SUMMER SCHOOL IN STATISTICS 2018
Los Andes Bogota

Introduction to Sensitivity Analysis
The Statistical Study of Sobol Indices

Thierry Klein
thierry.klein@math.univ-toulouse.fr
http://perso.math.univ-toulouse.fr/klein/

Universidad de los Andes Colombia

INSTITUT de MATHEMATIQUES de TOULOUSE

ENAC
ECOLE NATIONALE DE L'AVIATION CIVILE

Contents

1 Introduction

Mathematical models are used in many fields (one can think of environmental risk assessment, nuclear safety, Aeronautics) to model real phenomena. This modeling gives birth to some computer code. This code is used to perform some simulations of the model. Nevertheless in real application the code is very expensive in time. Those code representing physical phenomena take as inputs many numerical parameters, physical variables (those variables could be some real number, some vectors or even some functions) and give in general several outputs. Sensitivity Analysis (SA) is the part of applied mathematics which analysis these kind of code. In general the inputs parameters are not well known, one said that they are uncertain. In the statistical approach we model this uncertainty by considering the inputs as random objets (random variables, random vectors or even stochastic processes). One of the aim of sensitivity analysis is to study how the uncertainty in the output is related to the inputs uncertainty. Hence SA can be for example use to detect the most influent variables, to detect the variables that are not influent (and then fixed them to some nominal value), calibrate some model inputs. There exists many technics to perform some SA. The are local (or derivative) technics or some more global technics. In these lectures, we will focus on an particular aspect of SA, the one that is related to the ANOVA decomposition. This technic is based on a decomposition of the variance that gives raise to some indices (called the Sobol indices). As it will be shown later on, these indices can be seen as indicators on the importance of some inputs parameters.
In these notes, we will restrain our presentation to the statistical analysis of Sobol Indices.

2 Anova or the Hoeffdings decomposition of the variance

2.1 Linear models

Let $(X_1, \ldots, X_d)$ be some inputs random objects and $Y = f(X_1, \ldots, X_d)$ be the random output. Here f is assumed to be unknown. In some applications f a is computer code seen as a black box, if one gives to the computer some inputs, the code returns an answer, but we will assume that we don't have access to the code. In some others applications f can be some measurement of an real experience once the inputs are fixed. One of the first method used by statistician is to fit some linear model, that is to consider that

$$Y = \sum_{j=1}^{d} \beta_j X_j$$

in that case, if the inputs are independent

$$\mathrm{Var}(Y) = \sum_{j=1}^{d} \beta_j^2 \mathrm{Var}(X_j).$$

Hence $\beta_j^2 \frac{\mathrm{Var}(X_j)}{\mathrm{Var}(Y)}$ represents the part of the Variance of Y that is due to the input X_j. Now if the model is not linear, on can proceed an ANOVA type decomposition of the variance in order to quantify the importance of an input.

2.2 The ANOVA-Hoeffding decomposition of the variance

2.2.1 A simple example

In order to understand this decomposition, we will first consider a very simple example. Let $X_1 \in \{0, 1\}$ and $X_2 \in \{0, 1, 2\}$ be two independent random variables, having the uniform distribution respectively on $\{0, 1\}$ and on $\{0, 1, 2\}$. Let G be an application from $\{0, 1\} \times \{0, 1, 2\}$ to $\mathbb{R}$ then

$$G(X_1, X_2) = G_\emptyset + G_{\{1\}}(X_1) + G_{\{2\}}(X_2) + G_{\{1,2\}}(X_1, X_2). \tag{1}$$

Where

$$G_\emptyset = \frac{1}{6}\sum_{i=0}^{1}\sum_{j=0}^{2} G(i,j) \text{ is the mean value of the function}$$

$$G_{\{1\}}(x_1) = \frac{1}{3}\sum_{j=0}^{2} G(x_1,j) - G_\emptyset,\ \forall\ x_1 \in \{0,1\}$$

$$G_{\{2\}}(x_2) = \frac{1}{2}\sum_{i=0}^{1} G(i,x_2) - G_\emptyset\ \forall\ x_2 \in \{0,1,2\}$$

$$G_{\{1,2\}}(x_1,x_2) = G(x_1,x_2) - G_{\{1\}}(x_1) - G_{\{2\}}(x_2) - G_\emptyset.$$

One can see that

$$G_{\{1\}}(X_1) = \mathbb{E}\left(G(X_1,X_2)|X_1\right) - \mathbb{E}\left(G(X_1,X_2)\right),$$
$$G_{\{2\}}(X_2) = \mathbb{E}\left(G(X_1,X_2)|X_2\right) - \mathbb{E}\left(G(X_1,X_2)\right).$$

Now we have

$$\begin{aligned}
\mathbb{E}\left(G_\emptyset G_{\{1\}}(X_1)\right) &= G_\emptyset\left(\frac{1}{3}\sum_{j=0}^{2}\mathbb{E}\left(G(X_1,j)\right)\right) - G_\emptyset^2 \\
&= G_\emptyset\left(\frac{1}{6}\sum_{j=0}^{2}\sum_{i=0}^{1} G(i,j)\right) - G_\emptyset^2 = 0. \\
\mathbb{E}\left(G_\emptyset G_{\{2\}}(X_2)\right) &= 0 \text{ by symmetry.} \\
\mathbb{E}\left(G_{\{1\}}(X_1)G_{\{2\}}(X_2)\right) &= \left(\frac{1}{2}\sum_{i=0}^{1}\mathbb{E}\left(G(i,X_2)\right)\right)\left(\frac{1}{3}\sum_{j=0}^{2}\mathbb{E}\left(G(X_1,j)\right)\right) \\
&\quad - G_\emptyset\left(\frac{1}{3}\sum_{j=0}^{2}\mathbb{E}\left(G(X_1,j)\right)\right) - G_\emptyset\left(\frac{1}{2}\sum_{i=0}^{1}\mathbb{E}\left(G(i,X_2)\right)\right) + G_\emptyset^2 = 0 \\
\mathbb{E}\left(G_\emptyset G_{\{1,2\}}(X_1,X_2)\right) &= 0 \\
\mathbb{E}\left(G_{\{1\}}(X_1)G_{\{1,2\}}(X_1,X_2)\right) &= \mathbb{E}\left(G_{\{2\}}(X_2)G_{\{1,2\}}(X_1,X_2)\right) = 0.
\end{aligned}$$

Hence the variables appearing in decomposition (1) are orthogonal. We can then perform an L^2 decomposition of the variance

$$\operatorname{Var}\left(G(X_1,X_2)\right) = \operatorname{Var}\left(G_{\{1\}}(X_1)\right) + \operatorname{Var}\left(G_{\{2\}}(X_2)\right) + \operatorname{Var}\left(G_{\{1,2\}}(X_1,X_2)\right). \tag{2}$$

We will now generalize Equation (2) without specifying the law of the inputs.

2.2.2 The general model

Let $\mathcal{X} = (X_1,\ldots,X_d)$ be independent random variables, such that X_i belongs to some measurable Polish space $(E_i,\mathcal{B}(E_i))$.

Example 2.1. *Take for exemple $d=4$, X_1 a Poisson random variable with parameter $\lambda > 0$ ie for all $k \in \mathbb{N}$, $\mathbb{P}(X_1 = k) = e^{-\lambda}\frac{\lambda^k}{k!}$, $X_2 \sim \mathcal{N}(m,\sigma^2)$, the distribution of X_3 is the exponential law of parameter 1(with density $f(t) = \exp(-t)$, for $t \geq 0$), and X_4 has the Cauchy distribution on $\mathbb{R}$ (with density $h(t) = \frac{1}{\pi(1+x^2)}$).*

Example 2.2. *Take for exemple $d=3$, X_1 a Poisson random variable with parameter $\lambda > 0$ ie for all $k \in \mathbb{N}$, $\mathbb{P}(X_1 = k) = e^{-\lambda}\frac{\lambda^k}{k!}$, X_2 be some centered Gaussian vector of dimension 3 and X_3 be some brownian motion.*

We denote by $\mathbb{L}^2(P_{\mathcal{X}})$ the set of all measurable function f on $(E, \mathcal{E})$ such that $\mathbb{E}\left(f^2(\mathcal{X})\right) < +\infty$. Where $E = \prod_i^d E_i$ and $\mathcal{E} = \otimes_{i=1}^d \mathcal{B}(E_i)$. The space $\mathbb{L}^2(P_{\mathcal{X}})$ is an Hilbert space with inner product defined by for any $f \in \mathbb{L}^2(P_{\mathcal{X}})$ and $g \in \mathbb{L}^2(P_{\mathcal{X}})$

$$< f, g >= \mathbb{E}\left(f(\mathcal{X})g(\mathcal{X})\right).$$

For any $A \subset \{1, \ldots, d\}$ we set $\mathcal{X}_A = (X_i)_{i\in A}$ and $\mathbb{L}^2_A$ the subspace of $\mathbb{L}^2(P_{\mathcal{X}})$ that are E_A measurable ($E_A = \prod_{i\in A} E_i$) and

$$\mathbb{L}^2_{B\perp A} = \{f \in \mathbb{L}^2_B,\ \forall g \in \mathbb{L}^2_A,\ \mathbb{E}\left(f(\mathcal{X}_B)g(\mathcal{X}_A)\right) = 0\}.$$

Theorem 2.1 (Hoeffding)**.** *Let $G \in \mathbb{L}^2(P_{\mathcal{X}})$. Then G may be uniquely decomposed in $\mathbb{L}^2(P_{\mathcal{X}})$ as the following orthogonal expansion*

$$G(\mathcal{X}) = \sum_{A\subset\{1,\ldots d\}} G_A(\mathcal{X}_A)(a.s.) \tag{3}$$

where

1. $\forall A \subset \{1, \ldots d\}$, $G_A \in \mathbb{L}^2_A$.
2. $\forall A' \subsetneq A \subset \{1, \ldots d\}$, $G_A \in \mathbb{L}^2_{A\perp A'}$.
3. $\forall A',\ \forall A \subset \{1, \ldots d\}$, *with* $A' \cap A \neq A$ *and for* $f \in \mathbb{L}^2_A$, $\mathbb{E}\left(G_A(\mathcal{X}_A)f(\mathcal{X}_{A'})\right) = 0$.

proof We only prove the Theorem for $d = 2$, one can then give a general proof by induction. In fact the proof is just a generalization of what we did in Equation (1). We write

$$G(X_1, X_2) = G_\emptyset + G_{\{1\}}(X_1) + G_{\{2\}}(X_2) + G_{\{1,2\}}(X_1, X_2). \tag{4}$$

Where

$$\begin{aligned}
G_\emptyset &= \mathbb{E}\left(G(X_1, X_2)\right)\\
G_{\{1\}}(X_1) &= \mathbb{E}\left(G(X_1, X_2)|X_1\right) - \mathbb{E}\left(G(X_1, X_2)\right)\\
G_{\{2\}}(X_2) &= \mathbb{E}\left(G(X_1, X_2)|X_2\right) - \mathbb{E}\left(G(X_1, X_2)\right)\\
G_{\{1,2\}}(X_1, X_2) &= G(X_1, X_2) - G_{\{1\}}(X_1) - G_{\{2\}}(X_2) - G_\emptyset
\end{aligned}$$

The orthogonals properties are straightforward consequences that the inputs are independent and that all the functions in the decomposition are centered. □

Corollary 2.1. *Under the assumptions of Theorem 2.1, if we set* $V_A = \mathrm{Var}\left(G_A(\mathcal{X}_A)\right) = \mathbb{E}\left(G_A(\mathcal{X}_A)^2\right)$. *Then*

$$\mathrm{Var}\left(G(\mathcal{X})\right) = \sum_{A\subset\{1,\ldots,d\}} V_A.$$

and

$$1 = \frac{\sum_{A\subset\{1,\ldots,d\}} V_A}{\mathrm{Var}\left(G(\mathcal{X})\right)}.$$

Remark 2.1. *It is a well known fact that for $\mathbb{L}^2$ random variables the conditional expectation of $\mathbb{E}(Z|W)$ is a W- measurable random variable that is the best approximation in the $\mathbb{L}^2$ sence of Z by a W- measurable random variable. Hence G_A is the best aproximation of the function G in $\mathbb{L}^2_A$. So V_A can be seen as the quantification of the sensitivity of G with respect to the inputs $\mathcal{X}_A$. Now the quantity $V_A/\mathrm{Var}\left(G(\mathcal{X})\right)$ would be the key quantity for the study of sensitivity analysis for $\mathbb{L}^2$ random variables. In this lectures we will restrict our study to the studies of these quantities.*

2.3 Sobol indices

Definition 2.1. *Let $A \subset \{1, \ldots, d\}$, $\mathcal{X} = (X_1, \ldots, X_d)$ be independent random variables and G be a square integrable function of $\mathcal{X}$, then we define* ***Sobol' sensitivity index*** *(the Sobol' index) associated to A*

$$S^A := \frac{\mathrm{Var}\left(\mathbb{E}\left[G(\mathcal{X})|X_i,\ i \in A\right]\right)}{\mathrm{Var}\left(G(\mathcal{X})\right)}.$$

Exercise 1 (Ishigami function). *The Ishigami model is given by:*

$$Y = G(X_1, X_2, X_3) = \sin X_1 + 7\sin^2 X_2 + 0.1 X_3^4 \sin X_1 \tag{5}$$

where $(X_j)_{j=1,2,3}$ *are i.i.d. uniform random variables in* $[-\pi;\pi]$.
Show that

$$S^1 = 0.3139, \quad S^2 = 0.4424, \quad S^3 = 0.$$

Exercise 2 (Sobol G-function). *Assume that* $X_1, \ldots, X_d$ *are i.i.d random variables uniformly distributed on* $[0,1]$. *Now take* d *real numbers* $a_1, \ldots, a_d$ *and define the Sobol* $G-$*function by*

$$Y = g_{sobol}(X_1, \ldots, X_d) = \prod_{k=1}^{d} g_k(X_k) \tag{6}$$

with $g_k(X_k) = \frac{|4X_k - 2| + a_k}{1 + a_k}$.
Compute S^i *for* $i \in \{1, \ldots, d\}$.

In general, it is not possible to compute explicitly the Sobol's index. Indeed in most applications G is unknown or very complicated it is then impossible to perform analytic computations. The statistician would then want to give some estimation of these indices.

