Relative importance measures based on regression models performance metrics' allocation

[1]EDF R&D - *Département PRISME*
[2]Institut de Mathématiques de Toulouse
[3]SINCLAIR AI Lab

52èmes Journées de la Statistique de la SFdS
8 Juin 2021

Marouane Il Idrissi[1,2,3]
Bertrand Iooss[1,2,3]
Vincent Chabridon[1,3]

Introduction

Goal: Assess the importance of covariates in a regression model by allocating a share of model performance to each one.

Cooperative games provide a framework for producing relevant allocation schemes, **even when the covariates are correlated**.

In particular, *Shapley values* are widely used in ML interpretability... But are they **always** suitable?

Introduction

Goal: Assess the importance of covariates in a regression model by allocating a share of model performance to each one.

Cooperative games provide a framework for producing relevant allocation schemes, **even when the covariates are correlated**.

In particular, *Shapley values* are widely used in ML interpretability... But are they **always** suitable?

Challenges:

- How to formalize what *importance* means?
- How can cooperative games be used to produce relevant **importance measures**?
- How to assess whether an allocation scheme is **more suitable** than another one?

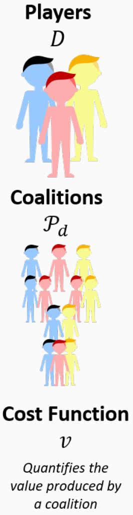
Cooperative games and allocations

Let:

- $D = \{1, \ldots, d\}$ be a finite and countable set of players ;
- $\mathcal{P}_d$ be the set of all coalitions of players ;
- $v : \mathcal{P}_d \to \mathbb{R}$ a cost function.

A **cooperative game** is formally defined by the couple (D, v).

Main question : How can one allocate shares of $v(D)$ to each players ?



Cooperative games and allocations

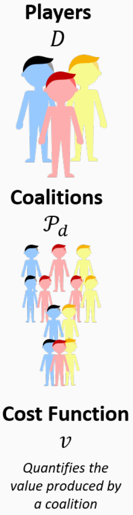
Let:

- $D = \{1, \dots, d\}$ be a finite and countable set of players ;
- $\mathcal{P}_d$ be the set of all coalitions of players ;
- $v : \mathcal{P}_d \to \mathbb{R}$ a cost function.

A **cooperative game** is formally defined by the couple (D, v).

Main question : How can one allocate shares of $v(D)$ to each players ?

An **allocation** is a function $\phi : D \to \mathbb{R}$ which allocates the quantity ϕ_i for every player $i \in D$.



Statistical cooperative games

Let $X \in \mathbb{R}^d$ be the covariates of $Y = \Theta(X, \beta)$, a parametric "nestable" regression model:

- $\Theta(D) \stackrel{\text{def}}{=} \Theta(X, \beta)$ is the "full" model ;
- $\forall S \in \mathcal{P}_d, \Theta(S) \stackrel{\text{def}}{=} \Theta(X_S, \beta_S)$ be a "nested" model.

Let $\mu_\Theta : \mathcal{P}_d \to \mathbb{R}^+$ be a **performance metric** (e.g., R^2, likelihood...). For $S \in \mathcal{P}_d$, $\mu_\Theta(S)$ denotes the performance of the nested model $\Theta(S)$.

A **statistical cooperative game** (Feldman 2005) is the cooperative game defined by (D, μ_Θ).

Covariates

D

$\{X_1, \dots, X_d\}$

Coalitions

$\mathcal{P}_d$

$$\left\{ \begin{matrix} \{X_1\}, \{X_2\}, \dots, \\ \{X_1, X_2\}, \{X_1, X_3\}, \dots, \\ \{X_1, \dots, X_d\} \end{matrix} \right\}$$

Cost Function

$\mu_\Theta(S)$

Performance of the nested model

$\Theta(X_S, \beta_S)$

Statistical cooperative games

Let $X \in \mathbb{R}^d$ be the covariates of $Y = \Theta(X, \beta)$, a parametric "nestable" regression model:

- $\Theta(D) \stackrel{\text{def}}{=} \Theta(X, \beta)$ is the "full" model ;
- $\forall S \in \mathcal{P}_d, \Theta(S) \stackrel{\text{def}}{=} \Theta(X_S, \beta_S)$ be a "nested" model.

