Università degli Studi di Padova

DIPARTIMENTO DI MATEMATICA "TULLIO LEVI-CIVITA"

Corso di Laurea Magistrale in Matematica

Tesi di laurea magistrale

Correlated Equilibria in Static Mean-Field Games

Relatore:
Prof. Markus Fischer

Candidato:
Laura Degl'Innocenti
Matricola 1155494

Anno Accademico 2017-2018
20 Luglio 2018

Contents

Ringraziamenti

Vorrei ringraziare innanzitutto il mio relatore, il Prof. Markus Fisher, che mi ha accompagnata e guidata in questi mesi di lavoro; lo ringrazio anche per la disponibilità e gentilezza che mi ha sempre dimostrato.
Un pensiero va anche ai miei compagni di avventura, Mariapia, Davide R., Alessandro, Davide P., Edoardo, Beatrice e tutti gli altri, troppi da elencare; stare vicino a voi ha reso piacevole anche la permanenza in Torre, cosa per niente scontata a parere di tutti. Spero che anche io sia stata tanto di aiuto a voi quanto voi lo siete stati per me. Un grazie anche ai miei compagni, ma soprattutto amici, del liceo con cui non solo è piacevole ricordare i divertimenti trascorsi sui banchi di scuola, ma anche crearne di nuovi.
Ringrazio poi Giulia, Giorgio, Silvia, ed Erica, coloro che hanno reso le abitazioni in cui sono stata delle vere Case, nidi padani in sostituzione alla mia vera casa collinare.
Greta e Scilla, a voi non c'è bisogno che dica quanto vi sia grata. Ormai sono troppi anni che mi sopportate; grazie per non dimenticarvi di me a chilometri di distanza, anche quando vi darei mille motivazioni per farlo.
Il ringraziamento speciale va alla mia famiglia, che mi ha permesso di conseguire ciò per cui ho sudato in questi cinque anni e soprattutto di incontrare persone fantastiche con cui ho condiviso molto. La prova più tangibile di quanto vi voglia bene è il fatto che, appena ne ho l'occasione, scappo per tornare a casa. Coloro che non incontro quasi mai né a Padova né a Rignano sono i miei cugini, cui però voglio dedicare queste due righe: i due ”grandi” sono stati per me ispirazione e modello per il mio percorso, i tre ”piccoli”, per i quali spero di essere io un tal figura. Un bacio anche a i miei nonni e i miei zii, parte per me fondamentale della vera famiglia.
Un ovvio, ma non banale, grazie va a Lucio, che mi ha veramente sopportata nonostante tutto e ha sempre teso le braccia e spronata quando pensavo di non farcela; probabilmente senza di te questa tesi non sarebbe finita adesso.
Infine ringrazio tutti voi, anche chi non ho nominato, collettivamente; spero che chi mi è vicino in questo traguardo, lo sarà anche per i prossimi futuri, nella speranza di attraversarne tanti insieme.

Introduction

In game theory the main role is typically played by Nash equilibria; the correlated is a type of equilibrium which generalizes the first one. It was largely developed by R.J. Aumann first in 1974 and then in 1987, [2] and [3]. Its idea is simple and maybe more intuitive than that of Nash, indeed in the latter every players choose his action independently one from the other, instead in correlated strategies all players observe the same random event and then choose their action. The most common example that clarifies the idea of the differences between these two concept, is the battle of sexes; in this case there are two players, husband and wife, who have to choose either to watch a baseball game or to go to the opera; the pay-off matrix is the following:

H/W	B	O
B	(2,1)	(0,0)
O	(0,0)	(1,2)

We can immediately observe that there are two pure Nash equilibria (B,B) and (O,O) whose pay-offs are respectively (2,1) and (1,2) with costs. The mixed one is instead $([2/3, 1/3], [1/3, 2/3])$, and the expected reward is 2/3 both for the husband and the wife. Suppose now that the two spouses call a ”mediator” who suggests to the players what to choose as a consequence of a random event; suppose that this random event is a coin flip, such that if its outcome is Head the spouses are told to choose Baseball and if its Tails they are told to choose the Opera. Now the expected pay-off for both is 3/2 which is strictly grater than those of Nash; this is due to the fact that the players don't play independently one from the other, but their are linked by a correlation device, that in this case is the coin flip. Hence, it is important to underline that such a ”mediator” doesn't imply that it is a cooperative game, on the contrary we are still on the framework of the non-cooperative games, since the players are not allowed to talk to each other when making their decisions. We can summarize the ideas of the previous example by saying that a Nash equilibrium, or more in general a mixed strategy, is a probability distribution over the set of the pure strategies for the single players, and a correlated strategy is a probability distribution over the product of all the strategy sets.
Now, the purpose of this thesis is to introduce the concept of correlated equilibrium in the theory of Mean-Field Games. This general theory was first introduced by J.-M. Lasry and P.-L. Lions between 2006 and 2007 in a series of papers of which we cite [14], and it is

dedicated to the analysis of games with a "large number of small players"; this means that while the number of agents tends to infinity, the impact of each one on the overall system tends to 0. They arise from the mean field models in mathematical physics which study the behaviour of many identical particles whose interactions depend on their empirical mean and this explains the name "mean field". Usually a solution of the mean field game is used to construct an approximated Nash equilibrium for the finite case, because when N is large enough, finding directly an equilibrium is not possible.

Fischer, see [9] in the reference therein, has shown how is possible to go also in the opposite direction, namely how to start for an ε-Nash equilibrium and obtain a solution for the mean-field equation. The idea is the following: consider an equation which describes the evolution of the private states of the N players

$$dX_i^N(t) = b(t, X_i^N(t), \mu^N(t), u_i(t))dt + \sigma(t, X_i^N(t), \mu^N(t))dW_i^N(t)$$

with $\mu^N := \frac{1}{N}\sum_i \delta_{X_i^N}$ the empirical mean of the private states and $u := (u_1, \dots, u_N)$ is the strategy vector and its cost is:

$$J_i^N(u) := \mathbb{E}\left[\int_0^T f(s, X_i^N(s), \mu^N(s), u_i(s))ds + F(X_i^N(T), \mu^N(T))\right]$$

heuristically if u is a Nash equilibrium with correspondent vector of the private states X^N, then the empirical measure μ^N should converge in distribution to a deterministic flow of measure m, induced by the solution of

$$dX(t) = b(t, X(t), m(t), u(t))dt + \sigma(t, X(t), m(t))dW(t) \tag{1}$$

where u realizes the following minimum:

$$\min J_m(v) := \mathbb{E}\left[\int_0^T f(s, X(s), m(s), v(s)ds + F(t, X(T), m(T)\right] \tag{2}$$

where the minimun is taken over all admissible v such that X solves under v.
The system (1)-(2) is the limit game for the N-player one, and works in this way: for each m we solve (2) to find the correspondent strategy u^m, then we solve (1) and get the correspondent X^m. We now choose $m(\cdot) = \text{Law}(X^m(\cdot))$. Thus the solution of the mean field game is identified with the pair $(\text{Law}(X^m(\cdot), u^m, W), m)$.

Motivated by the examples such as the "battle of sexes", we have ask ourselves if it would be worthy to study the behaviour of correlated equilibria of a N-person game with interactions of the mean-field type, when $N \to \infty$. We have confirmed this idea by finding an example that fits our model of a N-player game in which a correlated equilibrium is strictly better then the Nash ones for every N.
Thus, we have study the convergence in the case of static games, that could be an indicator of the possibility of a further generalization. The main difference between the Nash case is that in correlated strategy there isn't independence between the strategy (seen as a random vector which has values on the set of pure strategies) and so between the private

states of the players; we will see how this difference makes the difficulties rise when we are searching a mean-field equation which has to be satisfied by the equilibria.
It has to be notice that typically in game theory the final goal is to maximize the so called *expected pay-off*, on the contrary mean-field games are usually problems of minimization of the costs; in this work we go in this latter direction, indeed both the examples and the theoretic arguments have the intentions of minimizing a cost function.

The thesis is divided in two Chapters, which describe the finite and the limit case of the model. In Chapter 1 we describe the model, give the definitions, show the properties of the two types of equilibria we deal with, and prove their mutual relations. In Chapter 2 we study at first the behaviour of Nash and than that of correlated equilibria when $N \to \infty$ and find what type of equations they have to satisfy.

Chapter 1

N-player Games

1.1 Description of the model

Consider a static game with N players; let $S \subseteq \mathbb{R}^d$ be the (finite) set of pure strategies (or *actions*) for a single player; we'll consider only symmetric game, so the strategy set is the same for all the players and the general strategy set is $S_N := \times_N S$. Consider a probability space $(\Omega_N, \mathcal{A}_N, \mathbb{P}_N)$; a *strategy vector* is a random vector $\alpha^N = (\alpha_1^N, \dots, \alpha_N^N) : \Omega_N \to S_N$, where $\alpha_i^N : \Omega_N \to S$ is player i's strategy. The dynamic of player i's *private state* is

$$X_i^N = \phi(\alpha_i^N)$$

with $\phi : S \to \mathbb{R}^m$, so $X_i^N(\omega) = \phi(\alpha_i^N(\omega))$. We'll call $\mathcal{X} \subseteq \mathbb{R}^m$ the set $\mathcal{X} := \phi(S)$ The dependence among the players is given only by the cost function:

$$J_i^N(\alpha^N) = \mathbb{E}[g(X_i^N, \bar{\mu}_X^{N,i})] \tag{1.1}$$

where

$$\bar{\mu}_X^{N,i} := \frac{1}{N-1} \sum_{\substack{j=1 \\ j \neq i}}^{N} \delta_{X_j^N}$$

is called *empiric measure* and $\bar{\mu}_X^{N,i} : \Omega_N \to \mathcal{P}(\mathcal{X})$ with $\mathcal{P}(\cdot)$ indicates the probability space of a certain set.

For the calculations, it's often more useful to look at distributions instead of random variables. If we indicate with $p_{\alpha_i^N}$ the distribution over S of α_i^N and with p_{α^N} the joint distribution of α on S_N (i.e. $p_{\alpha_i^N}(s_i) = \mathbb{P}_N(\alpha_i^N = s_i)$ and $p_{\alpha^N}(s) = \mathbb{P}_N(\alpha^N = s) = \mathbb{P}_N(\alpha_1^N =$

$s_1, \dots, \alpha_N^N = s_N)$), then (1.1) could be written explicitly as

$$\begin{aligned} J_i^N(\alpha^N) =& \mathbb{E}\Big[g(\phi(\alpha_i^N), \bar{\mu}_X^{N,i})\Big] = \mathbb{E}\Big[g\Big(\phi(\alpha_i^N), \frac{1}{N-1}\sum_{j\neq i}\delta_{\phi(\alpha_j^N)}\Big)\Big] \\ =& \int_{\Omega_N} g\Big(\phi(\alpha_i^N(\omega)), \frac{1}{N-1}\sum_{j\neq i}\delta_{\phi(\alpha_j^N(\omega))}\Big)\mathbb{P}_N(d\omega) \\ =& \sum_{s\in S_N}\mathbb{P}_N(\alpha^N = s) g\Big(\phi(s_i), \frac{1}{N-1}\sum_{j\neq i}\delta_{\phi(s_j)}\Big) \\ =& \sum_{s\in S_N} p_{\alpha^N}(s) g\Big(\phi(s_i), \frac{1}{N-1}\sum_{j\neq i}\delta_{\phi(s_j)}\Big) =: J_i^N(p_\alpha) \end{aligned} \tag{1.2}$$

If we restrict the class of the strategies to those in which α_i^N are also independent, we're considering the so-called *mixed strategies*; so $\{\alpha_i^N\}_{i=1,\dots,N}$ is a family of independent random variables. We observe that, due to independence, the cost function in the case of mixed strategies becomes

$$\begin{aligned} J_i^N(\alpha^N) &= \sum_{s\in S_N} p_{\alpha^N}(s) g\Big(\phi(s_i), \frac{1}{N-1}\sum_{j\neq i}\delta_{\phi(s_j)}\Big) \\ &= \sum_{s\in S_N}\Big(\prod_{i=1}^N p_{\alpha_i^N}(s_i)\Big) g\Big(\phi(s_i), \frac{1}{N-1}\sum_{j\neq i}\delta_{\phi(s_j)}\Big) \end{aligned}$$

Consider now $\beta^N = (\beta_1^N, \dots, \beta_N^N) : \Omega_N \to S_N$ another strategy; we introduce the following notation:

$$(\alpha_{-i}^N, \beta_i^N) = (\alpha_1^N, \dots, \alpha_{i-1}^N, \beta_i^N, \alpha_{i+1}^N, \dots, \alpha_N^N).$$

Definition 1.1. A mixed strategy $\alpha^N = (\alpha_1^N, \dots, \alpha_N^N)$ is said **Nash equilibrium** if, for every other mixed strategy $\beta^N = (\beta_1^N, \dots, \beta_N^N)$

$$J_i^N(\alpha^N) \le J_i^N(\alpha_{-i}^N, \beta_i^N) \quad i = 1, \dots, N \tag{1.3}$$

We now introduce another type of equilibrium, the correlated one; we will give a definition which is similar to that of Aumann in [2] and we will show that is equivalent to another concept, the Bayes rationality. Thus, consider a probability space $(\Omega_N, \mathbb{P}_N)$. A *correlated strategy* is a random variable $\alpha : \Omega_N \to S_N$ (notice that on the contrary to the previous case, here we are not supposing the independence between the actions of the players).

Definition 1.2. A correlated strategy $\alpha^N = (\alpha_1^N, \dots, \alpha_N^N)$ is said **correlated equilibrium** if, for every i and for every other correlated strategy $\beta^N = (\beta_1^N, \dots, \beta_N^N)$ such that β_i is a function of α_i

$$J_i^N(\alpha^N) \le J_i^N(\alpha_{-i}^N, \beta_i^N) \tag{1.4}$$

This means that for any $d : S \to S$ and for any i

$$J_i^N(\alpha^N) \le J_i^N(\alpha_{-i}^N, d(\alpha_i^N))$$

Explicitly, calling p_α the distribution of α over S_N:

$$\begin{aligned}&\int_{\Omega_N} g(\phi(\alpha_i^N(\omega)), \bar{\mu}^{N,i}_{\phi(\alpha)}(\omega))\mathbb{P}(d\omega) \leq \int_{\Omega_N} g(\phi(d(\alpha_i^N(\omega))), \bar{\mu}^{N,i}_{\phi(\alpha)}(\omega))\mathbb{P}(d\omega)\\ \Leftrightarrow &\sum_{s\in S_N} p_\alpha(s) g(\phi(s_i), \bar{\mu}^{N,i}_s) \leq \sum_{s\in S_N} p_\alpha(s) g(\phi(d(s_i)), \bar{\mu}^{N,i}_s)\end{aligned} \tag{1.5}$$

Bayesian Rationality in Games: Given a N-person game, we assume also as given a probability space $(\Omega_N, \mathbb{P}_N)$, which represent all possible states ω of the world; for each player i a partition $\mathcal{P}_i^N$ of Ω_N, which is i's information partition. So, if the true state of the world is $\omega \in P_i^N \in \mathcal{P}_i^N$, then i knows that some element of P_i^N is the true state of the world but he/she doesn't know which one is it. Notice that conditional on a given ω, everybody knows everything, but, in general nobody knows which is the true ω.
Let $\alpha^N(\omega) := (\alpha_1^N(\omega), \dots, \alpha_N^N(\omega))$ be the n-tuple of actions chosen at state ω. We assume that α_i^N is $\mathcal{P}_i^N$-measurable for any i; this could be interpreted as the fact that each player knows the action he's choosing.

Definition 1.3. A player i is *Bayes rational* at ω if for any $s_i \in S$

$$\mathbb{E}[g(\phi(\alpha_i^N), \bar{\mu}^{N,i}_{\phi(\alpha)}) \mid \mathcal{P}_i] \leq \mathbb{E}[g(\phi(s_i), \bar{\mu}^{N,i}_{\phi(\alpha)}) \mid \mathcal{P}_i] \quad \text{a.s.} \tag{1.6}$$

This means that each player minimize the his cost given his information.

Theorem 1.4. If each player is Bayes rational at each state of the world, then the distribution of the action N-tuple α is a correlated equilibrium distribution.

Proof. If we take β_i as a function of α_i^N, since α_i^N is $\mathcal{P}_i^N$-measurable, β_i also is. The Bayes rationality implies that for each $P \in \mathcal{P}_i^N$

$$\mathbb{E}[g(\phi(\alpha_i^N), \bar{\mu}^{N,i}_{\phi(\alpha)}) \mid P] \leq \mathbb{E}[g(\phi(s_i), \bar{\mu}^{N,i}_{\phi(\alpha)}) \mid P]$$

Taking s_i to be constant value of β_i throughout P, yields:

$$\mathbb{E}[g(\phi(\alpha_i^N), \bar{\mu}^{N,i}_{\phi(\alpha)}) \mid P] \leq \mathbb{E}[g(\phi(\beta_i), \bar{\mu}^{N,i}_{\phi(\alpha)}) \mid P]$$

multiplying both sides by $\mathbb{P}(P)$ and summing over all $P \in \mathcal{P}_i^N$, we get

$$\mathbb{E}[g(\phi(\alpha_i^N), \bar{\mu}^{N,i}_{\phi(\alpha)})] \leq \mathbb{E}[g(\phi(\beta_i), \bar{\mu}^{N,i}_{\phi(\alpha)})]$$

and so, α^N is a correlated equilibrium. □

Also the converse holds, indeed:

Theorem 1.5. For each N-person game and each correlated equilibrium α^N, there is an information system $(\Omega_N, \mathbb{P}_N, \{\mathcal{P}_i^N\}_i)$ for which it is Bayes rational for the players to play in accordance with α^N.

Proof. We consider $(\Omega_N, \mathbb{P}_N)$ a probability space and $\alpha^N : \Omega_N \to S_N$ correlated equilibrium for 1.2. Let $\mathcal{P}_i$ be the partition generated by α^N, this means that ω and ω' belong to the same $P \in \mathcal{P}_i$ if and only if $\alpha_i^N(\omega) = \alpha_i^N(\omega')$.
We know that, for every $d : S \to S$:

$$\mathbb{E}[g(\phi(\alpha_i^N), \bar{\mu}^{N,i}_{\phi(\alpha)})] \le \mathbb{E}[g(\phi(d(\alpha_i^N)), \bar{\mu}^{N,i}_{\phi(\alpha)})]$$

thus, if there exists a $\tilde{s}_i \in S$ and a function $d : S \to S$ such that

$$\mathbb{E}[g(\phi(\alpha_i^N), \bar{\mu}^{N,i}_{\phi(\alpha)}) \mid \alpha_i^N = \tilde{s}_i] > \mathbb{E}[g(\phi(d(\alpha_i^N)), \bar{\mu}^{N,i}_{\phi(\alpha)}) \mid \alpha_i^N = \tilde{s}_i]$$

we define $h : S \to S$ as

$$h(\tilde{s}_i) = d(\tilde{s}_i)$$
$$h(s_i) = s_i \ \forall\, s_i \neq \tilde{s}_i$$

hence

$$\begin{aligned}
&\mathbb{E}[g(\phi(h(\alpha_i^N)), \bar{\mu}^{N,i}_{\phi(\alpha)})] = \sum_{s_i \in S} p_{\alpha_i^N}(s_i)\mathbb{E}[g(\phi(h(\alpha_i^N)), \bar{\mu}^{N,i}_{\phi(\alpha)}) \mid \alpha_i^N = s_i] \\
&= \sum_{s_i \neq \tilde{s}_i} p_{\alpha_i^N}(s_i)\mathbb{E}[g(\phi(h(s_i)), \bar{\mu}^{N,i}_{\phi(\alpha)}) \mid \alpha_i^N = s_i] + p_{\alpha_i^N}(\tilde{s}_i)\mathbb{E}[g(\phi(h(\tilde{s}_i)), \bar{\mu}^{N,i}_{\phi(\alpha)}) \mid \alpha_i^N = \tilde{s}_i] \\
&= \sum_{s_i \neq \tilde{s}_i} p_{\alpha_i^N}(s_i)\mathbb{E}[g(\phi(s_i), \bar{\mu}^{N,i}_{\phi(\alpha)}) \mid \alpha_i^N = s_i] + p_{\alpha_i^N}(\tilde{s}_i)\mathbb{E}[g(\phi(d(\tilde{s}_i)), \bar{\mu}^{N,i}_{\phi(\alpha)}) \mid \alpha_i^N = \tilde{s}_i] \\
&< \sum_{s_i \neq \tilde{s}_i} p_{\alpha_i^N}(s_i)\mathbb{E}[g(\phi(s_i), \bar{\mu}^{N,i}_{\phi(\alpha)}) \mid \alpha_i^N = s_i] + p_{\alpha_i^N}(\tilde{s}_i)\mathbb{E}[g(\phi(\tilde{s}_i), \bar{\mu}^{N,i}_{\phi(\alpha)}) \mid \alpha_i^N = \tilde{s}_i] \\
&= \mathbb{E}[g(\phi(\alpha_i^N), \bar{\mu}^{N,i}_{\phi(\alpha)})]
\end{aligned}$$

which contradicts the hypothesis.
Observe that since $\mathcal{P}_i^N$ is the partition generated by α_i^N, then$\{\alpha_i^N = \tilde{s}_i\} = P \in \mathcal{P}_i^N$ and also the vice-versa holds, this means that for every $P \in \mathcal{P}_i^N$:

$$\mathbb{E}[g(\phi(\alpha_i^N), \bar{\mu}^{N,i}_{\phi(\alpha)}) \mid P] \le \mathbb{E}[g(\phi(d(\alpha_i^N)), \bar{\mu}^{N,i}_{\phi(\alpha)}) \mid P]$$

which means that, since $d(\alpha_i^N)$ is constant in P and say it s_i, for every $P \in \mathcal{P}_i^N$ and for every $s_i \in S$

$$\mathbb{E}[g(\phi(\alpha_i^N), \bar{\mu}^{N,i}_{\phi(\alpha)}) \mid P] \le \mathbb{E}[g(\phi(s_i), \bar{\mu}^{N,i}_{\phi(\alpha)}) \mid P]$$

□

We have thus proved that there is a bijective relation between the correlated equilibria and the Bayes rationality of the information system. In the following we'll use predominantly definition 1.2 because it is operatively more convenient for our purposes but the Bayes rationality could give a more intuitive meaning of this type of equilibrium.

Remark 1.6. We observe that if α^N is a Nash equilibrium, then it is also a correlated one, indeed let's consider p_{α^N} its distribution and $d : S \to S$ a function, then $\beta^N := (d(\alpha_1^N), \dots, d(\alpha_N^N))$ is a mixed strategy, so:

$$\begin{aligned}
&\sum_{s\in S_N} p_{\alpha^N}(s) g\Big(\phi(s_i), \frac{1}{N-1}\sum_{j\neq i}\delta_{\phi(s_j)}\Big) = J_i^N(\alpha^N) \leq J_i^N(\alpha_{-i}^N, \beta_i^N) = \\
&= \int_{\Omega_N} g\Big(\phi(\beta_i^N(\omega)), \frac{1}{N-1}\sum_{j\neq i}\delta_{\phi(\alpha_j^N(\omega))}\Big)\mathbb{P}_N(d\omega) \\
&= \int_{\Omega_N} g\Big(\phi(d(\alpha_i^N(\omega)), \frac{1}{N-1}\sum_{j\neq i}\delta_{\phi(\alpha_j^N(\omega))}\Big)\mathbb{P}_N(d\omega) \\
&= \sum_{s\in S_N} p_{\alpha^N}(s) g\Big(\phi(d(s_i)), \frac{1}{N-1}\sum_{j\neq i}\delta_{\phi(s_j)}\Big).
\end{aligned}$$

Furthermore we have:

Proposition 1.7. The set of correlated equilibria is convex and compact.

Proof. In this proof we'll use the Definition 1.2 to characterize correlated equilibria. Denote with $\Delta_N := \{x \in \mathbb{R}^N : x_i \geq 0\ \forall i, \sum_{i=1}^N x_i = 1\}$ the simplex in $\mathbb{R}^N$, then the set of correlated equilibria, $\operatorname{corr}(S_N)$ is a subset of Δ_{dN}.
Take $p, \tilde{p} \in \operatorname{corr}(S_N)$, $\lambda \in [0,1]$, then for all i and $d : S \to S$:

$$\begin{aligned}
&\sum_{s\in S_N} (\lambda p(s) + (1-\lambda)\tilde{p}(s)) g\Big(\phi(s_i), \frac{1}{N-1}\sum_{j\neq i}\delta_{\phi(s_j)}\Big) \\
&= \lambda \sum_{s\in S_N} p(s) g\Big(\phi(s_i), \frac{1}{N-1}\sum_{j\neq i}\delta_{\phi(s_j)}\Big) + (1-\lambda)\sum_{s\in S_N} \tilde{p}(s) g\Big(\phi(s_i), \frac{1}{N-1}\sum_{j\neq i}\delta_{\phi(s_j)}\Big) \\
&\leq \lambda \sum_{s\in S_N} p(s) g\Big(\phi(d(s_i)), \frac{1}{N-1}\sum_{j\neq i}\delta_{\phi(s_j)}\Big) + (1-\lambda)\sum_{s\in S_N} \tilde{p}(s) g\Big(\phi(d(s_i)), \frac{1}{N-1}\sum_{j\neq i}\delta_{\phi(s_j)}\Big) \\
&= \sum_{s\in S_N} (\lambda p(s) + (1-\lambda)\tilde{p}(s)) g\Big(\phi(d(s_i)), \frac{1}{N-1}\sum_{j\neq i}\delta_{\phi(s_j)}\Big)
\end{aligned}$$

and so $\lambda p(s) + (1-\lambda)\tilde{p}(s) \in \operatorname{corr}(S)$.
Moreover, if we write condition (1.5) as:

$$\sum_{s\in S_N} p(s)\Big(g\Big(\phi(s_i), \frac{1}{N-1}\sum_{j\neq i}\delta_{\phi(s_j)} - g\Big(\phi(d(s_i)), \frac{1}{N-1}\sum_{j\neq i}\delta_{\phi(s_j)}\Big)\Big) \leq 0 \quad \forall\, i\ \forall\, d : S \to S$$

we see that $\operatorname{corr}(S_N)$ is a closed subset of Δ_{dN} which is compact; which implies that $\operatorname{corr}(S_N)$ is compact. □

Remark 1.8. Remark 1.6, together with Proposition 1.7, says that any convex composition of Nash equilibria is a correlated equilibrium. The vice-versa is not true in general, indeed if we consider the simple two person game, called "chicken" whose costs matrix is:

	D	T
D	-6,-6	-2,-7
T	-7,-2	0,0

In this game there are 3 Nash equilibria, which are (D,T), (T,D) and ([$\frac{2}{3}$T,$\frac{1}{3}$D], [$\frac{2}{3}$T,$\frac{1}{3}$D]) with payoff respectively (-2,-7), (-7,-2), ($-\frac{14}{3}$, $-\frac{14}{3}$). Their distributions as correlated strategies are:

	D	T
D	0	1
T	0	0

	D	T
D	0	0
T	1	0

	D	T
D	$\frac{4}{9}$	$\frac{2}{9}$
T	$\frac{2}{9}$	$\frac{1}{9}$

But we have that:

	D	T
D	$\frac{1}{3}$	$\frac{1}{3}$
T	$\frac{1}{3}$	0

is a correlated equilibrium whose payoff is (-5,-5) and it's outside the convex hull of Nash equilibria.