3 How to estimate Sobol index- the Sobol pick freeze Monte Carlo method

3.1 General framework

We will focus on the estimation of closed index, since if we know all closed index we can recover all indices.
As previously we consider a non necessarily linear regression model connecting an output $Y \in \mathbb{R}$ to independent random input vectors $\mathcal{X} = (X_1, \ldots X_d)$ with for $i = 1, \ldots d$, X_i belongs to some probability space $\mathcal{E}_i$. We denote

$$Y = f(\mathcal{X}) := f(X_1, \ldots, X_d) \tag{7}$$

where f is a deterministic real valued measurable function defined on $\mathcal{E} = \mathcal{E}_1 \times \ldots \mathcal{E}_d$. We assume that Y is square integrable and non deterministic ($\mathrm{Var}Y \neq 0$).
For applications, it is important to be able to estimate simultaneously several index, for this purpose let $\mathbf{u} := (u_1, \ldots, u_k)$ be k subsets of $I_d := \{1, \ldots, d\}$. The vector of closed Sobol indices is then

$$S^{\mathbf{u}} := \left(\frac{\mathrm{Var}(\mathbb{E}(Y|X_i, i \in u_1))}{\mathrm{Var}(Y)}, \ldots, \frac{\mathrm{Var}(\mathbb{E}(Y|X_i, i \in u_k))}{\mathrm{Var}(Y)} \right).$$

Example 3.1. *Assume* $d = 5$, $k = 3$ *and take* $\mathbf{u} := (\{1\}, \{1,3,5\}, \{2,4\})$ *in that case*

$$S^{\mathbf{u}} := \left(\frac{\mathrm{Var}(\mathbb{E}(Y|X_1)}{\mathrm{Var}(Y)}, \frac{\mathrm{Var}(\mathbb{E}(Y|X_1, X_3, X_5)}{\mathrm{Var}(Y)}, \frac{\mathrm{Var}(\mathbb{E}(Y|X_2, X_4))}{\mathrm{Var}(Y)} \right).$$

It is easy to estimate $\mathrm{Var}(Y)$, the problem here is to estimate quantities like $\mathrm{Var}(\mathbb{E}(Y|X_i, i \in u_1))$. Indeed in general, the estimation of conditional expectation is not an easy task. We will see in the next paragraph a very nice trick allowing to transform the variance of the conditional expectation into some covariance. For $\mathcal{X}$ and for any subset v of I_d we define $\mathcal{X}^v$ by the vector such that $X_i^v = X_i$ if $i \in v$ and $X_i^v = X_i'$ if $i \notin v$ where X_i' is an independent copy of X_i. We then set

$$Y^v := f(X^v).$$

Example 3.2. *Assume* $d = 2$ *and* $Y = f(X_1, X_2)$ *and assume* $v = \{1\}$, $\mathcal{X} = (X_1, X_2)$ *and* $\mathcal{X}^v = (X_1, X_2')$ *where* X_2' *is an independent copy of* X_2 *(*X_2' *is also independent of* X_1*),*

$$Y = (X_1, X_2) \text{ and } \mathrm{Y^v := f(X_1, X_2')}.$$

Remark 3.1. *The idea is*
you keep the variable if the index is in v and you take a new one if the index is not in v.

The next lemma shows how to express $S^{\mathbf{u}}$ in terms of covariances. This will lead to a natural estimator.

Lemma 3.1. *For any $u \subset I_p$, one has*

$$\mathrm{Var}(\mathbb{E}(Y|X_i, i \in u)) = \mathrm{Cov}\left(Y, Y^{\mathbf{u}}\right). \tag{8}$$

Proof It is easy to see that Y and $Y^{\mathbf{u}}$ have the same law, in addition we can assume without loss of generality that $\mathbb{E}(Y) = 0$. Now conditioning on the variables X_i, for $i \in u$, Y and $Y^{\mathbf{u}}$ are independent so

$$\begin{aligned}\mathrm{Cov}\left(Y, Y^{\mathbf{u}}\right) &= \mathbb{E}\left(YY^{\mathbf{u}}\right) = \mathbb{E}\left[\mathbb{E}\left(YY^{\mathbf{u}}|X_i, i \in u\right)\right] = \mathbb{E}\left[\mathbb{E}\left(Y|X_i, i \in u\right)\mathbb{E}\left(Y^{\mathbf{u}}|X_i, i \in u\right)\right] \\ &= \mathbb{E}\left[\mathbb{E}\left(Y|X_i, i \in u\right)^2\right] = \mathrm{Var}(\mathbb{E}(Y|X_i, i \in u)).\end{aligned}$$

□

Notation
From now on, we will denote $\mathrm{Var}(Y)$ by V, $\mathrm{Cov}(Y, Y^{\mathbf{u}})$ by C_u and $\overline{Z}_N$ the empirical mean of any N-sample $(Z_1, \ldots, Z_N)$ of Z.

A first estimation for $S^{\mathbf{u}}$. In view of Lemma 3.1, we are now able to define a first natural estimator of $S^{\mathbf{u}}$ (all sums are taken for i from 1 to N):

$$S^{\mathbf{u}}_{N,\mathrm{Cl}} = \left(\frac{\frac{1}{N}\sum Y_i Y_i^{u_1} - \left(\frac{1}{N}\sum Y_i\right)\left(\frac{1}{N}\sum Y_i^{u_1}\right)}{\frac{1}{N}\sum Y_i^2 - \left(\frac{1}{N}\sum Y_i\right)^2}, \ldots, \frac{\frac{1}{N}\sum Y_i Y_i^{u_k} - \left(\frac{1}{N}\sum Y_i\right)\left(\frac{1}{N}\sum Y_i^{u_k}\right)}{\frac{1}{N}\sum Y_i^2 - \left(\frac{1}{N}\sum Y_i\right)^2}\right). \tag{9}$$

A second estimation for $S^{\mathbf{u}}$. Since the observations consist in $(Y_i, Y_i^{u_1}, \ldots, Y_i^{u_k})_{(1 \le i \le N)}$, a more precise estimation of the first and second moments can be done and we are able to define a second estimator of $S^{\mathbf{u}}$ taking into account all the available information. Define

$$Z_i^{\mathbf{u}} = \frac{1}{k+1}\left(Y_i + \sum_{j=1}^{k} Y_i^{u_j}\right), \quad M_i^{\mathbf{u}} = \frac{1}{k+1}\left(Y_i^2 + \sum_{j=1}^{k}(Y_i^{u_j})^2\right).$$

The second estimator is then defined as

$$T^{\mathbf{u}}_{N,\mathrm{Cl}} = \left(\frac{\frac{1}{N}\sum Y_i Y_i^{u_1} - \left(\frac{1}{2N}\sum (Y_i + Y_i^{u_1})\right)^2}{\frac{1}{N}\sum M_i^{\mathbf{u}} - \left(\frac{1}{N}\sum Z_i^{\mathbf{u}}\right)^2}, \ldots, \frac{\frac{1}{N}\sum Y_i Y_i^{u_k} - \left(\frac{1}{2N}\sum (Y_i + Y_i^{u_k})\right)^2}{\frac{1}{N}\sum M_i^{\mathbf{u}} - \left(\frac{1}{N}\sum Z_i^{\mathbf{u}}\right)^2}\right). \tag{10}$$

Remark 3.2. *Let us just explain why the second estimator is going to be a little better. In $S^{\mathbf{u}}_{N,\mathrm{Cl}}$ in order to estimate the expected value of $\mathbb{E}(Y)$ we only use one of the sample we have that is we compute $\frac{1}{N}\sum Y_i$. Nevertheless, since we have 2 N sample, it seems reasonable to use all the information we have and consider $\frac{1}{2N}\sum (Y_i + Y_i^{u_1})$. We see that in the second case the variance of the estimator of the mean is reduced by a factor 2.*

4 Asymptotic properties of the Pick and Freeze estimators

In the previous section, we showed how to construct two estimators $S^{\mathbf{u}}_{N,\mathrm{Cl}}$ and $T^{\mathbf{u}}_{N,\mathrm{Cl}}$ of the Sobol's indices. We will focus our study on $S^{\mathbf{u}}_{N,\mathrm{Cl}}$, it is easy following the same road map to perform the same study for $T^{\mathbf{u}}_{N,\mathrm{Cl}}$. The two natural questions for a statistician is then

1. Are they consistant? That means do we have a.s. convergence of $S^{\mathbf{u}}_{N,\mathrm{Cl}}$?

2. If yes, do we have a central limit theorem?

The method develop to answer these questions is based on the so-called Delta-method. In the next sub-section, we provide the statistical background needed.

4.1 The Delta method

We recall here a well known result allowing to transfer a Central Limit Theorem via a differentiable functions.

4.1.1 Some basic facts about stochastic convergences

The results of this paragraph are some well known results concerning stochastics convergences. The proofs can be found for example in the book written by Van Der Vaart *Asymptotic Statistic*.

Theorem 4.1. *Let $(X_n)_n$, $(Y_n)_n$ and X, Y be some random vectors and c be a constant . Then*
i) If $X_n \xrightarrow[n]{p.s.} X$ then $X_n \xrightarrow[n]{Pr} X$.
ii) If $X_n \xrightarrow[n]{Pr} X$ thens $X_n \xrightarrow[n]{\mathcal{L}} X$.
iii) $X_n \xrightarrow[n]{Pr} c$ if and only if $X_n \xrightarrow[n]{\mathcal{L}} c$.
iv) If $X_n \xrightarrow[n]{\mathcal{L}} X$ and $d(X_n, Y_n) \xrightarrow[n]{Pr} 0$ then $Y_n \xrightarrow[n]{\mathcal{L}} X$.
v) (Slutsky) If $X_n \xrightarrow[n]{\mathcal{L}} X$ and $Y_n \xrightarrow[n]{Pr} c$ then $(X_n, Y_n) \xrightarrow[n]{\mathcal{L}} (X, c)$.
vi) If $X_n \xrightarrow[n]{Pr} X$ and $Y_n \xrightarrow[n]{Pr} Y$ then $(X_n, Y_n) \xrightarrow[n]{Pr} (X, Y)$.

We introduce here some useful notations

- $X_n = o_P(1)$ means that X_n converges to 0 in probability and $X_n = o_P(R_n)$ means that $X_n = Y_n R_n$ where Y_n converges to 0 in probability.
- $X_n = O_P(1)$ means that the family $(X_n)_n$ is uniformly tight and $X_n = O_P(R_n)$ means that $X_n = Y_n R_n$ where the family $(Y_n)_n$ is uniformly tight.

Lemma 4.1. *Let X_n be a sequence of random vectors going to zero in probability.Then for any $p > 0$, and any function R such that $R(0) = 0$,*

1. *$R(h) = o(\|h\|^p) \implies R(X_n) = o_P(\|X_n\|^p)$.*
2. *$R(h) = O(\|h\|^p) \implies R(X_n) = O_P(\|X_n\|^p)$.*

Theorem 4.2 (Classical C.L.T)**.** *Let $(Z_i)_{i\in\mathbb{N}^*}$ be i.id random variables such that $\mathbb{E}\left(Z_i^2\right) < \infty$, let $m = \mathbb{E}\left(Z_i\right)$ and $\sigma^2 = \mathrm{Var}(Z_i)$. Let $\overline{Z}_n = \frac{1}{n}\sum_{i=1}^n Z_i$. Then*

$$\sqrt{n}\left(\overline{Z}_n - m\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2).$$

Remark 4.1. *If the variable belongs to some $\mathbb{R}^k$ having the same distribution as $Z = (Z_1, \ldots, Z_k)$ the result is the same the limit distribution is the centered Gaussian vector with covariances matrix Σ defined for $1 \le i \le k$ and $1 \le j \le k$ by $(\Sigma)_{i,j} = \mathrm{Cov}(Z_i, Z_j)$.*

4.1.2 The Delta method

Now assume that you want to estimate some unknown parameter θ and that you know for some reason[1] that $\sqrt{n}\,(\ T_n - \theta) \xrightarrow{\mathcal{L}} X$. But unfortunately you are not really interested by the θ but by some transformation of θ let's say $\phi(\theta)$. The natural question would then be:
Do we still have something like $\sqrt{n}\,(\ \phi(T_n) - \phi(\theta)) \xrightarrow{\mathcal{L}}$???
The answer is obviously yes if ϕ is linear since the continuous mapping theorem insures that

$$\phi\left(\sqrt{n}\,(\ T_n - \theta)\right) \xrightarrow{\mathcal{L}} \phi(X)$$

and then by linearity

$$\sqrt{n}\,(\ \phi(T_n) - \phi(\theta)) \xrightarrow{\mathcal{L}} \phi(X).$$

The answer is not obvious in the general case. Nevertheless it's seems reasonable to think that if ϕ is differentiable, ϕ behaves locally as an linear mapping and the result should be true.

[1] For example the C.L.T

Theorem 4.3 (Delta method)**.** *Let ϕ be an application from $\mathbb{R}^k$ to $\mathbb{R}^m$ differentiable at the point θ. Let T_n be some random vectors in $\mathbb{R}^k$ and $(r_n)_n$ be a sequence of real numbers going to ∞. Then*

$$r_n\left(\phi(T_n)-\phi(\theta)\right) \xrightarrow{\mathcal{L}} D\phi(\theta)(T);$$

as soon as $r_n\left(T_n-\theta\right) \xrightarrow{\mathcal{L}} T$. Moreover the difference $r_n\left(\phi(T_n)-\phi(\theta)\right) - D\phi(\theta)\left(r_n(T_n-\theta)\right)$ converges to zero in probability.

