Direction des Statistiques démographiques et sociales

N°F1202

LA VALEUR DE SHAPLEY - COMMENT INDIVIDUALISER LE RÉSULTAT D'UN GROUPE

ALEXIS EIDELMAN

Document de travail

Institut National de la Statistique et des Études Économiques

INSTITUT NATIONAL DE LA STATISTIQUE ET DES ETUDES ECONOMIQUES

Série des Documents de Travail
de la

DIRECTION DES STATISTIQUES DEMOGRAPHIQUES ET SOCIALES

Division Etudes sociales

N° F1202

La valeur de Shapley

Comment individualiser le résultat d'un groupe

Alexis Eidelman

janvier 2012

Résumé

Il arrive fréquemment dans les sciences économiques que l'on cherche à individualiser un résultat collectif : quelle est la contribution d'un pays dans un résultat calculé au niveau d'une zone ? Quel est le rôle de chaque membre d'un ménage dans l'attribution d'une prestation ? Comment isoler l'effet d'un facteur dans une évolution ? Quel est le rôle d'un transfert dans la réduction globale des inégalités ? Quelle est l'influence d'une sous-population sur le taux de pauvreté ? Toutes ces questions permettent d'isoler un effet pour comprendre un mécanisme ou simplement de donner de l'information plus précise. De nombreux exemples de telles situations seront donnés dans ce document mais chacun pourra, en fonction de sa problématique, y trouver son intérêt propre. Le premier réflexe pour déterminer ce qu'apporte un élément donné à un groupe consiste en général à regarder sa contribution valeur, ce que fait le groupe avec cet individu et ce qu'il fait sans lui. Ce document montre que cette démarche n'est pas neutre et que les valeurs obtenues ne respectent pas certaines propriétés importantes.

On propose alors le recours à la valeur de Shapley qui est, en quelque sorte, une version améliorée de la valeur marginale prenant minutieusement en compte les interactions entre les différents agents. Son application requiert de connaitre le résultat du groupe mais aussi le résultat sur tous les sous-groupes que l'on peut constituer avec ses membres. Dans certains cas, cela constitue un obstacle fort ; toutefois, cette difficulté, quand elle se présente, peut parfois être contournée au moyen de la méthode d'Owen ou des Shapley emboités.

Abstract

Trying to get a personal contribution in a context of collective performance is a problem that quite frequently occurs in economics : what is a country's contribution to a region result ? What role is played by each agent in recieving an means-tested allocation ? How to isolate a single factor's effect to an evolution ? How to calculate the importance of a social transfer in reducing total income inequalities ? How each subpopulation influence the poverty rate ? These questions enable to understand a mecanism or to give more precise information. Many various examples are given in that document but anyone can find his or her own interest depending of his or her topic. The first idea to study individual importance is to calculate his marginal contribution which is the difference between what the team can make with that individual and what the team can do without him. It is shown in this document that this way of doing is not neutral and that the contributions obtained this way do not respect important properties.

Using the Sahpley value is then suggested. It is, in a way, a marginal contribution in an improved way which carefully takes into account all interactions between agents. To implement the Shapley value's procedure, results of all subgroup must be known. It can be a serious limit but sometimes Owen method or nested-Shapley are a solution to that limit.

Table des matières

1 Introduction

Ce document de travail a pour but de présenter simplement la valeur de Shapley sans chercher à être exhaustif. Elle peut aider le statisticien et l'économiste dans bien des domaines. On présentera des exemples et on explicitera les limites de l'utilisation de la valeur de Shapley afin que le lecteur puisse rapidement identifier les cas dans lesquels le recours à la valeur de Shapley est possible et utile à son analyse.

Pour donner un aperçu de l'étendue du champ d'application de la valeur de Shapley, disons qu'on peut l'utiliser pour mesurer la productivité d'un employé (ou d'un dirigeant d'entreprise), déterminer un pouvoir de vote, savoir ce qu'un joueur apporte vraiment à une équipe, individualiser des revenus même lorsqu'ils sont soumis à des conditions de ressources familiales ou encore déterminer le poids d'un pays dans la balance commerciale européenne.

Position du problème Il y a beaucoup d'exemples de la vie courante dans lesquels on observe seulement le résultat collectif d'un groupe : le bénéfice réalisé par un ensemble de travailleurs, la décision d'un corps électoral, le résultat d'une équipe sportive, etc. On peut se demander comment chaque individu contribue à forger ce résultat ? Que doit le résultat collectif à chacun ? C'est la question qui est étudiée ici. Dans la suite du document, on dira qu'il y a n joueurs d'un ensemble E qui s'associent pour obtenir un résultat. Ce résultat est une valeur $v(\mathrm{E})$ (dans $\mathbb{R}$). On veut déterminer une valeur $\phi(i)$ à chaque joueur i qui représente sa participation au résultat $v(\mathrm{E})$.

Pour illustrer le document, on se placera dans le cadre d'une entreprise rassemblant n individus devant se partager équitablement le bénéfice $v(E)$ gagné par l'entreprise. On suppose que l'on connait pour chaque sous-ensemble F de E, le bénéfice $v(\mathrm{F})$ que ferait l'entreprise constituée par les individus présents dans F. On cherche à déterminer la rémunération que doit recevoir chaque employé i de l'entreprise. C'est l'interprétation que l'on donne dans cet exemple à la valeur $\phi(i)$ de chaque joueur.

Dans la partie suivante, on établira les critères que l'on veut respecter dans l'attribution d'une rémunération juste à chacun. On verra ensuite que rémunérer les employés en fonction de leur valeur intrinsèque ou marginale ne respecte pas ces propriétés pourtant simples. La partie 4 introduira la valeur de Shapley qui permet de répartir un résultat collectif entre les joueurs ayant permis de l'obtenir. Des exemples d'application divers seront donnés ensuite. Ces cas pratiques auront aussi vocation à expliciter des problèmes que l'on peut rencontrer et les astuces que l'on peut employer. Cela amènera naturellement à une partie présentant les limites de l'utilisation de la valeur de Shapley et montrera quelques variantes utiles avant une conclusion.

2 Propriétés attendues de la valeur d'un individu

Les précisions fines et techniques seront données plus loin. Pour l'instant, en guise d'introduction, on peut s'accorder sur plusieurs propriétés que devraient avoir la fonction ϕ, donnant la contribution de chaque individu au résultat global du groupe.

2.1 Répartir le bénéfice

D'abord, et cela semble assez peu discutable, il faut répartir l'ensemble du bénéfice, il faut que la valeur de l'ensemble des joueurs soit égale à l'ensemble des valeurs de chaque joueur. Autrement dit, on souhaite que la valeur des joueurs respecte :

$$\sum_{i=1}^{n} \phi(i) = v(\mathrm{E})$$

Selon le contexte cette condition est appelée au choix : contrainte de budget, condition d'exactitude ou axiome de balance.

2.2 Ne pas récompenser les joueurs inutiles

Une infinité de fonction ϕ remplisse la condition ci-dessus. Une triviale est celle qui donne la même part à tout le monde : $\forall\ i, \phi(i) = \frac{v(\mathrm{E})}{n}$. On peut considérer, par principe, que tous les joueurs ont la même valeur et doivent recevoir le même montant. Toutefois, on rejette en général cette répartition égalitaire parce qu'elle ne tient aucunement compte de l'investissement de i et de son influence sur $v(\mathrm{E})$. C'est une évidence mais on souhaite que la valeur de i, le salaire qui lui est versé et le mérite qu'on lui accorde soient en lien avec sa contribution. Si i est transparent, c'est à dire s'il n'influence pas le niveau des bénéfices alors il ne doit rien avoir. Le sens de « influencer la valeur totale » est crucial dans l'attribution d'une valeur comme nous le verrons plus loin. Pour l'instant, donnons une formulation forte de la transparence : un joueur est dit transparent s'il n'influence jamais le score d'aucun sous groupe de $\mathrm{E} \setminus \{i\}$ qui est l'ensemble E sans l'élément i et que l'on note désormais E^i.

$$\forall\ \mathrm{F} \in \mathcal{P}(\mathrm{E}^i), v(\mathrm{F} \cup i) = v(\mathrm{F}) \Longrightarrow \phi(i) = 0$$

où $\mathcal{P}(\mathrm{E}^i)$ désigne l'ensemble des groupes ne contenant pas i que l'on peut former avec des éléments de E.

2.3 Être symétrique

C'est une condition fondamentale puisqu'on l'associe fortement au concept du justice. On traite tous les joueurs de la même façon. On parle parfois de condition d'anonymat pour insister sur le fait qu'on ne donne aucun privilège à une valeur donnée.

Une conséquence du fait que la valeur des joueurs ne dépend pas du nom i des joueurs est que si tous les joueurs ont un rôle symétrique, ils doivent tous avoir la même valeur. Cette valeur est alors évidemment $\frac{v(\mathrm{E})}{n}$.

2.4 Valeur sur des organisations particulières

Pour prolonger cette section sur les propriétés attendues de la façon de déterminer la valeur d'un individu, donnons des exemples d'organisations d'entreprise, des exemples d'interactions entre les joueurs. Ces exemples sont justement des cas où tous les joueurs jouent le même rôle et où l'on sait quelle valeur est attendue avec le point précédent : $\frac{v(\mathrm{E})}{n}$.

Une organisation particulière (en physique, on les appellerait des cas-limites) est celle d'une **production en parallèle** où les joueurs n'interagissent pas du tout entre eux. Chacun produit individuellement $\frac{v(\mathrm{E})}{n}$ sans modifier le travail des autres. Cela qui amène naturellement au fait que l'ensemble des joueurs produisent $n \times \frac{v(\mathrm{E})}{n} = v(\mathrm{E})$. Disons tout de suite qu'il y a peu de valeur ϕ raisonnable qui ne donne pas, dans ce cas, la valeur $\phi(i) = \frac{v(\mathrm{E})}{n}$ à chaque joueur.

Une autre organisation particulière est celle d'une **chaine de travail** dans laquelle d'une part un joueur ne peut rien produire seul $\forall i,\ v(\{i\}) = 0$ et où, dans le même temps, chaque joueur est absolument indispensable pour générer $v(\mathrm{E})$, on donc : $\forall i,\ v(\mathrm{E} \setminus \{i\}) = v(\mathrm{E}^i) = 0$ et même, plus précisément : $\forall\ \mathrm{F} \in \mathcal{P}(\mathrm{E}^i)), F \neq \mathrm{E},\ v(\mathrm{F}) = 0$.

Enfin, on peut imaginer une organisation que l'on peut qualifier de **partiellement redondante** dans laquelle tous les joueurs sont équivalents et dans laquelle l'absence d'un individu peut se combler. Si un joueur manque, chacun en fait un peu plus de façon à ce que la production du groupe reste $v(\mathrm{E})$. En revanche, il ne s'agit pas de dire que tous les joueurs sont transparents. En effet, une fois que i n'est plus dans E, si j quitte $E^i = \mathrm{E} \setminus \{i\}$ alors on la production n'est plus conservée : $v(\mathrm{E}) \neq v(\mathrm{E}^{ij})$.

3 La valeur marginale, la valeur intrinsèque et les interactions

Pour regarder la valeur d'un joueur deux idées viennent spontanément à l'esprit : regarder ce qu'il peut faire seul et regarder ce que le groupe fait sans lui. On peut appeler la première valeur, valeur intrinsèque, ϕ^{intra}. Il y a une certaine logique à valoriser un individu en fonction de ses qualités propres, c'est à dire en fonction de sa valeur lorsqu'il est seul. La seconde valeur est une marginale ϕ^{marg}, c'est ce que l'individu apporte au groupe s'il arrive le dernier. Cette valeur marginale est assez attirante. On apprend à s'en servir par exemple en biologie où pour déterminer la fonction d'un organe, on observe ce que l'organisme ne peut plus faire en son absence. On peut facilement définir mathématiquement ces deux valeurs comme suit :

$$\phi^{intra}(i) = v(\{i\}) \ ; \ \phi^{marg}(i) = v(\mathrm{E}) - v(\mathrm{E}^i)$$

On sent bien l'importance que peuvent avoir ces deux valeurs dans des négociations de répartition de bénéfice, cependant on va montrer qu'elles ne sont pas satisfaisantes. La première critique que l'on peut leur adresser c'est qu'elles ne vérifient pas en général la propriété d'efficacité (ou de budget ou de cohérence) $\sum_{i=1}^{n} \phi(i) = v(\mathrm{E})$. Il est néanmoins possible de répondre facilement à cette critique en renormalisant les valeurs afin qu'elles vérifient la propriété désirée. Elles deviennent donc :

$$\phi^{intra}(i) = \frac{v(\{i\})}{\sum_{j=1}^{n} v(\{j\})} v(\mathrm{E}) \ ; \ \phi^{marg}(i) =) \frac{v(\mathrm{E}) - v(\mathrm{E}^i)}{\sum_{j=1}^{n} v(\mathrm{E}) - v(\mathrm{E}^j)} v(\mathrm{E})$$

Cela étant fait, il faut noter que la valeur intrinsèque ne l'est plus vraiment puisqu'elle dépend de la valeur intrinsèque des autres joueurs. De la même manière, la valeur marginale n'est plus simplement ce que i apporte au reste du groupe puisque interviennent les valeurs marginales des autres joueurs calculées à partir d'un ensemble E^j dans lequel i est présent. On doit se résoudre à ne pas pouvoir étudier i isolément : ni sa seule présence, ni sa seule absence.

Plus embêtant, si les deux valeurs donnent bien la valeur attendue pour une organisation en parallèle (cf 2.4), la valeur intrinsèque n'est pas acceptable pour la chaine de travail :

$$\forall i, \ \phi^{intra}(i) = v(\{i\}) = 0$$

Une valeur nulle pour tous les joueurs ne respecte pas le critère d'efficacité et la version normalisée n'est même pas définie. De la même façon, dans un jeu partiellement redondant la valeur marginale n'est pas acceptable :

$$\phi^{marg}(i) = v(\mathrm{E}) - v(\mathrm{E}^i) = v(\mathrm{E}) - v(\mathrm{E}) = 0$$

Ces « organisations » qui montrent que certaines valeurs intuitives ne sont pas donc pas acceptables, peuvent sembler construites de toutes pièces et s'écarter de la généralité. Même si elles le sont effectivement un peu, elles soulignent les faiblesses de ces valeurs et permettent de comprendre et de sentir ce que doit être une valeur ϕ adaptée à notre problème de répartition d'un résultat collectif.