1.2 Existence of Nash and correlated Equilibria

In this section, we'll prove the existence of the two types of equilibria, as a matter of fact, one will follow from the other. In the following, we'll indicate by Δ_n the n-dimensional simplex.

Theorem 1.9 (Nash)**.** Every game with a finite number of players and where the set of the pure strategies is finite admits a Nash equilibrium in mixed strategies.

Proof. In this proof we'll use the equivalent definition of J_i^N as function of the probability distribution over S_N, that is, in the case of mixed strategies:

$$J_i^N(p) = \sum_{s\in S_N} p(s)g\Big(\phi(s_i), \frac{1}{N-1}\sum_{j\neq i}\delta_{\phi(s_j)}\Big) = \sum_{s\in S_N}\Big(\prod_{i=1}^{N} p_i(s_i)\Big)g\Big(\phi(s_i), \frac{1}{N-1}\sum_{j\neq i}\delta_{\phi(s_j)}\Big)$$

First of all we indicate with Σ the set of the mixed strategies over S ($\Rightarrow \Sigma = \Delta_d$) and $\Sigma_N := \times_N \Sigma$ ($\Rightarrow \Sigma \subset \Delta_{dN}$); we define player i's *reaction correspondence*, r_i as follows:

$$\begin{aligned} &r_i : \Sigma_N \to \Sigma \\ &r_i(p) = \arg\min_{\tilde{p}_i\in\Sigma}\{J_i^N(p_{-i}, \tilde{p}_i)\} \end{aligned} \tag{1.7}$$

This means that r_i is the set function which maps each strategy profile p to the set of mixed strategies that minimizes player i's cost function when his opponents plays p_{-i}. Now we take r as the Cartesian product of all r_i, $r = (r_1, \dots, r_N)$. So r is a set function $r : \Sigma_N \rightrightarrows \Sigma_N$. To prove Nash Theorem means to prove that r as a fixed point, that is a $p \in r(p)$; this could be done using Kakutani's theorem:

Theorem 1.10 (Kakutani)**.** Let S be a non-empty, compact and convex subset of some Euclidean space $\mathbb{R}^n$. Let $\phi : S \rightrightarrows S$ be a set-valued function on S with a closed graph and the property that $\phi(x)$ is non-empty and convex for all $x \in S$. Then ϕ has a fixed point.

We recall that to have a closed graph for r means that if $(p^n, \hat{p}^n) \to (p, \hat{p})$ and $\hat{p}^n \in r(p^n)$ then $\hat{p} \in r(p)$.

- Σ is convex and compact since it's the simplex in $\mathbb{R}^d$. So Σ_N is compact and convex since it's a finite product of compact and convex sets.

- $r(p)$ is non-empty for all p because the cost function is continuous (because linear with respect to mixed strategies) and Σ is compact so J_i^N attains its minimum for all i.

- $r(p)$ convex for all p: take $p', p'' \in r(p)$, that means $J_i^N(p_{-i}, p_i') = J_i^N(p_{-i}, p_i'') \leq J_i^N(p_{-i}, \tilde{p}_i)$ for all $\tilde{p} \in \Sigma$. Take $\lambda \in [0,1]$. For linearity of the cost function function:

$$\begin{aligned} &J_i^N(p_{-i}, \lambda p_i' + (1-\lambda)p_i'') = \lambda J_i^N(p_{-i}, p_i') + (1-\lambda) J_i^N(p_{-i}, p_i'') \\ =&\lambda J_i^N(p_{-i}, p_i') + (1-\lambda) J_i^N(p_{-i}, p_i') = J_i^N(p_{-i}, p_i') \leq J_i^N(p_{-i}, \tilde{p}_i) \text{ for all } \tilde{p} \in \Sigma \end{aligned}$$

  So also any convex combination of two best responses is a best response.

- $r(p)$ has a closed graph: assume that this condition is violated, so there is a sequence $(p^n, \hat{p}^n) \to (p, \hat{p})$ and $\hat{p}^n \in r(p^n)$ but $\hat{p} \notin r(p)$; this implies that there exists a i such that $\hat{p}_i \notin r_i(p)$. Let $p_i' \in r_i(p)$ (notice that such an p' always exists because we've shown, in the previous point, that $r(p)$ s non-empty for all p), so $\exists \varepsilon > 0$ s.t $J_i^N(p_{-i}, p_i') \leq J_i^N(p_{-i}, \hat{p}_i) - 3\varepsilon$. Since J_i^N is continuous and $(p^n, \hat{p}^n) \to (p, \hat{p})$, for n sufficiently large we have

$$\begin{aligned} J_i^N(p_{-i}^n, p_i') \leq J_i^N(p_{-i}, p_i') + \varepsilon \leq J_i^N(p_{-i}, \hat{p}_i) - 3\varepsilon + \varepsilon \leq& J_i^N(p_{-i}^n, \hat{p}_i^n) + \varepsilon - 2\varepsilon \\ =& J_i^N(p_{-i}^n, \hat{p}_i^n) - \varepsilon \end{aligned}$$

  Which means that p_i' is a strictly better response than $\hat{p}_i^n$ which contradicts the hypothesis $\hat{p}^n \in r(p^n)$.

We've verified all the hypothesis of Kakutani's theorem, so we've concluded the proof. □

We are interested in a particular type of game on which our specific model could be framed. On that we'll also show the existence of Nash equilibria:

Definition 1.11. Consider a game (S_i, J_i^N) where S_i is the i's strategy set and $J_i^N : S \to \mathbb{R}$ is the respective cost function. It is called a **symmetric game** if $S_1 = \ldots = S_N$ and if the cost functions are invariant for permutations, namely

$$J_i^N(\alpha_1^N, \ldots, \alpha_i^N, \ldots, \alpha_N^N) = J_{\pi(i)}^N(\alpha_{\pi(1)}^N, \ldots, \alpha_{\pi(i)}^N, \ldots, \alpha_{\pi(N)}^N) \quad \text{for any permutation } \pi.$$

It is now clear that our model fits Definition 1.11, so the following results will be very helpful.

Corollary 1.12. Every symmetric game with a finite number of players, where the set of strategies is finite, admits a symmetric Nash equilibrium.

Proof. We take r as in the definition of the proof of theorem 1.9 and we restrict his domain to

$$\tilde{\Sigma}_N := \left\{ \mathbf{p} = (p, \dots, p) \in \Sigma_N : p \in \Sigma \right\}$$

Due to the symmetry of the definition, $J_1^N(\tilde{p}, p, \dots, p) = \dots = J_i^N(p, \dots, p, \tilde{p}, p, \dots, p) = \dots = J_N^N(p, \dots, p, \tilde{p})$. Thus, if $\tilde{p} \in r_1(\mathbf{p})$ we have $J_1^N(\tilde{p}, p, \dots, p) \leq J_1^N(p', p, \dots, p)\ \forall p' \in \Sigma$ and so $J_i^N(p, \dots, p, \tilde{p}, p, \dots, p) \leq J_i^N(p, \dots, p, p', p, \dots, p)$ for every $p' \in \Sigma$ and for every i. This means that if $\tilde{p} \in r_1(\mathbf{p})$, $\tilde{p} \in r_i(\mathbf{p})$ for every i. Define:

$$\tilde{r}_1 : \Sigma \to \Sigma$$
$$\tilde{r}_1(p) := r_1(\mathbf{p})$$

We can now apply Kakutani's theorem to $\tilde{r}_1$, whose hypothesis are verified in the proof of the Theorem 1.9 also for this case. Thus, there exists a $p \in \Sigma$ such that $p \in \tilde{r}_1(p) = r_1(\mathbf{p})$, which means that $p \in \arg\min_{\tilde{p}} J_i^N(\mathbf{p}_{-i}, \tilde{p})$ and, because of the symmetry, $p \in \tilde{r}_i(p) := r_i(\mathbf{p})$ for all i. So there exists a $p \in \Sigma$ such that $\mathbf{p} \in r(\mathbf{p})$. □

Thanks to Remark 1.6, which tells us that every Nash equilibrium is also a correlated one, we have also proved the existence of correlated equilibria in the case of finite games.

1.3 Correlated vs Nash equilibria

In this section we are going to give three examples based on the same model that justify our interest to correlated equilibria. The first one allows us to understand the idea but it doesn't fit perfectly our setting, the second will increase fit the model but it will not be interesting when $N \to \infty$; in the last one we will add some difficulties but it will explain why what we are going to do in Chapter 2 might have an interest.

1.3.1 Example 1

We present a situation with N payers in which correlated strategies give a better result than mixed one, in the sense that the sum of the costs for a correlated equilibrium is greater that that of all the Nash's. This example is inspired from the one in Theorem 11 in [1]. We study a particular *simple congestion game*, in which there are N players who could choose between 2 strategies, 0 and 1 ($\Rightarrow S = \{0, 1\}$) and the private states coincide with the actions. The idea is that is the minimal cost is reached only if one player chooses differently from the others. Thus, we define $w_0^N(1 + \frac{j-1}{N-1})$, $j \in \{1, \dots, N\}$ the cost of each player who chooses 0 if exactly j players choose 0 and similarly $w_1^N(1 + \frac{j-1}{N-1})$. We define them as:

$$\begin{aligned}
&w_0^N(1) = w_1(1) = -1\\
&w_0^N(1+\frac{1}{N-1}) = \ldots = w_0^N(1+\frac{N-3}{N-1}) = w_1^N(1+\frac{1}{N-1}) = \ldots = w_1^N(1+\frac{N-3}{N-1}) = 0\\
&w_0^N(1+\frac{N-2}{N-1}) = w_1^N(1+\frac{N-2}{N-1}) = \frac{1}{N}\\
&w_0^N(2) = w_1^N(2) = 1
\end{aligned} \tag{1.8}$$

We rewrite the (1.8) as $w_0 = w_1 = (-1, 0, \ldots, 0, \frac{1}{N}, 1)$. With the notation used among the general theory, we have

$$g^N(s_i, \bar{\mu}_s^{N,i}) = w_{s_i}^N(1 + \bar{\mu}_s^{N,i}(\{s_i\}))$$

Observe that g is bounded.
Call $\pi = (1, N-1)$ the situation in which one player chooses 0 (resp, 1) and the others $N-1$ players choose 1 (resp. 0). Call $A_\pi \subseteq S_N$ the set all those strategies. We observe that:

- $|A_\pi| = 2N$
- each $s \in A_\pi$ gives the minimal global cost (meant as the sum of all the single costs). Indeed the possible situations are:
  - Every player chooses the same strategy (0 or 1) $\Rightarrow$ global cost $= N$
  - $2 \leq i \leq N-2$ players choose 0 (resp. 1) and the $N-i$ others choose 1 (resp.0) $\Rightarrow$ global cost $= 0$
  - 1 player chooses 0 (resp. 1) and the $N-1$ others choose 1 (resp. 1) $\Rightarrow$ global cost $= -1 + \frac{N-1}{N} = -\frac{1}{N} < \max\{0, N\}$

Consider the correlated strategy p having a uniform distribution over A_π so it assigns probability $\frac{1}{2N}$ to each $s \in A_\pi$, so

$$p(s) = \frac{1}{2N} \text{ iff } s \in A_\pi.$$

Then for all $i = 1, \ldots, N$

$$J_i^N(p) = \frac{1}{2N}(2 \cdot (-1) + 2\frac{1}{N}(N-1)) = -\frac{1}{N^2}.$$

So, the global cost is

$$J_{\text{corr}}^N(p) = \sum_i \frac{1}{N^2} = -\frac{1}{N^3}$$

Let's show that α is a correlated equilibrium by showing that relation (1.5) is verified. Because of the symmetry of the game, it's sufficient to verify it for a player i. There are four function $d : S \to S$: the identity, $d \equiv 0$, $d \equiv 1$ and that one such that $d(0) = 1$ and $d(1) = 0$. Thus, we just need to verify (1.5) for such a function d. There are two possible situations:

- if $s \in A_\pi$ is such that player i chooses 0 (resp. 1) and the others choose 1 (resp. 0) - there are only two $s \in A_\pi$ which verifies this condition - , than, if $d(s_i) = d(0) = 1$ (resp. $d(s_i) = d(1) = 0$), every player choose 1 (resp. 0) and everyone gets 1 as cost.
- if $s \in A_\pi$ is such that player i and other $N-1$ players choose 0 (resp. 1) and only one (different from i) chooses 1 (resp. 0) - there are $2N-2$ strategies $s \in A_\pi$ which verify this condition - , if $d(s_i) = d(0) = 1$ (resp. $d(s_i) = d(1) = 0$), everyone gets 0 as cost.

So, in the case of $d \equiv 0$, $d \equiv 1$, due to the symmetry of the game, the cost, indicated by $J_i^N(d(p))$, is

$$J_i^N(d(p)) = \frac{1}{2N}\left(-1 + \frac{1}{N}(N-1) + 1 + (N-1)\cdot 0\right) = 1 - \frac{1}{N}$$

and in the case of $d(0) = 1$ and $d(1) = 0$

$$J_i^N(d(p)) = \frac{1}{2N}(2 \cdot 1 + 2(N-1)\cdot 0) = \frac{1}{N}$$

We have:

$$J_i^N(p) - \min_d \{J_i^N(p_{-i}, d(p_i))\} = -\frac{1}{N^2} - \frac{1}{N} \leq 0.$$

So p is a correlated equilibrium. If we show that no Nash equilibrium can attain $J_{\text{corr}}^N(p)$, we would have finished.

First of all we observe that any strategy profile in A_π isn't an equilibrium, since every player who has chosen the strategy chosen by the other $N-2$ wish to deviate, because he/she would get a cost of 0 instead of $\frac{1}{N}$. The only Nash equilibria are those in which $2 \leq k \leq N-2$ players choose the same action, which means that the global cost is $0 < J_{\text{corr}}^N(p)$.

We observe the only strategies who attain the minimal global cost are those in A_π. So we want to prove that if at least one player (say i) plays a mixed strategy α, with distribution $p = (p^1, \dots, p^N)$, in which $p^i(0) > 0$, $p^i(1) > 0$, then there exists a profile $s \notin A_\pi$ such that $p(s) > 0$. Assume that $p(s) > 0 \Rightarrow s \in A_\pi$. Let s be such a strategy, then

$$p(s) > 0 \Rightarrow \prod_j p^j(s_j) > 0 \Rightarrow p^j(s_j) > 0 \text{ for all } j = 1, \dots, N$$

Since $p^i(0), p^i(1) > 0$, $p(s_{-i}, 0) = p^i(0)\prod_{j\neq i} p^j(s_j) > 0$ and $p(s_{-i}, 1) = p^i(1)\prod_{j\neq i} p^j(s_j) > 0$. So both $(s_{-i}, 0)$ and $(s_{-i}, 1)$ are in A_π, which is impossible since we've seen that if a player deviates from a strategy in A_π, we obtain a strategy not in A_π.

The problem with this example is the dependence of g from N, situation which is not included in our model.

In the following sections, we'll try to formulate some examples which improves this one keeping its structure.

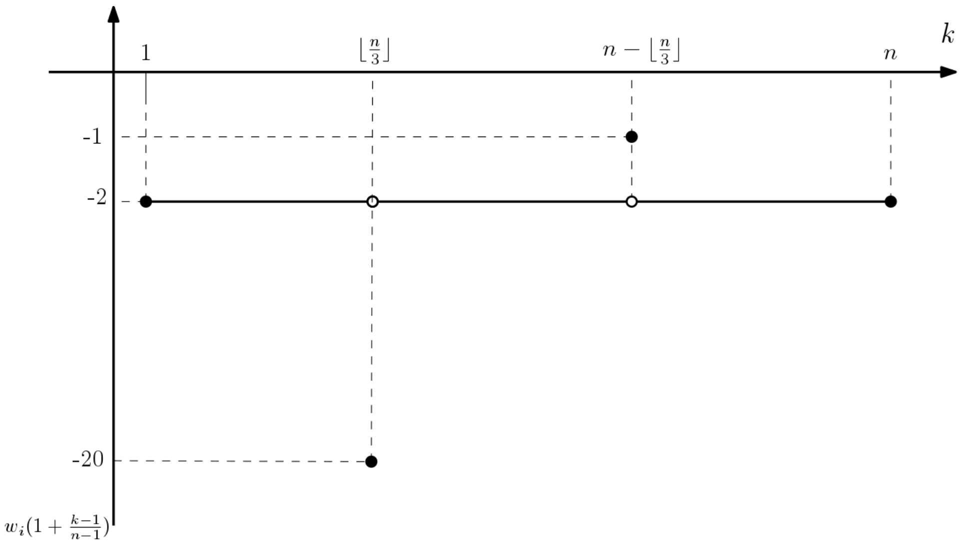
1.3.2 Example 2

Also in this example there are n players who could choose between 2 strategies, 0 and 1 ($\Rightarrow S = \{0,1\}$). We define $w_0(1+\frac{j-1}{n-1})$ $j \in \{1,\dots,n\}$ the cost of each player who chooses 0 if exactly j players choose 0 (in pure strategies) and similarly we define $w_1(1+\frac{j-1}{n-1})$. For simplicity we'll consider the n-players game when $n \geq 4$ and we recall that $\lfloor x \rfloor := \max\{m \in \mathbb{Z} : m \leq x\}$, it's such that

$$x - 1 \leq \lfloor x \rfloor \leq x. \tag{1.9}$$

We consider a game in which:

$$\begin{aligned} &w_0\Big(1+\frac{\lfloor \frac{n}{3} \rfloor - 1}{n-1}\Big) = w_1\Big(1+\frac{\lfloor \frac{n}{3} \rfloor - 1}{n-1}\Big) = -20 \\ &w_0\Big(1+\frac{n-1-\lfloor \frac{n}{3} \rfloor}{n-1}\Big) = w_1\Big(1+\frac{n-1-\lfloor \frac{n}{3} \rfloor}{n-1}\Big) = -1 \\ &w_0\Big(1+\frac{k-1}{n-1}\Big) = w_1\Big(1+\frac{k-1}{n-1}\Big) = -2 \quad \text{for } k \neq \{\Big\lfloor \frac{n}{3} \Big\rfloor, n - \Big\lfloor \frac{n}{3} \Big\rfloor\} \end{aligned} \tag{1.10}$$



Observe that:

$$g(s_i, \bar{\mu}_s^{n,i}) = w_{s_i}(1 + \bar{\mu}_s^{n,i}(\{s_i\}))$$

where g is no more dependent from n and it is bounded.
This means that if $\lfloor \frac{n}{3} \rfloor$ players (i.e. about one third of all players) choose 0 (resp. 1) they get -20 and the others $n - \lfloor \frac{n}{3} \rfloor$ get -1; in any other situation everybody gets -2. Thus, these two situations correspond to only two possible global costs:

- exactly $\lfloor \frac{n}{3} \rfloor$ players choose the same strategy:

$$J^n = -20\Big\lfloor \frac{n}{3} \Big\rfloor - 1(n - \Big\lfloor \frac{n}{3} \Big\rfloor) = -19\Big\lfloor \frac{n}{3} \Big\rfloor - n$$

- all the other situations

$$J'^n = -2n$$

$$J^n \overset{?}{\leq} J'^n \Leftrightarrow -19\left\lfloor\frac{n}{3}\right\rfloor - n \overset{?}{\leq} -2n \Leftrightarrow 19\left\lfloor\frac{n}{3}\right\rfloor - n \overset{?}{\geq} 0.$$

Using (1.9):

$$-19\left\lfloor\frac{n}{3}\right\rfloor - n \geq 19(\frac{n}{3} - 1) - n = \frac{19}{3}n - 19 - n = \frac{16}{3}n - 19 \overset{?}{\geq} 0 \Leftrightarrow 16n \overset{?}{\geq} 57$$

Since we've assumes $n \geq 4$ this is always true. This means that the pure strategies for which we get the minimal cost are those in which $\lfloor\frac{n}{3}\rfloor$ players choose 0 (resp. 1) and the other choose 1 (resp. 0). We'll denote with A the set of these strategies.
We now consider the correlated strategy p with uniform distribution over A.
Before proceeding further, we do some calculations:

1. the number of possible $\lfloor\frac{n}{3}\rfloor$ in a total of n are $\binom{n}{\lfloor\frac{n}{3}\rfloor} \Rightarrow |A| = 2\binom{n}{\lfloor\frac{n}{3}\rfloor}$

2. if we consider a generic player i, the number of sets of $\lfloor\frac{n}{3}\rfloor$ players in which he/she is contained is $\binom{n-1}{\lfloor\frac{n}{3}\rfloor-1}$

3. $$\binom{n}{\lfloor\frac{n}{3}\rfloor} - \binom{n-1}{\lfloor\frac{n}{3}\rfloor - 1} = \frac{n!}{\lfloor\frac{n}{3}\rfloor!(n - \lfloor\frac{n}{3}\rfloor)!} - \frac{(n-1)!}{(\lfloor\frac{n}{3}\rfloor - 1)!(n - 1 - \lfloor\frac{n}{3}\rfloor + 1)!}$$
$$= \frac{(n-1)!}{(\lfloor\frac{n}{3}\rfloor - 1)!(n - \lfloor\frac{n}{3}\rfloor)!}\left(\frac{n}{\lfloor\frac{n}{3}\rfloor} - 1\right) = \binom{n-1}{\lfloor\frac{n}{3}\rfloor - 1}\left(\frac{n}{\lfloor\frac{n}{3}\rfloor} - 1\right)$$

4. $$\frac{\binom{n}{\lfloor\frac{n}{3}\rfloor}}{\binom{n-1}{\lfloor\frac{n}{3}\rfloor-1}} = \frac{n!}{(n-1)!}\frac{(\lfloor\frac{n}{3}\rfloor - 1)!(n - \lfloor\frac{n}{3}\rfloor)!}{\lfloor\frac{n}{3}\rfloor!(n - \lfloor\frac{n}{3}\rfloor)!} = \frac{n}{\lfloor\frac{n}{3}\rfloor}$$

Now, because of the symmetry of the game, we can consider a generic player i and the calculations we'll be the same for all the other players. Thus:

$$J_i^n(p) \overset{1.+2.}{=} \frac{1}{2\binom{n}{\lfloor\frac{n}{3}\rfloor}}\left(2\binom{n-1}{\lfloor\frac{n}{3}\rfloor}(-20) + 2\left(\binom{n}{\lfloor\frac{n}{3}\rfloor} - \binom{n-1}{\lfloor\frac{n}{3}\rfloor - 1}\right)(-1)\right)$$
$$\overset{3.}{=} \frac{1}{\binom{n}{\lfloor\frac{n}{3}\rfloor}}\binom{n-1}{\lfloor\frac{n}{3}\rfloor - 1}\left(-20 - \frac{n}{\lfloor\frac{n}{3}\rfloor} + 1\right) \overset{4.}{=} \frac{\lfloor\frac{n}{3}\rfloor}{n}\left(-19 - \frac{n}{\lfloor\frac{n}{3}\rfloor}\right) = -\left(19\frac{\lfloor\frac{n}{3}\rfloor}{n} + 1\right)$$

Remark 1.13. Because of (1.9), we have:

$$\frac{\lfloor\frac{n}{3}\rfloor}{n} \leq \frac{\frac{n}{3}}{n} = \frac{1}{3}$$
$$\frac{\lfloor\frac{n}{3}\rfloor}{n} \geq \frac{\frac{n}{3} - 1}{n} = \frac{1}{3} - \frac{1}{n}$$
$$\Rightarrow -\frac{19}{3} - 1 = -\frac{22}{3} \leq J_i^n(p) \leq -19\left(\frac{1}{3} - \frac{1}{n} - 1\right) = -\frac{22}{3} + \frac{19}{n}$$

We now show that p defined as above is a correlated equilibrium; to do so recall that A probability distribution p over S_N is said correlated equilibrium if, for every i and for every function $d : S \to S$, we have

$$\sum_{s\in S_N} p(s)g\Big(\phi(s_i), \frac{1}{N-1}\sum_{j\neq i}\delta_{\phi(s_j)}\Big) \leq \sum_{s\in S_N} p(s)g\Big(\phi(d(s_i)), \frac{1}{N-1}\sum_{j\neq i}\delta_{\phi(s_j)}\Big) \qquad (1.11)$$

In this case, since $S = \{0, 1\}$ there are only four possible d, that are $d = id_S$, $d \equiv 0, d \equiv 1$ or $d(0) = 1, d(1) = 0$. In the first case, we have no modifications; in the last case:

$$g(d(s_i), \bar{\mu}_s^{n,i}) = \begin{cases} g(0, \bar{\mu}_s^{n,i}) & \text{if } s_i = 1 \\ g(1, \bar{\mu}_s^{n,i}) & \text{if } s_i = 0 \end{cases}$$

By symmetry we can consider just one of these two cases. Both in the situation in which the i-th player had chosen 0 with other $\lfloor \frac{n}{3} \rfloor - 1$ and in that in which he's in the remaining $n - \lfloor \frac{n}{3} \rfloor$, his deviation brings everybody to have cost -2. If we indicate with $J_i^n(d(p)) := \sum_{s\in S_N} p(s)g\Big(\phi(d(s_i)), \frac{1}{N-1}\sum_{j\neq i}\delta_{\phi(s_j)}\Big)$, we have:

$$\begin{aligned} J_i^n(d(p)) &= \frac{1}{2\binom{n}{\lfloor \frac{n}{3} \rfloor}}\binom{n-1}{\lfloor \frac{n}{3} \rfloor - 1}\Big(-2\cdot 2 - 2\cdot 2\Big(\frac{n}{\lfloor \frac{n}{3} \rfloor} - 1\Big)\Big) \\ &= \frac{\lfloor \frac{n}{3} \rfloor}{n}\Big(-2\frac{n}{\lfloor \frac{n}{3} \rfloor}\Big) = -2 \end{aligned}$$

In the case $d \equiv 0$, $d \equiv 1$:

$$\begin{aligned} J_i^n(d(p)) &= \frac{1}{2\binom{n}{\lfloor \frac{n}{3} \rfloor}}\binom{n-1}{\lfloor \frac{n}{3} \rfloor - 1}\Big(-20 - \Big(\frac{n}{\lfloor \frac{n}{3} \rfloor} - 1\Big) - 2 - 2\Big(\frac{n}{\lfloor \frac{n}{3} \rfloor} - 1\Big)\Big) \\ &= \frac{\lfloor \frac{n}{3} \rfloor}{2n}\Big(-19 - 3\frac{n}{\lfloor \frac{n}{3} \rfloor}\Big) = -\frac{1}{2}\Big(19\frac{\lfloor \frac{n}{3} \rfloor}{n} + 1 + 2\Big) \end{aligned}$$

$$\begin{aligned} 0 &\overset{?}{\geq} J_i^n(p) + 2 = -19\frac{\lfloor \frac{n}{3} \rfloor}{n} - 1 + 2 = -19\frac{\lfloor \frac{n}{3} \rfloor}{n} + 1 \\ &-19\frac{\lfloor \frac{n}{3} \rfloor}{n} + 1 \leq -\frac{19}{3} + \frac{19}{n} + 1 = -\frac{16}{3} + \frac{19}{n} \overset{?}{\leq} 0 \Leftrightarrow 16n \geq 57 \end{aligned}$$

since we have assumed $n \geq 4$, this is always true. This means that p is a correlated equilibrium.

