A short summary of the concept of Shapley decomposition

1) The Case of two variables:

Let $I(a,b)$ be a function depending on two variables $a$ and $b$. Such a function need not be linear. Although Chantreuil and Trannoy (1999) and Sastre et Trannoy (2002) limited their application of the Shapley value to the decomposition of income inequality, Shorrocks (1999) has shown that such a decomposition could be applied to any function.

The idea of the Shapley value is to consider all the possible sequences allowing us to eliminate the variables $a$ and $b$. Let us start with the elimination of the variable $a$. This variable may be the first one or the second one to be eliminated. If it is eliminated first, the function $I(a,b)$ will become equal to $I(a = 0; b \neq 0)$ since the variable $a$ has been eliminated so that in this case the contribution of $a$ to the function $I(a,b)$ is equal to $I(a \neq 0; b \neq 0) - I(a = 0; b \neq 0)$. If the variable $a$ is the second one to be eliminated the function $I$ will then be equal to $I(a \neq 0; b = 0)$. Since both elimination sequences are possible and assuming the probability of these two sequences is the same, we may conclude that the contribution $C(a)$ of the variable $a$ to the function $I(a,b)$ is equal to

$$C(a) = \frac{1}{2}\{I(a \neq 0; b \neq 0) - I(a = 0; b \neq 0)\}$$
$$+ \frac{1}{2}\{I(a \neq 0; b = 0) - I(a = 0; b = 0)\}$$

Similarly one can prove that the contribution $C(b)$ of the variable $b$ to the function $I(a,b)$ is

$$C(b) = \frac{1}{2}\{I(a \neq 0; b \neq 0) - I(a \neq 0; b = 0)\}$$
$$+ \frac{1}{2}\{I(a = 0; b \neq 0) - I(a = 0; b = 0)\}$$
Combining (1) and (2) we observe that

$$C(a) + C(b) = I(a,b)$$  \hspace{1cm} (3)

2) The case of three variables:

Assume an indicator $I$ is a function of three determinants $a, b, c$ and is written as $I = I(a, b, c)$.

There are obviously $3! = 6$ ways of ordering these three determinants $a, b$ and $c$:

$$(a,b,c), (a,c,b), (b,a,c), (b,c,a), (c,a,b), (c,b,a)$$  \hspace{1cm} (4)

Each of these three determinants may be eliminated first, second or third. The respective (marginal) contributions of the determinants $a, b, c$ will hence be a function of all the possible ways in which each of these determinants may be eliminated. Let for example $C(a)$ be the marginal contribution of $a$ to the indicator $I(a, b, c)$.

If $a$ is eliminated first its contribution to the overall value of the indicator $I$ will be expressed as $I(a, b, c) - I(b, c)$ where $I(b, c)$ corresponds to the case where $a$ is equal to zero. Since expression (4) indicates that there are two cases in which $a$ appears first and may thus be eliminated first we will give a weight of $(2/6)$ to this possibility.

If $a$ is eliminated second, it implies that another determinant has been eliminated first (and been assumed to be equal to 0). Expression (4) indicates that there are two cases in which this possibility occurs, the one denoted in (1) as $(b, a, c)$ and the one denoted $(c, a, b)$. In the first case the contribution of $a$ will be written as $I(a, c) - I(c)$ while in the second it is expressed as $I(a, b) - I(b)$. To each of these two cases we evidently give a weight of $(1/6)$.

Finally if $a$ is eliminated third, it implies that both $b$ and $c$ are assumed to be equal to 0. Expression (4) indicates that there are two such cases, the one denoted $(b, c, a)$ and
the one denoted \((c,b,a)\). Since we may assume that when each of the three determinants is equal to 0, the indicator \(I\) is equal to 0, we may write that the contribution of \(a\) in this case will be equal to \(I(a) - 0 = I(a)\) and evidently we have to give a weight of \((2/6)\) to such a possibility since there are two such cases.

