



HAL
open science

Regularizing nested Monte Carlo Sobol' index estimators to balance the trade-off between explorations and repetitions in global sensitivity analysis of stochastic models

Henri Mermoz Kouye, Gildas Mazo

► To cite this version:

Henri Mermoz Kouye, Gildas Mazo. Regularizing nested Monte Carlo Sobol' index estimators to balance the trade-off between explorations and repetitions in global sensitivity analysis of stochastic models. International Journal for Uncertainty Quantification, In press. <hal-03814478v6>

HAL Id: hal-03814478

<https://hal.science/hal-03814478v6>

Submitted on 26 Feb 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Regularizing nested Monte Carlo Sobol' index estimators to balance the trade-off between explorations and repetitions in global sensitivity analysis of stochastic models

Henri Mermoz KOUYE⁽¹⁾ and Gildas MAZO⁽¹⁾

⁽¹⁾Univ. Paris-Saclay, INRAE, MaIAGE, 78350, Jouy-en-Josas, France

February 26, 2025

Abstract

Sobol' sensitivity index estimators for stochastic models are functions of nested Monte Carlo estimators, which are estimators built from two nested Monte Carlo loops. The outer loop explores the input space and, for each of the explorations, the inner loop repeats model runs to estimate conditional expectations. Although the optimal allocation between explorations and repetitions of one's computational budget is well-known for nested Monte Carlo estimators, it is less clear how to deal with functions of nested Monte Carlo estimators, especially when those functions have unbounded Hessian matrices (with respect to the Frobenius norm), as it is the case for Sobol' index estimators. To address this problem, a regularization method is introduced to bound the mean squared error of functions of nested Monte Carlo estimators. Based on a heuristic, an allocation strategy that seeks to minimize a bias-variance trade-off is proposed. The behavior of the estimator obtained is investigated on a simulation study. As an illustrative use-case, the sensitivity analysis of a stochastic COVID-19 epidemic model is performed.

Keywords: nested Monte Carlo, Sobol' index, allocation, global sensitivity analysis, stochastic model

1 Introduction

The global sensitivity analysis of stochastic models is challenging because two sources of uncertainty operate: the uncertainty about the parameters and the intrinsic randomness of the model. We assume that the inputs are statistically mutually independent, and independent of the random noise endogenous to the model. In this paper, we shall use the words “inputs” and “parameters” interchangeably.

Stochastic models are of various kinds and entail different computational costs. For some, runtime is so long—a single run can take hours—that one must replace the model by a faster approximate model, called a surrogate (aka metamodel). In short, one generates a set of runs of the original model to construct a surrogate and then uses this surrogate *in lieu* of the original to perform the sensitivity analysis. The two main families of surrogates are polynomial chaos expansions (Sudret, 2008; Lüthen et al., 2021; Sudret et al., 2017) and Gaussian processes (Baker et al., 2022).

Although surrogate models are powerful tools in uncertainty quantification, they remain approximations. Some classes of surrogate models can be made arbitrarily close to the original model in theory. In practice, the number of runs is limited and assessing the closeness between the surrogate and the original is a reasearch question in itself (Tuo and Jeff Wu, 2016). What’s more, the construction of surrogates for stochastic models is a recent area of research (Castellan et al., 2020; Étoré et al., 2020; Jimenez et al., 2017; Fort et al., 2013; Janon et al., 2014; Panin, 2021), and many challenges remain yet to be addressed. For instance, surrogates based on the generalized lambda distribution (Zhu and Sudret, 2021) cannot fit multi-modal distributions (Zhu and Sudret, 2023).

But not all stochastic models run slow. Many interesting models run sufficiently fast—say, less than one second—so that they can be called a sufficiently large number of times to estimate sensitivity indices directly. Such models include ordinary differential equations (Jarrett et al., 2015; Mazo and Tournier, 2025), Boolean networks (Mazo and Tournier, 2025), and some Markov chains (Kouye et al., 2024; Courcoul et al., 2011; Le Maître et al., 2015; Navarro Jimenez et al., 2016). Three approaches to do the sensitivity analysis in this case are briefly described as follows. In the first approach, one chooses a deterministic quantity of interest (QoI)—usually the conditional expectation or variance of the output given the uncertain parameters—and use it to carry out the sensitivity analysis. Since the QoIs are deterministic functions of the uncertain parameters, the Sobol’-Hoeffding decomposition can be applied and the Sobol’ indices defined. Although this approach can result in loss of information in general (Hart et al., 2017), in some applications one is indeed interested in the sensitivity of the conditional expectation and/or conditional variance of the model’s output. In the second approach, one fixes the seed of the pseudo-random number generator and computes the Sobol’ indices corresponding to that seed. The final indices are averages over a sufficiently large number of seeds (Hart et al., 2017). In the third approach, one regards a stochastic model as a deterministic function with values in some probability space (Fort et al., 2021; Gamboa et al., 2021; da Veiga, 2021). Various sensitivity indices are then defined on such spaces.