On rejette la valeur intrinsèque parce qu'il est injuste qu'un individu qui participe activement à un travail ne reçoive rien au prétexte qu'il ne peut rien faire sans les autres : il faut tenir compte de l'environnement que représentent les autres joueurs. De la même façon, il est injuste qu'un individu dont l'absence à partir de la situation observée, où tous les joueurs sont réunis, peut être comblée, ne soit pas rémunéré alors qu'il n'est pas transparent et sert tout de même à l'entreprise. De plus, juger de l'inefficacité d'une personne et lui attribuer une valeur nulle, en ne regardant que la situation finale est quelque peu injuste et inefficace, surtout quand on sait que l'individu peut faire pression en s'associant avec une autre personne.

On sent que c'est dans la prise en compte des interactions entre joueurs que pêchent les valeurs marginale et intrinsèque. L'une ne regarde une interaction particulière, la seconde aucune interaction avec les autres joueurs.

4 La valeur de Shapley

4.1 L'exemple du manager

On commence à sentir, à la suite de la section précédente que la clé de voute d'une valeur, c'est la façon dont elle gère les interactions entre les joueurs et entre les groupes de joueurs. Un dernier exemple qui nous amènera naturellement à la valeur de Shapley est l'exemple du manager. L'exemple d'un meneur de troupe (un capitaine, un manager, etc.) est marquant parce il ne produit rien lorsqu'il est seul, en revanche son impact marginal est élevé, en particulier lorsque la taille de l'entreprise est grande. Ainsi, son impact marginal dépend de la présence des autres employés, on peut donc estimer que le mérite de cette valeur marginale leur revient un peu. Imaginons un exemple chiffré pour illustrer ce point :

Exemple 1

- Dans une entreprise, il y $n-1$ ouvriers. Ils fonctionnent « en parallèle », c'est-a-dire qu'ils n'interagissent pas entre eux et produisent chacun une unité.
- Il y a aussi un manager chargé d'organiser le travail, il a la faculté d'augmenter la productivité des autres employés et fait passer leur production individuelle de 1 à 2. C'est sa seule qualité et n'est pas capable de produire quelque chose seul.

Comment répartir le bénéfice total $v(E) = 2(n-1)$ entre les ouvriers et le manager ? La valeur intrinsèque d'un ouvrier est 1, sa valeur marginale est 2. La valeur intrinsèque du manager est 0, sa valeur marginale est $n-1$. Pour chaque ouvrier, le score de l'entreprise augmente de 2 unités, le mérite de la première unité ne semble devoir revenir à personne d'autre qu'à lui-même. En revanche la seconde doit être partagée. En effet, sans le manager, il n'y aurait pas cette unité supplémentaire, sans l'ouvrier non plus. Une façon intuitivement « juste » de répartir le bénéfice de l'entreprise est de partager cette unité en deux, en traitant ainsi sur un pied d'égalité vis-à-vis de cette unité supplémentaire le manager et l'ouvrier[1]. Au final, une idée peut être d'attribuer une de 1,5 à chaque ouvrier de $\frac{n-1}{2}$ au manager. Ici l'intuition donne une solution, ce peut ne plus être le cas dans l'exemple suivant, légèrement plus complexe.

Exemple 2

- On suppose qu'il n'y a que trois ouvriers, produisant toujours 1, indépendamment des autres.
- Le manager augmente toujours la productivité des trois ouvriers de 3 ; en revanche avec deux, un seul ou zéro ouvrier, il ne servirait à rien.

Difficile dans cet exemple de répartir les 3 de surplus du manager à égalité entre le manager et les ouvriers. En effet, ce manager a l'air moins efficace que celui de l'exemple 1 même si dans la situation observée, ils ont le même rôle. Dans la situation précédente, si un ouvrier partait, il faisait perdre 2 à l'entreprise ; dans ce nouvel exemple, il fait perdre 4 puisque les deux ouvriers et le manager restant ne peuvent plus produire que 2, le manager étant alors inefficace. En théorie des jeux, on dirait que l'ouvrier a un pouvoir de négociation plus fort, il va donc capter *a priori* une part plus importante du $n-1=3$ marginal du manager. Seulement, il faut noter que si la valeur marginale d'un ouvrier est plus forte, elle dépend maintenant de la configuration. Il est vrai que lorsqu'il y a trois joueurs sa valeur marginale est de 4, mais il doit cette situation au fait qu'il y a deux autres ouvriers, en effet, lorsqu'il n'y en a qu'un ou zéro, sa valeur marginale n'est plus de 4 ni même de deux comme elle l'était dans l'exemple 1 mais seulement de 1. Pour résumer, un ouvrier a un pouvoir de négociation plus fort mais il doit le partager avec les autres, exactement comme le pouvoir de négociation du manager (les 3 qu'il arrive à apporter à l'entreprise) doit se partager avec les ouvriers.

A ce stade, on commence à pressentir qu'il faut étudier non pas une situation donnée, mais tout un ensemble de situations qui permettent de mieux prendre en compte les interactions complexes entre les joueurs. L'exemple

[1] On peut imaginer que dans un cas concret, l'ordre d'arrivée dans l'entreprise soit important. On attribue l'unité supplémentaire à celui qui est arrivé le dernier. Ici, cela peut conduire à une répartition discutable dans laquelle deux ouvrier faisant le même travail seront considérés différemment s'ils sont arrivés avant ou après le manager. La valeur de Shapley peut se calculer et être appliquée lorsque l'on considère qu'il y a un ordre sur le groupe et que la valeur de (A,B,C) n'est pas la même que celle de (B,A,C) ; toutefois, par souci de simplicité, nous raisonnons ici sur des ensembles non ordonnées : l'ordre d'arrivée ni aucun autre ordre ne rentrent en compte dans le résultat d'un sous-ensemble dans la détermination de la valeur d'un joueur.

ci-dessous est peu différent des précédents mais il montre le pas à franchir pour déterminer une bonne valeur de chaque joueur.

Exemple 3

- On suppose qu'il n'y a que trois ouvriers, produisant toujours 1, indépendamment les uns des autres.
- Le manager augmente toujours la productivité des trois ouvriers de 3, avec un seul ou zéro ouvrier, il ne sert à rien. En revanche, comme dans le premier exemple, il augmente toujours la productivité de deux ouvriers de deux unités.

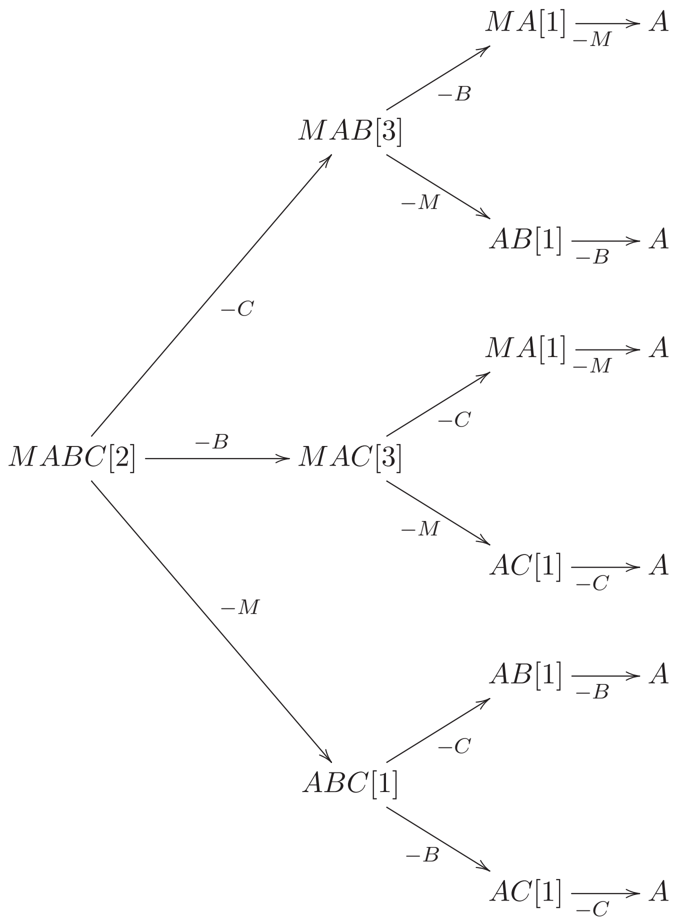
La proximité avec le premier exemple est renforcée par rapport au deuxième puisque non seulement le score de l'équipe est le même mais il ne change pas non plus si on retire un joueur, qu'il soit ouvrier ou manager. La valeur marginale est donc exactement la même si on ne regarde que la situation observée des quatre joueurs. Cependant, là encore, le manager semble moins performant que celui du premier exemple puisqu'il n'est pas capable de faire progresser un ouvrier lorsqu'il n'y en a qu'un. Il y a deux façons de voir qu'un ouvrier a un peu plus de poids que dans le premier exemple.

La première est qu'un duo d'ouvrier a une valeur marginale plus grande que dans le premier cas (un duo ouvrier-manager ayant toujours la même). En effet, la valeur marginale du duo était de 4 dans le premier exemple, elle est de 5 ici. Une approche pourrait donc être de ne pas regarder la valeur marginale d'un joueur mais celle de tous les sous-groupes le contenant afin de déterminer sa valeur $\phi(i)$.

Une autre manière de faire apparaitre la vraie importance d'un ouvrier est de regarder la perte que suscite son départ dans plusieurs situations et non uniquement à partir de la situation finale. En effet, le pouvoir de négociation d'un employé intègre les situations dans lesquelles d'autres joueurs sont déjà partis. Le graphique ci-dessous montre qu'à partir de la situation observée MABC (A,B et C étant le nom des trois ouvriers et M désigne le manager) pour observer le pouvoir de négociation de A, il faut étudier les configurations avec tous les départs successifs possibles. Le nombre entre crochet désigne la valeur marginale de A sur la configuration.

Dans l'exemple 3, la valeur marginale d'un ouvrier est toujours de 1 lorsqu'il n'y pas de manager ; mais si le manager est présent, la valeur marginale de l'ouvrier est de 1 lorsqu'il n'y a pas d'autre ouvrier, de 3 s'il y a un autre ouvrier, et de 2 s'il y a deux autres ouvriers. En regardant dans le détail de la valeur marginale de l'ouvrier sur différentes configurations, comme on vient de le faire, on voit que l'importance de l'ouvrier n'est pas identique à celle qu'il avait dans le premier exemple où, en présence du manager, sa valeur marginale était toujours de 2. En regardant ces exemples avec l'angle

Figure 1 – Arbre des séquences d'élimination pour calculer la valeur marginale d'un employé A



des pouvoirs de décisions, la menace que représente le départ de l'ouvrier inclut la perte sèche parce qu'il ne travaille plus mais aussi la perte qu'il pourrait faire subir suite au départ d'un autre ouvrier.

Ces deux approches se rejoignent. L'idée de Shapley est donc de regarder la valeur marginale sur toutes les coalitions de joueurs que l'on peut former à partir des n joueurs et de déterminer la valeur d'un joueur à partir de chacune de ces coalitions. On peut également la voir comme étant une valeur marginale calculée sur toute les séquences d'élimination de joueurs.

On peut profiter de l'exemple pour anticiper sur la formule mathématique de la valeur de Shapley. En effet, on peut observer que la valeur intrinsèque de A qui n'est autre que sa valeur marginale à partir de l'ensemble vide (quand A est seul sur le graphique) apparait 6 fois, les valeurs à partir d'un singleton (qu'il s'agisse de M, de B, ou de C) apparaissent deux fois. Les autres configurations n'ont l'air de n'apparaitre qu'une fois mais on verra plus tard qu'il faut multiplier ces configurations car elles réunissent plus de joueurs.

Pour conclure cet exemple du manager qui fait toucher la complexité du problème et entrevoir une solution, soulignons que nous sommes restés dans un cas simple où les ouvriers sont tous les mêmes. On notera que la complexité augmente fortement quand la valeur marginale d'un individu sur un groupe dépend précisément de la composition de ce groupe.

4.2 Principe

A la suite de la partie précédente, la valeur de Shapley se comprend aisément et sa philosophie est assez simple. Il s'agit, pour déterminer la valeur d'un joueur, de faire la moyenne de toute les interactions qu'il peut avoir avec des groupes formés par d'autres joueurs. Par interaction, on entend ce qu'il apporte à un groupe de joueurs, la différence entre la valeur obtenue par le groupe en sa présence et celle obtenue en son absence. Cela correspond à une sorte d'équilibre des menaces en théorie des jeux.

4.3 Formule

La formule donnant la valeur de Shapley est une somme sur tous les sous-ensembles de E qui contiennent i. On note s le nombre d'éléments d'un sous-groupe contenant i, s varie donc de 1 à n. La valeur de Shapley du joueur i est donnée par la formule suivante :

$$Sh(i) = \frac{1}{n!} \sum_{\mathrm{F} \ni i} (s-1)!(n-s)! \Big[v(\mathrm{F}) - v(\mathrm{F} \setminus \{i\}) \Big] \tag{1}$$

Justification Pour répartir le résultat collectif $v(\mathrm{E})$ entre le n joueurs, il existe une solution relativement simple. On peut attribuer au premier joueur $v(1) - v(0)$, au deuxième joueur $v(1,2) - v(1)$, au troisième joueur $v(1,2,3) - v(1,2)$ et ainsi de suite. Chacun est donc récompensé en fonction de sa valeur marginale calculée sur les joueurs placés avant lui. Bien sûr, il peut paraitre injuste que la valeur de chaque joueur dépende d'un ordre arbitraire, aussi, on prend une valeur moyenne sur tous les ordres possibles.

La formule mathématique de la valeur de Shapley d'unjoueur i s'obtient élégamment en suivant la procédure ci-dessus. Comme la conclusion des trois exemples ci-dessus le laissait présager, le point de départ n'est pas le fait que tous les sous-groupes sont équivalents mais que toutes les séquences d'éliminations de joueur à partir de la situation observée avec les n joueurs sont équivalentes.

- On étudie les $n!$ permutations R possibles des listes ordonnées constituées par un ordre sur les n joueurs **sans préférence** pour l'une ou l'autre des permutations.
- Pour chacune de ces permutations, on considère la liste S_i^R des éléments classés avant i en la finissant par i. C'est la liste d'une coalition potentielle de taille $s = |S_i^R|$ formée à partir des n joueurs.
- On calcule alors l'impact marginal $v(S_i^R) - v(S_i^R \setminus \{i\})$ de i sur cette liste.
- La valeur de Shapley de i est finalement la moyenne des $n!$ valeurs obtenues après avoir parcouru toutes les permutations.

