

A Prescriptive Approach for Playing Games

Michal Feldman

School of Business Administration and Center for the Study of Rationality

Hebrew University of Jerusalem

and Microsoft Israel R&D Center

1. INTRODUCTION

Decision-Making lies in the foundations of fields such as Economics, Operations Research, and Artificial Intelligence. The question of what should be the action to be taken by a decision-maker when facing an uncertain environment, potentially including other decision makers, is a fundamental problem which has led to a wide variety of models and solutions. One of the only types of situations (of more than a single agent) for which this question got an agreed-upon answer is in the context of two-player zero-sum games. This setting can model any situation in which a decision-maker aims at maximizing her guaranteed payoff. When mixed strategies are allowed, such desired behavior, termed an agent's maximin (or safety level) strategy, leads to a well defined expected payoff (known as the *value* of the game). Moreover, when presented explicitly in a matrix form, the computation of a maximin strategy is polynomial (by solving a linear program).

Various equilibrium concepts have been considered in the game-theoretic literature, but none of them provides prescriptive advice to a decision-maker which will be as acceptable as the maximin strategy solution in adversarial settings. Only little attention has been given over the years to the challenging task of coming up with game theoretic recommendations to the individual agent who plays in some strategic environment. For a review of some exceptions see [Tennenholtz 2008].

Even within the setting of two-person zero-sum games, various scenarios may arise which require new recommendations. In this letter we provide pointers to two recent contributions that provide a decision maker with a prescriptive advice on how to choose her action in two different adversarial settings. In [Alon et al. 2010] we study two-person zero-sum games in which the randomization phase of a randomized strategy is not completely private; rather, some information about the instantiation selected in the randomization phase leaks in an adversarial manner. In [Feldman et al. 2010] we consider repeated symmetric two-person zero-sum games in which one player never observes a single payoff (but does observe the opponent's play), where the opponent has full knowledge of the game.

2. INFORMATION LEAKAGE IN GAMES

Our model in [Alon et al. 2010] considers a two-player zero-sum game in strategic form with payoffs 1 or 0, where the MAX player is our decision-maker and the MIN player is the adversary. Both MAX and MIN have a set of (pure) strategies they

Author's address: mfeldman@huji.ac.il

can choose from (of sizes m and n , respectively). MAX chooses a mixed strategy, that is, a distribution vector over its pure strategies. MIN may base its action on the value of b binary predicates defined on MAX' pure strategies; each such predicate is a Boolean formula on the set of strategies, where the formula's value is determined according to the actual instantiation of MAX' mixed strategy. The parameter b can be thought of as the amount of information leakage regarding the instance of MAX' mixed strategy; MAX would like to maximize her guaranteed expected payoff against such b binary predicates.

We consider two settings, distinguished by the information structure assumed in them. In the *Strong Model* the MAX player chooses a mixed strategy, which is observable by the MIN player, who can then act upon it in determining the b predicates. In the *Weak Model*, the MIN player chooses the b predicates first, and MAX can observe it and act upon it in choosing her mixed strategy.

We study the following questions. What is the optimal mixed strategy for the MAX player in the two models? How well will the original maximin strategy of the game perform? Finally, what is the computational complexity of finding the optimal strategy under information leakage?

For the Strong Model, if the value of the original game is $q = 1 - \epsilon$ (for small positive ϵ) and 2^b is much smaller than $1/\epsilon$, then MAX can ensure value close to 1 (at least $1 - 2^b\epsilon$), and this is tight. To do so, she simply uses the maximin strategy (that is, the optimal mixed strategy for the original game with no predicates). On the other hand, if 2^b is much bigger than $1/\epsilon$, then for every mixed strategy of MAX, the MIN player can ensure value close to zero (at most $e^{-2^b\epsilon}$). Therefore, for EVERY such game with value $1 - \epsilon$, which is close to 1, a sharp transition occurs at b which is about $\log(1/\epsilon)$: if b is slightly smaller, the value stays close to 1; if it is slightly larger, the value drops to nearly zero.

For games with value q bounded away from 1, even one bit enables MIN to square the value and drop it to at most q^2 , and every additional bit squares the value again. There are also examples showing that this is essentially tight. Finally, for any fixed value $q < 1$, $\log \log m + O_q(1)$ bits suffice to enable MIN to drop the value to precisely 0.

For the Weak Model, the situation is different. Clearly, here MAX is in a better shape, hence if the value of the game is $q = 1 - \epsilon$ (for small positive ϵ), MAX can still ensure a value close to 1 if the number of bits is much smaller than $\log(1/\epsilon)$ as in the Strong Model. For games with value q bounded away from 1, however, there are examples in which she can do much better than in the Strong Model, and in fact can ensure no essential drop in the value as long as the number of leaking bits is somewhat smaller than $\log \log m$. More precisely, for any fixed value $0 < q < 1$ and for every large polynomially related m, n , there are examples of games represented by a binary m by n matrix with value $q + o(1)$, so that even if $b = \log \log m - O(1)$, MAX can ensure that the value will stay roughly q . This should be contrasted with the Strong Model, where every additional bit squares MAX' value.

