Variance reduction for estimation of Shapley effects and adaptation to unknown input distribution

François Bachoc

Institut de Mathématiques de Toulouse
Université Paul Sabatier
France

From the PhD thesis of Baptiste Broto
Also a joint work with Marine Depecker

SIAM UQ
April 2022

Framework and notation

- Random input vector $\boldsymbol{X} = (X_1, \dots, X_p)$ on space $\mathcal{X} = \mathcal{X}_1 \times \dots \times \mathcal{X}_p \subset \mathbb{R}^p$.
- $X_1, \dots, X_p$ are dependent.
- Deterministic black box function $f : \mathcal{X} \to \mathbb{R}$.
- Stochastic output

$$Y = f(\boldsymbol{X}).$$

- For $u \subseteq \{1, \dots, p\}$, let
  - $\boldsymbol{X}_u = (X_i, i \in u)$,
  - $-u = \{1, \dots, p\} \backslash u$,
  - $|u|$ be the cardinality of u.

Shapley effects

Conditional elements

For $u \subseteq \{1, \dots, p\}$, let

$$\mathrm{VE}_u = \mathrm{Var}\left(\mathbb{E}\left(Y|\boldsymbol{X}_u\right)\right)$$

and

$$\mathrm{EV}_u = \mathbb{E}\left(\mathrm{Var}\left(Y|\boldsymbol{X}_{-u}\right)\right).$$

Large VE_u or $\mathrm{EV}_u \Longrightarrow \boldsymbol{X}_u$ is important.

Shapley effects [Shapley, 1953, Owen, 2014, Iooss and Prieur, 2019]

For $i \in \{1, \dots, p\}$, the Shapley effect η_i is

$$\eta_i = \frac{1}{p\mathrm{Var}(Y)} \sum_{u \subset -\{i\}} \binom{p-1}{|u|}^{-1} \left(W_{u \cup \{i\}} - W_u\right),$$

with $W_u = \mathrm{VE}_u$ or $W_u = \mathrm{EV}_u$.

- $0 \leq \eta_i \leq 1$.
- $\sum_{i=1}^{p} \eta_i = 1$.
- Easy interpretation as percentage of importance even with dependent inputs.
- Even if f does not depend on variable i, we can have $\eta_i > 0$ if $\boldsymbol{X}_i$ is correlated with $\boldsymbol{X}_j$ and f depends on variable j.

Non-given data framework and double Monte Carlo

Non-given data framework

- We can choose any $\boldsymbol{x} \in \mathcal{X}$ and compute $f(\boldsymbol{x})$.
- We can sample from the conditional distributions of $\boldsymbol{X}$.

Estimation of EV_u by double Monte Carlo

Estimator

$$\widehat{\mathrm{EV}}_u = \frac{1}{N_u} \sum_{n=1}^{N_u} \frac{1}{N_I - 1} \sum_{k=1}^{N_I} \left(f\left(\boldsymbol{X}_{-u}^{(n)}, \boldsymbol{X}_u^{(n,k)}\right) - \overline{f\left(\boldsymbol{X}_{-u}^{(n)}\right)} \right)^2 .$$

- $\boldsymbol{X}_{-u}^{(1)}, \dots, \boldsymbol{X}_{-u}^{(N_u)}$ iid with law of $\boldsymbol{X}_{-u}$.
- $\boldsymbol{X}_u^{(n,1)}, \dots, \boldsymbol{X}_u^{(n,N_I)}$ iid with law of $\boldsymbol{X}_u$ conditionally to $\boldsymbol{X}_{-u}^{(n)}$.
- $\overline{f\left(\boldsymbol{X}_{-u}^{(n)}\right)}$ is the average of $f\left(\boldsymbol{X}_{-u}^{(n)}, \boldsymbol{X}_u^{(n,1)}\right), \dots, f\left(\boldsymbol{X}_{-u}^{(n)}, \boldsymbol{X}_u^{(n,N_I)}\right)$.

$\Longrightarrow$ Unbiased.
$\Longrightarrow$ We take $N_I = 3$ as in [Song et al., 2016].
$\Longrightarrow$ N_u is the budget/accuracy parameter.

Pick-freeze : an estimator of VE_u (not detailed explicitly here) [Janon et al., 2014].