If we show that no Nash equilibrium can attain the global cost of the correlated equilibrium, we would have finished.

First of all we observe that any strategy profile in A isn't a pure Nash equilibrium, since every player who has chosen the strategy chosen by $n - \binom{n}{\lfloor \frac{n}{3} \rfloor}$ wish to deviate, because he/she would get a cost of -2 instead of -1. The only pure Nash equilibria are those in which $1 \leq k \leq \lfloor \frac{n}{3} \rfloor - 2$ or $\lfloor \frac{n}{3} \rfloor + 2 \leq k \leq n - \lfloor \frac{n}{3} \rfloor - 2$ or $k \geq n - \lfloor \frac{n}{3} \rfloor + 2$ choose the same action. The cost is -2 for each player, but we have already proved that $J_i^n(p) + 2 < 0$.

We observe the only strategies who attain the minimal global cost are those in A. So we want to prove that if at least one player (say i) plays a mixed strategy α, with distribution

$q = (q^1, \dots, q^N)$, in which $q^i(0) > 0$, $q^i(1) > 0$, then there exists a profile $s \notin A$ such that $q(s) > 0$. Assume that $q(s) > 0 \Rightarrow s \in A$. Let s be such a strategy, then

$$q(s) > 0 \Rightarrow \prod_j q^j(s_j) > 0 \Rightarrow q^j(s_j) > 0 \text{ for all } j = 1, \dots, n$$

Since $q^i(0), q^i(1) > 0$, $q(s_{-i}, 0) = q^i(0) \prod_{j \neq i} q^j(s_j) > 0$ and $q(s_{-i}, 1) = q^i(1) \prod_{j \neq i} q^j(s_j) > 0$. So both $(s_{-i}, 0)$ and $(s_{-i}, 1)$ are in A, which is impossible since we've seen that if a player deviates from a strategy in A, we obtain a strategy not in A.
In this second example we've avoided the dependence of g from n, but we observe that the point with minimum cost is "isolated" thus, when $n \to \infty$, its presence becomes irrelevant.

1.3.3 Example 3

In this last example we will eliminate the problems concerning the other two, indeed there will not be any dependence of g upon N and any isolated points. The basic structure is always the same, that is a congestion game where $S = \{0, 1\}$. For sake of simplicity we consider $6n$ players, but it can be done for all n with some difficulties in the notation. The idea is the following: if less then one third and more then a "certain quantity" (which is n in this case) of the players chooses 0 (resp. 1), they get -7 and the remaining gets -1, otherwise everyone has cost -2. So in this type of game we will find a correlated equilibrium, which is not also a Nash one, that has less global cost with respect to pure Nash equilibria and symmetric mixed equilibria. We observe that, since among all the theory of Chapter 3, we study the behaviour of symmetric Nash equilibria and correlated ones when the number of players goes to infinity, this example is sufficient to justify our interest in this type of equilibria.

We now formalize the notation with that used in the rest of the thesis.

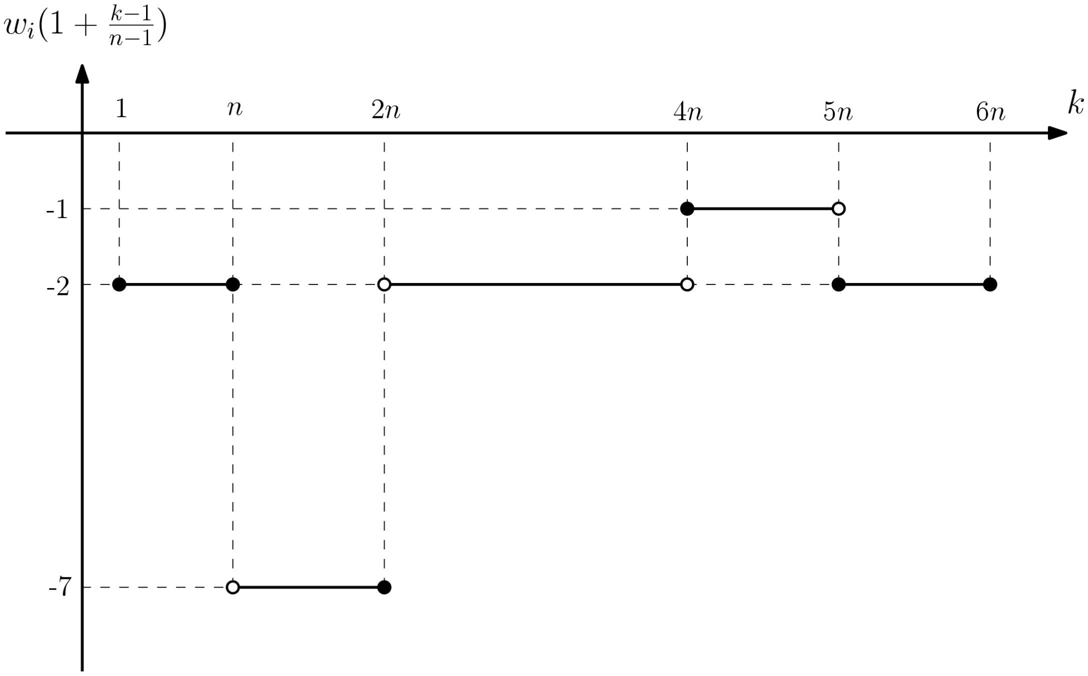
$$w_0\Big(1 + \frac{k-1}{6n-1}\Big) = w_1\Big(1 + \frac{k-1}{6n-1}\Big) = -7 \quad \text{if } k = n+1, \dots, 2n$$
$$w_0\Big(1 + \frac{k-1}{6n-1}\Big) = w_1\Big(1 + \frac{k-1}{6n-1}\Big) = -1 \quad \text{if } k = 4n, \dots, 5n-1$$
$$w_0\Big(1 + \frac{k-1}{6n-1}\Big) = w_1\Big(1 + \frac{k-1}{6n-1}\Big) = -2 \quad \text{if } k = 1, \dots, n,\ k = 5n, \dots, 6n$$

Thus, again:

$$g(s_i, \bar{\mu}_s^{n,i}) = w_{s_i}(1 + \bar{\mu}_s^{n,i}(\{s_i\}))$$

Consider A the set of vector strategies on which at least $n+1$ and at most $2n$ players choose the same action. We notice that $|A| = 2 \sum_{k=n+1}^{2n} \binom{6n}{k}$. We consider in A a correlated strategy $p = (p_k, k = n+1, \dots, 2n)$ depending only on the number of the exact k players choosing the same pure strategy. We will find some sufficient conditions under which p_k

- is a correlated equilibrium
- is strictly better then any pure Nash equilibrium
- is strictly better then any symmetric Nash equilibrium



Before starting to analyse the issues explained above, we observe that p defined as before can't be a non-symmetric mixed strategy, because, since p_k depends only on the number of player choosing a certain action, if we take a random vector having distribution p, it would be exchangeable and hence symmetric as a mixed strategy.

We do some preliminary observations and calculations:

1.

$$\binom{6n-1}{k-1}+\binom{6n-1}{k} = \frac{(6n-1)!}{(k-1)!(6n-k)!}+\frac{(18n-1)!}{k!(18n-k-1)!}$$
$$= \frac{(6n-1)!}{k!(6n-k)!}(6n-k+k) = \frac{(6n)!}{k!(6n-k)!} = \binom{6n}{k}$$

2.

$$\sum_{k=4n}^{5n-1}\binom{6n-1}{k-1}p_{6n-k} = \sum_{k=4n}^{5n-1}\binom{6n-1}{6n-k}p_{6n-k} = \sum_{k=n+1}^{2n}\binom{6n-1}{k}p_k \tag{1.12}$$

3. $2\sum_{k=n+1}^{2n}\binom{6n}{k}p_k = 1$ and, by 1.:

$$2\sum_{k=n+1}^{2n}\binom{6n-1}{k-1}p_k + 2\sum_{k=n+1}^{2n}\binom{6n-1}{k}p_k = 1 \tag{1.13}$$

By the symmetry of the game, we will refer to Player 1. We now calculate the cost of the correlated strategy p for this player. His/her possible situations are: 1. he has chosen 0 (resp. 1) with other $k = n, \dots, 2n-1$ players; 2. he has chosen 0 (resp. 1) with other $k = 4n-1, \dots, 5n-2$ players, thus the cost is:

$$J_1^n(p) = -14 \sum_{k=n+1}^{2n} \binom{6n-1}{k-1} p_k - 2 \sum_{k=4n}^{5n-1} \binom{6n-1}{k-1} p_{6n-k}$$
$$\overset{(1.12)}{=} -14 \sum_{k=n+1}^{2n} \binom{6n-1}{k-1} p_k - 2 \sum_{k=n+1}^{2n} \binom{6n-1}{k} p_k \overset{3.}{=} -12 \sum_{k=n+1}^{2n} \binom{6n-1}{k-1} p_k - 1$$

p is a correlated equilibrium Since $|S| = 2$ the only possible deviations d in the definition of the correlated equilibrium are $d = id_S$, $d(0) = 1, d(1) = 0$, $d \equiv 0$ or $d \equiv 1$. We'll indicate with $J_i^n(d(p))$ the deviated cost of the chosen d. First of all, we observe that the detailed possible situations deviation are in the table 1.1.

Initial cost	**n. players with 0 (or 1)**	**n. other players**		**n. players with 0 (or 1)**	**n. other players**	**Deviated cost**
-7	$n+1$	$5n-1$	$\rightarrow$	$5n$	n	-2
-7	$n+2$	$5n-2$	$\rightarrow$	$5n-1$	$n+1$	-1
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$
-7	$2n-1$	$4n+1$	$\rightarrow$	$4n+2$	$2n-2$	-1
-7	$2n$	$4n$	$\rightarrow$	$4n+1$	$2n-1$	-1
-1	$4n$	$2n$	$\rightarrow$	$2n+1$	$4n-1$	-2
-1	$4n+1$	$2n-1$	$\rightarrow$	$2n$	$4n$	-7
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$
-1	$5n-2$	$n+2$	$\rightarrow$	$n+3$	$5n-3$	-7
-1	$5n-1$	$n+1$	$\rightarrow$	$n+2$	$5n-2$	-7

Table 1.1: From $J_i^n(p)$ to $J_i^n(d(p))$

Thus, if $d(0) = 1, d(1) = 0$

$$
\begin{aligned}
J_1^n(d(p)) &= 2\Big(-7 \sum_{k=4n+1}^{5n-1} \binom{6n-1}{k-1} p_{6n-k} - \sum_{k=n+2}^{2n} \binom{6n-1}{k-1} p_k \\
&\quad - 2\binom{6n-1}{n} p_{n+1} - 2\binom{6n-1}{2n} p_{2n}\Big) \\
&= 2\Big(-7 \sum_{k=n+1}^{2n-1} \binom{6n-1}{k} p_k - \sum_{k=n+2}^{2n} \binom{6n-1}{k-1} p_k \\
&\quad - 2\binom{6n-1}{n} p_{n+1} - 2\binom{6n-1}{2n} p_{2n}\Big) \\
&= 2\Big(- \sum_{k=n+1}^{2n} \binom{6n-1}{k} p_k - \sum_{k=n+1}^{2n} \binom{6n-1}{k-1} p_k \\
&\quad - \binom{6n-1}{n} p_{n+1} - \binom{6n-1}{2n} p_{2n} - 6 \sum_{k=n+1}^{2n} \binom{6n-1}{k} p_k\Big) \\
&= 2\Big(-\frac{1}{2} - \binom{6n-1}{n} p_{n+1} - \binom{6n-1}{2n} p_{2n} - 6 \sum_{k=n+2}^{2n-1} \binom{6n-1}{k} p_k\Big) \\
&= 2\Big(-\frac{1}{2} - \binom{6n-1}{n} p_{n+1} - \binom{6n-1}{2n} p_{2n} - 6 \sum_{k=n+2}^{2n} \binom{6n-1}{k-1} p_{k-1}\Big)
\end{aligned}
\tag{1.14}
$$

and if $d \equiv 0$ or $d \equiv 1$, with analogue calculations as in (1.14):

$$
\begin{aligned}
J_1^n(d(p)) &= -7 \sum_{k=n+1}^{2n} \binom{6n-1}{k-1} p_k - \sum_{k=n+1}^{2n} \binom{6n-1}{k-1} p_k \\
&\quad - 7 \sum_{k=4n+1}^{5n-1} \binom{6n-1}{k-1} p_{6n-k} - \sum_{k=n+1}^{2n} \binom{6n-1}{k-1} p_k \\
&\quad - 2\binom{6n-1}{n} p_{n+1} - 2\binom{6n-1}{2n} p_{2n} \\
&= -7 \sum_{k=n+1}^{2n} \binom{6n-1}{k-1} p_k - \sum_{k=n+1}^{2n} \binom{6n-1}{k-1} p_k \\
&\quad - \frac{1}{2} - \binom{6n-1}{n} p_{n+1} - \binom{6n-1}{2n} p_{2n} - 6 \sum_{k=n+2}^{2n} \binom{6n-1}{k} p_k
\end{aligned}
$$

Now, it is clear that the condition we have to impose in order to achieve $\Delta := J_1^n(p) -$

$J_1^n(d(p)) \leq 0$ for any $d : S \to S$ is:

$$-6\sum_{k=n+1}^{2n}\binom{6n-1}{k-1}p_k \leq -\binom{6n-1}{n}p_{n+1} - \binom{6n-1}{2n}p_{2n} - 6\sum_{k=n+2}^{2n}\binom{6n-1}{k}p_k$$
$$\Rightarrow -6\sum_{k=n+2}^{2n}\binom{6n-1}{k-1}(p_k - p_{k-1}) + \binom{6n-1}{n}p_{n+1} + \binom{6n-1}{2n}p_{2n} - 6\binom{6n-1}{n}p_{n+1} \leq 0$$
$$\Rightarrow -6\sum_{k=n+2}^{2n-1}\binom{6n-1}{k-1}(p_k - p_{k-1}) - 6\binom{6n-1}{2n-1}p_{2n} + 6\binom{6n-1}{2n-1}p_{2n-1}$$
$$+\binom{6n-1}{2n}p_{2n} - 5\binom{6n-1}{n}p_{n+1} \leq 0$$

Now,

$$-6\binom{6n-1}{2n-1}p_{2n} + \binom{6n-1}{2n}p_{2n} = \frac{(6n-1)!}{(2n)!(4n!)}p_{2n}(-6\cdot 2n + 4n) = -4\binom{6n-1}{2n-1}p_{2n}$$

Thus, two sufficient condition for $\Delta \leq 0$ are:

- $k \mapsto p_k$ not-decreasing
- $4p_{2n} \geq 6p_{2n-1}$

Confronting correlated and pure Nash First of all, we observe that no vector strategy in A could be a pure Nash equilibrium, since all the players having cost -1 wish to deviate in order to decrease their cost to either -6 or -2. In this type of game, pure Nash equilibria are those in which exactly k, $k \in \{1, \dots n-1\} \cap \{2n+2, \dots, 4n-2\} \cap \{5n+1, \dots, 6n\}$ players choose 0 (resp. 1). The cost for each player is thus -2.
We're interested in having

$$\Delta' := J_1^n(p) + 2 = -12\sum_{k=n+1}^{2n}\binom{6n-1}{k-1}p_k + 1 < 0$$

Using (1.13) and the fact that the distribution we're interested in is not-decreasing, we find:

$$\begin{aligned}
\Delta' &= -10\sum_{k=n+1}^{2n}\binom{6n-1}{k-1}p_k + 2\sum_{k=n+1}^{2n}\binom{6n-1}{k}p_k \\
&= -10\sum_{k=n+1}^{2n}\binom{6n-1}{k-1}p_k + 2\sum_{k=n+2}^{2n+1}\binom{6n-1}{k-1}p_{k-1} \\
&= -10p_{n+1} - \sum_{k=n+2}^{2n}\binom{6n-1}{k-1}(10p_k - 2p_{k-1}) + 2\binom{6n-1}{2n}p_{2n} \\
&\leq -10p_{n+1} - 8\sum_{k=n+2}^{6n}\binom{6n-1}{k-1}p_k + 2\binom{6n-1}{2n}p_{2n}
\end{aligned}$$

Now,

$$-8\binom{6n-1}{2n-1}p_{2n}+2\binom{6n-1}{2n}p_{2n}=\frac{(6n-1)!}{(2n)!(4n)!}(-8\cdot 2n+4\cdot 2n)=-4\binom{6n-1}{2n-1}p_{2n}$$

and thus we conclude that $\Delta' < 0$.

Confronting correlated and symmetric Nash As we said the beginning of this section, we can restrict our analysis to symmetric mixed strategies. Call $0<\lambda<1$ the probability of each player of choosing 0 and consider player 1 as the reference one. Define $q_k := \lambda^k(1-\lambda)^{6n-k}$; without loss of generality, we can suppose $\lambda \geq 1-\lambda \Rightarrow \frac{1}{2}\leq\lambda\leq 1$. Now, his/her possibilities are

- he/she is in the group of k, $k=n+1,\dots,2n$ players choosing 0 (or 1).
- he/she is in the group of k, $k=4n,\dots,5n-1$ players choosing 0 (or 1).
- he/she is in the group of k, $k=\{1,\dots n\}\cap\{2n+1,\dots,4n-1\}\cap\{5n,\dots,6n\}$ players choosing 0 (or 1).

Thus, the cost is

$$\begin{aligned} J^n_{\text{mix}}(q) = &-7\sum_{k=n+1}^{2n}\binom{6n-1}{k-1}(q_k+q_{6n-k})-\sum_{k=4n}^{5n-1}\binom{6n-1}{k-1}(q_k+q_{6n-k})\\ &-2\sum_{k=2n+1}^{4n-1}\binom{6n-1}{k-1}(q_k+q_{6n-k})-2\sum_{k=1}^{n}\binom{6n-1}{k-1}(q_k+q_{6n-k})\\ &-2\sum_{k=5n}^{6n-1}\binom{6n-1}{k-1}(q_k+q_{6n-k})-2q_0-2q_{6n}\end{aligned} \tag{1.15}$$

Before going further we have to do some observations:

1. $\displaystyle\sum_{k=0}^{6n}\binom{6n}{k}q_k=1$
2. $\displaystyle\sum_{k=5n}^{6n-1}\binom{6n-1}{k-1}(q_k+q_{6n-k})=\sum_{k=5n}^{6n-1}\binom{6n-1}{6n-k}(q_k+q_{6n-k})=\sum_{k=1}^{n}\binom{6n-1}{k}(q_{6n-k}+q_k)$
3. $\displaystyle-2\sum_{k=1}^{n}\binom{6n-1}{k-1}(q_k+q_{6n-k})-2\sum_{k=5n}^{6n-1}\binom{6n-1}{k-1}(q_k+q_{6n-k})-2q_0-2q_{6n}$
   $\displaystyle\overset{3.}{=}-2\sum_{k=1}^{n}\binom{6n}{k}(q_k+q_{6n-k})-2q_0-2q_{6n}=-2\sum_{k=0}^{n}\binom{6n}{k}(q_k+q_{6n-k})$
4. $\displaystyle\sum_{k=2n+1}^{4n-1}\binom{6n-1}{k-1}(q_k+q_{6n-k})=\sum_{k=2n+1}^{4n-1}\binom{6n-1}{k-1}q_k+\sum_{k=2n+1}^{4n-1}\binom{6n-1}{6n-k}q_{6n-k}$
   $\displaystyle=\sum_{k=2n+1}^{4n-1}\binom{6n-1}{k-1}q_k+\sum_{k=2n+1}^{4n-1}\binom{6n-1}{k}q_k=\sum_{k=2n+1}^{4n-1}\binom{6n}{k}q_k$
5. $\displaystyle\sum_{k=j}^{h}\binom{6n-1}{k-1}q_{6n-k}=\sum_{k=j}^{h}\binom{6n-1}{6n-k}q_{6n-k}=\sum_{k=6n-h}^{6n-j}\binom{6n-1}{k}q_k$

We can rewrite the global mixed cost as:

$$\begin{aligned} J^n_{\text{mix}}(q) = &-7\sum_{k=n+1}^{2n}\binom{6n-1}{k-1}(q_k+q_{6n-k}) - \sum_{k=4n}^{5n-1}\binom{6n-1}{k-1}(q_k+q_{6n-k}) \\ &-2\sum_{k=2n+1}^{4n-1}\binom{6n}{k}q_k - 2\sum_{k=0}^{n}\binom{6n}{k}(q_k+q_{6n-k}) \end{aligned}$$

We want to prove that the global cost of the correlated equilibrium is strictly inferior that that of the Nash one, so, first of all we observe that the global cost for a generic strategy vector in A is:

$$-7k-(6n-k) = -6k-6n \quad k\in\{n+1,\dots,6n\}$$

but, if we take a strategy vector in A^C the only possible global cost is

$$-2\cdot 6n$$

Since $k>n$, we have

$$-2\cdot 6n > -6k-6n \Leftrightarrow -6k+6n>0.$$

This means that every strategy vector in A has an inferior global cost with respect to those in A^C; hence it is reasonable to think that the global costs for the Nash equilibria is greater then those of the correlated one; indeed mixed strategies assign a strictly positive probabilities to strategy vectors in A^C; on the converse p is only concentrated in A.