Let $\mu_\Theta : \mathcal{P}_d \to \mathbb{R}^+$ be a **performance metric** (e.g., R^2, likelihood...). For $S \in \mathcal{P}_d$, $\mu_\Theta(S)$ denotes the performance of the nested model $\Theta(S)$.

A **statistical cooperative game** (Feldman 2005) is the cooperative game defined by (D, μ_Θ).

How can one define an allocation for (D, μ_Θ)?

Covariates

D

$\{X_1, \dots, X_d\}$

Coalitions

$\mathcal{P}_d$

$$\left\{ \begin{matrix} \{X_1\}, \{X_2\}, \dots, \\ \{X_1, X_2\}, \{X_1, X_3\}, \dots, \\ \{X_1, \dots, X_d\} \end{matrix} \right\}$$

Cost Function

$\mu_\Theta(S)$

Performance of the nested model

$\Theta(X_S, \beta_S)$

Random order model allocations

One way of defining allocations is through **random order models** (Weber 1988).

Idea: Consider that players are combined in **random orders** (i.e., permutations of D), and every player is granted its marginal contribution to the **previous set of players**, weighted by **how likely the order is**.

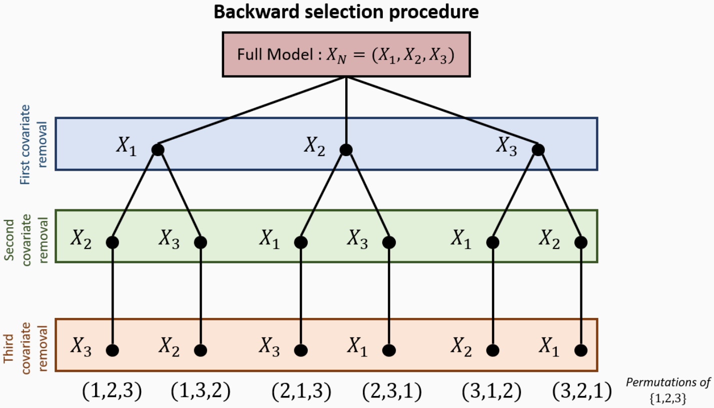
Let:

- $r = (r_1, \ldots, r_d) \in \mathcal{R}(D)$ be a permutation in the set of permutations of D;
- $r(j)$ be the position of the player j in the permutation r (i.e., $r_{r(j)} = j$);
- p be a *probability mass function* (pmf) defined on $\mathcal{R}(D)$.

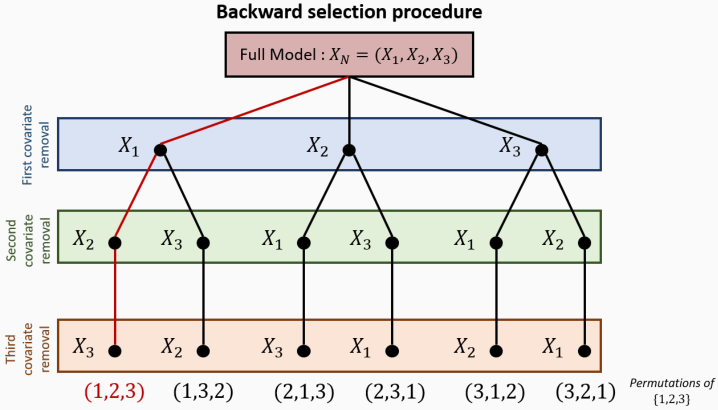
The **random order allocation** is given by, $\forall i \in D$:

$$\begin{aligned}
\phi_i &= \mathbb{E}_p\Big[\mu_\Theta\left(D \setminus \{r_1, \ldots, r_{r(i)-1}\}\right) - \mu_\Theta\left(D \setminus \{r_1, \ldots, r_{r(i)}\}\right)\Big] \\
&= \sum_{r \in \mathcal{R}(D)} p(r)\Big(\mu_\Theta\left(D \setminus \{r_1, \ldots, r_{r(i)-1}\}\right) - \mu_\Theta\left(D \setminus \{r_1, \ldots, r_{r(i)}\}\right)\Big)
\end{aligned}$$