On obtient bien par cette méthode, un calcul sur l'ensemble des permutations des n joueurs qui ne privilégie aucun ordre par rapport à une autre. La valeur de Shapley de i s'écrit finalement :

$$Sh(i) = \frac{1}{n!} \sum_{R} v(S_i^R) - v(S_i^R \setminus \{i\}) \tag{2}$$

Notons tout de suite qu'il y a beaucoup de redondance dans cette somme qui porte sur $n!$ termes alors qu'il n'y a pas $n!$ différentes listes S_i^R. Dans la procédure décrite ci-dessus, si les s premiers termes sont les mêmes avec i en s-ième position, on obtient toujours le même S_i^R de cardinal s quel quel que soit l'ordre des $n-s$ autres termes placés derrière i. Comme il y a $(n-s)!$ permutations de ces $n-s$ joueurs, on sait que S_i^R apparait $(n-s)!$ fois. Il en résulte que la formule précédente peut s'écrire sous la forme d'une somme sur l'ensemble des sous-liste de E ayant i en dernière position.

$$Sh(i) = \frac{1}{n!} \sum_{S_i^R} (n-s)! \Big[v(S_i^R) - v(S_i^R \setminus \{i\}) \Big]$$

La manière d'aboutir à la formule 1 ci-dessus explique pourquoi on dit que la valeur de Shapley du joueur i est : **la moyenne de sa valeur marginale calculé sur les joueurs placés avant lui**.

Formule simplifiée dans lorsque l'ordre des joueurs ne compte pas
Bien souvent, l'influence de C sur le groupe (A,B) est la même que sur le groupe (B,A), on peut dire que l'influence d'un joueur ne dépend pas de l'ordre des joueurs qui le précèdent. Si on considère que pour toutes les permutations possibles de S_i^R, l'influence de i est la même alors on peut

simplifier encore la formule précédente. En effet, un groupe F contenant i apparait autant de fois qu'on peut former de listes différentes avec ses éléments, une fois i fixé en dernière position. Il y a $(s-1)!$ telles listes S_i^R qui, dans notre cas, sont équivalentes, on peut donc écrire :

$$Sh(i) = \frac{1}{n!} \sum_{\mathrm{F} \ni i} (s-1)!(n-s)! \Big[v(\mathrm{F}) - v(\mathrm{F} \setminus \{i\}) \Big]$$

La somme ne contenant alors plus que sur 2^{n-1} ensembles tandis que la somme de la formule 2 portait sur $n!$ listes.

Bien évidemment si on note k le nombre de joueurs autre que i dans la coalition F, on a $k = s-1$ et on trouve pour la valeur de Shapley la forme suivante :

$$Sh(i) = \frac{1}{n!} \sum_{\mathrm{F} \ni i} k!(n-k-1)! \Big[v(\mathrm{F}) - v(\mathrm{F} \setminus \{i\}) \Big]$$

4.4 Exemples

Cas à deux joueurs Si on a deux joueurs, A et B, lors l'application de la formule de la valeur de Shapley non simplifiée donne :

$$\phi(A) = \frac{[v(B,A) - v(B)] + [v(A) - v(\emptyset)]}{2}$$

La seconde calculée lorsque l'ordre n'a pas d'importance s'écrit :

$$\phi(A) = 1! \times 0! \times \frac{[v(AB) - v(B)] + 0! \times 1! \times [v(A) - v(\emptyset)]}{2}$$

Cas à trois joueurs Si on a trois joueurs, A,B et C, alors la formule de la valeur de Shapley est :

$$\begin{aligned} \phi(A) &= \frac{1}{6}\Big([v(B,C,A) - v(B,C)] + [v(C,A) - v(C)] + [v(A) - v(\emptyset)] \\ &\quad + [v(C,B,A) - v(C,B)] + [v(B,A) - v(B)] + [v(A) - v(\emptyset)] \Big) \end{aligned}$$

En utilisant la version simplifiée, on a la formule suivante :

$$\begin{aligned} \phi(A) &= \frac{1}{6}\Big(2! \times 0! \times [v(ABC) - v(BC)] + 1! \times 1! \times [v(AB) - v(B)] \\ &\quad + 1! \times 1! \times [v(AC) - v(C)] + 0! \times 2! \times [v(A) - v(\emptyset)] \Big) \end{aligned}$$

Exemples du manager et des trois ouvriers On peut appliquer la valeur de Shapley aux ouvriers et au manager dans les exemples décrits en 4.1. Dans ces exemples, il y a quatre joueurs. La valeur du manager se calcule rapidement en relevant que les ouvriers jouent le même rôle. Sur un groupe de 3 ouvriers, il apporte marginalement 3, deux ouvriers 2, 1 ouvrier 1, zéro ouvrier zéro. Dans l'exemple 1, en partant du groupe le plus nombreux à gauche, on a[2] :

$$\begin{aligned} Sh(\text{manager}) &= \frac{1}{4!}[(3!0! \times [3]) + 3 \times (2!1! \times [2]) + 3 \times (1!2! \times [1]) + (0!3! \times [0])] \\ Sh(\text{manager}) &= 1,5 \end{aligned}$$

Le même calcul pour les exemples 2 et 3 montrent que la valeur du manager est 0,75 dans l'exemple 2 et 1,25 dans l'exemple 3.

L'application de la valeur de Shapley pour les ouvriers est plus compliquée, en effet, comme on l'a vu au 4.1 , la valeur marginale d'un ouvrier est différente selon que le manager est présent dans le sous-groupe ou non. En regardant d'abord les ensembles à trois éléments puis ceux à deux, puis ceux à un et en mettant toujours d'abord le résultat avec l'élimination du manager, on obtient le calcul suivant :

$$\begin{aligned} Sh(\text{ouvrier}) &= \frac{1}{4!}\Big[3!0! \times [3] + 1 \times (2!1! \times [2]) + 2 \times (2!1! \times [2]) \\ &\quad +1 \times (1!2! \times [1]) + 2 \times (1!2! \times [1]) + (0!3! \times [1])\Big] \\ &= \frac{1}{24}\Big[6 \times [2] + \underbrace{2 \times ([1] + 2 \times [2])}_{\text{2 ouvriers ou manager + 1 ouvrier}} + \underbrace{2 \times ([2] + 2 \times [1])}_{\text{manager ou 1 ouvrier}} + 6 \times [1]\Big] \\ Sh(\text{ouvrier}) &= 1,5 \end{aligned}$$

En étant vigilant sur les valeurs avec ou sans manager de la formule ci-dessus en fonction du nombre d'ouvriers, on trouve que la valeur d'un ouvrier de l'exemple 2 est 1,25 par le calcul suivant :

$$\begin{aligned} Sh(\text{ouvrier}) &= \frac{1}{24}\Big[6 \times [4] + 2 \times ([1] + 2 \times [1]) + 2 \times ([1] + 2 \times [1]) + 6 \times [1]\Big] \\ Sh(\text{ouvrier}) &= 1,75 \end{aligned}$$

Pour l'exemple 3, le bond de productivité marginale se fait lorsqu'un ouvrier s'ajoute à un manager et un ouvrier :

[2]Les deux premiers termes sont les factoriels de la formule 1 , le terme entre crochet suivant le signe $\times$ est la valeur marginale du joueur sur l'ensemble en question, lorsqu'il y a un terme multiplicatif devant cette série, cela correspond au fait qu'on a une répétition de ce terme suite au fait que les ouvriers ont un rôle symétrique

$$\begin{aligned} Sh(\text{ouvrier}) &= \frac{1}{24}\Big[6 \times [2] + 2 \times ([1] + 2 \times [3]) + 2 \times ([1] + 2 \times [1]) + 6 \times [1]\Big] \\ Sh(\text{ouvrier}) &= 1,583 \end{aligned}$$

On vérifie que la somme des valeurs des trois ouvriers et du manager vaut à chaque fois 6.

4.5 Propriétés et unicité

Les contributions mesurées par la valeur de Shapley respectent des propriétés satisfaisantes :

- $\sum_{i=1}^{n} Sh(i) = v(\mathrm{E})$. La décomposition est cohérente (ou efficace), c'est à dire que le résultat du groupe est réparti intégralement entre les joueurs.

- Si $\forall F \not\ni i, j$; $v(\mathrm{F} \cup \{i\}) = v(\mathrm{F} \cup \{j\})$ alors $Sh(i) = Sh(j)$. Deux joueurs ayant le même impact marginal sur tous les sous-ensembles ont la même valeur. C'est l'occasion de souligner qu'il suffit, pour deux joueurs, de ne pas avoir la même influence sur l'un des sous-groupes pour ne pas avoir exactement la même valeur.

- Si $\forall \mathrm{F} \not\ni i; v(\mathrm{F} \cup \{i\}) = v(\mathrm{F})$ alors $Sh(i) = 0$. C'est la propriété de valeur nulle pour les joueurs transparents que nous avions d'emblée considéré désirable.

- $Sh_{v+w}(i) = Sh_v(i) + Sh_w(i)$. où Sh_v désigne la valeur de Shapley associée à la valeur v et où w est un autre score sur les mêmes joueurs. C'est la propriété d'additivité. Sauf pour les experts, on ne pense pas immédiatement à cette propriété étudiant le cas de deux scores sur le même joueur, pourtant elle aurait mérité sa place dans les propriétés désirables d'une valeur ϕ. Si on sépare un jeu en deux jeux $v_{\frac{1}{2}}$ identiques tels que pour tous sous-ensemble : $v_{\frac{1}{2}}(\mathrm{F}) = \frac{v(\mathrm{F})}{2}$; il est rassurant d'avoir $\phi_v(i) = 2\phi_{v_{\frac{1}{2}}}(i)$ [3]. Mieux encore, cette propriété traduit le fait que si v et w sont deux étapes d'un même processus (de production par exemple) alors la valeur globale est la somme des valeurs calculées sur chacun des sous-processus.

La valeur de marginale, comme on l'a vu, ne respecte pas la première propriété. La valeur marginale normalisée la respecte mais ne respecte pas la propriété d'additivité. On peut illustrer ce point à l'aide d'un exemple de

Tableau 1 – Exemple de production de deux biens par deux joueurs

	A	B	A+B
J1 et J2	2	2	4
J1	1	0	1
J2	1	1	2

production de deux objets A et B par deux joueurs, J1 et J2. Le tableau de production est donné ci-dessous :

Les valeurs marginales et marginales normalisées selon le bien étudié sont données dans le tableau suivant :

Tableau 2 – Valeur marginale, normalisé ou non, sur les produits A et B et sur la quantité total A+B

	VM_A	VMN_A	VM_B	VMN_B	VM_{A+B}	VMN_{A+B}	$\mathrm{VMN}_A + \mathrm{VMN}_B$
J1	1	1	1	$\frac{2}{3}$	2	$\frac{8}{5}$	$\frac{5}{3}$
J2	1	1	2	$\frac{4}{3}$	3	$\frac{12}{5}$	$\frac{7}{3}$

La valeur marginale normalisée du joueur 1 est de $\frac{8}{5}$ sur l'ensemble de la production alors que lorsque l'on fait le détail sur les deux produits sa valeur marginale normalisée semble être de $\frac{5}{3}$. La valeur de Shapley du joueur 1 est $\frac{3}{2}$, c'est la somme de la valeur de Shapley calculée pour le produit A, qui est de 1, et de la valeur de Shapley calculée sur le produit B : 0,5. La valeur de Shapley respecte donc la propriété d'additivité sur cet exemple.

En fait, **Shapley (1953) démontre que la valeur qu'il introduit est la seule à respecter les quatre propriétés énoncées ci-dessus.**

4.6 Valeur de Shapley et justice

Dans la littérature, la valeur de Shapley est souvent associée aux notions de justice et d'équité. La répartition égalitaire qui répartit uniformément le résultat collectif entre les joueurs $\frac{v(\mathrm{E})}{n}$ est vue comme injuste car elle ne fait pas dépendre la valeur d'un joueur de son influence dans l'obtention du score collectif. La valeur intrinsèque et la valeur marginale sont contestées non pas parce qu'elles ne répartissent que rarement l'ensemble du score global mais parce qu'elles ne tiennent pas compte des interactions ou alors en privilégiant une situation particulière. Les exemples d'un manager avec des ouvriers montrent que l'on ne peut pas privilégier une seule situation. Finalement, la valeur de Shapley peut-être caractérisée par sa justice car elle tient compte des toutes les interactions possibles.

Plus fondamentalement, la valeur de Shapley considère que la constitution précise du groupe n'est pas le choix du joueur et ne doit donc pas avoir

[3] De même, cette propriété implique que $\phi_{-v}(i) = -\phi_v(i)$

une influence sur sa valeur. Dit autrement, le groupe effectivement observé ne doit pas avoir plus d'importance que tous les autres groupes que l'on pourrait imaginer. Cette notion de justice se retrouve dans le calcul en ne pondérant pas les permutations des n joueurs. On reviendra plus loin sur cette notion. En particulier, il est parfois reproché au calcul de la valeur de Shapley, le fait qu'elle évalue et de prend en compte des impacts marginaux dans des configurations que l'on juge, parfois, sans pertinence économique. c'est la notion de justice sans *a priori* qui justifie cette démarche de la valeur de Shapley.

5 Applications et exemples

La valeur de Shapley a de multiples applications. Dans ce document, elle a été introduite comme réponse à une question de répartition d'un score. Plus généralement, elle individualise un résultat collectif. Les exemples ci-dessous font référence à des thématiques variées. Ils abordent aussi des questions techniques de la valeur de Shapley. Ils montrent l'étendue des applications de la valeur de Shapley.

5.1 « -C'est mon RSA. -Non, c'est le mien »

Le RSA est un dispositif d'aides aux travailleurs pauvres et aux personnes sans ressources introduit en juin 2009 en France métropolitaine. Il est attribué à une famille au sens défini par la législation spécifique de ce dispositif. On se place dans le cas simple d'un RSA touché par un couple. Le calcul du RSA intègre le fait qu'il s'agit d'un couple et le montant perçu n'est pas la somme de deux RSA individuels.

On peut vouloir déterminer un revenu d'activité individuel ou simplement savoir quel est le rôle de chaque membre du couple dans l'obtention ou non d'un RSA. Comment alors répartir le montant de RSA calculé au niveau du couple entre les deux membres de ce couple ? On souhaite bien évidemment que le RSA du couple soit réparti intégralement entre les deux membres du couple (efficacité ou cohérence de la répartition). Si on souhaite déterminer un RSA par personne, attribuer la moitié du RSA à chaque membre n'est qu'un pis-aller car les deux conjoints n'ont pas forcément un rôle symétrique. La valeur de Shapley offre une réponse[4]. Il s'agit de prendre la moyenne des RSA marginaux sur toutes les configurations : la situation de célibat et la situation de couple. Autrement dit, le RSA individuel calculé par la valeur de Shapley est la demi-somme du RSA qui aurait été obtenu en étant célibataire et de la différence de RSA qui est induite sur le RSA du conjoint. De cette manière on répartit le montant de RSA perçu de manière

[4]Il y a deux joueurs $n = 2$ et la valeur d'une coalition est le montant de RSA qui lui est attribué en fonction de son revenu

entre les deux membres du couple sans supposer qu'ils ont a priori un rôle symétrique dans ce montant de RSA, en supposant par contre qu'ils sont symétriques dans le couple au sens où il ne s'agit pas d'une personne rejointe par une autre mais bien de deux personnes qui vivent conjointement. Enfin, cette méthode ne suppose pas que l'on a toujours une contribution positive ou nulle au RSA du foyer, elle ne recule pas devant le fait que certains conjoint font perdre du RSA et qu'au niveau individuel, la dépende de RSA qu'il génère est négative.

