COOPERATIVE GAMES: the SHAPLEY VALUE

The description of a cooperative game is still in terms of a **characteristic function** which specifies for every group of players the total payoff that the members of S can obtain by signing an agreement among themselves; this payoff is available for distribution among the members of the group.

DEFINITION. A **coalitional game with transferable payoff** (or characteristic function game) is a pair $\langle N, v\rangle$ where $N = \{1, ..., n\}$ is the set of players and for every subset S of I (called a ***coalition***) $v(S) \in \mathbb{R}$ is the total payoff that is available for division among the members of S (called the ***worth*** of S). We assume that the larger the coalition the higher the payoff (this property is called superadditivity):

$$\text{for all disjoint } S, T \subseteq N, \qquad v(S \cup T) \geq v(S) + v(T)$$

As before, an agreement is a list $(x_1, x_2, \ldots, x_n)$ where x_i is the proposed payoff to individual i. Shapley proposed some conditions (or axioms) that a solutions should satisfy and proved that there is a unique solution that meets those conditions. The solution, known as the **Shapley value**, has a nice interpretation in terms of **expected marginal contribution**. It is calculated by considering all the possible orders of arrival of the players into a room and giving each player his marginal contribution. The following examples illustrate this.

EXAMPLE 1. Suppose that there are two players and $v(\{1\}) = 10$, $v(\{2\}) = 12$ and $v(\{1,2\}) = 23$. There are two possible orders of arrival: (1) first 1 then 2, and (2) first 2 then 1.

If 1 comes first and then 2, 1's contribution is $v(\{1\}) = 10$; when 2 arrives the surplus increases from 10 to $v(\{1,2\}) = 23$ and therefore 2's marginal contribution is $v(\{1,2\}) - v(\{1\}) = 23 - 10 = 13$.

If 2 comes first and then 1, 2's contribution is $v(\{2\}) = 12$; when 1 arrives the surplus increases from 12 to $v(\{1,2\}) = 23$ and therefore 1's marginal contribution is $v(\{1,2\}) - v(\{2\}) = 23 - 12 = 11$.

Thus we have the following table:

Probability	Order of arrival	1's marginal contribution	2's marginal contribution
$\frac{1}{2}$	first 1 then 2	10	13
$\frac{1}{2}$	first 2 then 1	11	12

Thus 1's expected marginal contribution is: $\frac{1}{2}10 + \frac{1}{2}\ 11 = 10.5$ and 2's expected marginal contribution is $\frac{1}{2}13 + \frac{1}{2}\ 12 = 12.5$. This is the Shapley value: $x_1 = 10.5$ and $x_2 = 12.5$.

EXAMPLE 2. Suppose that there are three players now and $v(\{1\}) = 100$, $v(\{2\}) = 125$, $v(\{3\}) = 50$, $v(\{1,2\}) = 270$, $v(\{1,3\}) = 375$, $v(\{2,3\}) = 350$ and $v(\{1,2,3\}) = 500$. Then we have the following table:

v({1}) = 100, v({2}) =125, v({3}) = 50, v({1,2}) = 270, v({1,3}) = 375, v({2,3}) = 350 and v({1,2,3}) = 500

Probability	Order of arrival	1's marginal contribution	2's marginal contribution	3's marginal contribution
$\frac{1}{6}$	first 1 then 2 then 3: 123	v({1}) = 100	v({1,2}) − v({1}) = 270 − 100 = 170	v({1,2,3}) − v({1,2}) = 500 − 270 = 230
$\frac{1}{6}$	first 1 then 3 then 2: 132	v({1}) = 100	v({1,2,3}) − v({1,3}) = 500 − 375 = 125	v({1,3}) − v({1}) = 375 − 100 = 275
$\frac{1}{6}$	first 2 then 1 then 3: 213	v({1,2}) − v({2}) = 270 − 125 = 145	v({2}) =125	v({1,2,3}) − v({1,2}) = 500 − 270 = 230
$\frac{1}{6}$	first 2 then 3 then 1: 231	v({1,2,3}) − v({2,3}) = 500 − 350 = 150	v({2}) =125	v({2,3}) − v({2}) = 350 − 125 = 225
$\frac{1}{6}$	first 3 then 1 then 2: 312	v({1,3}) − v({3}) = 375 − 50 = 325	v({1,2,3}) − v({1,3}) = 500 − 375 = 125	v({3}) = 50
$\frac{1}{6}$	first 3 then 2 then 1: 321	v({1,2,3}) − v({2,3}) = 500 − 350 = 150	v({2,3}) − v({3}) = 350 − 50 = 300	v({3}) = 50

