CORVINUS UNIVERSITY OF BUDAPEST

Doctoral School of Economics, Business and Informatics

**THESIS BOOK**

to the PhD Thesis titled

# Complexity and estimation of the Shapley value

written by

Ferenc Illés

Supervisor:

Péter Csóka PhD

Corvinus University of Budapest, Institute of Finance

Budapest, 2023

# 1 Motivation of the research

Cooperative game theory is a model of many social dilemmas and financial and economic problems. It is usually used when the result achieved by a community exceeds the sum of the results achieved by individuals (synergy), or the total cost of an activity is lower than the sum of the costs of the sub-tasks (savings) or its total risk is lower than the aggregate risk of the sub-tasks (diversification). And the main question is, of course, how to distribute the surplus among the actors in some “fair” way.

## Applications of game theory

Social choice problems are usually solved by voting. The European Economic Community was founded in 1957 by six countries shown in Table 1. A positive decision required at least 12 yes votes out of 17 based on the weights in column 2. At first glance, Luxembourg seems to be over-represented with a voting power of a quarter of the West Germany, which is 150 times larger. If we take a closer look at the numbers, we can see that there is no such situation where Luxembourg's vote counts. With all other five weights being even, it is impossible to form a coalition of supporters of a proposal with a total weight of 11, which would be the only situation where Luxembourg could change the decision. Since the weights are clearly misleading, we would like to assign values to the players that reflect their true bargaining power.

| Country | Weight | Population | |
|---|---|---|---|
| Franciaország | 4 | 44 | million |
| Federal Republic of Germany | 4 | 51 | million |
| Italy | 4 | 49 | million |
| Belgium | 2 | 9 | million |
| Netherlands | 2 | 11 | million |
| Luxembourg | 1 | 308 | thousand |

Table 1: Weights and population of EEC countries

A financial application is as follows. There are thirty brokers in the treasury of a commercial bank, who can invest into any portfolio of thousands of assets available on the market. The department's total limit is HUF 30 billion, the capital requirement of the bank's aggregated portfolio cannot exceed this value at any time. The capital requirement is the Conditional Value at Risk at 99.9% level of confidence, which is estimated from the historical data, assuming a multidimensional normal distribution. The head of department approves a HUF 1 billion limit for each broker, which the colleagues respect individually, however, when the capital requirement of the bank's aggregated portfolio is calculated, it is only HUF 13 billion. The reason for the difference is that the capital requirement is not additive due to the diversification. The question is how to distribute the total risk of the portfolio among the individual assets? In the literature the problem is called “capital allocation” or “risk allocation” (see, for example, Colini-Baldeschi et al. (2018), Csóka et al. (2009) and Csóka and Pintér (2016)).

Cooperative games appear frequently in economics, finance, even in statistics and law. The general solution was given by Lloyd Stowell Shapley in 1953 in his famous paper *A value for n-person games*. Shapley's solution, which is now called the Shapley value, is the only general solution concept that satisfies certain fairness conditions. Unfortunately - despite the simple formula and the intuitive interpretation - we cannot calculate it precisely when the number of players is too large, because the number of terms in the formula is not polynomial in the number of players. The essence of this dissertation in one sentence is as follows. How can (or can't) we calculate or estimate effeciently the Shapley value of a cooperative game?

## 2 Prior research results

The origins of cooperative games go back to the book *Theory of Games and Economic Behaviour* (1944) by John von Neumann and Oskar Morgenstern.

**Definition 2.1.** *A cooperative n-person game in characteristic function form is (or TU-game for short) is an ordered pair $(N, v)$ where $N$ is a nonempty set and $v : 2^N \to \mathbb{Z}$ is a set function over $N$ such that $v(\emptyset) = 0$.*

The elements of $N$ are *players*, the subsets of $N$ are *coalitions*, function $v$ is called *characteristic function* and $v(S)$ is the *value* of coalition $S \subseteq N$. The value of a coalition $v(S)$ is the same for all players, meaning *transferable utility* or TU for short. Let $\mathcal{G}_N$, denote all TU games over player set $N$. This is a $\mathbb{Z}$-module with the pointwise (or "coalitionalwise") operations.

