applications would be an interesting problem for future work.

An additional motivation for our work relates to the study of Nash equilibria in *dynamic* n -player games, on which a vibrant literature has emerged recently. These games arise in a variety of settings and are harder to analyze than static games. Various law of large numbers type limit theorems and approximation results are now fairly well understood, both in discrete time [38, 2, 1, 20] and in continuous time [29, 12, 28, 19, 14], and are expressed in terms of associated dynamic games with a continuum of agents which largely go by the name of *mean field games*. The present work grew in part out of early efforts to understand large deviations in dynamic mean field games, especially in the case when the mean field game admits multiple equilibria. In many dynamic models, the random variables $\{W_i\}$ which we called *types* are better interpreted as *noises*. For instance, the continuous time models typically involve controlled diffusion processes driven by a sequence $\{W_i\}$ of i.i.d. Brownian motions. This *noise* interpretation is equally valid for the static games of this paper, if we think of W_i as a random shock to agent i . We hope that our large deviation analysis of static games be useful not only on its own merit but also as a first step toward understanding large deviations in dynamic games.

2. STATEMENTS OF MAIN RESULTS

In this section, we precisely state our results, the proofs of which are presented in Section 3, with some auxiliary results required for the proofs deferred to Appendices A and B. In what follows, given a metric space \mathcal{S} , we let $\mathcal{P}(\mathcal{S})$ denote the space of Borel probability measures on \mathcal{S} , equipped with the topology of weak convergence. We will refer to convergence in this topology as convergence in distribution, and denote this convergence by $m_n \rightarrow m$, which we recall means that $\int \varphi dm_n \rightarrow \int \varphi dm$ for every bounded continuous function φ on \mathcal{S} . We will most often consider the case $\mathcal{S} = \mathcal{W}$ or $\mathcal{S} = \mathcal{W} \times \mathcal{X}$. Throughout the paper, we make the following assumptions on the model.

Standing assumptions. The following model parameters are given:

- (1) The *action space* \mathcal{X} is a compact metric space.
- (2) The *type space* is a complete separable metric space \mathcal{W} .
- (3) The constraint map \mathcal{C} , which maps elements of \mathcal{W} to nonempty closed subsets of \mathcal{X} , is continuous. Here, continuity of the set-valued map \mathcal{C} means both that the graph $\text{Gr}(\mathcal{C}) = \{(w, x) \in \mathcal{W} \times \mathcal{X} : x \in \mathcal{C}(w)\}$ is closed and that, if $w_n \rightarrow w$ in \mathcal{W} and $x \in \mathcal{C}(w)$, then there exist n_k and $x_{n_k} \in \mathcal{C}(w_{n_k})$ such that $x_{n_k} \rightarrow x$.
- (4) The *objective function* $F : \mathcal{P}(\mathcal{W} \times \mathcal{X}) \times \mathcal{W} \times \mathcal{X} \rightarrow \mathbb{R}$ is bounded and continuous, where $\mathcal{P}(\mathcal{W} \times \mathcal{X}) \times \mathcal{W} \times \mathcal{X}$ is equipped with the product topology.

The compactness of \mathcal{X} assumed in (1) is important but could likely be replaced by a coercivity assumption on F . In our main application to congestion games, both sets \mathcal{W} and \mathcal{X} are finite, in which case the continuity assumptions (3) and (4) hold automatically. Fix throughout the paper an arbitrary probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and assume it is rich enough to support all of the random variables of interest. We also assume throughout that for each $n \geq 1$ and each type vector $\vec{w} = (w_1, \dots, w_n) \in \mathcal{W}^n$, the set of Nash equilibria with type vector \vec{w} is non-empty.

2.1. Nonatomic games and Cournot-Nash equilibria. Let $\hat{x}_i^n : \mathcal{W}^n \mapsto \mathcal{X}$ be measurable functions such that $(\hat{x}_1^n(\vec{w}), \dots, \hat{x}_n^n(\vec{w}))$ is a Nash equilibrium with type vector \vec{w} , for each $\vec{w} \in \mathcal{W}^n$ (it is shown in Lemma 3.4 that such a measurable selection always exists under our assumptions). Now, suppose that W_1, \dots, W_n are the i.i.d. types sampled from a distribution $\lambda_0 \in \mathcal{P}(\mathcal{W})$, which we fix once and for all. Let $X_i^n = \hat{x}_i^n(W_1, \dots, W_n)$ denote the associated

random Nash equilibrium vector. The equilibrium type-action distribution is the random probability measure (on $\mathcal{W} \times \mathcal{X}$) given by

$$\mu_n := \frac{1}{n} \sum_{i=1}^n \delta_{(W_i, X_i^n)}.$$

Our main results concern the asymptotic behavior of $\{\mu_n\}$, which, as mentioned in the introduction, is expressed in terms of equilibria for the corresponding *nonatomic game*, also called *Cournot-Nash equilibria*, defined as follows.

Definition 2.1 (Cournot-Nash equilibria). For $\lambda \in \mathcal{P}(\mathcal{W})$, the set $\mathcal{CN}(\lambda)$ of *Cournot-Nash equilibria with type distribution* λ is defined as the set of $m \in \mathcal{P}(\mathcal{W} \times \mathcal{X})$ with first marginal equal to λ that satisfy

$$m \left\{ (w, x) \in \mathcal{W} \times \mathcal{X} : x \in \mathcal{C}(w), F(m, w, x) = \inf_{y \in \mathcal{C}(w)} F(m, w, y) \right\} = 1,$$

that is, $x \in \mathcal{C}(w)$ and $F(m, w, x) = \inf_{y \in \mathcal{C}(w)} F(m, w, y)$ hold for m -almost every (w, x) .

Intuitively, a Cournot-Nash equilibrium $m \in \mathcal{CN}(\lambda)$ describes an equilibrium distribution of type-action pairs in a game consisting of a continuum of infinitesimally small agents. Although, in an n -player game (pure-strategy) Nash equilibria need not in general exist, a standard argument in Proposition 3.3 below (adapted from [30]) shows that there always exists a Cournot-Nash equilibrium, i.e., $\mathcal{CN}(\lambda) \neq \emptyset$ for all $\lambda \in \mathcal{P}(\mathcal{W})$.

2.2. Large deviation results for sequences of Nash equilibria. In a Cournot-Nash equilibrium no individual agent has direct influence on the equilibrium distribution m . Agents thus optimize independently, facing i.i.d. types, and a law of large numbers heuristic suggests that m should, in equilibrium, agree with the distribution of type-action pairs. This heuristic is justified by the following rigorous result:

Theorem 2.2. *Given that agent types are i.i.d. with distribution $\lambda_0 \in \mathcal{P}(\mathcal{W})$, for any metric d on $\mathcal{P}(\mathcal{W} \times \mathcal{X})$ compatible with weak convergence, it holds with probability one that $d(\mu_n, \mathcal{CN}(\lambda_0)) := \inf_{m \in \mathcal{CN}(\lambda_0)} d(\mu_n, m) \rightarrow 0$.*

We prove a somewhat more general form of this result in Section 3 (see Theorem 3.5(ii) therein), which allows for approximate Nash equilibria and correlated types, although we do not push this result to the utmost generality because it is not the main novelty of the paper.

Remark 2.3. For an idea of how to adapt Theorem 2.2 to mixed strategies, which we do not explore in this paper, see [7, Theorem 4.2].

We know from Theorem 2.2 that the limit points of $\{\mu_n\}$ lie in the set $\mathcal{CN}(\lambda_0)$. Our first main result, Theorem 2.4 below, lets us estimate how unlikely it is that μ_n remains “far” in some sense from this limiting set. To state the theorem precisely, we introduce some definitions. Write $\lambda \ll \lambda_0$ when λ is absolutely continuous with respect to λ_0 , and define the relative entropy as usual by

$$H(\lambda|\lambda_0) := \int_{\mathcal{W}} \frac{d\lambda}{d\lambda_0} \log \frac{d\lambda}{d\lambda_0} d\lambda, \text{ for } \lambda \ll \lambda_0, \quad H(\lambda|\lambda_0) = \infty \text{ otherwise.} \quad (2.1)$$

Define

$$\mathcal{CN} = \bigcup_{\lambda \in \mathcal{P}(\mathcal{W})} \mathcal{CN}(\lambda), \quad (2.2)$$

to be the set of all Cournot-Nash equilibria, with any type distribution. For $m \in \mathcal{P}(\mathcal{W} \times \mathcal{X})$ let m^w and m^x denote the first and second marginals, respectively, of m . Throughout the paper, we adopt the convention that $\inf \emptyset = \infty$ and $\sup \emptyset = -\infty$.

Theorem 2.4. *Assume that $\mathcal{CN}(\lambda)$ is a singleton for each $\lambda \in \mathcal{P}(\mathcal{W})$ with $\lambda \ll \lambda_0$. Then, for every measurable set $A \subset \mathcal{P}(\mathcal{W} \times \mathcal{X})$,*

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\mu_n \in A) &\leq - \inf_{m \in \overline{A} \cap \mathcal{CN}} H(m^w | \lambda_0), \\ \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\mu_n \in A) &\geq - \inf_{m \in A^\circ \cap \mathcal{CN}} H(m^w | \lambda_0), \end{aligned}$$

where A° and \overline{A} denote the interior and closure, respectively, of A . In other words, $\{\mu_n\}$ satisfies an LDP on $\mathcal{P}(\mathcal{W} \times \mathcal{X})$ with (good) rate function

$$m \mapsto \begin{cases} H(m^w | \lambda_0) & \text{if } m \in \mathcal{CN}, \\ \infty & \text{otherwise.} \end{cases}$$

Theorem 2.4 follows from a more general result, Theorem 3.9, proved in Section 3.4. In applications, one can use Theorem 2.4 to estimate the asymptotic probabilities of what are best interpreted as *rare equilibrium outcomes*. Given an event $A \subset \mathcal{P}(\mathcal{W} \times \mathcal{X})$ whose closure is disjoint from $\mathcal{CN}(\lambda_0)$, for example, $A = \{m \in \mathcal{P}(\mathcal{W} \times \mathcal{X}) : d(m, \mathcal{CN}(\lambda_0)) \geq \epsilon\}$, Theorem 2.2 says that $\mathbb{P}(\mu_n \in A) \rightarrow 0$, and Theorem 2.4 says that this happens exponentially quickly, making the event *rare* in the sense that roughly $\mathbb{P}(\mu_n \in A) \approx e^{-nc_A}$ for a constant $c_A > 0$. Indeed, it is easy to show (see Lemma 3.11 below) that $c_A := \inf_{m \in A \cap \mathcal{CN}} H(m^w | \lambda_0) > 0$ for the particular set A chosen above, so that the upper bound of Theorem 2.4 is nontrivial.

For a more tangible application, for a closed set $B \subset \mathcal{X}$ we can estimate the probability

$$\mathbb{P}(X_i^n \in B \text{ for some } i) = \mathbb{P}(\text{supp}(\mu_n^w) \cap B \neq \emptyset),$$

that the action of some agent belongs to the set B ; here $\text{supp}(m)$ denotes the support of a measure m . For instance, in a traffic congestion game, this event could represent some agent utilizing a seemingly inefficient or slow route. This event is “rare” as long as B does not intersect the support of m_0^x where m_0 is the unique element of $\mathcal{CN}(\lambda_0)$. Again, by “rare” we mean $\inf\{H(m^w | \lambda_0) : m \in \mathcal{CN}, \text{supp}(m^x) \cap B \neq \emptyset\} > 0$, so that the upper bound of Theorem 2.4 is nontrivial.

Theorem 2.4 is of course related to Sanov’s theorem and indeed reduces to it in degenerate cases (e.g., when \mathcal{X} is a singleton). Our framework also admits an analog of Cramér’s theorem: If \mathcal{X} is a subset of a Euclidean space, then we can estimate probabilities involving the *average* of agents’ actions, such as $\mathbb{P}(\frac{1}{n} \sum_{i=1}^n X_i^n \in B)$ for $B \subset \mathcal{X}$.

A full LDP, which explicitly characterizes asymptotic large deviation upper and lower bounds provides information [about the system] that cannot be obtained by just one-sided bounds. Specifically, in the spirit of the so-called Gibbs conditioning principle (see, for instance, [16, 17]), the LDP of Theorem 2.4 can be used to derive the following *conditional limit theorem*, which tells us about the typical behavior of μ_n given that a rare event of the form $\{\mu_n \in A\}$ occurs:

Theorem 2.5. *Let $A \subset \mathcal{P}(\mathcal{W} \times \mathcal{X})$ be measurable and define*

$$I(A) := \inf \{H(\lambda | \lambda_0) : \lambda \ll \lambda_0, \overline{A} \cap \mathcal{CN}(\lambda) \neq \emptyset\}, \quad (2.3)$$

$$S(A) := \{m \in \overline{A} \cap \mathcal{CN} : H(m^w | \lambda_0) = I(A)\}. \quad (2.4)$$

Suppose $I(A) < \infty$. Then $S(A)$ is nonempty and compact. Assume that

$$I(A) = \inf \{H(\lambda | \lambda_0) : \lambda \ll \lambda_0, \mathcal{CN}(\lambda) \subset A^\circ\}, \quad (2.5)$$

and also that $\mathbb{P}(\mu_n \in A)$ is nonzero for all sufficiently large n . Then, letting d denote any metric on $\mathcal{P}(\mathcal{W} \times \mathcal{X})$ compatible with weak convergence, for each $\epsilon > 0$ there exists $c > 0$ such that, for

all sufficiently large n ,

$$\mathbb{P}(d(\mu_n, S(A)) \geq \epsilon \mid \mu_n \in A) \leq e^{-cn}. \quad (2.6)$$

Remark 2.6. In the setting of Theorem 2.5, we can say that every limit point of the sequence of conditional distributions of μ_n given $\{\mu_n \in A\}$, $n \in \mathbb{N}$, is supported on the set $S(A)$. In particular, if $S(A) = \{\nu^*\}$ is a singleton, then these conditional distributions converge to the point mass at ν^* . Finally, if $\mathcal{CN}(\lambda)$ is a singleton for every $\lambda \ll \lambda_0$, then in fact (2.5) is equivalent to the following condition:

$$I(A) = \inf_{m \in A \cap \mathcal{CN}} H(m^w | \lambda_0) = \inf_{m \in A^\circ \cap \mathcal{CN}} H(m^w | \lambda_0). \quad (2.7)$$

One should interpret Theorem 2.5 as identifying the most likely manner in which the rare event $\{\mu_n \in A\}$ occurs. In the regime described in Remark 2.6, conditionally on $\{\mu_n \in A\}$ we expect μ_n to be very close to the unique minimizer ν^* of the rate function over A ; that is, $\nu^* = \arg \min_{\nu \in \bar{A} \cap \mathcal{CN}} H(m^w | \lambda_0)$, where we recognize the minimization problem from the LDP bounds of Theorem 2.4.

