Stat4DS / Homework 03

Pierpaolo Brutti

Due whenever you want before February 18 (on Moodle)

General Instructions

I expect you to upload your solutions on Moodle as a **single running** `R Markdown` file (`.rmd`) + its `html` output, **named with your surnames**. Alternatively, a `zip`-file with all the material inside will be fine too.

R Markdown Test

To be sure that everything is working fine, start `RStudio` and create an empty project called `HW1`. Now open a new `R Markdown` file (`File > New File > R Markdown...`); set the output to `HTML mode`, press `OK` and then click on `Knit HTML`. This should produce a web page with the knitting procedure executing the default code blocks. You can now start editing this file.

---

1. Background: Cooperative Games (more info)

1.1 Generalities

A *game* in the sense of *game theory* is an abstract mathematical model of a scenario in which some sort of agents *interact*. It is abstract in the sense that irrelevant detail is omitted: the game aims to capture only those features of the scenario that are relevant to the decisions that must be made by players within the game.
Agents can be anything depending on the application: typically real humans (e.g. investors, political parties, etc.), but also signals/towers in a communication network, genes in a biological setup or even variables in a predictive/learning problem.

The form of games we are interested in is the most basic and widely-studied model of **cooperative games**.
More specifically, our games are populated by a (non-empty) set $\mathcal{P} = \{1, \ldots, p\}$ of agents: the **players** of the game. A **coalition** C is simply *any* subset of the player-set $\mathcal{P}$. The **grand coalition** is the set $\mathcal{P}$ of *all* players.

All this said, let's define what we mean by a *characteristic function game*. In the following, with $2^{\mathcal{P}}$ we will denote the power-set, that is, the set of all subsets, of $\mathcal{P}$.

Definition 1. A characteristic function game G is given by a pair $(\mathcal{P}, \nu)$, where $\mathcal{P} = \{1, \ldots, p\}$ is a finite, non-empty set of agents, and $\nu : 2^{\mathcal{P}} \mapsto \mathbb{R}$ is a **characteristic function**, which maps each coalition $\mathsf{C} \subseteq \mathcal{P}$ to a real number $\nu(\mathsf{C})$.
The number $\nu(\mathsf{C})$ is usually referred to as the value of the coalition C.

There are many possible interpretations for the *characteristic function*, but note right away that characteristic function games assign values to a coalition as a whole, and **not** to its individual members. That is, the model behind a characteristic function game does **not** tell you how the coalition value $\nu(\mathsf{C})$ should be **divided** among the members of C.
In fact, the question of how to divide the coalition value is a fundamental research topic in cooperative game theory[1].

Notice also that, from a computational point of view, the naïve representation of a characteristic function game that consists in explicitly listing every coalition C together with the associated value $\nu(\mathsf{C})$, being of the order 2^p in size, is **not** practical unless the number of players is very small. On the other hand, most real-life interactions admit an encoding of size polynomial in p; such an encoding provides an **implicit** description of the characteristic function.

As an example, consider modeling the decision-making process in voting bodies.

Example 1. (*weighted voting games*)

- A country has a 101 seats in its parliament, and each representative belongs to one of four parties: Liberal (L $\rightsquigarrow$ 40 seats), Moderate (M $\rightsquigarrow$ 22 seats), Conservative (C $\rightsquigarrow$ 30 seats), or Green (G $\rightsquigarrow$ 9).
- The parliament needs to decide how to allocate "*1 billion euros*" of discretionary spending.

[1] An implicit assumption here: the coalition value $\nu(\mathsf{C})$ can be divided among the members of C in *any* way that the members of C choose. Formally, games with this property are said to be transferable utility games (TU games)

- The decision is made by a **simple majority vote**, and we assume that all representatives vote along the party lines.
- Parties can form coalitions; a coalition has **value** "*1 billion euros*" **IF** it can win the vote no matter what the other parties do, and value 0 otherwise.

