Sensitivity analysis for model risk management

Alexandre Boumezoued
Léo Tondolo

In the context of increasing operational complexity within the calculation chain, e.g., related to balance sheet calculation and financial reporting, it has become critical to assess efficiently the sensitivity of the output results to individual assumptions and inputs involved.

The purpose of this paper is to provide the methodological basis of sensitivity analysis and to demonstrate an example of the differentiated behaviour of Sobol and Shapley indices, as well as the performance of the Quasi-Monte Carlo (QMC) estimator in that context.

Introduction

Who has never confused two figures when copying a number? Who has never entered an overdrawn number by typing the wrong key on his keyboard? Who has never added the wrong attachment when sending an email or made a mistake when importing data? These situations are all familiar to us daily. Of course, today, a large part of these processes is automated, and in this context it can be even more challenging to detect potential errors.

Whether or not we are aware of errors, they are a risk that we must take into consideration and that we qualify as operational errors. Furthermore, it is possible to be aware of some errors, but how to be sure they have all been found? This paper details methods for knowing in advance the assumptions that will have the greatest impact on the final model outcome and thus to know what to prioritise in checking.

Model risk management

Before describing the different fields of model risk management, we need to give a definition of a model and explain what model risk is. A model is a process that links one or more input variables to one or more output variables. Mathematically or algorithmically, this process is called a function. This function may itself depend on several other functions in the case of a more or less complex model. Nowadays, insurers are developing increasingly complex models in order to represent their business and their risks as realistically as possible. Thus, they rely on production lines that combine different models and types of data.

Model risk is the uncertainty in the output variables of the model due to various upstream flaws. The role of model risk management is to identify, quantify and resolve potential defects throughout the production chain. Flaws can be divided into four fields:

- **Data quality:** The uncertainty in the input variables of the model. The error contained in each variable may come from an incorrect entry or can be the result of a flaw in another model. Uncertainty also resides in whether or not the data itself shows an underlying bias or the assumptions of the chosen model have been fulfilled.
- **Model's suitability:** The uncertainty contained in the model. It may be a lack of theoretical consistency, a mismatch with the reality or even more a mistake in implementation of the model.
- **Model's instability:** Here, we try to determine whether the model is not oversensitive to certain predefined assumptions.
- **Model's interpretation:** In this field, we consider the error caused by a misinterpretation or a misuse of the results. It is also the understanding of the interaction and impact of all the variables in the model on the output.

How can sensitivity analysis address model risk management?

In this paper, we study the impact of the structuring assumptions of a given model. The main purpose is to determine which parameters have the most influence on our output variable. This is the result of several findings. Increasingly large production lines imply longer processes with increasing computational times, from a few hours to even several days. Another point of attention is the fact that we are faced with a race for the best possible modelling. In this context, it is important to understand well what is at stake for each of the model's structuring assumptions and to determine the input variables that will have the greatest impact on the output.

Through sensitivity analysis, we can address the issue at stake. Indeed, sensitivity analysis can be seen as the study that seeks to determine how the uncertainty in the output variable can be distributed among the different sources of uncertainty in the input variables. Therefore, the idea is to measure the variability inherent to the output variable subject to the variation of each input variable.

Sensitivity analysis can address each component of model risk management:

- **Data quality:** Sensitivity analysis allows us to exhibit the most material input. Therefore, it orientates on those inputs that have to be free of material errors to avoid any major impact.
- **Model's suitability:** Through the understanding of the sensitivity indices, as we describe later, the modeller is informed about the key variables that are retained in the modelling exercise.
- **Model's instability:** By nature, sensitivity analysis provides the measure of the degree of response of the model to a change of any assumption.
- **Model's interpretation:** Sensitivity analysis involves the use of indices, e.g., Shapley, that are in fact widespread in the field of model interpretability (see Delcaillau et al., 2021).

In order to use sensitivity analysis to put the "black box" models behind us, it is necessary to be able to simulate the model a large number of times or alternatively to rely on approximate closed forms. In the following, we first present the Sobol and Shapley indices, then detail two methods of sensitivity analysis based on each strategy (simulations vs. closed forms).

A deep dive into sensitivity analysis

Among the range of sensitivity analysis methods, we were interested in methods based on variance decomposition. We describe below two well-known indices in sensitivity analysis: Sobol and Shapley.