The proof of this conditional limit theorem is given in Section 3.5. The challenge in applying Theorem 2.5 lies in checking the assumption (2.5), or equivalently (2.7) when there is uniqueness, and also showing that the set $S(A)$ of (2.4) is a singleton. The key difficulty is that the set \mathcal{CN} is never convex in nontrivial cases, which makes the minimization problems in (2.7) more difficult than those that arise from the usual Gibbs conditioning principle. However, these assumptions can be verified in several cases of interest. As an illustration, in Section 3.5 we discuss in detail a simple example of an entry game in which both assumptions can be verified.

Theorem 2.4 applies to a given sequence (more precisely, triangular array) of Nash equilibria $\{X_i^n, 1 \leq i \leq n\}_{n \in \mathbb{N}}$, under a crucial uniqueness assumption. Notice that the uniqueness assumption is imposed *only at the limit*, for the Cournot-Nash equilibrium, and no uniqueness is required of the equilibria of n -player games. It is evident that some kind of uniqueness assumption at the limit is necessary. Suppose, for instance, that \mathcal{X} contains at least two elements, that $\mathcal{C}(w) = \mathcal{X}$ for all w , and that the cost function is the trivial $F \equiv 0$. Then there is no hope for an LDP because *any choice of actions* is a Nash equilibrium. Uniqueness is known to hold in various particular models as well as for a broad class of games known as *potential games*, at least when the potential is strictly convex, and we will encounter a class of examples in our discussion of congestion games in Section 2.5. Nonetheless, uniqueness is not to be expected in general.

2.3. Large deviation results for the set of equilibria. We now address the case when there are multiple Cournot-Nash equilibria for the limiting nonatomic game. Let $\hat{\mathcal{N}}_n : \mathcal{W}^n \rightarrow 2^{\mathcal{P}(\mathcal{W} \times \mathcal{X})}$ denote the set-valued map that assigns to each type vector the corresponding set of equilibrium type-action distributions:

$$\hat{\mathcal{N}}_n(w_1, \dots, w_n) := \left\{ \frac{1}{n} \sum_{i=1}^n \delta_{(w_i, x_i)} : (x_1, \dots, x_n) \text{ is Nash for types } (w_1, \dots, w_n) \right\}. \quad (2.8)$$

Again, let $\{W_i\}$ be a sequence of i.i.d. \mathcal{W} -valued random variables with distribution λ_0 , and let $\mathcal{N}_n = \hat{\mathcal{N}}_n(W_1, \dots, W_n)$ denote the random set of equilibrium type-action distributions.

It is shown in Proposition 3.2 that $\hat{\mathcal{N}}_n(\vec{w})$ and $\mathcal{CN}(\lambda)$ are always closed sets. Thus, in the following theorem, we topologize the space \mathfrak{C} of closed subsets of $\mathcal{P}(\mathcal{W} \times \mathcal{X})$ with the *upper Vietoris topology*, generated by the base of open sets of the form $\{A \in \mathfrak{C} : A \subset E\}$, where E is an open subset of $\mathcal{P}(\mathcal{W} \times \mathcal{X})$. See Appendix A for a short discussion of the basic properties of this topology, the most important of which is that it topologizes *upper hemicontinuity* of set-valued maps. First, Theorem 2.7 states that \mathcal{N}_n converges almost surely to $\mathcal{CN}(\lambda_0)$, thus

establishing a law-of-large numbers result in the upper Vietoris topology, which we prove in Section 3.3.

Theorem 2.7. *The sequence of random sets $\{\mathcal{N}_n\}$ converges almost surely to $\mathcal{CN}(\lambda_0)$.*

It is worth distinguishing, however, this rather weak choice of topology from more standard notions of set-convergence. For instance, when one says that a sequence of sets A_n in some metric space converge to another set A , this typically means that for every $\epsilon > 0$ it holds for sufficiently large n that both (a) A_n is contained in an ϵ -enlargement of A and (b) A is contained in an ϵ -enlargement of A_n . It is fairly well known, however, that Theorem 2.7 fails if we demand that convergence hold in this stronger sense (see, e.g., [23]). On the other hand, convergence in the upper Vietoris topology corresponds only to (a), not (b). In words, the distinction is this: Theorem [23] says that every (subsequential) limit point of a sequence of n -player Nash equilibria (i.e., a sequence $\mu_n \in \mathcal{N}_n$) must lie in $\mathcal{CN}(\lambda_0)$, but it is not true in general that every element of $\mathcal{CN}(\lambda_0)$ can be achieved as the limit of some subsequence of n -player equilibria.

The next main result is an LDP for the *set of Nash equilibria*. This not only does away with the uniqueness assumption on Cournot-Nash equilibria imposed in Theorem 2.4, but it carries more information than Theorem 2.4 even when there is uniqueness. As shown in Remark 3.7, Theorem 2.8 follows from a more general result, Theorem 3.6, established in Section 3.4.

Theorem 2.8. *For Borel sets $\mathfrak{U} \subset \mathfrak{C}$,*

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\mathcal{N}_n \in \mathfrak{U}) &\leq -\inf\{H(\lambda|\lambda_0) : \lambda \in \mathcal{P}(\mathcal{W}), \mathcal{CN}(\lambda) \in \overline{\mathfrak{U}}\}, \\ \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\mathcal{N}_n \in \mathfrak{U}) &\geq -\inf\{H(\lambda|\lambda_0) : \lambda \in \mathcal{P}(\mathcal{W}), \mathcal{CN}(\lambda) \in \mathfrak{U}^\circ\}. \end{aligned}$$

In other words, $\{\mathcal{N}_n\}$ satisfies an LDP on \mathfrak{C} with (good) rate function

$$A \mapsto \inf\{H(\lambda|\lambda_0) : \lambda \in \mathcal{P}(\mathcal{W}), \mathcal{CN}(\lambda) = A\}. \quad (2.9)$$

At first, this theorem may appear too abstract to be useful, especially given that the upper Vietoris topology is rather coarse (even non-Hausdorff). On the contrary, it yields several interesting concrete results, a key example of which stems from the following simple corollary.

Corollary 2.9. *If $E \subset \mathcal{P}(\mathcal{W} \times \mathcal{X})$ is closed, then*

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\mathcal{N}_n \cap E \neq \emptyset) \leq -\inf\{H(m^w|\lambda_0) : m \in \mathcal{CN} \cap E\}. \quad (2.10)$$

If E is open, then

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\mathcal{N}_n \subset E) \geq -\inf\{H(\lambda|\lambda_0) : \lambda \in \mathcal{P}(\mathcal{W}), \mathcal{CN}_\lambda \subset E\}.$$

Corollary 2.9 can be interpreted in terms of *outliers*, or rare equilibria. Indeed, the left-hand side of (2.10) is the probability that there exists a Nash equilibrium for the n -player game that lies in the set E . If $\mathcal{CN}(\lambda_0) \cap E = \emptyset$, we know from Theorem 2.2 that equilibria in E should be rare when n is large in the sense that $\mathbb{P}(\mathcal{N}_n \cap E \neq \emptyset) \rightarrow 0$. The bound (2.10) shows that this probability decays exponentially and quantifies precisely the exponential decay rate. The proofs of the large deviations results in Theorem 2.8 can be found in Section 3.4 (see Theorem 3.6) and hinge on the well-known *contraction principle*, once the n -player games and the nonatomic game are set on a common topological space (as in Section 3.1).

It should also be mentioned that a map of the form $\mathfrak{C} \ni A \mapsto G(A) := \sup_{m \in A} g(m) \in \mathbb{R}$ is upper semicontinuous whenever g is upper semicontinuous. If g is continuous, and if it is constant on a set A , then G is continuous at A . These facts (proven in Lemma A.2) can be used to derive large deviation bounds for a sequence of random variables of the form $\sup_{m \in \mathcal{N}_n} g(m)$,

which we interpret as the worst case value of g , in equilibrium. The following section investigates a somewhat more complex instance of this observation.

2.4. Price of anarchy. We now provide a precise definition of the price of anarchy for both n -player and nonatomic games. We assume that $F \geq 0$, which is essentially without loss of generality due to the boundedness assumption (4). For each n and each type vector $\vec{w} = (w_1, \dots, w_n) \in \mathcal{W}^n$, define the set of *all admissible* type-action distributions by

$$\hat{\mathcal{A}}_n(w_1, \dots, w_n) := \left\{ \frac{1}{n} \sum_{k=1}^n \delta_{(w_k, x_k)} : x_i \in \mathcal{C}(w_i), i = 1, \dots, n \right\}. \quad (2.11)$$

The *average cost* of the game, for a fixed type-action distribution $m \in \mathcal{P}(\mathcal{W} \times \mathcal{X})$, is defined by

$$V(m) := \int_{\mathcal{W} \times \mathcal{X}} F(m, w, x) m(dw, dx). \quad (2.12)$$

Finally the price of anarchy is the ratio of the worst-case Nash equilibrium cost to the socially optimal cost, or

$$\text{PoA}_n(\vec{w}) := \frac{\sup_{m \in \hat{\mathcal{N}}_n(\vec{w})} V(m)}{\inf_{m \in \hat{\mathcal{A}}_n(\vec{w})} V(m)},$$

where recall the definition of $\hat{\mathcal{N}}_n$ from (2.8). Recall that $\hat{\mathcal{N}}_n(\vec{w})$, and thus $\hat{\mathcal{A}}_n(\vec{w})$, is non-empty due to our standing assumption on the existence of Nash equilibria for n -player games. Assume that V is strictly positive, which by continuity implies that V is bounded from below away from zero on the non-empty compact set $\hat{\mathcal{A}}_n(\vec{w})$, for each fixed n and $\vec{w} \in \mathcal{W}^n$. Moreover, V is bounded since F is bounded by our standing assumption (4). Thus, the numerator above is also a finite positive number. Hence, $\text{PoA}_n(\vec{w})$ is well defined.

Finally, define the price of anarchy for the nonatomic game as follows. For $\lambda \in \mathcal{P}(\mathcal{W})$, set

$$\mathcal{A}(\lambda) := \{m \in \mathcal{P}(\mathcal{W} \times \mathcal{X}) : m^w = \lambda, m\{(w, x) : x \in \mathcal{C}(w)\} = 1\}. \quad (2.13)$$

This is simply the set of all admissible type-action distributions for the nonatomic game with type distribution λ . The price of anarchy is then

$$\text{PoA}(\lambda) := \frac{\sup_{m \in \mathcal{CN}(\lambda)} V(m)}{\inf_{m \in \mathcal{A}(\lambda)} V(m)}.$$

Under our standing assumptions, V is bounded above and by Proposition 3.3, $\mathcal{CN}(\lambda) \neq \emptyset$ for each fixed $\lambda \in \mathcal{P}(\mathcal{W})$. Again, if $V > 0$ pointwise then by continuity V is bounded from below away from zero on the non-empty compact set $\mathcal{A}(\lambda)$, for each fixed $\lambda \in \mathcal{P}(\mathcal{W})$, and $\text{PoA}(\lambda)$ is well defined. As before, let $\{W_i\}$ be i.i.d. \mathcal{W} -valued random variables with distribution λ_0 . See Section 3.6 for the proof of the following:

Proposition 2.10. *Assume $V > 0$ pointwise. It holds almost surely that*

$$\limsup_{n \rightarrow \infty} \text{PoA}_n(W_1, \dots, W_n) \leq \text{PoA}(\lambda_0).$$

Moreover,¹ for each r ,

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\text{PoA}_n(W_1, \dots, W_n) \geq r) &\leq -\inf \{H(\lambda|\lambda_0) : \lambda \in \mathcal{P}(\mathcal{W}), \text{PoA}(\lambda) \geq r\}, \\ \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\text{PoA}_n(W_1, \dots, W_n) < r) &\geq -\inf \{H(\lambda|\lambda_0) : \lambda \in \mathcal{P}(\mathcal{W}), \text{PoA}(\lambda) < r\}. \end{aligned}$$

¹Equivalently, $\text{PoA}_n(W_1, \dots, W_n)$ satisfies an LDP on $(\mathbb{R} \cup \{-\infty\}, \tau)$ with good rate function $r \mapsto \inf \{H(\lambda|\lambda_0) : \lambda \in \mathcal{P}(\mathcal{W}), \text{PoA}(\lambda) = r\}$, where $\tau = \{[-\infty, a) : a \in \mathbb{R} \cup \{-\infty\}\}$ is the lower topology.

2.5. Congestion games. We now introduce the class of congestion games alluded to in Section 1.3. To specify the model, we work with a finite set \mathcal{W} of types. Given a finite set E of *elements*, the action space is the set $\mathcal{X} = 2^E \setminus \{\emptyset\}$ of nonempty subsets. The constraint map \mathcal{C} is arbitrary for the moment. A continuous increasing function $c_e : [0, \infty) \rightarrow [0, \infty)$ is given for each $e \in E$, which represents the *cost* faced by an agent when using element e , as a function of the current *load* or *congestion* on that element. The cost function F is defined by

$$F(m, w, x) := \sum_{e \in x} c_e(\ell_e(m)), \quad \text{where} \quad \ell_e(m) := m\{(w, x) \in \mathcal{W} \times \mathcal{X} : e \in x\}. \quad (2.14)$$

Here $\ell_e(m)$ is the *load* on the edge e imposed by the type-action distribution m , which is defined as the fraction of agents using the element e . The cost on a route is additive along edges, and the cost at each edge depends on the corresponding load. Notice that the type does not enter explicitly into F , and its only role is to govern the constraints.

A typical class of examples, representing a traffic network congestion game, originating with the seminal work of Wardrop [37], is as follows. The set E is the set of edges of some (directed) graph (V, E) , so that an action $x \in \mathcal{X}$ is a set of edges. The type space \mathcal{W} is a subset of V^2 , so that the *type* $w = (i, j)$ of an agent represents the *source* i and the *destination* j of this agent. The constraint set $\mathcal{C}(w)$ is the set of all (Hamiltonian) paths connecting the source i to the destination j , for $w = (i, j)$.

