Supplement to Global Sensitivity Analysis: a novel generation of mighty estimators based on rank statistics

Fabrice Gamboa[1], Pierre Gremaud[2], Thierry Klein[3], and Agnès Lagnoux[4]

[1]Institut de Mathématiques de Toulouse and ANITI; UMR5219. Université de Toulouse; CNRS. UT3, F-31062 Toulouse, France.
[2]Department of Mathematics. NC State University. Raleigh, North Carolina 27695, USA.
[3]Institut de Mathématiques de Toulouse; UMR5219. Université de Toulouse; ENAC - Ecole Nationale de l'Aviation Civile , Université de Toulouse, France
[4]Institut de Mathématiques de Toulouse; UMR5219. Université de Toulouse; CNRS. UT2J, F-31058 Toulouse, France.

June 30, 2023

Key words Global sensitivity analysis, Cramér-von-Mises distance, Pick-Freeze method, Chatterjee's coefficient of correlation, Sobol indices estimation.

AMS subject classification 62G05, 62G20, 62G30.

1 Technical results

1.1 Convergence of random measures

In the sequel, we will denote by $\mathcal{L}_Z$ the law of a random vector Z.

Lemma 1.1. *There exists a measurable set $\Pi \subset \Omega_W$ with $\mathbb{P}_W$-probability one such that for any $\omega_W \in \Pi$,*

$$\pi_n(\omega_W) := \frac{1}{n} \sum_{j=1}^{n-2} \delta_{\left(\frac{j}{n+1}, \frac{j+1}{n+1}, \frac{j+2}{n+1}, W_j(\omega_W), W_{j+1}(\omega_W)\right)} \Rightarrow \pi := \mathcal{L}_{(X,X,X)} \otimes \mathcal{L}_W \otimes \mathcal{L}_W,$$

as $n \to \infty$ where as before X is uniformly distributed on $[0,1]$ and $\Rightarrow$ stands for the weak convergence of measures.

Proof of Lemma 1.1. Let $\omega_W \in \Omega_W$. Let us consider the continuous and bounded functions defined on $\mathbb{R}^5$ by

$$g_{s,s',s'',t,t'}(x,x',x'',w,w') = \exp\{i(sx + s'x' + s''x'' + tw + t'w')\},$$

for any s, s', s'', t, and t' real numbers. To prove the weak convergence of the measures $(\pi_n(\omega_W))_n$, we show that $\pi_n(\omega_W)(g_{s,s',s'',t,t'})$ converges almost surely for any s, s', s'', t, and $t' \in \mathbb{Q}$ as $n \to \infty$. Finally, we will conclude by density of rational numbers in $\mathbb{R}$.
Let $(s,s',s'',t,t') \in \mathbb{Q}^5$ be fixed. To ease the reading, we use the shorthand notation g for $g_{s,s',s'',t,t'}$ and we omit the notation ω_W as classically done in probability.

One has

$$\pi_n(g) = \int g d\pi_n = \frac{1}{n}\sum_{j=1}^{n-2} e^{i\left(s\frac{j}{n+1} + s'\frac{j+1}{n+1} + s''\frac{j+2}{n+1} + tW_j + t'W_{j+1}\right)}.$$

Obviously, by the independence of the sequence W_n and the convergence theorem of Riemann sums,

$$\mathbb{E}\left[\pi_n(g)\right] = \mathbb{E}\left[e^{itW}\right]\mathbb{E}\left[e^{it'W}\right]\frac{1}{n}\sum_{j=1}^{n-2} e^{i\left(s\frac{j}{n+1} + s'\frac{j+1}{n+1} + s''\frac{j+2}{n+1}\right)} \underset{n\to\infty}{\to} \mathbb{E}\left[e^{itW}\right]\mathbb{E}\left[e^{it'W}\right]\int_0^1 e^{i(s+s'+s'')x}dx.$$

Observe that the almost sure convergence of π_n is equivalent to the almost sure convergence of its real part and that of its imaginary part. Setting

$$U_{n,j} = s\frac{j}{n+1} + s'\frac{j+1}{n+1} + s''\frac{j+2}{n+1} + tW_j + t'W_{j+1},$$

we have $\Re(\pi_n(g)) = \frac{1}{n}\sum_{j=1}^{n-2}\cos\left(U_{n,j}\right)$. In order to apply the Borel-Cantelli lemma, we need to control the fourth moment

$$\mathbb{E}\left[\left(\Re(\pi_n(g)) - \mathbb{E}[\Re(\pi_n(g))]\right)^4\right] = \frac{1}{n^4}\mathbb{E}\left[\left(\sum_{j=1}^{n-2}\cos\left(U_{n,j}\right) - \mathbb{E}[\cos\left(U_{n,j}\right)]\right)^4\right].$$