Random order model allocations



Random order model allocations

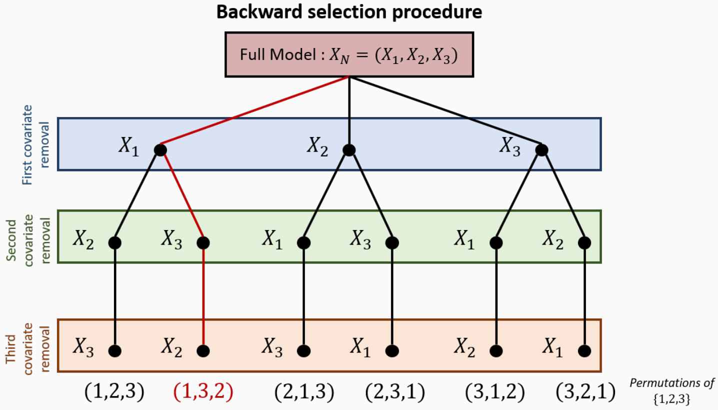


$$\phi_1 = p((1,2,3))(\mu_\Theta(D) - \mu_\Theta(D \setminus \{1\})) + \cdots$$

$$\phi_2 = p((1,2,3))(\mu_\Theta(D \setminus \{1\}) - \mu_\Theta(D \setminus \{1,2\})) + \cdots$$

$$\phi_3 = p((1,2,3))(\mu_\Theta(D \setminus \{1,2\}) - \mu_\Theta(\emptyset)) + \cdots$$

Random order model allocations

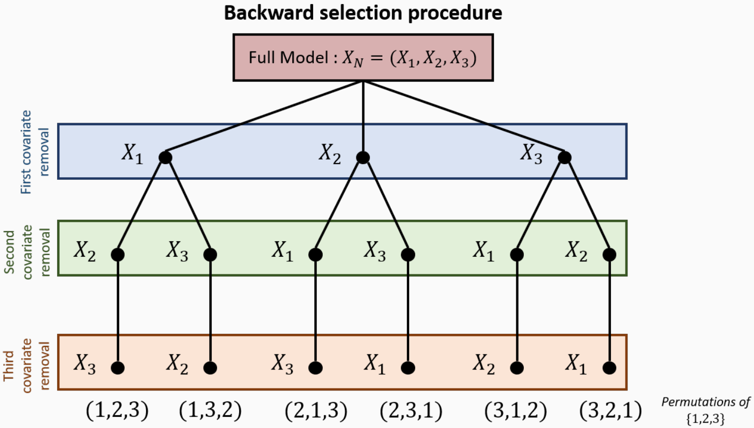


$$\phi_1 = p\big((1,2,3)\big)\big(\mu_\Theta(D) - \mu_\Theta(D \setminus \{1\})\big) + p\big((1,3,2)\big)\big(\mu_\Theta(D) - \mu_\Theta(D \setminus \{1\})\big) + \cdots$$

$$\phi_2 = p\big((1,2,3)\big)\big(\mu_\Theta(D \setminus \{1\}) - \mu_\Theta(D \setminus \{1,2\})\big) + p\big((1,3,2)\big)\big(\mu_\Theta(D \setminus \{1,3\}) - \mu_\Theta(\emptyset)\big) + \cdots$$

$$\phi_3 = p\big((1,2,3)\big)\big(\mu_\Theta(D \setminus \{1,2\}) - \mu_\Theta(\emptyset)\big) + p\big((1,3,2)\big)\big(\mu_\Theta(D \setminus \{1\}) - \mu_\Theta(D \setminus \{1,3\})\big) + \ \ldots$$