- **Sobol indices** were introduced in 1993. The first-order Sobol index measure the share of variability of the model output as a function of the variability of an input variable. We denote by $X = (X_1, \cdots, X_n) \in \mathbb{R}^n$ the random vector which represents the set of input variables and by $Y = f(X)$ the model output. Under these notations, the first-order Sobol index of the input variable X_i is given by:

  $$S_i = \frac{Var\left[\mathbb{E}[Y|X_i]\right]}{Var[Y]}.$$

  In other words, the index represents the share of uncertainty due to input X_i, because $\mathbb{E}[Y|X_i]$ corresponds to the most plausible value of the output, given outcome X_i.

- **Shapley indices** were defined in 1953 in the field of cooperative game theory. They have multiple uses nowadays. In actuarial science, for example, they can be used for capital allocation purposes. In data science, the Shapley Additive Explanations (SHAP) package provides insights related to model interpretability. Concerning sensitivity analysis, it is Owen in 2014 who uses them first. If we consider the same assumptions as above and denote by $-\{i\}$ the set of elements of $\{1, \cdots, n\}$ not containing i, then the Shapley index of an input variable X_i is defined as

  $$Sh_i = \frac{1}{n} \sum_{S \subseteq -\{i\}} \left[\frac{Var\left[\mathbb{E}[Y|X_{S\cup\{i\}}]\right]}{Var[Y]} - \frac{Var\left[\mathbb{E}[Y|X_S]\right]}{Var[Y]} \right].$$

  where X_S denotes the information $\{X_k, k \in S\}$ for any set S.

  In other words, the index represents the average uncertainty share of model $Y = f(X)$ associated with the additional knowledge of X_i.

In a general framework, these two sensitivity indices do not give the same values. However, in the very simplified case where the function is linear and the input variables are assumed independent, then both indices are equal. This property does not hold when the function moves away from linearity, as is usually the case in practice. Examples include cash flow model response to asset and liability assumptions, economic scenario generator response to economic variables and parameters or the calculations involved in reporting, such as risk aggregation to calculate Solvency Capital Requirement (SCR) under Solvency II.

When the input variables are dependent, Sobol and Shapley indices differ and the basic interpretation of Sobol indices as percentages of sensitivity (summation of all indices to 1) is lost while Shapley indices allow us to preserve this property. Indeed, the sum of Sobol indices is only equal to 1 when the input variables are independent whereas the sum of Shapley indices is always equal to 1, whether the input variables are dependent or not.

In a general context, we will therefore tend to favour the use of the Shapley over the Sobol index, because the former provides the relevant interpretation related to sensitivity to input parameters and data, even if input variables are assumed to be dependent. In the context of sensitivity analysis, correlation is a modelling feature that can be used to more properly reflect interactions between errors in inputs. Indeed, in practice, it is known that errors may occur simultaneously due to a common source or cause (for example, in the case where the input values depend on common user actions or when an operational error impacts a common set of inputs). In particular, positive dependence can be used to model the fact that the cause leads to a simultaneous increase or decrease of the inputs. In the following part, we will study the different estimators of Sobol and Shapley indices via a numerical example, while highlighting the impact of dependence on the analysis.

Estimators for sensitivity indices

The purpose is to introduce estimators of the indices presented in the last part. Two approaches will be presented.

SIMULATION APPROACHES

The first approach is based on simulations. As its name suggests, we will estimate the sensitivity indices with a method based on *Monte Carlo* simulations.

- **Sobol index estimation:** Here, we start from the mathematical definition of the Sobol index to derive the estimator. The estimation is performed in two steps, one for the variance and the other for the variance of the conditional expectation:

  Step 1: Estimation of $Var[Y]$:

  $$\widehat{Var}[Y] = \frac{1}{N-1}\sum_{k=1}^{N}\left[Y_k - \left(\frac{1}{N}\sum_{j=1}^{N} Y_j\right)\right]^2.$$

  Step 2: Estimation of $Var\big[\mathbb{E}[Y|X_i]\big] = \mathbb{E}[(\mathbb{E}[Y|X_i] - \mathbb{E}[Y])^2]$:

  $$\widehat{Var}\big[\mathbb{E}[Y|X_i]\big] = \frac{1}{N-1}\sum_{k=1}^{N}\left[\widehat{\mathbb{E}}[Y|X_i = x_{i_k}] - \left(\frac{1}{N}\sum_{j=1}^{N} Y_j\right)\right]^2,$$

  where $\widehat{\mathbb{E}}\big[Y\big|X_i = x_{i_k}\big]$ is estimated by *Monte Carlo* (with *M* simulations). This relies on the ability to simulate the distribution of the vector $(X_1, \dots, X_{i-1}, X_{i+1}, \dots X_n)$, which can be achieved using Gaussian conditioning formulas (see Tondolo [2019]).