The random variables $\cos\left(U_{n,j}\right) - \mathbb{E}\left[\cos\left(U_{n,j}\right)\right]$ are real-valued, centered, and bounded so that we can apply inequality (2.14) page 37 in [1]. Then we obtain

$$\mathbb{E}\left[\left(\sum_{j=1}^{n-2}\cos\left(U_{n,j}\right) - \mathbb{E}\left[\cos\left(U_{n,j}\right)\right]\right)^4\right] \leqslant 224n^2\left(\Lambda_2(\alpha^{-1})\right)^2 \qquad (1)$$

where

$$\Lambda_2(\alpha^{-1}) = \sup_{0\leqslant m<n}(m+1)(\alpha_m)^{\frac{1}{2}},$$

where $(\alpha_m)_m$ is the sequence f the strong mixing coefficients of the sequence $(U_{n,j})$. Now since the random variable Z_j^n only depends on (W_j, W_{j+1}), α_m equal zero as soon as $m \geqslant 2$. Hence, there exists a positive constant K such that

$$\frac{1}{n^4}\mathbb{E}\left[\left(\sum_{j=1}^{n-2}\cos\left(U_{n,j}\right) - \mathbb{E}\left[\cos\left(U_{n,j}\right)\right]\right)^4\right] \leqslant \frac{K}{n^2}.$$

It follows by Borel-Cantelli lemma that the real part of $\pi_n(g)$ converges almost surely. Since the imaginary part can be treated using the exact same steps, the proof of Lemma 1.1 is almost complete. Hence, there exists a Borel set $N_{s,s',s'',t,t'}$ with $\mathbb{P}(N_{s,s',s'',t,t'}) = 1$ so that the previous convergence holds on $\Omega_W \backslash N_{s,s',s'',t,t'}$. It remains to define $\Pi := \Omega_W \backslash \cup_{(s,s',s'',t,t')\in\mathbb{Q}^5} N_{s,s',s'',t,t'}$. Obviously, one has $\mathbb{P}(\Pi) = 1$ and the almost sure convergence holds on Π for all functions $g_{s,s',s'',t,t'}$ with $(s, s', s'', t, t') \in \mathbb{Q}^5$.

Finally, the result holding for any five-uplet $(s, s', s'', t, t') \in \mathbb{Q}^5$, we conclude to the required result by density of rational numbers in $\mathbb{R}$. □

The obvious following corollary is a direct consequence of Lemma 1.1.

Corollary 1.2. *We use the notation of Lemma 1.1. For any $\omega_W \in \Pi$, as $n \to \infty$,*

$$\eta_n := \frac{1}{n}\sum_{j=1}^{n-1} \delta_{\left(\frac{j}{n+1},\frac{j+1}{n+1}\right)} \Rightarrow \eta := \mathcal{L}_{(X,X)},$$

$$\kappa_n := \frac{1}{n}\sum_{j=1}^{n-1} \delta_{\left(\frac{j}{n+1},\frac{j+1}{n+1},\frac{j+2}{n+1}\right)} \Rightarrow \kappa := \mathcal{L}_{(X,X,X)},$$

$$\mu_n(\omega_W) := \frac{1}{n}\sum_{j=1}^{n} \delta_{\left(\frac{j}{n+1},W_j(\omega_W)\right)} \Rightarrow \mu := \mathcal{L}_X \otimes \mathcal{L}_W,$$

$$\nu_n(\omega_W) := \frac{1}{n}\sum_{j=1}^{n-1} \delta_{\left(\frac{j}{n+1},\frac{j+1}{n+1},W_j(\omega_W),W_{j+1}(\omega_W)\right)} \Rightarrow \nu := \mathcal{L}_{(X,X)} \otimes \mathcal{L}_W \otimes \mathcal{L}_W.$$

1.2 Generalized L-Statistics

Lemma 1.3. *Let $(E_i)_{i\geqslant 1}$ be a sequence of i.i.d. random variables with standard exponential distribution and let ψ be a bounded measurable function on $[0,1]$. We assume that the set of discontinuity points of ψ has null Lebesgue measure. Then, the sequence*

$$\left(n^{-1/2}\sum_{j=1}^{n-1}\psi(j/n)(E_j-1)\right)_{n\in\mathbb{N}^*}$$

converges in distribution to a centered Gaussian law with asymptotic variance: $\sigma_\psi^2 := \int_{[0,1]}\psi^2(x)dx$.