Pour un couple sans revenu, le RSA est de 700 euros et, par symétrie, chacun perçoit la moitié soit 350 euros. Le tableau ci-dessous donne le RSA individuel mesuré par la valeur de Shapley dans le cas de 3 couples qui ne sont pas dans cette situation et où les deux revenus d'activité sont différents :

	revenu A	revenu B	RSA de A seul	RSA de B seul	RSA du couple	Sh(A)	Sh(B)
Couple 1	600	400	239	315	320	122	198
Couple 2	1350	0	0	467	187	-140	327
Couple 3	1121	721	41	193	0	-76	76

Tableau 3 – Note : Chiffre du RSA selon le barème 2011. Couple sans enfant, sans allocation logement, etc. Le RSA du couple est calculé par la formule RSA $= \max(700 - 0,38(R_A + R_B); 0)$ où R_A et R_B désignent les revenus d'activité de l'un et l'autre des conjoints, le RSA d'un célibataire est donné par la formule : RSA $= \max(467 - 0,38R_A; 0)$. La valeur de Shapley est la moité de la colonne RSA célibataire et de la différence entre RSA du couple et RSA du conjoint seul

Le cas du couple 3 montre que même lorsque le montant du RSA du couple est nul, il peut y avoir un RSA individuel, ici, l'un des membres aurait droit à un RSA de 193 euros mais les revenus de son conjoint l'excluent du dispositif, on peut donc considérer que ce dernier fait perdre du RSA au couple et que son RSA individuel est négatif.

Notons que lorsqu'un enfant d'un couple travaille, ses revenus peuvent aussi modifier le montant de RSA perçu. De plus, le RSA doit alors être réparti entre trois joueurs. Il devient alors compliqué d'attribuer un tiers du RSA à chacun et la valeur de Shapley s'impose encore plus naturellement dans ce cas plus complexe.

Ce procédé peut bien entendu s'appliquer à d'autres transferts qu'il peut être intéressant de calculer individuellement : l'impôt sur le revenu, les allocations logements, la prime pour l'emploi, etc. Avec cette individualisation impartiale des ressources, on peut déterminer un revenu individuel et ensuite déterminer une incitation financière à cohabiter ou à s'unir légalement. Cette répartition peut permettre d'analyser le passage d'une prestation « familialisée » à une prestation individualisée. Elle peut aussi déterminer un niveau de vie individuel dans un ménage dans lequel on ne suppose pas qu'il y a solidarité parfaite (Roy 2006).

5.2 Les exportations de la zone euro

L'union européenne ou la zone euro fournissent des cas, désormais fréquent dans lesquels on s'interroge sur les résultats d'un groupe et où l'on peut vouloir répartir ce résultat entre chaque membre du groupe. On peut prendre l'exemple des exportations de la zone euro et vouloir quantifier le rôle de chaque pays dans ces exportations hors zone euro[5].

La réponse classique consiste à regarder ce que chaque pays exporte en dehors de la zone. Cette approche est logique et intuitive, de plus, la somme des exportations des pays vaut bien l'exportation générale de la zone ce qui est appréciable. Elle fournit une réponse à la question : « Quelles sont les exportations de chaque partenaire à l'exportation l'extérieur de la zone ? ». L'usage de la valeur de Shapley ne s'impose pas dans ce cas. Il est toutefois rassurant de noter que l'on trouve les mêmes contribution pour chaque pays, en utilisant la valeur de Shapley en associant à un sous-groupe, la somme des exportations des pays de ce sous-groupe[6].

L'évaluation précédente, par certains points, peut paraitre simpliste. En effet, prenons le cas d'un avion exporté par la France en dehors de la zone euro. Pourquoi inclure entièrement cette exportation dans le bilan de la France si l'empennage est produit en Espagne et le fuselage en Allemagne[7] ? Avec la méthode précédente, parce que l'assemblage se fait en France, ses exportations en dehors de la zone sont surestimées et celles de l'Allemagne et de l'Espagne, sans qui la France ne pourrait exporter un tel avion, sont sous-estimées.

Il semble alors logique de prendre en compte les exportations intra-zone pour quantifier le vrai rôle d'un pays aux exportations de la zone. Puisqu'une partie des exportations peut-être due à la présence d'autres pays dans la zone. On commence à se demander ce que serait les exportations si on ôtait (au moins fictivement) le pays de la zone pour prendre en compte les interactions au sein de la zone.

---

[5] Fin 2011, les exportations peuvent être remplacées par les dettes de la zone euros qui font plus l'actualité

[6] En effet, lorsque le score d'un sous-ensemble est la somme des exportations hors-zone de chaque pays du sous-ensemble lorsque l'on fait l'hypothèse que ces exportations ne seraient pas modifiées par le départ d'un pays de la zone ; la valeur marginale d'un joueur est alors la même sur tous les sous-ensembles.

[7] Cet exemple est bien sûr tiré des avions Airbus

Avec la valeur de Shapley, en calculant le niveau d'exportation de chaque sous-ensemble de pays de la zone et en regardant l'influence marginale d'un pays particulier sur ces sous-ensembles, on détermine une valeur ayant de bonnes propriétés. Le cas de production commune et d'exportation individuelle, comme dans l'exemple de l'avion, est géré par cette méthode (c'est une de ses grandes forces). Et on a toujours, la répartition totale des exportations de la zone entre les pays membres. Cette méthode permet aussi, si on le souhaite, de mettre sur un pied d'égalité les pays aux frontières de la zone qui vont naturellement exporter plus en dehors de la zone et les pays dont les voisins sont eux-aussi dans la zone. En effet, en modifiant virtuellement la zone, les pays normalement intérieurs vont dans certaines configurations se trouver à la frontière avec le hors-zone. Pour résumer l'apport de la valeur de Shapley dans ce cas précis : l'approche intuitive regarde l'impact d'un pays en supposant qu'en son absence les exportations des autres pays seraient strictement identique, la valeur de Shapley permet de considérer les échanges à l'intérieur de la zone comme devenant parfois des échanges à l'extérieur et même d'inclure des modèles plus compliqués. Par exemple, on peut imaginer des échanges commerciaux plus compliqués prenant en compte le fait que l'appartenance d'un pays à la zone change l'intensité des liens commerciaux entre ce pays et les membres de la zone. Si on a un modèle qui prévoit ces changements d'intensité (inspiré d'un modèle de gravité par exemple), la valeur de Shapley l'intègre dans son calcul et devient presque incontournable. C'est à l'économiste, en fonction de ce qu'il désire montrer, de choisir le score qui convient pour capter les interactions qu'il désire étudier. La valeur de Shapley a donc une certaine souplesse qui lui permet d'être adaptée au problème précis en question.

On peut reprocher à cette procédure de considérer des cas que l'on juge impossible ou sans sens économique. Par exemple, on n'imagine pas la Belgique et la Grèce construire une zone commune à eux-seuls. On verra bien plus loin que l'on peut contourner cette difficulté et adapter la valeur de Shapley pour ne regarder que les configurations que l'on juge acceptable. Mais, avant de vouloir contourner ce point, il faut avoir conscience que ce caractère fictif de certaines configurations est le prix à payer pour prendre pleinement en compte les interactions des pays-membre sans *a priori*.

5.3 Vote

Dans une assemblée, il peut être intéressant de savoir quel est le poids de chaque électeur dans un vote. Deux cas sont triviaux, celui où toutes les voix ont le même poids et celui où la décision dépend d'un seul individu (cas du dictateur ou de l'électeur médian lorsque l'on peut ordonner les électeurs). Cependant, on peut penser à des vote plus complexes, par exemple aux votes à l'Union Européenne dans lesquels chaque État s'exprime par une seule voix mais où celle-ci est pondérée selon la taille de sa population. La question est alors de savoir comment quantifier l'importance du vote d'un électeur donné.

Pour une position donnée concernant la réponse au vote, on regarde l'ensemble des personnes qui votent comme le joueur i. On dit qu'elles remportent le vote si leur position est retenue par l'assemblée et le score est de 1. Dans le cas contraire, si une autre position est retenue à l'issue du vote, la valeur de la coalition est nulle. Intuitivement, un membre a de l'importance lorsque le résultat du vote dépend essentiellement de son choix. C'est exactement ce que va comptabiliser le calcul de la valeur de Shapley. En effet, les seuls cas où la valeur marginale du joueur i n'est pas nulle sont ceux où sa coalition gagne mais perdrait sans lui. On dit alors que i fait basculer le vote. En faisant la moyenne des valeurs marginales du joueur i, la valeur de Shapley respecte bien l'intuition et permet de donner à chaque votant une importance dans le vote. Dans ce cas précis, la valeur de Shapley prend le nom d'indice de Shapley-Shubik la suite à leur article de référence (Shapley et Shubik 1954).

Exemple On considère une assemblée votant à la majorité des individus. Le vote est une question fermée sur laquelle il n'y a que deux alternatives : pour et contre. Il n'y a pas d'abstention, tous les électeurs s'expriment. Les électeurs sont regroupés en partis politiques et tous les membres d'un même parti défendent toujours la même opinion. Ainsi, on peut considérer que les partis sont les votants chacun ayant un poid liés au nombre d'électeurs qui les constituent. Finalement, on veut connaitre le poids de ces partis.

Imaginons que nous soyons dans le cas d'une assemblée similaire à l'Assemblée Nationale en France en 1973[8] avec 6 partis ayant les effectifs suivants[9] :

Nom	A	B	C	D	E	F
Effectif	30	30	51	73	100	162

Il y a 446 députés dans cette assemblée. Pour mesurer l'importance d'un groupe, on regarde la proportion de cas dans lequel il est décisif. En procédant de cette manière, on attribue une valeur à chaque groupe en étant complètement neutre, c'est-à-dire en ne privilégiant pas certaines associations par rapport à d'autres. On obtient le résultat suivant :

Dans 43,3 % des cas, le choix du groupe F fait basculer le vote. Remarquons que même si le groupe E a plus de membres que le groupe D, la configuration de l'assemblée fait que ces deux groupes ont exactement le même poids. En effet, lorsque l'un est décisif sur une configuration, l'autre

[8]Les cas plus récents sont moins intéressants car l'un des partis est en majorité absolu et détient tout le pouvoir. Le calcul est alors trivial

[9] Les députés apparentés et non rattachés à un parti ne sont pas pris en compte dans cette version simplifiée

Nom	A	B	C	D	E	F
Effectif	30	30	51	73	100	162
Effectif (en %)	6,7	6,7	11,4	16,4	22,4	36,3
Valeur (en %)	3,3	3,3	10	20	20	43,3

l'est aussi : si F est dans un groupe, les deux groupes, D et E, peuvent en le rejoignant dépasser la majorité de 224 électeurs (ce n'est pas le cas du groupe C par exemple). S'ils ne rejoignent pas le groupe F, ils peuvent chacun acquérir une majorité uniquement si aucun autre groupe n'a rejoint F. Au final, les deux groupes D et E jouent bien le même rôle dans toutes les situations et il est logique qu'ils aient la même valeur. On vérifie bien la deuxième propriété de la valeur de Shapley (cf. 4.5). Par curiosité, on peut noter qu'en modifiant les effectifs des deux plus petits partis pour que la somme de leurs effectifs soit inférieure à 40, alors les importances attribuées selon la procédure de Shapley sont modifiées. Les deux plus petits partis ont une valeur nulle (ils ne sont jamais décisifs), le plus grand à la moitié du pouvoir de vote et les trois autres ont chacun un sixième du pouvoir.

5.4 Valeur d'un dirigeant et partage de la valeur ajoutée

On se demande régulièrement ce qui justifie les salaires des dirigeants d'entreprise. On pourrait penser appliquer la valeur de Shapley mais là encore, il est difficile de savoir quels seraient les bénéfices sur toutes les configurations possibles de l'entreprise. La valeur marginale, qui requiert de savoir quel serait le résultat de l'entreprise sans le dirigeant, est déjà elle-même compliquée. On peut alors penser à appliquer une méthode contrefactuelle et évaluer la participation d'un dirigeant à la valeur ajoutée en comparaison d'un autre dirigeant mais là-aussi on se heurte souvent à un problème d'observation.

C'est finalement à des niveaux agrégés que la valeur de Shapley peut donner des indications sur la répartition de la valeur ajoutée entre par exemple l'ensemble des actionnaires, l'ensemble des employés et l'ensemble des clients. Sur cette thématique, on peut lire une application de la valeur de Shapley qui décompose le taux de croissance des taux de marge en fonction des contributions du coût salarial de l'unité d'excédent brut d'exploitation, de l'inflation et du taux de croissance du coût salarial (Mussard et Philippe 2007). L'analyse produite dans plusieurs pays donnent des éléments d'explication intéressants sur l'évolution du chômage entre 1961 et 2000.

5.5 Sélection de modèle économétrique

On peut imaginer des applications de la valeur de Shapley dans des problèmes de sélection de modèle en économétrie. En effet, quand on veut déterminer l'apport de l'ajout d'une variable explicative à un modèle, on raisonne souvent en regardant une valeur marginale : la variation du R^2. Pourtant, on a bien montré dans ce document que cette démarche peut être remise en cause lorsque les variables sont corrélées et que la modification du R^2 s'explique de fait par la présence de plusieurs variables.

On peut accepter de ne pas être neutre quand il s'agit de variables et de considérer que certaines sont prioritaires. Mais dans l'arbitrage sur la sélection des variables d'un modèle, on peut vouloir déterminer indépendamment de tout ordre d'élimination la contribution au R^2 de chaque variable et utiliser la valeur de Shapley. Cette démarche est appliquée par Osnat (2007) qui défend l'intérêt de cette méthode en particulier en présence de variables indicatrices, de termes d'interactions ou de dépendance non linéaires.

5.6 Analyse d'une évolution

On cherche souvent en économie à expliquer des variations, souvent des variations temporelles.

Illustrons ceci par l'évolution d'une grandeur P s'expliquant par deux facteurs μ et L ; il s'agit des notations de Shorrocks (1999) pour qui P est le taux de pauvreté, μ le revenu moyen et L la courbe de Lorenz des revenus. Entre deux périodes 1 et 2, on écrit souvent :

$$P(\mu_2, L_2) - P(\mu_1, L_1) = [P(\mu_2, L_2) - P(\mu_1, L_2)] + [P(\mu_1, L_2) - P(\mu_1, L_1)]$$

On considère ensuite que le premier terme caractérise l'influence de l'évolution de μ sur P tandis que le second mesure l'influence de l'évolution de L sur P. L'écriture ci-dessus n'est pas symétrique et l'on procède en fait en faisant comme si l'évolution de μ se produisait après celle de L et en lui affectant sa valeur marginale. De quoi outrer les adeptes de la valeur de Shapley.