  However, this simulation approach remains computationally intensive, because *N x M* vectors of *n* random variables have to be simulated, with N outer scenarios called primary simulations and M inner simulations called secondary simulations.

  Some attempts to accelerate this estimation in the context of linear functions have been proposed in the literature, see, e.g., Kucherenko et al. (2012). By rewriting the Sobol index, they proposed a formula involving a single *Monte Carlo* loop that allows us to avoid the nested nature of the estimator. However, for each simulation, it is required to simulate two independent copies of the random vector of input variables. Our work has led us to conclude that this will ultimately result in increased computing time for nonlinear functions due to a lower convergence in this context.

- **Shapley index estimation:** In order to estimate the Shapley indices, one possible option would be to proceed as before. However, in that case, the estimator would not converge in a polynomial time. Instead, we can consider a method introduced by Song et al. (2016), which is based on the reformulation of the Shapley index calculation by considering all the permutations of the input variables. Once again, if we consider all the $n!$ permutations, in the context of a large dimension, it becomes impossible to estimate the indices. This is why Song et al. (2016) propose an improvement to reduce the cost of calculation. This is performed by generating only a limited set of permutations and arranging the order of the calculations to obtain an estimator with faster convergence. The Shapley's indices estimation algorithm that accelerates calculations and converges in polynomial time is detailed in Tondolo (2019).

CLOSED-FORM APPROACHES

The so-called closed-form approach presented below is based on an approximation of the variance known as the Delta method. The virtue of such an approach is that it avoids performing intensive simulations to estimate the sensitivity indices.

Within the general framework of the previous notations and considering a smooth function f, we are able to express $Var[Y]$ as a function of $\mu = (\mathbb{E}[X_1], \cdots, \mathbb{E}[X_n])$ and $\sum$ the variance-covariance matrix of X. The Delta method gives us the following approximation:

$$Var[Y] = Var[f(X)] \approx \nabla f(\mu)^t \textstyle\sum \nabla f(\mu),$$

where the gradient of the function f is expressed as:

$$\nabla f(x_1, \cdots, x_n) = \left(\frac{\partial f}{\partial x_1}, \cdots, \frac{\partial f}{\partial x_n}\right).$$

In the case where X is one-dimensional, we obtain the following based on the representation $X = \mu + \sigma Z$ and using an expansion in the volatility parameter:

$$\begin{aligned} &Var[Y] = Var[f(X)] \\ &\approx Var[f(\mu) + \sigma Z f'(\mu)] \\ &= \sigma^2\big(f'(\mu)\big)^2. \end{aligned}$$

The Delta method also allows us to derive an approximation of the variance of the conditional expectation $Var(\mathbb{E}[\, f(X) \mid X_i \,])$. Indeed, we can use this second-order approximation to first demonstrate the following approximation:

$$Var\big[\mathbb{E}[f(\boldsymbol{X})|X_i]\big] \approx Var[f_i(X_i)],$$

where f_i corresponds to the function f taken as the vector of conditional expectation with respect to the input variable X_i. Furthermore, conditional expectation being a function of the conditioned variable, f_i is a function of the single variable X_i. The one-dimensional approximation of the Delta method leads us to the following approximation:

$$Var\big[\mathbb{E}[f(X)|X_i]\big] \approx \sigma_i^2[f_i'(\mu_i)]^2.$$

Thus, we have a closed-form formula for the Sobol index by considering the ratio of the approximations.

Remark: Because Shapley indices can be seen as a sum of differences between two Sobol indices, it is possible to use the above approximation, but this requires computing a large number of approximations, which becomes complex in large dimensions.

The table in Figure 1 provides a summary of when to use one approach or the other, depending on our constraints.

FIGURE 1: COMPARISON OF APPROACHES

	Advantages	Disadvantages
Simulation approaches	• Implementable in large dimensions and regardless the function used	• Calculation time
Closed form approaches	• Calculation time	• Needs to modify the calculations with each new model • Cannot be implamanted in large dimensions and with complex model

Before moving on to numerical applications and comparisons, let us look at one last theoretical point. This concerns the way to simulate random numbers for *Monte Carlo* estimators.