Proof Using Prohorov's Theorem, we know that since the sequence $r_n(T_n-\theta) \xrightarrow{\mathcal{L}} T$, she is uniformly tight. Moreover Slutsky Theorem's shows that $T_n-\theta \xrightarrow{\mathbb{P}} 0$. Consider now $R(h)=\phi(\theta+h)-\phi(\theta)-D\phi(\theta)(h)$, since ϕ is differentiable we know that $R(h)=o(\|h\|)$. Applying now Lemma 4.1,

$$\phi(T_n)-\phi(\theta)-D\phi(\theta)(T_n-\theta)=R(T_n-\theta)=o_P(\|T_n-\theta\|).$$

Multiplying both sides by r_n, one gets

$$r_n\phi(T_n)-r_n\phi(\theta)-r_nD\phi(\theta)(T_n-\theta)=r_no_P(\|T_n-\theta\|).$$

$r_no_P(\|T_n-\theta\|)=o_P(r_n\|T_n-\theta\|)$. In addition since $r_n(T_n-\theta)$ is uniformly tight, we have that $o_P(r_n\|T_n-\theta\|)=o_P(1)$ [2]. We have just proved the last part of the Theorem. Now $D\phi(\theta)$ is a continuous linear mapping, hence by the continuity mapping Theorem we have

$$r_nD\phi(\theta)(T_n-\theta) \xrightarrow{\mathcal{L}} D\phi(\theta)(T).$$

We conclure using Theorem 4.1, point 4. □

Example 4.1 (Fondamental exemple)**.** *If $\sqrt{n}\left(T_n-\theta\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0,\Sigma)$. Then*

$$\sqrt{n}\left(\phi(T_n)-\phi(\theta)\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0,D\phi(\theta),\Sigma D\phi(\theta)^T)$$

Example 4.2. *Let (X_i) be a sequence of i.i.d random variables distributed as $\mathcal{E}(\lambda)$, here λ is an unknown parameters in $]0,+\infty[$. Then by the CLT we have*

$$\sqrt{n}\left(\overline{X}_n-\frac{1}{\lambda}\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0,\frac{1}{\lambda^2}).$$

Now applying the Delta method with $\phi(x)=\frac{1}{x}$ we get

$$\sqrt{n}\left(\frac{1}{\overline{X}_n}-\lambda\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0,\lambda^2).$$

4.2 Consistency and CLT for $S^{\mathbf{u}}_{N,\mathrm{Cl}}$

Theorem 4.4. *If $\mathbb{E}(Y^2)<+\infty$ then*
$S^{\mathbf{u}}_{N,\mathrm{Cl}}$ and $T^{\mathbf{u}}_{N,\mathrm{Cl}}$converge a.s. to $S^{\mathbf{u}}_{\mathrm{Cl}}$ when goes to infinity.

Proof It is a simple application of the strong law of large numbers and the continuity mapping Theorem. □

Theorem 4.5. *Assume that $\mathbb{E}(Y^4)<\infty$. Then:*

1. $$\sqrt{N}\left(S^{\mathbf{u}}_{N,\mathrm{Cl}}-S^{\mathbf{u}}_{\mathrm{Cl}}\right) \underset{N\to\infty}{\xrightarrow{\mathcal{L}}} \mathcal{N}_k\left(0,\Gamma_{\mathbf{u},S}\right) \tag{11}$$

   where $\Gamma_{\mathbf{u},S}=((\Gamma_{\mathbf{u},S})_{l,j})_{1\le l,j\le k}$ with

   $$(\Gamma_{\mathbf{u},S})_{l,j}=\frac{\mathrm{Cov}(YY^{u_l},YY^{u_j})-S^{u_l}_{\mathrm{Cl}}\mathrm{Cov}(YY^{u_j},Y^2)-S^{u_j}_{\mathrm{Cl}}\mathrm{Cov}(YY^{u_l},Y^2)+S^{u_j}_{\mathrm{Cl}}S^{u_l}_{\mathrm{Cl}}\mathrm{Var}(Y^2)}{\left(\mathrm{Var}(Y)\right)^2}$$

[2]One shall write $o_P(r_n\|T_n-\theta\|)=r_n\|T_n-\theta\|Z_n$ with $Z_n=o_P(1)$ then for an $\epsilon>0$ fixed, we take M such that $\mathbb{P}\left(r_n\|T_n-\theta\|>M\right)<\epsilon$. It is then easy to see that $\forall\eta>0$, $\mathbb{P}\left(r_n\|T_n-\theta\|Z_n>\eta\right)\to 0$.

2.
$$\sqrt{N}\left(T^{\mathbf{u}}_{N,\mathrm{Cl}} - S^{\mathbf{u}}_{\mathrm{Cl}}\right) \underset{N\to\infty}{\overset{\mathcal{L}}{\to}} \mathcal{N}_k\left(0, \Gamma_{\mathbf{u},T}\right) \tag{12}$$
where $\Gamma_{\mathbf{u},T} = ((\Gamma_{\mathbf{u},T})_{l,j})_{1\le l,j\le k}$ *with*
$$(\Gamma_{\mathbf{u},T})_{l,j} = \frac{\mathrm{Cov}(YY^{u_l}, YY^{u_j}) - S^{u_l}_{\mathrm{Cl}}\mathrm{Cov}(YY^{u_j}, M^{\mathbf{u}}) - S^{u_j}_{\mathrm{Cl}}\mathrm{Cov}(YY^{u_l}, M^{\mathbf{u}}) + S^{u_j}_{\mathrm{Cl}}S^{u_l}_{\mathrm{Cl}}\mathrm{Var}(M^{\mathbf{u}})}{\left(\mathrm{Var}(Y)\right)^2}.$$

Remark 4.2. *Note that in Theorem 4.5, we had the stronger assumption t* $\mathbb{E}(Y^4) < \infty$*. But since, we want a C.L.T for Sums of quantities like* Y_i^2*, it is necessary to impose that* Y_i^2 *has a second order moment that is* $\mathbb{E}(Y^4) < \infty$*.*

Example 4.3. *1. Assume* $k = p$*,* $u = (\{1\}, \ldots, \{p\})$ *and* $\mathbb{E}(Y^4) < \infty$*. We denote* $Y_i^{\{j\}}$ *by* Y_i^j*. Here*
$$S^{\mathbf{u}} = \left(\frac{\mathrm{Var}(\mathbb{E}(Y|X_1))}{\mathrm{Var}(Y)}, \ldots, \frac{\mathrm{Var}(\mathbb{E}(Y|X_p))}{\mathrm{Var}(Y)}\right)$$
and
$$T^{\mathbf{u}}_{N,\mathrm{Cl}} = \left(\frac{\frac{1}{N}\sum Y_iY_i^1 - \left(\frac{1}{2N}\sum(Y_i + Y_i^1)\right)^2}{\frac{1}{N}\sum M_i^{\mathbf{u}} - \left(\frac{1}{N}\sum Z_i^{\mathbf{u}}\right)^2}, \ldots, \frac{\frac{1}{N}\sum Y_iY_i^p - \left(\frac{1}{2N}\sum(Y_i + Y_i^p)\right)^2}{\frac{1}{N}\sum M_i^{\mathbf{u}} - \left(\frac{1}{N}\sum Z_i^{\mathbf{u}}\right)^2}\right).$$
The CLT becomes
$$\sqrt{N}\left(T^{\mathbf{u}}_{N,\mathrm{Cl}} - S^{\mathbf{u}}_{\mathrm{Cl}}\right) \underset{N\to\infty}{\overset{\mathcal{L}}{\to}} \mathcal{N}_p\left(0, \Gamma_{\mathbf{u},T}\right)$$
where $\Gamma_{\mathbf{u},T} = ((\Gamma_{\mathbf{u},T})_{l,j})_{1\le l,j\le k}$ *with*
$$\left(\mathrm{Var}(Y)\right)^2 (\Gamma_{\mathbf{u},T})_{l,j} = \mathrm{Cov}(YY^l, YY^j) - S^l_{\mathrm{Cl}}\mathrm{Cov}(YY^j, M^{\mathbf{u}}) - S^j_{\mathrm{Cl}}\mathrm{Cov}(YY^l, M^{\mathbf{u}}) + S^j_{\mathrm{Cl}}S^l_{\mathrm{Cl}}\mathrm{Var}(M^{\mathbf{u}}).$$

2. We can obviously have a CLT for any index of order 2. Indeed if we take $k = 1$ *and* $(i, j) \in \{1, \ldots, p\}^2$ *with* $i \neq j$ *and* $u = \{i, j\}$*. We get* $Z^{\mathbf{u}} = \frac{1}{2}(Y + Y^{\mathbf{u}})$ *and* $M^{\mathbf{u}} = \frac{1}{2}\left(Y^2 + (Y^{\mathbf{u}})^2\right)$*; thus*
$$S^{\mathbf{u}} = \frac{\mathrm{Var}(\mathbb{E}(Y|X_i, X_j))}{\mathrm{Var}(Y)} \text{ and } T^{\mathbf{u}}_{N,\mathrm{Cl}} = \frac{\frac{1}{N}\sum Y_iY_i^{\mathbf{u}} - \left(\frac{1}{2N}\sum(Y_i + Y_i^{\mathbf{u}})\right)^2}{\frac{1}{2N}\sum\left(Y^2 + (Y^{\mathbf{u}})^2\right) - \left(\frac{1}{2N}\sum(Y_i + Y_i^{\mathbf{u}})\right)^2}.$$
The CLT becomes
$$\sqrt{N}\left(T^{\mathbf{u}}_{N,\mathrm{Cl}} - S^{\mathbf{u}}_{\mathrm{Cl}}\right) \underset{N\to\infty}{\overset{\mathcal{L}}{\to}} \mathcal{N}_1\left(0, \Gamma_{\mathbf{u},T}\right)$$
with
$$\left(\mathrm{Var}(Y)\right)^2 (\Gamma_{\mathbf{u},T}) = \mathrm{Var}(YY^{\mathbf{u}}) - 2S^{\mathbf{u}}_{\mathrm{Cl}}\mathrm{Cov}(YY^{\mathbf{u}}, Y^2) + \frac{(S^{\mathbf{u}}_{\mathrm{Cl}})^2}{2}\left(\mathrm{Var}(Y^2) + \mathrm{Cov}(Y^2, (Y^{\mathbf{u}})^2)\right).$$

3. One can also straightforwardly deduce the joint distribution of the vector of all indices of order 2. For example, if $p = 3$ *take* $k = 3$ *and* $\mathbf{u} = (\{1, 2\}, \{1, 3\}, \{2, 3\})$ *and apply Theorem 4.5.*

Exercise 3. *Show that* $S^{\mathbf{u}}_{N,\mathrm{Cl}}$ *is invariant by any centering (translation) of the* Y_i*'s and* $Y_i^{u_j}$*'s for* $j = 1, \ldots, k$*.*

Proof of Theorem 4.5 Since $S^{\mathbf{u}}_{N,\mathrm{Cl}}$ and $T^{\mathbf{u}}_{N,\mathrm{Cl}}$ are invariant by any centering (translation) of the Y_i's and $Y_i^{u_j}$'s for $j = 1, \ldots, k$, we can simplify the next calculations translating by $\mathbb{E}(Y)$. For the sake of simplicity, Y_i and $Y_i^{u_j}$ now denote the centered random variables.
Proof of (11) :
Recall that
$$S^{\mathbf{u}}_{N,\mathrm{Cl}} - S^{\mathbf{u}}_{\mathrm{Cl}} = \left(\frac{\frac{1}{N}\sum Y_iY_i^{u_1} - \left(\frac{1}{N}\sum Y_i\right)\left(\frac{1}{N}\sum Y_i^{u_1}\right)}{\frac{1}{N}\sum Y_i^2 - \left(\frac{1}{N}\sum Y_i\right)^2} - S^{u_1}_{\mathrm{Cl}}, \ldots, \frac{\frac{1}{N}\sum Y_iY_i^{u_k} - \left(\frac{1}{N}\sum Y_i\right)\left(\frac{1}{N}\sum Y_i^{u_k}\right)}{\frac{1}{N}\sum Y_i^2 - \left(\frac{1}{N}\sum Y_i\right)^2} - S^{u_k}_{\mathrm{Cl}}\right).$$

Let $W_i = (Y_i Y_i^{u_j}, j=1,\dots,k, Y_i, Y_i^{u_j}, j=1\dots,k, Y_i^2)^t$ $(i=1,\dots)$ and g the mapping from $\mathbb{R}^{2k+2}$ to $\mathbb{R}^k$ defined by

$$g(x_1,\dots,x_k,y,y_1,\dots,y_k,z) = \left(\frac{x_1 - yy_1}{z-y^2},\dots,\frac{x_k - yy_k}{z-y^2}\right).$$

Let Σ denote the covariance matrix of W_i and set

$$E=\mathbb{E}(Y), V=\mathrm{Var}(Y), C_Z=\mathrm{Cov}(Y,Y^Z), C_X=\mathrm{Cov}(Y,Y^{\mathbf{u}}), C=\mathrm{Cov}(Y^Z,Y^{\mathbf{u}}), W=(YY^{\mathbf{u}},Y,Y^2,Y^{\mathbf{u}},YY^Z,Y^Z)^t$$

and

$$\Sigma = \begin{pmatrix} \mathrm{Var}(YY^{\mathbf{u}}) & \mathrm{Cov}(YY^{\mathbf{u}},Y) & \mathrm{Cov}(YY^{\mathbf{u}},Y^2) & \mathrm{Cov}(YY^{\mathbf{u}},Y^{\mathbf{u}}) & \mathrm{Cov}(YY^{\mathbf{u}},YY^Z) & \mathrm{Cov}(YY^{\mathbf{u}},Y^Z) \\ \mathrm{Cov}(Y,YY^{\mathbf{u}}) & V & \mathrm{Cov}(Y,Y^2) & C_X & \mathrm{Cov}(Y,YY^Z) & C_Z \\ \mathrm{Cov}(Y^2,YY^{\mathbf{u}}) & \mathrm{Cov}(Y^2,Y) & \mathrm{Var}(Y^2) & \mathrm{Cov}(Y^2,Y^{\mathbf{u}}) & \mathrm{Cov}(Y^2,YY^Z) & \mathrm{Cov}(Y^2,Y^Z) \\ \mathrm{Cov}(Y^{\mathbf{u}},YY^{\mathbf{u}}) & C_X & \mathrm{Cov}(Y^{\mathbf{u}},Y^2) & V & \mathrm{Cov}(Y^{\mathbf{u}},YY^Z) & C \\ \mathrm{Cov}(YY^Z,YY^{\mathbf{u}}) & \mathrm{Cov}(YY^Z,Y) & \mathrm{Cov}(YY^Z,Y^2) & \mathrm{Cov}(YY^Z,Y^{\mathbf{u}}) & \mathrm{Var}(YY^Z) & \mathrm{Cov}(YY^Z,Y^Z) \\ \mathrm{Cov}(Y^Z,YY^{\mathbf{u}}) & C_Z & \mathrm{Cov}(Y^Z,Y^2) & C & \mathrm{Cov}(Y^Z,YY^Z) & V \end{pmatrix}$$

First, the following central limit theorem holds

$$\sqrt{N}\left(\frac{1}{N}\sum W_i - \mathbb{E}(W)\right) \underset{N\to\infty}{\overset{\mathcal{L}}{\to}} \mathcal{N}_{2k+2}(0,\Sigma)$$

We then apply the so-called Delta method to W and g so that

$$\sqrt{N}\left(g(\overline{W}_N) - g(\mathbb{E}(W))\right) \underset{N\to\infty}{\overset{\mathcal{L}}{\to}} \mathcal{N}\left(0, J_g(\mathbb{E}(W))\Sigma J_g(\mathbb{E}(W))^t\right)$$

with $J_g(\mathbb{E}(W))$ the Jacobian of g at point $\mathbb{E}(W)$.
Define $(g_1,\dots,g_k) := g$. For $i=1,\dots,k$, $j=1,\dots,k$,

$$\begin{cases} \frac{\partial g_j}{\partial x_i}(\mathbb{E}(W)) = \frac{1}{V}\delta_{i,j} \\ \frac{\partial g_j}{\partial y}(\mathbb{E}(W)) = 0 \\ \frac{\partial g_j}{\partial y_i}(\mathbb{E}(W)) = 0 \\ \frac{\partial g_j}{\partial z}(\mathbb{E}(W)) = -\frac{S_{\mathrm{Cl}}^{u_j}}{V} \end{cases}$$

with $\delta_{i,i}=1$ and $\delta_{i,j}=0$ if $i\neq j$. Thus $\Gamma_{\mathbf{u},S} = J_g(\mathbb{E}(W))\Sigma J_g(\mathbb{E}(W))^t$ is as stated in Theorem 4.5.