By symmetry, the global costs are

$$\begin{aligned} J^{\text{corr}}_{\text{glob}}(p) &= 6n\cdot 2\left(-7\sum_{k=n+1}^{2n}\binom{6n-1}{k-1}p_k - \sum_{k=n+1}^{2n}\binom{6n-1}{k}p_k\right) \\ &= 2\sum_{k=n+1}^{2n}\frac{(6n)!}{k!(6n-k)!}(-7k-6n+k)\,p_k = -6\sum_{k=n+1}^{2n}\binom{6n}{k}(k+n)2p_k \end{aligned}$$

and by analogous calculations we get:

$$\begin{aligned} J^{\text{mix}}_{\text{glob}}(q) = &-6\sum_{k=n+1}^{2n}\binom{6n}{k}(k+n)(q_k+q_{6n-k}) \\ &-2\cdot 6n\left(\sum_{k=0}^{n}\binom{6n}{k}(q_k+q_{6n-k}) + \sum_{k=2n+1}^{4n-1}\binom{6n}{k}q_k\right) \end{aligned}$$

Using the fact that $\sum_{k=n+1}^{2n}\binom{6n}{k}2p_k = \sum_{k=0}^{6n}\binom{6n}{k}q_k$, we get.

$$\begin{aligned} J^{\text{mix}}_{\text{glob}}(q) = &-6\sum_{k=n+1}^{2n}\binom{6n}{k}(k+n)(q_k+q_{6n-k}) \\ &-2\cdot 6n\left(\sum_{k=n+1}^{2n}\binom{6n}{k}2p_k + \sum_{k=n+1}^{2n}\binom{6n}{k}(q_k+q_{6n-k})\right) \end{aligned}$$

Thus, calling $\Delta' := J^{\text{corr}}_{\text{glob}}(p) - J^{\text{mix}}_{\text{glob}}(q)$, we have:

$$\begin{aligned}
\Delta' =& -6\sum_{k=n+1}^{2n}\binom{6n}{k}(k+n)2p_k + -6\sum_{k=n+1}^{2n}\binom{6n}{k}(k+n)(q_k+q_{6n-k}) \\
&+ 2\cdot 6n\left(\sum_{k=n+1}^{2n}\binom{6n}{k}2p_k + \sum_{k=n+1}^{2n}\binom{6n}{k}(q_k+q_{6n-k})\right) \\
=& -6\sum_{k=n+1}^{2n}\binom{6n}{k}(k-n)2p_k + 6\sum_{k=n+1}^{2n}\binom{6n}{k}(k-n)(q_k+q_{6n-k}) \\
=& -6\sum_{k=n+1}^{2n}\binom{6n}{k}(k-n)(2p_k-(q_k+q_{6n-k}))
\end{aligned}$$

If we find a particular p that satisfies the previous conditions and such that $\Delta' < 0$ for every λ, then we've achieved our purpose. We define

$$p_{n+1} = \ldots = p_{2n-1} = 0$$

then:

$$2\binom{6n}{2n}p_{2n} = 1 \Rightarrow p_{2n} = \frac{1}{2\binom{6n}{2n}}$$

Now:

$$\begin{aligned}
\Delta' =& -6\sum_{k=n+1}^{2n}\binom{6n}{k}(k-n)(2p_k-(q_k+q_{6n-k})) \\
=& \sum_{k=n+1}^{2n-1}\binom{6n}{k}(k-n)(q_k+q_{6n-k}) - 6\binom{6n}{2n}n(2p_{2n}-(q_{2n}+q_{4n})) \\
\leq& \sum_{k=n+1}^{2n-1}\binom{6n}{k}n(q_k+q_{6n-k}) - 6\binom{6n}{2n}n(2p_{2n}-(q_{2n}+q_{4n})) \\
=& -6n\sum_{k=n+1}^{2n-1}\binom{6n}{k}(2p_k-(q_k+q_{6n-k})) < 0
\end{aligned}$$

Where the last strict inequality follows from the fact that:

$$1 = 2\sum_{k=n+1}^{6n}\binom{6n}{k}p_k = \sum_{k=0}^{6n}\binom{6n}{k}q_k > \sum_{k=n+1}^{2n-1}\binom{6n}{k}(q_k+q_{6n-k})$$

$$\Rightarrow \sum_{k=n+1}^{6n}\binom{6n}{k}(2p_k-(q_k+q_{6n-k})) > 0$$

Chapter 2

Solutions of Static Mean-Field Equations

2.1 Strong Solutions

In this section, we want to study the behaviour of Nash equilibria when $N \to \infty$ and to find and to find an equation that they satisfy in our static model. In the work of Fischer [9], there is the more general case of stochastic games. Usually the procedure goes in the opposite direction, namely we start from a mean-field equation, we find a solution for it and then we go back to the finite case, see [7] for a comprehensive argument.

Consider a mixed strategy $(\alpha_1^n, \ldots, \alpha_n^n)$; recall that the $\{\alpha_i^n\}_{i=1,\ldots,n}$ are independent. In Corollary 1.12 we've shown the existence of symmetric Nash equilibria, so it is reasonable to restrict our study to those strategies $\{\alpha_i^n\}_{i=1,\ldots,n}$ which are also identically distributed, with common distribution ν^n. If we consider the private states $\{X_i^n = \phi(\alpha_i^n)\}_{i=1,\ldots,n}$, it is clear that they are also i.i.d. with common distribution υ^n. In the following we are going to indicate the empiric measures as $\bar{\mu}_X^{n,i} := \frac{1}{n-1} \sum\limits_{j \neq i} \delta_{X_j^n}$ and $\bar{\mu}_X^n := \frac{1}{n} \sum\limits_{j=1}^{n} \delta_{X_j^n}$. We make a first simple observation about the convergence in distribution of these random measures to stress the fact that for the purpose of convergence (in distribution) it is indifferent considering one or the other.

Remark 2.1. Notice that $\bar{\mu}_X^{n,i} \to \mu$ in distribution $\Leftrightarrow \bar{\mu}_X^n \to \mu$ in distribution, indeed

$$\bar{\mu}_X^n = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i^n} = \frac{1}{n} \sum_{i \neq j} \delta_{X_j^n} + \frac{1}{n} \delta_{X_i^n} = \frac{n-1}{n} \frac{1}{n-1} \sum_{i \neq j} \delta_{X_j^n} + \frac{1}{n} \delta_{X_i^n} = \frac{n-1}{n} \bar{\mu}_X^{n,i} + \frac{1}{n} \delta_{X_i^n}$$

Now, the convergence in distribution is closed under the summation and by Corollary A.32, we have the two implications, indeed call μ and μ the limit of $\bar{\mu}_X^n$ and $\bar{\mu}_X^{n,i}$ respectively (they exists because of the tightness of the sequences).

$$\mu \leftarrow \bar{\mu}_X^n = \frac{n-1}{n} \bar{\mu}_X^{n,i} + \frac{1}{n} \delta_{X_i^n} \to 1 \cdot \mu' + 0 \cdot \upsilon = \mu' \quad \Rightarrow \quad \mu = \mu'$$

With an abuse of notation, we will denote the distribution both of $\bar{\mu}_X^{n,i}$ and $\bar{\mu}_X^n$ with m^n. Now, fixing a particular i, we have that $X_i^n : \Omega_n \to \mathcal{X}$ and $\bar{\mu}_X^{n,i} : \Omega_n \to \mathcal{P}(\mathcal{X})$ are

independent by definition. Moreover, we notice that $\mathcal{X}$ and $\mathcal{P}(\mathcal{X})$ are compact because the first is a finite set and the second is a simplex; so we consider the random vector $(X_i^n, \bar{\mu}_X^{i,n}) : \Omega_n \to \mathcal{X} \times \mathcal{P}(\mathcal{X})$; it has values in a compact space because product of two compact spaces, so it is tight. As a consequence of Prokhorov theorem A.26, there exists a sub-sequence, that we are going to denote as $(X_i^n, \bar{\mu}_X^{i,n})$ with abuse of notation, which converges in distribution to a random vector $(X, \mu) : \Omega \to \mathcal{X} \times \mathcal{P}(\mathcal{X})$ for a certain probability space $(\Omega, \mathcal{A}, \mathbb{P})$.
Since $\{X_i^n\}_n$ are $\bar{\mu}_X^{n,i}$ are both tight they also are convergent in distribution, up to sub-sequences. X_i^n and $\bar{\mu}_X^{n,i}$ are respectively the projection on the first and on the second component of the vector $(X_i^n, \bar{\mu}_X^{i,n})$ and, by the Continuous mapping theorem A.17, it has to be $X_i^n \to X$ and $\bar{\mu}_X^{n,i} \to \mu$ in distribution. Thus, calling υ and m the distributions of X and μ, $\upsilon^n \to \upsilon$ and $m^n \to m$ weakly.
Observe that, due to the independence, the distribution of the vector $(X_i^n, \bar{\mu}_X^{i,n})$ is the product measure $\upsilon^n \otimes m^n$ and by Proposition A.30, the distribution of (X, μ) is $\upsilon \otimes m$.

Proposition 2.2 (Strong solutions of the static Mean-Field Equation)**.** Let g be continue and bounded. Let, for any n, $(\tilde{\alpha}_1^n, \dots, \tilde{\alpha}_1^n)$ be a symmetric Nash equilibrium, and $(\tilde{X}_1^n, \dots, \tilde{X}_n^n)$ the correspondent private states with distribution $(\tilde{\upsilon}^n, \dots, \tilde{\upsilon}^n)$. Then, up to a subsequence $\{\tilde{\upsilon}^n\}$ converges weakly to a measure υ satisfying:

$$\int_{\mathcal{X}} g(x, \upsilon)\upsilon(dx) = \inf_{\sigma \in \mathcal{P}(\mathcal{X})} \int_{\mathcal{X}} g(x, \upsilon)\sigma(dx). \tag{2.1}$$

Moreover the empiric measure $\bar{\mu}_X^N$, up to sub-sequences, converges in distribution to $\delta_\upsilon \in \mathcal{P}(\mathcal{P}(\mathcal{X}))$.

Remark 2.3. Before starting the proof, we want to see what does it means to be a Nash equilibrium in terms of the distributions of the private states. We want to prove that Definition 1.1 is equivalent to: $(\tilde{\alpha}_1^n, \dots, \tilde{\alpha}_1^n)$ is a symmetric Nash equilibrium, with $(\tilde{X}_1^n, \dots, \tilde{X}_n^n)$ the correspondent private states with distribution $(\tilde{\upsilon}^n, \dots, \tilde{\upsilon}^n)$ if

$$\int_{\mathcal{P}(\mathcal{X}) \times \mathcal{X}} g(x, y)\tilde{\upsilon}^n(dx)dm^n(dy) \leq \int_{\mathcal{P}(\mathcal{X}) \times \mathcal{X}} g(x, y)\sigma(dx)m^n(dy) \ \ \forall \sigma \in \mathcal{P}(\mathcal{X}) \tag{2.2}$$

(2.2) $\Rightarrow$ (1.3): if $(\tilde{\alpha}_1^n, \dots, \tilde{\alpha}_1^n)$ is a symmetric Nash equilibrium (in the sense of (1.3)), with $(\tilde{X}_1^n, \dots, \tilde{X}_n^n)$ the correspondent private states with distribution $(\tilde{\upsilon}^n, \dots, \tilde{\upsilon}^n)$, consider a generic $\sigma \in \mathcal{P}(\mathcal{X})$ and consider a random variable $X^n : \Omega_n \to \mathcal{X}$ with distribution σ. Since $\phi : S \to \mathcal{X}$ is surjective, because $\mathcal{X}$ is defined as $\mathcal{X} := \phi(S)$ there exists a $\beta : \Omega_n \to S$ such that $X^n = \phi(\beta)$ and so by (1.1)

$$\begin{aligned}
&\int_{\mathcal{P}(\mathcal{X}) \times \mathcal{X}} g(x, y)\tilde{\upsilon}^n(dx)m^n(dy) = \int_{\Omega_n} g\Big(\tilde{X}_i^n(\omega), \bar{\mu}^{n,i}_{\tilde{X}(\omega)}\Big)\mathbb{P}_n(d\omega) \\
&= \int_{\Omega_n} g\Big(\phi(\tilde{\alpha}_i^n(\omega)), \bar{\mu}^{n,i}_{\tilde{X}(\omega)}\Big)\mathbb{P}_n(d\omega) \leq \int_{\Omega_n} g\Big(\phi(\beta_i^n(\omega)), \bar{\mu}^{n,i}_{\tilde{X}(\omega)}\Big)\mathbb{P}_n(d\omega) \\
&= \int_{\Omega_n} g\Big(X^n(\omega)), \bar{\mu}^{n,i}_{\tilde{X}(\omega)}\Big)\mathbb{P}_n(d\omega) = \int_{\mathcal{P}(\mathcal{X}) \times \mathcal{X}} g(x, y)\sigma(dx)m^n(dy)
\end{aligned}$$

(1.3) ⇒ (2.2): for any mixed strategy β define $X = \phi(\beta)$ and let σ be its distribution, then:

$$\int_{\Omega_n} g\Big(\phi(\tilde{\alpha}_i^n(\omega)), \bar{\mu}^{n,i}_{\tilde{X}(\omega)}\Big)\mathbb{P}_n(d\omega) = \int_{\mathcal{P}(\mathcal{X})\times\mathcal{X}} g(x,y)\tilde{v}^n(dx)m^n(dy)$$
$$\leq \int_{\mathcal{P}(\mathcal{X})\times\mathcal{X}} g(x,y)\sigma(dx)m^n(dy) = \int_{\Omega_n} g\Big(\phi(\beta_i^n(\omega)), \bar{\mu}^{n,i}_{\tilde{X}(\omega)}\Big)\mathbb{P}_n(d\omega)$$

Proof. In the following paragraphs we'll study $\bar{\mu}^n_X = \frac{1}{N}\sum_{i=1}^n \delta_{X_i^n}$ and we'll indicate again with $\bar{\mu}^n_X$ the converging subsequence, that exists since $\mathcal{P}(\mathcal{X})$ is compact. We'll call μ the random variable to whom the empiric measure converges. By definition of convergence in distribution, we have that $\forall\ f : \mathcal{P}(\mathcal{X}) \to \mathbb{R}$ continue and bounded:

$$\lim_{n\to\infty}\int_{\Omega_n} f\circ\bar{\mu}^n_X(\omega)\mathbb{P}(d\omega) = \mathbb{E}_n[f(\bar{\mu}^n_X)] = \mathbb{E}_P[f(\mu)] = \int_\Omega f\circ\mu(\omega)\mathbb{P}(d\omega)$$

First of all we observe that, since $\mathcal{X}$ is compact, $\mathcal{C}_b(\mathcal{X})$ is separable[1], thus there exists a countable and dense subset, say $\mathcal{T}$. For the topology induced by the weak convergence, because of Proposition A.8, if two measures, μ and $\nu \in \mathcal{P}(\mathcal{X})$, coincide if tested on $\mathcal{T}$ than $\sigma = \lambda$. This means that if $\langle\sigma, f\rangle = \int_\mathcal{X} f d\sigma = \int_\mathcal{X} f d\lambda = \langle\lambda, f\rangle$, for all $f \in \mathcal{T}$, than $\sigma = \lambda$. Let $\Psi \in \mathcal{T}$ and define

$$m_\Psi := \mathbb{E}_P\Big[\int_\mathcal{X}\Psi(x)\mu(dx)\Big] = \int_\Omega\int_\mathcal{X}\Psi(x)\mu_\omega(dx)\mathbb{P}(d\omega)$$
$$v_\Psi := \mathbb{E}_P\Big[\Big(\int_\mathcal{X}\Psi(x)\mu(dx) - m_\Psi\Big)^2\Big]$$
$$m^n_\Psi := \mathbb{E}_n\Big[\int_\mathcal{X}\Psi(x)\bar{\mu}^n(dx)\Big] = \int_{\Omega_n}\int_\mathcal{X}\Psi(x)\bar{\mu}^n_\omega(dx)\mathbb{P}_n(d\omega)$$

where $\mathbb{E}_P$ and $\mathbb{E}_n$ are the expected values respectively on $(\Omega, \mathcal{A}, \mathbb{P})$ and $(\Omega_n, \mathcal{A}_n, \mathbb{P}_n)$.
Notice that the mapping $F : \sigma \mapsto \int \Psi d\sigma$ is a continuous function in $\mathcal{P}(\mathcal{X})$ (for the topology induced by the weak convergence), indeed, consider $\sigma_n \in \mathcal{P}(\mathcal{X})$ for all n, then $F(\sigma_n) = \int \psi d\sigma_n \to \int \psi d\sigma = F(\sigma)$. So, by the convergence of $\bar{\mu}^n$ to μ, $\mathbb{E}_n[F(\bar{\mu}^n)] \to \mathbb{E}_P[F(\mu)]$, which means that $m^n_\Psi \to m_\Psi$ for every $\Psi \in \mathcal{T}$. Moreover, because of the Continuous mapping theorem in the version A.18:

$$v_\Psi = \lim_{n\to\infty}\mathbb{E}_n\Big[\Big(\int_\mathcal{X}\Psi(x)\bar{\mu}^n(dx) - m^n_\Psi\Big)^2\Big] = \lim_{n\to\infty}\mathbb{E}_n\Big[\Big(\int_\mathcal{X}\Psi(x)d\Big(\frac{1}{n}\sum_{i=1}^n\delta_{X_i^n}\Big)(x) - m^n_\Psi\Big)^2\Big]$$
$$= \lim_{n\to\infty}\mathbb{E}_n\Big[\Big(\frac{1}{n}\sum_i\Psi(X_i^n) - m^n_\Psi\Big)^2\Big].$$

Moreover, we observe that, since $\{X_i^n\}_i$ are identically distributed:

$$m^n_\Psi = \mathbb{E}_n\Big[\int_\mathcal{X}\Psi(x)\bar{\mu}^n(dx)\Big] = \mathbb{E}_n\Big[\frac{1}{n}\sum_i\Psi(X_i^n)\Big] = \mathbb{E}_n[\Psi(X_1^n)]$$
$$\Rightarrow\quad \mathbb{E}_n\Big[\frac{1}{n}\sum_i\Psi(X_i^n) - m^n_\Psi\Big] = 0\quad\forall n \tag{2.3}$$

[1] For a complete proof see [8]

Now, since X_i^n are i.i.d. for $i = 1, \dots, n$, so do the $\Psi(X_i^n)$ and the $\Psi(X_i^n)^2$, thus, using (2.3):

$$
\begin{aligned}
&\mathbb{E}_n\Big[\Big(\frac{1}{n}\sum_i \Psi(X_i^n) - m_\Psi^n\Big)^2\Big] = \mathbb{E}_n\Big[\Big(\frac{1}{n}\sum_i \Psi(X_i^n) - m_\Psi^n\Big)\Big(\frac{1}{n}\sum_i \Psi(X_i^n) - m_\Psi^n\Big)\Big] \\
&= \mathbb{E}_n\Big[\Big(\frac{1}{n}\sum_i \Psi(X_i^n)\Big)\Big(\frac{1}{n}\sum_i \Psi(X_i^n) - m_\Psi^n\Big)\Big] - m_\Psi^n \mathbb{E}_n\Big[\frac{1}{n}\sum_i \Psi(X_i^n) - m_\Psi^n\Big] \\
&= \mathbb{E}_n\Big[\Big(\frac{1}{n}\sum_i \Psi(X_i^n)\Big)\Big(\frac{1}{n}\sum_i \Psi(X_i^n) - m_\Psi^n\Big)\Big] \\
&= \mathbb{E}_n\Big[\Big(\frac{1}{n}\sum_i \Psi(X_i^n)\Big)^2\Big] - m_\Psi^n \mathbb{E}_n\Big[\frac{1}{n}\sum_i \Psi(X_i^n)\Big] \\
&= \frac{1}{n^2}\mathbb{E}_n\Big[\Big(\sum_i \Psi(X_i^n)\Big)^2\Big] - (m_\Psi^n)^2 \\
&= \frac{1}{n^2}\mathbb{E}_n\Big[\sum_i \Psi(X_i^n)^2\Big] + \frac{2}{n^2}\mathbb{E}_n\Big[\sum_{i=1}^n \sum_{j<i} \Psi(X_i^n)\Psi(X_j^n)\Big] - (m_\Psi^n)^2 \\
&= \frac{1}{n^2}\; n\; \mathbb{E}_n\big[\Psi(X_1^n)^2\big] + \frac{2}{n^2} \cdot \frac{n(n-1)}{2}\mathbb{E}_n\big[\Psi(X_1^n)\big]^2 - (m_\Psi^n)^2 \\
&= \frac{1}{n}\mathbb{E}_n\big[\Psi(X_1^n)^2\big] - \frac{1}{n}\mathbb{E}_n\big[\Psi(X_1^n)\big]^2 \xrightarrow{n\to\infty} 0.
\end{aligned}
$$

The convergence toward 0 follows from the boundedness of Ψ. We've found that $v_\Psi = 0 \Rightarrow \mathbb{E}_P[(\int_{\mathcal{X}} \Psi(x) d\mu(x) - m_\Psi)^2] = 0$. Hence for all $\Psi \in \mathcal{T}$, $\mathbb{P}$- almost surely

$$\int_{\mathcal{X}} \Psi(x)\mu(dx) = m_\Psi$$

but, since $\mathcal{T}$ is countable we can exchange the quantifiers and get that $\mathbb{P}$- almost surely

$$\int_{\mathcal{X}} \Psi(x)\mu(dx) = m_\Psi \quad \text{for all } \Psi \in \mathcal{T}$$

which means that there exists a set $A \in \Omega$ with probability one, such that for any $\omega \in A$

$$\int_{\mathcal{X}} \Psi(x)\mu_\omega(dx) = m_\Psi \quad \text{for all } \Psi \in \mathcal{T} \tag{2.4}$$

thus, there exists a unique measure $\sigma \in \mathcal{P}(\mathcal{X})$ such that

$$\int_{\mathcal{X}} \Psi(x)\sigma(dx) = m_\Psi \quad \text{for all } \Psi \in \mathcal{T}.$$

indeed, the uniqueness follows from: let $\sigma, \tilde{\sigma}$ be two measures such that $\int_{\mathcal{X}} \Psi d\sigma = \int_{\mathcal{X}} \Psi d\tilde{\sigma} = m_\Psi$ for all $\Psi \in \mathcal{T}$, so they coincide if tested on all functions in $\mathcal{T} \Rightarrow \sigma = \tilde{\sigma}$. For the existence, it is sufficient to define σ as $\sigma := \mu_\omega$ for an $\omega \in A$, observing that the definition does not depend on the choice of the particular ω since (2.4) holds for every $\omega \in A$.

We now show that $\sigma = \upsilon$ which is the limit distribution of υ^n (up to sub-sequences). As said before $m^n_\Psi \to m_\Psi$ for every $\Psi \in \mathcal{T}$, but we've also that

$$\begin{aligned} m^n_\Psi &= \mathbb{E}_n\Big[\int_{\mathcal{X}} \Psi(x)\bar{\mu}^n(dx)\Big] = \mathbb{E}_n[\Psi(X^n_1)] \\ &= \int_{\Omega_n} \Psi(X^n_1(\omega))\mathbb{P}(d\omega) = \int_{\mathcal{X}} \Psi(x)\upsilon^n(dx) \xrightarrow{n\to\infty} \int_{\mathcal{X}} \Psi(x)\upsilon(dx) \quad \text{for all } \Psi \in \mathcal{T} \end{aligned}$$

By the uniqueness of the limit in distribution:

$$\int_{\mathcal{X}} \Psi(x)\upsilon(dx) = m_\Psi = \int_{\mathcal{X}} \Psi(x)\sigma(dx) \quad \text{for all } \Psi \in \mathcal{T}$$

$\Rightarrow \sigma = \upsilon$.
So if $F \in \mathcal{C}(\mathcal{P}(\mathcal{X}))$ and m the distribution of μ:

$$\begin{aligned} \langle \delta_\upsilon, F\rangle &= \int_{\mathcal{P}(\mathcal{X})} F(x)\delta_\upsilon(dx) = F(\upsilon) \\ \langle m, F\rangle &= \int_{\mathcal{P}(\mathcal{X})} F(x)m(dx) = \int_\Omega F(\mu_\omega)\mathbb{P}(d\omega) = \int_A F(\mu_\omega)\mathbb{P}(d\omega) + \int_{A^C} F(\mu_\omega)\mathbb{P}(d\omega) \\ &= \int_A F(\mu_\omega)\mathbb{P}(d\omega) = \int_A F(\upsilon)\mathbb{P}(d\omega) = F(\upsilon)\mathbb{P}(A) = F(\upsilon) \end{aligned}$$

$\Rightarrow m = \delta_\upsilon$.
Now we have to prove (2.1). We know that $(\tilde{\alpha}^n_1, \dots, \tilde{\alpha}^n_1)$ is a Nash equilibrium for all n, so by (2.2) we find:

$$\int_{\mathcal{P}(\mathcal{X})\times\mathcal{X}} g(x,y)\upsilon^n(dx)m^n(dy) \le \int_{\mathcal{P}(\mathcal{X})\times\mathcal{X}} g(x,y)\sigma(dx)m^n(dy) \ \ \forall \sigma \in \mathcal{P}(\mathcal{X})$$

Observe that, because of the boundedness of g we can apply Fubini-Tonelli's theorem . Moreover, since $\mathcal{P}(\mathcal{X}) \times \mathcal{X}$ is separable (because compact), the product measure $\upsilon^n \otimes m^n$ is weakly convergent and by Proposition A.30, $\upsilon^n \otimes m^n \to \upsilon \otimes m$. Now:

$$\begin{aligned} \lim_{n\to\infty} \int_{\mathcal{P}(\mathcal{X})\times\mathcal{X}} g(x,y)m^n(dy)\upsilon^n(dx) &= \int_{\mathcal{P}(\mathcal{X})\times\mathcal{X}} g(x,y)m(dy)\upsilon(dx) \\ = \int_{\mathcal{P}(\mathcal{X})\times\mathcal{X}} g(x,y)\delta_\upsilon(dy)\upsilon(dx) &= \int_{\mathcal{X}} g(x,\upsilon)\upsilon(dx) \end{aligned}$$

and for all $\sigma \in \mathcal{P}(\mathcal{X})$

$$\begin{aligned} \lim_{n\to\infty} \int_{\mathcal{P}(\mathcal{X})\times\mathcal{X}} g(x,y)\sigma(dx)m^n(dy) &= \int_{\mathcal{P}(\mathcal{X})\times\mathcal{X}} g(x,y)\sigma(dx)m(dy) \\ = \int_{\mathcal{P}(\mathcal{X})\times\mathcal{X}} g(x,y)\sigma(dx)\delta_\upsilon(dy) &= \int_{\mathcal{X}} g(x,\upsilon)\sigma(dx) \end{aligned}$$

$\Rightarrow \int_{\mathcal{X}} g(x,\upsilon)\upsilon(dx) \le \int_{\mathcal{X}} g(x,\upsilon)\sigma(dx)$ for all $\sigma \in \mathcal{P}(\mathcal{X})$, but, since υ realizes the equality, we have:

$$\int_{\mathcal{X}} g(x,\upsilon)\upsilon(dx) = \inf_{\sigma\in\mathcal{P}(\mathcal{X})} \int_{\mathcal{X}} g(x,\upsilon)\sigma(dx)$$

which concludes the proof. □

We call (2.1) *mean field equation*. It allows us to study a continuous problem, indeed a υ that satisfies (2.1) is called a *Nash equilibrium for the continuous game*. Moreover, we observe that that condition is equivalent to:

$$\upsilon(\{x \mid g(x,\upsilon) \leq g(y,\upsilon) \; \forall \, y \in \mathcal{X}\}) = 1. \tag{2.5}$$

Heuristically it means that, if the other players play with "mean strategy" υ, the most convenient thing to do for me is to play υ, namely it is useful to act as the others.