Random order model allocations



$$\phi_i = \sum_{r \in \mathcal{R}(D)} p(r) \left(\mu_\Theta(D \backslash \{r_1, \ldots r_{r(i)-1}\}) - \mu_\Theta(D \backslash \{r_1, \ldots r_{r(i)}\}) \right)$$

Random order model allocations

If we assume that μ_Θ is *weakly monotonic* (i.e., a "smaller" model cannot outperform a "bigger" model), then, for all pmf p defined on $\mathcal{R}(D)$, the random order model allocation ϕ is:

- **Efficient:** $\sum_{i \in D} \phi_i = \mu_\Theta(D)$;
- **Non-negative:** $\forall i \in D, \phi_i \geq 0$;

and thus provides a **decomposition** of the full model's performance ($\mu_\Theta(D)$) according to **each covariate**.

Random order model allocations

If we assume that μ_Θ is *weakly monotonic* (i.e., a “smaller” model cannot outperform a “bigger” model), then, for all pmf p defined on $\mathcal{R}(D)$, the random order model allocation ϕ is:

- **Efficient:** $\sum_{i \in D} \phi_i = \mu_\Theta(D)$;
- **Non-negative:** $\forall i \in D, \phi_i \geq 0$;

and thus provides a **decomposition** of the full model’s performance ($\mu_\Theta(D)$) according to **each covariate**.

But how can one define a suitable probability mass function p?

Relative importance

Relative importance can be modeled as a complete and transitive binary relation $\prec$ on D.

A **relative importance measure** is then a representation (utility function) ϕ of this binary relation, such that:

$$i \preceq j \iff \phi_i \leq \phi_j.$$

Relative importance can be modeled as a complete and transitive binary relation $\prec$ on D.

A **relative importance measure** is then a representation (utility function) ϕ of this binary relation, such that:

$$i \preceq j \iff \phi_i \leq \phi_j .$$

Four admissibility criteria (Cox 1985; Johnson and Lebreton 2004; Feldman 2005; Grömping 2007) can be defined:

- **Non-negativity:** $\forall i \in D, \phi(i) \geq 0$;
- **Proper exclusion:** If, for $\Theta(D)$, one has $\beta_i = 0$, then $\phi_i = 0$;
- **Proper inclusion:** If, for $\Theta(D)$, $\beta_i \neq 0$, then $\phi_i > 0$;
- **Efficiency/Total contribution:** $\sum_{i=1}^{d} \phi_i = \mu_\Theta(D)$.

Shapley and proportional values

Shapley values (Shapley 1951)

$$p(r) = \frac{1}{d!}$$

Shapley values are a maximum entropy/uniform prior choice.

The Shapley values of (D, μ_Θ) violate the proper exclusion criterion .

Proportional values (Ortmann 2000)

$p(r)$ is defined axiomatically, with **relative importance** in mind:
If $r_1 \preceq r_2 \preceq \cdots \preceq r_d$, then $p(r)$ should be of high probability.

The proportional values of (D, μ_Θ) respect all four admissibility criteria.

Shapley and proportional values

Shapley values (Shapley 1951)

$$p(r) = \frac{1}{d!}$$

Shapley values are a maximum entropy/uniform prior choice.

The Shapley values of (D, μ_Θ) violate the proper exclusion criterion .

Proportional values (Ortmann 2000)

$p(r)$ is defined axiomatically, with **relative importance** in mind:
If $r_1 \preceq r_2 \preceq \cdots \preceq r_d$, then $p(r)$ should be of high probability.

The proportional values of (D, μ_Θ) respect all four admissibility criteria.

But how do they handle covariate correlation?

Relative importance measure for linear models - Analytical results

In the context of linear models, with Gaussian covariates, i.e.:

$$Y = \sum_{i=1}^{d} \beta_i X_i, \quad X \sim \mathcal{N}_d(0, \Sigma),$$

and the **coefficient of determination** R^2 as a performance metric, one has:

- Shapley values **grant importance to correlated exogenous variables** (Shapley's Joke). Proportional values **do not**;
- Shapley values **distribute correlation effects equally** among all the variables. Proportional values favor covariates with high β;
- Shapley values and proportional values **are equal** when:
  - The covariates have the same β value ;
  - The covariates are independent.