Thus 1's expected marginal contribution is: $\frac{1}{6}\left(100+100+145+150+325+150\right)=\frac{970}{6}$

2's expected marginal contribution is $\frac{1}{6}170+\frac{1}{6}125+\frac{1}{6}125+\frac{1}{6}125+\frac{1}{6}125+\frac{1}{6}300=\frac{970}{6}$

3's expected marginal contribution is $\frac{1}{6}230+\frac{1}{6}275+\frac{1}{6}230+\frac{1}{6}225+\frac{1}{6}50+\frac{1}{6}50=\frac{1060}{6}$

The sum, of course, is $\frac{3000}{6}=500=v(\{1,2,3\})$

COOPERATIVE GAMES: the CORE

So far we have looked at non-cooperative games, characterized by the fact that the individuals involved cannot sign binding agreements and therefore any suggested outcome has to be self-enforcing (i.e. a Nash equilibrium) for the players to be willing to go along with it. We now turn to **cooperative** games, where **binding agreements** are possible. The central question then becomes: *what agreement would the individuals involved be willing to subscribe to?*

The description of a cooperative game is in terms of a **characteristic function** which specifies for every group of players (i.e. every group of individuals who might enter into a binding agreement) the total payoff (e.g. money) that the members of S can obtain by signing an agreement among themselves; this payoff is available for distribution among the members of the group.

DEFINITION. A **coalitional game with transferable payoff** (or characteristic function game) is a pair $\langle N, v \rangle$ where $N = \{1, ..., n\}$ is the set of players and for every subset S of N (called a ***coalition***) $v(S) \in \mathbb{R}$ is the total payoff that is available for division among the members of S (called the ***worth*** of S). We assume that the larger the coalition the higher the payoff (this property is called superadditivity):

$$\text{for all disjoint } S, T \subseteq N, \qquad v(S \cup T) \geq v(S) + v(T)$$

What kind of agreement do we expect individuals to get to? An agreement can be thought of as a list $(x_1, x_2, \ldots, x_n)$ where x_i is the proposed payoff to individual i. Let us try to determine the set of **acceptable agreements** by eliminating those that are unacceptable. First of all, an agreement $(x_1, x_2, \ldots, x_n)$ must be feasible, i.e. it cannot be such that $x_1 + x_2 + \ldots + x_n > v(N)$. Thus,

first necessary condition for acceptability:	$x_1 + x_2 + \ldots + v_n \leq v(N)$ (feasibility condition).

Secondly, an agreement $(x_1, x_2, \ldots, x_n)$ would be unacceptable to individual i if $x_i < v(\{i\})$, because if such an agreement were proposed, individual i would do better by refusing to be part of the agreement and acting by herself [thus guaranteeing herself a payoff of $v(\{i\})$]. Thus

second necessary condition for acceptability:	$x_i \geq v(\{i\})$ for all i (individual rationality condition).

Thirdly, an agreement $(x_1, x_2, \ldots, x_n)$ would also be unacceptable if $x_1 + x_2 + \ldots + x_n < v(N)$, because it would require some potential surplus to be wasted.. Thus

third necessary condition for acceptability:	$x_1 + x_2 + \ldots + x_n \geq v(N)$ (Pareto optimality).

Note that the first and third condition together require $x_1 + x_2 + \ldots + x_n = v(N)$.

EXAMPLE. Consider the following game: N = {1,2,3} and

$$v(\{1\}) = 100$$
$$v(\{2\}) = 125$$
$$v(\{3\}) = 50$$
$$v(\{1,2\}) = 270$$
$$v(\{1,3\}) = 375$$
$$v(\{2,3\}) = 350$$
$$v(\{1,2,3\}) = 500$$

Then the following agreement satisfies the three necessary conditions listed above: $x_1 = 120$, $x_2 = 250$, $x_3 = 130$. Is such an agreement likely to be accepted, if proposed to the three individuals? The answer is No, because individuals 1 and 3 would be better off if they walked out of the negotiations and acted independently of individual 2: 1 and 3 together (and without individual 2) can get 375 and they could split this sum in such a way that they are both better off than in the proposed agreement, e.g. 1 gets 180 and 3 gets 195. Thus we need to add further restrictions.