Remark 2.4. Notice that in our case $\mathcal{X}$ is finite and thus compact and separable and also $\mathcal{P}(\mathcal{X})$ is compact and thus separable. But Proposition 2.2 holds even if $\mathcal{X}$ is simply compact, indeed it follows from Proposition A.15.

2.2 Weak Solutions

In this section, we're going to study the behaviour of correlated equilibria when $N \to \infty$. The purpose is to achieve an equation similar to (2.1), with thw difference that here we can't exploit the independence between the variables.
If considering a correlated strategy $\alpha^n = (\alpha_1^n, \dots, \alpha_n^n)$ in certain probability space $(\Omega_n \mathbb{P}_n)$, due to the symmetry of the problem, in analogy with what we have done in the Nash case, we can restrict our analysis to the exchangeable ones, which means that $(\alpha_1^n, \dots, \alpha_n^n) \overset{D}{=} (\alpha_{\sigma(1)}^n, \dots, \alpha_{\sigma(n)}^n)$ in distribution, for any permutation σ; this implies that also the private states are exchangeable $\Rightarrow$ $(X_1^n, \dots, X_n^n) \overset{D}{=} (X_{\sigma(1)}^n, \dots, X_{\sigma(n)}^n)$. Observe that for every finite game it always exists al least one exchangeable correlated equilibrium, indeed we know that there always exists a symmetric Nash equilibrium $(\alpha_1, \dots, \alpha_n)$, which is, thus a collection of i.i.d. random variables, which implies that, seen as a correlated, the vector $(\alpha_1, \dots, \alpha_n)$ is exchangeable.

2.2.1 Exchangeability

Before going further, we want to give some useful characterizations and properties of exchangeable random variables.

Definition 2.5. Let I be an arbitrary index set and E be a Polish space. A family $(X_i)_{i \in I}$ of random variables with values in E is called **exchangeable** if

$$\text{Law}[(X_i)_{i \in I}] = \text{Law}[(X_{\sigma(i)})_{i \in I}]$$

for any finite permutation $\sigma : I \to I$.

Remark 2.6. It follows directly from the definition that the following are equivalent:

- $(X_i)_{i \in I}$ are exchangeable
- Let $n \in \mathbb{N}$ and $i_1, \dots, i_n \in I$, $j_1, \dots, j_n$ be pairwise distinct. Then $\text{Law}[(X_{i_1}, \dots, X_{i_n})] = \text{Law}[(X_{j_1}, \dots, X_{j_n})]$.

Observe that, if we consider $n = 1$ we have that exchangeable random variables are identical distributed.
The converse is not true in general. Consider X, Y who take values on $\{0, 1, 2\}$ with the following distribution:

X \Y	0	1	2
0	$\frac{1}{7}$	0	$\frac{1}{7}$
1	$\frac{1}{7}$	$\frac{2}{7}$	0
2	0	$\frac{1}{7}$	$\frac{2}{7}$

Making the sum over the rows and the columns we have that $\mathbb{P}(X = 0) = \mathbb{P}(Y = 0) = \frac{2}{7}$, $\mathbb{P}(X = 1) = \mathbb{P}(Y = 1) = \frac{3}{7}$, $\mathbb{P}(X = 2) = \mathbb{P}(Y = 2) = \frac{3}{7}$, but $\mathbb{P}(X = 0, Y = 1) = 0 \neq \frac{1}{7} = \mathbb{P}(X = 1, Y = 0)$.
This means that the exchangeability is a stronger property than to be identically distributed.

Let $X = (X_n)_{n\in\mathbb{N}}$ be a stochastic process with values in a Polish space E. Let $S(n)$ be the set of permutation $\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}$. We consider σ also as a map $\mathbb{N} \to \mathbb{N}$ by defining $\sigma(k) = k$ for $k > n$. For $\sigma \in S(n)$ and $x = (x_1, \ldots, x_n) \in E^n$, denote $x_\sigma = (x_{\sigma(1)}, \ldots, x_{\sigma(n)})$. Similarly, for $x \in E^{\mathbb{N}}$, define $x_\sigma = (x_{\sigma(1)}, x_{\sigma(2)}, \ldots) \in E^{\mathbb{N}}$. Let E' be another Polish space. For measurable maps $f : E^n \to E'$ and $F : E^{\mathbb{N}} \to E'$, define the maps f_σ, F_σ by $f_\sigma(x) = f(x_\sigma), F_\sigma(x) = F(x_\sigma)$. In the following, we might write $f(x) = (x_1, \ldots, x_n)$ both for $x \in E^n$ and $x \in E^{\mathbb{N}}$.

Definition 2.7.
- A map $f : E^n \to E'$ is called **symmetric** if $f_\sigma = f$ for all $\sigma \in S(n)$
- A map $F : E^{\mathbb{N}} \to E'$ is called n**-symmetric** id $F_\sigma = F$ for all $\sigma \in S(n)$. F is **symmetric** if is n-symmetric for all $n \in \mathbb{N}$.

Definition 2.8. Let $X = (X_n)_{n\in\mathbb{N}}$ be a stochastic process with values in E. For $n \in \mathbb{N}$, define the σ-algebra $\mathcal{E}'_n$ in $E^{\mathbb{N}}$ as:

$$\mathcal{E}'_n := \sigma(F \mid F : E^{\mathbb{N}} \to \mathbb{R} \text{ measurable and } n\text{-symmetric})$$

and let $\mathcal{E}_n := X^{-1}(\mathcal{E}'_n)$ be the σ-algebra of the events that are invariant for X under all permutations $\sigma \in S(n)$. Further, let

$$\mathcal{E}' = \bigcap_{n=1}^{\infty} \mathcal{E}'_n = \sigma(F \mid F : E^{\mathbb{N}} \to \mathbb{R} \text{ measurabe and symmetric})$$

and let $\mathcal{E} = \bigcap_{n=1}^{\infty} \mathcal{E}_n = X^{-1}(\mathcal{E}')$ be the σ-algebra of exchangeable events for X, or briefly the *exchangeable σ-algebra.*

Remark 2.9. Let $n \in \mathbb{N}$. Then every symmetric function $f : E^n \to \mathbb{R}$ can be written in the form $f(x) = g(\frac{1}{n}\sum_{i=1}^{n} \delta_{x_i})$. Indeed, for any f symmetric, define

$$\begin{aligned} g_f : &\mathcal{P}(E) \to E' \\ &\frac{1}{n}\sum_{i=1}^{n} \delta_{x_i} \mapsto f(x_1, \ldots, x_n) \end{aligned}$$

and such that for any other $\nu \in \mathcal{P}(E)$ be measurable. Such g_f is well-defined because of the symmetry of f:

$$g_f\left(\frac{1}{n}\sum_{i=1}^{n}\delta_{x_i}\right) = f(x_1,\dots,x_n) = f(x_{\sigma(1)},\dots,x_{\sigma(n)}) = g_f\left(\frac{1}{n}\sum_{i=1}^{n}\delta_{x_{\sigma(i)}}\right).$$

Thus, we've associated for every f symmetric, a function g as demanded.

Now, because of Remark 2.9, we have that $\mathcal{E}_n = \sigma(\bar{\mu}^n_X)$.

Theorem 2.10. Let $X = (X_n)_n$ be exchangeable. If $\varphi : E^{\mathbb{N}} \to \mathbb{R}$ is measurable and if $\mathbb{E}[|\varphi(X)|] < \infty$, then, for all $n \in \mathbb{N}$ and all $\sigma \in S(n)$

$$\mathbb{E}[\varphi(X) \mid \mathcal{E}_n] = \mathbb{E}[\varphi(X_\sigma) \mid \mathcal{E}_n]$$

In particular

$$\mathbb{E}[\varphi(X) \mid \mathcal{E}_n] = \frac{1}{n!}\sum_{\sigma\in S(n)}\varphi(X_\sigma) =: A_n(\varphi)$$

Proof. Let $A \in \mathcal{E}_n$. Then there exists a $B \in \mathcal{E}'_n$ such that $A = X^{-1}(B)$. Let $F = \mathbf{1}_B$; then $F \circ X = \mathbf{1}_A$. By the definition of $\mathcal{E}_n$, the map $F : E^{\mathbb{N}} \to \mathbb{R}$ is measurable, n-symmetric and bounded. Therefore, for any $\sigma \in S(n)$

$$\mathbb{E}[\varphi(X)\mathbf{1}_A] = \mathbb{E}[\varphi(X)F(X)] = \mathbb{E}[\varphi(X_\sigma)F(X_\sigma)] = \mathbb{E}[\varphi(X_\sigma)F(X)] = \mathbb{E}[\varphi(X_\sigma)\mathbf{1}_A].$$

Where we've used the exchageability of X in the first equality and the symmetry of F in the second one. We now observe that $A_n(\varphi)$ is $\mathcal{E}_n$-measurable and hence

$$\mathbb{E}[\varphi(X) \mid \mathcal{E}_n] = \mathbb{E}\Big[\frac{1}{n!}\sum_{\sigma\in S(n)}\varphi(X_\sigma) \mid \mathcal{E}_n\Big] = \frac{1}{n!}\sum_{\sigma\in S(n)}\varphi(X_\sigma).$$

□

Miming the proof of Theorem 2.10, we have the result also for finite sequence of exchangeable random variables:

Corollary 2.11. Let $X = (X_1,\dots,X_n)$ be exchangeable If $\varphi : E^n \to \mathbb{R}$ is measurable and if $\mathbb{E}[|\varphi(X)|] < \infty$, then, for all $\sigma \in S(n)$

$$\mathbb{E}[\varphi(X) \mid \mathcal{E}_n] = \mathbb{E}[\varphi(X_\sigma) \mid \mathcal{E}_n]$$

In particular

$$\mathbb{E}[\varphi(X) \mid \mathcal{E}_n] = \frac{1}{n!}\sum_{\sigma\in S(n)}\varphi(X_\sigma) =: A_n(\varphi)$$

Remark 2.12. We notice that for a finite sequence of exchangeable random variables, we could not apply directly Theorem 2.10, because it isn't true that every finite exchangeable sequence could be extended to an infinite exchangeable; as a matter of fact, for any $n \in \mathbb{N} \setminus \{1\}$, there is an exchangeable family of random variables $X_1,\dots,X_n$ that cannot be extended to an infinite exchangeable family $X_1, X_2, \ldots$. Take for example X_1, X_2, X_3 which take values in $\{0, 1\}$ such that:

$$\mathbb{P}(X_i = 1) = \mathbb{P}(X_j = 0) = \frac{1}{2} \qquad \forall\, j = 1, 2, 3$$

$$\mathbb{P}(X_1 = 1, X_1 = 2) = \mathbb{P}(X_1 = 1, X_1 = 3) = \mathbb{P}(X_1 = 2, X_1 = 3) = \frac{1}{k}$$

Suppose that X_1, X_2, X_3 could be extended to an infinite sequence of exchangeable random variables, thus, for every n we should have:

$$\begin{aligned}
0 \le \operatorname{Var}\Big(\sum_{i=1}^{n} \mathbf{1}_{\{X_i=1\}}\Big) &= \mathbb{E}\Big[\Big(\sum_{i=1}^{n} \mathbf{1}_{\{X_i=1\}}\Big)^2\Big] - \Big(\mathbb{E}\Big[\sum_{i=1}^{n} \mathbf{1}_{\{X_i=1\}}\Big]\Big)^2 \\
&= \sum_{i=1}^{n} \mathbb{E}\left[\mathbf{1}_{\{X_i=1\}}\right] + \sum_{i=1}^{n}\sum_{j\neq i} \mathbb{E}\left[\mathbf{1}_{\{X_i=1\}}\mathbf{1}_{\{X_j=1\}}\right] - (n\mathbb{P}(X_j = 1))^2 \\
&= \frac{n}{2} + \frac{n(n-1)}{k} - \frac{n^2}{4} = \frac{2nk + 4n^2 - 4n - kn^2}{4k} = \frac{n(2k - 4 - (k-4)n)}{4k} \\
&\Rightarrow 0 \le n \le \frac{2k-4}{k-4}
\end{aligned}$$

The existence of a superior bound for n implies that the sequence could not be extended to an infinite one.

2.2.2 Probability Kernels

Before starting the analysis of our particular case we need to introduce a concept that will generalize the conditional probability to the cases in which we are conditioning by an event of probability 0. For example, if X is a uniform random variable in $[0, 1]$ and A is an event, we would like to give sense to an expression of the type $\mathbb{P}(A \mid X = \frac{1}{2})$. Moreover, if X is a random variable which takes values on a measurable space $(E, \mathcal{E})$ and $A \in \mathcal{A}$ is an event in a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, we want also $\mathbb{P}(\cdot \mid X = x)$ to be a probability measure for all $x \in E$ such that $\forall A \in \mathcal{A}$, $\mathbb{P}(A \mid X)(\omega) = \mathbb{P}(A \mid X = x)$ on $\{\omega : X(\omega) = x\}$.

We begin with a general lemma; we will indicate by $\mathcal{B}(S)$ the Borel σ-algebra over a set S.

Lemma 2.13 (Factorization Lemma)**.** Let $(\Omega', \mathcal{A}')$ be a measurable space and let Ω be a non-empty set. Let $f : \Omega \to \Omega'$ be a map. A map $g : \Omega \to \bar{\mathbb{R}}$ is $\sigma(f)$ - $\mathcal{B}(\mathbb{R})$-measurable if and only if there is a measurable map $\phi : (\Omega', \mathcal{A}') \to (\bar{\mathbb{R}}, \mathcal{B}(\bar{\mathbb{R}}))$ such that $g = \phi \circ f$.

Proof. $\Leftarrow$: If ϕ is measurable and $g = \phi \circ f$, then g is clearly $\sigma(f)$-$\mathcal{B}(\bar{\mathbb{R}})$-measurable.

$\Rightarrow$: Assume g to be $\sigma(f)$-$\mathcal{B}(\bar{\mathbb{R}})$-measurable. First consider the case $g \ge 0$. Then there exist measurable sets $A_1, A_2, \ldots \in \sigma(f)$ as well as numbers $\alpha_1, \alpha_2, \ldots \in [0, +\infty)$ such that $g = \sum_{n=1}^{\infty} \alpha_n \mathbf{1}_{A_n}$. By the definition of $\sigma(f)$, for any $n \in \mathbb{N}$ there is a set $B_n \in \mathcal{A}'$ such that $f^{-1}(B_n) = A_n$; that is, $\mathbf{1}_{A_n} = \mathbf{1}_{B_n} \circ f$. Define $\phi : \Omega' \to \bar{\mathbb{R}}$ as $\phi = \sum_{n=1}^{\infty} \alpha_n \mathbf{1}_{B_n}$, thus ϕ is $\mathcal{A}'$-$\mathcal{B}(\bar{\mathbb{R}})$-measurable and $g = \phi \circ f$.

For a general g, consider g^+ and g^-, its positive and negative parts. As said above there exist ϕ^+ and ϕ^- such that $g^+ = \phi^+ \circ f$ and $g^- = \phi^- \circ f$, hence $\phi = \phi^+ - \phi^-$ does the trick.

□

Let X be as above and let Z be a $\sigma(X)$-measurable real random variable; by Lemma 2.13 (with $g = Z$ and $X = f$), there exists a map $\phi : E \to \mathbb{R}$ such that

$$\phi \text{ is } \mathcal{E}\text{-}\mathcal{B}(\bar{\mathbb{R}})\text{-measurable and } \phi(X) = Z. \tag{2.6}$$

We'll indicate, with abuse of notation, $Z \circ X^{-1} := \phi$.

Definition 2.14. Let $X \in L^1(\Omega)$ and $Y : (\Omega, \mathcal{A}) \to (E, \mathcal{E})$. We define the **conditional expectation** of X given $\{Y = x\}$ by $\mathbb{E}[X \mid Y = x] := \phi(x)$, where ϕ is the function on (2.6) with $Z = \mathbb{E}[X \mid Y]$.
Analogously we define $\mathbb{P}(A \mid Y = x) := \mathbb{E}[\mathbf{1}_A \mid X = x]$ for $A \in \mathcal{A}$.

Definition 2.15. Let $(\Omega_1, \mathcal{A}_1)$, $(\Omega_2, \mathcal{A}_2)$ be two measurable spaces. A map $k : \Omega_1 \times \mathcal{A}_2 \to [0, \infty]$ is called a (σ-) finite **transition kernel** (from Ω_1 to Ω_2) if:

- $\omega_1 \mapsto k(\omega_1, A_2)$ is $\mathcal{A}_1$-measurable for any $A_2 \in \mathcal{A}_2$.
- $A_2 \mapsto k(\omega_1, A_2)$ is a (σ-) finite measure on $(\Omega_2, \mathcal{A}_2)$ for any $\omega_1 \in \Omega_1$.

If in the second condition the measure is a probability measure for all $\omega_1 \in \Omega_1$ then k is called **probability** or **Markov kernel**.

Remark 2.16. Notice that a random measure verifies the conditions in Definition 2.15, which means that it is always a probability kernel.

Definition 2.17. Let X be a random variable with values in a measurable space $(E, \mathcal{E})$ and let $\mathcal{F} \subset \mathcal{A}$ be a sub-σ-algebra. A probability kernel $k_{X,\mathcal{F}}$ from $(\Omega, \mathcal{F})$ to $(E, \mathcal{E})$ is called a **regular conditional probability** of X given $\mathcal{F}$ if for $\mathbb{P}$-almost all $\omega \in \Omega$ and for all $B \in \mathcal{E}$

$$k_{X,\mathcal{F}}(\omega, B) = \mathbb{P}(X \in B \mid \mathcal{F})(\omega)$$

that is,

$$\int_\Omega \mathbf{1}_B(X(\omega))\mathbf{1}_A \mathbb{P}(d\omega) = \int_\Omega k_{X,\mathcal{F}}(\omega, B)\mathbf{1}_A \mathbb{P}(d\omega) \text{ for all } A \in \mathcal{F}, B \in \mathcal{E} \tag{2.7}$$

Remark 2.18. Consider the special case where $\mathcal{F} = \sigma(Y)$ for a random variable Y with values in a measurable space $(E', \mathcal{E}')$; then the probability kernel $k_{X,Y}$ is defined by

$$(x, A) \mapsto k_{X,Y}(x, A) = \mathbb{P}(X \in A \mid Y = x) = k_{X,\sigma(Y)}(Y^{-1}(x), A)$$

and it's called a regular conditional distribution of X given Y. Observe that this map is the function arising from the Factorization Lemma with an arbitrary value for $x \notin Y(\Omega)$. To be more precise:

$$k_{X,\sigma(Y)} : \Omega \times \mathcal{E} \to \mathbb{R}$$
$$k_{X,Y} : E' \times \mathcal{E} \to \mathbb{R}.$$

given $B \in \mathcal{E}$, let $\phi(\cdot, B)$ be a measurable function such that $k_{X,\sigma(Y)}(\omega, B) = \phi(Y(\omega), B)$ for all $\omega \in \Omega$. Notice that $k_{X,\sigma(Y)}(\omega, B)$ is constant on the set $\{\omega \mid Y(\omega) = x\}$ if $x \in E'$, hence $\phi(x, \cdot)$ is a probability measure on $(E, \mathcal{E})$ for every $x \in Y(\omega)$. For every $x \notin Y(\Omega)$, set $\phi(x, \cdot) = \mu(\cdot)$, where μ is an arbitrary probability measure on $(E, \mathcal{E})$. Therefore ϕ is a probability kernel from $(E', \mathcal{E}')$ to $(E, \mathcal{E})$. By definition of $k_{X,\sigma(Y)}$:

$$\mathbb{P}(\{X \in B\} \cap A) = \int_A k_{X,\sigma(Y)}(\cdot, B) d\mathbb{P} = \int_A \phi(Y(\cdot), B) d\mathbb{P} = \int_{Y^{-1}(A)} \phi(\cdot, B) d\mathrm{Law}(Y)$$

for every $A \in \sigma(Y)$. Also, by definition:

$$\begin{aligned} \mathbb{P}(\{X \in B\} \cap A) &= \mathbb{E}[\mathbf{1}_{\{X \in B\}} \mathbf{1}_A] = \mathbb{E}[\mathbb{E}[\mathbf{1}_{\{X \in B\}} \mid \sigma(Y)]\mathbf{A}] = \int_A \mathbb{E}[\mathbf{1}_{\{X \in B\}} \mid \sigma(Y)] d\mathbb{P} \\ &= \int_A \mathbb{P}(X \in B \mid \sigma(Y))(\cdot) d\mathbb{P} = \int_A \mathbb{P}(X \in B \mid Y = x)_{x = Y(\cdot)} d\mathbb{P} \\ &= \int_{Y^{-1}(A)} \mathbb{P}(X \in B \mid Y = \cdot) d\mathrm{Law}(Y) \end{aligned}$$

This means that $k_{X,Y}(x, B) := \phi(x, B) = k_{X,\sigma(Y)}(Y^{-1}(x), B) = \mathbb{P}(X \in B \mid Y = x)$ for every $x \in Y(\Omega)$.
Moreover, for every $B \in \mathcal{E}$, we can see $k_{X,Y}(\cdot, B)$ as a random measure in the following way:

$$\begin{aligned} & k_{X,Y}(\cdot, B) : \Omega \to \mathbb{R} \\ & k_{X,Y}(\cdot, B)(\omega) = k_{X,Y}(Y(\omega), B) = k_{X,\sigma(Y)}(\omega, B) \end{aligned} \tag{2.8}$$

Theorem 2.19 (Existence of the regular conditional distribution)**.** Fix a Borel space E and let Y be a random element in E and $\mathcal{F} \subset \mathcal{A}$ a sub-σ-algebra. Then, there exists a regular conditional distribution $k_{X,\mathcal{F}}$ of X given $\mathcal{F}$.