Algerian Forest Fires

Algerian Forest Fires (Abid and Izeboudjen 2020) dataset: 244 observations of 8 covariates in two regions (Bejaia et Sidi Bel-Abbes).

Goal: Predict the occurence of forest fires using a logistic regression.
Challenge: The covariates are **highly correlated**.

We consider the statistical cooperative game (N, R^2) where R^2 denotes the **generalized coefficient of determination** given by:

$$R^2(S) = 1 - \frac{\mathrm{Dev}(S)}{\mathrm{Dev}(\emptyset)}$$

where $\mathrm{Dev}(S)$ denotes the deviance of the nested model $\Theta(S)$ and $\mathrm{Dev}(\emptyset)$ the deviance of the null model (i.e., only considering an intercept).

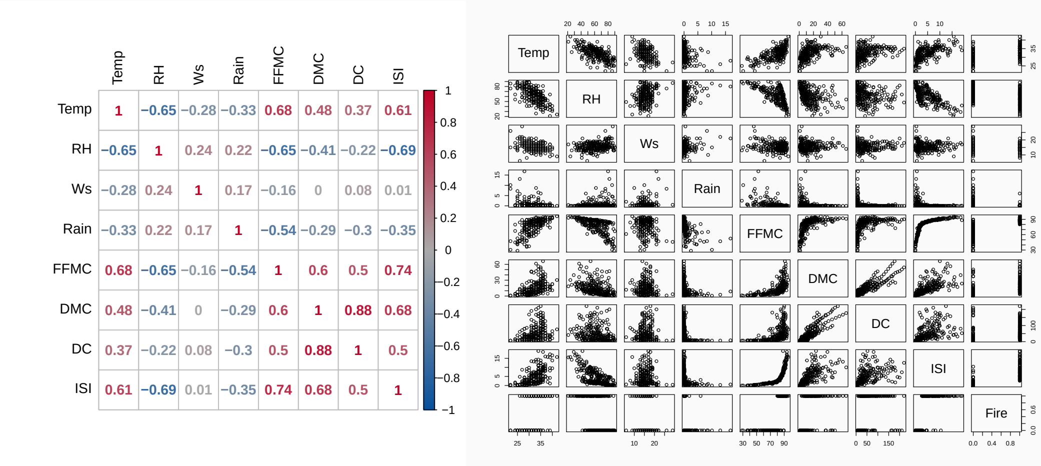


Figure 1: Correlation matrix (left) and scatterplot (right) of the *Algerian Forest Fires* dataset.

Algerian Forest Fires

Estimated full model performance: $R^2 \simeq 0.803$ and $Q^2 \simeq 0.79$ (predictivity coefficient).

Covariates	Temp	RH	Ws	Rain	FFMC	DMC	DC	ISI	Total
VIF	1.36	1.90	1.72	1.44	7.08	8.04	6.24	5.04	-
Sh (%)	4.5	3.7	0.4	5.5	33.3	6.2	3.2	23.5	80.3
PMD (%)	0.4	0	0	0.7	69.7	6.4	0	3.1	80.3

Table 1: Multicollinearity and relative importance measures for the *Algerian Forest Fires* dataset.

Contributions and perspectives

Contributions:

- Better understanding of the use of cooperative games on importance measures;
- Extension to the case of (penalized) logistic regressions;
- Illustration on a public dataset;
- Application to an industrial EDF use-case (fission products release in the primary circuit of a PWR) (Remy et al. 2018) ;
- Efficient and parallel implementation available in the open source R package available from the CRAN website: `sensitivity` (functions `lmg()` for Shapley values and `emvd()` for proportional values).

Perspectives:

- Extension to other types of models;
- Exploration of other performance metrics;
- Development of new context-based allocations for meaningful interpretations.

Abid, F., and N. Izeboudjen. 2020. "Predicting Forest Fire in Algeria Using Data Mining Techniques: Case Study of the Decision Tree Algorithm." In *Advanced Intelligent Systems for Sustainable Development,* 363–370. Springer International Publishing. ISBN: 978-3-030-36674-2.