2.5.1. Existence and uniqueness of equilibrium. Congestion games are well known to belong to the class of *potential games* [32], for which (pure-strategy) Nash equilibria always exist, and for which the uniqueness assumption of Theorem 2.4 can be established simply by proving a certain function is strictly convex. Consider the function $U : \mathcal{P}(\mathcal{W} \times \mathcal{X}) \rightarrow \mathbb{R}$ given by

$$U(m) = \sum_{e \in E} \int_0^{\ell_e(m)} c_e(s) ds. \quad (2.15)$$

Because $c_e \geq 0$ is increasing, the function $t \mapsto \int_0^t c_e(s) ds$ is convex, and thus U is itself convex. Moreover, recalling the definition of $\mathcal{A}(\lambda)$ from (2.13), it can be shown that for each $\lambda \in \mathcal{P}(\mathcal{W})$ the set of minimizers of U on the set $\mathcal{A}(\lambda)$ is precisely $\mathcal{CN}(\lambda)$, the set of Cournot-Nash equilibria with type distribution λ . Hence, when U is strictly convex, the set $\mathcal{CN}(\lambda)$ is a singleton for every λ . The following two propositions justify and elaborate on these claims. At least the first of the two is well known, but we provide the short proofs in Appendix B to keep the paper self-contained. In the following, $|E|$ denotes the cardinality of a set E and for a statement H , 1_H is 1 if the statement H holds and is zero otherwise.

Proposition 2.11. *Fix $\lambda \in \mathcal{P}(\mathcal{W})$. Then m minimizes $U(\cdot)$ on $\mathcal{A}(\lambda)$ if and only if $m \in \mathcal{CN}(\lambda)$.*

The final proposition, regarding uniqueness of the Cournot-Nash equilibrium, is likely sub-optimal but is merely meant to illustrate that uniqueness is not an unreasonable request of a congestion game:

Proposition 2.12. *Enumerate $\mathcal{W} = \{w_1, \dots, w_{|\mathcal{W}|}\}$ and $\mathcal{X} = \{x_1, \dots, x_{|\mathcal{X}|}\}$. Let \mathbb{T} denote the space of $|\mathcal{W}| \times |\mathcal{X}|$ stochastic matrices, i.e., matrices with nonnegative entries whose columns sum to one. Assume c_e is differentiable with a strictly positive derivative. Suppose $\lambda \in \mathcal{P}(\mathcal{W})$ is such that the span of $\{(\lambda\{w_i\}1_{\{e \in x_j\}})_{i,j} : e \in E\}$ contains \mathbb{T} . Then U has a unique minimizer on $\mathcal{A}(\lambda)$.*

2.5.2. Price of anarchy. There is a rich literature on *worst-case* bounds, which are typically valid for a large class of cost functions and model specifications. For instance, for the class of linear cost functions, the seminal paper of Roughgarden and Tardos [35, Theorem 4.5] provides

a worst-case bound of $4/3$ for the PoA in nonatomic games. More precisely, if c_e is linear for each e , then $\text{PoA}(\lambda) \leq 4/3$ for all $\lambda \in \mathcal{P}(\mathcal{W})$. On the other hand, for finite games with linear cost functions, Christodoulou and Koutsoupias showed in [15, Theorem 1] that the worst-case bound on the PoA is $5/2$. That is, if c_e is linear for each e , then $\text{PoA}_n(\vec{w}) \leq 5/2$ for all n and all $\vec{w} \in \mathcal{W}^n$. These PoA bounds are sharp in the sense that there exist linear cost functions and type distributions for which the bound holds with equality. Nonetheless, the following result asserts that for a *fixed* choice of linear cost functions $\{c_e\}_{e \in E}$, the probability of the PoA in the n -player exceeding $4/3$ decays super-exponentially in n .

Corollary 2.13. *In the congestion game model described above, let $R = \sup_{\lambda \ll \lambda_0} \text{PoA}(\lambda)$, and assume that for each $w \in \mathcal{W}$ and every $x \in \mathcal{C}(w)$ there exists $e \in x$ such that $c_e(t) > 0$ for all $t > 0$. Suppose $\{W_i\}$ is an i.i.d. sequence of types with distribution λ_0 . Then, for every $\epsilon > 0$ and $c > 0$, there exists N such that, for all $n \geq N$,*

$$\mathbb{P}(\text{PoA}_n(W_1, \dots, W_n) \geq R + \epsilon) \leq e^{-cn}.$$

Remark 2.14. The assumption in Corollary 2.13 is not very restrictive; it means that if an admissible route for a given agent has a nonzero load on every edge, then the route has nonzero travel time. This holds, for instance, if $c_e(t) > 0$ for all $t > 0$ and for all $e \in E$.

Proof of Corollary 2.13. In this model, since \mathcal{W} and \mathcal{X} are finite, using (2.12) and (2.14), for any $m \in \mathcal{P}(\mathcal{W} \times \mathcal{X})$ we can write

$$V(m) = \sum_{w \in \mathcal{W}} \sum_{x \in \mathcal{C}(w)} m\{(w, x)\} \sum_{e \in x} c_e(\ell_e(m)).$$

Choose $w \in \mathcal{W}$ and $x \in \mathcal{C}(w)$ such that $m\{(w, x)\} > 0$. By assumption, we may find $e \in x$ such that $c_e(t) > 0$ for all $t > 0$. Then

$$\ell_e(m) = \sum_{w' \in \mathcal{W}} \sum_{x' \in \mathcal{C}(w')} 1_{e \in x'} m\{(w', x')\} \geq 1_{e \in x} m\{(w, x)\} > 0,$$

which implies

$$V(m) \geq m\{(w, x)\} c_e(\ell_e(m)) > 0.$$

We are now in a position to apply Proposition 2.10. Because $\inf \emptyset = \infty$ by convention,

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\text{PoA}_n(W_1, \dots, W_n) \geq R + \epsilon) \\ \leq -\inf \{H(\lambda | \lambda_0) : \lambda \ll \lambda_0, \text{PoA}(\lambda) \geq R + \epsilon\} \\ = -\infty. \end{aligned}$$

□

As discussed above, Roughgarden and Tardos showed that the constant R of Corollary 2.13 is at most $4/3$ when c_e is linear for each e . Even though the finite n -player game worst-case PoA_n bound of $5/2$ is optimal *among the class of linear cost functions*, our results show that for large n , it is highly unlikely for any *fixed collection of linear cost functions* $\{c_e\}_{e \in E}$ to produce a PoA over $4/3$ when sampling i.i.d. random types. More generally, Corollary 2.13 produces a high-probability PoA bound for a large but finite population game from a PoA bound for the corresponding class of nonatomic congestion games.

2.5.3. When does high congestion occur? Using the conditional limit result, Theorem 2.5, we illustrate how to describe the typical equilibrium outcome given that a certain route is highly congested. Fix a threshold $r > 0$, and let $P \subset E$ denote some set of elements (i.e., a path through the graph). We might say that the route P is *highly congested* if, in equilibrium, the load on P exceeds r , meaning

$$\ell_P(\mu_n) := \sum_{e \in P} c_e(\ell_e(\mu_n)) \geq r.$$

This event is *rare* if $\ell_P(m) < r$ for $m \in \mathcal{CN}(\lambda_0)$; indeed, under this assumption, it can be shown using Corollary 2.9 that

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\ell_P(\mu_n) \geq r) \leq -\inf \{H(m^w | \lambda_0) : m \in \mathcal{CN}, \ell_P(m) \geq r\}, \quad (2.16)$$

and we deduce from Lemma 3.11 that the right-hand side is strictly negative. Hence, $\mathbb{P}(\ell_P(\mu_n) \geq r)$ decays to zero exponentially. Moreover, the minimizer(s) of the infimum in the right-hand side of (2.16) describe the behavior of μ_n conditioned on the rare event $\{\ell_P(\mu_n) \geq r\}$. That is, if $S \subset \mathcal{CN}$ denotes the set of minimizers of (2.16), then we should expect μ_n to be near S if $\ell_P(\mu_n) \geq r$. That is, the set S describes the most likely equilibrium network flows, given that the congestion on the route P exceeds the threshold r . A caveat is due, as the minimization problem (2.16) is non-convex (because \mathcal{CN} is not convex) and difficult to approach analytically, though it should be computationally feasible if the sets \mathcal{W} and \mathcal{X} are not too large.

3. EXTENSIONS AND PROOFS OF MAIN RESULTS

We begin our analysis in Section 3.1 by embedding the n -player games and the associated nonatomic game on a common space, inspired by a construction of Housman [23]. Then, in Sections 3.3 and 3.4 we prove, respectively, the law of large numbers and large deviation results. Finally, we prove the conditional limit theorem in Section 3.5 and our results on the price of anarchy in Section 3.6.

3.1. A common embedding of n -player and nonatomic games. Let $\text{Gr}(\mathcal{C})$ denote the graph of the constraint set-valued map \mathcal{C} :

$$\text{Gr}(\mathcal{C}) = \{(w, x) \in \mathcal{W} \times \mathcal{X} : x \in \mathcal{C}(w)\}.$$

We wish to define an *equilibrium map* $\mathcal{N} = \mathcal{N}(\lambda, \epsilon, u)$, which maps certain elements of $\mathcal{P}(\mathcal{W}) \times [0, \infty) \times [0, 1]$ to subsets of $\mathcal{P}(\mathcal{W} \times \mathcal{X})$. The first input parameter, $\lambda \in \mathcal{P}(\mathcal{W})$, denotes the distribution of types, while the parameter $\epsilon \in [0, \infty)$ signifies that we are interested in ϵ -Nash equilibria (defined precisely in Remark 3.1(4) below). Finally, the parameter $u \in [0, 1]$ is interpreted as the *size* (or degree of influence) of an agent. We are only interested in sizes belonging to $\overline{\mathbb{N}}^{-1} := \{1/n : n = 1, 2, \dots\} \cup \{0\}$. When the size is $1/n$ we are only interested in discrete probability distributions of the form $m = \frac{1}{n} \sum_{i=1}^n \delta_{(w_i, x_i)}$, where $(w_i, x_i) \in \text{Gr}(\mathcal{C})$, $i = 1, \dots, n$. Thus, the domain of the map \mathcal{N} is a certain subset of $\mathcal{P}(\mathcal{W}) \times [0, \infty) \times [0, 1]$, whose definition requires the following notation. For any set \mathcal{S} , and positive integer n , let $\mathcal{E}_{1/n}(\mathcal{S}) := \{\frac{1}{n} \sum_{i=1}^n \delta_{e_i} : e_i \in \mathcal{S}\}$ denote the set of empirical distributions of n points in \mathcal{S} . When \mathcal{S} is a metric space, the convention $\mathcal{E}_0(\mathcal{S}) := \mathcal{P}(\mathcal{S})$ will be useful as well, where as usual, $\mathcal{P}(\mathcal{S})$ is the set of Borel probability measures on \mathcal{S} . Define $\mathcal{D}_{\mathcal{N}}$ to be the set of $(\lambda, \epsilon, u) \in \mathcal{P}(\mathcal{W}) \times [0, \infty) \times [0, 1]$ such that $u \in \overline{\mathbb{N}}^{-1}$ and $\lambda \in \mathcal{E}_u(\mathcal{W})$. That is,

$$\mathcal{D}_{\mathcal{N}} := \bigcup_{u \in \overline{\mathbb{N}}^{-1}} (\mathcal{E}_u(\mathcal{W}) \times [0, \infty) \times \{u\}) \quad (3.1)$$

Next, define a real-valued function G by

$$G(m, u, w, x) := F(m, w, x) - \inf_{y \in \mathcal{C}(w)} F(m + u(\delta_{(w,y)} - \delta_{(w,x)}), w, y), \quad (3.2)$$

for $((m, u), w, x)$ in $\mathcal{D} \times \mathcal{W} \times \mathcal{X}$, where

$$\mathcal{D} := \bigcup_{u \in \mathbb{N}^{-1}} (\mathcal{E}_u(\mathcal{W} \times \mathcal{X}) \times \{u\}). \quad (3.3)$$

Finally, define the equilibrium map \mathcal{N} on $\mathcal{D}_{\mathcal{N}}$ by

$$\mathcal{N}(\lambda, \epsilon, u) = \{m \in \mathcal{E}_u(\mathcal{W} \times \mathcal{X}) : m(\text{Gr}(\mathcal{C})) = 1, m^w = \lambda, G(m, u, w, x) \leq \epsilon \text{ for } m\text{-a.e. } (w, x)\}. \quad (3.4)$$

Remark 3.1. Several comments are in order here.

- (1) $\mathcal{N}(\lambda, 0, 0)$ is precisely the set $\mathcal{CN}(\lambda)$ of Cournot-Nash equilibria; here the “error” parameter ϵ and the “size” parameter u are both zero, which means that $\mathcal{E}_0(\mathcal{W} \times \mathcal{X}) = \mathcal{P}(\mathcal{W} \times \mathcal{X})$ contains all probability distributions on $\mathcal{W} \times \mathcal{X}$.
- (2) When $u = 1/n > 0$ for some positive integer n , there are n agents, each of “size” $1/n$, and $\mathcal{N}(\lambda, \epsilon, u)$ is a subset of $\mathcal{E}_u(\mathcal{W} \times \mathcal{X})$, the empirical distributions of n points in $\mathcal{W} \times \mathcal{X}$.
- (3) The term $u(\delta_{(w,y)} - \delta_{(w,x)})$ appearing in F in the definition (3.2) of G accounts for the effect on the distribution m of agents when an agent of size u changes its strategy.
- (4) If $(w_1, \dots, w_n) \in \mathcal{W}^n$ is a type vector for the n -player game, it is straightforward to see that $\mathcal{N}(\frac{1}{n} \sum_{i=1}^n \delta_{w_i}, \epsilon, 1/n)$ is precisely the set of empirical distributions $\frac{1}{n} \sum_{i=1}^n \delta_{(w_i, x_i)}$, where (x_1, \dots, x_n) is an ϵ -Nash equilibrium with type vector (w_1, \dots, w_n) , in the sense that $G(\frac{1}{n} \sum_{i=1}^n \delta_{(w_i, x_i)}, 1/n, w_i, x_i) \leq \epsilon$ for every i . Most importantly, recalling the definition of $\hat{\mathcal{N}}_n(w_1, \dots, w_n)$ from (2.8), we have

$$\mathcal{N}\left(\frac{1}{n} \sum_{i=1}^n \delta_{w_i}, 0, \frac{1}{n}\right) = \hat{\mathcal{N}}_n(w_1, \dots, w_n).$$

The key result of this section, inspired by [23], is that the map \mathcal{N} is *upper hemicontinuous*, a crucial property that is used in the proofs of most of the main results. The basic definitions and properties of continuity for set-valued functions are recalled in Appendix A.