Proof. Since E is a Borel space, we may assume $S \in \mathcal{B}(\mathbb{R})$. The strategy will consist in constructing a measurable version of the distribution function of the conditional distribution of X by first defining it for rational values and then extending it for the real numbers.
For $r \in \mathbb{Q}$, let $F(r, \cdot)$ be a modification of the conditional probability $\mathbb{P}(X \in (-\infty, r] \mid \mathcal{F})$. For $r \leq s$ $\mathbf{1}_{\{X \in (-\infty, r]\}} \leq \mathbf{1}_{\{X \in (-\infty, s]\}}$, thus, by the monotonicity of the condition expectation, there is a null probability set $A_{r,s} \in \mathcal{F}$ such that

$$F(r, \omega) \leq F(s, \omega) \quad \text{for all } \omega \in \Omega \setminus A_{r,s}. \tag{2.9}$$

By dominated convergence (for the conditional expectation), for any $r \in \mathbb{Q}$ there exists a null probability set $B_r \in \mathcal{F}$ and $C \in \mathcal{F}$ such that

$$\lim_{n \to \infty} F\big(r + \frac{1}{n}, \omega\big) = F(r, \omega) \quad \text{for all } \omega \in \Omega \setminus B_r$$

as well as

$$\inf_{n \in \mathbb{N}} F(-n, \omega) = 0 \quad \text{and} \quad \sup_{n \in \mathbb{N}} F(n, \omega) = 1 \quad \text{for all } \omega \in \Omega \setminus C.$$

Let $N := (\bigcup_{r,s\in\mathbb{Q}}) \cup (\bigcup_{r\in\mathbb{Q}} B_r) \cup C$. For $\omega \in \Omega \setminus N$ define

$$\tilde{F}(z,\omega) := \inf\{F(r,\omega) : r \in \mathbb{Q}, r > z\} \quad \text{for all } z \in \mathbb{R}$$

By construction $\tilde{F}(\cdot,\omega)$ is monotone increasing, right continuous with limit 1 in ∞ and 0 in $-\infty$. This means that $\tilde{F}(\cdot,\omega)$ is a distribution function for any $\omega \in \Omega \setminus N$. For $\omega \in N$ define $\tilde{F}(\cdot,\omega) = F_0$ with F_0 an arbitrary but fixed distribution function.
For any $\omega \in \Omega$ let $k(\omega,\cdot)$ be the probability measure on $(\Omega,\mathcal{A})$ with distribution function $\tilde{F}(\cdot,\omega)$. Then, for $r \in \mathbb{Q}$ and $B = (-\infty, r]$,

$$\omega \mapsto k(\omega,B) = F(r,\omega)\mathbf{1}_{N^C}(\omega) + F_0(r)\mathbf{1}_N(\omega) \tag{2.10}$$

is $\mathcal{F}$-measurable. Now $\{(-\infty,r] : r \in \mathbb{Q}\}$ is a π-system that generates $\mathcal{B}(\mathbb{R})$, thus the measurability holds for all $B \in \mathcal{B}(\mathbb{R})$ and hence k is identified as a probability kernel.
We now show that k is a modification of the conditional distribution. For $A \in \mathcal{F}, r \in \mathbb{Q}$ and $B = (-\infty, r]$, by (2.10),

$$\int_A k(\omega,B)\mathbb{P}(d\omega) = \int_A F(r,\omega)\mathbb{P}(d\omega) = \int_A \mathbb{P}[X \in B \mid \mathcal{F}]\mathbb{P}(d\omega) = \mathbb{P}[A \cap \{X \in B\}]$$

As function of B both sides are finite measure on $\mathcal{B}(\mathbb{R})$ that coincide on the π-system $\{(-\infty,r] : r \in \mathbb{Q}\}$. Thus, we have equality for all $B \in \mathbb{R}$. Hence $\mathbb{P}$-a.s. $k(\cdot,B) = \mathbb{P}[X \in B \mid \mathcal{F}]$, that is $k = k_{X,\mathcal{F}}$. □

Proposition 2.20. Let X be a random variable with values on a Borel space $(E,\mathcal{E})$. Let $\mathcal{F} \subset \mathcal{A}$ a σ-algebra and lef $k_{X,\mathcal{F}}$ be a regular conditional distribution of X given $\mathcal{F}$. Let $f : E \to \mathbb{R}$ be measurable and $\mathbb{E}[|f(X)|] < \infty$. Then

$$\mathbb{E}[f(X)|\mathcal{F}] = \int_E f(x)k_{X,\mathcal{F}}(\cdot,dx) \;\text{ for } \mathbb{P}\text{-almost all } \omega \tag{2.11}$$

More in general:

Theorem 2.21 (Disintegration Formula)**.** Fix two measurable spaces $(E,\mathcal{E})$ and $(T,\mathcal{T})$, a σ-field $\mathcal{F} \subset \mathcal{A}$ and a random element X in E such that $\mathbb{P}[X \in \cdot \mid \mathcal{F}]$ has a regular version $k_{X,\mathcal{F}}$. Further consider a $\mathcal{F}$-measurable random element Y in T and a measurable function $f : E \times T \to \mathbb{R}$ with $\mathbb{E}[|f(X,Y)|] < \infty$. Then

$$\mathbb{E}[f(X,Y) \mid \mathcal{F}] = \int_E f(x,Y)k_{X,\mathcal{F}}(\cdot,dx) \quad \text{a.s.} \tag{2.12}$$

Proof. Take $B \in \mathcal{E}$ and $C \in \mathcal{T}$, then , recalling that X is $\mathcal{F}$-measurable:

$$\mathbb{P}[X \in B, X \in C] = \mathbb{E}[\mathbb{P}[X \in B, X \in C \mid \mathcal{F}]] = \mathbb{E}[\mathbb{P}[X \in B \mid \mathcal{F}]\mathbf{1}_{\{Y \in C\}}]$$
$$= \mathbb{E}[k_{X,\mathcal{F}}(\cdot,B)\mathbf{1}_{\{Y\in C\}}] = \mathbb{E}\Big[\int_E k_{X,\mathcal{F}}(\cdot,dx)\mathbf{1}_{\{y\in B, Y\in C\}}\Big]$$

Thus, we've proven that

$$\mathbb{E}[f(X,Y)] = \mathbb{E}[\int_E f(x,Y)k_{X,\mathcal{F}}(\cdot,dx)] \tag{2.13}$$

with $f(X,y) = \mathbf{1}_{\{X\in B, Y\in C\}}$. The formula shows the measurability of the integral on the right-hand side of (2.12) and extends, by linearity and monotone convergence, (2.13) to any measurable function $f \geq 0$.
Fix now a measurable function $f : E \times T \to \mathbb{R}_+$ with $\mathbb{E}[f(X,Y)] < \infty$ and let $A \in \mathcal{F}$ be arbitrary. Observe that $(X, \mathbf{1}_B)$ is a $\mathcal{F}$-measurable random element in $T \times [0,1]$, so $g(X,(Y,\mathbf{1}_A)) := f(X,Y)\mathbf{1}_A$ satisfies our hypothesis, thus by (2.13):

$$\mathbb{E}[f(X,Y)\mathbf{1}_A] = \mathbb{E}[\int_E f(x,Y)\mathbf{1}_A k_{X,\mathcal{F}}(\cdot,dx)] \quad A \in \mathcal{F}$$

By definition and uniqueness of conditional expectation

$$\int_E f(x,Y)k_{X,\mathcal{F}}(\cdot,dx) = \mathbb{E}[f(X,Y) \mid \mathcal{F}] \quad \text{a.s.}$$

For general f (and not only $f \geq 0$), it's sufficient taking differences. □

Remark 2.22. In the special case in when $\mathcal{F} = \sigma(Y)$ and $\mathbb{P}[X = \cdot \mid Y] = k_{X,Y}(\cdot,\cdot)$, (2.12) becomes

$$\mathbb{E}[f(X,Y) \mid Y] = \int_E f(x,Y)k_{X,Y}(\cdot,dx) \quad \text{a.s.} \tag{2.14}$$

2.2.3 Convergence of Correlated Equilibria

Unlike the Nash case, here we don't have the independence between the private states, so we have to study the convergence of the random vector $(X_i^n, \bar{\mu}_X^{n,i}) : \Omega_n \to \mathcal{X} \times \mathcal{P}(\mathcal{X})$. Due to the compactness of $\mathcal{X}, \mathcal{P}(\mathcal{X})$ and so $\mathcal{X} \times \mathcal{P}(\mathcal{X})$, up to subsequences $\{(X_i^n, \bar{\mu}_X^{n,i})\}_n, \{X_i^n\}_n, \{\bar{\mu}_X^{n,i}\}_n$ are convergent in distribution. Denote with (X,μ) the limit of the random vector and with $p, p^n \in \mathcal{P}(\mathcal{X} \times \mathcal{P}(\mathcal{X}))$ the respective distributions. Thus:

$$\begin{aligned} (X_i^n, \bar{\mu}_X^{n,i}) &\xrightarrow{n\to\infty} (X,\mu) \quad \text{in distribution} \\ p^n &\xrightarrow{n\to\infty} p \qquad\quad \text{weakly} \end{aligned}$$

Now, X_i^n and $\bar{\mu}_X^{n,i}$ are the marginals of the vector, so they are the composition with a continuous function, the projection; thus, by the Continuous Mapping Theorem A.17, we find, said $\upsilon^n, m^n, \upsilon, m$ the distributions of $X_i^n, \bar{\mu}_X^{n,i}, X, \mu$:

$$\begin{aligned} X_i^n &\xrightarrow{n\to\infty} X \quad \text{in distribution} \qquad & \bar{\mu}_X^{n,i} &\xrightarrow{n\to\infty} \mu \quad \text{in distribution} \\ \upsilon^n &\xrightarrow{n\to\infty} \upsilon \quad \text{weakly} & m^n &\xrightarrow{n\to\infty} m \quad \text{weakly} \end{aligned}$$

Let us consider $\Psi \in \mathcal{C}_b(\mathcal{X})$ (observe that both the continuity and the boundedness hypothesis is redundant because of the compactness of $\mathcal{X}$, but it is an hypothesis which has to be kept if we only suppose the compactness of $\mathcal{X}$), then, as in section 2.1, $F : \mathcal{P}(\mathcal{X}) \to \mathbb{R}, \sigma \mapsto \int_{\mathcal{X}} \Psi d\sigma$ is continuous with respect to the weak topology. Thus , if $i \neq 1$, using the fact that the X_i^n are identically distributed:

$$\mathbb{E}_n\Big[\int_{\mathcal{X}} \Psi d\bar{\mu}_X^{n,i}\Big] = \mathbb{E}_n[F(\bar{\mu}_X^{n,i})] \xrightarrow{n\to\infty} \mathbb{E}[F(\mu)]$$

$$\mathbb{E}_n[\int_{\mathcal{X}} \Psi d\bar{\mu}_X^{n,i}] = \mathbb{E}_n\Big[\int_{\mathcal{X}} \Psi d\Big(\frac{1}{n-1}\sum_{j\neq i}\delta_{X_j^N}\Big)\Big] = \mathbb{E}_n\Big[\frac{1}{n-1}\sum_{j\neq i}\Psi(X_j^n)\Big]$$

$$=\frac{1}{n-1}(n-1)\mathbb{E}_n[\Psi(X_1^n)] = \mathbb{E}_n[\Psi(X_1^n)] \xrightarrow{n\to\infty} \mathbb{E}[\Psi(X)] = \sum_{x\in\mathcal{X}}\Psi(x)\mathbb{P}(X=x)$$

By the uniqueness of the convergence in distribution:

$$\mathbb{E}\Big[\int_{\mathcal{X}} \Psi d\mu\Big] = \mathbb{E}[\Psi(X)] = \sum_{x\in\mathcal{X}}\Psi(x)\mathbb{P}(X=x) \tag{2.15}$$

for any $\Psi \in \mathcal{C}_b(\mathcal{X})$.
We have thus obtained that $\mathbb{E}[\mu] = \upsilon$. But we can find a better result by calculating the regular conditional probability of X_i^n given $\bar{\mu}_X^n$.
Recall that, by Remark 2.1, studying the convergence (in distribution) of $\bar{\mu}_X^{n,i}$ is equivalent to study that of $\bar{\mu}_X^n$ and their limit is the same.

Proposition 2.23. Let $(X_1^n, \ldots, X_n^n)$ exchangeable with $X_i^n : \Omega_n \to \mathcal{X}$ for every i, let $\bar{\mu}_X^n = \frac{1}{n}\sum_i \delta_{X_i}$ be its empirical measure. Then

$$k_{X_i^n,\bar{\mu}_X^n}(\cdot,\cdot) = \bar{\mu}_X^n \tag{2.16}$$

Proof. We need to prove that $\bar{\mu}_X^n$ verifies (2.7) for every $B \in \mathcal{B}(\mathcal{X})$ and for every A that is $\sigma(\mu)$-measurable, i.e.:

$$\mathbb{E}_n[\mathbf{1}_{\{X_i^n=x\}} \mid \bar{\mu}_X^n] = \bar{\mu}_X^n(x) \quad \text{a.s.}$$

Now, since $(X_1^n, \ldots, X_n^n)$ is exchangeable, as a consequence of Corollary 2.11, taking $\varphi = \delta_{\pi_1(X^n)}(x)$, we have:

$$\mathbb{E}_n[\varphi(X^n) \mid \mathcal{E}_n] = \mathbb{E}_n[\mathbf{1}_{\{X_1^n=x\}} \mid \bar{\mu}_X^n] = \ldots = \mathbb{E}_n[\mathbf{1}_{\{X_n^n=x\}} \mid \bar{\mu}_X^n] \quad \text{a.s.}$$

thus

$$\begin{aligned}\mathbb{E}_n[\mathbf{1}_{\{X_i^n=x\}} \mid \bar{\mu}_X^n] &= \frac{1}{n}\sum_{i=1}^n \mathbb{E}_n[\mathbf{1}_{\{X_i^n=x\}} \mid \bar{\mu}_X^n] = \mathbb{E}_n\Big[\frac{1}{n}\sum_{i=1}^n \mathbf{1}_{\{X_i^n=x\}} \mid \bar{\mu}_X^n\Big] \\ &=\mathbb{E}_n\Big[\frac{1}{n}\sum_{i=1}^n \delta_{X_i^n}(x) \mid \bar{\mu}_X^n\Big] = \mathbb{E}_n[\bar{\mu}_X^n(x) \mid \bar{\mu}_X^n] = \mathbb{E}_n[\langle \mathbf{1}_x, \bar{\mu}_X^n\rangle \mid \bar{\mu}_X^n] = \bar{\mu}_X^n(x)\end{aligned} \tag{2.17}$$

□

By definition of the regular conditional probability we've find that $k_{X_i^n,\sigma(\bar{\mu}_X^n)} = \bar{\mu}_X^n$; thus for every $\theta \in \mathcal{P}(\mathcal{X})$, $k_{X_i^n,\bar{\mu}_X^n}(\theta,\cdot) = k_{X_i^n,\sigma(\bar{\mu}_X^n)}((\bar{\mu}_X^n)^{-1}(\theta),\cdot) = \theta(\cdot) = \mathbb{P}(X_i^n \in \cdot \mid \bar{\mu}_X^n = \theta)$. Briefly we could formally write: $\mathrm{Law}(X_i^n \mid \bar{\mu}_X^n = \theta) = \theta$.

If we show that $k_{X_i^n,\bar{\mu}_X^n}(\cdot,B)$, seen as a real random variable, converges in distribution to $k_{X,\mu}(\cdot, B)$ for all $B \in \mathcal{B}(\mathcal{X})$, since $\bar{\mu}^n_X$ converge in distribution to μ, we have that $k_{X,\mu}(\cdot,\cdot) = \mu$, and hence $\text{Law}(X \mid \mu = \theta) = \theta$. In order to achieve this result, we first show the convergence of the conditional expectation; consider $f : \mathcal{P}(\mathcal{X}) \to \mathbb{R}$ a continuous function, then $\mathbb{E}[\mathbf{1}_{\{X=x\}}|\mu]$ is the (a.s.) unique function Z such that $\mathbb{E}[\mathbf{1}_{\{X=x\}} f(\mu)] = \mathbb{E}[Zf(\mu)]$. Because of the convergence in distribution of $(X^n_i, \bar{\mu}^n_X)$, we have:

$$\mathbb{E}_n[\mathbf{1}_{\{X^n_i=x\}} f(\bar{\mu}^n_X)] \to \mathbb{E}[\mathbf{1}_{\{X=x\}} f(\mu)]$$

but also

$$\begin{aligned}\mathbb{E}_n[\mathbf{1}_{\{X^n_i=x\}} f(\bar{\mu}^n_X)] &= \mathbb{E}_n[\mathbb{E}_n[\mathbf{1}_{\{X^n_i=x\}} f(\bar{\mu}^n_X) \mid \bar{\mu}^n_X]] = \mathbb{E}_n[f(\bar{\mu}^n_X)\mathbb{E}_n[\mathbf{1}_{\{X^n_i=x\}} \mid \bar{\mu}^n_X]] \\ &\overset{(2.17)}{=} \mathbb{E}_n[f(\bar{\mu}^n_X)\bar{\mu}^n_X(x)] = \mathbb{E}_n[\bar{\mu}^n_X(x) f(\bar{\mu}^n_X)] \to \mathbb{E}[\mu(x) f(\mu)]\end{aligned}$$

This means that for all $f : \mathcal{P}(\mathcal{X}) \to \mathbb{R}$ continuous, by the uniqueness of the convergence in distribution:

$$\mathbb{E}[\mathbf{1}_{\{X=x\}} f(\mu)] = \mathbb{E}[\mu(x) f(\mu)] \Rightarrow \mathbb{E}[\mathbf{1}_{\{X=x\}}|\mu] = \mu(x)$$

Going back to the particular case of a correlated equilibrium, first of all, we suppose that the function ϕ that links the chosen strategy within the private state is injective, and thus bijective. In this situation every deviation from the private state corresponds to a deviation from the strategy (observe that the converse it is always true), indeed if $d : \mathcal{X} \to \mathcal{X}$ than, we can define $\tilde{d} : S \to S$ as $\tilde{d} : \phi^{-1} \circ d \circ \phi$, hence $\tilde{d}(\alpha) = \phi^{-1} \circ d \circ \phi(\alpha) = \phi^{-1}(d(X))$.
We could than rewrite the definition of the correlated equilibrium, for a fixed n, in this particular situation in the following way: a correlated strategy $\alpha^n = (\alpha^n_1, \dots, \alpha^n_n)$ is a correlated equilibrium if, for every i and for every $d : \mathcal{X} \to \mathcal{X}$

$$\mathbb{E}_n[g(X^n_i, \bar{\mu}^{n,i}_X)] \le \mathbb{E}_n[g(d(X^n_i), \bar{\mu}^{n,i}_X)] \tag{2.18}$$

Proposition 2.24 (Weak solutions of the static Mean-Field equation)**.** Let ϕ be bijective. Let $g : \mathcal{X} \times \mathcal{P}(\mathcal{X}) \to \mathbb{R}$ be continuous (and bounded). Let $\tilde{\alpha}^n = (\tilde{\alpha}^n_1, \dots, \tilde{\alpha}^n_n)$ an exchangeable correlated equilibrium when varying n, and $(\tilde{X}^n_1, \dots, \tilde{X}^n_n)$ the correspondent private states. Then, up to a subsequence, the random vector $\{(\tilde{X}^n_i, \bar{\mu}^{i,n}_{\tilde{X}})\}$ converges in distribution to (X, μ) satisfying:

$$\mathbb{E}\Big[\int_{\mathcal{X}} g(x,\mu)\mu(dx)\Big] = \inf_{d:\mathcal{X}\to\mathcal{X}} \Big\{\mathbb{E}\Big[\int_{\mathcal{X}} g(s,\mu)k_d(\cdot, ds)\Big]\Big\} \tag{2.19}$$

where the infimum is taken over the probability kernels $k_d(\cdot,\cdot) = \mathbb{P}(d(X) \in \cdot \mid \mu)$ a.s..

Proof. We want to study the convergence of the cost function $J^n_i(\tilde{\alpha})$. Since $(\tilde{X}^n_i, \bar{\mu}^{n,i}_{\tilde{X}})$, up to sub-sequences converges in distribution, to (X, μ), say, we have:

$$J^n_i(\tilde{\alpha}) = \mathbb{E}_n[g(\tilde{X}^n_i, \bar{\mu}^{n,i}_{\tilde{X}})] \xrightarrow{n\to\infty} \mathbb{E}[g(X,\mu)]$$

We can apply the Disintegration Formula (2.12) and obtain:

$$\begin{aligned}J^n_i(\tilde{\alpha}) &\xrightarrow{n\to\infty} \mathbb{E}[g(X,\mu)] = \mathbb{E}[\mathbb{E}[g(X,\mu)|\mu]] \\ &= \mathbb{E}\Big[\int_{\mathcal{X}} g(x,\mu)k_{X,\mu}(\cdot, dx)\Big] = \mathbb{E}\Big[\int_{\mathcal{X}} g(x,\mu)\mu(dx)\Big]\end{aligned}$$

Since ϕ is bijective and for any n, $\tilde{\alpha}^n = (\tilde{\alpha}_1^n, \dots, \tilde{\alpha}_n^n)$ is a correlated equilibrium, by equation (2.18), for every i and for every $d : \mathcal{X} \to \mathcal{X}$, $\mathbb{E}_n[g(\tilde{X}_i^n, \bar{\mu}_{\tilde{X}}^{n,i})] \leq \mathbb{E}_n[g(\phi(d(\tilde{\alpha}_i^n)), \bar{\mu}_{\tilde{X}}^{n,i}))]$. We write $Z_n := \phi(d(\tilde{\alpha}_i^n))$, hence the inequality becomes: $\mathbb{E}_n[g(\tilde{X}_i^n, \bar{\mu}_{\tilde{X}}^{n,i})] \leq \mathbb{E}_n[g(Z_n, \bar{\mu}_{\tilde{X}}^{n,i}))]$. Because of the compactness of the spaces we have again $(Z_n, \bar{\mu}_{\tilde{X}}^{n,i}) \to (Z, \mu)$ and $Z_n \to Z$ in distribution (up to sub-sequences):

$$\mathbb{E}_n[g(Z_n, \bar{\mu}_{\tilde{X}}^{n,i}))] \xrightarrow{n\to\infty} \mathbb{E}[g(Z, \mu)]$$

Now, $\mathbb{P}[Z \in \cdot \mid \mu] = k_{Z,\mu}(\cdot, \cdot)$ for some probability kernel $k_{Z,\mu}$, thus, integrating (2.14), we obtain:

$$\mathbb{E}[g(Z, \mu)] = \mathbb{E}[\mathbb{E}[g(Z, \mu) \mid \mu]] = \mathbb{E}\Big[\int_{\mathcal{X}} g(s, \mu) k_{Z,\mu}(\cdot, ds)\Big] \tag{2.20}$$

So, we have that for every $d : \mathcal{X} \to \mathcal{X}$

$$\mathbb{E}\Big[\int_{\mathcal{X}} g(x, \mu)\mu(dx)\Big] \leq \mathbb{E}\Big[\int_{\mathcal{X}} g(s, \mu) k_d(\cdot, ds)\Big].$$

Where $k_d(\cdot, \cdot) = \mathbb{P}(d(X) \in \cdot \mid \mu)$ a.s.. But since $d = id_{\mathcal{X}}$ realizes the equality, we finally have:

$$\mathbb{E}\Big[\int_{\mathcal{X}} g(x, \mu)\mu(dx)\Big] = \inf_{d:\mathcal{X}\to\mathcal{X}} \Big\{\mathbb{E}\Big[\int_{\mathcal{X}} g(s, \mu) k_d(\cdot, ds)\Big]\Big\}$$

□

Equation (2.19) tells us that μ, which could be interpreted as the law of X given μ, minimizes the limit cost function all over the laws of the deviations from X always given μ.

We know want to verify that a strong solution in the sense of (2.1) is also a weak solution in the sense of (2.19). Take $\mu = \theta \in \mathcal{P}(\mathcal{X})$ a.s., but, since in (2.15) we've proved that $\mathbb{E}[\mu] = \upsilon$, with $\upsilon = \text{Law}(X)$, we have $\theta = \mathbb{E}[\theta] = \mathbb{E}[\mu] = \upsilon \Rightarrow \mu = \upsilon$ a.s. . The left-hand side in (2.19) becomes

$$\mathbb{E}\Big[\int_{\mathcal{X}} g(x, \mu)\mu(dx)\Big] = \int_{\mathcal{X}} g(x, \upsilon)\upsilon(dx) = \mathbb{E}[g(X, \upsilon)]$$

which is exactly the left-hand side in (2.1). Considering now the right-hand side, we find:

$$k_d(\theta, B) = \mathbb{P}(d(X) \in B \mid \mu = \theta) \neq 0 \Leftrightarrow \theta = \upsilon$$

in that case we obtain, since $\{\mu = \upsilon\}$ has probability one, $k_d(\upsilon, B) = \mathbb{P}(d(X) \in B \mid \mu = \upsilon) = \mathbb{P}(d(X) \in B)$, hence

$$k_d(\cdot, dx)(\omega) = k_d(\mu_\omega, dx) = k_d(\upsilon, dx) = \mathbb{P}(d(X) \in dx) \quad \text{for a.e. } \omega \in \Omega$$

$$\mathbb{E}\Big[\int_{\mathcal{X}} g(x, \mu) k_d(\cdot, dx)\Big] = \int_{\mathcal{X}} g(x, \upsilon) \text{Law}_{d(X)}(dx)$$

Thus, we observe that a strong solution in the sense of (2.1) is a weak solution in the sense of (2.19). On the contrary, if in (2.19) we use the hypothesis of the Nash's case (μ

deterministic), we do not obtain exactly (2.1), indeed in the latter case we have that the infimum is taken over all the maps $d : \mathcal{X} \to \mathcal{X}$, but if X doesn't have "full image", which means that if $\mathrm{Im}(X) \subsetneq \mathcal{X}$, we can't obtain all the distributions over $\mathcal{X}$ when varying d, but only those over $\mathrm{Im}(X)$.

In Observation 1.8, we've seen that, in the finite case, every linear combination of Nash equilibria is still a correlated equilibrium; now, we want to see if this behaviour is kept also in the mean-field case. So, let $\theta_1, \theta_2 \in \mathcal{P}(\mathcal{X})$, let m, the distribution of μ, be defined as $m = p\delta_{\theta_1} + (1-p)\delta_{\theta_2} \in \mathcal{P}(\mathcal{P}(\mathcal{X}))$. This means that

$$\mathbb{P}(\mu = \theta_1) = p \qquad \mathbb{P}(\mu = \theta_2) = 1 - p$$

Then, on the left-hand side, we have:

$$\mathbb{E}\Big[\int_{\mathcal{X}} g(x,\mu)\mu(dx)\Big] = p\int_{\mathcal{X}} g(x,\theta_1)\theta_1(dx) + (1-p)\int_{\mathcal{X}} g(x,\theta_2)\theta_2(dx)$$

and on the right-hand side:

$$\begin{aligned}
&\mathbb{E}\Big[\int_{\mathcal{X}} g(x,\mu)k_d(\cdot,dx)\Big] = \int_{\Omega}\int_{\mathcal{X}} g(x,\mu_\omega)k_d(\cdot,dx)(\omega)\mathbb{P}(d\omega)\\
=&\int_{\mu^{-1}(\theta_1)}\int_{\mathcal{X}} g(x,\mu_\omega)k_d(\mu_\omega,dx)\mathbb{P}(d\omega) + \int_{\mu^{-1}(\theta_2)}\int_{\mathcal{X}} g(x,\mu_\omega)k_d(\mu_\omega,d\omega)\mathbb{P}(dx)\\
=&\,p\int_{\mathcal{X}} g(x,\theta_1)k_d(\theta_1,dx) + (1-p)\int_{\mathcal{X}} g(x,\theta_2)k_d(\theta_2,dx)\\
=&\,p\int_{\mathcal{X}} g(x,\theta_1)\mathbb{P}(d(X)\in dx \mid \mu=\theta_1) + (1-p)\int_{\mathcal{X}} g(x,\theta_2)\mathbb{P}(d(X)\in dx \mid \mu=\theta_2)\\
=&\,p\int_{\mathcal{X}} g(x,\theta_1)\mathrm{Law}_{(d(X)|\mu=\theta_1)}(dx) + (1-p)\int_{\mathcal{X}} g(x,\theta_2)\mathrm{Law}_{(d(X)|\mu=\theta_2)}(dx)
\end{aligned}$$

Now, if, for $i = 1,2$, $\int_{\mathcal{X}} g(x,\theta_i)\theta_i(dx) = \inf_\sigma \int_{\mathcal{X}} g(x,\theta_i)\sigma(dx)$, we'll have:

$$\begin{aligned}
&p\int_{\mathcal{X}} g(x,\theta_1)\theta_1(dx) + (1-p)\int_{\mathcal{X}} g(x,\theta_2)\theta_2(dx)\\
=&\,p\inf_\sigma\int_{\mathcal{X}} g(x,\theta_1)\sigma(dx) + (1-p)\inf_\sigma\int_{\mathcal{X}} g(x,\theta_2)\sigma(dx)\\
\le&\inf_\sigma\{\int_{\mathcal{X}} g(x,\theta_1)\sigma(dx) + (1-p)\int_{\mathcal{X}} g(x,\theta_2)\sigma(dx)\}
\end{aligned} \tag{2.21}$$

Thus, if σ realizes the infimum in the sense of (2.21), it will also realizes the infimum in the sense of weak solution of a static mean-field game.
Here the difference between Nash's and correlated cases are two: (1) as before (in the case of μ deterministic) the infimum is taken for the modifications of X, (2) we have to consider the conditioning by the value of μ, that is could be ignored since X and μ are not independent and $\{\mu = \theta_i\}$, for $i = 1,2$, is not a trivial event.