En effet, on peut considérer que l'on doit expliquer un score global : $\Delta P = P(\mu_2, L_2) - P(\mu_1, L_1)$ et répartir ce score en deux joueurs $\Delta\mu = \mu_2 - \mu_1$ et $\Delta L = L_2 - L_1$. La méthode asymétrique ci-dessus est bancale car la contribution ϕ de $\Delta\mu$ à l'évolution de P est donnée par :

$$\phi(\Delta\mu) = F(\Delta\mu, \Delta L) - F(0, \Delta L)$$

où $F(a, b) = P(\mu_1 + a, L_1 + b) - P(\mu_1, L_1)$, alors que la valeur de Shapley préconise la formule symétrique :

$$\phi(\Delta\mu) = \frac{1}{2}\Big\{[F(\Delta\mu, \Delta L) - F(0, \Delta L)] + [F(\Delta\mu, 0) - F(0, 0)]\Big\}$$

Il s'agit finalement de la moyenne de l'influence de μ avant et après l'influence de l'évolution de L.

Dans le même article, Shorrocks (1999) détaille la décomposition d'un indice par sous-population y compris quand il y a plusieurs critères de distinction (il prend l'exemple d'une décomposition par âge et par région). Plus intéressant encore, il poursuit l'analyse de l'évolution d'une grandeur en fonction des évolutions dans plusieurs sous-population. La population générale est réparties en K sous-groupes et on peut décomposer une valeur calculée sur la population P en $P_1, \cdots, P_K$ calculées sur chacune des sous-population. Par la méthode de Shapley, par exemple, on peut construire une expression de type :

$$P = \sum_{k=1}^{K} \nu_k P_k \ \big(= \sum_{k=1}^{K} Sh(k) \big)$$

où P_k est l'équivalent de la grandeur P calculée au niveau de la sous-population k et ν_k est la contribution de cette grandeur à la valeur globale P. On notera à cette occasion que la méthode de Shapley permet d'éliminer les termes croisés entre les sous-populations ce que l'on cherche parfois.

L'évolution de P peut s'expliquer par des variations de la population k de deux façons, ou bien la grandeur P_k évolue ou bien c'est la manière dont la sous-population k influence la population totale qui change. C'est par exemple le cas lorsque la taille de la sous-population grandit (on peut penser aux retraités ou aux parents isolés). On peut écrire :

$$\Delta P = \sum_{k=1}^{K} [\nu_{k2} P_{k2} - \nu_{k1} P_{k1}]$$

L'étape suivante consiste à utiliser l'approche de Shapley en ayant comme score une fonction F bien adaptée comme dans l'exemple où $k = 1$ présenté au début de cette section et des joueurs $\Delta\nu_1, \Delta P_1, \cdots, \Delta\nu_K, \Delta P_K$. Shorrocks montre que la valeur de Shapley aboutit alors à l'expression suivante :

$$\Delta P = \sum_{k=1}^{K} \frac{\nu_{k1} + \nu_{k2}}{2} \Delta P_k + \sum_{k=1}^{K} \frac{P_{k1} + P_{k2}}{2} \Delta \nu_k$$

L'évolution globale s'explique par la somme, dans chaque sous-population, de la variation de la grandeur pondérée par le poids moyen entre la situation de départ et la situation d'arrivée et l'évolution du poids de la contribution multipliée par la moyenne de la valeur sur la période. Cette formule reste relativement simple dans un cas avec deux périodes et deux facteurs, on peut envisager des situations plus complexes dans lesquelles la valeur de Shapley s'avère très puissante et presque incontournable.

On pourra lire une comparaison de cette méthode avec d'autres et une application à l'explication des variations de salaire par trois déterminant dans (Devicienti 2010). La valeur de Shapley pourrait aussi s'appliquer à déterminer un pouvoir d'achat ou une inflation par catégorie socio-professionnelles et par région tout en laissant la possibilité de parler de l'influence d'une région ou d'une CSP. Enfin, on trouvera une illustration de la valeur de Shapley à une étude du taux de pauvreté en annexe 9.1.

5.7 Valeur d'un sportif dans une équipe

On aurait envie de quantifier l'apport d'un sportif à une équipe. Le problème est que décrire l'impact d'un joueur sur un sous-ensemble n'a bien souvent pas de sens puisque les règles même du sport en question imposent de jouer à un nombre constant n. On ne peut donc pas *a priori* procéder à des séquences d'éliminations de joueurs et appliquer la valeur de Shapley dans ce cas.

Cependant, en se distançant de l'exemple de l'entreprise, rien n'interdit d'inventer un remplaçant à chaque joueur. Pour une équipe de basketball de cinq joueurs A,B,C,D et E, par exemple, on peut définir des joueurs de référence A',B',C',D' et E'. Évaluer un score sur une coalition, par exemple, (ABD) revient alors à évaluer le résultat de l'équipe A,B,C',D,E'. La notion de valeur du joueur devient alors relative, il s'agit de sa performance par rapport à son remplaçant. Il est du ressort de l'économiste (où du supporter dans ce cas) de choisir judicieusement les remplaçants de façon à avoir une évaluation neutre. Il pourra s'agir de remplaçant fictif se définissant comme représentatif d'un niveau moyen. Lorsque cela est possible un remplaçant identique pour tous les joueurs permet de comparer directement les joueurs.

Lorsqu'on n'a pas de modèle qui prédit le résultat de chaque équipe, il faut observer toutes les configurations ce qui n'est parfois pas possible. Nous reviendrons sur ce point plus loin au 6.3. Il faut retenir de cet exemple d'individualisation de valeur collective la notion de remplaçant. On parlera plus généralement de contrefactuel.

5.8 Contribution à la réduction des inégalités

Comme cela a été dit dès l'introduction de ce document, la valeur de Shapley est la solution d'un problème délicat qui se pose lorsque l'on étudie les inégalités. Il s'agit du problème de la décomposition des inégalités par sources de revenus. L'idée est que la réduction des inégalités entre le revenu avant redistribution et le revenu après redistribution est la résultante de plusieurs transferts et on aimerait déterminer quelle est la contribution de chaque transfert à la réduction globale de l'inégalité.

La méthode la plus couramment employé est celle que l'on qualifie de méthode du pseudo-Gini. Cette méthode a plusieurs défauts. L'un d'entre-eux est que cette méthode impose de choisir un indice d'inégalité et donc une sensibilité aux inégalités dans la catégorie restreinte des indices de Gini généralisés[10]. On trouvera une application de la méthode du pseudo-Gini dans l'édition 2011 de France, Portrait Social (Cazenave, Duval, Eidelman, Langumier, et Vicard 2011).

La valeur de Shapley offre une méthode robuste ayant de nombreux avantages comparés par rapport au pseudo-Gini[11]. Elle a pour la première fois été évoquée par Auvray et Trannoy (1992). On reprend ici, les mêmes définitions et les mêmes outils que dans l'article évoqué plus haut, on sélectionne l'indice de Gini pour mesurer les inégalités. Pour détailler la méthode de Shapley, appelons R^0 le revenu avant redistribution. Le score d'un sous-ensemble F est la différence d'inégalités induite par les transferts de ce sous-ensemble. Il se calcule par la formule suivante :

$$v(\mathrm{F}) = G(R^0 + \mathrm{F}) - G(R^0)$$

Il s'en suit que la valeur marginale d'un transfert T mesurée sur un ensemble F ne contenant pas T vaut :

$$\begin{aligned}
\mathrm{VM}(T|\mathrm{F}) &= v(\mathrm{F}+T) - v(\mathrm{F}) \\
&= [G(R^0 + \mathrm{F} + T) - G(R^0)] - [G(R^0 + \mathrm{F}) - G(R^0)] \\
\mathrm{VM}(T|\mathrm{F}) &= G(R^0 + \mathrm{F} + T) - G(R^0 + \mathrm{F})
\end{aligned}$$

On remarque que le terme $G(R^0)$ se simplifie et n'intervient plus dans le calcul.

Soulignons que si on choisit une procédure d'élimination, on aurait aussi pu choisir de comparer un transfert à un « remplaçant » pour reprendre l'expression du 5.7. C'est une question que discutent Sastre et Trannoy (2002), ils proposent non pas de supprimer le revenu associé à une source mais de le remplacer par un revenu constant (la moyenne du transfert). Ceci leur permet de ne pas avoir de contribution négative pour un transfert égalitaire. On pourrait aussi penser à un revenu plus neutre en matière de redistribution (proportionnel à la distribution initial des revenus R^0) comme contrefactuel. De cette manière, on fait comme si les autres transferts ne réduisaient pas les inégalités et non comme s'ils n'étaient pas là. L'influence des montants globaux mis en jeu dans chaque transferts est alors différent.

Le tableau suivant donne les résultats et compare avec ceux de Cazenave, Duval, Eidelman, Langumier, et Vicard (2011).

Sur cet exemple, les contributions de chacun des transferts à la réduction des inégalités sont assez proches qu'elles soient mesurée par la valeur de Shapley ou par la méthode du pseudo-Gini. Ceci n'est cependant pas une règle générale.

[10] Voir (Gajdos 2001)

[11] Nous ne faisons pas ici la liste de ces avantages, ils feront l'objet d'une note interne.

Tableau 4 – Contribution à la réduction des inégalités des transferts et comparaison au pseudo-Gini

	Pseudo-Gini	Shapley associé à Gini
Cotisations redistributives (famille, logement)	4,8	7,6
Contributions sociales	3,1	4,6
Impôt sur le revenu (avant PPE)	22,4	25,1
Prime pour l'emploi (PPE) et autres crédits d'impôt	3,8	3,4
Taxe d'habitation	-0,3	0,1
Allocations familiales	10,6	10,0
Prestations familiales sans condition de ressources	5,3	3,3
Prestations sous conditions de ressources	9,9	9,3
Aides au logement	18,5	17,0
Minima sociaux	17,4	16,0
RSA "activité"	1,2	1,2
Apa	3,4	2,4
Total	100	100

5.9 Variante continue

La littérature se référant à la valeur de Shapley est conséquente. Certains articles mathématiques font des développements en plusieurs dimensions, en changeant légèrement les propriétés désirées ou en regardant les liens entre la valeur de Shapley le noyau en théorie des jeux, par exemple. On peut lire à ce sujet Winter (2002) ou Ghintran (2010), ces articles montrent combien la recherche a été développée autours de la valeur de Shapley. Un livre entièrement consacré à la valeur de Shapley (Roth et al 1988) aborde de nombreux points et souligne de nombreuses vertus à cette valeur. D'autres articles concernent l'application de la valeur de Shapley et les astuces ou variantes qu'il faut parfois utiliser.

Une extension importante de la valeur de Shapley est sa variante continue quand les joueurs ne se caractérisent pas seulement par le fait d'être présent ou non dans une coalition mais quand ils peuvent avoir un niveau d'engagement continu variant entre 0 et 1.

Un exemple Pensons par exemple à une entreprise produisant un produit a partir de plusieurs composants $\underline{x} = (x_1, x_2, \cdots, x_n)$ selon la fonction de production $y = f(x_1, x_2, \cdots, x_n) = f(\underline{x})$. A partir du $\underline{x}$ réellement utilisé pour produire y, on souhaite affecter une valeur à chaque input x_i par exemple pour répartir un bénéfice ou pour des raisons de comptabilité.

La valeur de chaque facteur x_i ne peut se déterminer uniquement en fonction de ce que serait la production avec $x_i = 0$. Nous sommes ici dans un cas continu où les inputs n'influencent pas uniquement par leur présence ou leur absence mais aussi par leur niveau. Peut-on adapter la valeur de Shapley à cette situation ?

On remarquera que la fonction de production tient le rôle de score. Elle attribue un score à chaque sous-ensemble, ce qui est primordial dans la démarche de Shapley. Dans le cas où il n'y a pas de coût fixe $f(\underline{0}) = 0$ et où la fonction de production f a de bonnes propriétés de régularités, une réponse positive est donnée par la valeur d'Aumann-Shapley suite à leur livre (Aumann et Shapley 1974).

Le prix Aumann-Shapley P_i du bien x_i associé au niveau $\underline{x}$ et à la fonction de production f se calcule par la formule :

$$P_i(f, \underline{x}) = \int_0^1 \frac{\partial f}{\partial x_i}(t\underline{x}) \, \mathrm{d}t$$

On retrouve des éléments de la valeur de Shapley discrète comme la somme sur des sous-ensemble et la dimension marginale du calcul. En effet, la valeur d'Aumann-Shapley consiste à attribuer à une entité x_i la somme (l'intégrale) de la variation marginale qui lui est associée le long d'un rayon qui va de l'origine au point qui représente le vecteur final. En discret, il s'agit bien d'une somme de valeur marginale associée à x_i. Le chemin sur lequel portent la somme est le segment reliant l'origine au point observé ; de la même manière que les sous-listes traitent de façon symétrique les joueurs présents dans le cas discret, on n'accorde ici pas plus de poids à un joueur et à une configuration donnée qu'aux autres. Boyer, Moreaux, et Truchon (2006) ajoutent que l'on peut imaginer d'autres sentiers le long desquels mesurer les impacts marginaux même en conservant une certaine symétrie du chemin, synonyme d'équité entre les joueurs. La valeur qui en découle possède alors des propriétés différentes.

Exemples d'applications

Avec un mécanisme de détermination de prix comme celui présenté rapidement ci-dessus, il est tentant d'étudier les équilibres d'une économie. On peut lire (Mirman et Tauman 1981) pour une présentation des axiomes qui définissent la valeur d'Aumann-Shapley, une discussion et une application de ces prix à la recherche d'équilibre. Plus généralement, les prix sont étudiés dans plusieurs problèmes concrets.

Économie publique On rencontre en économie appliquée des problèmes de répartition de coût. Des communes se regroupant pour construire un équipement important : un incinérateur, une centrale électrique ou autre, doivent décider comment répartir le coût de la construction. On trouvera dans l'article (Zaccour 1988) des références à plusieurs articles appliqués dans le domaine.

Tarification La détermination du prix de vente est un problème central en l'économie. Si la notion de marché et de concurrence (parfaite ou non) apporte des solutions dans la plupart des cas, il faut utiliser une stratégie différente dans le cas de monopole et en particulier dans le cas de service public. En effet, dans ce cas, l'objectif n'est plus un maximum de profit mais la réalisation d'objectifs sociaux dans lesquels l'égalité des usagers joue un rôle prépondérant.