Hence, after some thinking, we see that the associated 4-players game has $\mathcal{P} = \{\mathrm{L}, \mathrm{M}, \mathrm{C}, \mathrm{G}\}$ and characteristic function

$$\nu(\mathsf{S}) = \begin{cases} 0 & \text{if } (\#\mathsf{S} \leqslant 1) \text{ or } (\mathrm{G} \in \mathsf{S}) \\ 10^9 & \text{otherwise} \end{cases} \qquad \text{where} \qquad \#\mathsf{S} = \{\text{cardinality of the coalition } \mathsf{S}\}.$$

1.2 Solution Concepts: the Shapley Value

A key problem in game theory is to try to understand what the **outcomes** of a game will be. In our cooperative framework, an *outcome* of a game G consists of two parts:

1. a **coalition structure**, that is, a **partition** $\mathrm{CS} = \{\mathrm{C}_1, \ldots, \mathrm{C}_m\}$ of the player-set $\mathcal{P} = \{1, \ldots, p\}$ into coalitions;
2. a **payoff vector** $\boldsymbol{x} = (x_1, \ldots, x_p) \in \mathbb{R}^p$ for a coalition structure $\mathrm{CS} = \{\mathrm{C}_1, \ldots, \mathrm{C}_m\}$, which distributes the value $\nu(\mathsf{C}_j)$ of each coalition among its members. Any legit payoff vector $\boldsymbol{x}$ must satisfy the following natural conditions:
   - $x_j \geqslant 0$ for all $j \in \mathcal{P}$;
   - $\sum_{r \in \mathsf{C}_j} x_r \leqslant \nu(\mathsf{C}_j)$ for any $j \in \{1, \ldots, m\}$

Now, any partition of agents into coalitions and any payoff vector that respects this partition corresponds to a *plausible* outcome of a characteristic function game. However, not all outcomes are equally desirable.
For instance, if all agents contribute equally to the value of a coalition, a payoff vector that allocates the entire payoff to just one of the agents is less appealing than the one that shares the profits equally among all agents.
Similarly, an outcome that push all agents to work together is (typically) preferable to an outcome that some of the agents want to deviate from.

More broadly, one can evaluate outcomes according to two criteria: **fairness** (i.e., how well each agent's payoff reflects his contribution), and **stability** (i.e., what are the incentives for the agents to stay in the coalition structure). These criteria give rise to two families of solution concepts having as most notable members the **Shapley Value** and the **Core** respectively.

Given its broad success for defining **variable importance** in supervised learning problems (see here, here, here, here, and also here, just to mention a few) in the following we will focus on the former, the *Shapley Value*, forged in the '50s by the one and only, sir Lloyd S. Shapley.

The Shapley value is a solution concept that is usually formulated with respect to the grand coalition: it defines a way of distributing the value $\nu(\mathcal{P})$ that could be obtained by the grand coalition, and it is based on the intuition that the payment that each agent receives should be *proportional to his contribution*.

Idea v1.0: a naïve implementation of this idea would be to pay each agent according to how much he increases the value of the coalition of all other players when he joins it, i.e., set the payoff of player j to $\nu(\mathcal{P}) - \nu(\mathcal{P} \backslash \{j\})$.
Problem: Under this payoff scheme the total payoff assigned to the agents may differ from the value of the grand coalition.

Idea v2.0: we can **fix an ordering** of the agents and pay each agent according to how much he contributes to the coalition formed by his predecessors in this ordering: *Agent 1* receives $\nu(\{1\})$, *Agent 2* receives $\nu(\{1,2\}) - \nu(\{1\})$, and so on.
It is easy to see that this payoff scheme distributes the value of the grand coalition among the agents.
Problem: agents that play symmetric roles in the game may receive different payoffs depending on their position in the order.

Shapley's insight: the dependence on the agents ordering can be eliminated by **averaging over all possible orderings** (or permutations) of the players.