Proof of (12) :
The proof is similar to the one of (11). We now define $W_i = (Y_i Y_i^{u_j}, j=1,\dots,k, Y_i, Y_i^{u_j}, j=1\dots,k, \overline{(Y_i^{\mathbf{u}})^2})^t$. We apply the delta method to g from $\mathbb{R}^{2k+2}$ into $\mathbb{R}^k$ defined by

$$g(x_1,\dots,x_k,y,y_1,\dots,y_k,z) = \left(\frac{x_1 - \left(\frac{y+y_1}{2}\right)^2}{z - \left(\frac{y+y_1+\ldots+y_k}{k+1}\right)^2},\dots,\frac{x_k - \left(\frac{y+y_k}{2}\right)^2}{z - \left(\frac{y+y_1+\ldots+y_k}{k+1}\right)^2}\right).$$

For $i=1,\dots,k$, $j=1,\dots,k$,

$$\begin{cases} \frac{\partial g_j}{\partial x_i}u(\mathbb{E}(W)) = \frac{1}{V}\delta_{i,j} \\ \frac{\partial g_j}{\partial y}(\mathbb{E}(W)) = 0 \\ \frac{\partial g_j}{\partial y_i}(\mathbb{E}(W)) = 0 \\ \frac{\partial g_j}{\partial z}(\mathbb{E}(W)) = -\frac{S_{\mathrm{Cl}}^{u_j}}{V}. \end{cases}$$

□

Exercise 4. *Let $Y = X_1 + X_2$ with X_1 and X_2 i.i.d. $\mathcal{N}(0,1)$ distributed. Let $\mathbf{u} = (\{1\},\{2\}$*

$$S^{\mathbf{u}} = \left(\frac{\mathrm{Var}(\mathbb{E}(Y|X_1))}{\mathrm{Var}(Y)}, \frac{\mathrm{Var}(\mathbb{E}(Y|X_2))}{\mathrm{Var}(Y)}\right)$$

Give an explicit formula for the covariance matrices of Theorem 4.5.

4.3 Application to significance Test

In order to simplify the notation we will write the vectors $S^{\mathbf{u}}$ as column vectors. In this section, we give a general procedure to build significance tests of level α and then illustrate this procedure on two examples.

Let $\mathbf{u} := (u_1, \ldots, u_k)$ so that for any $i = 1, \ldots, k$, u_i is a subset of $I_p := \{1, \ldots, p\}$. Similarly, let $\mathbf{v} := (v_1, \ldots, v_l)$ and $\mathbf{w} := (w_1, \ldots, w_l)$ be l be so that for any $i = 1, \ldots, l$, $v_i \subseteq I_p$ and $w_i \subseteq I_p$.

Consider the following general testing problem

$$H_0 : S^{\mathbf{u}} = 0 \text{ and } S^{\mathbf{v}} = S^{\mathbf{w}} \quad \text{against} \quad H_1 : H_0 \text{ is not true.}$$

Remark 4.3. *Note that one can also test*

$$H_0 : S^{\mathbf{u}} \leq s \quad \textit{against} \quad H_1 : S^{\mathbf{u}} > s,$$

or

$$H_0 : S^{\mathbf{u}} \leq S^{\mathbf{v}} \quad \textit{against} \quad H_1 : S^{\mathbf{u}} > S^{\mathbf{v}}.$$

Appling Theorem 4.5 we have

$$G_N := \sqrt{N}\left(\begin{pmatrix} S^{\mathbf{u}}_{N,\mathrm{Cl}} \\ S^{\mathbf{v}}_{N,\mathrm{Cl}} - S^{\mathbf{w}}_{N,\mathrm{Cl}} \end{pmatrix} - \begin{pmatrix} S^{\mathbf{u}}_{\mathrm{Cl}} \\ S^{\mathbf{v}}_{\mathrm{Cl}} - S^{\mathbf{w}}_{\mathrm{Cl}} \end{pmatrix}\right) \underset{N\to\infty}{\overset{\mathcal{L}}{\to}} \mathcal{N}_{k+l}\left(0, \Gamma\right). \tag{13}$$

Since we have an explicit expression of Γ we may build an estimator Γ_N of Γ thanks to empirical means. Note that $(\Gamma_N)_N$ converges a.s. to Γ. Define

$$\widetilde{G}_N := \sqrt{N}\begin{pmatrix} S^{\mathbf{u}}_{N,\mathrm{Cl}} \\ S^{\mathbf{v}}_{N,\mathrm{Cl}} - S^{\mathbf{w}}_{N,\mathrm{Cl}} \end{pmatrix}.$$

Then:

$$G_N = \widetilde{G}_N - \begin{pmatrix} S^{\mathbf{u}}_{\mathrm{Cl}} \\ S^{\mathbf{v}}_{\mathrm{Cl}} - S^{\mathbf{w}}_{\mathrm{Cl}} \end{pmatrix}.$$

Corollary 4.1. *Under H_0, $\widetilde{G}_N \underset{N\to\infty}{\overset{\mathcal{L}}{\to}} \mathcal{N}_{k+l}\left(0, \Gamma\right)$.*
Under H_1, $|\widetilde{G}_N(1)| + |\widetilde{G}_N(2)| \underset{N\to\infty}{\overset{a.s.}{\to}} \infty$.

This corollary allows us to construct several tests. It is a well-known fact that in the case of a vectorial null hypothesis "there exists no uniformly most powerful test, not even among the unbiased tests". In practice, we return to the dimension 1 introducing a function $F : \mathbb{R}^{k+l} \to \mathbb{R}$ and testing $H_0(F) : F(h) = 0$ (respectively $H_1(F) : F(h) \neq 0$) instead of $H_0 : h = 0$ (resp. $H_1 : h \neq 0$). The choice of a reasonable test "depends on the alternatives at which we wish a high power".

Remark 4.4. *If we take as test statistic $T_N = A\widetilde{G}_N$ where A is a linear form defined on $\mathbb{R}^{l+k}$, under H_0, $T_N \underset{N\to\infty}{\overset{\mathcal{L}}{\to}} \mathcal{N}\left(0, A\Gamma A'\right)$. Replacing Γ by Γ_N and using Slutsky's lemma we get*

$$(A\Gamma_N A')^{-1/2} T_N \underset{N\to\infty}{\overset{\mathcal{L}}{\to}} \mathcal{N}\left(0, 1\right).$$

Thus we reject H_0 if $(A\Gamma_N A')^{-1/2} T_N \geq z_\alpha$ where z_α is the $1-\alpha$ quantile of a standard Gaussian random variable.
One can have a similar result when A is not anymore linear but only C^1 by applying the so-called Delta method.

4.3.1 Numerical applications: toy examples

Example 1 In this first toy example, we compare 5 different test statistics through their power function. Let $X = (X_1, X_2) \sim \mathcal{N}(0, I_2)$, and

$$Y = f(X) = \lambda_1 X_1 + \lambda_1 X_2 + \lambda_2 X_1 X_2,$$

with $2\lambda_1^2+\lambda_2^2=1$. We consider here the following testing problem

$$H_0 : S^{\mathbf{1}} = S^{\mathbf{2}} = \lambda_1^2 = 0 \quad \text{against} \quad H_1 : \ \lambda_1 \neq 0.$$

Then, computations lead to

$$\begin{aligned} \Gamma(1,1) = \Gamma(2,2) &= 3 - 2\lambda_1^2 - 11\lambda_1^4 + 24\lambda_1^6 - 24\lambda_1^8 \\ \Gamma(2,1) = \Gamma(1,2) &= -7\lambda_1^4 + 24\lambda_1^6 - 24\lambda_1^8. \end{aligned}$$

The Gaussian limit in Theorem 4.5 is $\mathcal{N}_2(0, 3Id_2)$ under H_0 while it is asymptotically distributed as $\mathcal{N}_2(0,\Gamma)$ under H_1.

Test 1: we take as test statistic $T_{N,1} = \widetilde{G}_N(1) + \widetilde{G}_N(2)$.
Under H_0, $T_{N,1} \underset{N\to\infty}{\overset{\mathcal{L}}{\to}} \mathcal{N}(0,6)$ so we reject H_0 if $T_{N,1} > z_\alpha$ where $z_\alpha/\sqrt{6}$ is the $(1-\alpha)$ quantile of a standard Gaussian random variable.While under H_1, following the procedure of Remark 4.4 with $A = (1\ 1)$.

$$\left(T_{N,1} - 2\sqrt{N}\lambda_1^2\right)/(2[\Gamma(1,1)+\Gamma(1,2)])^{1/2} \underset{N\to\infty}{\overset{\mathcal{L}}{\to}} \mathcal{N}(0,1).$$

It is then easy to compute the theoretical power function.

Test 2: since the Sobol indices are non negative, the testing problem is naturally unilateral. However in view of more general contexts we introduce the test statistic $T_{N,2} = |\widetilde{G}_N(1)| + |\widetilde{G}_N(2)|$. We reject H_0 if $T_{N,2} > z_\alpha$ where $z_\alpha/\sqrt{3}$ is the $(1-\alpha)$ quantile of the random variable having

$$\frac{2}{\sqrt{\pi}} e^{-u^2/4} \Phi(u/\sqrt{2}) \mathbb{1}_{\mathbb{R}_+}(u)$$

as density (Φ being the distribution function of a standard Gaussian random variable). Under H_1, the power function of $T_{N,2}$ and the limit variance are estimated using Monte Carlo technics.
Test 3: in the same spirit, we introduce the test statistic $T_{N,3} = |\widetilde{G}_N(1) + \widetilde{G}_N(2)|$. We reject H_0 if $T_{N,3} > z_\alpha$ where $z_\alpha/\sqrt{6}$ is the $(1-\alpha/2)$ quantile of a standard Gaussian random variable.Under H_1, the power function of $T_{N,3}$ and the limit variance are estimated using Monte Carlo technics.
Test 4: we use the L^2 norm and consider $T_{N,4} = (G_N(1))^2 + (G_N(2))^2$. Under H_0, $T_{N,4}/3 \underset{N\to\infty}{\overset{\mathcal{L}}{\to}} \chi_2(2)$ so we reject H_0 if $T_{N,4} > z_\alpha$ where $z_\alpha/3$ is the $(1-\alpha)$ quantile of a χ_2 random variablewith 2 degrees of freedom. Under H_1, the power function of $T_{N,4}$ and the limit variance are estimated using Monte Carlo technics.

Test 5: we use the infinity norm and consider $T_{N,5} = max(|G_N(1)|; |G_N(2)|)$. We reject H_0 if $T_{N,5} > z_\alpha$ where $z_\alpha/\sqrt{3}$ is the $[1+\sqrt{1-\alpha}]/2$ quantile of a standard Gaussian random variable.Under H_1, the power function of $T_{N,5}$ and the limit variance are estimated using Monte Carlo technics.

Example 2 Let $X = (X_1, X_2, X_3) \sim \mathcal{N}(0, I_3)$, $2\lambda_1^2 + \lambda_2^2 = 1$ and

$$Y = f(X) = \lambda_1(X_2 + X_3) + \lambda_2 X_1 X_2.$$

Let us test if X_1 has any influence ie $H_0 : S^{\{\mathbf{1}\}} = 0$, $S^{\{\mathbf{1,2}\}} = S_{\text{Cl}}^{\{\mathbf{2}\}}$ and $S_{\text{Cl}}^{\{\mathbf{1,3}\}} = S_{\text{Cl}}^{\{\mathbf{3}\}}$. Applying Theorem 4.5 we easily get

$$G_N := \sqrt{N}\left(\begin{pmatrix} S^1_{N,\text{Cl}} \\ S^{1,2}_{N,\text{Cl}} - S^2_{N,\text{Cl}} \\ S^{1,3}_{N,\text{Cl}} - S^3_{N,\text{Cl}} \end{pmatrix} - \begin{pmatrix} S^1_{\text{Cl}} \\ S^{1,2}_{\text{Cl}} - S^2_{\text{Cl}} \\ S^{1,3}_{\text{Cl}} - S^3_{\text{Cl}} \end{pmatrix}\right) \underset{N\to\infty}{\overset{\mathcal{L}}{\to}} \mathcal{N}_3\left(0,\Gamma\right).$$

Here under H_0 the covariance limit Γ in Theorem 4.5 is the identity matrix. Under H_1 we use its explicit expression given in Theorem 4.5 to compute an empirical estimator Γ_N.

N	Min	Mean	Max
10	0.041	0.0463	0.048
50	0.042	0.0482	0.050
100	0.044	0.0489	0.051
500	0.047	0.0510	0.053
1000	0.049	0.0510	0.055

Table 1: Results for the Ishigami function

Ishigami function The Ishigami model is given by:

$$Y = f(X_1, X_2, X_3) = \sin X_1 + 7\sin^2 X_2 + 0.1 X_3^4 \sin X_1 \tag{14}$$

for $(X_j)_{j=1,2,3}$ are i.i.d. uniform random variables in $[-\pi; \pi]$. Exact values of these indices are analytically known:

$$S^{\{1\}} = 0.3139, \quad S^{\{2\}} = 0.4424, \quad S^{\{3\}} = 0.$$

We perform simulations in order to show that our test procedure allows us to recover the fact that $S^{\{3\}} = 0$, even for relatively small values of N. In Table 1, we present the simulated confidence levels obtained for $N \in \{10, 50, 100, 500, 1000\}$ by the following procedure. For each value of N, we use a 1000 sample to estimate the confidence level and we repeat this scheme 20 times. We give in Table 1 the minimum, the mean and the maximum of these 20 distinct simulated values of the confidence levels.