L'article appliqué à la tarification de coût du téléphone à l'université de Cornell de (Billera, Heath, et Raanan 1978) est une référence dans le domaine. Il donne une solution pratique en choisissant judicieusement la fonction de production dans un cas, où l'on souhaite déterminer des prix en fonction de l'heure, du type de communication et du type de jour (semaine ou weekend). Sur la même thème, la tarification et la répartition des coûts dans les télécommunication, l'analyse de Moreaux et Encaoua (1987) compare les modes de tarification en mettant l'équité et l'efficacité au centre de leur problématique.

Coût conjoint Une entreprise rassemblant plusieurs filières chacune dédiée à un produit différent peut vouloir déterminer le coût de production de chaque élément. Pour cela, il faut répartir les coûts conjoints que sont les coûts informatiques, la gestion des ressources humaines, etc. La valeur de Shapley permet de répartir ces coûts dont la mise en commun permet de réaliser des économies (Hamlen, Hamlen, et Tschirhart 1980).

L'article (Littlechild et Thompson 1977) étudie des problématiques que l'on rencontre dans les aéroports et en particulier celui de Birmingham. La question centrale est de savoir comme répartir les coûts associés aux pistes d'atterrissage entre les compagnies qui les utilisent. Cet article ingénieux a permis de faire prendre conscience de l'intérêt pratique de la valeur de Shapley. L'ouvrage (Boyer, Moreaux, et Truchon 2006) couvre un champ important avec des explications claires et des exemples appliqués des méthodes de répartition des coûts et de tarification.

Cette partie a présenté un grand nombre d'applications de la valeur de Shapley. On trouvera dans (Moretti et Patrone 2008), une revue de bien d'autres domaines dans lesquels la valeur Shapley est utilisée : l'étude des réseaux sociaux et de la « centralité » dans ces réseaux ; l'appréciation des facteurs risques dans les épidémiologiques ; la détermination de l'influence des gènes dans certains mécanismes biologiques ; ou encore la mesure de la sécurité associée à chaque composant d'un système complexe (un circuit électrique). On peut bien sûr en imaginer d'autres en fonction de ses sujets d'études.

Si les exemples d'applications sont nombreux, il est important de relever les limites de la valeur de Shapley. C'est ce qui sera fait dans la section suivante. On montrera aussi que lorsque l'on a une organisation préalable des joueurs, on peut adapter la valeur de Shapley et ainsi contourner un certain nombre de problèmes.

6 Limites et extensions

6.1 L'étude de tous les sous-ensembles

Comme on l'a déjà mentionné, on peut être gêner par l'idée de considérer tous les sous-ensembles possibles, y compris certains n'ayant que très peu de sens 5.2. Il faut rappeler que la procédure de Shapley inclut toutes les combinaisons possibles pour ne pas juger avec un *a priori* l'influence des différents joueurs. Si cette configuration est fictive est irréaliste, il faut tout de même l'introduire dans le calcul pour tenir compte des interactions particulières avec ces joueurs. Pour reprendre l'exemple d'une zone euro formée de la Belgique et de la Grèce à partir de laquelle on calculerait l'impact marginal de la France sur les exportation, par exemple, Shapley ne demande pas de considéré que cette situation est possible, sa valeur étudie simplement ce cas afin de prendre en compte l'interaction particulière du duo Belgique-Grèce sur la France lorsqu'aucun autre pays ne perturbe cette relation. Ce procédé permet d'assurer les propriétés d'équité *a priori* entre les joueurs.

On peut voir la valeur marginale comme un cas extrême de cette logique. La valeur marginale est en effet une valeur de Shapley qui interdit les sous-ensembles de cardinal inférieur à $n-1$. On sait que dans ces cas on n'assure plus la stricte répartition du résultat collectif entre tous les joueurs.

Si l'argument d'équité *a priori* et si le recours à des « remplaçants » continuent de bloquer l'utilisation de la valeur de Shapley, une solution est donnée plus bas dans un cas précis.

6.2 Connaitre le score sur chaque sous-ensemble

On a supposé dans tout ce document que l'on connaissait le score qui serait obtenue par chaque sous-ensemble de E. On n'a pas insisté sur cette condition, pourtant elle est souvent dirimante. En effet, il y a un certains nombres de cas dans lesquels le nombre de joueur est fixé et où évaluer le score sur un sous-ensemble n'a pas de sens[12]. On a vu au 5.7 que l'on pouvait en général contourner cet obstacle en invoquant une situation de référence contrefactuelle.

[12] La valeur marginale pure aussi peut, dans ce cas, ne pas être définie ou pas connue

Malgré cette possibilité de contournement, il reste en pratique à déterminer le score du sous-ensemble. Ce score demande parfois une observation : il faut connaitre le résultat d'une équipe en fonction de la composition des joueurs, il faut connaitre le résultat d'une entreprise avec des remplaçants, etc. On butte alors sur le fait que le remplaçant est souvent fictif ou qu'on n'observe jamais par exemple tous les remplaçants en même temps : on est alors dans l'impossibilité de déterminer un score sur tous les sous-ensembles et la valeur de Shapley ne peut s'appliquer.

Heureusement, la valeur de Shapley peut être utilisée dans des situations où un calcul donne simplement les valeurs sur tous les sous-ensemble. C'est le cas du vote ou de la réduction des inégalités par exemple. En économie, on est souvent capable de déterminer la grandeur à séparer dans toutes les configurations possibles. Le cas échéant, c'est l'utilisation de modèles et un jeu de prévision qui permettent tout de même de passer outre la méconnaissance du score sur toutes les configurations possibles.

6.3 Limite pratique sur le nombre de joueurs

Même dans les cas où il n'y a pas d'ordre sur les joueurs, le nombre de sous-ensembles croit exponentiellement avec le nombre de joueurs (2^n évaluations de score sont nécessaires pour calculer la valeur de Shapley d'un joueur lorsque qu'il y en a n). Aussi, on atteint généralement rapidement des limites au niveau de la puissance de calcul ou dans le stockage des données lorsque l'on a un nombre de joueur trop grand. Par exemple pour 15 joueurs, il y a 32 768 sous-ensembles sur lesquels calculer un score. On dépasse le million de sous-ensembles pour vingt joueurs.

Les temps de calcul peuvent être long et lorsqu'il faut observer des situations pratiques, le nombre de joueurs ne peut être que très limités. Pour un équipe de basket de cinq joueurs avec cinq remplaçants, il y a 32 équipes différentes à évaluer, c'est un travail fastidieux.

6.4 Niveau de désagrégation

Pour solutionner le problème du temps de calcul vu à l'instant, une idée est de regrouper les joueurs (par exemple ceux ayant le même rôle : ouvriers, partis politiques avec le même poids, transferts monétaires proches conceptuellement, etc.). Cette méthode se heurte à un obstacle : **la valeur de Shapley n'est pas indépendante du niveau de désagrégation**.

Problème Si l'on regroupe deux joueurs en un seul et qu'on l'on applique à nouveau la procédure de Shapley, la valeur de ce duo ne sera pas nécessairement la somme des valeurs des deux joueurs indépendants qui le constituent. On comprend en effet qu'une fois les deux joueurs regroupés on ne calcule plus la valeur marginale lorsque l'un est présent et l'autre non. Si ce résultat est heureux en termes de justice sans *a priori* et d'équité entre les joueurs, il a l'inconvénient de ne pas permettre de simplifier les calculs, par exemple en regroupant deux joueurs identiques.

Exemple 1 Par exemple, dans le contexte du vote à l'assemblée nationale (se reporter à 5.3), si les groupes E et F fusionnent, ils forment un parti majoritaire détenant tout le pouvoir de vote de cette assemblée donc 100% du pouvoir et non la somme de leur pouvoir avant fusion de 63,3%. En effet, le fait que D puisse parfois faire basculer le vote lorsque E et F sont en désaccord n'existe plus si bien que son pouvoir est nul. Dans les faits, ce sont les groupes A et B qui ont fusionné, cela leur a permis d'avoir la même importance que le groupe C à 6,7% : à l'évidence, il y a d'autres considérations en politique que l'optimisation du pouvoir de vote préconisée par la valeur de Shapley.

Exemple 2 Le calcul de contributions à la réduction des inégalités mesurées par la valeur de Shapley, présenté au 5.8 est un calcul sur 12 joueurs-transferts. Même s'il pourrait certainement être optimisé, le calcul total nécessite pour chaque joueurs une évaluation de l'indice de Gini sur $2^{12} = 4096$ situations et donc le duodecuple pour l'ensemble des transferts. Le tableau 5 suivant donne les résultats de la procédure de Shapley lorsque l'on regroupe les transferts pour n'avoir plus que six catégories de transferts : les cotisations et les contributions distributives sont regroupées, on ne sépare pas les impôts sur le revenu négatif et ceux positifs, les prestations familiales sont rassemblées (celles avec ou sans condition de ressource et les allocations familiales auxquelles on ajoute l'Apa), enfin, le RSA activité est considéré comme faisant partie des minima sociaux. Avec ces six catégories, le calcul qui dépassait 36 heures ne prend plus que deux minutes. Il faut dire qu'il n'y a plus que $64 \times 6 = 384$ indices de Gini à calculer. Une colonne donne la somme des valeurs individuelles des transferts de chaque catégorie telles que calculés avec la décomposition la plus large.

Les résultats de la méthode agrégée sont ici proches de ceux de calculés sans regroupement des transferts. Cela est cependant loin d'être une règle générale. L'exemple 1 le montre et on pourra trouver dans (Sastre et Trannoy 2002) des exemples dans lesquels l'agrégation de transferts modifie fortement la contribution de l'ensemble. Ici, la stabilité est due au fait que les transferts regroupés sont assez similaires et la valeur marginale des aides au logement n'est pas modifiée fortement lorsque le RSA activité n'est jamais présent sans les minima (et vice-et-versa).

	Sommes des valeurs de Shapley individuelles	Valeur de Shapley du groupe
Financement de la protection sociale	12,2	12,1
Impôt sur le revenu	28,5	28,5
Taxe d'habitation	0,1	0,1
Prestations familiales	25,1	25,2
Aides au logement	17,0	17,0
Minima sociaux	17,2	17,2

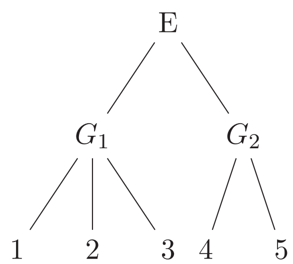
Tableau 5 – Calcul de la valeur de Shapley des transferts lorsque ceux-ci sont regroupés par catégorie. La colonne de gauche donne la somme des contributions des joueurs qui sont dans cette catégorie telles que calculées en 5.8

6.5 Deux solutions donnant un peu plus de souplesse à la valeur de Shapley

Une solution à l'ensemble des problèmes posés plus haut peut consister à définir des coalitions de joueurs. Il s'agit de regrouper les joueurs en famille ou en coalition. On étudie alors les interactions à l'intérieure de chaque coalition en voyant les autres coalitions comme des joueurs indissociables. Si on arrive à résoudre la difficulté lié au niveau de désagrégation ci-dessus, on pourra limiter le nombre de coalitions observées et donc le temps de calcul. On pourra, en choisissant judicieusement, la partition des ensembles éviter des coalitions que l'on juge aberrante et que l'on tient à éliminer malgré les réserves faites sur cette méthode au 6.1.

Le graphique ci-dessous illustre une partition de cinq joueur en deux groupes.

FIGURE 2 – Partition d'un ensemble E en deux sous-groupes



L'idée est que pour déterminer la valeur d'un des joueurs de G_1, on regarde sa valeur marginale quand G_2 dans son ensemble est présent ou absent. Par exemple, on ne regarde pas la configuration (2,4) pour déterminer la valeur de 1.

Formellement, on se place donc dans le cadre de n joueurs d'un ensemble E mais qui peuvent être regroupés en m groupes de joueurs. Notons $G = (G_1, G_2, \cdots, G_m)$ la partition qui en découle. Par définition on a : $\cup_{k=1}^m G_k = \mathrm{E}$ et $\forall k, l;\ G_k \cap G_l = \emptyset$.

On aimerait voir respectée une condition dite d'efficacité relative : la valeur d'un groupe G_k doit se répartir intégralement entre les joueurs qui le composent G_k :

$$\forall k, \sum_{i \in G_k} \phi(i) = \phi(G_k)$$

Bien évidemment, puisque l'on souhaite communiquer sur la valeur des groupes, il faut que ces valeurs respectent les propriétés de la valeur de Shapley. On n'a pas de marge de manœuvre et on sait que la seule façon de répartir le score global $v(\mathrm{E})$ parmi les ensembles $(G_1, G_2, \cdots, G_m)$ sans céder sur les propriétés désirables énoncées plus haut dans ce document est d'utiliser la valeur de Shapley en assimilant les groupes à des joueurs. On note $Sh_{|G}$ cette valeur pour indiquer qu'elle est calculée au niveau des groupes.

C'est donc sur la valeur des « feuilles », des joueurs constituant les groupes qu'il faut travailler. Il faut répartir la valeur $Sh_{|G}(G_k)$ parmi les joueurs qui font partie de ce groupe en notant qu'en général, cette valeur calculée dans le jeu formé par les groupes ne vaut pas la sommes des valeurs de Shapley de joueurs calculé dans le jeu total.

6.5.1 La méthode d'Owen

Une belle idée pour résoudre le problème précédent, due à Owen (1977), ravira ceux qui bien compris la valeur de Shapley. En effet, le problème posé rentre exactement dans le cadre de la valeur de Shapley, il s'agit bien de répartir une valeur collective $Sh_{|G}(G_k)$ sur un ensemble de joueurs. $Sh_{|G}(G_k)$ est donc la valeur d'un groupe. Tête baissée, on calcule une valeur de Shapley pour chaque joueur en mesurant comment il impacte marginalement $Sh_{|G}(G_k)$ en moyenne sur toutes les constitutions possibles du groupe G_k. Dit autrement, pour un joueur i dans G_k, on regarde pour chaque sous-ensemble constitué à partir des éléments de G_k et ne contenant pas i ce que serait la valeur de G_k dans le jeu $Sh_{|G}$ s'il n'y avait que ce sous-ensemble dans G_k et on la compare avec ce qu'elle serait en ajoutant i à ce sous-ensemble. En fait, $Sh_{|G}(G_k = \mathrm{F}) = v(\mathrm{F})$ est le score d'un sous-ensemble F : le score d'un sous-ensemble est lui-même une valeur de Shapley !