Proposition 3.2. *The sets $\mathcal{D}_{\mathcal{N}}$ and \mathcal{D} in (3.1) and (3.3) are closed. The set-valued map \mathcal{N} in (3.4) is upper hemicontinuous on $\mathcal{D}_{\mathcal{N}}$ with compact values, and the function G in (3.2) is continuous. In particular, $\mathcal{N}(\lambda, 0, 0) = \mathcal{CN}(\lambda)$ is closed for all $\lambda \in \mathcal{P}(\mathcal{W})$.*

Proof. Let $(\lambda_n, \epsilon_n, u_n) \in \mathcal{D}_{\mathcal{N}}$ and $(\lambda_\infty, \epsilon_\infty, u_\infty) \in \mathcal{P}(\mathcal{W}) \times [0, \infty) \times [0, 1]$ with $(\lambda_n, \epsilon_n, u_n) \rightarrow (\lambda_\infty, \epsilon_\infty, u_\infty)$. If $u_\infty = 0$, then trivially $\lambda_\infty \in \mathcal{E}_0(\mathcal{W}) = \mathcal{P}(\mathcal{W})$, and so $(\lambda_\infty, \epsilon_\infty, u_\infty) \in \mathcal{D}_{\mathcal{N}}$. If $u_\infty \neq 0$, then there exists N such that $u_\infty = u_n = u$ for all $n \geq N$. But then λ_n belongs to the closed set $\mathcal{E}_{u_\infty}(\mathcal{W})$ for all $n \geq N$, and thus so does λ_∞ . Moreover, by definition $\epsilon_\infty \in [0, \infty)$. This shows that $\mathcal{D}_{\mathcal{N}}$ is closed, and the same argument shows that \mathcal{D} is closed.

To show that \mathcal{N} is upper hemicontinuous, we use the sequential characterization described above. Let $(\lambda_n, \epsilon_n, u_n) \in \mathcal{D}_{\mathcal{N}}$ with $(\lambda_n, \epsilon_n, u_n) \rightarrow (\lambda_\infty, \epsilon_\infty, u_\infty)$, and let $m_n \in \mathcal{N}(\lambda_n, \epsilon_n, u_n)$ for every n . First, note that $m_n^w = \lambda_n$ for each n , which implies that $\{m_n^w\} \subset \mathcal{P}(\mathcal{W})$ is tight by Prokhorov’s theorem and our standing assumption (2) that \mathcal{W} is a complete separable metric space. Because \mathcal{X} is compact, $\{m_n^x\} \subset \mathcal{P}(\mathcal{X})$ is also tight. Thus $\{m_n\} \subset \mathcal{P}(\mathcal{W} \times \mathcal{X})$ is tight, and by Prokhorov’s theorem it admits a subsequential limit point m_∞ . It remains to show that $m_\infty \in \mathcal{N}(\lambda_\infty, \epsilon_\infty, u_\infty)$.

We will abuse notation somewhat by assuming $m_n \rightarrow m_\infty$. Because $m_n(\text{Gr}(\mathcal{C})) = 1$ for each n and $\text{Gr}(\mathcal{C})$ is closed, the Portmanteau theorem yields $m_\infty(\text{Gr}(\mathcal{C})) = 1$. It is clear also that

$$\lambda_\infty = \lim_{n \rightarrow \infty} \lambda_n = \lim_{n \rightarrow \infty} m_n^w = m_\infty^w,$$

where the limits are in the sense of weak convergence. The continuity of F and \mathcal{C} of standing assumptions (3-4) implies the continuity of G by Berge's theorem [3, Theorem 17.31]. Define measures η_n on $\mathcal{D} \times \mathcal{W} \times \mathcal{X}$ by

$$\eta_n(dm, du, dw, dx) = \delta_{(m_n, u_n)}(dm, du) m_n(dw, dx).$$

Then $\eta_n \rightarrow \eta_\infty$ because $(m_n, u_n) \rightarrow (m_\infty, u_\infty)$, and it follows from the Portmanteau theorem that for any $\Delta > 0$,

$$\begin{aligned} m_\infty \{(w, x) : G(m_\infty, u_\infty, w, x) \leq \epsilon_\infty + \Delta\} &= \eta_\infty \{(m, u, w, x) : G(m, u, w, x) \leq \epsilon_\infty + \Delta\} \\ &\geq \limsup_{n \rightarrow \infty} \eta_n \{(m, u, w, x) : G(m, u, w, x) \leq \epsilon_\infty + \Delta\} \\ &= \limsup_{n \rightarrow \infty} m_n \{(w, x) : G(m_n, u_n, w, x) \leq \epsilon_\infty + \Delta\} \\ &\geq \limsup_{n \rightarrow \infty} m_n \{(w, x) : G(m_n, u_n, w, x) \leq \epsilon_n\} \\ &= 1, \end{aligned}$$

where the last inequality uses the fact that, since $\epsilon_n \rightarrow \epsilon_\infty$, $\epsilon_n \leq \epsilon_\infty + \Delta$ for all sufficiently large n , and the last equality holds because $m_n \in \mathcal{N}(\lambda_n, \epsilon_n, u_n)$. Since $\Delta > 0$ is arbitrary, it follows that $G(m_\infty, u_\infty, w, x) \leq \epsilon_\infty$ for m_∞ a.e. (w, x) . It remains to check that m_∞ belongs to $\mathcal{E}_{u_\infty}(\mathcal{W} \times \mathcal{X})$. If $u_\infty = 0$, there is nothing to prove because of the convention $\mathcal{E}_0(\mathcal{W} \times \mathcal{X}) = \mathcal{P}(\mathcal{W} \times \mathcal{X})$. If $u_\infty > 0$, then there exists N such that $u_n = u_\infty$ for all $n \geq N$. Then m_n belongs to the closed set $\mathcal{E}_{u_\infty}(\mathcal{W} \times \mathcal{X})$ for all $n \geq N$, and hence, so does m_∞ . \square

3.2. Existence of Cournot-Nash equilibria. Under our standing assumptions, there always exist Cournot-Nash equilibria for the nonatomic game. The proof uses a simple argument due to Mas-Colell [30], appropriately modified to incorporate the constraint map \mathcal{C} .

Proposition 3.3. *For each $\lambda \in \mathcal{P}(\mathcal{W})$, $\mathcal{CN}(\lambda) \neq \emptyset$.*

Proof. Let $\text{Gr}(\mathcal{C}) = \{(w, x) \in \mathcal{W} \times \mathcal{X} : x \in \mathcal{C}(w)\}$ as before, and define $\mathcal{A}(\lambda)$ as in (2.13). Note that $\mathcal{A}(\lambda)$ is closed, as $\text{Gr}(\mathcal{C})$ is closed by our standing assumption (3). Because \mathcal{X} is compact and $\mathcal{W} \times \mathcal{X}$ is a complete separable metric space by our standing assumptions (1-2), it is straightforward to check that $\mathcal{A}(\lambda)$ is tight and thus compact. Consider the map Φ from $\mathcal{A}(\lambda)$ into subsets of $\mathcal{A}(\lambda)$, given by

$$\Phi(m) = \left\{ \tilde{m} \in \mathcal{A}(\lambda) : \int_{\mathcal{W} \times \mathcal{X}} G(m, 0, w, x) \tilde{m}(dw, dx) \leq 0 \right\}$$

for $m \in \mathcal{A}(\lambda)$. Note that $m \in \mathcal{P}(\mathcal{W} \times \mathcal{X})$ is a Cournot-Nash equilibrium with type distribution λ if and only if $m \in \Phi(m)$, i.e., m is a fixed point of Φ . Clearly $\mathcal{A}(\lambda)$ is convex, and hence $\Phi(m)$ is convex for each m . The graph $\text{Gr}(\Phi) = \{(m, \tilde{m}) \in \mathcal{A}(\lambda) \times \mathcal{A}(\lambda) : \tilde{m} \in \Phi(m)\}$ is easily seen to be closed, using the fact that G is continuous (due to Proposition 3.2) and bounded (by standing assumption (4)). To check that $\Phi(m)$ is nonempty for each m , note that there exists (e.g., by [3, Theorem 18.19]) a measurable function $\hat{x} : \mathcal{A}(\lambda) \times \mathcal{W} \rightarrow \mathcal{X}$ such that $\hat{x}(m, w) \in \mathcal{C}(w)$ and $F(m, w, \hat{x}(m, w)) = \inf_{y \in \mathcal{C}(w)} F(m, w, y)$ for each $(m, w) \in \mathcal{A}(\lambda) \times \mathcal{W}$. Then $\hat{m}(dw, dx) = \lambda(dw) \delta_{\hat{x}(m, w)}(dx)$ always belongs to $\Phi(m)$. Because Φ has a closed graph and nonempty convex values, it admits a fixed point by Kakutani's theorem [3, Corollary 17.55]. \square

3.3. Proof of laws of large numbers. Using Proposition 3.2, we give streamlined proofs of Theorems 2.7 and 2.2, and even an extension of the latter.

Proof of Theorem 2.7. Proposition 3.2 shows that \mathcal{N} is upper hemicontinuous. According to Lemma A.1, this implies that \mathcal{N} is continuous as a map from $\mathcal{D}_{\mathcal{N}}$ to the space \mathfrak{C} of closed subsets of $\mathcal{P}(\mathcal{W} \times \mathcal{X})$ endowed with the upper Vietoris topology. Because $(\frac{1}{n}\delta_{W_i}, 0, \frac{1}{n})$ converges almost surely to $(\lambda_0, 0, 0)$, it follows (see Remark 3.1(4) for the first equality) that, almost surely,

$$\widehat{\mathcal{N}}_n(W_1, \dots, W_n) = \mathcal{N}\left(\frac{1}{n}\delta_{W_i}, 0, \frac{1}{n}\right) \rightarrow \mathcal{N}(\lambda_0, 0, 0) = \mathcal{CN}(\lambda_0).$$

□

We next turn to the proof of Theorem 2.2, which we precede with a reassuring technical lemma. First, for $\epsilon \geq 0$ and $n \in \mathbb{N}$, let $N_n^\epsilon(\vec{w})$ denote the set of ϵ -Nash equilibria with type vector $\vec{w} \in \mathcal{W}^n$. By Remark 3.1(4), this can be expressed in terms of the function G of (3.2) as

$$N_n^\epsilon(w_1, \dots, w_n) = \left\{ \vec{x} \in \mathcal{X}^n : x_i \in \mathcal{C}(w_i) \text{ and } G\left(\frac{1}{n} \sum_{i=1}^n \delta_{(w_i, x_i)}, \frac{1}{n}, w_i, x_i\right) \leq \epsilon, \forall i = 1, \dots, n \right\}.$$

Also, let D_n^ϵ be the set of $\vec{w} \in \mathcal{W}^n$ for which there exists an associated ϵ -Nash equilibrium:

$$D_n^\epsilon = \{\vec{w} \in \mathcal{W}^n : N_n^\epsilon(\vec{w}) \neq \emptyset\}.$$

Lemma 3.4. *For each n and $\epsilon \geq 0$, the set D_n^ϵ is closed. Moreover, there exists a universally measurable map $\hat{x} : D_n^\epsilon \rightarrow \mathcal{X}^n$ such that $\hat{x}(\vec{w}) \in N_n^\epsilon(\vec{w})$ for each $\vec{w} \in D_n^\epsilon$.*

Proof. Continuity of G (proven in Proposition 3.2) and closedness of the graph of \mathcal{C} (one of our standing assumptions) together imply that the graph

$$\text{Gr}(N_n^\epsilon) = \{(\vec{w}, \vec{x}) \in \mathcal{W}^n \times \mathcal{X}^n : \vec{x} \in N_n^\epsilon(\vec{w})\}$$

is closed. The projection from $\mathcal{W} \times \mathcal{X}$ to \mathcal{W} is a closed map, since \mathcal{X} is compact, which shows that D_n^ϵ is closed. The existence of \hat{x} follows from the Jankov-von Neumann theorem [6, Proposition 7.49]. □

Theorem 3.5. *Let $\epsilon_n \geq 0$ be such that $\epsilon_n \rightarrow 0$. Suppose, for each n and each $\vec{w} \in \mathcal{W}^n$, we are given an ϵ_n -Nash equilibrium $\hat{x}^n(\vec{w}) = (\hat{x}_1^n(\vec{w}), \dots, \hat{x}_n^n(\vec{w}))$ with type vector \vec{w} . By Lemma 3.4 we may assume each \hat{x}_i^n is universally measurable. Finally, suppose $\vec{W}^n = (W_1^n, \dots, W_n^n)$ is a \mathcal{W}^n -valued random vector, and define the random empirical distributions*

$$\mu_n = \frac{1}{n} \sum_{i=1}^n \delta_{(W_i^n, \hat{x}_i^n(\vec{W}^n))}, \quad \mu_n^w := \frac{1}{n} \sum_{i=1}^n \delta_{W_i^n}.$$

Then the following hold:

- (i) *If the sequence $\{\mu_n^w\}$ is tight, then so is the sequence $\{\mu_n\}$, and every subsequential limit in distribution μ of $\{\mu_n\}$ satisfies $\mu \in \mathcal{CN}(\mu^w)$, almost surely.*
- (ii) *If $\mu_n^w \rightarrow \lambda_0$ in distribution, where $\lambda_0 \in \mathcal{P}(\mathcal{W})$ is deterministic, then every subsequential limit in distribution of $\{\mu_n\}$ is supported on $\mathcal{CN}(\lambda_0)$. In particular,*

$$\lim_{n \rightarrow \infty} \mathbb{P}(d(\mu_n, \mathcal{CN}(\lambda_0)) \geq \epsilon) = 0, \tag{3.5}$$

for every $\epsilon > 0$, and any metric d on $\mathcal{P}(\mathcal{W} \times \mathcal{X})$ compatible with weak convergence.

- (iii) *If W_1^n, \dots, W_n^n are i.i.d. with distribution $\lambda_0 \in \mathcal{P}(\mathcal{W})$, for each n , then $d(\mu_n, \mathcal{CN}(\lambda_0)) \rightarrow 0$ almost surely, for d as in (ii).*

Proof.