5 Concentration Inequalities

5.1 Motivation

The starting point is the STRONG LAW OF LARGE NUMBER

Theorem 5.1. *Assume $(X_n)_{n\geq 1}$ is a sequence of i.i.d random variables such that $\mathbb{E}\left(|X_n|\right) < +\infty$ then*

$$\frac{X_1 + \ldots + X_n}{n} \xrightarrow[n\to\infty]{p.s.} \mathbb{E}(X_1).$$

For the statistician $\mathbb{E}(X_1)$ represents an unknown quantity to be estimated and $\frac{X_1+\ldots+X_n}{n}$ is an natural estimator. In the real life n never goes to infinity, we only have a finite number of observations ($n = 100$, $n = 1000$). It is then natural to wonder for a fixed n if $\frac{X_1+\ldots+X_n}{n}$ is close or far from $\mathbb{E}(X_1)$. The speed of convergence is also unnatural question we can be interested in.
The first answer concerning the rate of convergence is given by the central limit theorem

Theorem 5.2. *Let $(X_n)_{n\geq 1}$ be a sequence of i.i.d random variables such that the variance σ^2 exists (i.e.$\mathbb{E}\left(X_n^2\right) < +\infty$) then*

$$\sqrt{n}\Big(\frac{X_1 + \ldots + X_n}{n} - \mathbb{E}(X_1)\Big) \xrightarrow[n\to\infty]{Loi} \mathcal{N}(0, \sigma^2).$$

Roughly speaking this theorem tells us that $\frac{X_1+\ldots+X_n}{n}$ goes at rate $\sqrt{n}$ to $\mathbb{E}(X_1)$. Nevertheless, this is an asymptotic result and gives us nothing when n is fixed (in particular if n is small).
The aim of concentration inequalities is to give non asymptotic results allowing to quantify the error $\frac{X_1+\ldots+X_n}{n} - \mathbb{E}(X_1)$ for a fixed n. There exists several concentration inequalities, we will only present the one needed for our purpose.

5.2 Bennett's inequality

Theorem 5.3 (Bennett's inequality)**.** *Let $X_1, \ldots, X_n$ be n independent random variables with finite variance. Assume that for all index i, $X_i \leq b$. Set*

$$S = \sum_{i=1}^{n} (X_i - \mathbb{E}(X_i))$$

and

$$v=\sum_{i=1}^{n}\mathbb{E}(X_i^2).$$

For $u\in\mathbb{R}$, *set* $\phi(u)=e^u-u-1$ *and for* $u\geq -1$, $h(u)=(1+u)\log(1+u)-u$ *Then*

1. *For* $t>0$

$$\psi_S(t):=\log\left(\mathbb{E}\left(e^{tS}\right)\right)\leq n\log\left(1+\frac{v}{nb^2\phi(bt)}\right)\leq\frac{v}{b^2}\phi(bt)$$

2. *For* $x>0$,

$$\mathbb{P}\left(S\geq x\right)\leq\exp\left(-\frac{v}{b^2}h\left(\frac{bx}{v}\right)\right)$$

Proof

1.
   Step 1: One can assume (without loss of generality) that $b=1$.
   Step 2: Note first that $u\mapsto\frac{\phi(u)}{u^2}$ is increasing. Hence since $X_i\leq 1$ it is obvious that

$$\phi(tX_i)\leq t^2X_i^2\phi(t)=X_i^2\left(e^t-t-1\right)$$
$$e^{tX_i}\leq tX_i+1+X_i^2\left(e^t-t-1\right)$$

   Step 3: We computs $\psi_S(t)$ and use step 2

$$\psi_S(t)=\sum_{i=1}^{n}\log\left(\mathbb{E}\left[e^{t(X_i-\mathbb{E}(X_i))}\right]\right)=\sum_{i=1}^{n}\left(\log\left(\mathbb{E}\left[e^{tX_i}\right]\right)-t\mathbb{E}(X_i)\right)$$
$$\leq\sum_{i=1}^{n}\left(\log\left(1+t\mathbb{E}(X_i)+\mathbb{E}(X_i^2)\left(e^t-t-1\right)\right)-t\mathbb{E}(X_i)\right)$$

   using the concavity of $u\mapsto\log(1+u)$ we have

$$\psi_S(t)\leq n\left(\log\left(1+t\frac{\sum_{i=1}^{n}\mathbb{E}(X_i)}{n}+\frac{v}{n}\left(e^t-t-1\right)\right)-t\frac{\sum_{i=1}^{n}\mathbb{E}(X_i)}{n}\right)$$
$$\leq v\left(e^t-t-1\right).$$

   Which proves the first point.

2. We use the Cramér's method[3]

$$\mathbb{P}\left(S\geq x\right)\leq e^{-tx+\psi_S(t)}\leq e^{-tx+v\left(e^t-t-1\right)}$$

   The right hand side is optimized for $t=\log\left(1+\frac{x}{v}\right)$, we then get

$$\mathbb{P}\left(S\geq x\right)\leq e^{-\log\left(1+\frac{x}{v}\right)x+v\left(\frac{x}{v}-\log\left(1+\frac{x}{v}\right)\right)}$$
$$\leq e^{-v\left[\left(1+\frac{x}{v}\right)\log\left(1+\frac{x}{v}\right)-\frac{x}{v}\right]}.$$

□

Remark 5.1. *One can see that*

$$h(u)\geq\frac{u^2}{2(1+u/3)}.$$

which provides

$$\mathbb{P}\left(S\geq x\right)\leq e^{-\frac{x^2}{2(v+bx/3)}}.$$

Notation
V will denote $\mathrm{Var}(Y)$ and as previously h is defined for $x>-1$ by

$$h(x)=(1+x)\ln(1+x)-x.$$

[3] We recall that if h is a positive function then Markov inequality says that $\mathbb{P}\left(h(X)\geq h(x)\right)\leq\frac{\mathbb{E}(h(X))}{h(x)}$, and consider $h(x)=e^{tx}$

5.3 Concentration inequalities for $S^{\mathbf{u}}_{N,\mathrm{Cl}}$

Let us introduce the following random variables

$$U_i^{\pm} = Y_i Y_i^{\mathbf{u}} - (S^{\mathbf{u}} \pm y)(Y_i)^2 \text{ et } J_i^{\pm} = (S^{\mathbf{u}} \pm y) Y_i - Y_i^{\mathbf{u}}$$

Set V_U^+ (resp. V_U^-, V_J^+ and V_J^-) the moment of order 2 of the variables U_i^+ (resp. U_i^-, J_i^+ and J_i^-).

Theorem 5.4. *Soit $b > 0$ et $y > 0$. We assume that Y_i and $Y_i^{\mathbf{u}}$belongs to $[-b, b]$. Then*

$$\mathbb{P}\left(S^{\mathbf{u}}_{N,\mathrm{Cl}} \geq S^{\mathbf{u}} + y\right) \leq M_1 + 2M_2 + 2M_3, \tag{15}$$

$$\mathbb{P}\left(S^{\mathbf{u}}_{N,\mathrm{Cl}} \leq S^{\mathbf{u}} - y\right) \leq M_4 + 2M_2 + 2M_5, \tag{16}$$

where

$$M_1 = \exp\left\{-\frac{NV_U^+}{b_U^2} h\left(\frac{b_U}{V_U^+}\frac{yV}{2}\right)\right\} \qquad M_3 = \exp\left\{-\frac{NV_J^+ b^2}{b_U^2} h\left(\frac{b_U}{bV_J^+}\sqrt{\frac{yV}{2}}\right)\right\}$$

$$M_2 = \exp\left\{-\frac{NV}{b^2} h\left(\frac{b}{V}\sqrt{\frac{yV}{2}}\right)\right\} \qquad M_4 = \exp\left\{-\frac{NV_U^-}{b_U^2} h\left(\frac{b_U}{V_U^-}\frac{yV}{2}\right)\right\}$$

$$M_5 = \exp\left\{-\frac{NV_J^- b^2}{b_U^2} h\left(\frac{b_U}{bV_J^-}\sqrt{\frac{yV}{2}}\right)\right\}$$

and $b_U = b^2(1 + S^{\mathbf{u}} + y)$.

Proof

Since $S^{\mathbf{u}}$ and $S^{\mathbf{u}}_{N,\mathrm{Cl}}$ are invariant when one translate the variables Y and $Y^{\mathbf{u}}$ we can assume that $\mathbb{E}(Y) = 0$.

1. U_i^+ et $U_i^-$are bounded by b_U, J_i^+ and J_i^- by b_U/b, moreover

   $$\mathbb{E}(U_i^+) = -yV \qquad \mathbb{E}(J_i^+) = 0$$
   $$\mathbb{E}(U_i^-) = yV \qquad \mathbb{E}(J_i^-) = 0$$

   and

   $$V_U^{\pm} = \mathrm{Var}(YY^{\mathbf{u}}) + (S^{\mathbf{u}} + y)^2 \mathrm{Var}(Y^2) - 2(S^{\mathbf{u}} \pm y)\mathrm{Cov}(YY^{\mathbf{u}}, Y^2) + y^2V^2$$

   $$V_J^{\pm} = ((S^{\mathbf{u}} \pm y)^2 + 1)V - 2(S^{\mathbf{u}} \pm y)C_u.$$

2. Proof of (15). As

   $$\{a + b \geq c\} \subset \{a \geq c/2\} \cup \{b \geq c/2\} \quad \text{et} \quad \{ab \geq c\} \subset \{|a| \geq \sqrt{c}\} \cup \{|b| \geq \sqrt{c}\}$$

   we have

   $$\begin{aligned}
   \mathbb{P}\left(S^{\mathbf{u}}_{N,\mathrm{Cl}} \geq S^{\mathbf{u}} + y\right) &= \mathbb{P}\left(\frac{\frac{1}{N}\sum_{i=1}^N Y_i Y_i^{\mathbf{u}} - \overline{Y}_N \overline{Y}^{\mathbf{u}}_N}{\frac{1}{N}\sum_{i=1}^N (Y_i)^2 - \left(\overline{Y}_N\right)^2} \geq S^{\mathbf{u}} + y\right) \\
   &= \mathbb{P}\left(\frac{1}{N}\sum_{i=1}^N \left(U_i^+ - \mathbb{E}(U^+)\right) + \overline{Y}_N \overline{J}^+_N \geq yV\right) \\
   &\leq \mathbb{P}\left(\sum_{i=1}^N \left(U_i^+ - \mathbb{E}(U^+)\right)] \geq N\frac{yV}{2}\right) + \mathbb{P}\left(\sum_{i=1}^N Y_i \geq N\sqrt{\frac{yV}{2}}\right) \\
   &\quad + \mathbb{P}\left(\sum_{i=1}^N (-Y_i) \geq N\sqrt{\frac{yV}{2}}\right) + \mathbb{P}\left(\sum_{i=1}^N J_i^+ \geq N\sqrt{\frac{yV}{2}}\right) \\
   &\quad + \mathbb{P}\left(\sum_{i=1}^N (-J_i^+) \geq N\sqrt{\frac{yV}{2}}\right).
   \end{aligned}$$

   Inequality (15) comes from the application of Bennett's inequality (apply Bennett's result five time).

3. Proof (16). Similarly we have

$$\begin{aligned}\mathbb{P}\left(S^{\mathbf{u}}_{N,\mathrm{Cl}} \le S^{\mathbf{u}} - y\right) &= \mathbb{P}\left(\frac{1}{N}\sum_{i=1}^{N}\left(-U_i^- + \mathbb{E}(U^-)\right) + (-\overline{Y}_N)\overline{J}_N^- \ge yV\right)\\ &\le \mathbb{P}\left(\sum_{i=1}^{N}\left(-U_i^- + \mathbb{E}(U^-)\right) \ge N\frac{yV}{2}\right) + \mathbb{P}\left(\sum_{i=1}^{N} Y_i \ge N\sqrt{\frac{yV}{2}}\right)\\ &\quad + \mathbb{P}\left(\sum_{i=1}^{N}(-Y_i) \ge N\sqrt{\frac{yV}{2}}\right) + \mathbb{P}\left(\sum_{i=1}^{N} J_i^- \ge N\sqrt{\frac{yV}{2}}\right)\\ &\quad + \mathbb{P}\left(\sum_{i=1}^{N}(-J_i^-) \ge N\sqrt{\frac{yV}{2}}\right).\end{aligned}$$

Inequality (16) comes from the application of Bennett's inequality (apply Bennett's result five time).

□

Exercise 5. *Let $Y = X_1 + X_2$ where X_1 and X_2 are i.i.d. uniformly distributed on $[0,1]$. Let $u = \{1\}$, and compute in that case $S^{\mathbf{u}}$ and the bound M_1, M_2 and M_3.*

6 Case of Vectorial outputs

6.1 Motivation

We begin by considering two examples that enlighten the need for a proper definition of sensitivity indices for multivariate outputs.

Example 6.1. *Let us consider the following nonlinear model*

$$Y = f^{a,b}(X_1, X_2) := \begin{pmatrix} f_1^{a,b}(X_1,X_2)\\ f_2^{a,b}(X_1,X_2)\end{pmatrix} = \begin{pmatrix} X_1 + X_1X_2 + X_2\\ aX_1 + bX_1X_2 + X_2\end{pmatrix}$$

where X_1 and X_2 are assumed to be i.i.d. standard Gaussian random variables (r.v.s).
First, we compute the one-dimensional Sobol indices $S^{\mathbf{j}}(f_i^{a,b})$ of $f_i^{a,b}$ with respect to X_j $(i,j = 1,2)$. We get

$$\begin{aligned}(S^{\mathbf{1}}(f_1^{a,b}), S^{\mathbf{1}}(f_2^{a,b})) &= (1/3, a^2/(1+a^2+b^2))\\ (S^{\mathbf{2}}(f_1^{a,b}), S^{\mathbf{2}}(f_2^{a,b})) &= (1/3, 1/(1+a^2+b^2)).\end{aligned}$$

So that, the ratios

$$\frac{S^{\mathbf{1}}(f_i^{a,b})}{S^{\mathbf{2}}(f_i^{a,b})},\ i = 1,2$$

do not depend on b. Moreover, for $|a| > 1$, as this ratio is greater than or equal to 1, X_1 seems to have more influence on the output.
Now let us perform a sensitivity analysis on $\|Y\|^2$. Straightforward calculus lead to

$$S^{\mathbf{1}}(\|Y\|^2) \ge S^{\mathbf{2}}(\|Y\|^2) \iff (a-1)(a^3 + a^2 + 5a + 5 - 4b) \ge 0.$$

For the quantity $\|Y\|^2$, the region where X_1 is the most influent variable depends on the value of b. This region is not very intuitive.