**Definition 2.2.** *A value for an n-person game is a $\varphi : \mathcal{G}_N \to \mathbb{Q}^N$ function.*

A value for a game is a solution concept, but it might be completely meaningless without further assumptions. To define the "fairness" of a value we need to introduce a the concept of *marginal contribution.*

**Definition 2.3.** *Let $g \in \mathcal{G}_N$ be a game, $S \subseteq N$ a coaltion and $i \in N$ a player such that $i \notin S$. Player i's marginal contribution to coalition $S$ is*

$$v'_i(S) = v(S \cup \{i\}) - v(S).$$

**Definition 2.4.** *A value $\varphi : \mathcal{G}_N \to \mathbb{Q}^N$*

- *satisfies equal treatment property (ETP), if for each $g \in \mathcal{G}_N$ for all $i$ and $j$ players*

$$(v'_i(S) = v'_j(S)\ \forall\ i, j \notin S \subseteq N) \Longrightarrow \varphi_i = \varphi_j,$$

- *is efficient (EFF), if $\sum \varphi_i = v(N)\ \forall g \in \mathcal{G}_N$,*
- *satisfies null player property (NPP) if $\forall g \in \mathcal{G}_N$*

$$v'_i(S) = 0 \quad \forall i \notin S \subseteq N \Longrightarrow \varphi_i = 0,$$

- *is additive (ADD) if $\varphi(g_1 + g_2) = \varphi(g_1) + \varphi(g_2) \quad \forall g_1, g_2 \in \mathcal{G}_N$.*

According to the following result these properties characterize the value of cooperative games.

**Theorem 2.1** (Shapley, 1953)**.** *For all $N \neq \emptyset$ $\exists!\varphi$ value that satisfies properties ADD, ETP, NPP and EFF.*

Shapley did not only prove the existence and uniqueness of such a value but deduced a formula to calculate it.

**Definition 2.5.** *Let $N$ be a set of players in game $g = (N, v)$. An order of arrival of players is an $o : N \leftrightarrow \{1, 2, \ldots, n\}$ bijective function. For $i \in N$ and $j \in \mathbb{N}$ $o(i) = j \in \mathbb{N}$ means that i's position is $j$ in $o$. Let $S_o(i) = \{i' \in N \mid o(i') < o(i)\}$ denote the players preceding $i$ in $o$. The Shapley value of game $g = (N, v)$ is*

$$Sh_i = \sum_o \frac{v'_i(S_o(i))}{n!} \qquad \forall i \in N. \tag{1}$$

$Sh_i$ is the average marginal contribution of player $i$ to the coalition of players arrived before $i$, considering all of the possible order of arrivals of the $n = |N|$ players, hence it is a sum of $n!$ terms. As the marginal contribution does not depend on the order of players if they form the same coalition, the formula contains a lot of equal terms. If we combine them we get that

$$Sh_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|! \cdot (n - |S| - 1)!}{n!} \cdot v_i'(S) = \frac{1}{n} \sum_{S \subseteq N \setminus \{i\}} \frac{1}{\binom{n-1}{|S|}} \cdot v_i'(S). \tag{2}$$

Since the number of coalitions in a game of $n$ players is $2^n$ and the number of terms in formula (2) is $2^{n-1}$, the Shapley value is computationally easy to calculate based on the list of coalition values. However if we want to define a game with hundred players it is impractical to list the values of all of the non-empty coalitions. In most cases the characteristic function is defined by a simple formula and a "small" set of parameters as the following examples show.