Cox, Louis A. 1985. "A New Measure of Attributable Risk for Public Health Applications" [in English]. *Management Science* 31, no. 7 (July): 800–813. ISSN: 0025-1909, 1526-5501. https://doi.org/10.1287/mnsc.31.7.800. http://pubsonline.informs.org/doi/abs/10.1287/mnsc.31.7.800.

Feldman, B. E. 2005. "Relative Importance and Value" [in English]. *SSRN Electronic Journal,* ISSN: 1556-5068. https://doi.org/10.2139/ssrn.2255827. http://www.ssrn.com/abstract=2255827.

Grömping, U. 2007. "Estimators of Relative Importance in Linear Regression Based on Variance Decomposition" [in English]. *The American Statistician* 61, no. 2 (May): 139–147. ISSN: 0003-1305, 1537-2731. https://doi.org/10.1198/000313007X188252. http://www.tandfonline.com/doi/abs/10.1198/000313007X188252.

Johnson, J. W., and J. M. Lebreton. 2004. "History and Use of Relative Importance Indices in Organizational Research" [in English]. *Organizational Research Methods* 7, no. 3 (July): 238–257. ISSN: 1094-4281, 1552-7425. https://doi.org/10.1177/1094428104266510. http://journals.sagepub.com/doi/10.1177/1094428104266510.

Lindeman, R. H., P. F. Merenda, and R. Z. Gold. 1980. *Introduction to Bivariate and Multivariate Analysis* [in English]. Scott, Foresman. ISBN: 978-0-673-15099-8. https://books.google.cz/books?id=-hfvAAAAMAAJ.

Ortmann, K. M. 2000. "The proportional value for positive cooperative games" [in English]. *Mathematical Methods of Operations Research (ZOR)* 51 (2): 235–248. ISSN: 1432-2994, 1432-5217, accessed December 4, 2020. https://doi.org/10.1007/s001860050086. http://link.springer.com/10.1007/s001860050086.

Remy, E., E. Dautrême, C. Talon, Y. Dirat, and C. Dinse Le Strat. 2018. "Comparison of machine learning algorithms on data from the nuclear industry." In *Safety and Reliability – Safe Societies in a Changing World.* Num Pages: 8. CRC Press. ISBN: 978-1-351-17466-4.

Shapley, L. S. 1951. *Notes on the n-Person Game – II: The Value of an n-Person Game* [in English]. Research Memorandum ATI 210720. Santa Monica, California: RAND Corporation, August. https://www.rand.org/content/dam/rand/pubs/research_memoranda/2008/RM670.pdf.

Weber, R. J. 1988. "Probabilistic values for games" [in English]. In *The Shapley Value,* 1st ed., 101–120. Cambridge University Press, October. ISBN: 978-0-521-36177-4 978-0-521-02133-3 978-0-511-52844-6. https://doi.org/10.1017/CBO9780511528446.008. https://www.cambridge.org/core/product/identifier/CBO9780511528446A014/type/book_part.

Thank you for attending!

Any questions?

Relative importance measure for linear models

In the context of linear models, with Gaussian covariates, i.e.:

$$Y = \sum_{i=1}^{d} \beta_i X_i, \quad X \sim \mathcal{N}_d\left(0, \Sigma\right),$$

and for the statistical cooperative game (D, R^2), the Shapley values are known as the LMG indices (Lindeman, Merenda, and Gold 1980), and the proportional values as the PMVD indices (Grömping 2007). When $d = 2$, and for $\mathrm{Cov}(X_1, X_2) = \sigma_1\sigma_2\rho$, one has that:

Shapley values of (N, R^2):

$$\mathrm{LMG}_1 = \frac{1}{\mathbb{V}(Y)}\left(\beta_1^2\sigma_1^2 + \beta_1\beta_2\sigma_1\sigma_2\rho + \frac{\rho^2}{2}(\beta_2^2\sigma_2^2 - \beta_1^2\sigma_1^2)\right)$$

$$\mathrm{LMG}_2 = \frac{1}{\mathbb{V}(Y)}\left(\beta_2^2\sigma_2^2 + \beta_1\beta_2\sigma_1\sigma_2\rho + \frac{\rho^2}{2}(\beta_1^2\sigma_1^2 - \beta_2^2\sigma_2^2)\right)$$