Example 6.2. *Here, we study the following two-dimensional model*

$$Y = f(X_1, X_2) = \begin{pmatrix} X_1 \cos X_2\\ X_1 \sin X_2\end{pmatrix}$$

with $(X_1, X_2) \sim$ Unif$([0;10]) \otimes$ Unif$([0;\pi/2])$.

We obviously get

$$S^{\mathbf{1}}(f_1^{a,b}) = S^{\mathbf{1}}(f_2^{a,b}) = \frac{10}{5\pi^2 - 30} \approx 0.52$$
$$S^{\mathbf{2}}(f_1^{a,b}) = S^{\mathbf{2}}(f_2^{a,b}) = \frac{3(\pi^2 - 8)}{4(\pi^2 - 6)} \approx 0.36.$$

So that X_1 seems to have more influence on the output than X_2.
If we consider $\|Y\|^2$, we straightforwardly get $\|Y\|^2 = X_1^2$ that does not depend on X_2.

A last motivation to introduce new Sobol indices is related to the statistical problem of their estimation. As the dimension increases the statistical estimation of the whole vector of scalar Sobol indices becomes more and more expensive. Moreover, the interpretation of such a large vector is not easy. This strengthens the fact that one needs to introduce Sobol indices of small dimension, which condense all the information contained in a large collection of scalars.
In the next section we define new Sobol indices generalizing the scalar ones and containing all the information.

6.2 Definition of the new indices

We denote by $X := (X_1, \ldots, X_d)$ the random input, defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and valued in some measurable space $E = E_1 \times \cdots \times E_d$. We denote also by Y the output

$$Y = f(X_1, \ldots, X_d),$$

where $f : E \to \mathbb{R}^k$ is an unknown measurable function (d and k are positive integers). We assume that $X_1, \ldots, X_d$ are independent and that Y is square integrable (i.e. $\mathbb{E}(\|Y\|^2) < \infty$). We also assume, without loss of generality, that the covariance matrix of Y is positive definite.
Let $\mathbf{u}$ be a subset of $\{1, \ldots, d\}$ and denote by $\sim\mathbf{u}$ its complement in $\{1, \ldots, d\}$. Further, we set $X_{\mathbf{u}} = (X_i, i \in \mathbf{u})$ and $E_{\mathbf{u}} = \prod_{i \in \mathbf{u}} E_i$.
As the inputs $X_1, \ldots, X_d$ are independent, f may be decomposed through the Hoeffding decomposition see Theorem 2.1

$$f(X) = c + f_{\mathbf{u}}(X_{\mathbf{u}}) + f_{\sim\mathbf{u}}(X_{\sim\mathbf{u}}) + f_{\mathbf{u},\sim\mathbf{u}}(X_{\mathbf{u}}, X_{\sim\mathbf{u}}), \tag{17}$$

where $c \in \mathbb{R}^k$, $f_{\mathbf{u}} : E_{\mathbf{u}} \to \mathbb{R}^k$, $f_{\sim\mathbf{u}} : E_{\sim\mathbf{u}} \to \mathbb{R}^k$ and $f_{\mathbf{u},\sim\mathbf{u}} : E \to \mathbb{R}^k$ are given by

$$c = \mathbb{E}(Y),\ f_{\mathbf{u}} = \mathbb{E}(Y|X_{\mathbf{u}}) - c,\ f_{\sim\mathbf{u}} = \mathbb{E}(Y|X_{\sim\mathbf{u}}) - c,\ f_{u,\sim\mathbf{u}} = Y - f_{\mathbf{u}} - f_{\sim\mathbf{u}} - c.$$

Thanks to L^2-orthogonality, computing the covariance matrix of both sides of (17) leads to

$$\Sigma = C_{\mathbf{u}} + C_{\sim\mathbf{u}} + C_{\mathbf{u},\sim\mathbf{u}}. \tag{18}$$

Here Σ, $C_{\mathbf{u}}$, $C_{\sim\mathbf{u}}$ and $C_{\mathbf{u},\sim\mathbf{u}}$ are denoting respectively the covariance matrices of Y, $f_{\mathbf{u}}(X_{\mathbf{u}})$, $f_{\sim\mathbf{u}}(X_{\sim\mathbf{u}})$ and $f_{\mathbf{u},\sim\mathbf{u}}(X_{\mathbf{u}}, X_{\sim\mathbf{u}})$.

Remark 6.1. *Notice that for scalar outputs (i.e. when $k = 1$), the covariance matrices are scalar (variances), so that (18) may be interpreted as the decomposition of the total variance of Y. The summands traduce the fluctuation induced by the input factors $X_{\mathbf{u}}$ and $X_{\sim\mathbf{u}}$, and the interactions between them. The (univariate) Sobol index $S^{\mathbf{u}}(f) = \mathrm{Var}(\mathbb{E}(Y|X_{\mathbf{u}}))/\mathrm{Var}(Y)$ is then interpreted as the sensibility of Y with respect to $X_{\mathbf{u}}$. Due to non-commutativity of the matrix product, a direct generalization of this index is not straightforward.*

In the general case ($k \geq 2$), for any square matrix M of size k, the equation (18) can be scalarized in the following way

$$\mathrm{Tr}(M\Sigma) = \mathrm{Tr}(MC_{\mathbf{u}}) + \mathrm{Tr}(MC_{\sim\mathbf{u}}) + \mathrm{Tr}(MC_{\mathbf{u},\sim\mathbf{u}}).$$

This suggests to define as soon as $\mathrm{Tr}(M\Sigma) \neq 0$ the M-sensitivity measure of Y with respect to $X_{\mathbf{u}}$ as

$$S^{\mathbf{u}}(M; f) = \frac{\mathrm{Tr}(MC_{\mathbf{u}})}{\mathrm{Tr}(M\Sigma)}.$$

Of course we can analogously define

$$S^{\sim\mathbf{u}}(M;f) = \frac{\mathrm{Tr}(MC_{\sim\mathbf{u}})}{\mathrm{Tr}(M\Sigma)}, \quad S^{\mathbf{u},\sim\mathbf{u}}(M;f) = \frac{\mathrm{Tr}(MC_{\mathbf{u},\sim\mathbf{u}})}{\mathrm{Tr}(M\Sigma)}.$$

The following lemma is obvious.

Lemma 6.1.

1. *The generalized sensitivity measures sum up to 1*

$$S^{\mathbf{u}}(M;f) + S^{\sim\mathbf{u}}(M;f) + S^{\mathbf{u},\sim\mathbf{u}}(M;f) = 1. \tag{19}$$

2. $0 \le S^{\mathbf{u}}(M;f) \le 1$.

3. *Left-composing f by a linear operator O of $\mathbb{R}^k$ changes the sensitivity measure accordingly to*

$$S^{\mathbf{u}}(M;Of) = \frac{\mathrm{Tr}(MOC_{\mathbf{u}}O^t)}{\mathrm{Tr}(MO\Sigma O^t)} = \frac{\mathrm{Tr}(O^tMOC_{\mathbf{u}})}{\mathrm{Tr}(O^tMO\Sigma)} = S^{\mathbf{u}}(O^tMO;f). \tag{20}$$

4. *For $k = 1$ and for any $M \neq 0$, we have $S^{\mathbf{u}}(M;f) = S^{\mathbf{u}}(f)$.*

6.3 The important identity case

We now consider the special case $M = \mathrm{Id}_k$ (the identity matrix of dimension k). We set $S^{\mathbf{u}}(f) = S^{\mathbf{u}}(\mathrm{Id}_k;f)$. The index $S^{\mathbf{u}}(f)$ has the following obvious properties

Proposition 6.1.

1. *$S^{\mathbf{u}}(f)$ is invariant by left-composition of f by any isometry of $\mathbb{R}^k$ i.e.*

$$\textit{for any square matrix } O \textit{ of size } k \textit{ s.t. } O^tO = Id_k, \quad S^{\mathbf{u}}(Of) = S^{\mathbf{u}}(f);$$

2. *$S^{\mathbf{u}}(f)$ is invariant by left-composition by any nonzero scaling of f i.e.*

$$\textit{for any } \lambda \in \mathbb{R}, \quad S^{\mathbf{u}}(\lambda f) = S^{\mathbf{u}}(f);$$

Remark 6.2. *The properties in this proposition are natural requirements for a sensitivity measure. In the next section, we will show that these requirements can be fulfilled by $S^{\mathbf{u}}(M;f)$ only when $M = \lambda Id_k$ ($\lambda \in \mathbb{R}^*$). Hence, the canonical choice among indices of the form $S^{\mathbf{u}}(M;f)$ is the sensitivity index $S^{\mathbf{u}}(f)$.*

6.4 Identity is the only good choice

The following proposition can be seen as a kind of reciprocal of Proposition 6.1.

Proposition 6.2. *Let M be a square matrix of size k such that*

1. *M does not depend neither on f nor $\mathbf{u}$;*

2. *M has full rank;*

3. *$S^{\mathbf{u}}(M;f)$ is invariant by left-composition of f by any isometry of $\mathbb{R}^k$.*

Then $S^{\mathbf{u}}(M;\cdot) = S^{\mathbf{u}}(\cdot)$.

Proof We can write $M = M_{Sym} + M_{Antisym}$ where $M_{Sym}^t = M_{Sym}$ and $M_{Antisym}^t = -M_{Antisym}$. Since, for any symmetric matrix V, we have $\mathrm{Tr}(M_{Antisym}V) = 0$, we deduce that $S^{\mathbf{u}}(M;f) = S^{\mathbf{u}}(M_{Sym};f)$ (C_u and Σ being symmetric matrices). Thus we assume, without loss of generality, that M is symmetric.
We diagonalize M in an orthonormal basis: $M = PDP^t$, where $P^tP = \mathrm{Id}_k$ and D diagonal. We have

$$S^{\mathbf{u}}(M;f) = \frac{\mathrm{Tr}(PDP^tC_{\mathbf{u}})}{\mathrm{Tr}(PDP^t\Sigma)} = \frac{\mathrm{Tr}(DP^tC_{\mathbf{u}}P)}{\mathrm{Tr}(DP^t\Sigma P)} = S^{\mathbf{u}}(D;P^tf).$$

By assumption 1. and 3., M can be assumed to be diagonal.
Now we want to show that $M = \lambda \mathrm{Id}_k$ for some $\lambda \in \mathbb{R}^*$. Suppose, by contradiction, that M has two different diagonal coefficients $\lambda_1 \neq \lambda_2$. It is clearly sufficient to consider the case $k = 2$. Choose $f = \mathrm{Id}_2$ (hence, $p = 2$), and $\mathbf{u} = \{1\}$. We have $\Sigma = \mathrm{Id}_2$ and $C_{\mathbf{u}} = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right)$. Hence on one hand $S^{\mathbf{u}}(M; f) = \frac{\lambda_1}{\lambda_1+\lambda_2}$. On the other hand, let O be the isometry which exchanges the two vectors of the canonical basis of $\mathbb{R}^2$. We have $S^{\mathbf{u}}(M; Of) = \frac{\lambda_2}{\lambda_1+\lambda_2}$. Thus 3. is contradicted if $\lambda_1 \neq \lambda_2$. The case $\lambda = 0$ is forbidden by 2. Finally, it is easy to check that, for any $\lambda \in \mathbb{R}^*$, $S^{\mathbf{u}}(\lambda \mathrm{Id}_k; \cdot) = S^{\mathbf{u}}(\mathrm{Id}_k; \cdot) = S^{\mathbf{u}}(\cdot)$. □
We now give two toy examples to illustrate our definition.

Example 6.3. *We consider as first example*

$$Y = f^a(X_1, X_2) = \begin{pmatrix} aX_1 \\ X_2 \end{pmatrix},$$

with X_1 and X_2 i.i.d. standard Gaussian random variables. We easily get

$$S^{\mathbf{1}}(f^a) = \frac{a^2}{a^2+1} \quad \text{and} \quad S^{\mathbf{2}}(f^a) = \frac{1}{a^2+1} = 1 - S^{\mathbf{1}}(f).$$

Example 6.4. *We consider Example 6.1*

$$Y = f^{a,b}(X_1, X_2) = \begin{pmatrix} X_1 + X_1X_2 + X_2 \\ aX_1 + bX_1X_2 + X_2 \end{pmatrix}.$$

We have

$$S^{\mathbf{1}}(f^{a,b}) = \frac{1+a^2}{4+a^2+b^2} \quad \text{and} \quad S^{\mathbf{2}}(f^{a,b}) = \frac{2}{4+a^2+b^2}$$

and obviously

$$S^{\mathbf{1}}(f^{a,b}) \geq S^{\mathbf{2}}(f^{a,b}) \iff a^2 \geq 1.$$

This result has the natural interpretation that, as X_1 is scaled by a, it has more influence if and only if this scaling enlarges X_1's support i.e. $|a| > 1$.

