1. Shapeley value theorem:

We assume that the payoff function below is definition of the function of the Shapeley value theorem:

$$\phi(i,N) = \sum_{S\subseteq M\setminus\{i\}} \frac{|S|!\,(|N|-|S|-1)!}{|N|!}\left(v(S\cup\{i\}) - v(S)\right) \qquad (1.1)$$

let R represent a permutation of pleyes set N

let P_i^R represent the set of players permutate before player i in permutation R. than (1.1) is equlvalent to:

$$\phi(i,N) = \sum_R \frac{1}{|N|!}\left(v(P_i^R \cup\{i\}) - v(P_i^R)\right) \quad (1.2)$$

1). Symmetry: $v(S\cup\{i\}) = v(S\cup\{j\})$, if $\forall S$, $i,j \notin S$ (1.3)

Based on (1,2):

$$\phi(i,N) = \sum_R \frac{1}{|N|!}\left(v(P_i^R \cup\{i\}) - v(P_i^R)\right) \quad (1.4)$$

$$\phi(j,N) = \sum_R \frac{1}{|N|!}\left(v(P_j^R \cup\{i\}) - v(P_j^R)\right) \quad (1.5)$$

Let the permutation whose i before j be R_1,

Let the permutation whose j before i be R_2,

Knowably, $R = R_1 \cup R_2$ with $R_1 \cap R_2 = \phi$ (1.6)

for $\forall r \in R_1$, $r = [\underbrace{\ldots}_{r_0} i, \underbrace{\ldots}_{r_1}, j \underbrace{\ldots}_{r_2}]$, correspondingly we have

and only have one $r' \in R_2$, $r' = [\underbrace{\ldots}_{r_0}, j \underbrace{\ldots}_{r_1}, i \underbrace{\ldots}_{r_2}]$ with the part-permutation r_0, r_1 and r_2 are the same.

Obviously, $P_i^{r} = P_j^{r'}$, based on (1.3), we have:

$v(P_i^{r} \cup \{i\}) - v(P_i^{r}) = v(P_j^{r'} \cup \{j\}) - v(P_j^{r'})$ (one-to-one mapping between r & r')

$$\Rightarrow \sum_{R_1} \frac{1}{|N|!} \left(v(P_i^{R'} \cup \{i\}) - v(P_i^{R'}) \right) = \sum_{R_2} \frac{1}{|N|!} \left(v(P_j^{R^2} \cup \{i\}) - v(P_j^{R^2}) \right) \quad (1.7)$$

Similarly, $\sum_{R_2} \frac{1}{|N|!} \left(v(P_i^{R^2} \cup \{i\}) - v(P_i^{R^2}) \right) = \sum_{R_1} \frac{1}{|N|!} \left(v(P_j^{R'} \cup \{i\}) - v(P_j^{R'}) \right)$ (1.8)

Based on (1.6), (1.4) and (1.5) can be transformed into

$$\begin{cases} \phi(i,N) = \sum_{R_1} \frac{1}{|N|!} \left(v(P_i^{R'} \cup \{i\}) - v(P_i^{R'}) \right) + \sum_{R_2} \frac{1}{|N|!} \left(v(P_i^{R^2} \cup \{i\}) - v(P_i^{R^2}) \right) & (1.9) \\ \phi(j,N) = \sum_{R_1} \frac{1}{|N|!} \left(v(P_j^{R'} \cup \{i\}) - v(P_j^{R'}) \right) + \sum_{R_2} \frac{1}{|N|!} \left(v(P_j^{R^2} \cup \{i\}) - v(P_j^{R^2}) \right) & (1.10) \end{cases}$$

Brining (1.7) & (1.8) into (1.9) respectively yields (1.10)

$\Rightarrow \phi(i,N) = \phi(j,N)$, Symmetry satisfied.