Proportional values of (N, R^2):

$$\mathrm{PMVD}_1 = \frac{\beta_1^2\sigma_1^2}{\beta_1^2\sigma_1^2 + \beta_2^2\sigma_2^2}$$

$$\mathrm{PMVD}_2 = \frac{\beta_2^2\sigma_2^2}{\beta_1^2\sigma_1^2 + \beta_2^2\sigma_2^2}$$

When $d = 3$, $\mathrm{Cov}(X_2, X_3) = \sigma_2\sigma_3\rho$, and $\beta_1 = \beta_2 = 1$:

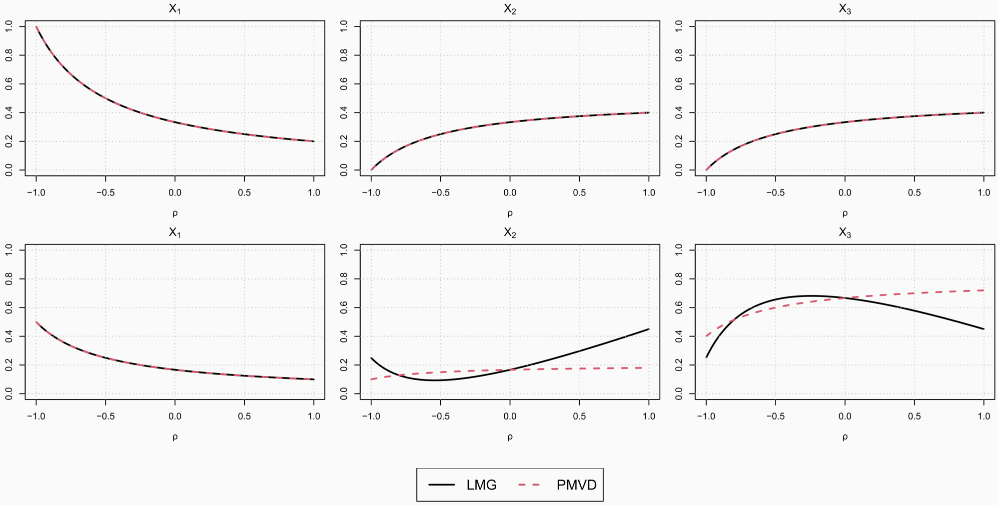


Figure 2: LMG and PMVD with respect to ρ, when $\beta_3 = 1$ (top row), and $\beta_3 = 2$ (bottom row).

Random order models

Let $i \in N$ be a player, in a game (N, v), the marginal contribution of i, denoted by $w(\{i\})$, and sometimes called the *worth* of i, is defined by:

$$w(\{i\}) = v(N) - v(N \setminus \{i\}). \tag{1}$$

The concept of marginal contribution can be extended to coalitions of players (i.e., subsets of players). Let $S \in \mathcal{P}_n$ be a coalition of players, and let its worth be defined by:

$$w(S) = v(N) - v(N \setminus S). \tag{2}$$

Let $r = (r_1, \ldots, r_n) \in \mathcal{R}(N)$ be a specific ordering/permutation of N. Let S_k^r be the set of the k first players in the order r. One can then define the *individual positional marginal contribution*, which represents the marginal contribution of a player in the i^{th} position in an order r to the set of players that precedes him in the same ordering, noted $M_i(r)$, with the convention that $S_0^r = \emptyset$:

$$\begin{aligned} M_i(r) &= w(S_i^r) - w(S_{i-1}^r) && (3) \\ &= v(N \setminus S_{i-1}^r) - v(N \setminus S_i^r). && (4) \end{aligned}$$