Proof of Lemma 1.3. For $k \in \mathbb{N}^*$, let cum_k denotes the cumulant of order k of

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n-1}\psi(j/n)(E_j-1).$$

Obviously, $\mathrm{cum}_1 = 0$ and, for $k \geqslant 2$, $\mathrm{cum}_k = n^{-k/2}\sum_{j=1}^{n-1}(\psi(j/n))^k$. So that, $\lim_{n\to\infty}\mathrm{cum}_2 = \int \psi^2(x)dx$ while, for $k \geqslant 3$, $\lim_{n\to\infty}\mathrm{cum}_k = 0$. □

Remark 1.4. The previous lemma obviously extends to the case of a continuous function $\Psi = (\psi_i)$ valued in $\mathbb{R}^d$ ($d \geqslant 1$). In this case, the asymptotic covariance matrix Σ_Ψ is the Gram matrix $\left(\int_{[0,1]}\psi_i(x)\psi_j(x)dx;\ 1 \leqslant i,j \leqslant d\right)$. Indeed, the previous lemma holds for any linear combination of such random vector sequence. A direct computation of the asymptotic variance leads to the quadratic form built on Σ_Ψ.

The next lemma is a generalization of the CLT for a L-statistics (see, e.g., [3, Chapter 22]).

Lemma 1.5. *Let $(U, \mathbb{B}(U))$ be a Polish space where $\mathbb{B}(U)$ denotes the Borel σ algebra of U. We consider a sequence $(\chi_j)_{1\leqslant j\leqslant n,\, n\in\mathbb{N}^*}$ valued in U and Q a probability measure on $U\times[0,1]$. We assume that the sequence of empirical measures $\left(\frac{1}{n}\sum_{j=1}^{n-1}\delta_{\chi_j, j/n}\right)_{n\in\mathbb{N}^*}$ converges in distribution to Q.*
Let ψ be a bounded measurable real function on $U\times[0,1]$. We assume that the set of discontinuity points of ψ has null Q-probability. Then,

$$D_n := \frac{1}{\sqrt{n}}\sum_{j=1}^{n-1}\psi\left(\chi_j, j/n\right)\left(X_{\sigma_n(j)} - \frac{j}{n+1}\right) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\left(0, s_\psi^2\right),$$

where the asymptotic variance s_ψ^2 is given in (3).

Proof of Lemma 1.5. Recall that the sequence (E_i) has been defined in Lemma 1.3. We have

$$\begin{aligned} X_{\sigma_n(j)} - \frac{j}{n+1} &\overset{\mathcal{L}}{=} \frac{\sum_{i=1}^{j}E_i}{\sum_{i=1}^{n+1}E_i} - \frac{j}{n+1} = \frac{1}{\frac{1}{n+1}\sum_{i=1}^{n+1}E_i}\left(\frac{1}{n+1}\sum_{i=1}^{j}E_i - \frac{j}{(n+1)^2}\sum_{i=1}^{n+1}E_i\right) \\ &= \frac{1}{\frac{1}{n+1}\sum_{i=1}^{n+1}E_i}\left(\frac{1}{n+1}\sum_{i=1}^{j}(E_i-1) - \frac{j}{(n+1)^2}\sum_{i=1}^{n+1}(E_i-1)\right), \end{aligned}$$

so that,

$$D_n \overset{\mathcal{L}}{=} \frac{1}{\sqrt{n}(n+1)}\frac{1}{\frac{1}{n+1}\sum_{i=1}^{n+1}E_i}\sum_{j=1}^{n-1}\psi\left(\chi_j, j/n\right)\left(\sum_{i=1}^{j}(E_i-1) - \frac{j}{n+1}\sum_{i=1}^{n+1}(E_i-1)\right).$$

Using the assumption on the empirical measure, we get

$$\frac{1}{n}\sum_{j=1}^{n}\psi\left(\chi_j, j/n\right)\frac{j}{n+1} \to I := \int_{U\times[0,1]} x\psi(\chi, x)dQ(\chi, x).$$

Further, by the weak law of large numbers, $(1/(n+1))\sum_{i=1}^{n+1}E_i$ converges in probability to $\mathbb{E}[E_1] = 1$. Hence, by Slutsky's lemma, we are led to consider the random vector

$$V_n := \frac{1}{\sqrt{n}}\begin{pmatrix}\frac{1}{n+1}\sum_{j=1}^{n-1}\psi\left(\chi_j, j/n\right)\sum_{i=1}^{j}(E_i-1) \\ \sum_{i=1}^{n+1}(E_i-1)\end{pmatrix}.$$