2.2.4 Additional remarks

We observe that in (2.19), it would seem natural to have

$$\mathbb{E}\left[\int_{\mathcal{X}} g(x,\mu)\mu(dx)\right] = \inf_{f:\mathcal{X}\times\mathcal{P}(\mathcal{X})\to\mathcal{X}} \mathbb{E}\left[\int_{\mathcal{X}} g(f(x,\mu),\mu)\mu(dx)\right]$$

which is equivalent to

$$\mathbb{E}\left[g(X,\mu)\right] = \inf_{f:\mathcal{X}\times\mathcal{P}(\mathcal{X})\to\mathcal{X}} \mathbb{E}\left[g(f(X,\mu),\mu)\right] \tag{2.22}$$

this intuition is given by the fact that it seems reasonable to think that an equilibrium is also an "a posteriori" concept, that means that a player reaches a stable situation when, after seeing what the other players choose (in this case only by means of their empirical measure), he/she chooses his action.

We weren't be able to prove (2.22), but it appears to be false by the following motivations. First of all, we notice that, since a Nash equilibrium is also a correlated one, if (2.22) is valid, it has to be verified also in the case of the strong solutions; thus, we start from the case of Nash equilibria when $N \to \infty$. We know that in this case, $\mu = \upsilon = \text{Law}(X)$ a.s., hence we have that any random function $f(\cdot,\mu) = f(\cdot,\upsilon)$ a.s., which means that there a bijective relation between any deviation $f : \mathcal{X}\times\mathcal{P}(\mathcal{X}) \to \mathcal{X}$ and those of the form $d : \mathcal{X} \to \mathcal{X}$, it follows that, in this situation:

$$\mathbb{E}\left[g(X,\mu)\right] = \inf_{d:\mathcal{X}\to\mathcal{X}} \mathbb{E}[g(d(X),\mu)] = \inf_{f:\mathcal{X}\times\mathcal{P}(\mathcal{X})\to\mathcal{X}} \mathbb{E}\left[g(f(X,\mu),\mu)\right]$$

If we consider instead the case of N finite, we manage to give a very simple example in which the deviation $f : \mathcal{X}\times\mathcal{P}(\mathcal{X}) \to \mathcal{X}$ is not allowed. In fact, consider a 2-persons game where every players can choose between two actions, say $S = \{0,1\}$, and the private state coincides with the respective strategy. The cost matrix we are considering is:

1/2	0	1
0	(-2,-2)	(-3,-3)
1	(-3,-3)	(-2,-2)

In this case, since $\bar{\mu}_X^{i,n} = \delta_{X_j}$, we can simply consider $f : S\times S \to S$. We also observe that, by symmetry, the choice of the player is indifferent; in the following, we will in fact refer to a generic player.
A symmetric Nash equilibrium is $p = (\frac{1}{2},\frac{1}{2})$, indeed:

$$\frac{d}{dx}(-2x^2 - 6x(1-x) - 2(1-x)^2) = -4x - 6 + 12x + 4 - 4x = 4x - 2$$

which is minimum when $x = \frac{1}{2}$. In this case:

$$J = \mathbb{E}[p] = \frac{1}{4}(-2-3-3-2) = -\frac{10}{4} = -\frac{5}{2}.$$

Now, we take $f : S \times S \to S$ defined as follows:

$$f(0,0) = 1, f(1,0) = 1, f(0,1) = 0, f(1,1) = 0 \tag{2.23}$$

So:

$$\begin{aligned}\mathbb{E}[\tilde{p}] =& \frac{1}{4}(g(f(0,0),0) + g(f(1,0),0) + g(f(0,1),1) + g(f(1,1),1)) \\ =& \frac{1}{4}(g(1,0) + g(1,0) + g(0,1) + g(0,1))\frac{1}{4}(-3-3-3-3) = -3 < -\frac{5}{2}\end{aligned}$$

The next step is to find a correlated equilibrium, which is not also a Nash one, whose f-deviation, for a suitable f, decreases the cost. We will consider f as in (2.23).
A simple correlated equilibrium $\tilde{p}$ is

1/2	0	1
0	0	$\frac{1}{2}$
1	$\frac{1}{2}$	0

Here $\mathbb{E}\left[g(f(\alpha_1, \alpha_2), \alpha_2)\right] = \mathbb{E}\left[g(\alpha_1, \alpha_2)\right]$. By Proposition 1.7, we now that any linear combination of correlated equilibria is still a correlated equilibrium, we take $p' = \frac{1}{2}p + \frac{1}{2}\tilde{p}$, which is

1/2	0	1
0	$\frac{1}{8}$	$\frac{3}{8}$
1	$\frac{3}{8}$	$\frac{1}{8}$

whose cost is $J(p') = \frac{1}{8}(-2-2) + \frac{3}{8}(-3-3) = -\frac{22}{8}$. Applying f we get, $J(f(p')) = \frac{1}{8}(g(f(0,0),0) + g(f(1,1),1)) + \frac{3}{8}(g(f(0,1),1) + g(f(1,0),0)) = \frac{1}{8}(-3-3) + \frac{3}{8}(-3-3) = -3 < -\frac{22}{8}$. If we now show that p' is not a Nash equilibrium, we would have reached our purpose; if it is so, we call p_i the probability of the ith player of choosing 0, hence: $\frac{1}{8}p_1 p_2 = (1-p_1)(1-p_2) = 1 - p_1 p_2 + p_1 p_2 \Rightarrow p_1 = 1 - p_2$, $\frac{3}{8} = (1-p_1)p_2 = p_1(1-p_2) \Rightarrow p_1 = p_2 \Rightarrow p_1 = p_2 = \frac{1}{2}$; but $\frac{1}{4} \neq \frac{1}{8}$.

Summarizing, we could say that these observations lead us to the following conclusions:

- the intuition is that (2.22) does not work in the correlated case, even if we manage to obtain it in the limit case of Nash. Indeed, we frame this situation in the very special case of μ a.s. constant and the examples we have given go in the opposite direction.
- Another idea could be that of defining at the beginning a correlated equilibrium as a distribution which minimizes the cost even for modification of the form $f : \mathcal{X} \times \mathcal{P}(\mathcal{X}) \to \mathcal{X}$, but this should be coherent with the concept Nash equilibrium, which is not since we have shown that such a modification is not allowed in that case.

Appendix A

Convergence of measures

A.1 Weak Convergence

Consider a metric space (S, d) and a sequences of probability measures $(P_n)_{n\in\mathbb{N}}$ in S. We'll indicate the space of the probability measures over S with $\mathcal{P}(S)$.

Definition A.1. We say that $(P_n)_n$ **converges weakly** to $P \in \mathcal{P}(S)$ if, for every continuous bounded mapping $f : S \to \mathbb{R}$

$$\int_S f dP_n \xrightarrow{n\to\infty} \int_S f dP.$$

Consider a sequence of random variable $X_{nn\ in\mathbb{N}}$ which take values is a metric space (S, d); denote with P_n their distribution. Observe that the probability space on which they're defined could be different.

Definition A.2. We say that $\{X_n\}_{n\in\mathbb{N}}$ **converges in distribution** to the random element X, whose distribution is P, if $P_n \xrightarrow{n\to\infty} P$ weakly.

Explicitly we can write that $X_n \to X$ in distribution if, for every $f \in \mathcal{C}_b(S)$

$$\begin{aligned}\int_S f(x)P_n(dx) &= \int_{\Omega_n} f(X_n(\omega))\mathbb{P}(d\omega) = \mathbb{E}_n[f(X_n)] \to \mathbb{E}[f(X)] \\ &= \int_{\Omega} f(X(\omega))\mathbb{P}(d\omega) = \int_S f(x)P(dx)\end{aligned}$$

Definition A.3. A σ-finite measure μ on $(\mathcal{S}, S)$ is called

1. **inner regular** if

$$\mu(A) = \sup\{\mu(C) \mid C \subset A \text{ is closed}\} \quad \text{for all } A \in \mathcal{S} \tag{A.1}$$

2. **outer regular** if

$$\mu(A) = \sup\{\mu(G) \mid G \supset A \text{ is open}\} \quad \text{for all } A \in \mathcal{S} \tag{A.2}$$

3. **regular** if it's both inner and outer regular.

Proposition A.4. Let (S, d) be a metric space with the Borel σ-algebra $\mathcal{S}$. Then every probability measure $P \in \mathcal{P}(S)$ is regular.

Proof. Define

$$\mathcal{G} := \{A \in \mathcal{S} \mid \mu(A) = \sup\{\mu(C) \mid C \subset A \text{ is closed }\} = \sup\{\mu(G) \mid G \supset A \text{ is open}\}\}.$$

If we show that $\mathcal{G}$ is a σ-algebra that contains all the closed sets, we've finished. Let A be closed. It's clear that the first inequality is satisfied. In order to verify the second one, we define $G_n := \{x \in S \mid d(x, A) < \frac{1}{n}\}$, which is a system of open neighbourhood of A. Since A is closed, $\bigcap_{n\in\mathbb{N}_{>0}} G = A$ and because of the upper semi-continuity of P, $\inf_n P(G_n) = P(\bigcap_{n\in\mathbb{N}_{>0}}) = P(A)$, which proves the second equality.
We now show that $\mathcal{G}$ is a σ-algebra. It's clear that $S \in \mathcal{G}$ and that $\mathcal{G}$ is stable under complementation. Let (A_n) be a sequence of elements in $\mathcal{G}$ and let $A := \bigcup_{n\in\mathbb{N}} A_n$. By definition, for every n there exists a closed set $C_n \subset A_n$ and an open set $G_n \supset A_n$ such that $P(C_n) \geq P(A_n) - \frac{\varepsilon}{2^n}$ and $P(G_n) \leq P(A_n) + \frac{\varepsilon}{2^n}$; thus, $P(G_n \setminus A_n) \leq \frac{\varepsilon}{2^n}$ and $P(A_n \setminus C_n) \leq \frac{\varepsilon}{2^n}$. We set $G := \bigcup_{n\in\mathbb{N}} G_n$ and $C := \bigcup_{n=1}^{n_0} C_n$ where n_0 is such that $P(\bigcup_{n\in\mathbb{N}} C_n \setminus C) \leq \varepsilon$. Thus G and F are respectively an open and a closed set such that $C \subset A \subset G$.

$$\left.\begin{array}{l} P(G \setminus A) \leq P(\bigcup_{n\in\mathbb{N}}(G_n \setminus A_n)) \leq \sum_{n\in\mathbb{N}} P(G_n \setminus A_n) = \varepsilon \\ P(A \setminus C) \leq P(\bigcup_{n\in\mathbb{N}} C_n \setminus C) + P(A \setminus \bigcup_{n\in\mathbb{N}} C_n) \leq 2\varepsilon \end{array}\right\} \Rightarrow A \in \mathcal{G}.$$

□

Corollary A.5. Let $P, \tilde{P}$ be two probability measures over (S, d), then:

1. if $P(C) = \tilde{P}(C)$ for all C closed (or open) subset of S, then $P = \tilde{P}$.
2. if $\int_S f dP = \int_S f d\tilde{P}$ for every $f \in \mathcal{C}(S)$, then $P = \tilde{P}$.

Proof. 1. Direct consequence of Proposition A.4
2. Let $C \in \mathcal{S}$ be closed. We define a function $\varphi\mathbb{R} \to [0, 1]$ by

$$\varphi(x) := \begin{cases} 1 & \text{if } x \leq 0 \\ 1 - x & \text{if } 0 \leq x \leq 1 \\ 0 & \text{if } x \geq 1 \end{cases}$$

For every $n \geq 1$ and every $x \in S$, let $f_n(x) := \varphi(nd(x, C))$. Now, $(f_n)_n$ is a decreasing sequence of continue functions over S, whose limit is 1_C. By dominated convergence, $\int_S f_n dP \to P(C)$ and $\int_S f_n d\tilde{P} \to \tilde{P}(C)$. By 1., we conclude.

□

Remark A.6. Observe that, as a consequence of Corollary A.5, the weak limit is unique.

Definition A.7. Let $\Pi \subset \mathcal{P}(S)$ be a family of probability measures. A family $\mathcal{C}$ of measurable maps $S \to \mathbb{R}$ is called a **separating family** for Π if, for any two measures $P, \tilde{P} \in \Pi$

$$\int_S f dP = \int_S f d\tilde{P} \text{ for all } f \in \mathcal{C} \cap L^1(P) \cap L^1(\tilde{P}) \Rightarrow P = \tilde{P}$$

Proposition A.8. Let $\mathcal{C}$ be a dense subset of $\mathcal{C}_b(S)$ equipped with the sup norm $\|\cdot\|_\infty$, then $\mathcal{C}$ is a separating family for $\mathcal{P}(S)$.

Proof. We would like to prove that if for every $f, \tilde{f} \in \mathcal{C}$ $\int_S f dP = \int_S \tilde{f} d\tilde{P}$, then for every $g, \tilde{g} \in \mathcal{C}_b(S)$ $\int_S g dP = \int_S \tilde{g} d\tilde{P}$. Since $\mathcal{C}$ is dense in $\mathcal{C}_b(S)$, taken $g, \tilde{g} \in \mathcal{C}_b(S)$, we have that for any $\varepsilon > 0$, there exist $f, \tilde{f} \in \mathcal{C}$ such that

$$\|g - f\|_\infty \leq \varepsilon \qquad \|\tilde{g} - \tilde{f}\|_\infty \leq \varepsilon \tag{A.3}$$

Now

$$\begin{aligned}
\int_S g dP &= \int_S (g-f) dP + \int_S f dP - \int_S \tilde{f} d\tilde{P} + \int_S (\tilde{f} - \tilde{g}) d\tilde{P} + \int_S \tilde{g} d\tilde{P} \\
&\leq \int_S \|g - f\|_\infty dP + \int_S f dP - \int_S \tilde{f} d\tilde{P} + \int_S \|\tilde{f} - \tilde{g}\| d\tilde{P} + \int_S \tilde{g} d\tilde{P} \\
&\leq \|g - f\|_\infty + \int_S f dP - \int_S \tilde{f} d\tilde{P} + \|\tilde{f} - \tilde{g}\| + \int_S \tilde{g} d\tilde{P} \leq 2\varepsilon + \int_S \tilde{g} d\tilde{P}
\end{aligned}$$

Were the last inequality is a consequence of (A.3) and the fact that $f, \tilde{f} \in \mathcal{C}$ $\int_S f dP = \int_S \tilde{f} d\tilde{P}$. Analogously, we show that $\int_S \tilde{g} d\tilde{P} \leq 2\varepsilon + \int_S g dP$. By the arbitrariness of ε, we have that for every $g, \tilde{g} \in \mathcal{C}_b(S)$ $\int_S g dP = \int_S \tilde{g} d\tilde{P}$. Thanks to Corollary A.5, we conclude. □

Remark A.9. Weak convergence induces on $\mathcal{P}(S)$ the *weak topology* τ_w; this is the coarsest topology such that for all $f \in \mathcal{C}_b(S)$, the map $\mathcal{P}(S) \to \mathbb{R}, P \mapsto \int_S f dP$ is continuous.

We now introduce two distances in $\mathcal{P}(S)$, proving their equivalence by showing that they induce the same topology τ_w over $\mathcal{P}(S)$.
The first one is the **Bounded Lipschitz distance**, defined as follows:

$$\rho_{bL}(P, Q) := \sup\{\int_A f d(P - Q) \mid f : A \to \mathbb{R}, \|f\|_{bL} \leq 1\} \tag{A.4}$$

where

$$\|f\|_{bL} = \sup_{a \in A} |f(a)| + \sup_{a, \tilde{a} \in A} \frac{|f(a) - f(\tilde{a})|}{d_A(a, \tilde{a})} = \|f\|_\infty + \mathrm{Lip}(f)$$

$(\mathcal{P}(S), \tau_w)$ is metrized by bounded Lipschitz distance; indeed, let us consider $\{P_n\}$ and P such that $\rho_{bL}(P_n, P) \to 0$ as $n \to \infty$. Then for every f bounded and Lipschitz, $\int_A f dP_n \to \int_A f dP$, with $\|f\|_{bL} \leq 1$ and, by linearity, that would be true for any Lipschitz bounded function (indeed, if h is m-Lip., then h/m is 1-Lip. and if $\|h\|_\infty = c$ then, $(\|h\|_\infty)/c \leq 1$). Now, since bounded Lipschitz function are dense in $\mathcal{C}_b(A)$, by Proposition A.3, we have $\int_A f dP_n \to \int_A f dP$ for any $f \in \mathcal{C}_b(A)$, that is the definition of weak convergence. The other implication is trivial since any Lipschitz bounded function is also continuous and bounded.
The second one is the **Prokhorov distance**, defined as follows:

$$\rho(P, Q) := \inf\{\varepsilon \geq 0 \mid P(A) \leq Q(A^\varepsilon) + \varepsilon, Q(A) \leq P(A^\varepsilon) + \varepsilon \text{ for all } A \in \mathcal{S}\} \tag{A.5}$$

where $A^\varepsilon := \{x \in S \mid d(x, A) < \varepsilon\}$. Observe that for $\varepsilon \searrow 0, A^\varepsilon \searrow \bar{A}$.

Remark A.10. • If $\varepsilon \geq \rho(P,Q)$, then for every $A \in \mathcal{S}, P(A) \leq Q(A^\varepsilon) + \varepsilon$.

• (A.5) is equivalent to

$$\rho(P,Q) = \inf\{\varepsilon \geq 0 \mid P(A) \leq Q(A^\varepsilon) + \varepsilon \text{ for all } A \in \mathcal{S}\}$$

indeed, if we suppose that $P(A) \leq Q(A^\varepsilon)+\varepsilon$ for every $A \in \mathcal{S}$, we define $B = (A^\varepsilon)^C \in \mathcal{S} \Rightarrow P(B) \leq Q(B^C)+\varepsilon \Rightarrow 1-P(A^\varepsilon) \leq 1-Q(A^\varepsilon)+\varepsilon \Rightarrow Q(A) \leq Q(A^\varepsilon) \leq P(A^\varepsilon)+\varepsilon$.

$\rho(P_n, P) \to 0 \Rightarrow P_n \to P$ weakly: for every n, we set $\varepsilon_n := \rho(P_n, P) + \frac{1}{n}$. By observation A.10, $\varepsilon_n \geq \rho(P_n, P) \Rightarrow P_n(A) \leq P(A^\varepsilon) + \varepsilon$. Let $f \in \mathcal{C}_b(S)$, $M := \|f\|_\infty$

$$\int_S f dP_n = \int_0^M P_n(\{f \geq a\})da \leq \int_0^M (P(\{f \geq a\}^{\varepsilon_n} + \varepsilon_n)da = \int_0^M P(\{f \geq a\}^{\varepsilon_n})da + M\varepsilon_n$$

For $n \to \infty, \rho(P_n, P) \to 0 \Rightarrow \varepsilon_n \to 0$ and so $\{f \geq a\}^\varepsilon_n \searrow \{f \geq a\}$ (because $\{f \geq a\}^\varepsilon_n$ is closed). By dominated convergence

$$\lim_{n\to\infty} \int_S f dP_n \leq \lim_{n\to\infty} \int_0^M P(\{f \geq a\}^{\varepsilon_n})da + M\varepsilon_n = \int_0^M P(\{f \geq a\})da = \int_S f dP.$$

This implies that $\limsup_{n\to\infty} \int_S f dP_n \leq \int_S f dP$. Now, for every $f \in \mathcal{C}_b(S)$, define $g := 1 - f \in \mathcal{C}_b(S)$, so $\liminf_{n\to\infty} \int_S f dP_n = \liminf_{n\to\infty} \int_S (1-g)dP_n = 1 - \limsup_{n\to\infty} \int_S f dP_n \geq 1 - \int_S f dP = \int_S f dP$. Thus, we conclude that $\lim_{n\to\infty} \int_S f dP_n = \int_S f dP$ for every bounded continuous function.
$P_n \to P \Rightarrow \rho(P_n, P) \to 0$ (if S is separable): Let $\varepsilon > 0$. Consider a $\mathcal{S}$-partition of S $\{E_k\}_k$ such that $\operatorname{diam}(E_k) < \varepsilon$ for all k. Let be N such that $P(\bigcup_{i\geq N} E_i) \leq \varepsilon$. Define $\mathcal{G} := \{G = (\bigcup_{i\in I} E_i)^\varepsilon \mid I \subset \{1, \dots, N\}\}$. Since $\#\mathcal{G} < \infty$, as a consequence of portmanteau theorem (Theorem A.11), there exist n_0 such that $P(G) \leq P_n(G)+\varepsilon$ for all $G \in \mathcal{G}$ and for all $n \geq n_0$ (it's sufficient to take the maximum over the n_0 that verifies this equality for the singles $G \in \mathcal{G}$). Let $E \in \mathcal{S}$, consider the set J of all the $i \leq N$ such that $E \cap E_i \neq \emptyset$ and define $E_0 := \bigcup_{i\in J} E_i$. For $n \geq n_0$, by noting that $E_0 \in \mathcal{G}$, we have:

$$\begin{aligned} P(E) &= P(E \cap \bigcup_{i\leq N} E_i) + P(E \cap \bigcup_{i\geq N} E_i) \leq P(E \cap \bigcup_{i\in J} E_i) + P(\bigcup_{i\geq N} E_i) \\ &\leq P(E_0) + \varepsilon \leq P(E_0^\varepsilon) + \varepsilon \leq P_n(E_0^\varepsilon) + \varepsilon + \varepsilon \leq P_n(E^{2\varepsilon}) + 2\varepsilon \end{aligned}$$

The last inequality is justified by the fact that $\operatorname{diam}(E_i) < \varepsilon$, $E_i \cap E \neq \emptyset$ for $i \in J$ and thus $E_i \subset E^\varepsilon$. We've found that, for any $n \geq n_0$, $\rho(P_n, P) \leq 2\varepsilon$.

We say that a set $A \in \mathcal{S}$ is a *P-continuity set* if $P(\partial A) = 0$. Observe that for every A, $\partial A = \bar{A} \setminus \mathring{A}$ is always measurable because it is closed.

Theorem A.11 (Portmanteau)**.** The following are equivalent:

(i) $P_n \to P$ weakly.

(ii) For all bounded and Lipschitz f, $\int_S f dP_n \to \int_S f dP$.

(iii) For every closed set $C \subset S$, $\limsup_{n\to\infty} P_n(C) \leq P(C)$.

(iv) For every open set $G \subset S$, $\liminf_{n\to\infty} P_n(G) \geq P(G)$.

(v) For every P-continuity set A, $\lim_{n\to\infty} P_n(A) = P(A)$.

Proof. (i) ⇒ (ii) and (iii) ⇔ (iv) are trivial.
(ii) ⇒ (iii): Define $f_k(x) := \varphi(kd(x, C))$ as in the proof of Corollary A.5. We have that f_k is Lipschitz because φ and d are too and that $f_k \searrow \mathbf{1}_C$ simply. By hypothesis, for every k, $\int_S f_k dP_n \xrightarrow{n\to\infty} \int_S f_k dP$. But also, since $f_k \geq \mathbf{1}_C$, we have $\limsup_{n\to\infty} P_n(C) = \limsup_{n\to\infty} \int_S \mathbf{1}_C dP_n \leq \limsup_{n\to\infty} \int_S f_k dP_n = \int_S f_k dP$. By the dominated convergence theorem when $k \to \infty$, we conclude.
(iii) + (iv) ⇒ (v): Let A be a P-continuity set, then:

$$\limsup_{n\to\infty} P_n(A) \leq \limsup_{n\to\infty} P_n(\bar{A}) \leq P(\bar{A}) = P(\mathring{A}) \leq \liminf_{n\to\infty} P_n(\mathring{A}) \leq \liminf_{n\to\infty} P_n(A).$$

(v) ⇒ (i): Let $f : S \to \mathbb{R}$ be continuous and bounded. Without loss of generality, we can suppose f with values over $\mathbb{R}_+$, otherwise we consider separately the positive and negative parts. Let $A := \{(x, a) \in S \times \mathbb{R}_+ \mid f(x) > a\} = \bigcup_{r\in\mathbb{Q}_+} (\{f > r\} \times [0, r]) \in \mathcal{S} \otimes \mathcal{B}(\mathbb{R}_+)$. If we indicate by λ the Lebesgue measure and $M := \|f\|_\infty < \infty$, we have

$$\left.\begin{array}{l} (P \otimes \lambda)(A) = \displaystyle\int_{S\times\mathbb{R}_+} \mathbf{1}_A d(P \otimes \lambda) = \int_{\mathbb{R}_+} P(\{f > a\})da \\ (P \otimes \lambda)(A) = \displaystyle\int_S f dP \end{array}\right. \Rightarrow \int_{\mathbb{R}_+} P(\{f > a\})da = \int_S f dP$$

Thus, $\int_{\mathbb{R}_+} P_n(\{f > a\})da = \int_S f dP_n = \int_0^M P_n(\{f > a\})da$. Since f is continuous, $\partial\{f > a\} \subset f = a$. The set $\{a \geq 0 \mid P(\{f = a\}) > 0\}$ is at most countable, indeed, for a fixed n, there are at most n a such that $P(\{f = a\}) \geq \frac{1}{n}$, otherwise $1 = P(S) = \sum_{a\in S} P(\{f = a\}) \geq \frac{1}{n}\#\{a \mid P(\{f = a\}) \geq \frac{1}{n}\}$. Now, for λ-a.e. a, $P(\partial\{f > a\}) = 0$, and so, for λ-a.e. a $P_n(\{f > a\}) \to P(\{f > a\})$. By dominated convergence, $\int_0^M P_n(\{f > a\})da \to \int_0^M P(\{f > a\})da$, which means that $\int_S f dP_n \to \int_S f dP$. □

Before proceeding further in the analysis of the convergence, we give some topological results about S and that link the properties of the space S with those of $\mathcal{P}(S)$.