**Example 2.1.**

**weighted voging game:** Let $n \in \mathbb{N}$ be a positive integer, $N = \{1, 2, \dots, n\}$, $w = (w_1, w_2, \dots, w_n) \in \mathbb{N}^n$ and $q \in \mathbb{Z}$. Let the value of each coalition $S \subseteq N$ be defined by formula

$$v(S) = \begin{cases} 1, & \text{ha } \sum_{i \in S} w_i \geq q, \\ 0, & \text{ha } \sum_{i \in S} w_i < q. \end{cases}$$

**airport game:** Let $n \in \mathbb{N}$ be a positive integer, let $N$ be a set of $n$ items for each $i \in N$ let $c_i \in \mathbb{N}$ be a positive parameter and consider the characteristic function

$$v(S) = \max\{c_i | i \in S\}.$$

**bankruptcy game:** Let $A \in \mathbb{Z}^+$ be a non-negative integer, $n \in \mathbb{N}$, $l_1, l_2, \dots l_n \in \mathbb{N}$ and consider the characteristic function

$$v(S) = \max\{0, A - \sum_{j \notin S} l_j\}.$$

The computational complexity of the Shapley value of a voting game given by the weights of the players is a meaningful question, because the input is a sequence of $n$ integers and the output is $n$ rational numbers, but to evaluate formula (2) it requires the calculation of a sum of $O(2^n)$ terms. The problem is similar in case of most games.

One of the oldest results was published by Megiddo (1978). Let the set of players be $N = \{1, 2, \dots, n\}$ and let $G(N \cup \{0\}, E, d)$ be a rooted tree with weights on the edges $d(i, j) \geq 0$ $(i, j \in N \cup \{0\})$ where 0 denotes the root. For an $S \subseteq N$ coalition, let $v(S)$ denote the sum of weights of edges on a path from the root to a node in $i \in S$. The Shapley value of these cost sharing games can be calculated in polinomial time using a simple graph-searching algorithm.

On the other hand, Deng and Papadimitriou (1994) states that the Shapley value of voting games is #P-complete. Furthermore, it is easy to see that to decide if the Shapley value of a player in a voting game is 0 is a generalization of the SUBSET SUM problem, which is well-kown to be NP-hard.

The complexity results limit the possibilities for calculating the Shapley value, but there are two options that can work in practice, both proposed by Shapley himself back in the '60s.

- Though they never used this terminology, Mann and Shapley (1962) give a pseudopolynomial algorithm for the case of voting games, based on Cantor's idea, by recursively computing a generator function. The algorithm is polynomial on sub-game classes where the weights are polynomial in the number of players.
- Mann and Shapley (1960) suggests estimating the Shapley value using Monte-Carlo simulation.

# 3 Main results

The contribution to the literature is as follows.

- A complexity of the Shapley value for a new class of games. We present some minor results on the approximation of the Shapley value and the representation of games.
- A new Monte-Carlo method is introduced for estimating the Shapley value.
- We introduce the concept of linearly representable games and present an algorithm equivalent in complexity to the method proposed by Mann and Shapley (1962) for voting games, but more general, and its relation to airport games.

## 3.1 Representation and complexity of cooperative games

The calulcation of the Shapley value as a computational problem can be meaningfully defined as input-output pairs. As it is impractical to consider the characteristic function a gigantic list of coalition values, the first step is to define what we mean by passing a game to an algorithm. Let $\bar{n}$ denote the $\bar{n} = \{1, 2, 3, \ldots, n\}$ the set of the first $n$ natural numbers.

**Definition 3.1.** *A game class $G \subseteq \bigcup\limits_{n\in\mathbb{N}} \mathcal{G}_{\bar{n}}$ is polinomially representable, if*

- *$\exists\Sigma$ finite alphabet, that contains an empty symbol denoted by a star $(*)$ and a 0 and 1 symbol,*[1]
- *$\exists T_G$ Turing machine over $\Sigma$ with at least three tapes and*
- *$\forall g = (\bar{n_g}, v_g) \in G\ \exists x_g \in \Sigma_0^* = \bigcup_{i=0}^{\infty} (\Sigma \setminus \{*\})^i$ such that*
  - *the length of $x_g$ is polinomial in $n_g$ and*
  - *if we pass an $n_g$-long 0-1 sequence $u$ and $x_g$ to $T_G$ as input, it calculates the value of the coalition represented by $u$ in game $g$: $T_G(u, x_g) = v_g(S)$.*

A polynomial representation means that a class of games can be sufficiently "compressed" to make its Shapley value complexity a meaningful question. The games of example 2.1. are given by such a concise representation (if the parameters are at most exponentially large in the number of players), for example, in the case of voting games, there are $n + 1$ parameters, the players' weights, and the quota.