2). Dummy: $v(S \cup \{i\}) = v(S) + v(\{i\})$, $\forall S, i \notin S$ (1.11)

Based on (1.2) and (1.11)

$$\begin{aligned} \phi(i,N) &= \sum_{R} \frac{1}{|N|!} \left(v(P_i^{R} \cup \{i\}) - v(P_i^{R}) \right) \\ &= \sum_{R} \frac{1}{|N|!} \left(v(P_i^{R}) + v(\{i\}) - v(P_i^{R}) \right) \\ &= \frac{1}{|N|!} v(\{i\}) \cdot n_R \quad (1.12) \end{aligned}$$

n_R is a full permutation of N, so $n_R = |N|!$ (1.13)

Brining (1.13) into (1.12), we have:

$\phi(i,N) = v(\{i\})$, Dummy satisfied

3) Additivity: $V = V_1 + V_2$ (1.14).

Based on (1.2) and (1.14):

$$\phi(i,N) = \sum_R \frac{1}{|N|!}\left(v(P_i^R \cup \{i\}) - v(P_i^R)\right)$$

$$= \sum_R \frac{1}{|N|!}\left[v_1(P_i^R \cup \{i\}) + v_2(P_i^R \cup \{i\}) - (v_1(P_i^R) + v_2(P_i^R))\right]$$

$$= \sum_R \frac{1}{|N|!}\left[(v_1(P_i^R \cup \{i\}) - v_1(P_i^R)) + (v_2(P_i^R \cup \{i\}) - v_2(P_i^R))\right]$$

$$= \sum_R \frac{1}{|N|!}(v_1(P_i^R \cup \{i\}) - v_1(P_i^R)) + \sum_R \frac{1}{|N|!}(v_2(P_i^R \cup \{i\}) - v_2(P_i^R))$$

$= \phi_1(i,N) + \phi_2(i,N)$, Additivity satisfied

Because the payoff function (1.1) satisfies 1). Symmetry. 2). Dummy 3) Additivity, it is the unique definition of the Shapeley value

2. Integrated Gradient Theorem:

Definition of IG: $\frac{\partial f(\vec{x}' + \alpha(\vec{x} - \vec{x}'))}{\partial [x_i' + \alpha(x_i - x_i')]}$

$$\phi(i,\vec{x}) = (x_i - x_i')\int_{\alpha=0}^{1} \partial_{x_i} f(\vec{x}' + \alpha(\vec{x} - \vec{x}'))\, d\alpha \quad (2.1)$$

1) Completeness: let Path $\vec{\gamma}(\alpha) = \vec{x}' + \alpha(\vec{x} - \vec{x}')$ $\alpha \in [0,1]$ (2.2)

Base on (2.1) and (2.2). we have:

$$f(\vec{x}) - f(\vec{x}') = f(\vec{\gamma}(1)) - f(\vec{\gamma}(0)) = \int_{\vec{\gamma}(0)}^{\vec{\gamma}(1)} \frac{d(f(\vec{\gamma}(\alpha)))}{d(\vec{\gamma}(\alpha))} d(\vec{\gamma}(\alpha))$$

$$= \int_{\alpha=0}^{1} \frac{d(f(\vec{\gamma}(\alpha)))}{d(\vec{\gamma}(\alpha))} \frac{d(\vec{\gamma}(\alpha))}{d\alpha} d\alpha = \int_{\alpha=0}^{1} \langle \nabla_{\vec{\gamma}} f(\vec{\gamma}(\alpha)), \vec{\gamma}'(\alpha) \rangle d\alpha$$

$$(\vec{a} = [a_1, \dots a_n]^T, \vec{\beta} = [b_1, \dots, b_n]^T, \langle \vec{\alpha}, \vec{\beta} \rangle = \vec{\alpha}^T \vec{\beta} = \sum_{i=1}^{n} a_i b_i)$$

$$= \sum_i \int_0^1 [\nabla_{\vec{\gamma}} f(\vec{\gamma}(\alpha))]_i \cdot [\vec{\gamma}'(\alpha)]_i \, d\alpha = \sum_i \int_0^1 \frac{\partial f(\vec{\gamma}(\alpha))}{\partial [\vec{\gamma}(\alpha)]_i} \cdot [x_i - x_i'] d\alpha$$

$$= \sum_i [x_i - x_i'] \int_0^1 \partial_i f(\vec{x}' + \alpha(\vec{x} - \vec{x}')) d\alpha = \sum_i \phi(i, \vec{x})$$

$\Rightarrow f(\vec{x})-f(\vec{x}') = \sum_i \phi(i,\vec{x})$. (2.3) completeness satisfied.