Now, to formally define the Shapley value, we need some additional notation.

1. Fix a characteristic function game $G = (\mathcal{P}, \nu)$ and let $\Pi_{\mathcal{P}}$ be the set of all permutations of $\mathcal{P}$. Given a specific permutation $\pi \in \Pi_{\mathcal{P}}$, we denote by $S_\pi(j)$ the set of all predecessors of player j in π, i.e., we set $S_\pi(j) = \{r \in \mathcal{P} \text{ such that } \pi(r) < \pi(j)\}$. For example, if $\mathcal{P} = \{1, 2, 3\}$, then

   $$\Pi_{\mathcal{P}} = \big\{(1,2,3), (1,3,2), (2,1,3), (3,1,2), (3,2,1)\big\},$$

   moreover, if $\pi = (3,1,2)$, then $S_\pi(3) = \emptyset$, $S_\pi(1) = \{3\}$ and $S_\pi(2) = \{1,3\}$.
2. The **marginal contribution** of an agent j with respect to a permutation π in a game $G = (\mathcal{P}, \nu)$ is denoted by $\Delta_\pi^G(j)$ and measures how much player j increases the value of the coalition consisting of its predecessor in π. More specifically:

   $$\Delta_\pi^G(j) = \nu\big(S_\pi(j) \cup \{j\}\big) - \nu\big(S_\pi(j)\big). \tag{1}$$

We can now define the Shapley value $\psi^G(j)$ of player j: it is simply his **average marginal contribution**, where the average is taken over **all** permutations of the player-set (assumed to be all equally likely):

$$\psi^G(j) \overset{\text{def.}}{=} \mathbb{E}_\pi\big(\Delta_\pi^G(j)\big) = \frac{1}{p!}\sum_{\pi\in\Pi_{\mathcal{P}}} \Delta_\pi^G(j) \overset{(\heartsuit)}{=} \sum_{\mathsf{C}\,:\, j\notin\mathsf{C}} \frac{(\#\mathsf{C})!(p-1-\#\mathsf{C})!}{p!}\big(\nu(\mathsf{C}\cup\{j\})-\nu(\mathsf{C})\big). \tag{2}$$

The second version ($\heartsuit$) can be derived by noting that the marginal contributions $\Delta_\pi^G(j)$ are of the form $\nu(\mathsf{C}\cup\{j\})-\nu(\mathsf{C})$ where C is a coalition not containing j. Now, for how many orderings does one have $S_\pi(j)=\mathsf{C}$? There are $(\#\mathsf{C})!$ possible orderings of C and $(p-1-\#\mathsf{C})!$ orderings of $\mathcal{P}\backslash(\mathsf{C}\cup\{j\})$ $\rightsquigarrow$ another nice probabilistic interpretation (see page 307 here).

It can be shown that the Shapley value is in fact **the only** payoff division scheme that has **all** these four properties:

- *Efficiency*: $\sum_j \psi^G(j) = \nu(\mathcal{P})$.
- *Dummy player*: if player j is *dummy* (i.e., $\nu(\mathsf{C}\cup\{j\})=\nu(\mathsf{C})$ for *any* coalition C), then $\psi^G(j)=0$.
- *Symmetry*: if j_1 and j_2 are *symmetric* players (i.e., $\nu(\mathsf{C}\cup\{j_1\})=\nu(\mathsf{C}\cup\{j_2\})$ for *any* coalition C), then $\psi^G(j_1)=\psi^G(j_2)$.
- *Additivity*: given any two characteristic function games $G_1=(\mathcal{P},\nu_1)$ and $G_2=(\mathcal{P},\nu_2)$, and their *sum* $G^+=(\mathcal{P},\nu_1+\nu_2)$,

$$\psi^{G^+}(j) = \psi^{G_1}(j)+\psi^{G_2}(j) \quad \text{for all } \; j\in\mathcal{P}.$$