Somewhat surprisingly, once the number of leaking bits is slightly larger, that is, $b = \log \log m + O(1)$, the MIN player can already ensure value 0 in any game with a fixed value $q < 1$. Thus, in the examples above a sharp transition occurs at nearly $b = \log \log m$ under the Weak Model: nearly $\log \log m$ bits have essentially

no effect on the value, while slightly more bits suffice to drop the value to 0.

Note that, in contrast to leakage-free settings, where no advantage is gained by observing the opponent's mixed strategy (due to the minimax theorem), in settings of adversarial leakage, such information can contribute a great deal to the informed player, reflected by the advantage obtained by MAX in the Weak Model compared with the Strong Model.

With respect to computational complexity, computing the optimal strategy in the Strong Model (for the MAX player) against b leaking bits is poly-time for any fixed b , while this problem becomes NP-hard, or even to approximate within any factor, for a general b . In the Weak Model, the optimal strategy of MAX can be computed in polynomial time for every b .

3. PLAYING GAMES WITHOUT OBSERVING PAYOFFS

Our model in [Feldman et al. 2010] considers a symmetric two-person game that is repeated T times, where one player – our decision maker – never observes a single payoff (but observes the opponent's play), while her opponent has full knowledge of the game. Naturally, the decision maker should attempt to mimic the opponent's play. However, one has to be careful about how one mimics opponents who may know that they are being mimicked. As we show, a good copycat can reap tremendous rewards without ever observing a single payoff, while a poor copycat may perform worse than making random decisions.

We give a simple (easy to compute) repeated-game strategy for the uninformed player using only the history of observed play of the opponent. The strategy in no way depends on prior knowledge of the game (not even the number of strategies). No assumptions are made about the opponent. Hence the strategy may, in extreme cases, even be used by a player that is completely oblivious to payoffs, one who never observes a single payoff and has no idea about what these might even be, against an opponent that has complete knowledge of the game. The strategy has the following guarantee: for any finite symmetric zero-sum n by n game $A \in R^{n \times n}$, any number of rounds $T \geq 1$, and any opponent's strategy, the expected average payoff of the uninformed player in the first T rounds is at least $-\frac{n}{\sqrt{2T}} \max_{i,j} |a_{ij}|$.

The COPYCAT strategy achieving the above bound is the following: on period t , let $V_t \in R^{n \times n}$ be the frequency matrix where $V_t(i, j)$ is the number of times (i, j) has been played on periods $1, 2, \dots, t-1$. The copycat strategy plays, on period t , a min-max mixed-strategy of the zero-sum game with payoffs $P_t = V_t^T - V_t$.

We then consider several extensions. We first show that the uninformed player need not even know the set of strategies in advance. Second, in general-sum symmetric games, we show that the COPYCAT strategy guarantees the uninformed player an expected payoff that is nearly-equal that of the informed player. Third, we note that the COPYCAT strategy yields *learning equilibria* [Brafman and Tennenholtz 2004], a type of equilibrium for an entire family of games (a non-Bayesian notion of equilibrium for games with incomplete information).

While two-player symmetric games are a central object of study in game theory, our model of extreme informational asymmetry is new. Of course, such extremes are less common than other situations. But the point here is to test the limits of how much feedback a player needs to play a game. If the informed player has less

knowledge, or the uninformed player has more feedback, the algorithm and analysis can of course still be applied. In cases where determining the payoff is difficult or costly, it is good to know that it is possible to simply copy. While this payoff-oblivious setting is less common than bandit settings [Cesa-Bianchi and Lugosi 2006], we do feel that there are sufficient applications to merit study. Furthermore, understanding this setting helps expand our general understanding of optimization and game play under uncertainty.

REFERENCES

- ALON, N., EMEK, Y., FELDMAN, M., AND TENNENHOLTZ, M. 2010. Information leakage in games. In *Proceedings of the first Conference on Innovations in Computer Science*.
- BRAFMAN, R. AND TENNENHOLTZ, M. 2004. Efficient learning equilibrium. *Artificial Intelligence* 159, 27–47.
- CESA-BIANCHI, N. AND LUGOSI, G. 2006. *Prediction, Learning, and Games*. Cambridge University Press.
- FELDMAN, M., KALAI, A., AND TENNENHOLTZ, M. 2010. Playing games without observing payoffs. In *Proceedings of the first Conference on Innovations in Computer Science*.
- TENNENHOLTZ, M. 2008. Game-theoretic recommendations: some progress in an uphill battle. In *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems*.