**Definition 3.2.** *Let $G \subseteq \bigcup\limits_{n\in\mathbb{N}} \mathcal{G}_{\bar{n}}$ be a polynomially representable class of games. Then the words $x_g$ form a language $L_G = \{x_g \in \Sigma_0^* \mid g \in G\}$. We say that the ordered pair $(L_G, T_G)$, where $T_G$ is the Turing-machine that calculates the coalition values is polinomial representation of $G$.*

Though it is a tautology that a game class is polynomially representable iff it has a(t least one) polynomial representation, the definition is not superfluous, because a game class can have multiple polynomial representation, which leads to the problem of *representational equivalence*: it is not clear how to check if two formally different polynomial representation defines the same game. In certain cases this can even be intractable.

**Theorem 3.1.** *It is co-NP-complete to decide whether two linear representations in the form of example 2.1. represent the same voting games.*

[1] Thus we can represent integer numbers in binary format;

*Proof.* Given $2n+2$ positive integers $w_1, w_2, \ldots, w_n, q$ and $w'_1, w'_2, \ldots, w'_n, q'$ we have to decide if for every $S \subseteq \{1, 2, \ldots, n\}$

$$\sum_{i \in S} w_i \geq q \Leftrightarrow \sum_{i \in S} w'_i \geq q'.$$

The special case when $w_i = w'_i$ and $q' = q - 1$ it is equivalent to the

$$\nexists S \subseteq \{1, 2, \ldots, n\} \text{ such that } \sum_{i \in S} w_i = q$$

subset sum problem. Furthermore, the problem is obviously in co-NP. □

Using the polynomial representation and representability we can formally define what we mean by the efficient calculation of the Shapley value for a game class.

**Definition 3.3.** *We say that the Shapley value can be computed in polynomial time for the game class* $G \subseteq \bigcup_{n \in \mathbb{N}} \mathcal{G}_{\bar{n}}$ *if*

- $\exists (L_G, T_G)$ *polynomial representation of* $G$ *and*
- *a Turing machine* $T$ *with at least three tapes, such that*
- *for each* $g \in G$ *game* $T$ *calculates its Shapely value in polynomial time based on its representing word* $x_g \in L_G$ *using* $T_G$ *as a subroutine.*

**Remark.** The definition only requires that a class of games have (at least one) polynomial representation that can be used to efficiently calculate the Shapley value. Theoretically, it is possible that a game is representable in several ways, but the Shapley value can be calculated from one of them but not from the other. This is not a contradiction, but in this case one representation cannot be easily converted to the oter.

The results related to the complexity of the Shapley value are of two types: it is polynomial for certain games (Megiddo, 1978), in other cases it is a generalization of an NP-hard problem (Deng and Papadimitriou, 1994), however, the dichotomy is still an open question. A new result presented in the dissertation is that for a new game class *liability games* introduced by Csóka and Herings (2019),the Shapley value is NP-hard to compute. This result was published by (Csóka et al., 2022) in 2022.

## 3.2 Estimation of the Shapley's value using Monte Carlo simulation

The Shapley value, as the average marginal contribution of the players, is the expected value of a discrete random variable, and thus – even if we cannot calculate it exactly – it can be estimated with a Monte-Carlo simulation. The idea was first proposed by Shapley himself in 1960 (Mann and Shapley, 1960) and he examined several possible methods for reducing the estimation error (variance) compared to the simple random sampling method, which can be considered a natural benchmark.

In the fourth chapter of the dissertation, a new MC method is dicussed. The estimation by *ergodic sampling* introduced by Illés and Kerényi (2019) generates a pair of samples that are not independent (but negatively correlated to reduce the variance), and yet follow the law of large numbers (so the estimate is consistent).