Example 2. Suppose that `Ciccio-Pharma` (C) is a small biotech company who discovered a new vaccine but, to manufacture and market it, needs to team up with a larger partner. Two candidates: `Aristo-Medical` (A) and `BruttiMaBuoni-Inc` (B). If A or B teams up with C, the big firm will split "*1 billion*" with C. Here is a *possible* characteristic function

$$\nu(A)=\nu(B)=\nu(C)=\nu(AB)=0 \qquad \text{and} \qquad \nu(AC)=\nu(BC)=\nu(ABC)=1.$$

Since there're only 3 players, to get Shapley's payoffs we can make a table indicating the value brought to a coalition by each player on the way to formation of the gran coalition:

Permutation	Player A	Player B	Player C
ABC	0	0	1
ACB	0	0	1
BAC	0	0	1
BCA	0	0	1
CAB	1	0	0
CBA	0	1	0
TOTAL VALUE	1	1	4

Since in the derivation of the Shapley value it is assumed that the $3!=6$ permutations/arrival sequences are all *equally likely*, the average value of each biotech company is simply:

$$\psi(A)=\tfrac{1}{6}, \quad \psi(B)=\tfrac{1}{6}, \quad \psi(C)=\tfrac{4}{6}=\tfrac{2}{3}.$$

So `Ciccio-Pharma`, the drug discoverer, will get two-thirds of the billion, and the big companies split the remaining third.

Example 3. (*Shapley in Simple Games*)
A game $G=(\mathcal{P},\nu)$ is said to be **simple** if it is *monotone* (i.e., $\nu(\mathsf{C}_1)\leqslant\nu(\mathsf{C}_2)$ if $\mathsf{C}_1\subseteq\mathsf{C}_2$) and its characteristic function only takes values 0 and 1. The game in Example 1 is clearly *simple* as soon as we rescale the payoffs so that $\nu(\mathcal{P})=1$.

In simple games, the Shapley (a.k.a. *Shapley–Shubik power index* in this context) have a particularly attractive interpretation: it measures the **power of a player**, i.e., the probability that she can influence the outcome of the game.

Indeed, the Shapley value of a player j in a simple game $G=(\mathcal{P},\nu)$ with $\#\mathcal{P}=p$ can be rewritten as follows:

$$\psi^G(j) = \{\text{proportion of permutations where } j \text{ is } \textbf{pivotal}\} = \frac{\#\big\{\pi\in\Pi_{\mathcal{P}} \text{ such that } \nu\big(S_\pi(j)\big)=0 \text{ and } \nu\big(S_\pi(j)\cup\{j\}\big)=1\big\}}{p!} \tag{3}$$

In other words, if agents join the coalition in a random order, $\psi^G(j)$ is exactly the **probability that player j turns a losing coalition into a winning one**.

Example 4. (*Shapley for Variable Importance*)
Suppose we are dealing with a generic (supervised) learning problem where we need to predict a response Y based on a bunch of covariates $\boldsymbol{X}=(X_1,\dots,X_p)$. The goal here is to measure the **importance** of X_j in this task.

We can cast this problem into a suitable characteristic function game having the p covariates as players $\mathcal{P}=\{1,\dots,p\}$, and some measure of fit as characteristic function ν. To be more specific, for any coalition (of covariates) C, let $\boldsymbol{X}_{\mathsf{C}}=(X_j : j\in\mathsf{C})$ and $\widehat{Y}(\mathsf{C})=\mathbb{E}(Y\,|\,\boldsymbol{X}_{\mathsf{C}})$, the ideal optimal predictor (under a squared risk) based on the covariates in $\boldsymbol{X}_{\mathsf{C}}$. Now, if we take

$$\nu(\mathsf{C}) = -\mathbb{E}\Big(Y-\widehat{Y}(\mathsf{C})\Big)^2 \overset{\text{then}}{\rightsquigarrow} \psi(j) = \frac{1}{p!}\sum_{\pi}\mathbb{E}\Big[\widehat{Y}\big(S_\pi(j)\big)-\widehat{Y}\big(S_\pi(j)\cup\{j\}\big)\Big]^2 \rightsquigarrow \text{population quantity to be estimated!} \tag{4}$$

This is just an averaged version (over all possible submodels) of LOCO, another, recently introduced var-importance measure. It is instructive to try to rephrase in this context the previous 4 properties that characterize the Shapley value.