Notice that the first coordinate of V_n can be rewritten as (up to the normalizing factor $n^{-1/2}$)

$$\sum_{i=1}^{n-1}\left(\frac{1}{n+1}\sum_{j=1}^{n-1}\psi\left(\chi_j, j/n\right)\mathbb{1}_{i\leqslant j}\right)(E_i-1).$$

For $t\in[0,1]$, let $\phi(t) := \int_{U\times[t,1]}\psi(\chi, x)dQ(\chi, x)$. We will show below that

$$\lim_n \sup_{t\in[0,1]}\left|\left(\frac{1}{n+1}\sum_{j=1}^{n-1}\psi\left(\chi_j, j/n\right)\mathbb{1}_{i\leqslant j}\right) - \phi(t)\right| = 0. \tag{2}$$

Let assume for a while that this result holds. Then, in our study, we may replace V_n by

$$\widehat{V}_n := \frac{1}{\sqrt{n}} \begin{pmatrix} \frac{1}{n+1}\sum_{i=1}^{n-1} \phi(i/n)(E_i - 1) \\ \sum_{i=1}^{n+1}(E_i - 1) \end{pmatrix}$$

since (2) implies that $\lim_{n\to\infty} \mathbb{E}\|V_n - \widehat{V}_n\|^2 = 0$. Using Remark 1.4, we obtain that the sequence $(\widehat{V}_n)_{n\in\mathbb{N}^*}$ converges in distribution to a centered Gaussian vector with covariance matrix

$$\begin{pmatrix} \int_0^1 \phi^2(t)dt & \int_0^1 \phi(t)dt \\ \int_0^1 \phi(t)dt & 1 \end{pmatrix}.$$

Finally, using the so-called delta method [3, Theorem 3.1], $(D_n)_{n\in\mathbb{N}^*}$ converges in distribution to a centered Gaussian variable with variance

$$s_\psi^2 = \int_0^1 (\phi(t) - I)^2 dt. \tag{3}$$

It remains to show that (2) holds. First let assume that $\psi \geqslant 0$. Set, for $j = 1, \dots n$, $\phi_n(j/n) := (1/(n+1)) \sum_{j=1}^{n-1} \psi(\chi_j, j/n)$ and consider the piece-wise linear extension ϕ_n defined on $[0,1]$. The second Dini's theorem [2] allows to conclude that the sequence of functions $(\phi_n)_{n\in\mathbb{N}^*}$ converges uniformly to ϕ yielding the result. In the general case, we may mimic this reasoning on $\psi^+ = \sup(\psi, 0)$ and $\psi^- = \sup(-\psi, 0)$ and so conclude. □

Notice that, using the definitions of ϕ and I and applying Fubini's theorem, s_ψ^2 can be explicited as follows:

$$\begin{aligned} s_\psi^2 &= \int_0^1 (\phi(t) - I)^2 dt = \int_0^1 \left(\int_{U\times[0,1]} \psi(\chi, x)(\mathbb{1}_{t\leqslant x} - x) dQ(\chi, x) \right)^2 dt \\ &= \int_0^1 \iint_{(U\times[0,1])^2} \psi(\chi_1, x_1)\psi(\chi_2, x_2)(\mathbb{1}_{t\leqslant x_1} - x_1)(\mathbb{1}_{t\leqslant x_2} - x_2) dQ(\chi_1, x_1) dQ(\chi_2, x_2) dt \\ &= \iint_{(U\times[0,1])^2} \psi(\chi_1, x_1)\psi(\chi_2, x_2) \int_0^1 (\mathbb{1}_{t\leqslant x_1} - x_1)(\mathbb{1}_{t\leqslant x_2} - x_2) dt dQ(\chi_1, x_1) dQ(\chi_2, x_2) \\ &= \iint_{(U\times[0,1])^2} \psi(\chi_1, x_1)\psi(\chi_2, x_2)(x_1 \wedge x_2 - x_1 x_2) dQ(\chi_1, x_1) dQ(\chi_2, x_2). \end{aligned} \tag{4}$$

References

[1] E. Rio. *Théorie asymptotique des processus aléatoires faiblement dépendants*, volume 31. Springer Science & Business Media, 1999.

[2] W. Rudin. Real and complex analysis. 1987. *Cited on*, 156, 1987.

[3] A. W. van der Vaart. *Asymptotic statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 1998.