Pour un joueur i appartenant au groupe G_k de taille s_k, on appelle, pour chaque sous-ensemble F de G_k contenant i, $\mathrm{F} \in \mathcal{P}(G_k^i) \cup \{i\}$, $G_k(\mathrm{F})$ le groupe G_k réduit aux f joueurs de F. La valeur du joueur i s'écrit alors :

$$\phi^{\mathrm{Owen}}(i) = \frac{1}{s_k!} \sum_{\mathrm{F} \in \mathcal{P}(G_k^i)} (s_k - f)!f!\Big[Sh_{|G}(G_k(\mathrm{F})) - Sh_{|G}(G_k(\mathrm{F} \setminus \{i\})\Big]$$

Rappelons que dans cette expression, les termes $Sh_{|G}$ cachent une évaluation de la valeur de Shapley d'un groupe. Les deux valeurs se simplifient toutefois car pour chaque sous-ensemble de G ne contenant pas G_k, on fait en fait la différence de l'impact marginal de $G_k(\mathrm{F})$ et de $G_k(\mathrm{F} \setminus \{i\}$.

6.5.2 La méthode de Nested-Shapley

L'autre méthode pour donner une valeur à un joueur au sein d'une coalition consiste à conserver le même score v que dans le jeu sans groupe mais à n'étudier que l'intérieur du groupe contenant le joueur i, on peut appeler $Sh_{|G_k}(i)$ la valeur des joueurs i du groupe G_k ainsi calculé. De cette manière, on répartit $v(G_k)$ entre les joueurs, or, comme en général $v(G_k)$ n'est pas égal à $Sh_{|G}(G_k)$, on n'a pas l'efficacité relative directement. L'ajustement consiste à répartir cette différence également entre les joueurs ou bien en ajoutant $\frac{v(G_k)-Sh_{|G}(G_k)}{s_k}$ à chaque joueur, ou bien en respectant un peu plus les contributions mesurée par Shapley en multipliant les contribution par $\frac{Sh_{|G}(G_k)}{v(G_k)}$. Cette méthode dite *Nested*-Shapley ou Shapley emboités, détaillée par (Chantreuil et Trannoy 1999) est moins pure que la méthode d'Owen. Elle n'en a pas toutes ses propriétés de bon comportement et n'est pas solution unique dans une recherche de valeur respectant certaines propriétés(se référer à (Winter 2002) pour plus de détails).

Cependant, dans certains cas, cette méthode des Shapley emboités peut être préférées à la méthode d'Owen. En effet, parfois, on peut trouver que le fait de comparer un joueur à un groupe dans la méthode d'Owen est faible en sens économique (voir (Sastre et Trannoy 2002)). Cet argument est relativement faible mais il va dans le sens d'un argument pratique important. La méthode des Shapley emboités est plus simple à mettre en place et demande moins de calculs que la méthode d'Owen. Lorsqu'il y a K partitions et s_k joueurs dans le groupe de i alors la méthode des Shapley-emboité demande 2^{K-1} valeurs marginales (et donc le double de scores) pour trouver la valeur de Shapley du groupe G_k et 2^{s_k-1} pour la valeur de Shapley du joueur i. Pour la méthode d'Owen, il faut pour chacun des 2^{s_k-1} sous-ensembles de G_k contenant i déterminer la différence de deux valeurs de Shapley. Comme on l'a dit plus haut, ce calcul se simplifiant, il n'y a que l'équivalent d'un valeur à calculer ce qui fait 2^K évaluation du score.

Au final, les 2^n score à calculer lorsqu'on ne réunit pas de joueurs en groupe ne sont plus que 2^{K+s_k-1} avec la méthode d'Owen et seulement $2^K + 2^{s_k}$ avec la méthode des Shapley emboités. On comprend l'avantage de cette dernière méthode dans des cas complexes.

Signalons que l'on peut s'autoriser plusieurs niveaux d'agrégation, il y a alors autant d'étapes que de niveaux dans la méthode des Shapley emboités. La méthode d'Owen peut elle-aussi être généralisée à plusieurs niveau de groupes, sous-groupes, etc. On peut aussi, pourquoi pas, mixer ces deux méthodes dans un arbitrage entre justesse et temps de calcul.

Notons enfin qu'il faut une certaine expertise dans le domaine traité pour déterminer l'arbre de regroupement. Il faut trouver une proximité entre les joueurs. Une bonne maitrise du fonctionnement de la valeur de Shapley dans des cas simples permet en général de comprendre les évolutions des contributions.

7 Conclusion

La valeur de Shapley a été introduite progressivement dans ce document. On peut noter en conclusion qu'elle respecte l'intuition que l'on associe à la valeur marginale mais en y ajoutant de la justesse et de la justice. A chaque fois que l'on s'interroge sur ce qu'apporte un élément à un groupe, on peut commencer par raisonner en terme de valeur marginale pour préciser ce que l'on souhaite mesurer pour en déduire un score et ensuite appliquer la valeur de Shapley.

Ce document a tenté de montrer l'ampleur de la recherche qui s'est développé depuis l'article de Shapley en 1953. La fonction de la valeur de Shapley est de répartir un résultat collectif entre plusieurs individus. Comme on peut le deviner à la lecture de ce document, le champ d'application de cette méthode est vaste. L'étendue du domaine d'application de la valeur de Shapley s'explique certainement par sa souplesse. En effet, même si elle se présente comme un concept rigide de théorie des jeux, la valeur de Shapley laisse une grande liberté à celui qui l'utilise. La liberté de choisir le score est la plus importante, elle permet de s'adapter à des situations complexes. Par exemple, la possibilité de raisonner sur des contrefactuels ou d'utiliser des arbres d'appartenance à des groupes permet d'évaluer des contributions même lorsque connaitre le résultat d'un sous-groupe est délicat.

Pour terminer, on sait qu'il ne faut pas négliger la communication qu'on peut avoir autours d'une méthode. En particulier, la valeur de Shapley n'est pas connue autant qu'elle le devrait et nécessite souvent un développement. Si on veut éviter d'y consacrer trop de temps, par exemple dans une diffusion grand public, on peut certainement se contenter de dire qu'il s'agit de ce qu'apporte en moyenne un individu aux autres. Selon le contexte, on pourra insister sur le fait qu'elle tient compte de toutes les interactions, qu'elle n'a aucun *a priori* sur les joueurs ou bien que c'est la seule méthode à vérifier certaines propriétés.

8 Bibliographie

Références

Aumann, R. J., et L. S. Shapley (1974) : *Values of Non-Atomic Games.* Princeton University Press.

Auvray, C., et A. Trannoy (1992) : "Décomposition par source de l'inégalité des revenus à l'aide de la Valeur Shapley," Journées de Microéconomie appliquée. Sfax.

Billera, L. J., D. C. Heath, et J. Raanan (1978) : "Internal Telephone Billing Rates-A Novel Application of Non-Atomic Game Theory," *Operations Research*, 26(6), pp. 956–965.

Boyer, M., M. Moreaux, et M. Truchon (2006) : *Partage des coîts et tarification des infrastructures*, no. 2006-01 in CIRANO Monographs. CIRANO.

Cazenave, M.-C., J. Duval, A. Eidelman, F. Langumier, et A. Vicard (2011) : "La redistribution : état des lieux en 2010 et évolution depuis vingt ans," dans *France, portrait social*. INSEE Référence.

Chantreuil, F., et A. Trannoy (1999) : "Inequality decomposition values : the trade-off between marginality and consistency," THEMA Working Papers 99-24, THEMA (THéorie Economique, Modélisation et Applications), Université de Cergy-Pontoise.

Devicienti, F. (2010) : "Shapley-value decompositions of changes in wage distributions : a note," *Journal of Economic Inequality*, 8, 35–45.

Gajdos, T. (2001) : "Les fondements axiomatiques de la mesure normative des inégalités," Université paris1 panthéon-sorbonne, HAL.

Ghintran, A. (2010) : "Marginalism and the Shapley value," Working Paper Series 1007, Obuda University, Keleti Faculty of Economics.

Hamlen, S. S., J. Hamlen, William A., et J. Tschirhart (1980) : "The Use of the Generalized Shapley Allocation in Joint Cost Allocation," *The Accounting Review*, 55(2), pp. 269–287.

Littlechild, S. C., et G. F. Thompson (1977) : "Aircraft Landing Fees : A Game Theory Approach.," *Bell Journal of Economics*, 8(1), 186 – 204.

Mirman, L. J., et Y. Tauman (1981) : "Valeur de Shapley et répartition équitable des couts de production," *Cahiers du Séminaire d'Économétrie*, (23), pp. 121–151.

Moreaux, M., et D. Encaoua (1987) : "L'analyse théorique des problèmes de tarification et d'allocation des coûts dans les télécommunications," *Revue Économique*, 38(2), 375–414.

Moretti, S., et F. Patrone (2008) : "Transversality of the Shapley value," *TOP*, 16, 1–41.

Moulin, H. (1992) : "An Application of the Shapley Value to Fair Division with Money," *Econometrica*, 60(6), pp. 1331–1349.

Mussard, S., et B. Philippe (2007) : "Une évaluation du rôle des déterminants du partage de la valeur ajoutée," Cahiers de recherche 07-14, Département d'Économie de la Faculté d'administration à l'Université de Sherbrooke.

Osnat, I. (2007) : "A Shapley-based decomposition of the R-Square of a linear regression," *Journal of Economic Inequality*, 5(2), 199–212.

Owen, G. (1977) : *Essays in Mathematical Economics and Game Theory*chap. Values of Game with a priori Unions, pp. 76–88. Springer-Verlag.

Roth, A. E., et al (1988) : *The Shapley value : essays in honor of Lloyd S. Shapley*. Cambridge University Press, Cambridge [Cambridgeshire] ; New York :.

Roy, D. (2006) : "L'argent du « ménage », qui paie quoi ?," *Travail, genre et sociétés*, 1(15), 101–119.

Sastre, M., et A. Trannoy (2002) : "Shapley inequality decomposition by factor components : Some methodological issues," *Journal of Economics*, 9(1), 51–89.

Shapley, L. (1953) : "A value for n-person games," dans *Contributions to the Theory of Games, Vol. II*, ed. H. Kuhn, et A. Tucker, vol. 28 of *Annals of Mathematics Studies*, pp. 307–317. Princeton University Press, Princeton, NJ.

Shapley, L. S., et M. Shubik (1954) : "A Method for Evaluating the Distribution of Power in a Committee System," *The American Political Science Review*, 48(3), 787–792.

Shorrocks, A. (1999) : "Decomposition Procedures for Distributional Analysis : A Unified Framework Based on the Shapley Value," *Mimeo,Essex University*.

Winter, E. (2002) : "The shapley value," dans *Handbook of Game Theory with Economic Applications*, ed. R. Aumann, et S. Hart, vol. 3 of *Handbook* *of Game Theory with Economic Applications*, chap. 53, pp. 2025–2054. Elsevier.

ZACCOUR, G. (1988) : "Valeur de Shapley et partage équitable des ressources," *L'Actualité Economique*, 64(1), 96–121.

9 Annexes

9.1 Exemple d'étude : le taux de pauvreté

Dans cette annexe, nous nous penchons sur le taux de pauvreté calculé à partir de l'ERFS 2008 comme étant la proportion d'individu vivant en dessous du seuil de 60 % du revenu médian. La valeur de Shapley peut décomposer ce taux de pauvreté et surtout quantifier des éléments en évolution.

Cet exemple a simplement vocation à illustrer ce que la valeur de Shapley peut apporter à une étude. Il s'agit bien d'un exemple appliqué et non réellement d'une étude : les chiffres n'ont pas été vérifiés, les choix ne sont pas probablement pas les plus pertinents et une conclusion ou une analyse à partir de ces calculs bruts serait douteuse. Il s'agit simplement de montrer le discours que permet de tenir la valeur de Shapley.

9.1.1 Décomposition du taux de pauvreté

Lorsque l'on étudie la pauvreté, on veut connaitre les caractéristiques socio-démographiques des pauvres. On regarde alors la composition des personnes vivant sous le seuil de pauvreté ou la proportion de pauvres pour différentes catégories de personnes (retraités, familles monoparentale, ...). Dans cette démarche le seuil de pauvreté calculé sur la population générale sert de référence. On peut se demander quelle est l'influence de chaque catégorie de personne sur ce seuil de pauvreté et sur le taux de pauvreté qui en découle. Quelle est l'influence des enfants sur le taux de pauvreté ? Comment l'évolution des populations influence ce taux ? Ce sont des questions qui peuvent trouver des réponses grâce à la valeur de Shapley [13].

On utilise donc la valeur de Shapley en ayant pour joueurs des catégories de populations prédéfinies et pour score d'un sous-ensemble, le taux de pauvreté mesuré sur ce sous-ensemble[14][15]. A titre d'exemple, on choisit deux type de répartition de la population : la répartition hommes/femmes et une répartition par âge[16]. On pourrait en imaginer d'autres par exemple les ménages bénéficiaires du RSA ou selon la configuration familiale et bien sûr on peut appliquer la valeur de Shapley aux croisements de ces catégories pour obtenir une influence du RSA sur le taux de pauvreté selon la configuration

[13] Encore une fois, cette approche sert d'illustration et n'est peut-être pas la plus pertinente.

[14] On choisit que le taux de pauvreté de l'ensemble vide est nul.

[15] On pourrait aller plus loin en recalculant les niveaux de vie sur chaque population, en effet, si on retire les enfants alors on peut calculer un nouveau niveau de vie pour leur parent ne tenant pas compte des dépenses associées à ces enfants.

[16] Les catégories ont été choisies pour faire être comparer au tableau disponible sur le site de l'Insee pour 2008 : `http://www.insee.fr/fr/themes/tableau.asp?reg_id=0&ref_id=NATTEF04416`

familiale.

On peut donc décomposer le taux de pauvreté par sous-population. Ces calculs sont présentés plus bas dans le tableau 6. Les hommes sont responsables de 7,2 des 13,1 pour cent du taux de pauvreté, cela représente 55,6 %. Les femmes augmentent en moyenne le taux de pauvreté de 5,8 points. Cette moyenne est logiquement plus influencée par le taux de pauvreté intrinsèque car on compare alors le taux de pauvreté sur la population des femmes 12,3 % à zéro. Il ne s'agit pas de penser que les femmes augmente le taux de pauvreté des hommes de 5,8.

Il faut être prudent sur l'analyse de tels résultats. On peut montrer que la contribution des familles monoparentales au taux de pauvreté est négative. Elles font baisser le taux de pauvreté alors que 29,6 % d'entre elles sont en dessous du seuil de pauvreté. L'explication réside en ce que lorsque l'on introduit les familles monoparentales dans la population, un bon nombre d'entre elles est pauvre mais en baissant le niveau du revenu médian et donc du seuil de pauvreté, certains ménages se retrouvent mécaniquement au-dessus du seuil alors qu'ils sont en dessous sans les familles monoparentales. Le seuil de pauvreté est endogène et la valeur de Shapley capte cet effet. On peut critiquer une telle approche dans ce cas précis et considérer qu'elle n'a pas de sens. Ce n'est cependant pas la valeur de Shapley qu'il faut remettre en cause mais son application rapide et purement illustrative dans cet exemple.