Generation of random numbers

We have seen that we estimate the sensitivity coefficients using a *Monte Carlo* (MC) method. Here, we will discuss how to reduce the estimation error of this approximation by modifying the generation of random numbers.

Commonly, when we talk about the *Monte Carlo* (MC) method, we imply a simulation of random numbers with the uniform law on $[0, 1]$ that we call pseudo-random. In this case, we know that the estimation error is $O\left(\frac{1}{\sqrt{N}}\right)$, where N represents the number of simulations. Thus, we must generate a large number of simulations to obtain an estimator of quality. In many cases, a unique call to the function f can last several hours, which prevents the generation of many random paths. Consequently, the *Monte Carlo* estimation will suffer from a large estimation error. One way to reduce the estimation error by keeping the number of simulations N constant is to make sure that we cover the interval $[0, 1]$ as “best” as possible, as described in the following. Therefore, we turn to another method of random number generation.

The second method is based on low-discrepancy sequences. The low-discrepancy sequences are built more to minimise discrepancy than to construct more evenly generated sequences. We denote them by quasi-randomly generated random numbers and the underlying method of estimation is called *Quasi-Monte Carlo* (QMC). Several low-discrepancy sequences exist. In the sequel, we will use a popular one often applied in practice in finance, which is the Sobol sequence. In his article, Sobol (1967) provides a theorem that ensures a lower estimation error by using quasi-random number generation with the Sobol sequence rather than using a pseudo-random method with a uniform law. However, one should be cautious as the Sobol sequence is known to be very efficient for small dimensions, but it decreases in quality when the dimension increases.

Let us now look at a graphical comparison between the two methods of random number generation mentioned above. In Figure 2, we have represented the generation of the 10,000 points of the Sobol sequence in dimension 2. In Figure 3, the generation is done using the uniform law.

FIGURE 2: QUASI-RANDOM GENERATION: SOBOL’S SEQUENCE

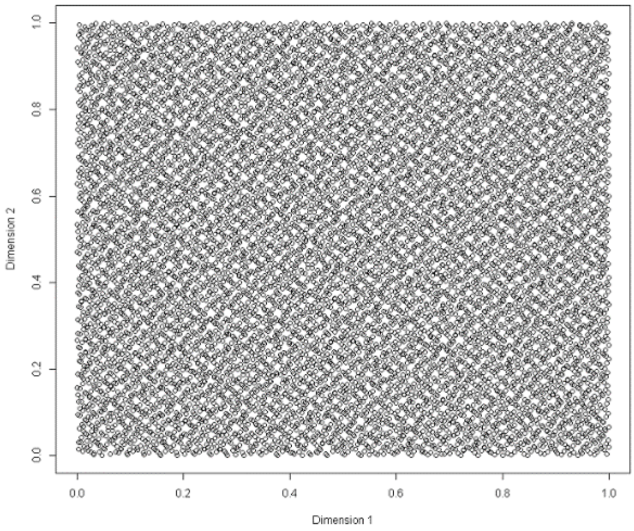
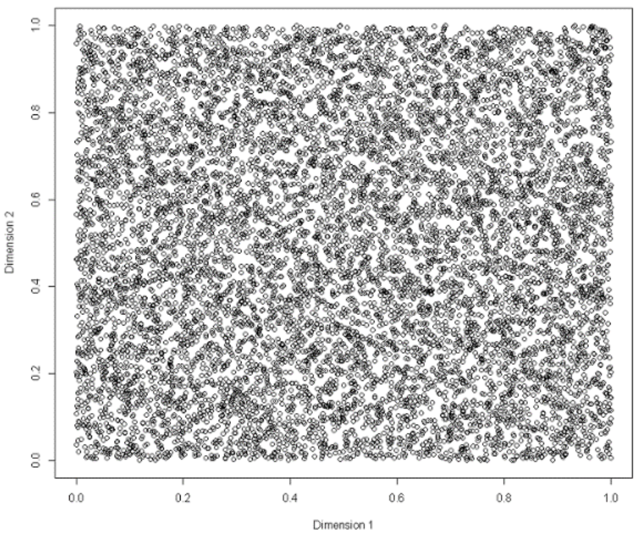


FIGURE 3: PSEUDO-RANDOM GENERATION: UNIFORM LAW



We note that the quality of the distribution is not satisfactory when the numbers are generated pseudo-randomly (Figure 3). Indeed, we observe some areas containing many points while others are completely empty. Conversely, if we focus on those generated quasi-randomly with the Sobol sequence (Figure 2), they seem very well distributed on the interval $[0, 1]$.