6.5 Estimation of $S^{\mathbf{u}}(f)$

6.5.1 The Pick and Freeze estimator

In practice, the covariance matrices $C_{\mathbf{u}}$ and Σ are not analytically available. So as in the scalar case ($k = 1$),we will estimate $S^{\mathbf{u}}(f)$ by using a Monte-Carlo Pick and Freeze method, which uses a finite sample of evaluations of f.
For this purpose we set $Y^{\mathbf{u}} = f(X_{\mathbf{u}}, X'_{\sim\mathbf{u}})$ where $X'_{\sim\mathbf{u}}$ is an independent copy of $X_{\sim\mathbf{u}}$ which is still independent of $X_{\mathbf{u}}$. Let N be an integer. We take N independent copies $Y_1, \ldots, Y_N$ (resp. $Y_1^{\mathbf{u}}, \ldots, Y_N^{\mathbf{u}}$) of Y (resp. $Y^{\mathbf{u}}$). For $l = 1, \ldots, k$, and $i = 1, \ldots, N$, we also denote by $Y_{i,l}$ (resp. $Y_{i,l}^{\mathbf{u}}$) the l^{th} component of Y_i (resp. $Y_i^{\mathbf{u}}$). We then define the following estimator of $S^{\mathbf{u}}(f)$

$$S_{\mathbf{u},N} = \frac{\sum_{l=1}^k \left(\frac{1}{N}\sum_{i=1}^N Y_{i,l}Y_{i,l}^{\mathbf{u}} - \left(\frac{1}{N}\sum_{i=1}^N \frac{Y_{i,l}+Y_{i,l}^{\mathbf{u}}}{2}\right)^2 \right)}{\sum_{l=1}^k \left(\frac{1}{N}\sum_{i=1}^N \frac{Y_{i,l}^2+(Y_{i,l}^{\mathbf{u}})^2}{2} - \left(\frac{1}{N}\sum_{i=1}^N \frac{Y_{i,l}+Y_{i,l}^{\mathbf{u}}}{2}\right)^2 \right)}. \tag{21}$$

Remark 6.3. *Note that this estimator can be written*

$$S_{\mathbf{u},N} = \frac{\mathrm{Tr}\left(C_{\mathbf{u},N}\right)}{\mathrm{Tr}\left(\Sigma_N\right)} \tag{22}$$

where $C_{\mathbf{u},N}$ and Σ_N are the empirical estimators of $C_{\mathbf{u}} = \mathrm{Cov}(Y, Y^{\mathbf{u}})$ and $\Sigma = \mathrm{Var}(Y)$ defined by

$$C_{\mathbf{u},N} = \frac{1}{N}\sum_{i=1}^N Y_i^{\mathbf{u}} Y_i^t - \left(\frac{1}{N}\sum_{i=1}^N \frac{Y_i + Y_i^{\mathbf{u}}}{2}\right)\left(\frac{1}{N}\sum_{i=1}^N \frac{Y_i + Y_i^{\mathbf{u}}}{2}\right)^t$$

and

$$\Sigma_N = \frac{1}{N}\sum_{i=1}^N \frac{Y_iY_i^t + Y_i^{\mathbf{u}}(Y_i^{\mathbf{u}})^t}{2} - \left(\frac{1}{N}\sum_{i=1}^N \frac{Y_i + Y_i^{\mathbf{u}}}{2}\right)\left(\frac{1}{N}\sum_{i=1}^N \frac{Y_i + Y_i^{\mathbf{u}}}{2}\right)^t.$$

6.5.2 Asymptotic properties

A straightforward application of the Strong Law of Large Numbers leads to

Proposition 6.3 (Consistency)**.** *$S_{\mathbf{u},N}$ converges almost surely to $S^{\mathbf{u}}(f)$ when $N \to +\infty$.*

We now study to the asymptotic normality of $(S_{\mathbf{u},N})_N$.

Proposition 6.4 (Asymptotic normality)**.** *Assume $\mathbb{E}(Y_l^4) < \infty$ for all $l = 1, \ldots, k$. For $l = 1, \ldots, k$, we set*

$$U_l = (Y_{1,l} - \mathbb{E}(Y_l))(Y_{1,l}^{\mathbf{u}} - \mathbb{E}(Y_l)), \qquad V_l = (Y_{1,l} - \mathbb{E}(Y_l))^2 + (Y_{1,l}^{\mathbf{u}} - \mathbb{E}(Y_l))^2.$$

Then

$$\sqrt{N}\left(S_{\mathbf{u},N} - S^{\mathbf{u}}(f)\right) \underset{N\to\infty}{\overset{\mathcal{L}}{\to}} \mathcal{N}_1\left(0, \sigma^2\right) \tag{23}$$

where

$$\begin{aligned}\sigma^2 = a^2 \sum_{l,l' \in \{1,\ldots,k\}} \operatorname{Cov}(U_l, U_{l'}) + b^2 \sum_{l,l' \in \{1,\ldots,k\}} \operatorname{Cov}(V_l, V_{l'}) \\ + 2ab \sum_{l,l' \in \{1,\ldots,k\}} \operatorname{Cov}(U_l, V_{l'}),\end{aligned} \tag{24}$$

with

$$a = \frac{1}{\sum_{l=1}^k \operatorname{Var}(Y_l)}, \qquad b = -\frac{a}{2} S^{\mathbf{u}}(f).$$

6.6 Numerical illustrations

In this section, we provide numerical simulations for the sensitivity indices $S^{\mathbf{u}}(f)$ defined in Section **??**. We consider again Example 6.4 with $k = p = 2$, $a = 2$ and $b = 3$ which leads to the following model

$$Y = f(X_1, X_2) = \begin{pmatrix} X_1 + X_2 + X_1 X_2 \\ 2X_1 + 3X_1 X_2 + X_2 \end{pmatrix}.$$

In the "Gaussian case" (respectively "Uniform case"), we take X_1 and X_2 independent standard Gaussian random variables (resp. independent uniform random variables on $[0, 1]$). In these two cases, a simple analytic calculation yields the true values of the sensitivity indices $S^1(f)$ and $S^2(f)$.

7 A first approach for indices based on the whole distribution

We consider, here a numerical code Y seen as a function of the vector of the distributed input $(X_r)_{r=1,\cdots,d}$ ($d \in \mathbb{N}^*$),

$$Y = f(X_1, \ldots, X_d), \tag{25}$$

where f is a regular unknown numerical function on the state space $E_1 \times E_2 \times \ldots \times E_d$ on which the distributed variables $(X_1, \ldots, X_d)$ are living. The random inputs are assumed to be independent. We recall that thanks to the so-called Hoeffding decomposition, f is expanded as an L^2-sum of uncorrelated functions involving only a part of the random inputs. For any subset v of $I_d = \{1, \ldots, d\}$, this leads to an index called the Sobol index that measures the amount of *randomness* of Y carried in the subset of input variables $(X_i)_{i\in v}$. Since nothing has been assumed on the nature of the inputs, one can consider the vector $(X_i)_{i\in v}$ as a single input. Thus without loss of generality, let us consider the case where v reduces to a singleton. The numerator H_v of the Sobol index related to the input X_v is

$$H_v = \operatorname{Var}\left(\mathbb{E}\left[Y|X_v\right]\right) = \operatorname{Var}(Y) - \mathbb{E}\left[\left(Y - \mathbb{E}\left[Y|X_v\right]\right)^2\right] \tag{26}$$

while the denominator of the index is nothing more than the variance of Y. In order to estimate H_v we saw the the clever trick of the Pick and Freeze method. More precisely, let X^v be the random vector such that $X_v^v = X_v$ and $X_i^v = X_i'$ if $i \neq v$ where X_i' is an independent copy of X_i. Then, setting

$$Y^v := f(X^v) \tag{27}$$

an obvious computation leads to the nice relationship

$$\mathrm{Var}(\mathbb{E}(Y|X_v)) = \mathrm{Cov}\,(Y, Y^v)\,. \tag{28}$$

The last equality leads to a natural Monte Carlo estimator (Pick and Freeze estimator)

$$T^v_{N,\mathrm{Cl}} = \frac{1}{N}\sum_{j=1}^N Y_j Y_j^v - \left(\frac{1}{2N}\sum_{j=1}^N (Y_j + Y_j^v)\right)^2 \tag{29}$$

where for $j = 1, \cdots, N$, Y_j (resp. Y_j^v) are independent copies of Y (resp. Y^v). As pointed out before, Sobol indices are based on L^2 decomposition. As a matter of fact, Sobol indices are well adapted to measure the contribution of an input on the deviation around the mean of Y.
We introduce a new sensitivity index that is based on the conditional distribution of the output and requires only $3 \times N$.
The code will be denoted by $Z = f(X_1, \ldots, X_d) \in \mathbb{R}^k$. Let F be the distribution function of Z. For any $t = (t_1, \ldots, t_k) \in \mathbb{R}^k$,

$$F(t) = \mathbb{P}\,(Z \leqslant t) = \mathbb{E}\left[\mathbb{1}_{\{Z\leqslant t\}}\right]$$

and $F^v(t)$ the conditional distribution function of Z conditionally on X_v:

$$F^v(t) = \mathbb{P}\,(Z \leqslant t|X_v,) = \mathbb{E}\left[\mathbb{1}_{\{Z\leqslant t\}}|X_v\right].$$

Notice that $\{Z \leqslant t\}$ means that $\{Z_1 \leqslant t_1, \ldots, Z_k \leqslant t_k\}$. Obviously, $\mathbb{E}\left[F^v(t)\right] = F(t)$. Now, we apply the previous framework with $Y(t) = \mathbb{1}_{\{Z\leqslant t\}}$ and $p = 2$. Hence, for $t \in \mathbb{R}^k$ fixed, we have a consistent and asymptotically normal estimation procedure for the estimation of

$$\mathbb{E}\left[(F(t) - F^v(t))^2\right].$$

We define a Cramér Von Mises type distance of order 2 between $\mathcal{L}\,(Z)$ and $\mathcal{L}\,(Z|X_v)$ by

$$D^v_{2,CVM} := \int_{\mathbb{R}^k} \mathbb{E}\left[(F(t) - F^v(t))^2\right] dF(t). \tag{30}$$

The aim of the rest of the section is dedicated to the estimation of $D^v_{2,CVM}$ and the study of the asymptotic properties of the estimator. Notice that

$$D^v_{2,CVM} = \mathbb{E}\left[\mathbb{E}\left[(F(Z) - F^v(Z))^2\right]\right]. \tag{31}$$

Let us note that these indices are naturally adapted to multivariate outputs.

Remark 7.1. *Unlike the procedure for $p = 2$, we did not normalize the generalized Sobol index of $Y(t)$. The purpose, that becomes clear in this section, is to avoid numerical explosion during the estimation procedure. Indeed, the normalizing term would be $F(t)(1 - F(t))$, like in the Anderson-Darling statistic, canceling for small and large values of t. Nevertheless, in view of the following proposition, one can consider $4D^v_{2,CVM}$ instead of $D^v_{2,CVM}$ in order to have an index bounded by 1 as for the Sobol index. The asymptotic properties will not be affected by this renormalizing factor, so we still consider $D^v_{2,CVM}$.*

Proposition 7.1. *One has the following properties.*

1. *$0 \leqslant D^v_{2,CVM} \leqslant \frac{1}{4}$. Moreover, if $k = 1$ and F is continuous, we have $0 \leqslant D^v_{2,CVM} \leqslant \frac{1}{6}$.*
2. *$D^v_{2,CVM}$ is invariant by translation, by left-composition by any nonzero scaling of Y.*

We then proceed to a double Monte-Carlo scheme for the estimation of $D^v_{2,CVM}$ and consider the following design of experiment consisting in:

1. two N-samples of Z: $(Z_j^{v,1}, Z_j^{v,2})$, $1 \leqslant j \leqslant N$;
2. a third N-sample of Z independent of $(Z_j^{v,1}, Z_j^{v,2})_{1\leqslant j\leqslant N}$: W_k, $1 \leqslant k \leqslant N$.

The empirical estimator of $D^v_{2,CVM}$ is then given by

$$\widehat{D}^v_{2,CVM} = \frac{1}{N}\sum_{k=1}^{N}\left\{\frac{1}{N}\sum_{j=1}^{N}\mathbb{1}_{\{Z_j^{v,1}\leqslant W_k\}}\mathbb{1}_{\{Z_j^{v,2}\leqslant W_k\}} - \left[\frac{1}{2N}\sum_{j=1}^{N}\left(\mathbb{1}_{\{Z_j^{v,1}\leqslant W_k\}} + \mathbb{1}_{\{Z_j^{v,2}\leqslant W_k\}}\right)\right]^2\right\}. \quad (32)$$

The consistency of $\widehat{D}^v_{2,CVM}$ follows directly from the following lemma:

Lemma 7.1. *Let G and H be two L^1-measurable functions. Let $(U_j)_{j\in I_N}$ and $(V_k)_{k\in I_N}$ be two independent samples of iid rv such that $\mathbb{E}[G(U_1,V_1)] = 0$ and $\mathbb{E}[H(U_1,U_2,V_1)] = 0$. We define S_N and T_N by*

$$S_N = \frac{1}{N^2}\sum_{j,k=1}^{N} G(U_j,V_k) \quad \textit{and} \quad T_N = \frac{1}{N^3}\sum_{i,j,k=1}^{N} H(U_i,U_j,V_k).$$

Then S_N and T_N converge a.s. to 0 as N goes to infinity.

Proof. (i) If we prove that $\mathbb{E}[S_N^4] = O\left(\frac{1}{N^2}\right)$, we then apply Borel-Cantelli lemma to deduce the almost sure convergence of S_N to 0. Clearly,

$$\mathbb{E}[S_N^4] = \frac{1}{N^8}\sum \mathbb{E}[G(U_{i_1},V_{j_1})G(U_{i_2},V_{j_2})G(U_{i_3},V_{j_3})G(U_{i_4},V_{j_4})]$$

where the sum is taken over all the indices i_1, i_2, i_3, i_4, j_1, j_2, j_3, j_4 from 1 to N. The only scenarii that could lead to terms in $O\left(\frac{1}{N}\right)$ or even $O\left(1\right)$ appear when we sum over indices all different except 2 i's or 2 j's or over indices all different. Nevertheless, in those cases, at least one term of the form $\mathbb{E}[G(U_i,V_j)]$ appears. Since the function G is centered, those scenarii are then discarded.