Remark: this is all nice and good but, of course, Shapley for variable importance is not perfect. For example, it defines variable importance with respect to all submodels, but most of those submodels are not of interest and, in addition, it is strongly influenced by the correlation between covariates.

2. The Exercise: Let's get together and feel all right...

⇝ Your job ⇜

1. **Introductory**
To check your understanding of Shapley, let's start easy by playing around with a tiny game having only 3 players. So, imagine (!) there're three students curiously named: Antwohnette (A), BadellPadel (B) and Chumbawamba (C). Exactly one of them needs to be working (not necessarily all day long) on this HW to complete it. Here's their working-hours:

A	Antwohnette	14:00-17:00
B	BadellPadel	11:00-16:00
C	Chumbawamba	9:00-13:00

A coalition C is an agreement by one or more students as to the times they will be *really* working[2]. Its values will be given by: $\nu(\mathsf{C})=\{\#$ of hours potentially saved by well organized coalition$\}$

Clearly $\nu(A)=\nu(B)=\nu(C)=0$ and $\nu(ABC)=4$. **Complete the definition of ν and find Shapley** (see Example 2).

2. **Statistical**
Imagine you are a financial investor currently dealing with a **portfolio of p stocks** whose **returns** (see notes) are modeled by a set of random variables $\{X_1,\dots,X_p\}$. In this setup, it may be important to allocate to each asset of the portfolio its *contribution* to the **total utility** defined as $U_\omega\big(\sum_{j=1}^p X_j\big)$, where the **utility function** $U_\omega(\cdot)$ for us will be a linear combination of the portfolio average return and volatility (Markowitz'style!, see here), that is

$$U_\omega(X)=\mathbb{E}(X)-\omega\cdot\mathbb{V}\mathrm{ar}(X)\quad\text{for some weight}\quad \omega>0.$$

In a recent paper, the Authors tackle the problem from a cooperative game theory perspective by defining a suitable **variance game** over $\mathcal{P}=\{1,\dots,p\}$, whose Shapley value turns out to be very simple, intuitive (and familiar).
More specifically, let C be any coalition of the p assets and $R_{\mathsf{C}}=\sum_{j\in\mathsf{C}}X_j$ their return. Now, for any $\mathsf{C}\subset\mathcal{P}$, define

$$\nu(\mathsf{C})=U_\omega(R_{\mathsf{C}})=\mathbb{E}(R_{\mathsf{C}})-\omega\cdot\mathbb{V}\mathrm{ar}(R_{\mathsf{C}})=\nu_{\mathbb{E}}(\mathsf{C})-\omega\cdot\nu_{\mathbb{V}}(\mathsf{C}).$$

Hence, the game $G=(\mathcal{P},\nu)$ is a linear combination of two other games, $G_{\mathbb{E}}=(\mathcal{P},\nu_{\mathbb{E}})$ and $G_{\mathbb{V}}=(\mathcal{P},\nu_{\mathbb{V}})$ $\implies$ by the additivity of Shapley, $\psi^G(j)=\psi^{G_{\mathbb{E}}}(j)-\omega\cdot\psi^{G_{\mathbb{V}}}(j)$. By the linearity of expectation, the first term is trivially equal to $\psi^{G_{\mathbb{E}}}(j)=\mathbb{E}(X_j)$, whereas the second one, after elementary manipulations, turns out to be equal to $\psi^{G_{\mathbb{V}}}(j)=\mathbb{C}\mathrm{ov}(X_j,R_{\mathcal{P}})$.
In conclusion, once we pick a suitable weight $\omega>0$, the Shapley allocation to each stock $j\in\{1,\dots,p\}$ is given by