Subset aggregation of estimators of conditional elements

- For $u \subseteq \{1, \ldots, p\}$, consider the estimator $\widehat{\mathrm{EV}}_u$ of EV_u.
- Then subset aggregation simply means summing over all subsets :

$$\widehat{\eta}_i = \frac{1}{p\widehat{\mathrm{Var}}(Y)} \sum_{u \subset -\{i\}} \binom{p-1}{|u|}^{-1} \left(\widehat{\mathrm{EV}}_{u \cup \{i\}} - \widehat{\mathrm{EV}}_u \right).$$

Question : For each $u \subseteq \{1, \ldots, p\}$, which budget (number of samples N_u) to allocate to $\widehat{\mathrm{EV}}_u$?

Contribution : optimal budget allocation

The optimal choice of $(N_u)_u$ subject to $\sum_{\substack{u \subseteq \{1,\ldots,p\} \\ u \neq \varnothing, \{1,\ldots,p\}}} N_u = N_{\mathrm{tot}}$ is

$$N_u^* \propto \sqrt{(p - |u|)!|u|!(p - |u| - 1)!(|u| - 1)!\mathrm{Var}(\widehat{\mathrm{EV}}_u^{(1)})},$$

with $\widehat{\mathrm{EV}}_u^{(1)}$ computed with budget $N_u = 1$.
$\Longrightarrow$ Issue : $\mathrm{Var}(\widehat{\mathrm{EV}}_u^{(1)})$ unknown.
$\Longrightarrow$ For practice (with some approximations) we take

$$N_u^* \propto \mathrm{Round}\left(\binom{p}{|u|}^{-1} \right).$$

Prospect : For large p, have $N_u^* = 0$ for many u to be computationally scalable ?

Random-permutation aggregation

- For a permutation σ on $\{1, \dots, p\}$ and $i \in \{1, \dots, p\}$, we let $P_i(\sigma) = \{\sigma(j);\ j = 1, \dots, \sigma^{-1}(i) - 1\}$. We have [Song et al., 2016]

$$\eta_i = \frac{1}{p!\mathrm{Var}(Y)} \sum_{\substack{\text{permutations} \\ \sigma}} (\mathrm{EV}_{P_i(\sigma) \cup \{i\}} - \mathrm{EV}_{P_i(\sigma)}).$$

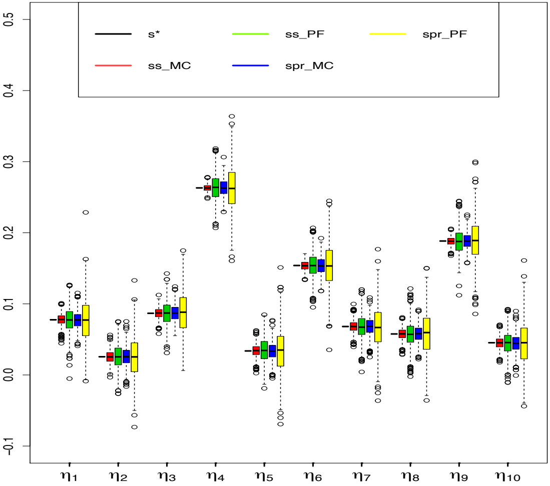
- Hence the random-permutation aggregation :

$$\widehat{\eta}_i = \frac{1}{\widehat{\mathrm{Var}}(Y)} \frac{1}{M} \sum_{j=1}^{M} (\widehat{\mathrm{EV}}_{P_i(\sigma_j) \cup \{i\}} - \widehat{\mathrm{EV}}_{P_i(\sigma_j)}),$$

  with $\sigma_1, \dots, \sigma_M$ iid random permutations (uniform).
- [Song et al., 2016] suggests to take the budget for each $\widehat{\mathrm{EV}}_u$ minimal ($N_u = 1$) and to take M maximal.

A Gaussian linear example.

- $s^\star$: true value.
- *ss* : subset aggregation.
- *spr* : random permutation.
- *MC* : double Monte Carlo for $\widehat{\mathrm{EV}}_u$.
- *PF* : pick-freeze for $\widehat{\mathrm{VE}}_u$.