(ii) Analogously, it suffices to show that $\mathbb{E}[T_N^4] = O\left(\frac{1}{N^2}\right)$. The only scenarii that could lead to terms in $O\left(\frac{1}{N}\right)$ or even $O\left(1\right)$ appear when we sum over indices all different except 2 i's, 2 j's or 2 k's or over indices all different. Nevertheless, in those cases, at least one term of the form $\mathbb{E}[H(U_i,U_j,V_k)]$ appears. Since the function H is centered, those scenarii are then discarded. □

Corollary 7.1. *$\widehat{D}^v_{2,CVM}$ is strongly consistent as N goes to infinity.*

Proof. The proof is based on Lemma 7.1. First, we define $Z_j = \left(Z_j^{v,1}, Z_j^{v,2}\right)$, $G(Z_j,W_k) = \mathbb{1}_{\{Z_j^{v,1}\leqslant W_k\}}\mathbb{1}_{\{Z_j^{v,2}\leqslant W_k\}}$, $F(Z_j,W_k) = \frac{1}{2}\left(\mathbb{1}_{\{Z_j^{v,1}\leqslant W_k\}} + \mathbb{1}_{\{Z_j^{v,2}\leqslant W_k\}}\right)$ and $H(Z_i,Z_j,W_k) = F(Z_i,W_k)F(Z_j,W_k)$. Second we proceed to the following decomposition

$$
\begin{aligned}
\widehat{D}^v_{2,CVM} &= \frac{1}{N}\sum_{k=1}^N \left\{ \frac{1}{N}\sum_{j=1}^N \mathbb{1}_{\{Z_j^{v,1}\leqslant W_k\}}\mathbb{1}_{\{Z_j^{v,2}\leqslant W_k\}} - \left[\frac{1}{2N}\sum_{j=1}^N \left(\mathbb{1}_{\{Z_j^{v,1}\leqslant W_k\}} + \mathbb{1}_{\{Z_j^{v,2}\leqslant W_k\}}\right)\right]^2 \right\} \\
&= \frac{1}{N^2}\sum_{j,k=1}^N \mathbb{1}_{\{Z_j^{v,1}\leqslant W_k\}}\mathbb{1}_{\{Z_j^{v,2}\leqslant W_k\}} - \frac{1}{4N^3}\sum_{i,j,k=1}^N \left(\mathbb{1}_{\{Z_i^{v,1}\leqslant W_k\}} + \mathbb{1}_{\{Z_i^{v,2}\leqslant W_k\}}\right)\left(\mathbb{1}_{\{Z_j^{v,1}\leqslant W_k\}} + \mathbb{1}_{\{Z_j^{v,2}\leqslant W_k\}}\right) \\
&= \frac{1}{N^2}\sum_{j,k=1}^N G(Z_j,W_k) - \frac{1}{N^3}\sum_{i,j,k=1}^N H(Z_i,Z_j,W_k) \\
&= \frac{1}{N^2}\sum_{j,k=1}^N \{G(Z_j,W_k) - \mathbb{E}[G(Z_j,W_k)]\} - \frac{1}{N^3}\sum_{i,j,k=1}^N \{H(Z_i,Z_j,W_k) - \mathbb{E}[H(Z_i,Z_j,W_k)]\} \\
&\quad + \frac{1}{N^2}\sum_{j,k=1}^N \mathbb{E}[G(Z_j,W_k)] - \frac{1}{N^3}\sum_{i,j,k=1}^N \mathbb{E}[H(Z_i,Z_j,W_k)] \\
&= \frac{1}{N^2}\sum_{j,k=1}^N \{G(Z_j,W_k) - \mathbb{E}[G(Z_j,W_k)]\} - \frac{1}{N^3}\sum_{i,j,k=1}^N \{H(Z_i,Z_j,W_k) - \mathbb{E}[H(Z_i,Z_j,W_k)]\} \\
&\quad + \mathbb{E}[G(Z_1,W_1)] - \left(1-\frac{1}{N}\right)\mathbb{E}[H(Z_1,Z_2,W_1)] - \frac{1}{N}\mathbb{E}[H(Z_1,Z_1,W_1)].
\end{aligned}
$$

The two first sums converges almost surely to 0 by Lemma 7.1. The remaining term goes to $\mathbb{E}[G(Z_1,W_1)] - \mathbb{E}[H(Z_1,Z_2,W_1)]$ as N goes to infinity.

It remains to show that $D^v_{2,CVM} = \mathbb{E}[G(Z_1,W_1)] - \mathbb{E}[H(Z_1,Z_2,W_1)]$. On the one hand,

$$
\begin{aligned}
D^v_{2,CVM} &= \int_{\mathbb{R}} \mathbb{E}[(F(t)-F^v(t))^2]dF(t) = \mathbb{E}[H_v^2(W)] \\
&= \mathbb{E}[\mathrm{Cov}(\mathbb{1}_{\{Z_1^{v,1}\leqslant W_1\}}, \mathbb{1}_{\{Z_1^{v,2}\leqslant W_1\}})] \\
&= \mathbb{E}_W[\mathbb{E}_Z[\mathbb{1}_{\{Z_1^{v,1}\leqslant W_1\}}\mathbb{1}_{\{Z_1^{v,2}\leqslant W_1\}}] - \mathbb{E}_Z[\mathbb{1}_{\{Z_1^{v,1}\leqslant W_1\}}]^2].
\end{aligned}
$$

On the other hand,

$$
\begin{aligned}
&\mathbb{E}[G(Z_1,W_1)] - \mathbb{E}[H(Z_1,Z_2,W_1)] \\
&= \mathbb{E}[\mathbb{1}_{\{Z_1^{v,1}\leqslant W_1\}}\mathbb{1}_{\{Z_1^{v,2}\leqslant W_1\}}] - \frac{1}{4}\mathbb{E}[\left(\mathbb{1}_{\{Z_1^{v,1}\leqslant W_1\}} + \mathbb{1}_{\{Z_1^{v,2}\leqslant W_1\}}\right)\left(\mathbb{1}_{\{Z_2^{v,1}\leqslant W_1\}} + \mathbb{1}_{\{Z_2^{v,2}\leqslant W_1\}}\right)] \\
&= \mathbb{E}_W[\mathbb{E}_Z[\mathbb{1}_{\{Z_1^{v,1}\leqslant W_1\}}\mathbb{1}_{\{Z_1^{v,2}\leqslant W_1\}}]] - \mathbb{E}[\mathbb{1}_{\{Z_1^{v,1}\leqslant W_1\}}\mathbb{1}_{\{Z_2^{v,2}\leqslant W_1\}}] \\
&= \mathbb{E}_W[\mathbb{E}_Z[\mathbb{1}_{\{Z_1^{v,1}\leqslant W_1\}}\mathbb{1}_{\{Z_1^{v,2}\leqslant W_1\}}]] - \mathbb{E}[\mathbb{E}[\mathbb{1}_{\{Z_1^{v,1}\leqslant W_1\}}\mathbb{1}_{\{Z_2^{v,2}\leqslant W_1\}}|W_1]] \\
&= \mathbb{E}_W[\mathbb{E}_Z[\mathbb{1}_{\{Z_1^{v,1}\leqslant W_1\}}\mathbb{1}_{\{Z_1^{v,2}\leqslant W_1\}}]] - \mathbb{E}[\mathbb{E}[\mathbb{1}_{\{Z_1^{v,1}\leqslant W_1\}}|W_1]\mathbb{E}[\mathbb{1}_{\{Z_2^{v,2}\leqslant W_1\}}|W_1]] \\
&= \mathbb{E}_W[\mathbb{E}_Z[\mathbb{1}_{\{Z_1^{v,1}\leqslant W_1\}}\mathbb{1}_{\{Z_1^{v,2}\leqslant W_1\}}]] - \mathbb{E}[\mathbb{E}[\mathbb{1}_{\{Z_1^{v,1}\leqslant W_1\}}|W_1]]\mathbb{E}[\mathbb{E}[\mathbb{1}_{\{Z_2^{v,2}\leqslant W_1\}}|W_1]] \\
&= \mathbb{E}_W[\mathbb{E}_Z[\mathbb{1}_{\{Z_1^{v,1}\leqslant W_1\}}\mathbb{1}_{\{Z_1^{v,2}\leqslant W_1\}}]] - \mathbb{E}[\mathbb{1}_{\{Z_1^{v,1}\leqslant W_1\}}]\mathbb{E}[\mathbb{1}_{\{Z_2^{v,2}\leqslant W_1\}}] \\
&= \mathbb{E}_W[\mathbb{E}_Z[\mathbb{1}_{\{Z_1^{v,1}\leqslant W_1\}}\mathbb{1}_{\{Z_1^{v,2}\leqslant W_1\}}]] - \mathbb{E}[\mathbb{1}_{\{Z_1^{v,1}\leqslant W_1\}}]^2 \\
&= \mathbb{E}_W[\mathbb{E}_Z[\mathbb{1}_{\{Z_1^{v,1}\leqslant W_1\}}\mathbb{1}_{\{Z_1^{v,2}\leqslant W_1\}}] - \mathbb{E}_Z[\mathbb{1}_{\{Z_1^{v,1}\leqslant W_1\}}]^2].
\end{aligned}
$$

□

We now turn to the asymptotic normality of $\widehat{D}^v_{2,CVM}$. We follow van der Vaart **?** to establish the following proposition (more precisely Theorems 20.8 and 20.9, Lemma 20.10 and Example 20.11).

Theorem 7.1. *The sequence of estimators $\widehat{D}^v_{2,CVM}$ is asymptotically Gaussian in estimating $D^v_{2,CVM}$ that is $\sqrt{N}\left(\widehat{D}^v_{2,CVM} - D^v_{2,CVM}\right)$ is weakly convergent to a Gaussian centered variable with variance ξ^2 given by (33).*

Proof. We define

$$\mathbb{G}^i_N(t) = \frac{1}{N}\sum_{j=1}^{N} \mathbb{1}_{\{Z_j^{v,i} \leqslant s\}},\ i = 1, 2,$$

$$\mathbb{G}^{1,2}_N(t,t) = \frac{1}{N}\sum_{j=1}^{N} \mathbb{1}_{\{Z_j^{v,1} \leqslant t\}} \mathbb{1}_{\{Z_j^{v,2} \leqslant t\}},$$

$$\mathbb{F}_N(t) = \frac{1}{N}\sum_{k=1}^{N} \mathbb{1}_{\{W_k \leqslant t\}}.$$

and rewrite $\widehat{D}^v_{2,CVM}$ as a regular function depending on the four empirical processes defined behind:

$$\widehat{D}^v_{2,CVM} = \int \left[\mathbb{G}^{1,2}_N - \left(\frac{\mathbb{G}^1_N + \mathbb{G}^2_N}{2}\right)^2\right] d\mathbb{F}_N.$$

Since these processes are cad-lag functions of bounded variation, we introduce the maps $\psi_1,\ \phi_2$: $BV_1[-\infty,+\infty]^2 \mapsto \mathbb{R}$ and $\Psi : BV_1[-\infty,+\infty]^4 \mapsto \mathbb{R}$ by

$$\psi_i(F_1, F_2) = \int (F_1)^i dF_2 \quad \text{and} \quad \Psi(F_1,F_2,F_3,F_4) = \psi_1(F_1,F_4) - \psi_2\left(\frac{F_2+F_3}{2}, F_4\right),$$

where set $BV_M[a,b]$ is the set of cÃ d-lÃ g functions of variation bounded by M.

By Donsker's theorem,

$$\sqrt{N}\left(\mathbb{G}^1_N - F, \mathbb{G}^2_N - F, \mathbb{G}^{1,2}_N - \widetilde{G}, \mathbb{F}_N - F\right) \underset{N\to\infty}{\overset{\mathcal{L}}{\to}} \mathbb{G}$$

where $G(t,s) = \mathbb{P}\left(Z^{v,1} \leqslant t,\ Z^{v,2} \leqslant s\right)$, $\widetilde{G}(t) = G(t,t)$ and $\mathbb{G}$ is a centered Gaussian process of dimension 4 with covariance function defined for $(t,s) \in \mathbb{R}^2$ by

$$\Pi(t,s) = \mathbb{E}\left(X_t X_s^T\right) - \mathbb{E}\left(X_t\right)\mathbb{E}\left(X_s\right)^T$$

and $X_t := \left(\mathbb{1}_{\{Z^{v,1} \leqslant t\}}, \mathbb{1}_{\{Z^{v,2} \leqslant t\}}, \mathbb{1}_{\{Z^{v,1} \leqslant t\}}\mathbb{1}_{\{Z^{v,2} \leqslant t\}}, \mathbb{1}_{\{W \leqslant t\}}\right)^T$.

Using the chain rule 20.9 and Lemma 20.10 in **?**, the map Ψ is Hadamard-differentiable from the domain $BV_1[-\infty,+\infty]^4$ into $\mathbb{R}$. The derivative is given by

$$(h_1,h_2,h_3,h_4) \mapsto \psi'_{(F_3,F_4)}(h_3,h_4) - \psi'_{\left(\frac{F_1+F_2}{2},F_4\right)}\left(\frac{h_1+h_2}{2}, h_4\right)$$

where the derivative of ψ (resp. ϕ) are given by Lemma 20.10:

$$(h_1,h_2) \mapsto h_2\varphi\circ F_1|_{-\infty}^{+\infty} - \int h_{2-}d\varphi\circ F_1 + \int \varphi'(F_1)h_1 dF_2$$

taking $\varphi \equiv Id$ (resp. $\varphi(x) = x^2$) and h_- is the left-continuous version of a cÃ d-lÃ g function h.
Since

$$\widehat{D}^v_{2,CVM} = \Psi\left(\mathbb{G}^1_N, \mathbb{G}^2_N, \mathbb{G}^{1,2}_N, \mathbb{F}_N\right),$$

we apply the functional delta method 20.8 in **?** to get limit distribution of $\sqrt{N}\left(\widehat{D}^v_{2,CVM}-D^v_{2,CVM}\right)$ converges weakly to the following limit distribution

$$\int h_{4-}d(F^2-\widetilde{G})+\int h_3 dF-\int F(h_1+h_2)dF.$$

Since the map Ψ is defined and continuous on the whole space $BV_1[-\infty,+\infty]^4$, the delta method in its stronger form 20.8 in **?** implies that the limit variable is the limit in distribution of the sequence

$$\begin{aligned}&\Psi'_{(F,F,\widetilde{G},F)}\left(\sqrt{N}\left(\mathbb{G}^1_N-F,\mathbb{G}^2_N-F,\mathbb{G}^{1,2}_N-\widetilde{G},\mathbb{F}_N-F\right)\right)\\ &=\sqrt{N}\left[\int(\mathbb{F}_N-F)_-\,d\left(F^2-\widetilde{G}\right)+\int\left(\mathbb{G}^{1,2}_N-\widetilde{G}-F\left(\mathbb{G}^1_N+\mathbb{G}^2_N-2F\right)\right)dF\right].\end{aligned}$$

We define

$$\begin{aligned}U&:=\int 1\!\!1_{\{W<t\}}d\left(F^2(t)-G(t,t)\right)=G(W_+,W_+)-F(W_+)^2,\\ V&:=\int\left[1\!\!1_{\{Z^{v,1}\leqslant t\}}1\!\!1_{\{Z^{v,2}\leqslant t\}}-\left(1\!\!1_{\{Z^{v,1}\leqslant t\}}+1\!\!1_{\{Z^{v,2}\leqslant t\}}\right)F(t)\right]dF(t)=\frac{1}{2}\left(F(Z^{v,1})^2+F(Z^{v,2})^2\right)-F(Z^{v,1}\vee Z^{v,2}).\end{aligned}$$

Obviously,

$$\begin{aligned}\mathbb{E}(U)&=\int\left(G(t_+,t_+)-F(t_+)^2\right)dF(t),\\ \mathbb{E}(U^2)&=\int\left(G(t_+,t_+)-F(t_+)^2\right)^2dF(t),\\ \mathbb{E}(V)&=\int\left(F(t)^2-G(t,t)\right)dF(t),\\ \mathbb{E}(V^2)&=\frac{1}{2}\int F(t)^4dF(t)+\iint\left[F(t\vee s)\left(F(t\vee s)-F(t)^2-F(s)^2\right)+\frac{1}{2}F(t)^2F(s)^2\right]dG(t,s).\end{aligned}$$

By independence, the limiting variance ξ^2 is

$$\xi^2=\mathrm{Var}U+\mathrm{Var}V. \tag{33}$$

□