- (i) By [36, Proposition 2.2(ii)], tightness of $\{\mathbb{P} \circ (\mu_n^w)^{-1}\} \subset \mathcal{P}(\mathcal{P}(\mathcal{W}))$ is equivalent to tightness of the sequence of mean measures $\{\mathbb{E}[\mu_n^w]\} \subset \mathcal{P}(\mathcal{W})$, where $\mathbb{E}[\mu_n^w](\cdot) := \mathbb{E}[\mu_n^w(\cdot)]$. The mean measure $\mathbb{E}[\mu_n]$ has first marginal $\mathbb{E}[\mu_n^w]$, and because \mathcal{X} is compact we conclude that $\{\mathbb{E}[\mu_n]\} \subset \mathcal{P}(\mathcal{W} \times \mathcal{X})$ is tight. Again using [36, Proposition 2.2(ii)], we conclude that $\{\mathbb{P} \circ \mu_n^{-1}\} \subset \mathcal{P}(\mathcal{P}(\mathcal{W} \times \mathcal{X}))$ is tight. Now, by Skorohod's representation theorem, we may assume that (along a subsequence) μ_n converges almost surely to a random element μ of $\mathcal{P}(\mathcal{W} \times \mathcal{X})$. This implies $\mu_n^w \rightarrow \mu^w$ a.s. Since $\mu_n \in \mathcal{N}(\mu_n^w, \epsilon_n, 1/n)$ by assumption, the upper hemicontinuity of \mathcal{N} implies that a.s., μ must belong to $\mathcal{N}(\mu^w, 0, 0) = \mathcal{CN}(\mu^w)$, where the last equality holds by Remark 3.1(1).
- (ii) Suppose the random measure μ is a subsequential limit in distribution of $\{\mu_n\}$. Because $\mu_n^w \rightarrow \lambda_0$, we must have $\mu^w = \lambda_0$ a.s. We conclude from (i) that $\mathbb{P}(\mu \in \mathcal{CN}(\lambda_0)) = 1$. Thus, for any subsequential limit μ of $\{\mu_n\}$ and $\epsilon > 0$, we have $\mathbb{P}(d(\mu, \mathcal{CN}(\lambda_0)) \geq \epsilon) = 0$. When combined with the Portmanteau theorem, the closedness of the set $\mathcal{CN}(\lambda_0)$ established in Proposition 3.2 and the consequent closedness of $\{m \in \mathcal{P}(\mathcal{W} \times \mathcal{X}) : d(m, \mathcal{CN}(\lambda_0)) \geq \epsilon\}$ imply the claim (3.5).
- (iii) Almost surely, the following holds: Because $\mu_n^w \rightarrow \lambda_0$ due to $\{W_i^n\}$ being i.i.d., and $\mu_n \in \mathcal{N}(\mu_n^w, \epsilon_n, 1/n)$, upper hemicontinuity of \mathcal{N} implies that the limit $\lim_{k \rightarrow \infty} \mu_{n_k}$ exists along some subsequence, and every such limit belongs to $\mathcal{N}(\lambda_0, 0, 0)$. By Remark 3.1(1), this is enough to show $d(\mu_n, \mathcal{CN}(\lambda_0)) \rightarrow 0$.

□

3.4. Proofs of large deviation results. We are now prepared to state and prove an extension of our main result (Theorem 2.8) that allows for approximate equilibria. Recall from Section 2.3 the definition of the space \mathfrak{C} , equipped with the upper Vietoris topology. Having identified the suitable space, topology and mappings, the proof of this extension follows from a simple application of the contraction principle from large deviations theory. As we will use it on several occasions, it is worth recalling here the general definition of an LDP. We say that a sequence of Borel probability measures $\{\nu_n\}$ on a topological space S satisfies an LDP with good rate function $I : S \rightarrow [0, \infty]$ if the level set $\{s \in S : I(s) \leq c\}$ is compact for each $c \geq 0$ and if the following holds for every Borel set $A \subset S$:

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{1}{n} \log \nu_n(A) &\leq - \inf_{s \in \bar{A}} I(s), \\ \liminf_{n \rightarrow \infty} \frac{1}{n} \log \nu_n(A) &\geq - \inf_{s \in A^\circ} I(s), \end{aligned}$$

where \bar{A} and A° denote the closure and interior. We say a sequence of S -valued random variables satisfies an LDP if the corresponding sequence of probability measures does. In the following, recall the definition of the relative entropy H from (2.1).

Theorem 3.6. *Suppose $\epsilon_n \rightarrow 0$ and $\{W_i\}$ is an i.i.d. sequence of \mathcal{W} -valued random variables with common type distribution λ_0 . Then the sequence of sets of ϵ_n -Nash equilibria*

$$\mathcal{N}\left(\frac{1}{n} \sum_{i=1}^n \delta_{W_i}, \epsilon_n, \frac{1}{n}\right), \quad n \in \mathbb{N},$$

satisfies an LDP on \mathfrak{C} with good rate function

$$J(A) = \inf \{H(\lambda|\lambda_0) : \lambda \in \mathcal{P}(\mathcal{W}), \mathcal{CN}(\lambda) = A\}. \quad (3.6)$$

Proof. First recall from Proposition 3.2 that $\mathcal{N}(\lambda, \epsilon, u)$ is closed and thus belongs to \mathfrak{C} , for every $(\lambda, \epsilon, u) \in \mathcal{D}_{\mathcal{N}}$. Define two $\mathcal{D}_{\mathcal{N}}$ -valued random variables

$$M_n = \left(\frac{1}{n} \sum_{i=1}^n \delta_{W_i}, \epsilon_n, \frac{1}{n} \right), \quad M_n^0 = \left(\frac{1}{n} \sum_{i=1}^n \delta_{W_i}, 0, 0 \right).$$

By Sanov's theorem and the contraction principle [17, Theorem 4.2.1], applied to the continuous map $\mathcal{P}(\mathcal{W}) \ni \lambda \mapsto (\lambda, 0, 0) \in \mathcal{P}(\mathcal{W}) \times [0, \infty) \times [0, 1]$, $\{M_n^0\}$ satisfies an LDP on $\mathcal{P}(\mathcal{W}) \times [0, \infty) \times [0, 1]$ with good rate function

$$(\lambda, \epsilon, u) \mapsto \begin{cases} H(\lambda | \lambda_0) & \text{if } \epsilon = u = 0, \\ \infty & \text{otherwise.} \end{cases}$$

The sequences $\{M_n\}$ and $\{M_n^0\}$ are exponentially equivalent in the sense that (c.f. [17, 4.2.10])

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\bar{d}(M_n, M_n^0) \geq a) = -\infty, \text{ for all } a > 0, \quad (3.7)$$

where we define the metric \bar{d} on $\mathcal{P}(\mathcal{W}) \times [0, \infty) \times [0, 1]$ by

$$\bar{d}((\lambda', \epsilon', u'), (\lambda, \epsilon, u)) = d(\lambda, \lambda') + |\epsilon - \epsilon'| + |u - u'|,$$

where d is any metric on $\mathcal{P}(\mathcal{W})$ compatible with weak convergence. In fact, the probability in (3.7) is zero for sufficiently large n . Thus, $\{M_n\}$ satisfies an LDP with the same rate function [17, Theorem 4.2.13]. Because \mathcal{N} is upper hemicontinuous as a set-valued map (by Proposition 3.2), it is continuous as a map from $\mathcal{D}_{\mathcal{N}}$ to \mathfrak{C} , equipped with the upper Vietoris topology (see Lemma A.1). Thus, the contraction principle (see [17, Theorem 4.2.1], which does not actually need the spaces to be Hausdorff) implies that $\{\mathcal{N}(M_n) = \mathcal{N}(\frac{1}{n} \sum_{i=1}^n \delta_{W_i}, \epsilon_n, \frac{1}{n})\}$ satisfies an LDP on \mathfrak{C} with good rate function

$$\begin{aligned} A &\mapsto \inf \{H(\lambda | \lambda_0) : \lambda \in \mathcal{P}(\mathcal{W}), \mathcal{N}(\lambda, 0, 0) = A\} \\ &= \inf \{H(\lambda | \lambda_0) : \lambda \in \mathcal{P}(\mathcal{W}), \mathcal{CN}(\lambda) = A\} \\ &= J(A), \end{aligned}$$

where the first equality uses the fact that $\mathcal{N}(\lambda, 0, 0) = \mathcal{CN}(\lambda)$ from Remark 3.1(4). \square

Remark 3.7. Recalling from Remark 3.1(4) that $\widehat{\mathcal{N}}_n(w_1, \dots, w_n) = \mathcal{N}(\frac{1}{n} \sum_{i=1}^n \delta_{w_i}, 0, 1/n)$, Theorem 2.8 is an immediate corollary of Theorem 3.6.

Remark 3.8. Because the proof of Theorem 3.6 relies on the contraction principle, a similar result holds if we weaken the assumptions on the type sequence $\{W_i\}$. They need not be i.i.d., as long as the sequence of empirical distributions $\frac{1}{n} \sum_{i=1}^n \delta_{W_i}$ satisfies some LDP.

We next state an extension of Theorem 2.4 and prove it using Theorem 3.6 and some elementary properties of the upper Vietoris topology, reviewed in Appendix A. Interestingly, even without uniqueness we find upper and lower bounds, although they do not match in general.

Theorem 3.9. *Use the notation and assumptions of Theorem 3.5, and assume also that $W_i^n = W_i$ for $1 \leq i \leq n$, where $\{W_i\}$ is an i.i.d. sequence with distribution $\lambda_0 \in \mathcal{P}(\mathcal{W})$. Then we have the following bounds, valid for every measurable set $A \subset \mathcal{P}(\mathcal{W} \times \mathcal{X})$:*

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\mu_n \in A) \leq -\inf \{H(m^w | \lambda_0) : m \in \overline{A} \cap \mathcal{CN}\}, \quad (3.8)$$

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\mu_n \in A) \geq -\inf \{H(\lambda | \lambda_0) : \lambda \in \mathcal{P}(\mathcal{W}), \mathcal{CN}(\lambda) \subset A^\circ\}. \quad (3.9)$$

Moreover, if $\mathcal{CN}(\lambda)$ is a singleton for every $\lambda \ll \lambda_0$, then

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\mu_n \in A) \geq -\inf \{H(m^w | \lambda_0) : m \in A^\circ \cap \mathcal{CN}\}. \quad (3.10)$$

Proof. Suppose A is closed. Then $\mathfrak{U} := \{E \in \mathfrak{C} : E \subset A^c\} = \{E \in \mathfrak{C} : E \cap A = \emptyset\}$ is open in the upper Vietoris topology, so its complement is closed. Thus, using the upper bound of Theorem 3.6,

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P} \left(\mathcal{N} \left(\frac{1}{n} \sum_{i=1}^n \delta_{W_i}, \epsilon_n, \frac{1}{n} \right) \cap A \neq \emptyset \right) \\ \leq - \inf_{B \in \mathfrak{U}^c} J(B) \\ = - \inf_{B \in \mathfrak{U}^c} \inf \{H(\lambda | \lambda_0) : \lambda \in \mathcal{P}(\mathcal{W}), \mathcal{CN}(\lambda) = B\} \\ = - \inf \{H(\lambda | \lambda_0) : \lambda \in \mathcal{P}(\mathcal{W}), \mathcal{CN}(\lambda) \cap A \neq \emptyset\} \\ = - \inf \{H(m^w | \lambda_0) : m \in \mathcal{CN} \cap A\}. \end{aligned}$$

Indeed, this last equality follows from two simple observations: If $\mathcal{CN}(\lambda) \cap A \neq \emptyset$, then there exists $m \in \mathcal{CN}(\lambda) \cap A \subset \mathcal{CN} \cap A$ such that $m^w = \lambda$. On the other hand, if $m \in \mathcal{CN} \cap A$, then $m \in \mathcal{CN}(m^w)$, so $\mathcal{CN}(m^w) \cap A \neq \emptyset$. Finally, the upper bound (3.8) follows from the inequality

$$\mathbb{P}(\mu_n \in A) \leq \mathbb{P} \left(\mathcal{N} \left(\frac{1}{n} \sum_{i=1}^n \delta_{W_i}, \epsilon_n, \frac{1}{n} \right) \cap A \neq \emptyset \right),$$

which holds because, by Remark 3.1(4), $\mu_n \in \mathcal{N} \left(\frac{1}{n} \sum_{i=1}^n \delta_{W_i}, \epsilon_n, \frac{1}{n} \right)$ a.s.

To prove the lower bound, let A be open, and notice that $\mathfrak{U} = \{E \in \mathfrak{C} : E \subset A\}$ is open in the upper Vietoris topology. Theorem 3.6 then implies

$$\begin{aligned} \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P} \left(\mathcal{N} \left(\frac{1}{n} \sum_{i=1}^n \delta_{W_i}, \epsilon_n, \frac{1}{n} \right) \subset A \right) \\ \geq - \inf_{B \in \mathfrak{U}} J(B) \\ = - \inf_{B \in \mathfrak{U}} \inf \{H(\lambda | \lambda_0) : \lambda \in \mathcal{P}(\mathcal{W}), \mathcal{CN}(\lambda) = B\} \\ = - \inf \{H(\lambda | \lambda_0) : \lambda \in \mathcal{P}(\mathcal{W}), \mathcal{CN}(\lambda) \subset A\}, \end{aligned}$$

where the last equality uses the property that $\mathcal{CN}(\lambda)$ is closed for every $\lambda \in \mathcal{P}(\mathcal{W})$ (see Proposition 3.2). Then the lower bound (3.9) follows from the inequality

$$\mathbb{P}(\mu_n \in A) \geq \mathbb{P} \left(\mathcal{N} \left(\frac{1}{n} \sum_{i=1}^n \delta_{W_i}, \epsilon_n, \frac{1}{n} \right) \subset A \right),$$

which again holds because $\mu_n \in \mathcal{N} \left(\frac{1}{n} \sum_{i=1}^n \delta_{W_i}, \epsilon_n, \frac{1}{n} \right)$ a.s.; see Remark 3.1(4).