1 Shapley effects

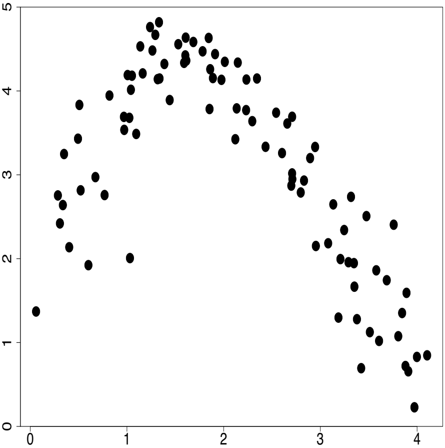
2 Improving estimation in non-given data framework

3 Given data framework with nearest-neighbors

Given data framework

- Only a data set of the input variables is available : $\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(n)}$ iid with the distribution of $\boldsymbol{X}$.
- The corresponding $Y_1 = f(\boldsymbol{X}^{(1)}), \ldots, Y_n = f(\boldsymbol{X}^{(n)})$ are available
  OR
  f can be called at any input $\boldsymbol{x}$.
- The exact conditional sampling needed by the previous estimators is not available.
- We will mimick this conditional sampling using nearest neighbors.

Nearest-neighbor approximation of conditional distributions



Estimator of EV_u with nearest-neighbors

- Recall
$$\mathrm{EV}_u = \mathbb{E}\left(\mathrm{Var}\left(Y|\boldsymbol{X}_{-u}\right)\right).$$
- Let $k_n^{-u}(i,\ell)$ be the ℓ-th nearest neighbor of $\boldsymbol{X}_{-u}^{(i)}$ among $\boldsymbol{X}_{-u}^{(1)}, \dots, \boldsymbol{X}_{-u}^{(n)}$.
- The estimator is

$$\begin{aligned}
\widehat{\mathrm{EV}}_u &= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{N_I - 1}\sum_{\ell=1}^{N_I}\left(f\left(\boldsymbol{X}_{-u}^{(k_n^{-u}(i,\ell))}, \boldsymbol{X}_{u}^{(k_n^{-u}(i,\ell))}\right) - \overline{f\left(\boldsymbol{X}_{-u}^{(i)}\right)}\right)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{N_I - 1}\sum_{\ell=1}^{N_I}\left(Y_{k_n^{-u}(i,\ell)} - \overline{Y_i}\right)^2
\end{aligned}$$

with

$$\overline{f\left(\boldsymbol{X}_{-u}^{(i)}\right)} = \text{ average}\left(f\left(\boldsymbol{X}_{-u}^{(k_n^{-u}(i,1))}, \boldsymbol{X}_{u}^{(k_n^{-u}(i,1))}\right), \dots, f\left(\boldsymbol{X}_{-u}^{(k_n^{-u}(i,N_I))}, \boldsymbol{X}_{u}^{(k_n^{-u}(i,N_I))}\right)\right)$$

$$\overline{Y_i} = \text{ average}\left(Y_{k_n^{-u}(i,1)}, \dots, Y_{k_n^{-u}(i,N_I)}\right).$$

- Other variants available when we can call $f(\boldsymbol{x})$ at new $\boldsymbol{x}$.
- This time, good choice of N_I is more open. Not necessarily $N_I = 3$.
- Similarities with rank methods for $|-u| = 1$ [Gamboa et al., 2020].

Rate of convergence

- Consider a fixed $u \subseteq \{1, \ldots, p\}$.

Condition

The function f is C^1, $\mathcal{X}$ is compact in $\mathbb{R}^p$, and $\boldsymbol{X}$ has a density $f_{\boldsymbol{X}}$ with respect to Lebesgue measure that is lower and upper bounded and Lipschitz continuous.

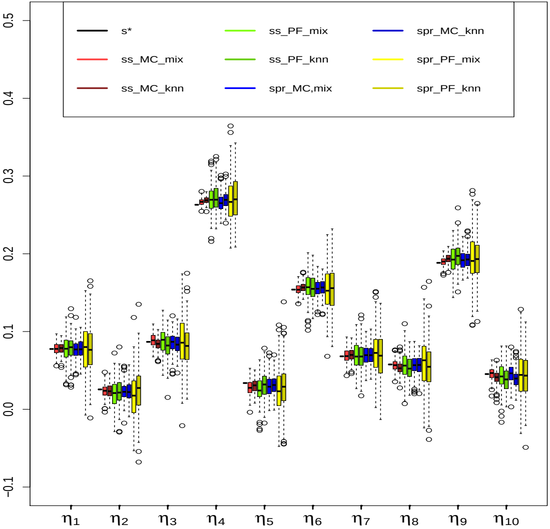
Contribution : rate of convergence

We have, for each $\delta > 0$,

$$\left| \widehat{\mathrm{EV}}_u - \mathrm{EV}_u \right| = o_p \left(\frac{1}{n^{\frac{1}{2(p-|u|)} - \delta}} \right).$$