We may therefore summarize what we have just explained by stating that the marginal contribution \(C(a)\) of the determinant \(a\) to the overall value of the indicator \(I\) may be written as

\[
C(a) = \frac{2}{6}[I(a = 0; b = 0; c = 0) - I(a = 0; b = 0; c = 0)] \\
+ \frac{1}{6}[I(a = 0; b = 0; c = 0) - I(a = 0; b = 0; c = 0)] \\
+ \frac{1}{6}[I(a = 0; b = 0; c = 0) - I(a = 0; b = 0; c = 0)] \\
+ \frac{2}{6}[I(a = 0; b = 0; c = 0) - I(a = 0; b = 0; c = 0)]
\]

(5)

One can similarly determine the marginal contribution \(C(b)\) of \(b\) and \(C(c)\) of \(c\) and then find out that

\[
I(a,b,c) = C(a) + C(b) + C(c)
\]

(6)

3) The case of four variables:

Let us give a simple illustration where the index \(I\) depends on four determinants \(a,b,c,d\). Table 1 below gives all the possible ways of ordering these four elements.
Table 1: The 24 ways of ordering four elements

<table>
<thead>
<tr>
<th>a appears in the first position</th>
<th>a appears in the second position</th>
<th>a appears in the third position</th>
<th>a appears in the fourth position</th>
</tr>
</thead>
<tbody>
<tr>
<td>abcd</td>
<td>bacd</td>
<td>bcad</td>
<td>Bcd</td>
</tr>
<tr>
<td>abdc</td>
<td>badc</td>
<td>bdac</td>
<td>Bdeca</td>
</tr>
<tr>
<td>acbd</td>
<td>cabd</td>
<td>cbad</td>
<td>Cbda</td>
</tr>
<tr>
<td>acdb</td>
<td>cadb</td>
<td>cdab</td>
<td>Cdbba</td>
</tr>
<tr>
<td>adbc</td>
<td>dabc</td>
<td>dbac</td>
<td>Dbca</td>
</tr>
<tr>
<td>adcb</td>
<td>Daacb</td>
<td>dcab</td>
<td>Dcbb</td>
</tr>
</tbody>
</table>

The so-called Shapley decomposition looks at all possible elimination sequences. As Table 1 indicates, there are 6 cases where a appears in the first position, 6 where it appears in the second, etc…

If we look at the various ways of eliminating a, we can say that if a is eliminated first its contribution to the indicator will be equal to the difference between the value of the indicator when all four determinants are different from zero and its value when a is equal to zero. In that case (a eliminated first) the contribution of a will be written as

\[ C(a/a \text{ eliminated first}) = [I(a \neq 0, b \neq 0, c \neq 0, d \neq 0) - I(a=0, b \neq 0, c \neq 0, d \neq 0)] \]

Clearly, looking at the first column of the table above, there are six such cases out of 24 possible orderings.
The case where \( a \) is eliminated second is based on the orderings given in the second column of this table. There are three possibilities:

- \( b \) is eliminated first and \( a \) second, so that the contribution of \( a \) in this case will be written as

\[
C(a/a\text{ eliminated second and } b \text{ first}) = [I(a \neq 0, b = 0, c \neq 0, d \neq 0) - I(a = 0, b = 0, c \neq 0, d \neq 0)]
\]

and this case appears twice in the second column of the table.

- \( c \) is eliminated first and \( a \) second so that the contribution of \( a \) in this case will be written as

\[
C(a/a\text{ eliminated second and } c \text{ first}) = [I(a \neq 0, b \neq 0, c = 0, d \neq 0) - I(a = 0, b \neq 0, c = 0, d \neq 0)]
\]

and this case appears also twice in the second column of the table.

- \( d \) is eliminated first and \( a \) second so that the contribution of \( a \) in this case will be written as

\[
C(a/a\text{ eliminated second and } d \text{ first}) = [I(a \neq 0, b \neq 0, c \neq 0, d = 0) - I(a = 0, b \neq 0, c \neq 0, d = 0)]
\]

and this case appears also twice in the second column of the table.
The case where \( a \) is eliminated third is based on the orderings given in the third column of this table. There are again three possibilities:

- \( b \) is eliminated first, \( c \) second and a third or \( c \) is eliminated first, \( b \) second and \( a \) third. In both cases the contribution of \( a \) will be written as

\[
C(a/a \text{ eliminated third and } b \text{ and } c \text{ first or second}) = [I(a \neq 0, b=0, c=0, d \neq 0) - I(a=0, b=0, c=0, d \neq 0)]
\]

- \( b \) is eliminated first, \( d \) second and \( a \) third or \( d \) is eliminated first, \( b \) second and \( a \) third. In both cases the contribution of \( a \) will be written as

\[
C(a/a \text{ eliminated third and } b \text{ and } d \text{ first or second}) = [I(a \neq 0, b=0, c \neq 0, d=0) - I(a=0, b=0, c \neq 0, d=0)]
\]