Finally, we deduce (3.10) from (3.9). Again let A be open and note first that $H(\lambda | \lambda_0) < \infty$ only if $\lambda \ll \lambda_0$. Supposing $\mathcal{CN}(\lambda) = \{M[\lambda]\}$ is a singleton for all $\lambda \ll \lambda_0$, then trivially $m = M[m^w]$ for all $m \in \mathcal{CN}$, and

$$\begin{aligned} \inf \{H(\lambda | \lambda_0) : \lambda \in \mathcal{P}(\mathcal{W}), \mathcal{CN}(\lambda) \subset A\} &= \inf \{H(\lambda | \lambda_0) : \lambda \ll \lambda_0, M[\lambda] \in A\} \\ &= \inf \{H(m^w | \lambda_0) : m \in \mathcal{CN} \cap A, m^w \ll \lambda_0\} \\ &= \inf \{H(m^w | \lambda_0) : m \in \mathcal{CN} \cap A\}. \end{aligned}$$

□

In applications, it is important to know if the bounds in the large deviation principles of Theorems 3.9 and 3.6 are nonzero. The following lemmas help to check this, and their straightforward proofs are deferred to Appendix C.

Lemma 3.10. *Let J be as in (3.6), and let $\mathfrak{U} \subset \mathfrak{C}$ be a closed set with $\mathcal{CN}(\lambda_0) \notin \mathfrak{U}$. Then $\inf_{A \in \mathfrak{U}} J(A) > 0$.*

Lemma 3.11. *If $A \subset \mathcal{P}(\mathcal{W} \times \mathcal{X})$ is closed and $A \cap \mathcal{CN}(\lambda_0) = \emptyset$, then $\inf_{m \in A \cap \mathcal{CN}} H(m^w | \lambda_0) > 0$.*

3.5. The conditional limit theorem and entry games. In this section we first prove Theorem 2.5. Then, to illustrate the tractability of the assumptions, we apply the theorem to an example from the class of entry games discussed in Section 1.3.

Proof of Theorem 2.5. Fix a measurable set $A \subset \mathcal{P}(\mathcal{W} \times \mathcal{X})$. Since $I(A) < \infty$, the closedness of \mathcal{CN} and the compactness of the sub-level sets $\{m \in \mathcal{P}(\mathcal{W} \times \mathcal{X}) : H(m^w | \lambda_0) \leq c\}$ established in Lemma C.1 imply that the set $S(A)$ of minimizers in (2.4) is non-empty and compact. Next, use the lower bound of Theorem 3.9 to get

$$\begin{aligned} \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\mu_n \in A) &\geq -\inf \{H(\lambda | \lambda_0) : \lambda \in \mathcal{P}(\mathcal{W}), \mathcal{CN}(\lambda) \subset A^\circ\} \\ &= -I(A) \\ &= -\inf \{H(\lambda | \lambda_0) : \lambda \in \mathcal{P}(\mathcal{W}), \mathcal{CN}(\lambda) \cap \overline{A} \neq \emptyset\} \\ &= -\inf \{H(m^w | \lambda_0) : m \in \mathcal{CN} \cap \overline{A}\}, \end{aligned}$$

where we have used the assumption (2.5) in the third line. Because the set

$$A_\epsilon := \{m \in A : d(m, S(A)) \geq \epsilon\}$$

is closed, the upper bound of Theorem 3.9 yields

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(d(\mu_n, S(A)) \geq \epsilon \mid \mu_n \in A) \\ &= \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\mu_n \in A_\epsilon) - \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\mu_n \in A) \\ &\leq \inf_{m \in \overline{A} \cap \mathcal{CN}} H(m^w | \lambda_0) - \inf_{m \in A_\epsilon \cap \mathcal{CN}} H(m^w | \lambda_0) \\ &=: C, \end{aligned}$$

Then clearly $C \leq 0$, as $\overline{A} \supset A_\epsilon$. If $C = 0$, then there exists $m \in A_\epsilon \cap \mathcal{CN}$ such that $H(m^w | \lambda_0) = \inf_{m \in \overline{A} \cap \mathcal{CN}} H(m^w | \lambda_0)$. But this implies $m \in S(A)$, which contradicts the fact that $S(A)$ and A_ϵ are disjoint. Thus $C < 0$, and the proof of (2.6) is complete.

Finally, if $\mathcal{CN}(\lambda) = \{M(\lambda)\}$ is a singleton for every $\lambda \ll \lambda_0$, then the identity $(M(\lambda))^w = \lambda$ implies

$$\begin{aligned} \inf \{H(\lambda | \lambda_0) : \lambda \ll \lambda_0, \mathcal{CN}(\lambda) \subset A^\circ\} &= \inf \{H(\lambda | \lambda_0) : \lambda \ll \lambda_0, M(\lambda) \in A^\circ\} \\ &= \inf \{H(m^w | \lambda_0) : m \in A^\circ \cap \mathcal{CN}, m^w \ll \lambda_0\} \\ &= \inf \{H(m^w | \lambda_0) : m \in A^\circ \cap \mathcal{CN}\}. \end{aligned}$$

Similarly, we noted above already that

$$I(A) = \inf \{H(\lambda | \lambda_0) : \lambda \ll \lambda_0, \mathcal{CN}(\lambda) \cap \overline{A} \neq \emptyset\} = \inf \{H(m^w | \lambda_0) : m \in \mathcal{CN} \cap \overline{A}\}.$$

□

Let us now consider a simple entry game, specified as follows. There are two types and two actions, with $\mathcal{W} = \{1, 2\}$ and $\mathcal{X} = \{0, 1\}$, there are no constraints in this model, so $\mathcal{C}(w) = \mathcal{X}$ for all $w \in \mathcal{W}$, and the objective function is given by

$$F(m, w, x) = -x(-3m^x\{1\} + w),$$

for $m \in \mathcal{P}(\mathcal{W} \times \mathcal{X})$, $w \in \mathcal{W}$, and $x \in \mathcal{X}$. Think of each agent as facing a fixed payoff w from entering the market, i.e., choosing $x = 1$. This payoff is offset by a loss of $3m^x\{1\}$ which increases with the fraction of agents entering the market. If the net payoff is negative, the agent will choose $x = 0$ and not enter the market.

For $q \in [0, 1]$, let $\lambda_q = q\delta_2 + (1 - q)\delta_1$ denote the type distribution in which the fraction of type-2 agents is q . Of course, $\{\lambda_q : q \in [0, 1]\}$ exhausts all of $\mathcal{P}(\mathcal{W})$. To apply our conditional limit theorem, we first characterize all possible Cournot-Nash equilibria:

Proposition 3.12. *For the entry game described above, for each $q \in [0, 1]$ the Cournot-Nash equilibrium is unique. That is, $\mathcal{CN}(\lambda_q) = \{m_q\}$, where $m_q \in \mathcal{P}(\mathcal{W} \times \mathcal{X})$ is defined by*

$$\begin{pmatrix} m_q\{(1, 0)\} & m_q\{(1, 1)\} \\ m_q\{(2, 0)\} & m_q\{(2, 1)\} \end{pmatrix} = \begin{cases} \begin{pmatrix} 2/3 & 1/3 - q \\ 0 & q \end{pmatrix} & \text{if } q \leq 1/3, \\ \begin{pmatrix} 1 - q & 0 \\ 0 & q \end{pmatrix} & \text{if } 1/3 < q < 2/3, \\ \begin{pmatrix} 1 - q & 0 \\ q - 2/3 & 2/3 \end{pmatrix} & \text{if } q \geq 2/3. \end{cases}$$

Proof. We know $\mathcal{CN}(\lambda_q) \neq \emptyset$ for each $q \in [0, 1]$, thanks to Proposition 3.3. Fix $q \in [0, 1]$ and $m \in \mathcal{CN}(\lambda_q)$, and abbreviate $p = m^x\{1\}$. We will show that $m = m_q$. Note first that

$$\arg \min_{x \in \{0, 1\}} F(m, w, x) = \begin{cases} \{1\} & \text{if } w > 3p, \\ \{0\} & \text{if } w < 3p, \\ \{0, 1\} & \text{if } w = 3p. \end{cases}$$

Next, there are three cases to check. First, if $3p \notin \{1, 2\}$, then $m\{(w, 1)\} = 1_{\{w > 3p\}}$ for each w . Thus,

$$p = m^x\{1\} = q1_{\{2 > 3p\}} + (1 - q)1_{\{1 > 3p\}} = \begin{cases} 0 & \text{if } p > 2/3 \\ q & \text{if } 2/3 > p > 1/3 \\ 1 & \text{if } p < 1/3. \end{cases}$$

This can only hold if $p = q$ and $1/3 < p < 2/3$. For the second case, suppose $p = 1/3$. Then all type-2 agents enter since $2 > 3p$; that is, $m\{(2, 1)\} = q$ and $m\{(2, 0)\} = 0$. Therefore, we have

$$1/3 = p = m\{(1, 1)\} + m\{(2, 1)\} = m\{(1, 1)\} + q,$$

which implies $m\{(1, 1)\} = 1/3 - q$, which only makes sense for $q \leq 1/3$. For the final case, suppose $p = 2/3$. Then type-1 agents do not enter since $1 < 3p$; that is, $m\{(1, 0)\} = 1 - q$ and $m\{(1, 1)\} = 0$. This implies

$$2/3 = p = m\{(1, 1)\} + m\{(2, 1)\} = m\{(2, 1)\}.$$

Since $q = m\{(2, 0)\} + m\{(2, 1)\} = m\{(2, 0)\} + 2/3$, we must have $q \geq 2/3$. \square

Similarly, in the n -player game, we can argue that there exists a Nash equilibrium with type vector \vec{w} , for every fixed $\vec{w} = (w_1, \dots, w_n) \in \mathcal{W}^n$. To construct an example, there are three cases, depending again on the fraction q of (w_1, \dots, w_n) which equal 2. In each case, we

construct one example (though there may be more) of an equilibrium, recalling that agent i enters the market if $x_i = 1$:

- (1) Suppose $q \in [1/3, 2/3]$. All type-2 agents enter, while none of the type-1 agents enter.
- (2) Suppose $q < 1/3$. All type-2 agents enter. Let k be the greatest integer less than or equal to $n(1/3 - q)$. Then k of the type-1 agents enter, and the rest do not.
- (3) Suppose $q > 2/3$. All type-1 agents choose not to enter. Let k be the greatest integer less than or equal to $2/3$. Then k of the type-2 agents enter, and the rest do not.

Note that we have constructed multiple equilibria in the latter cases, although they share a common type-action distribution.

Now that we have computed the Cournot-Nash equilibria and are confident that n -player equilibria exist, we are ready to apply Theorem 2.5. Now, let μ_n denote the empirical type-action distribution as usual, where the types are i.i.d. samples from the distribution $\lambda_{2/3}$. That is, each of the n agents is independently assigned type 2 with probability $2/3$ and type 1 with probability $1/3$. By Theorem 3.5 and Proposition 3.12, we know that μ_n converges a.s. to the unique element $m_{2/3}$ of $\mathcal{CN}(\lambda_{2/3})$. Let us show using Theorem 2.5 (and Remark 2.6 thereafter) that, for $r \in (1/3, 2/3)$, if we condition on the rare event $\{\mu_n^x(1) \leq r\}$, then $\mu_n \rightarrow m_r$. More precisely, the conditional law of μ_n converges to the point mass at m_r . Intuitively, this rare event means that no more than a fraction of r of the agents enters the market, and the most likely way for this to happen (asymptotically) is for precisely a fraction of r of the agents to enter.

Let $1/3 < r < 2/3$, and consider the set

$$A = \{m \in \mathcal{P}(\mathcal{W} \times \mathcal{X}) : m^x\{1\} \leq r\}.$$

We then compute

$$\begin{aligned} \inf_{m \in A \cap \mathcal{CN}} H(m^w | \lambda_{2/3}) &= \inf \{H(m_q^w | \lambda_{2/3}) : q \in [0, 1], m_q \in A\} \\ &= \inf \{H(\lambda_q | \lambda_{2/3}) : q \in [0, r]\} \\ &= \inf \{H(\lambda_q | \lambda_{2/3}) : q \in [0, r)\} \\ &= \inf_{m \in A^\circ \cap \mathcal{CN}} H(m^w | \lambda_{2/3}), \end{aligned}$$

where the second to last equality holds by continuity of $q \mapsto H(\lambda_q | \lambda_{2/3})$ at $q = r$. Moreover, the unique minimizer on the left-hand side is m_r , since $q \mapsto H(\lambda_q | \lambda_{2/3})$ is strictly decreasing for $0 < q < 2/3$. This shows that the assumption (2.7) of Remark 2.6 holds, and also that the set $S(A)$ therein is simply the singleton $\{m_r\}$.

Remark 3.13. Interestingly, a simple variant of the above game yields a tractable example in which the Cournot-Nash equilibria are not unique and yet Theorem 2.5 can be applied. For instance, suppose $\mathcal{W} = \{-1, 1\}$ and $\mathcal{X} = \{0, 1\}$, with

$$F(m, w, x) = -x(2m^x\{1\} + w).$$

Note that there is no minus sign in front of $2m^x\{1\}$, so it is not really an entry game; agents are now *encouraged* to participate (i.e., choose $x = 1$) when other agents participate. Setting $m_q^1(dw, dx) = \lambda_q(dw)\delta_1(dx)$, it can be checked that $m_q^1 \in \mathcal{CN}(\lambda_q)$ for each $q \in [0, 1]$; that is, it is always an equilibrium for every agent to participate. However, there are two (resp. three) equilibria for $q = 1/2$ (resp. $q < 1/2$). Nonetheless, if μ_n is the empirical type-action distribution when types are sampled from λ_p , where $p > 1/2$, we can find a limit theorem for the law of μ_n conditioned on the event $\{\mu_n \in A\}$ where $A = \{m : m\{(-1, 1)\} \leq 1 - r\}$, for $r \in (p, 1)$ close to p . Indeed, we can check that the assumption (2.5) holds, and the unique element of $S(A)$ is m_r . There is even a critical value of r for which (2.5) holds, but $S(A)$ is no longer a singleton. We

omit the details of these calculations, with the remark mainly serving to illustrate the need for the generality of Theorem 2.5.