- $p - |u| = |-u|$ is the dimension of the nearest-neighbor approximation $\Longrightarrow$ curse of dimensionality.
- When $p - |u| = 1$, essentially we reach parametric rate $n^{-1/2} \Longrightarrow$ optimal.
- For $p - |u| > 1$, is the rate optimal ?

- $s^{\star}$: true value.
- *ss* : subset aggregation.
- *spr* : random permutation.
- *MC* : double Monte Carlo for $\widehat{\mathrm{EV}}_u$.
- *PF* : pick-freeze for $\widehat{\mathrm{VE}}_u$.
- *knn* : nearest neighbor estimation.
- *mix* : a variant with new calls to f.



Conclusion

Conclusion :

- $\eta_1, \ldots, \eta_p$ are percentages of importance for dependent inputs.
- Estimation by aggregating $\widehat{\mathrm{EV}}_u$, $u \subseteq \{1, \ldots, p\}$.
- Can be beneficial to tune budgets N_u, $u \subseteq \{1, \ldots, p\}$.
- Nearest neighbors to mimick conditional distributions.

The paper :

B. Broto, F. Bachoc and M. Depecker, Variance reduction for estimation of Shapley effects and adaptation to unknown input distribution, *SIAM/ASA Journal on Uncertainty Quantification*,8(2) : 693–716, 2020

The R implementation : Function `shapleySubsetMC` in R package `sensitivity`.

Some follow-up works : [Broto et al., 2022, Demange-Chryst et al., 2022].

Many other ongoing work and prospects :

- Large dimension p, link with Hilbert-Schmidt Independence Criterion, interactions with machine learning...
- [Da Veiga, 2021, Ghorbani and Zou, 2019],...

Thank you for your attention !

Broto, B., Bachoc, F., Clouvel, L., and Martinez, J.-M. (2022).
Block-diagonal covariance estimation and application to the Shapley effects in sensitivity analysis.
SIAM/ASA Journal on Uncertainty Quantification, 10(1) :379–403.

Da Veiga, S. (2021).
Kernel-based anova decomposition and Shapley effects–application to global sensitivity analysis.
arXiv preprint arXiv :2101.05487.

Demange-Chryst, J., Bachoc, F., and Morio, J. (2022).
Shapley effect estimation in reliability-oriented sensitivity analysis with correlated inputs by importance sampling.
arXiv preprint arXiv :2202.12679.

Gamboa, F., Gremaud, P., Klein, T., and Lagnoux, A. (2020).
Global sensitivity analysis : a new generation of mighty estimators based on rank statistics.
arXiv preprint arXiv :2003.01772.

Ghorbani, A. and Zou, J. (2019).
Data Shapley : Equitable valuation of data for machine learning.
In *International Conference on Machine Learning*, pages 2242–2251. PMLR.

Iooss, B. and Prieur, C. (2019).
Shapley effects for sensitivity analysis with dependent inputs : comparisons with Sobol' indices, numerical estimation and applications.
International Journal for Uncertainty Quantification, 9(5).

Janon, A., Klein, T., Lagnoux, A., Nodet, M., and Prieur, C. (2014).
Asymptotic normality and efficiency of two Sobol index estimators.
ESAIM : Probability and Statistics, 18 :342–364.

Owen, A. B. (2014).
Sobol' Indices and Shapley Value.
SIAM/ASA Journal on Uncertainty Quantification, 2(1) :245–251.

Shapley, L. S. (1953).
A value for n-person games. Contribution to the Theory of Games.
Annals of Mathematics Studies, 2 :28.

Song, E., Nelson, B. L., and Staum, J. (2016).
Shapley effects for global sensitivity analysis : theory and computation.
SIAM/ASA Journal on Uncertainty Quantification, 4(1) :1060–1083.