2) linearity: $f = a_1 f_1 + a_2 f_2$ (2.4) based on (2.3):

$f_1(\vec{x})-f_1(\vec{x}') = \sum_i \phi_1(i,\vec{x})$. (2.5) $\quad f_2(\vec{x})-f_2(\vec{x}') = \sum_i \phi_2(i,\vec{x})$. (2.6)

based on (2.4), (2.5), (2.6), we have:

$$\sum_i \phi(i,\vec{x}) = f(\vec{x}) - f(\vec{x}') = a_1 f_1(\vec{x}) + a_2 f_2(\vec{x}) - a_1 f_1(\vec{x}') - a_2 f_2(\vec{x}')$$

$$= a_1 f_1(\vec{x}) - a_1 f_1(\vec{x}') + a_2 f_2(\vec{x}) - a_2 f_2(\vec{x}') = a_1 \sum_i \phi_1(i,\vec{x}) + a_2 \sum_i \phi_2(i,\vec{x})$$

$$= \sum_i [a_1 \phi_1(i,\vec{x}) + a_2 \phi_2(i,\vec{x})]$$

$\Rightarrow \phi(i,\vec{x}) = a_1 \phi_1(i,\vec{x}) + a_2 \phi_2(i,\vec{x})$ (2.7) linearity satisfied.

3) Symmetry: f is Symmetric $\Rightarrow f(x_1, \dots, x_j, \dots x_k, \dots, x_n)$

$= f(x_1, \dots, x_k, \dots x_j, \dots, x_n)$, for $\forall j \neq k \in [1, n]$ (2.8)

let $\vec{x}_{j,k} = (x_1, \dots, x_j, \dots x_k, \dots, x_n)$, $\vec{x}'_{j,k} = (x_1', \dots, x_j', \dots x_k', \dots, x_n')$

$\Rightarrow \vec{x}_{k,j} = (x_1, \dots, x_k, \dots x_j, \dots, x_n)$, $\vec{x}'_{k,j} = (x_1', \dots, x_k', \dots x_j', \dots, x_n')$

It is easy to prove: $f(\vec{x}_{j,k}' + \alpha(\vec{x}_{j,k} - \vec{x}_{j,k}')) = f(\vec{x}_{k,j}' + \alpha(\vec{x}_{k,j} - \vec{x}_{k,j}'))$ (2.9)

based on (2.1) & (2.9) we have:

1° for $\forall i \notin \{j,k\}$, $\phi(i,\vec{x}_{k,j}) = (x_i - x_i') \int_{\alpha=0}^{1} \partial_{x_i} f(\vec{x}_{k,j}' + \alpha(\vec{x}_{k,j} - \vec{x}_{k,j}'))\, d\alpha$

$= (x_i - x_i') \int_{\alpha=0}^{1} \partial_{x_i} f(\vec{x}_{j,k}' + \alpha(\vec{x}_{j,k} - \vec{x}_{j,k}'))\, d\alpha = \phi(i, \vec{x}_{j,k})$

2° for $i \in \{j,k\}$, $\phi(j,\vec{x}_{k,j}) = \phi(k,\vec{x}_{j,k})$, $\phi(k,\vec{x}_{k,j}) = \phi(j,\vec{x}_{j,k})$

$\Rightarrow$ for $\forall$ individual $\phi(i,\vec{x})$, symmetry satisfied

4) Sensitivity:

if $f(x)$ does not depend on i, than in (2.1)

$\partial_{x_i} f(x' + \alpha(x - x')) = 0 \Rightarrow \phi(i,x) = 0$, Sensitivity satisfied.

5). Implementation Invariance:

for different implementation of f, the math formular of f will keep its mapping relationship from each input and its corresponding output, so base on (2.1). $\phi(i, \vec{x})$ will just be related to the input-output mapping relationship of the numerical value of function. Implementation Invariance satisfied.

Because the payoff function (2.1) satisfies 1) Completeness 2) linearity 3) Symmetry 4) Sensitivity 5) Implementation Invariance, it's the unique definition of payoff function of Integrated Gradient Theorem