Let $r(j)$ be the position of the player j in the ordering r. Let $p(r)$ be a probability mass function over the set of permutations of $\mathcal{R}(N)$. Let $\Delta_{n!}$ be the set of all probability mass functions over $\mathcal{R}(N)$. A random order model allocation can then be defined as, for $p \in \Delta_{n!}$, $i = 1 \ldots, n$ and for a game (N, v):

$$\begin{aligned} \phi_i &= \mathbb{E}_p\left[M_{r(i)}(r)\right] && (5) \\ &= \sum_{r \in \mathcal{R}(N)} p(r) M_{r(i)}(r). && (6) \end{aligned}$$

Axiomatic definition of PMD

Axiom (Anonymity)

Let r and r^* be two different permutations in $\mathcal{R}(N)$. If $\mathrm{MC}(r^*) = \mathrm{MC}(r)$ then $L(r^*) = L(r)$.

Axiom (Limit Proper Exclusion)

Let w be defined by a model Θ and performance measure μ where $\beta_i^* = 0$. Consider a sequence of games w_k, where $\beta_j^k = \beta_j^*$ for $j \neq i$. Assume that $\beta_i^1 > 0$ and $\beta_i^k \to 0$. Then:

$$\lim_{k \to \infty} PMD_i(w_k) = 0. \tag{7}$$

Axiom (Equal proportional effect)

Let $r \in \mathcal{R}(N)$, and let $S \in r$. Then:

$$\left| \frac{\partial \ln L(r)}{\partial \ln w(S)} \right| = 1. \tag{8}$$

PMD identification

- From the Limit Proper Exclusion, one has that $\frac{\partial \ln L(r)}{\partial \ln w(S)} < 0$.
- Since $\frac{\partial X}{\partial Y} = \frac{\partial X}{\partial \ln Y} \times \frac{1}{Y}$, this leads, to $-\frac{\partial \ln L(r)}{\partial w(S)} = \frac{1}{w(S)}$.

In turn, these observations lead to:

$$-\ln L(r) = c_r + \sum_{S \in r} \int_0^{w(S)} \frac{1}{x} dx = c_r + \sum_{S \in r} \ln w(S). \tag{9}$$

where c_r is a multiplicative factor dependent of r. However, the anonymity axiom requires that c_r should be constant for all $r \in \mathcal{R}(N)$, and appears both in the numerator and denominator of $p(r)$. One could subsequently assume that $c_r = 0$. This leads to the unique identification of $L(r), \forall r \in \mathcal{R}(N)$ as being:

$$L(r) = \left(\prod_{S \in r} w(S) \right)^{-1} \tag{10}$$

Axiomatic definition of the Shapley Values

Let (N, v) be a cooperative game. The unique allocation of (N, v) respecting the following set of axioms:

1. (Efficiency) $\sum_{j=1}^{d} \phi_j = \mathrm{val}(\{1 \ldots, d\})$, meaning that the sum of the allocated values have to be equal to the value produced by the cooperation of all the players ;
2. (Symmetry) If $\mathrm{val}(A \cup \{i\}) = \mathrm{val}(A \cup \{j\})$ for all $A \in \mathcal{P}_d$, then $\phi_i = \phi_j$, meaning that if two players allow for the same contribution to every coalition, their attribution should be the same ;
3. (Dummy) If $\mathrm{val}(A \cup \{i\}) = \mathrm{val}(A)$ for all $A \in \mathcal{P}_d$, then $\phi_i = 0$, meaning that if a player does not contribute the the production of resources for all coalition, he should not be attributed any resources ;
4. (Additivity) If val and val' have Shapley Values ϕ and ϕ' respectively, then the game with cost function $\mathrm{val} + \mathrm{val}'$ has Shapley values $\phi_j + \phi_j'$ for $j \in \{1, \ldots, d\}$;

is the Shapley value, defined by:

$$\mathrm{Sh}_i = \frac{1}{n!} \sum_{r \in \mathcal{R}(N)} M_{r(i)}(r) \tag{11}$$

$$= \frac{1}{n} \sum_{A \subset N \setminus \{i\}} \binom{n-1}{|A|}^{-1} \big(v(A \cup \{i\}) - v(A)\big). \tag{12}$$