## 3.3 Exact pseudopolynomial algorithms

An algorithm is called pseudo-polynomial if its running time is a polynomial function of the numerical value of the input (not necessarily the size of the input). The concept only makes sense for problems where the input is a number number or a sequences of numbers. Since it is typical in case of the Shapley value, it makes sence to develop algorithms, which are not necessarily polynomial as a function of the size of the input, but the running time, unlike the brute force method, does not increase exponentially with the number of players. In the last chapter of the dissertation, a new algorithm is discussed for a special, but wide class of games *linearly representable games*. This game class and the algorithm proposed by Illés (2022).

# 4 Summary of the main results

- We have introduced the concept of polynomially representable games, a framework in which the computational complexity of the Shapley value is a meaningful question.
- We have shown that the calculation of the Shapley value for the class of the recently introduced liability games (Csóka and Herings (2019)) cannot be solved in polynomial time (unless P = NP).
- We presented a new method of estimating the Shapley value using Monte-Carlo simulation.
- We introduced the concept of linearly representable games and showed that the calculation of the Shapley value for them is pseudopolynomial. This is a generalization of several previous results and covers unexamined cases, because both voting games, bankruptcy games, and liability games are linearly representable.
- We have shown that the algorithm also works for non-linearly represented airport games, and even runs in linear time after a great simplification, which is a new proof of a known result.

Most of the results have been published in one article and two working papers:

- Csóka, Péter, Ferenc Illés, and Tamás Solymosi. "On the Shapley value of liability games." *European Journal of Operational Research* 300.1 (2022): 378-386.
- Illés, Ferenc, and Péter Kerényi. "Estimation of the Shapley value by ergodic sampling." *arXiv preprint arXiv:1906.05224* (2019).
- Illés, Ferenc. "Linearly representable games and pseudo-polynomial calculation of the Shapley value." *arXiv preprint arXiv:2205.12324* (2022).

# References


Riccardo Colini-Baldeschi, Marco Scarsini, and Stefano Vaccari. Variance allocation and Shapley value. *Methodology and Computing in Applied Probability*, 20(3):919–933, 2018.

Péter Csóka and P Jean-Jacques Herings. Liability games. *Games and Economic Behavior*, 116:260–268, 2019. URL `https://doi.org/10.1016/j.geb.2019.05.007`.

Péter Csóka and Miklós Pintér. On the impossibility of fair risk allocation. *The BE Journal of Theoretical Economics*, 16(1):143–158, 2016.

Péter Csóka, P Jean-Jacques Herings, and László Á Kóczy. Stable allocations of risk. *Games and Economic Behavior*, 67(1):266–276, 2009.

Péter Csóka, Ferenc Illés, and Tamás Solymosi. On the shapley value of liability games. *European Journal of Operational Research*, 300(1):378–386, July 2022. doi: 10.1016/j.ejor.2021.10.012. URL `https://doi.org/10.1016/j.ejor.2021.10.012`.

Xiaotie Deng and Christos H Papadimitriou. On the complexity of cooperative solution concepts. *Mathematics of Operations Research*, 19(2):257–266, 1994. URL `http://www.jstor.org/stable/3690220`.

Ferenc Illés. Linearly representable games and pseudo-polynomial calculation of the shapley value. *arXiv preprint arXiv:2205.12324*, 2022.

Ferenc Illés and Péter Kerényi. Estimation of the Shapley value by ergodic sampling. *arXiv preprint arXiv:1906.05224*, 2019.

Irwin Mann and Lloyd S Shapley. Values of large games, iv: Evaluating the electoral college by montecarlo techniques. 1960.

Irwin Mann and Lloyd S Shapley. Values of large games, vi: Evaluating the electoral college exactly. 1962.

Nimrod Megiddo. Computational complexity of the game theory approach to cost allocation for a tree. *Mathematics of Operations Research*, 3(3):189–196, 1978. doi: 10.1287/moor.3.3.189. URL `https://doi.org/10.1287/moor.3.3.189`.

Lloyd S Shapley. A value for n-person games. *Contributions to the Theory of Games*, 2(28): 307–317, 1953.

Alan D Taylor and Allison M Pacelli. *Mathematics and politics: strategy, voting, power, and proof.* Springer Science & Business Media, 2008.

J. von Neumann and O. Morgenstern. *Theory of games and economic behavior.* Princeton University Press, 1944.