$$\psi(j)^G=\mathbb{E}(X_j)-\omega\cdot\mathbb{C}\mathrm{ov}(X_j,R_{\mathcal{P}})=\mathbb{E}(X_j)-\omega\cdot\sum_{r=1}^p\mathbb{C}\mathrm{ov}(X_j,X_r)\quad\rightsquigarrow\quad\text{a population quantity to be estimated!}\tag{5}$$

In other words, to learn the Shapley in this financial game, we need to study the marginal correlation graph among some standard measure of stock *relative performance* (see `Appendix (B)` in the notes). To this end, we may collect the *daily closing prices* for p stocks[3], selected within those consistently in the S&P500 index.
The stocks are categorized into 10 *Global Industry Classification Standard* (GICS) sectors, including `Consumer Discretionary`, `Energy`, `Financials`, `Materials`, `Health Care`, `Utilities`, `Industrials`, `Information Technology`, `Consumer Staples`, and `Telecommunications Services`. It is expected that stocks from the same GICS sectors should tend to be clustered together, since they (*allegedly*) tend to interact more with each other[4].
So, although cherry picking stocks to optimize our portfolio is not our goal here, ideally we want to collect few stocks from different GICS to boost the overall value/utility of the portfolio[5]. Here's the details:

- Select a sensibly sized[6] portfolio of p stocks and take data from the "*Covid-Age*": **January 1, 2020 till the end of 2023** (see notes). More specifically, if $c_{t,j}$ is the *closing price* of stock j on day t, let $x_{t,j}=\log(c_{t,j}/c_{t-1,j})$.
- Based on the data matrix $\mathbb{X}=\big[x_{t,j}\big]_{t,j}$ with *time* on the rows and *stocks* on the columns, we want to estimate the Pearson-based correlation graph over stocks (i.e. each node in the graph is a stock), by simply treating the instances $\{x_{t,j}\}_t$ as independent samples even though they are **not** (they clearly form a time series!).
In particular we place an (undirected) edge between stock j_1 and stock j_2 only if their (estimated) Pearson correlation is, statistically speaking, *large enough*.
More specifically, for any chosen threshold $\tau>0$, we place an edge between j_1 and j_2 if $C_n^{j_1,j_2}(\alpha)\cap[-\tau,+\tau]=\emptyset$, where $C_n^{j_1,j_2}(\alpha)$ denotes the usual asymptotic Normal-based confidence interval for the Pearson-correlation.
You choose the threhold τ (**explain** your choice) and the level α of the many CI's ⇝ beware of multiplicity!

[2] This is an example of cooperative game theory applied to a resource allocation problem.
[3] The number of stocks p will affect the computational time and the statistical performance, hence, as usual, choose it wisely!
[4] I wrote "allegedly" since this is a hypothesis we should verify empirically instead of taking it for granted.
[5] *Diversification* in investing is key, somebody once said.
HINT: keep this in mind when talking about *esemble tecniques* like *random forest* to tackle predictive problems.
[6] Large but not too large, depending on your laptop.

- *Artistically* visualize the resulting graph (see here, and here), and draw some conclusion based on the decisions made along the way (i.e. the number and sector of the stocks, the threshold τ, the level α, etc.).

- Finally, always based on the data matrix $\mathbb{X}$, provide and possibly visualize (bootstrapped) confidence intervals for the Shapley values of the p stocks in your portfolio. As before, you choose the weight(s) ω, the level α, and the interval type. Comment the results based on the choice you made.
  Remark: as done before, also here make the simplifying assumption that, for each stock, we are dealing with independent replicates.