We will thus use quasi-random number generation for our following numerical applications.

Case study

We will now focus on numerical results and graphical comparisons to compare the different sensitivity indices and their estimators.

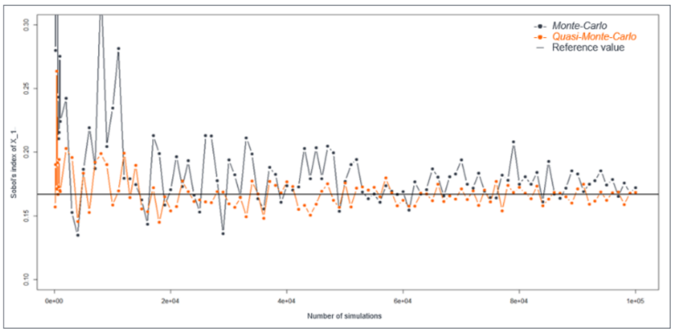
In terms of methodology, we must first define the law of the random vector of the input variable and its structure of dependency. In our case, we consider it to be a Gaussian vector. In a second step, we have to simulate a large number of random vectors with the previous setting. The sensitivity coefficients are then obtained using the estimators presented above.

As a first step, let us look at a numerical comparison of random number generation methods for the linear function:

$$f(X_1, \dots, X_n) = X_1 + \dots + X_n. \quad (1)$$

As a reference of comparison, we have considered the exact value obtained by closed-form approach.

FIGURE 4: MONTE CARLO VS. QUASI-MONTE CARLO



The graph in Figure 4 confirms what we found concerning the comparison on the quality of the distribution between the MC and QMC methods. The two methods seem to converge towards the exact value of the Sobol index. However, we note that the method with pseudo-random generation (black curve) fluctuates even with many simulations and slightly overestimates the Sobol index of X_1. On the other hand, the QMC method seems to oscillate in a reduced range around the exact value. This is true for a not too high number of simulations, which saves computing time in order to have good precision. Due to the good properties of the QMC random sequence, we consider this setting in the following application.

Let us now consider the Solvency II risk aggregation formula:

$$f(X_1, \dots, X_n) = \sqrt{\sum_{1 \le i,j \le n} \rho_{i,j} \,.\, X_i \,.\, X_j}\,, \quad (2)$$

where $\rho_{i,j}$ is the correlation coefficient allowing the aggregation of economic capitals for the set of n risks.

For the numerical application, we will set the values below for the reference market, default, life, non-life and health SCRs.

$$\mu = \begin{pmatrix} SCR_{Market} \\ SCR_{Default} \\ SCR_{Life} \\ SCR_{Non-Life} \\ SCR_{Health} \end{pmatrix} = \begin{pmatrix} 35 \\ 14 \\ 27 \\ 21 \\ 3 \end{pmatrix}. \quad (3)$$

For Sobol indices, we used 1,000 primary (N) and 1,000 secondary (M) simulations. In order to consider the same number of points generated for Shapley's indices, we used the $n! = 120$ permutations of the five input variables as well as 100 primary and secondary simulations. The case of dependency includes a 50% correlation for health and life SCRs and health and non-life SCRs. This is to illustrate that, due to the similarity in the way calculations are produced for those risks (tools, underlying data, methods etc.), it is likely that some errors may be common to the health and life/non-life values. Results are presented with a confidence interval obtained by a nonparametric bootstrap method. Confidence intervals are calculated on $B = 1000$ simulations and at the level of confidence $\alpha = 95\%$. The results are depicted in Figure 5.

FIGURE 5: PSEUDO-RANDOM GENERATION: UNIFORM LAW

Comparison of Sobol and Shapley indices - *BSCR*					
		Sobol index	CI Sobol	Shapley index	CC Shapley
Independence	*SCR Market*	*0.6050*	*[0.6030, 0.6070]*	*0.6040*	*[0.6040, 0.6040]*
	SCR Default	*0.0612*	*[0.0609, 0.0613]*	*0.0611*	*[0.0610, 0.0612]*
	SCR Life	*0.2190*	*[0.2190, 0.2200]*	*0.2190*	*[0.2190, 0.2190]*
	SCR Non-Life	*0.1120*	*[0.1120, 0.1130]*	*0.1140*	*[0.1140, 0.1140]*
	SCR Health	*0.0008*	*[0.0008, 0.0008]*	*0.0013*	*[0.0013, 0.0013]*
Dependence	*SCR Market*	*0.5720*	*[0.5720, 0.5760]*	*0.5940*	*[0.5940, 0.5940]*
	SCR Default	*0.0564*	*[0.0562, 0.0564]*	*0.0589*	*[0.0588, 0.0590]*
	SCR Life	*0.2200*	*[0.2200, 0.2210]*	*0.1720*	*[0.1710, 0.1720]*
	SCR Non-Life	*0.1150*	*[0.1150, 0.1150]*	*0.0823*	*[0.0820, 0.0826]*
	SCR Health	*0.1840*	*[0.1840, 0.1850]*	*0.0936*	*[0.0928, 0.0935]*