3.6. Proof of probabilistic bounds on the price of anarchy. To prove Proposition 2.10, we rework the notation of Section 2.4 as we did in Section 3.1. Recall the notation $\text{Gr}(\mathcal{C}) = \{(w, x) : x \in \mathcal{C}(w)\}$. For (λ, u) belonging to the domain \mathcal{D} defined in (3.3), define $\mathcal{A}(\lambda, u) \subset \mathcal{P}(\mathcal{W} \times \mathcal{X})$ by

$$\mathcal{A}(\lambda, u) := \{m \in \mathcal{E}_u(\mathcal{W} \times \mathcal{X}) : m(\text{Gr}(\mathcal{C})) = 1, m^w = \lambda\},$$

Interpret $\mathcal{A}(\lambda, u)$ as the set of admissible type-action distributions. Recall the notation

$$V(m) = \int_{\mathcal{W} \times \mathcal{X}} F(m, w, x) m(dw, dx),$$

for $m \in \mathcal{P}(\mathcal{W} \times \mathcal{X})$. The price of anarchy is now defined as the function $\mathfrak{P} : \mathcal{D} \rightarrow [1, \infty]$ given by

$$\mathfrak{P}(\lambda, u) = \sup_{m \in \mathcal{N}(\lambda, 0, u)} V(m) \Bigg/ \inf_{m \in \mathcal{A}(\lambda, u)} V(m).$$

Lemma 3.14. *Suppose $V > 0$ pointwise. Then \mathfrak{P} is upper semicontinuous on \mathcal{D} .*

Proof. First, $(\lambda, u) \mapsto \sup_{m \in \mathcal{N}(\lambda, 0, u)} V(m)$ is upper semicontinuous because V is continuous and because, by Proposition 3.2, \mathcal{N} is upper hemicontinuous and has compact values (see [3, Lemma 17.30]). It suffices (since $V > 0$) to show that the denominator $\inf_{m \in \mathcal{A}(\lambda, u)} V(m)$ is lower semicontinuous and strictly positive. For both of these claims it suffices to show that the set-valued map \mathcal{A} is upper hemicontinuous and has compact values (again by [3, Lemma 17.30]). To prove this, we again use the sequential characterization of upper hemicontinuity. Fix a convergent sequence $(\lambda_n, u_n) \rightarrow (\lambda, u)$ in \mathcal{D} , and let $m_n \in \mathcal{A}(\lambda_n, u_n)$ for each n . We must show that there exist $m \in \mathcal{A}(\lambda, u)$ and a subsequence $\{m_{n_k}\}$ which converges to m . Because $m_n^w = \lambda_n$, the sequence $\{m_n^w\} \subset \mathcal{P}(\mathcal{W})$ is tight. Because \mathcal{X} is compact, the sequence $\{m_n\} \subset \mathcal{P}(\mathcal{W} \times \mathcal{X})$ is tight and thus precompact by Prokhorov's theorem. Let m denote any limit point, and abuse notation by assuming $m_n \rightarrow m$.

It remains to show that m belongs to $\mathcal{A}(\lambda, u)$. As $\text{Gr}(\mathcal{C})$ is closed, the Portmanteau theorem implies $m(\text{Gr}(\mathcal{C})) = \lim_n m_n(\text{Gr}(\mathcal{C})) = 1$. Clearly

$$m^w = \lim_n m_n^w = \lim_n \lambda_n = \lambda,$$

where the limits are in distribution. Finally, to check that $m \in \mathcal{E}_u(\mathcal{W} \times \mathcal{X})$, there are two cases. If $u = 0$, then $\mathcal{E}_u(\mathcal{W} \times \mathcal{X}) = \mathcal{P}(\mathcal{W} \times \mathcal{X})$ and there is nothing to prove. Otherwise, $u_n = u$ for all sufficiently large n , which implies m_n and thus m belong to the closed set $\mathcal{E}_u(\mathcal{W} \times \mathcal{X})$. \square

Proof of Proposition 2.10. The notation of Proposition 2.10 translates as follows to the present notation, for $n \geq 1$ and $w_1, \dots, w_n \in \mathcal{W}$:

$$\text{PoA}_n(w_1, \dots, w_n) = \mathfrak{P}\left(\frac{1}{n} \sum_{i=1}^n \delta_{w_i}, \frac{1}{n}\right), \quad \text{and} \quad \text{PoA}(\lambda) = \mathfrak{P}(\lambda, 0).$$

When $(W_i)_{i=1}^\infty$ are i.i.d. with distribution λ_0 , we know that $\frac{1}{n} \sum_{i=1}^n \delta_{W_i} \rightarrow \lambda_0$ almost surely. By Lemma 3.14,

$$\limsup_{n \rightarrow \infty} \text{PoA}_n(W_1, \dots, W_n) \leq \text{PoA}(\lambda_0), \text{ a.s.}$$

To prove the second claim, we apply Sanov's theorem. Consider the set

$$B = \{(\lambda, u) \in \mathcal{D} : \mathfrak{P}(\lambda, u) \geq r\}.$$

Because \mathfrak{P} is upper semicontinuous on \mathcal{D} , the set B is closed in \mathcal{D} and thus in $\mathcal{P}(\mathcal{W}) \times [0, 1]$. By Sanov's theorem, $(\frac{1}{n} \sum_{i=1}^n \delta_{W_i}, \frac{1}{n})$ satisfies an LDP on $\mathcal{P}(\mathcal{W}) \times [0, 1]$ with good rate function

$$J(\lambda, u) = \begin{cases} H(\lambda|\lambda_0) & \text{if } \lambda \in \mathcal{P}(\mathcal{W}), u = 0, \\ \infty & \text{otherwise.} \end{cases}$$

Thus, we have

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\text{PoA}_n(W_1, \dots, W_n) \geq r) &= \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P} \left(\left(\frac{1}{n} \sum_{i=1}^n \delta_{W_i}, \frac{1}{n} \right) \in B \right) \\ &\leq - \inf_{(\lambda, u) \in B} J(\lambda, u) \\ &= - \inf \{ H(\lambda|\lambda_0) : \lambda \in \mathcal{P}(\mathcal{W}), \text{PoA}(\lambda) \geq r \}. \end{aligned}$$

To prove the lower bound, simply apply the lower bound of Sanov's theorem to the set B^c . \square

APPENDIX A. HEMICONTINUITY AND THE UPPER VIETORIS TOPOLOGY

Let us first recall some basic definitions regarding set-valued functions. Let \mathcal{Z} and \mathcal{Y} be topological spaces throughout this section. Let $\Gamma : \mathcal{Z} \rightarrow 2^{\mathcal{Y}}$ map points in \mathcal{Z} to subsets of \mathcal{Y} . We say that the set-valued map Γ is *upper hemicontinuous* if $\{z \in \mathcal{Z} : \Gamma(z) \subset A\}$ is open in \mathcal{Z} for every open set $A \subset \mathcal{Y}$, and we say that Γ is *lower hemicontinuous* if $\{z \in \mathcal{Z} : \Gamma(z) \cap A \neq \emptyset\}$ is open in \mathcal{Z} for every open set $A \subset \mathcal{Y}$. Say that Γ is *continuous* if it is both upper and lower hemicontinuous. If \mathcal{Y} is compact Hausdorff, and if $\Gamma(z)$ is closed for each z , then Γ is upper hemicontinuous if and only if its graph $\text{Gr}(\Gamma) = \{(z, y) \in \mathcal{Z} \times \mathcal{Y} : y \in \Gamma(z)\}$ is closed [3, Theorem 17.11].

If \mathcal{Z} and \mathcal{Y} are metric spaces, there is a useful sequential characterization (c.f. Theorems 17.16 and 17.19 of [3]): first, Γ is lower hemicontinuous if and only if, whenever $z_n \rightarrow z$ in \mathcal{Z} and $y \in \Gamma(z)$, there exist integers $1 \leq n_1 < n_2 < \dots$ and $y_{n_k} \in \Gamma(z_{n_k})$ such that $y_{n_k} \rightarrow y$. Second, a map Γ with compact values is upper hemicontinuous if and only if, whenever $z_n \rightarrow z$ in \mathcal{Z} and $y_n \in \Gamma(z_n)$ for all n , the sequence $\{y_n\}$ is precompact, and every limit point belongs to $\Gamma(z)$.

The *upper Vietoris topology* on $\mathfrak{C}(\mathcal{Y})$ is the one generated by the base $\{E^+ : E \subset \mathcal{Y} \text{ is open}\}$, where we define $E^+ := \{A \in \mathfrak{C}(\mathcal{Y}) : A \subset E\}$ for sets $E \subset \mathcal{Y}$. The rest of this section collects a few basic facts about this topology. First, notice that $\mathfrak{C}(\mathcal{Y})$ is not Hausdorff, because if A_1 and A_2 are two distinct closed subsets of \mathcal{Y} with $A_1 \cap A_2 \neq \emptyset$, then A_1 and A_2 cannot be separated by open sets; indeed, if $E_i \subset \mathcal{Y}$ is open with $A_i \in E_i^+$ for $i = 1, 2$, then $A_i \subset E_i$, and $A_1 \cap A_2 \subset E_1 \cap E_2$ implies that E_1^+ and E_2^+ are not disjoint.

The upper Vietoris topology corresponds to upper hemicontinuity in the following sense. Consider a map $\Gamma : \mathcal{Z} \rightarrow \mathfrak{C}(\mathcal{Y})$. Then Γ is upper hemicontinuous if and only if $\{z \in \mathcal{Z} : \Gamma(z) \subset E\}$ is open for every open set $E \subset \mathcal{Y}$. Equivalently, Γ is upper hemicontinuous if and only if $\Gamma^{-1}(E^+) := \{z \in \mathcal{Z} : \Gamma(z) \in E^+\}$ is open for every open set $E \subset \mathcal{Y}$. Because $\{E^+ : E \subset \mathcal{Y} \text{ is open}\}$ is a base for the upper Vietoris topology, this immediately proves the following observation:

Lemma A.1. *Let \mathcal{Z} be another topological space. A mapping $\Gamma : \mathcal{Z} \rightarrow \mathfrak{C}(\mathcal{Y})$ is continuous with respect to the upper Vietoris topology if and only if it is upper hemicontinuous as a set-valued map.*

Lemma A.2. *Let $g : \mathcal{Y} \rightarrow \mathbb{R}$, and define $G : \mathfrak{C}(\mathcal{Y}) \rightarrow \mathbb{R}$ by $G(A) = \sup_{y \in A} g(y)$. If g is upper semicontinuous, then so is G . If g is continuous, and if $A \in \mathfrak{C}(\mathcal{Y})$ is a compact set such that g is constant on A (i.e., $g(y) = g(y')$ for all $y, y' \in A$), then G is continuous at A .*

Proof. The identity map on $\mathfrak{C}(\mathcal{Y})$ is continuous and thus, by Lemma A.1, can be seen as an upper hemicontinuous set-valued map. The first claim then follows from [3, Lemma 17.30]. To prove the second claim, define a set-valued map $\Gamma : \mathfrak{C}(\mathcal{Y}) \rightarrow 2^{\mathbb{R}}$ by $\Gamma(B) = \{g(y) : y \in B\}$. Then we may write $G(B) = \sup_{r \in \Gamma(B)} r$ for every $B \in \mathfrak{C}(\mathcal{Y})$. Because g is continuous, Γ is upper hemicontinuous [3, Theorem 17.23]. Note $\Gamma(A) = \{r_0\}$ is a singleton, by assumption.

To prove G is lower semicontinuous at A , suppose A_α is a net in $\mathfrak{C}(\mathcal{Y})$ converging to A . Choose arbitrarily $r_\alpha \in \Gamma(A_\alpha)$ for each α . By upper hemicontinuity of Γ and compactness of $\Gamma(A)$, the net (r_α) has a limit point in $\Gamma(A)$ by [3, Theorem 17.16]. Hence, $r_\alpha \rightarrow r_0$. Thus

$$\liminf_{\alpha} G(A_\alpha) \geq \liminf_{\alpha} r_\alpha = r_0 = G(A).$$

□

Next, we identify the interiors and closures of certain subsets of $\mathfrak{C}(\mathcal{Y})$. Recall that the interior of a set is simply the union of its open subsets, and the closure of a set is the intersection of all closed sets containing it. Take note also that $\{E^+ : E \subset \mathcal{Y} \text{ is open}\}$ is a base, and so every open set in $\mathfrak{C}(\mathcal{Y})$ can be written as a union of these base elements. In the following lemma, especially the proof, we will be applying repeated complements, interiors, and $+$ operations, and we prefer to keep parentheses to a minimum by writing, e.g., $E^{c \circ +}$ in place of $((E^c)^\circ)^+$.

Lemma A.3. *Let $E \subset \mathcal{Y}$ be any set, and define*

$$\mathfrak{U} := \{A \in \mathfrak{C}(\mathcal{Y}) : A \cap E \neq \emptyset\} = E^{c+c}, \text{ and}$$

$$\mathfrak{W} := \{A \in \mathfrak{C}(\mathcal{Y}) : A \subset E\} = E^+.$$

The following hold:

- (i) $\mathfrak{U}^\circ = \mathfrak{W}^\circ = \{A \in \mathfrak{C}(\mathcal{Y}) : A \subset E^\circ\}$. In other words, $E^{c+c \circ} = E^{+ \circ} = E^{\circ +}$.
- (ii) $\overline{\mathfrak{U}} = \overline{\mathfrak{W}} = \{A \in \mathfrak{C}(\mathcal{Y}) : A \cap \overline{E} \neq \emptyset\}$. In other words, $\overline{E^{c+c}} = \overline{E^+} = \overline{E}^{c+c}$.

Proof.

