GAME-THEORETIC LEARNING USING THE IMPRECISE DIRICHLET MODEL

ERIK QUAEGHEBEUR & GERT DE COOMAN

SYSTeMS Research Group

Department of Electrical Energy, Systems & Automation, Ghent University
Technologiepark 914, B-9052 Zwijnaarde, Belgium

UNIVERSITEIT GENT

Introduction

For a specific class of games, and when players use a precise Dirichlet model (PDM) to learn the strategy used by their opponent, Fudenberg *et al.* have proved a number of interesting convergence results; [2, 3].
We present a generalisation of this learning model that uses the imprecise Dirichlet model (IDM); [5]. We also generalise the convergence results.

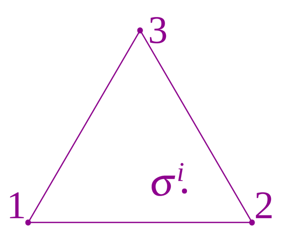
The game: two players competing against each other

We're considering *two-player games*: player i has one opponent $-i$.
The *rules* are simple: each player chooses a strategy every time he plays. These strategies then completely determine their rewards (i.e., a possibly negative pay-off).
We're only considering *strictly competitive games*: if one player gets more by changing his strategy, his opponent will get less.

Two strategy types: pure and mixed

Each player has a finite set S^i of *pure strategies* s^i.
He can also use a *mixed strategy* $\sigma^i \in \Sigma^i$, which is a probability mass function over the set of pure strategies, and let his pure strategy be randomly chosen accordingly.
Strategies can be represented on a unit simplex, where pure strategies correspond to the vertices and convex combinations of vertices correspond to mixed strategies.

The simplex Σ^i of player i, with $S^i = \{1,2,3\}$ and $\sigma^i = (\frac{1}{5}, \frac{3}{5}, \frac{1}{5})$.



What pay-off to expect

After every game, both players receive a *pay-off* $X_{s^i}(s^{-i})$ which is completely determined by the played strategies.
When mixed strategies are played, we can only calculate an *expected pay-off*

$$X_{\sigma^i}(\sigma^{-i}) = \sum_{s^i \in S^i} \sum_{s^{-i} \in S^{-i}} X_{s^i}(s^{-i}) \sigma^i(s^i) \sigma^{-i}(s^{-i}).$$

The uncertainty about the opponent's strategy

Each player supposes that his opponent plays a fixed mixed strategy (*fictitious play*).

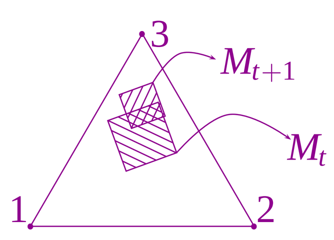
An IDM describes the uncertainty about the opponent's strategy

The player uses a *set of Dirichlet distributions* on the simplex Σ^{-i} as the basis for probabilistic statements that relate to his opponent's unknown fixed strategy choice.
An *IDM is the lower prevision $\underline{P}(\cdot \mid \beta, M)$ determined by these distributions*.
This set of distributions is parameterised by a number of *(pseudo)counts* β and a *subset* $M \subseteq \operatorname{int}(\Sigma^{-i})$. Every strategy in M corresponds to the expectation value of the fixed strategy under one distribution.

Updating the IDM after playing a game

After observing the pure strategy played by the opponent, the set of distributions is updated using *Bayes' rule*, which, here, comes down to moving and shrinking M.
Initially, when the player hasn't a clue about his opponent's strategy, he can use $M_0 = \operatorname{int}(\Sigma^{-i})$. The corresponding IDM is a vacuous prevision.

The simplex Σ^{-i} of the opponent, with $S^{-i} = \{1,2,3\}$.
Updating after observing $s^{-i} = 3$ in game t.



The player chooses his own strategy, if he can...

On the basis of the information at his disposal, the player will want to choose an *optimal strategy*. Optimal in the sense that it maximises his immediate expected pay-off and possibly minimises his risk. In general, the player will find sets of not always comparable optimal strategies. Out of these he can only make an arbitrary choice.

Best replies and maximin strategies

There are two types of optimal strategy choices when a player is sure that his opponent will choose a strategy in a subset $M \subseteq \Sigma^{-i}$.
Best replies $BR^i(M)$ are all the player's strategies σ^i that maximise his expected pay-off $X_{\sigma^i}(\sigma^{-i})$ for a strategy $\sigma^{-i} \in M$.
M-maximin-strategies are those strategies τ^i that maximise his minimal expected pay-off, i.e., for which

$$\tau^i \in \operatorname*{argmax}_{\sigma^i \in \Sigma^i} \inf_{\sigma^{-i} \in M} X_{\sigma^i}(\sigma^{-i}).$$

A player's optimal strategies under uncertainty

When an IDM $\underline{P}(\cdot \mid \beta, M)$ describes the information available about the opponent's fixed strategy (now, we are only sure of the fact that this strategy lies on Σ^{-i}), there are two analogous types of optimal strategies as seen above; [4, §3.9].
Maximal strategies are those strategies σ^i for which

$$\min_{\tau^i \in \Sigma^i} \overline{P}(X_{\sigma^i} - X_{\tau^i} \mid \beta, M) \geq 0,$$

where $\overline{P}(\cdot \mid \beta, M)$ is the conjugate upper prevision of $\underline{P}(\cdot \mid \beta, M)$. We have found that *maximal strategies are best replies* $BR^i(\overline{\text{co}}(M))$ to the closed convex hull of M.
A *$\underline{P}(\cdot \mid \beta, M)$-maximin strategy* is a strategy σ^i that maximises $\underline{P}(X_{\sigma^i} \mid \beta, M)$. We have found that *$\underline{P}(\cdot \mid \beta, M)$-maximin strategies are $\overline{\text{co}}(M)$-maximin strategies*.

What can happen if the game is repeatedly played

A *strategy profile* σ is a couple of strategies (σ^i, σ^{-i}) of the player and his opponent.
An *equilibrium* σ_* is a strategy profile for which $\sigma_* \in BR(\sigma_*)$, each component is a best reply to the other; [1]. For a *strict equilibrium* s_*, it holds that $s_* = BR(s_*)$.

When the played strategy profile converges, it is to an equilibrium

We have proved that if players use an IDM as described above (or a PDM, which is an IDM with M a singleton), it holds that:

- If a strict equilibrium s_* must be played once, it will always be played subsequently.
- If in an infinite sequence of games a strategy profile s_* is played from a certain game onward, it is an equilibrium.
- If in an infinite sequence of games the frequencies of played pure strategies converges to a strategy profile σ_*, then this is an equilibrium.

Conclusion

We have successfully generalised a game-theoretic learning model that uses a PDM to one that uses an IDM. Its main advantages are the possibility of representing initial or intermediary ignorance. A disadvantage is its increased complexity.

References

[1] FRIEDMAN, J. W. *Game Theory with Applications to Economics*. Oxford University Press, New York, 1989.

[2] FUDENBERG, D., AND KREPS, D. M. Learning mixed equilibria. *Games and Economic Behaviour 5* (1993), 320–367.

[3] FUDENBERG, D., AND LEVINE, D. K. *The Theory of Learning in Games*, vol. 2 of *The MIT Press Series on Economic Learning and Social Evolution*. The MIT Press, Cambridge, Massachusets and London, England, 1998.

[4] WALLEY, P. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London, 1991.

[5] WALLEY, P. Inferences from multinomial data: learning about a bag of marbles. *Journal of the Royal Statistical Society, Series B 58* (1996), 3–57. With discussion.