In a first step, we illustrate on the toy independent case the main theoretical properties that have been previously described. In our example, we indeed observe that in the independent case between the input variables (SCR per risk), the ordering between the risks in terms of Sobol or Shapley index is preserved compared to the values of the SCRs themselves; see values (3) above. As expected, in this case the values of the Sobol and Shapley indices differ, although the difference appears negligible, which can be due to the fact that the function (2) is in practice not "too far" from linearity (1). From an interpretation standpoint, both indices then quantify the percentage of variance explained by each input SCR; as an example, the market SCR represents 35% of the total sum of SCR per risk but explains 60% of the diversified SCR variation.

We now turn to the main example in this study that considers dependence between input variables, interpreted as potential common underlying errors in the calculation chain to produce those inputs. In this setting, we see that the difference between Sobol and Shapley indices increases. As already discussed, Sobol indices do not add up to 1, therefore they can no longer be interpreted conveniently as percentages, making particularly challenging any comparison with the independent case. Looking at Shapley indices for which the interpretation in terms of percentage of impact on overall uncertainty is relevant, it appears that the ranking is changed. Indeed, although the overall contribution of life, non-life and health SCRs input remains roughly stable between the independent and dependent configurations, the dependent case shows a strong increase in the Shapley index for the health SCR input. In terms of value, the health SCR is negligible (3% of total SCR values). However, when dependence is introduced, any variation (interpreted as an error) in health SCR calculation is likely to lead to a simultaneous error in the life or non-life SCR calculations, leading in turn to a share in overall variation of 10%. This provides evidence of how even negligible inputs in terms of value can in fact represent a reasonable share of impact on the output, when dependencies are in play. This also shows how sensitivity testing with dependence can allow us to identify the critical inputs in the calculation process, going beyond classical materiality arguments.

Such an approach has the potential for a significant range of applications, including as examples balance sheet valuation, Standard Formula and Internal Models processes within Solvency II, International Financial Reporting Standard (IFRS) 17 and financial reporting in general. This approach could also be included directly within tools and software that make up those calculation chains, as well as within any risk management procedure for a better monitoring of model risk and its components, from data quality to model stability, suitability and interpretation.

References

- Delcaillau, D., Ly, A., Papp, A., & Vermet, F. 2021. Model Transparency and Interpretability: Survey and Application to the Insurance Industry. https://arxiv.org/pdf/2007.12919.
- Lee, R. D., & Carter, L. R. 1992. Modeling and Forecasting U. S. Mortality. Journal of the American Statistical Association, 87(419), 659.
- Kucherenko, S., Tarantola, S., and Annoni, P. 2012. Estimation of global sensitivity indices for models with dependent variables. *Computer Physics Communications*, 183(4) :937–946.
- 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. *Contributions to the Theory of Games*, 2(28) :307–317.
- Sobol, I. M. 1967. On the distribution of points in a cube and the approximate evaluation of integrals. *Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki*, 7(4) :784–802.
- Sobol, I.M. 1993. Sensitivity estimates for nonlinear mathematical models. Mortality. *Mathematical modelling and computational experiments*, 1(4) :407–414.
- 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.
- Tondolo, L. 2019. Quantification de l'impact des erreurs opérationnelles sur les variables d'entrée d'un outil de calcul et de reporting Solvabilité 2 : approche par analyse de sensibilité. Actuarial thesis for the French Actuarial Institute. Available at http://www.ressources-actuarielles.net/C12574E200674F5B/0/D9E87CF116F7E59DC1258485006AAD2C.
- Ver Hoef, J. M. (2012). Who invented the delta method?. *The American Statistician*, 66(2), 124-127.

Milliman

CONTACT

Alexandre Boumezoued
alexandre.boumezoued@milliman.com

Léo Tondolo
leo.tondolo@milliman.com