DEFINITION. A proposed agreement $(x_1, x_2, \ldots, x_n)$ is *blocked by coalition S* if there exists a vector $(y_1, y_2, \ldots, y_n)$ such that:

(1) $y_i > x_i$ for all $i \in S$, (each member of S is better off under the alternative *y*)

(2) $\sum_{i \in S} y_i \leq v(S)$ (alternative *y* is feasible for the coalition S.

Thus the coalition S blocks agreement $(x_1, x_2, \ldots, x_n)$ if the members of S can withdraw from the negotiations with the rest of the players and achieve among themselves a better allocation of payoffs. Thus,

fourth necessary condition for acceptability:	there is no coalition that blocks the proposed agreement.

DEFINITION. The **core** is the set of proposed agreements that satisfy the above four conditions.

How do we find the core? The following theorem gives us the answer.

THEOREM. A feasible agreement $(x_1, x_2, \dots, x_n)$ is in the core if and only if

$$\sum_{i \in S} x_i \geq v(S) \quad \text{for all } S \subseteq N \quad (S \neq \varnothing)$$

For the intellectually ambitious here is a simple-to-understand proof.

Proof. Let $(x_1, x_2, \dots, x_n)$ be a feasible allocation that satisfies the above property. Then taking $S = \{i\}$ we get individual rationality and taking $S = N$ we get Pareto optimality. On the other hand, if there were a coalition S that could block $(x_1, x_2, \dots, x_n)$ with $(y_1, y_2, \dots, y_n)$ then we would have that $y_i > x_i$ for all $i \in S$ and $\sum_{i \in S} y_i \leq v(S)$. But $y_i > x_i$ for all $i \in S$ implies that $\sum_{i \in S} y_i > \sum_{i \in S} x_i$. By hypothesis $\sum_{i \in S} x_i \geq v(S)$. Thus $\sum_{i \in S} y_i > v(S)$ yielding a contradiction.

Conversely, let $(x_1, x_2, \dots, x_n)$ be an allocation in the core. We want to show that it must satisfy the property that $\sum_{i \in S} x_i \geq v(S)$ for all $S \subseteq N$ $(S \neq \varnothing)$. Suppose not. Then there exists an $S \subseteq N$ such that $\sum_{i \in S} x_i < v(S)$. Let $a = v(S) - \sum_{i \in S} x_i > 0$ and consider the following allocation $(y_1, \dots, y_n)$:

$$y_i = \begin{cases} x_i + \dfrac{a}{|S|} & \text{if } i \in S \\ y_i = v(\{i\}) & \text{if } i \notin S \end{cases} \quad \text{(where } |S| \text{ denotes the number of elements in S)}$$

Then $(y_1, \dots, y_n)$ is feasible (by the superadditivity condition) and S blocks $(x_1, x_2, \dots, x_n)$ with $(y_1, \dots, y_n)$, since $y_i > x_i$ for all $i \in S$ and $\sum_{i \in S} y_i = v(S)$. Thus $(x_1, x_2, \dots, x_n)$ cannot be in the core, yielding a contradiction. ■

EXAMPLE. In the above example, by the theorem the core consists of all the triples (x_1,x_2,x_3) such that:

$x_1 \geq v(\{1\}) = 100$	(1)
$x_2 \geq v(\{2\}) = 125$	(2)
$x_3 \geq v(\{3\}) = 50$	(3)
$x_1 + x_2 \geq v(\{1,2\}) = 270$	(4)
$x_1 + x_3 \geq v(\{1,3\}) = 375$	(5)
$x_2 + x_3 \geq v(\{2,3\}) = 350$	(6)
$x_1 + x_2 + x_3 = v(\{1,2,3\}) = 500$	(7)

From (5) and (7) we get that $x_2 \leq 125$. This, together with (2), gives

$$x_2 = 125. \quad (8)$$

From (7) and (8) we get that $x_1 + x_3 = 375$ so that

$$x_1 = 375 - x_3. \quad (9)$$

From (4) and (8) we get that

$$x_1 \geq 270 - 125 = 145. \quad (10)$$

From (9) and (10) we get that $375 - x_3 \geq 145$ i.e. $x_3 \leq 230$.

From (6) and (8) we get that

$$x_3 \geq 225. \quad (10)$$

Thus the core is the set of triples (x_1, x_2, x_3) such that $x_1 = 375 - x_3$, $x_2 = 125$ and $225 \leq x_3 \leq 230$.