- \( c \) is eliminated first, \( d \) second and \( a \) third or \( d \) is eliminated first, \( c \) second and \( a \) third. In both cases the contribution of \( a \) will be written as

\[
C(a/a \text{ eliminated third and } c \text{ and } d \text{ first or second}) = [I(a \neq 0, b \neq 0, c=0, d=0) - I(a=0, b \neq 0, c=0, d=0)]
\]

Finally the cases where \( a \) is eliminated last (fourth) appear in the fourth column of the table above. There are 6 such cases but in each of these cases the contribution of \( a \) may be expressed as
C(a/a eliminated fourth while b, c and d eliminated first, second or third)

= \[I(a \neq 0, b = 0, c = 0, d = 0) - I(a = 0, b = 0, c = 0, d = 0)\]

Taking all the 24 cases into account we therefore conclude that the overall contribution of \(a\) to the indicator \(I\) may be expressed as

\[
C(a) = \{ (6/24) \left[ I(a \neq 0, b \neq 0, c \neq 0, d \neq 0) - I(a = 0, b = 0, c = 0, d = 0) \right] \\
+ (2/24) \left[ I(a \neq 0, b = 0, c \neq 0, d \neq 0) - I(a = 0, b = 0, c = 0, d \neq 0) \right] \\
+ (2/24) \left[ I(a \neq 0, b = 0, c = 0, d \neq 0) - I(a = 0, b = 0, c = 0, d \neq 0) \right] \\
+ (2/24) \left[ I(a \neq 0, b = 0, c = 0, d = 0) - I(a = 0, b = 0, c = 0, d = 0) \right] \\
+ (2/24) \left[ I(a \neq 0, b = 0, c = 0, d = 0) - I(a = 0, b = 0, c = 0, d = 0) \right] \\
+ (6/24) \left[ I(a \neq 0, b = 0, c = 0, d = 0) - I(a = 0, b = 0, c = 0, d = 0) \right] \}
\]

One can naturally derive in a similar way the overall contributions \(C(b)\) of \(b\), \(C(c)\) of \(c\) and \(C(d)\) of \(d\) to the value of the indicator \(I\). Moreover it is also easy to verify that

\[
C(a) + C(b) + C(c) + C(d) = I(a \neq 0, b \neq 0, c \neq 0, d \neq 0)
\]

where \(I(a \neq 0, b \neq 0, c \neq 0, d \neq 0)\) is in fact the original value of the index \(I\).
4) The case of five variables:

Let us give a simple illustration where the index $I$ depends on five determinants $a,b,c,d,e$. Table 2 below gives all the possible ways of ordering these five elements.

**Table 1: The 120 ways of ordering five elements**

<table>
<thead>
<tr>
<th>$a$ appears in the first position</th>
<th>$a$ appears in the second position</th>
<th>$a$ appears in the third position</th>
<th>$a$ appears in the fourth position</th>
<th>$a$ appears in the fifth position</th>
</tr>
</thead>
<tbody>
<tr>
<td>abcde</td>
<td>bacde</td>
<td>bcade</td>
<td>bcdae</td>
<td>bcdea</td>
</tr>
<tr>
<td>abced</td>
<td>baced</td>
<td>bcaed</td>
<td>bcdea</td>
<td>bceda</td>
</tr>
<tr>
<td>abdec</td>
<td>badce</td>
<td>bdace</td>
<td>bdace</td>
<td>bdcea</td>
</tr>
<tr>
<td>abdce</td>
<td>badec</td>
<td>bdaec</td>
<td>bdaec</td>
<td>bdeca</td>
</tr>
<tr>
<td>abedc</td>
<td>baecd</td>
<td>beacd</td>
<td>beacd</td>
<td>becdca</td>
</tr>
<tr>
<td>acbde</td>
<td>cabde</td>
<td>cbade</td>
<td>cbade</td>
<td>cbdea</td>
</tr>
<tr>
<td>acbed</td>
<td>cabde</td>
<td>cbade</td>
<td>cbade</td>
<td>cbdea</td>
</tr>
<tr>
<td>acdbe</td>
<td>cadbe</td>
<td>cdaeb</td>
<td>cdaeb</td>
<td>cdeba</td>
</tr>
<tr>
<td>acdeb</td>
<td>cadeb</td>
<td>cdaeb</td>
<td>cdeab</td>
<td>cdeba</td>
</tr>
<tr>
<td>acbed</td>
<td>caedc</td>
<td>ceabd</td>
<td>cebad</td>
<td>cebda</td>
</tr>
<tr>
<td>acebd</td>
<td>caedc</td>
<td>ceabd</td>
<td>cebad</td>
<td>cebda</td>
</tr>
<tr>
<td>acedb</td>
<td>caedc</td>
<td>ceabd</td>
<td>cedab</td>
<td>cedba</td>
</tr>
<tr>
<td>adebc</td>
<td>daebc</td>
<td>dcbae</td>
<td>dcaeb</td>
<td>dcbea</td>
</tr>
<tr>
<td>adebc</td>
<td>daebc</td>
<td>dcbae</td>
<td>dcbae</td>
<td>dcbea</td>
</tr>
<tr>
<td>adebc</td>
<td>daebc</td>
<td>dcbae</td>
<td>dcbae</td>
<td>dcbea</td>
</tr>
<tr>
<td>adebc</td>
<td>daebc</td>
<td>dcbae</td>
<td>dcbae</td>
<td>dcbea</td>
</tr>
</tbody>
</table>
From this table it is easy to deduct that the contribution of determinant \( a \), written as