- (i) Suppose $\tilde{E} \subset \mathcal{Y}$ is open. Then $\tilde{E}^+ \subset \mathfrak{U}$ if and only if $A \cap E \neq \emptyset$ for every closed set $A \subset \tilde{E}$. By considering $A = \{x\}$ for $x \in \tilde{E}$ (which is closed because \mathcal{Y} is Hausdorff), we see that $\tilde{E}^+ \subset \mathfrak{U}$ if and only if $\tilde{E} \subset E$. Thus the interior of \mathfrak{U} is the union over all sets \tilde{E}^+ such that \tilde{E} is open and $\tilde{E} \subset E$, and the largest such set is given by $\tilde{E} = E^\circ$. This shows $\mathfrak{U}^\circ = E^{\circ +} = \{A \in \mathfrak{C}(\mathcal{Y}) : A \subset E^\circ\}$. To show that $\mathfrak{W}^\circ = E^{\circ +}$, note first that clearly $E^{\circ +} \subset \mathfrak{W}^\circ$. On the other hand, if $\tilde{E} \subset \mathcal{Y}$ is any open set such that $\tilde{E}^+ \subset \mathfrak{W}$, then $A \subset \tilde{E}$ implies $A \subset E$ for every closed set $A \subset \mathcal{Y}$. Taking $A = \{x\}$, we conclude that $\tilde{E}^+ \subset \mathfrak{W}$ implies $\tilde{E} \subset E$, which in turn implies $\tilde{E} \subset E^\circ$ and $\tilde{E}^+ \subset E^{\circ +}$.
- (ii) Recall the identities $\overline{A} = A^{c \circ c}$ and $\overline{A^c} = A^{c \circ}$, valid for any set A in any topological space. Namely, apply (i) with E^c in place of E to get $E^{c \circ +} = E^{c \circ c}$, and thus

$$\overline{\mathfrak{U}} = \overline{E^{c+c}} = E^{c+c \circ c} = E^{c \circ c} = E^{c \circ +} = \overline{E}^{c+c}.$$

Similarly, apply (i) with E^c in place of E to get $E^{+ \circ} = E^{c \circ c} = E^{c \circ +}$, and thus

$$\overline{\mathfrak{W}} = \overline{E^+} = E^{+ \circ c} = E^{c \circ +} = \overline{E}^{c+c}.$$

□

APPENDIX B. NONATOMIC CONGESTION GAMES

This section is devoted to existence and uniqueness results for nonatomic congestion games, namely the proofs of Propositions 2.11 and 2.12.

Proof of Proposition 2.11. For each $m, \tilde{m} \in \mathcal{P}(\mathcal{W})$, define the directional derivative

$$D_{\tilde{m}}U(m) := \frac{d}{d\epsilon}U(m + \epsilon(\tilde{m} - m))|_{\epsilon=0}.$$

Noting that $\ell_e(m) = \int 1_{\{e \in x\}} m(dw, dx)$ for each $e \in E$, we compute

$$\begin{aligned} D_{\tilde{m}}U(m) &= \sum_{e \in E} c_e(\ell_e(m)) \frac{d}{d\epsilon} \ell_e(m + \epsilon(\tilde{m} - m))|_{\epsilon=0} \\ &= \sum_{e \in E} c_e(\ell_e(m)) \int 1_{\{e \in x\}} (\tilde{m} - m)(dw, dx) \\ &= \int \sum_{e \in x} c_e(\ell_e(m)) (\tilde{m} - m)(dw, dx) \\ &= \int F(m, x) (\tilde{m} - m)(dw, dx). \end{aligned}$$

By definition, $m \in \mathcal{CN}(\lambda)$ if and only if $\int F(m, x) (\tilde{m} - m)(dw, dx) \geq 0$ for every $\tilde{m} \in \mathcal{A}(\lambda)$. On the other hand, m minimizes $U(\cdot)$ on $\mathcal{A}(\lambda)$ if and only if $D_{\tilde{m}}U(m) \geq 0$ for every $\tilde{m} \in \mathcal{A}(\lambda)$ \square

Proof of Proposition 2.12. Denote a generic element of \mathbb{T} by $K = (K_{i,j})_{i,j}$, where $i = 1, \dots, |\mathcal{W}|$ and $j = 1, \dots, |\mathcal{X}|$. Define $U_\lambda : \mathbb{T} \rightarrow \mathbb{R}$ by $U_\lambda(K) = U(m_K)$, where $m_K \in \mathcal{A}(\lambda)$ is given by $m_K\{(w_i, x_j)\} = \lambda_i K_{i,j}$, where $\lambda_i := \lambda\{w_i\}$. Note that

$$\ell_e(K) := \ell_e(m_K) = \sum_{i=1}^{|\mathcal{W}|} \sum_{j=1}^{|\mathcal{X}|} 1_{\{e \in x_j\}} \lambda_i K_{i,j}.$$

Then $\partial_{K_{i,j}} \ell_e(K) = \lambda_i 1_{\{e \in x_j\}}$, and so

$$\begin{aligned} \partial_{K_{i,j}} U_\lambda(K) &= \partial_{K_{i,j}} \sum_{e \in E} \int_0^{\ell_e(K)} c_e(s) ds = \sum_{e \in E} c_e(\ell_e(K)) \lambda_i 1_{\{e \in x_j\}}, \\ \partial_{K_{i',j'}} \partial_{K_{i,j}} U_\lambda(K) &= \sum_{e \in E} c'_e(\ell_e(K)) \lambda_i \lambda_{i'} 1_{\{e \in x_j\}} 1_{\{e \in x_{j'}\}}. \end{aligned}$$

Hence, for any $T = (T_{i,j}) \in \mathbb{T}$, we have

$$\begin{aligned} \sum_{i,j} \sum_{i',j'} T_{i,j} T_{i',j'} \partial_{K_{i',j'}} \partial_{K_{i,j}} U_\lambda(K) &= \sum_{e \in E} c'_e(\ell_e(K)) \sum_{i,j} \sum_{i',j'} T_{i,j} T_{i',j'} \lambda_i \lambda_{i'} 1_{\{e \in x_j\}} 1_{\{e \in x_{j'}\}} \\ &= \sum_{e \in E} c'_e(\ell_e(K)) \left(\sum_{i,j} T_{i,j} \lambda_i 1_{\{e \in x_j\}} \right)^2. \end{aligned}$$

Note that $c'_e > 0$ for each $e \in E$ by assumption, and also the squared sum in the last expression is strictly positive for some e since $T \neq 0$ belongs to the span of $(\lambda_i 1_{\{e \in x_j\}})_{i,j}$ by assumption. This shows that the Hessian of U_λ is positive definite everywhere, and so U_λ has a unique minimizer on the compact convex set \mathbb{T} . \square

APPENDIX C. PROOFS OF LEMMAS 3.10 AND 3.11

Proof of Lemma 3.10. Note that

$$\inf_{A \in \mathfrak{U}} J(A) = \inf \{H(\lambda|\lambda_0) : \lambda \in \mathcal{P}(\mathcal{W}), \mathcal{CN}(\lambda) \in \mathfrak{U}\}.$$

Recall that $\lambda \mapsto \mathcal{CN}(\lambda) = \mathcal{N}(\lambda, 0, 0) \in \mathfrak{C}$ is continuous by Proposition 3.2 and Lemma A.1. Hence, the set $S = \{\lambda \in \mathcal{P}(\mathcal{W}) : \mathcal{CN}(\lambda) \in \mathfrak{U}\}$ is closed because \mathfrak{U} is. Because $\lambda \mapsto H(\lambda|\lambda_0)$ is lower semicontinuous and has compact sub-level sets, there exists $\lambda^* \in S$ such that $H(\lambda^*|\lambda_0) = \inf_{A \in \mathfrak{U}} J(A)$. But $\mathcal{CN}(\lambda^*) \in \mathfrak{U}$ implies $\lambda^* \neq \lambda_0$, and thus $H(\lambda^*|\lambda_0) > 0$. \square

To prove Lemma 3.11, we first need a preparatory lemma:

Lemma C.1. *The set \mathcal{CN} is closed. Moreover, the sub-level set $\{m \in \mathcal{P}(\mathcal{W} \times \mathcal{X}) : H(m^w|\lambda_0) \leq c\}$ is compact for every $c < \infty$.*

Proof. Suppose $m_n \in \mathcal{CN}$ and $m \in \mathcal{P}(\mathcal{W} \times \mathcal{X})$ with $m_n \rightarrow m$. Then $\lambda_n := m_n^w$ converges to $\lambda := m^w$. Then $m_n \in \mathcal{CN}(\lambda_n) = \mathcal{N}(\lambda_n, 0, 0)$, and the upper hemicontinuity of \mathcal{N} (proven in Proposition 3.2) implies that the unique limit point m must belong to $\mathcal{N}(\lambda, 0, 0) \subset \mathcal{CN}$. This proves that \mathcal{CN} is closed. The second statement follows because \mathcal{X} is compact and the sub-level set $\{\lambda \in \mathcal{P}(\mathcal{W}) : H(\lambda|\lambda_0) \leq c\}$ is compact for each $c < \infty$ [18, Lemma 1.4.3(c)]. \square

Proof of Lemma 3.11. Since \mathcal{CN} is closed and the sublevel sets $\{m : H(m^w|\lambda_0) \leq c\}$ are compact by Lemma C.1, $A \cap \mathcal{CN}$ is closed and there exists $m_* \in A \cap \mathcal{CN}$ such that $H(m_*^w|\lambda_0) = \inf_{m \in A \cap \mathcal{CN}} H(m|\lambda_0)$. But $A \cap \mathcal{CN}(\lambda_0) = \emptyset$ implies $m_*^w \neq \lambda_0$, and thus, $H(m_*^w|\lambda_0) > 0$. \square

REFERENCES

1. S. Adlakha and R. Johari, *Mean field equilibrium in dynamic games with strategic complementarities*, Operations Research **61** (2013), no. 4, 971–989.
2. S. Adlakha, R. Johari, G.Y. Weintraub, and A. Goldsmith, *Oblivious equilibrium for large-scale stochastic games with unbounded costs*, Decision and Control, 2008. CDC 2008. 47th IEEE Conference on, IEEE, 2008, pp. 5531–5538.
3. C. Aliprantis and K. Border, *Infinite dimensional analysis: A hitchhiker’s guide*, 3 ed., Springer, 2007.
4. P. Bajari, H. Hong, and S.P. Ryan, *Identification and estimation of a discrete game of complete information*, Econometrica **78** (2010), no. 5, 1529–1568.
5. S. Berry, *Estimation of a model of entry in the airline industry*, Econometrica: Journal of the Econometric Society (1992), 889–917.
6. D. Bertsekas and S. Shreve, *Stochastic optimal control: The discrete time case*, Athena Scientific, 1996.
7. A. Blanchet and G. Carlier, *From Nash to Cournot-Nash equilibria via the Monge-Kantorovich problem*, Phil. Trans. R. Soc. A **372** (2014), no. 2028, 20130398.
8. ———, *Remarks on existence and uniqueness of Cournot-Nash equilibria in the non-potential case*, Mathematics and Financial Economics **8** (2014), no. 4, 417–433.
9. ———, *Optimal transport and Cournot-Nash equilibria*, Mathematics of Operations Research **41** (2015), no. 1, 125–145.
10. A. Blanchet, G. Carlier, and L. Nenna, *Computation of Cournot-Nash equilibria by entropic regularization*, (2016).
11. T.F. Bresnahan and P.C. Reiss, *Empirical models of discrete games*, Journal of Econometrics **48** (1991), no. 1, 57–81.
12. P. Cardaliaguet, F. Delarue, J.-M. Lasry, and P.-L. Lions, *The master equation and the convergence problem in mean field games*, arXiv preprint arXiv:1509.02505 (2015).
13. G. Carmona, *Nash equilibria of games with a continuum of players*, (2004).
14. R. Carmona and F. Delarue, *Probabilistic analysis of mean field games*, SIAM Journal on Control and Optimization **51** (2013), 2705–2734.
15. G. Christodoulou and E. Koutsoupias, *The price of anarchy of finite congestion games*, Proceedings of the thirty-seventh annual ACM symposium on Theory of computing, ACM, 2005, pp. 67–73.
16. I. Csiszár, *Sanov property, generalized I-projection and a conditional limit theorem*, The Annals of Probability (1984), 768–793.
17. A. Dembo and O. Zeitouni, *Large deviations techniques and applications*, Stochastic Modelling and Applied Probability, Springer, 2009.
18. P. Dupuis and R.S. Ellis, *A weak convergence approach to the theory of large deviations*, vol. 902, John Wiley & Sons, 2011.

19. M. Fischer, *On the connection between symmetric n -player games and mean field games*, arXiv preprint arXiv:1405.1345 (2014).
20. D.A. Gomes, J. Mohr, and R.R. Souza, *Discrete time, finite state space mean field games*, Journal de mathématiques pures et appliquées **93** (2010), no. 3, 308–328.
21. E.J. Green, *Continuum and finite-player noncooperative models of competition*, Econometrica: Journal of the Econometric Society (1984), 975–993.
22. J.C. Harsanyi, *Games with incomplete information played by "Bayesian" players, i-iii*, Management Science **14** (1967), no. 3, 159–182.
23. D. Housman, *Infinite player noncooperative games and the continuity of the Nash equilibrium correspondence*, Mathematics of Operations Research **13** (1988), no. 3, 488–496.
24. E. Kalai, *Large robust games*, Econometrica **72** (2004), no. 6, 1631–1665.
25. P. Klemperer, *Auction theory: A guide to the literature*, Journal of economic surveys **13** (1999), no. 3, 227–286.
26. E. Koutsoupias and C. Papadimitriou, *Worst-case equilibria*, Annual Symposium on Theoretical Aspects of Computer Science, Springer, 1999, pp. 404–413.
27. V. Krishna, *Auction theory*, 2 ed., Academic press, 2009.
28. D. Lacker, *A general characterization of the mean field limit for stochastic differential games*, Probability Theory and Related Fields, 1–68.
29. J.M. Lasry and P.L. Lions, *Mean field games*, Japanese Journal of Mathematics **2** (2007), 229–260.
30. A. Mas-Colell, *On a theorem of Schmeidler*, Journal of Mathematical Economics **13** (1984), no. 3, 201–206.
31. K. Menzel, *Inference for games with many players*, The Review of Economic Studies **83** (2016), no. 1, 306–337.
32. D. Monderer and L.S. Shapley, *Potential games*, Games and economic behavior **14** (1996), no. 1, 124–143.
33. N. Nisan, R. Roughgarden, E. Tardos, and V.V. Vazirani, *Algorithmic game theory*, vol. 1, Cambridge University Press Cambridge, 2007.
34. R.W. Rosenthal, *A class of games possessing pure-strategy Nash equilibria*, International Journal of Game Theory **2** (1973), no. 1, 65–67.
35. T. Roughgarden and É. Tardos, *How bad is selfish routing?*, Journal of the ACM (JACM) **49** (2002), no. 2, 236–259.
36. A.S. Sznitman, *Topics in propagation of chaos*, Ecole d'été de Probabilités de Saint-Flour XIX - 1989, Lecture Notes in Mathematics, vol. 1464, Springer Berlin / Heidelberg, 1991, pp. 165–251.
37. J.G. Wardrop, *Some theoretical aspects of road traffic research*, Proceedings of the Institution of Civil Engineers **1** (1952), no. 3, 325–378.
38. G.Y. Weintraub, C.L. Benkard, and B. Van Roy, *Markov perfect industry dynamics with many firms*, Econometrica **76** (2008), no. 6, 1375–1411.