La contribution des différentes populations au taux de pauvreté peut être décomposée en plusieurs facteurs. Par exemple, il semble naturel de rapporter les contributions des sous-populations à leur taille m_k. En effet, il est logique qu'une population importante influence plus le taux de pauvreté qu'une population plus petite. On décompose la contribution au taux de pauvreté de chaque sous-population k en une partie $\nu_k = \frac{m_k}{m} \times P$ qui caractérise ce que serait sa contribution si la pauvreté était répartie de façon homogène entre les populations[17] et une partie I_k qui mesure justement l'influence dans le taux de pauvreté relativement à sa masse.

$$\Delta T = \sum_{k=1}^{n} \nu_k I_k$$

Si on le désir, on peut ensuite décomposer les contributions de Shapley comme étant le résultat d'une valeur intrinsèque et d'un terme d'interaction avec les autres joueurs. Il peut être multiplicatif ou additif. Même si cela n'a pas nécessairement une grande signification économique ici, on décompose la contribution d'une population par sa masse relative $\mu_k = \frac{m_k}{m}$, son taux

[17] m_k désigne le nombre de personne dans la sous-population k et m désigne la population totale.

de pauvreté intrinsèque P_k, calculé sur le sous-ensemble et un coefficient d'influence sur la société G_k :

$$\Delta T = \sum_{k=1}^{n} m_k P_k G_k$$

L'interprétation que l'on peut donner est la suivante : le premier terme donne le nombre de personnes dans la population, le deuxième terme donne une idée de la répartition à l'intérieur du groupe et le troisième caractérise la place du centre de gravité du groupe par rapport à la distribution des niveaux de vie dans l'ensemble de la population[18].

Les contributions par catégorie ainsi que les décompositions en deux ou trois termes sont données dans le tableau suivant :

	$Sh(k)$ (en %)	ν_k (en %)	I_k	$m_k(\times 100)$	P_k	G_k
Homme	55,6	48,4	1,15	48,4	13,8%	1,09
Femme	44,4	51,6	0,86	51,6	12,3%	0,91
- de 18 ans	2,7	22,3	0,12	22,3	10,3%	0,15
18-29 ans	9,5	14,3	0,66	14,3	11,2%	0,77
30-49 ans	22,9	27,7	0,83	27,7	14,6%	0,74
50-59 ans	34,9	19,8	2,57	19,8	19,8%	1,69
60-74 ans	22,0	22,0	1,59	22,0	14,9%	1,39
75 ans et +	8,0	8,0	0,98	8,0	9,8%	1,30

Tableau 6 – La première colonne indique l'influence de la sous-population sur le taux de pauvreté. Les colonnes suivantes décomposent cette contribution en fonction de sa masse relative et pour les trois dernières colonnes du taux de pauvreté intrinsèque. A noter que ν_k est le nombre de personnes en pourcentage de la quantité $\frac{P}{m}$ tandis que m_k est le nombre de personnes rapporté à la quantité m

Les hommes sont moins nombreux mais influencent plus le taux de pauvreté que les femmes, leur influence relative, I_k ou G_k, est supérieure à 1. Les femmes contribuent moins au taux de pauvreté, cela ne s'explique pas par leur nombre et que peu par le faible taux de pauvreté intrinsèque [19].

L'influence selon l'âge sur la pauvreté semble croître depuis un niveau très faible des - de 18 ans qui semblent donc bien répartis dans la population, jusqu'à la tranche 50-59 ans. Cette tranche, avec un taux de pauvreté

[18] Il y a, là encore, à redire sur cette décomposition, mais encore une fois, il ne s'agit pas d'une étude. Une idée peut-être plus naturelle serait de mettre à la place du taux de pauvreté intrinsèque qui n'a pas grand sens, le rapport entre le niveau de vie moyen du groupe et le seuil de pauvreté.

[19] $1 < G_k < I_k$: si m_k et P_k expliquait parfaitement le rôle des femmes, on aurait $G_k = 1$ ce n'est pas la cas mais comme G_k est plus proche de 1 que ne l'est I_k, on peut se dire que P_k explique tout de même une partie des choses.

intrinsèque élevé semble se distinguer par une forte densité de population avec un niveau de vie très bas. Encore une fois, il faudrait plus de temps pour approfondir et valider ces résultats.

9.1.2 Évolution

Une fois le constat fait pour l'année 2008, on peut ensuite avoir envie de regarder l'évolution du taux de pauvreté. On regarde ici, l'évolution depuis 2004. Le taux de pauvreté était cette année-là de 11,7% (c'est 13,1 % en 2008). On calcule désormais des contributions à l'évolution du taux de pauvreté en prenant pour score collectif l'évolution du taux de pauvreté mesuré en 2004 et en 2008 sur la sous-population.

En fait, grâce à la propriété de l'additivité que respecte la valeur de Shapley, on peut calculer les contributions au taux de pauvreté en 2004 avec le même score que précédemment et simplement les soustraire à celles mesurées en 2008 ($Sh_{\Delta P} = Sh_{P_{2008}} - Sh_{P_{2004}}$).

Ce qui est intéressant avec la valeur de Shapley ici, c'est qu'elle permet ensuite d'étudier l'évolution de l'influence d'une sous-population, en fonction de chacun des paramètres. Ainsi, si on regardait les parents isolés, la valeur de Shapley saurait faire la distinction entre l'évolution du nombre de parents isolés et l'évolution associée au niveau de vie de ces parents.

Le tableau suivant donne la contribution de chaque sous-population à l'évolution du taux de pauvreté dans la première colonne. Les colonnes de droite donne la contribution de l'évolution de chacun des facteurs à l'évolution générale de la contribution de la population (tous les chiffres sont exprimés en %) :

On peut lire que l'évolution du taux pauvreté s'explique à 63,0% par l'évolution de la population des hommes. Cette contribution n'est pas due à l'évolution de m_k qui est faible et qui a plutôt tendance à baisser le taux de pauvreté mais bien plus par l'évolution de la répartition des hommes qui ont un taux de pauvreté plus élevé en 2008 qu'en 2004. Il faut noter deux choses : le rôle de G_k n'est pas évident à analyser en évolution ; en effet presque par construction, G_k est un élément résiduel. Enfin, dans la décomposition à deux terme qui est fait, le paramètre $\nu_k = \frac{m_k}{m} \times P$ intègre le taux de pauvreté. Il est lui-aussi compliqué à analyser et serait sans doute à écarter dans une étude.

On peut tout de même noter que l'évolution du taux de pauvreté ne s'explique pas par l'évolution de la tranche 50-59 ans (contribution négative) pourtant tranche la plus importante dans la détermination de ce taux (34,9% en 2008). Sa masse relative augmente ce qui devrait baisser le niveau de vie et lui donner une contribution positive à l'évolution du taux de pauvreté mais dans le même temps, l'intensité de son influence diminue et donne, au final, un amoindrissement de la tendance à la hausse de la pauvreté. Cet effet est capté en partie par son taux de pauvreté intrinsèque qui diminue

Catégories	Part dans l'évolution	rôle de ν_k	rôle de I_k	rôle de m_k	rôle de P_k	rôle de G_k
Homme	63,0	83,6	16,4	-4,0	90,2	13,8
Femme	37,0	126,4	-26,4	5,2	105,4	-10,5
- de 18 ans	7,1	30,1	69,9	-3,7	40,3	63,4
18-29 ans	-13,7	-68,2	168,2	10,6	-31,4	120,8
30-49 ans	13,9	105,7	-5,7	-62,5	136,0	26,5
50-59 ans	-10,1	-395,7	495,7	-24,9	43,7	81,1
60-74 ans	77,4	39,3	60,7	15,1	55,6	29,3
75 ans et +	25,3	47,6	52,4	19,9	40,4	39,7

Tableau 7 – La première colonne indique l'influence de la sous-population dans l'évolution globale entre 2004 et 2008. Les colonnes suivantes explique l'évolution d'une sous-population en fonction de l'évolution des deux critères ou des trois qui la caractérisent.

sur la période, on peut penser qu'il y a un peu moins de personnes au niveaux de vie très faible dans la population des 50-59 ans en 2008 qu'en 2004, ceci introduirait an moyenne une baisse du taux de pauvreté malgré l'augmentation générale de la taille de cette population.

9.2 Programmes

Dans cette annexe, on présente des programmes SAS qui permettent de calculer une valeur de Shapley de façon automatique. En dehors de la liste des joueurs, l'utilisateur de ces programmes a uniquement besoin de définir la fonction score. Elle doit s'écrire sous la forme d'une macro-fonction ayant un nom donné et prenant pour argument une liste sur laquelle elle calcule une valeur réelle. Cette valeur doit être enregistrée dans une macro-variable portant le même nom que la macro-fonction.

Pour commencer (Programme 1), on définit trois petites sous-macro-fonctions qui seront utiles dans le calcul. On comprend vite pourquoi : la première calcule le nombre de mots d'une liste, la deuxième rend le factoriel d'un nombre entier et la troisième donne à partir d'une sous-liste, la même sous-liste ne contenant pas un individu donné en argument.

On crée ensuite une table nommé `temp` ayant une ligne par sous-ensemble ne contenant pas le joueur. C'est la macro-fonction `% listsousens` qui construit cette table une fois celle-ci initialisée (voir Programme 2).

On calcule ensuite la valeur de Shapley (Programme 3). En argument de la macro-fonction, on a besoin du strict nécessaire : la liste des joueurs, le joueur dont on veut la valeur et le score. Le code SAS est normalement suffisamment détaillé pour suivre le calcul. Les instructions d'affichage permettent de suivre l'avancée du calcul, y compris avec l'option textttnonotes.

Enfin, la dernière macro-fonction (Programme 4) réalise le calcule de la valeur de Shapley pour tous les joueurs et inscrit le résultat dans un tableau de sortie.

```
************ 1 - création des sous-macros ******************;

/**une petite procédure qui compte le nombre d'éléments d'une liste;*/
%macro nbmot(liste);
%global nb;
%let nb=0;
%do %while(%scan(&liste,%eval(&nb+1)) ne  );
      %let nb=%eval(&nb+1);
%end;
%mend nbmot;
/*la macro factoriel;*/
%macro fact(n);
%global factoriel;
%if %eval(&n)<=0 %then %do; %let  factoriel=%eval(1); %end;
%else %do;
 %fact(%eval(&n-1));
      %let factoriel=%eval(&n*&factoriel);
      %end;
%mend fact;

/*** un fonction qui donne l ensemble ne contenant pas lvar **;*/
%macro contenant(varcont,listcont);
%global contenant;
%let contenant= ;
%nbmot(&listcont);
%if &nb>0 %then %do;
      %do i=1 %to &nb ;
            %if ( %scan(&listcont,%eval(&i)) ne &varcont )
            %then %do; %let contenant = &contenant
%scan(&listcont,%eval(&i)); %end;
      %end;
%end;
%mend contenant;
```

Programme SAS 1: Trois sous-macro-fonctions

```
/*la table temp a vocation à contenir une ligne pour chaque sous-ensemble
ne contenant pas le
joueur*/
data temp;
format ens $32499.;
input ens;
cards;
.
;
run;
%macro sousens(listsousens);

/*on compte le nombre de mots de listsousens*/
%nbmot(&listsousens);

/* initialisation en _n_=1, on a l'ensemble vide*/
data temp; set temp; if _n_=1; run;

%if &nb>0 %then %do;
      %do i=1 %to &nb ; /* pour chaque joueur de listsousens...*/
      /*...on ajoute le joueur à tous les ensembles déjà dans temp...*/
            data temp1; set temp; ens="%scan(&listsousens,&i)"!!'
'!!ens;run;
      /*...puis on assemble ces nouveaux ensembles aux anciens sans le
joueur i.*/
            data temp; set temp temp1; run;
      %end;
%end;
%mend sousens;
```

Programme SAS 2: Création de la liste des sous-ensemble à partir desquels on calculera une valeur marginale

```
 /*varshap est le nom du joueur dont on calcule la valeur,
  listshap est la liste des joueurs,
  et valeurshap est le score associé à chaque configuration*/
%global shap;
%let shap =0;
%global nbjoueur;
%nbmot(&listshap);%let nbjoueur = &nb;
%contenant(&varshap, &listshap);
%sousens(&contenant);

/*on compte le nombre de ligne de la table temp;*/
proc sql noprint;
   select count(*) into : cnt
      from work.temp ;
quit;
%put le calcul se fera sur &cnt sous-ensemble ne contenant
pas le joueur et donc sur le double de &cnt;

data temp2; set temp;
ens2="&varshap"!!' '!!ens;run;

%do ensemble = 1 %to &cnt;/*pour chaque sous-ensemble*/
      %local list1;
      %local list2;
/*on récupère le sous-ensemble et le sous-ensemble avec le joueur*/
      data _null_; set temp2; if _n_= &ensemble;
      call symput( "list1", ens);
      call symput( "list2", ens2);
      %put &list2;
      run;
/*calcul de la pondération de la valeur marginale sur le sous-groupe à
l'aide
de la fonction factoriel*/
      %fact(%eval(&nb-1));
      %let coeff=&factoriel;
      %fact(%eval(&nbjoueur-&nb));
      %let coeff=%sysevalf(&factoriel*&coeff);
      %fact(&nbjoueur);
      %let coeff=%sysevalf(&coeff/&factoriel);
 /*calcul de la valeur marginale*/
      %&valeurshap(&list2);%nbmot(&list2);
      %let shap=%sysevalf( &shap + %unquote(%str(&)&valeurshap)*&coeff ) ;
      %&valeurshap(&list1);
      %let shap=%sysevalf( &shap - %unquote(%str(&)&valeurshap)*&coeff ) ;
%end;
%put &shap;
%mend shapley;
```

Programme SAS 3: Calcul de la valeur de Shapley d'un joueur

```
%macro tableshap(liste,score,sortie=resultat);
option nonotes;
%shapley(varshap=%scan(&liste,1),listshap=&liste, valeurshap=&score);
data &sortie; set temp(keep=); if _n_=1; joueur="%scan(&liste,1)";
valeur=%sysevalf( &shap);
run;

%nbmot(&liste);
%do joueur=2 %to %eval(&nb);
%shapley(varshap=%scan(&liste,%eval(&joueur)),listshap=&liste,
valeurshap=&score);
data res; set temp(keep=); if _n_=1; joueur="%scan(&liste,%eval(&joueur))"
valeur=%sysevalf( &shap); run;
data &sortie; set &sortie res; run;
%end;
option notes;
%mend tableshap;
```

Programme SAS 4: Calcul de la valeur de Shapley pour tous les joueurs et sauvegarde dans un tableau de sortie