Definition A.12.
- $A \subset S$ is **relatively compact** if $\bar{A}$ is compact.
- $A \subset S$ is **totally bounded** if, for every $\varepsilon > 0$, there exist $x_1, \ldots, x_N \in S$ such that $A \subset \bigcup_{n=1}^N B(x_n, \varepsilon)$.

Proposition A.13.
- If $A \subset S$ is relatively compact, then A is totally bounded.
- If S is complete, then A relatively compact if and only if A is totally bounded.

Proof.
- Consider $\varepsilon > 0$ and an open covering $\bar{A} \subset \bigcup_{n=1}^\infty B(x_n, \varepsilon)$, then there exist a finite sub-covering $\bar{A} \subset \bigcup_{n=1}^N B(x_n, \varepsilon) \Rightarrow A \subset \bar{A} \subset \bigcup_{n=1}^\infty B(x_n, \varepsilon)$.
- Let S be complete and A be totally bounded. Observe that A totally bounded ⇒ $\bar{A}$ totally bounded, indeed, for $\varepsilon > 0$, there exist a finite covering $A \subset \bigcup_{n=1}^N B(x_n, \frac{\varepsilon}{2})$, since a finite union of closed sets is closed, we have $\bar{A} \subset \bigcup_{n=1}^N \bar{B}(x_n, \frac{\varepsilon}{2}) \subset \bigcup_{n=1}^N B(x_n, \varepsilon)$.

$\bar{A}$ totally bounded implies that every sequence has a Cauchy sub-sequence, indeed, if $(x_n)_n$ is a sequence, than it has a subsequence contained in a ball of radius $\frac{1}{2}$ (because $\bar{A} \subset \bigcup_{n=1}^{N} B(y_n, \varepsilon)$ for every ε) then this sub-sequence is contained in a ball of radius $\frac{1}{3}$ and so on. By a diagonal argument, take the first element of these sub-sequences and call the new sequence $(x_{n_k})_k$. If $k > h$, $d(x_{n_k}, x_{n_h}) < \frac{2}{k+1} \xrightarrow{k\to\infty} 0 \Rightarrow (x_{n_k})_k$ is Cauchy sub-sequence.
Now $\bar{A}$ is totally bounded and complete. Thus, for every sequence, there is a Cauchy sub-sequence that is convergent since $\bar{A}$ is complete $\Rightarrow$ $\bar{A}$ is sequentially compact and thus compact.

□

Proposition A.14. If (S, d) is a compact metric space, then S is also separable.

Proof. Consider the following open covering:

$$I_n = \left\{ B(x, \frac{1}{n}) | x \in S \right\}$$

since S is compact, each I_n has a countable sub-covering $J_n = \{B(x_m^n, \frac{1}{n}) | m \in \mathbb{N}, x_m \in S\}$. Now $A := \{x_m^n | m, n \in \mathbb{N}\}$ is a countable subset of S, which is dense, indeed, since J_n are coverings for S for all n, taken an $x \in S$, for all $n \in \mathbb{N}$, there exists a $m \in \mathbb{N}$ such that $d(x, x_n^m) < \frac{1}{n}$. This implies that S is separable. □

Proposition A.15. S compact metric space $\Rightarrow$ $\mathcal{P}(S)$ compact.

Proof. We've seen that with Bounded Lipschitz distance (A.4), $\mathcal{P}(S)$ is a metric space (with the τ_w topology), thus, compactness and sequential compactness are here equivalent; we will show in fact that $\mathcal{P}(S)$ is sequentially compact, that is for any sequence $P_n \in \mathcal{P}(S)$, $\{P_n\}$ has a convergent subsequence. For convenience, we shall write $P(f) := \langle P, f \rangle = \int_S f dP$.
Because of the separability of $\mathcal{C}(\mathcal{S})$, we can choose a countable dense subset, say $\{f_i\}_{i\in\mathbb{N}_{>0}}$. Consider the sequence $P_n(f_1) \in \mathbb{R}$ of real numbers. We have $|P_n(f_1)| = |\int_S f_1 dP_n| \leq \|f_1\|_\infty < \infty$ for all n, so $P_n(f_1)$ is a bounded sequence of real numbers $\Rightarrow$ it has a convergent subsequence $P_n^{(1)}(f_1)$.
Consider the sequence $P_n^{(1)}(f_2)$; as before, it's a bounded sequence of real numbers, so it has a convergent subsequence $P_n^{(2)}(f_2)$.
Iterating this process, we obtain, for each $i \geq 1$, nested sub-sequences $\{P_n^{(i)}\} \subseteq \{P_n^{(i-i)}\}$, such that $P_n^{(i)}(f_j)$ converges for $1 \leq j \leq i$. Consider now the diagonal sequence $\{P_n^{(n)}$. Since, for $n \geq i$ $P_n^{(n)}$ is a subsequence of $P_n^{(i)}$ and $P_n^{(i)}(f_i)$ converges for all i, then $P_n^{(n)}(f_i)$ converges for every $i \geq 1$. By the density of $f_{i i\in\mathbb{N}_{>0}}$, we have that $P_n^{(n)}(f)$ converges for all $f \in \mathcal{C}(\mathcal{S})$.
We write $w(f) = \lim_{n\to\infty} P_n^{(n)}(f)$. to complete the proof we need to show that there exists a probability measure P such that $w(f) = \int_S f dP$; to do so we'll use Riesz Representation Theorem.

Theorem A.16 (Riesz Representation Theorem)**.** Let S a locally compact Hausdorff space. For any positive bounded linear functional ψ on $\mathcal{C}_b(S)$, such that $\psi(1) = 1$, then there is a unique Borel probability measure P on S such that

$$\psi(f) = \int_S f(x)dP(x)$$

for all $f \in \mathcal{C}_b(S)$.

- w is a linear mapping by construction.
- $|w(f)| \leq \|f\|_\infty$, so w is bounded.
- If $f \geq 0$, $w(f) \geq 0$, so w is positive.
- $w(1) = \lim_{n\to\infty} P_n^{(n)}(1) = \lim_{n\to\infty} 1 = 1$.

We can apply Theorem A.16 and obtain a measure $P \in \mathcal{P}(S)$ such that $w(f) = \int_S f dP$. This means that $P_n^{(n)}$ converges weakly to P. □

The following theorem extends in a certain sense the continuous mapping theorem, which guaranties the weak convergence when we compose weakly convergent sequences of probability measures with continuous functions. Consider h_n and h measurable mappings from (S, d) to (S', d'), two separable spaces, and consider $P_n, P \in \mathcal{P}(S)$ such that $P_n \to P$ weakly. Suppose that h_n converges to h in some sense. Define $E := \{x \in S | h_n(x_n) \nrightarrow h(x)$ for some sequnce $x_n \to x\}$; now $x \in E^C$ if and only if $\forall\ \varepsilon > 0\ \exists\ k\ \exists\ \delta > 0$ such that $n \geq k,\ d(x, y) < \delta \Rightarrow d'(h(x), h_n(y)) < \varepsilon$.

Theorem A.17 (Continuous mapping theorem)**.** Assume $P_n \to P$ weakly. Assume $h : S \to S'$ to be a measurable function between two metric space, and assume that its discontinuity set, D_h, is such that $P(D_h) = 0$; then $P_n \circ h_n^{-1} \to P \circ h^{-1}$ weakly.

Theorem A.18 (Continuous mapping theorem (Extended version))**.** If $P_n \to P$ weakly and $P(E) = 0$, then $P_n \circ h_n^{-1} \to P \circ h^{-1}$ weakly.

Proof. Using Portmanteau Theorem, we shall show that $P(h^{-1}(G)) \leq \liminf_{n\to\infty} P_n(h_n^{-1}(G))$ for any $G \subset S'$ open. Now, let us take $x \in E^C$ such that $h(x) \in G$. Thus, $h(x) \in G \Rightarrow \exists \varepsilon' > 0$ s.t. $B'(h(x), \varepsilon') \subset G \Rightarrow \forall\ y$ s.t. $d'(h(x), y) < \varepsilon' \Rightarrow y \in G$; $x \in E^C \Rightarrow \forall\ \varepsilon > 0\ \exists\ k\ \exists\ \delta > 0$ such that $n \geq k,\ d(x, y) < \delta \Rightarrow d'(h(x), h_n(y)) < \varepsilon$. Take $\varepsilon := \varepsilon' \Rightarrow h_n(y) \in G\ \forall n \geq k,\ \exists k,\ \forall y$ s.t. $d(x, y) < \delta\ \exists \delta$.
Define $T_k := \bigcap_{n\geq k} h_n^{-1}(G)$. So $\{x \in E^C \mid h(x) \in G\} \subseteq \mathring{T}_k$, thus $h^{-1}(G) \subset E \cup \bigcup_k \mathring{T}_k$. Since $P(E) = 0$, $P(h^{-1}(G)) \leq P(\bigcup_k \mathring{T}_k)$; but $\mathring{T}_k \subset \mathring{T}_{k+1}$, so by the continuity from below $\lim_{k\to\infty} P(\mathring{T}_k) = P(\bigcup_k \mathring{T}_k)$, thus for a given $\varepsilon > 0$ and for k sufficiently large $P(h^{-1}(G)) < P(\bigcup_k \mathring{T}_k) + \varepsilon$. From $P_n \to P$ weakly, it follows that $P(\mathring{T}_k) \leq \liminf_{n\to\infty} P_n(h_n^{-1}(G)) \Rightarrow P(h^{-1}(G)) \leq \liminf_{n\to\infty} P_n(h_n^{-1}(G)) + \varepsilon$, which concludes the proof because ε is arbitrary. Observe that we don't have problems of mesurability, because $P(E) = 0$ so it's measurable, and all the T_k are countable intersection of reverse-images of measurable functions. □

Remark A.19.
- In Theorem A.18, if $h_n = h$ for all n, this results reduces to Theorem A.17, indeed in such a case $E = D_h$.

- If h is everywhere continuous and $h_n \to h$ uniformly on compact sets, then E is empty and so, the hypothesis of Theorem A.18 are satisfied.

A.1.1 Tightness

Definition A.20.
- A probability P on A is said to be **tight** if, for every $\varepsilon > 0$, there exists a compact subset $K \subseteq A$ such that $P(K^c) \leq \varepsilon$.
- A family of probability Π on A is said to be **tight** if, for every $\varepsilon > 0$, there exist a compact subset $K \subseteq A$ such that $P(K^c) \leq \varepsilon$ for every $P \in \Pi$.

Remark A.21. We observe that in Definition A.20, it is sufficient for K to be relatively compact, because, in this case, we can take $\bar{K}$ instead of K.

Definition A.22. Let Π be a family of probability measures over S. We say that Π is **relatively compact** if, for every sequence $\{P_n\}$ of Π, we can extract a sub-sequence $\{P_{n_k}\}$ which is weakly convergent.

Remark A.23. If S is separable we know that weak convergence is equivalent to the convergence under the Prokhorov distance. Thus A.22 coincides with the usual notion of sequentially relative compactness on $(\mathcal{P}(S), \rho)$.

Proposition A.24. If Π is a tight family on (A, d) and if $h : A \to A'$ is a continuous mapping, then, $\{P \circ h^{-1} | P \in \Pi\}$ is a tight family on (A, d').

Proof. Given ε, choose in A a compact set K such that $P(K) > 1 - \varepsilon$ for any $P \in \Pi$. Then, $K' := h(K)$ is compact since h is continuous and $h^{-1}(K') \subseteq K$, so $P(h^{-1}(K')) \geq P(K) > 1 - \varepsilon$. □

Theorem A.25 (Ulam)**.** If S is Polish, then every probability $P \in \mathcal{P}(S)$ is tight.

Proof. Let $(x_n)_n$ be a dense sequence in S and let $\varepsilon > 0$. For every $f \geq 1$, $S = \bigcup_{n=1}^{\infty} B(x_n, \frac{1}{k})$, thus, there exist a $N_k < \infty$ such that $P\Big(\bigcup_{n=1}^{N_k} B(x_n, \frac{1}{k})\Big) \geq 1 - 2^{-k}\varepsilon$. Let $K := \bigcap_{k=1}^{\infty} \bigcup_{n=1}^{N_k} B(x_n, \frac{1}{k})$ that is totally bounded and thus relatively compact, because S complete (Proposition A.13).
Since $P(K^C) \leq P\Big(\bigcup_{k=1}^{\infty} (\bigcup_{n=1}^{N_k} B(x_n, \frac{1}{k}))^C\Big) \leq \sum_{k=1}^{\infty} \frac{\varepsilon}{2^k} = \varepsilon$, P is tight. □

Consider now

Theorem A.26 (Prokhorov)**.** Let Π be a family of probability measures on S. Then:

- If Π is tight, then it is also relatively compact.
- Suppose S separable and complete. If Π is relatively compact, than it is also tight.

Proof. Since the proof of the first point is out of the scope of these thesis we will omit it.

- For the full proof see [4].

- Let $\varepsilon > 0$. Since Π is relatively compact on $(\mathcal{P}(S), \rho)$ and thus totally bounded (look at Remark A.21 and Proposition A.13). Hence, there exist, for every $n \geq 1$, $Q_1, \dots, Q_{N_n} \in \mathcal{P}(S)$ such that for all $P \in \Pi$, $\exists\, i* = i*(n, P) \leq N_n$ with $\rho(P, Q_{i*}) < \frac{\varepsilon}{2^{n+1}}$.

  Since S is Polish, by Ulam Theorem, for every n, there exists $K_n \subset S$ compact such that $\max_{1 \leq n \leq N_n} Q_i(K_n^c) \leq \frac{\varepsilon}{2^{n+1}}$. So, for any $P \in \Pi$:

$$\begin{aligned} P(K_n^{\frac{\varepsilon}{2^{n+1}}}) &\geq Q_{i*}(K_n) - \frac{\varepsilon}{2^{n+1}} \qquad \text{because } \rho(P, Q_{i*}) < \frac{\varepsilon}{2^{n+1}} \\ &\geq 1 - \frac{\varepsilon}{2^n} \end{aligned}$$

  Let $K := \bigcap_{n \geq 1} K_n^{\frac{\varepsilon}{2^{n+1}}}$, which is relatively compact, because totally bounded, and $P(K) \geq 1 - \sum_{n \in \mathbb{N}} \frac{\varepsilon}{2^n} = 1 - \varepsilon \Rightarrow \Pi$ is tight.

□

Corollary A.27. If a sequence of probabilities $(P_n)_n$ is tight and if all convergent sebsequences (in the sense of weak convergence) have the same limit P, than P_n converge weakly to P.

Proof. Since P_n is tight, Prokhorov Theorem tells us that all subsequences have a subsubsequence that converges weakly, so its limit is forcedly P by hypothesis. So for all functions f continues and bounded all subsequences $\left(\int f dP_{n_k}\right)$ have a sub-subsequence $\left(\int f dP_{n_{k_l}}\right)$ who converges toward $\int f dP$. So $\int f dP_n \to \int f dP$, but that's the definition of weakly convergence of P_n. □

Proposition A.28.
- If S is separable, then $\mathcal{P}(S)$ is separable.
- If S is Polish, then $\mathcal{P}(S)$ is Polish.

Proof.
- Fix $n \in \mathbb{N}$. Let $\{E_i\}_i$ be a $\mathcal{S}$-partition of S such that $\operatorname{diam}(E_i) < \frac{1}{n}$ for all i. If $E_i \neq \emptyset$, choose $x_i \in E_i$. Define

$$\Pi_n := \{\sum_{i \leq k} r_i \delta_{x_i} \mid k \geq 1, r_i \mathbb{Q}\} \subset \mathcal{P}(S)$$

  Observe that Π_n is countable for all n.

  Given $P \in \mathcal{P}(S)$, choose k such that $P(\bigcup_{i > k} E_i) < \frac{1}{n}$; such a k exists because $\{E_i\}$ is a partition of S. Choose $r_1, \dots, r_k$ such that

$$\sum_{i \leq k} r_i = 1 \qquad \sum_{i \leq k} |r_i - P(E_i)| < \frac{1}{n}$$

  Put $Q := \sum_{i \leq k} r_i \delta_{x_i}$.

Given $E \subset S$, define $J := \{i \leq k \mid E_i \bigcap E \neq \emptyset\}$. $E_0 := \bigcup_{i \in J}$, then:

$$\begin{aligned} P(E) &= P(E \cap \bigcup_{i \leq k} E_i) + P(E \cap \bigcup_{i > k} E_i) = P(E \cap \bigcup_{i \in J} E_i) + P(E \cap \bigcup_{i > k} E_i) \\ &\leq P(\bigcup_{i \in J} E_i) + P(\bigcup_{i > k} E_i) \leq P(E_0) + \frac{1}{n} = \sum_{i \in J} P(E_i) + \frac{1}{n} \\ &\leq \sum_{i \in J} |P(E_i) - r_i| + \sum_{i \in J} r_i + \varepsilon \leq \sum_{i \in J} r_i + 2\varepsilon = Q(E_0) + 2\varepsilon \end{aligned}$$

Since $\mathrm{diam}(E_i) < \frac{1}{n} \forall i$, $E_i \cap E \neq \emptyset \forall i \in J \Rightarrow E_i \subset E^{\frac{1}{n}} \forall i \in J \Rightarrow \bigcup_{i \in J} E_i \subset E^{\frac{1}{n}}$. Thus

$$P(E) \leq Q(E_0) + 2\frac{1}{n} \leq Q(E^{\frac{1}{n}}) + 2\frac{1}{n} \Rightarrow \rho(P, Q) \leq \frac{2}{n}$$

This means that Π_n is a countable, dense family of $\mathcal{P}(S)$.

- Suppose that $\{P_n\}$ is a Cauchy sequence, recalling that a Cauchy sequence is convergent if and only if it has a convergent sub-sequence, if we show that is tight, then we've concluded, indeed, by Prokhorov Theorem A.26, $\{P_n\}$ is also relatively compact and, hence it has a convergent sub-sequence.

  Fix $\varepsilon > 0$ and $\delta > 0$; our goal is to find a finite number of balls of radius $\frac{\delta}{2}$ (we'll call them δ-balls), such that

  $$P_n(C_1 \cup \ldots \cup C_m) > 1 - \varepsilon \quad \forall n$$

  - choose η such that $0 < 2\eta < \varepsilon \wedge \delta$
  - choose n_0 such that $n \geq n_0 \Rightarrow \rho(P_{n_0}, P_n) < \eta$
  - cover S by balls $A_i := B(x_i, \eta)$ and choose m such that $P_n(A_1 \cup \ldots \cup A_m) > 1 - \eta$ for $n \leq n_0$
  - $B_i := B(x_i, 2\eta)$. If $n \geq n_0$ (they exist by Ulam Theorem A.25 because we're considering a finite number of probability measures)

    $$P_n(B_1 \cup \ldots \cup B_m) \geq P_n((A_1 \cup \ldots \cup A_m)^\eta) \geq P_{n_0}(A_1 \cup \ldots \cup A_m) - \eta \geq 1 - 2\eta$$

    If $n \leq n_0$

    $$P_n(B_1 \cup \ldots \cup B_m) \geq P_n(A_1 \cup \ldots \cup A_m) \geq 1 - \eta \geq 1 - 2\eta$$

  We take $C_i := B(x_i, \delta)$ if $i \leq m$, and so, we've found that for every n

  $$P_n(C_1 \cup \ldots \cup C_m) > 1 - 2\eta \geq 1 - \varepsilon \quad \forall n$$

  Now we take $K := \overline{\bigcap_{m \geq 1} \bigcup_{i=1}^m C_i} \Rightarrow K$ is compact because $\bigcap_{m \geq 1} \bigcup_{i=1}^m C_i$ is totally bounded and, by Proposition A.13, relatively compact. Thus for all n

  $$P_n(K) \geq 1 - \varepsilon$$

□

A.1.2 Convergence in product spaces

During the thesis, we frequently have to study some convergences of random variable which take values in product spaces, it is hence useful to give some basics results in such spaces, which will clarify the situation in our model. Let $S := S' \times S''$ be the product of two metric spaces. If S is separable (which means that S and S' have t be separable), then the Borel σ-algebras verifies $\mathcal{S} = \mathcal{S}' \times \mathcal{S}''$ (we omit the proof of this fact because it is outside the scope of our argumentation; for a detailed proof see [4]). The marginal distributions are defined by $P'(A') := P(A' \times S''), A \in \mathcal{S}'$, and $P''(A'') := P(S' \times A''), A'' \in \mathcal{S}''$. Moreover we say that a set $A \in \mathcal{S}$ is a *P-continuity set* if $P(\partial A) = 0$.

Theorem A.29. If S is separable, and P_n, P are probability measure on S, then $P_n \to P$ weakly if and only if $P_n(A' \times A'') \to P(A' \times A'')$, for each P'-continuity set A' and each P''-continuity set A''.

Proof. We're going to use the portmanteau theorem A.11.

- $\Rightarrow$: Denote with $\partial, \partial', \partial''$ the boundary operators respectively on S, S', S''. It's sufficient to observe that $\partial(A' \times A'') \subseteq \big((\partial' A') \times S'') \cup (S' \times \partial'' A'')\big)$ to conclude.
- $\Leftarrow$: We observe that the class $\mathcal{A}$ of sets $A' \times A''$ is a π-system for the P-continuity sets. The hypothesis $P_n(A' \times A'') \to P(A' \times A'')$ for all P'-continuity set A' and each P''-continuity set A'', let us conclude.

□

Theorem A.30. If $S = S' \times S''$ is separable, then $P_n' \times P_n'' \to P \times P''$ weakly if and only if $P_n' \to P', P_n'' \to P''$ weakly.

Proof. $\Rightarrow$: It follows directly from Theorem A.17 since $P_n' = P_n \circ \pi_{S'}^{-1}$ and $P_n'' = P_n \circ \pi_{S''}^{-1}$ with $P_n = P_n' \times P_n''$ and because of the continuity of the projections.
$\Leftarrow$: Direct consequence of Theorem A.29.

□

Theorem A.31 (Slutzky's Theorem)**.** Suppose $S = S' \times S''$ separable. If X_n and Y_n are random elements of S' and S'' such that $X_n \to X$ and $Y_n \to a$ in distribution, where $a \in S''$ is a constant, then $(X_n, Y_n) \to (X, a)$ in distribution.

Proof. Suppose that A' is an X-continuity set and $A'' \subset S''$ such that $a \notin \partial A''$. If $a \in A''$, then $\mathbb{P}(Y_n \notin A'') \to 0$ and so

$$
\begin{array}{ccccc}
\mathbb{P}(X_n \in A') - \mathbb{P}(Y_n \notin A'') \leq & & \mathbb{P}(X_n \in A', Y_n \in A'') \leq & & \mathbb{P}(X_n \in A') \\
\downarrow & & & & \downarrow \\
\mathbb{P}(X \in A') = & & \mathbb{P}(X \in A', Y \in A'') & & \mathbb{P}(X \in A')
\end{array}
$$

If $a \notin A''$, then $\mathbb{P}(X_n \in A', Y_n \in A'') \leq \mathbb{P}(Y_n \in A'') \to 0 = \mathbb{P}(X \in A', Y \in A'')$ □

Corollary A.32. Suppose $S = S' \times S''$ separable. If X_n and Y_n are random elements of S' and S'' such that $X_n \to X$ and $Y_n \to a$ in distribution, where $a \in S''$ is a constant, then $X_n Y_n \to aX$ in distribution.

Proof. By Theorem A.31, we have $(X_n, Y_n) \to (X, a)$ weakly. Consider $g(x, y) = xy$; g is a continuous function and so, by the continuous mapping theorem we have $X_n Y_n \to aX$ in distribution. □

Bibliography

[1] I. Ashlagi, D. Monderer, M. Tennenholtz, *On the Value of Correlation*, Journal of Artificial Intelligence Research 33 (2008), 575-613.

[2] R. J. Aumann, *Correlated Equilibrium as an Expression of Bayesian Rationality*, *Econometrica*, Vol. 55, No. 1 (Jan., 1987), 1-18.

[3] R.J. Aumann, *Subjectivity and Correlation in Randomized Strategies*, Journal of Mathematical Economics 1 (1974) 67-96.

[4] P. Billingsley, *Convergence of Probability Measures*, Wiley Series in Probability and Statistic, 1999.

[5] P. Cardilaguet, *Notes on Mean Field Games*, 2012.

[6] R. Carmona, F. Delarue, *Probabilistic Theory of Mean Field Games with Applications I*, Springer, 2018.

[7] R. Carmona, F. Delarue, A. Lachapelle, *Control of McKean-Vlasov dynamics versus mea n field games*, Math. Finan. Econ. (2013), 7:131-166.

[8] J.B. Conway, *A course in Functional Analysis, Second Edition*, Springer, 1997

[9] M. Fischer, *On the connection between symmetric N-player games and mean field games*, Ann. Appl. Probab., 27(2), 757-810, 2017.

[10] D. Fudenberg, J. Tirole, *Game Theory*, The MIT Press, 1991.

[11] O. Kallenberg, *Foundations of Modern Probability*, Springer, 1997.

[12] A. R. Karlin, Y. Peres, *Game Theory, Alive*, American Mathematical Society, 2016

[13] A. Klenke, *Probability Theory*, Springer, 2104.

[14] J.-M. Lasry, P.-L. Lions, Mean field games, Jpn. J. Math. 2 (2007), no. 1, 229-260.

[15] M. Maschler, E. Solan, S. Zamir, *Game Theory*, Cambridge University Press, 2013.