\[
C(a) = (24/120)[I(a \neq 0; b \neq 0; c \neq 0; d \neq 0; e \neq 0) - I(a = 0; b \neq 0; c \neq 0; d \neq 0; e \neq 0)]
\] 
\[
+ (6/120)[I(a \neq 0; b = 0; c \neq 0; d \neq 0; e \neq 0) - I(a = 0; b = 0; c \neq 0; d \neq 0; e \neq 0)]
\] 
\[
+ (6/120)[I(a \neq 0; b \neq 0; c = 0; d \neq 0; e \neq 0) - I(a = 0; b \neq 0; c = 0; d \neq 0; e \neq 0)]
\] 
\[
+ (6/120)[I(a \neq 0; b \neq 0; c \neq 0; d = 0; e \neq 0) - I(a = 0; b \neq 0; c \neq 0; d = 0; e \neq 0)]
\] 
\[
+ (6/120)[I(a \neq 0; b \neq 0; c \neq 0; d \neq 0; e = 0) - I(a = 0; b \neq 0; c \neq 0; d \neq 0; e = 0)]
\] 
\[
+ (4/120)[I(a \neq 0; b = 0; c \neq 0; d \neq 0; e = 0) - I(a = 0; b = 0; c \neq 0; d \neq 0; e = 0)]
\] 
\[
+ (4/120)[I(a \neq 0; b = 0; c \neq 0; d = 0; e \neq 0) - I(a = 0; b = 0; c \neq 0; d = 0; e \neq 0)]
\] 
\[
+ (4/120)[I(a \neq 0; b = 0; c = 0; d \neq 0; e = 0) - I(a = 0; b = 0; c = 0; d \neq 0; e = 0)]
\] 
\[
+ (4/120)[I(a \neq 0; b = 0; c = 0; d = 0; e = 0) - I(a = 0; b = 0; c = 0; d = 0; e = 0)]
\] 
\[
+ (6/120)[I(a \neq 0; b = 0; c = 0; d \neq 0; e = 0) - I(a = 0; b = 0; c = 0; d \neq 0; e = 0)]
\] 
\[
+ (6/120)[I(a \neq 0; b = 0; c = 0; d = 0; e \neq 0) - I(a = 0; b = 0; c = 0; d = 0; e \neq 0)]
\] 
\[
+ (6/120)[I(a \neq 0; b = 0; c = 0; d = 0; e = 0) - I(a = 0; b = 0; c = 0; d = 0; e = 0)]
\] 
\[
+ (6/120)[I(a \neq 0; b = 0; c = 0; d = 0; e = 0) - I(a = 0; b = 0; c = 0; d = 0; e = 0)]
\] 
\[
+ (6/120)[I(a \neq 0; b = 0; c = 0; d = 0; e \neq 0) - I(a = 0; b = 0; c = 0; d = 0; e \neq 0)]
\] 
\[
+ (6/120)[I(a \neq 0; b = 0; c = 0; d = 0; e \neq 0) - I(a = 0; b = 0; c = 0; d = 0; e \neq 0)]
\] 
\[
+ (6/120)[I(a \neq 0; b = 0; c = 0; d = 0; e \neq 0) - I(a = 0; b = 0; c = 0; d = 0; e \neq 0)]
\] 
\[
+ (24/120)[I(a \neq 0; b = 0; c = 0; d = 0; e = 0) - I(a = 0; b = 0; c = 0; d = 0; e = 0)]
\]
One can similarly derive the contributions of the other four determinants \((C(b),C(c),C(d) \text{ and } C(e))\).