The estimation of the sensitivity indices for the conditional expectation of the QoI given the parameters can be performed with various methods (Mycek and De Lozzo, 2019; da Veiga and Gamboa, 2013; da Veiga et al., 2017). A simple and widely-used one is based on two nested Monte Carlo loops (Mazo, 2021; Courcoul et al., 2011). An outer loop explores the input space and an inner loop “repeats” the model at each of the exploration points to approximate the conditional expectation. Although we assume the stochastic model of interest to be computationally inexpensive, computational resources are nonetheless limited. Hence, it is important to optimize the allocation of model runs between the outer and inner loops. In principle, to do so, one chooses an error function (usually the mean squared error) and minimize it with respect to all possible numbers of explorations and repetitions that keep fixed the total number of runs available to the practitioner. See Mazo (2021) for an example in sensitivity analysis and Rainforth et al. (2018); Giles and Haji-Ali (2019); Giorgi et al. (2020); Hong and Juneja (2009); Chen and Zhou (2014, 2017); Binois et al. (2018, 2019) for examples in other contexts. However, the error function is often intractable and one strategy is then to calculate a tractable bound which is then minimized to yield a solution of the allocation problem. Such a strategy was followed in Mazo (2021) with an error function penalizing bad rankings of the estimated sensitivity indices. The calculation of a solution for the mean squared error (MSE) of the nested Monte Carlo Sobol’ index estimator, however, remains an open question. The difficulty lies in that the Sobol’

index estimator is in fact a combination of three nested Monte Carlo estimators and, while the analysis of each of them separately can be done with techniques existing in the literature, the analysis of the combination of them together faces obstacles.

In the present work, a regularization parameter is introduced to bound the MSE of the nested Monte Carlo Sobol' index estimator. The convergence rate of the bound is obtained as the sizes of the outer and inner loops go to infinity. Under this asymptotic setting, a solution of the allocation problem is calculated and used to guide the balance between the numbers of explorations and repetitions in practice. The behavior of the MSE under various scenarios is investigated in a simulation study for two toy models. As an illustrative use-case, the Sobol' indices of the parameters of a COVID-19 epidemic stochastic model are estimated.

The manuscript is organized as follows. Section 2 presents Sobol' indices associated with expected outputs of stochastic models. Section 3 presents nested Monte Carlo estimation and its application to the estimation of Sobol' indices. The method to bound the MSE through regularization is discussed in 4. Section 5 extends the previous section by letting the regularization parameter vanish as the numbers of explorations and repetitions go to infinity. Section 6 contains the simulation study. Section 7 contains the use-case. A Conclusion closes the paper.

2 Sensitivity analysis for expected function of stochastic model outputs

A stochastic model with inputs $(X_1, \dots, X_p) =: \mathbf{X} \in \mathbb{R}^p$ and output $Y \in \mathbb{R}$ is seen as a function f of \mathbf{X} and some collection of latent random variables Z such that

$$Y = f(\mathbf{X}, Z).$$

Throughout this paper, we assume that \mathbf{X} and Z are independent. In addition, the components X_1, \dots, X_p of \mathbf{X} are also assumed to be mutually independent. By a slight abuse of language, we sometimes refer to f as the stochastic model. Note that the stochastic model run with some particular input $\mathbf{X} = \mathbf{x}$ is represented by the real-valued random variable $f(\mathbf{x}, Z)$. Thus, the intrinsic randomness of the stochastic model comes from this random variable Z , which, for this reason, is referred to as the intrinsic randomness. To do the sensitivity analysis of such stochastic models, one approach consists in defining QoIs as conditional expectations. Since the conditional expectation is a function of the inputs alone, we apply the Sobol' decomposition to this function and define Sobol indices.

For a stochastic model f with inputs \mathbf{X} and intrinsic randomness Z , we consider a QoI Q of the form:

$$Q(\mathbf{x}) = \mathbb{E}[\varphi(\mathbf{X}, Z) \mid \mathbf{X} = \mathbf{x}],$$

where $\varphi(\mathbf{X}, Z)$ is a function of \mathbf{X} and Z . For instance, if $\varphi = f$ then $Q(\mathbf{X})$ is the conditional expectation of the model and if $\varphi(\mathbf{X}, Z) = (f(\mathbf{X}, Z) - \mathbb{E}[f(\mathbf{X}, Z) \mid \mathbf{X}])^2$ then $Q(\mathbf{X})$ is the conditional variance, two common choices in practice. Assume that $\text{Var}(Q(\mathbf{X})) < +\infty$. If u is a subset of $\{1, \dots, p\}$, let \mathbf{X}_u denote the group of inputs $\{X_i, i \in u\}$. Let $\mathbf{X}_{\sim u}$ denote the group of inputs $\{X_i, i \notin u\}$. The first-order and total

Sobol' indices S_u and T_u of \mathbf{X}_u are respectively defined as:

$$S_u = \frac{\text{Var}(\mathbb{E}[Q(\mathbf{X}) | \mathbf{X}_u])}{\text{Var}(Q(\mathbf{X}))}$$

$$T_u = 1 - \frac{\text{Var}(\mathbb{E}[Q(\mathbf{X}) | \mathbf{X}_{\sim u}])}{\text{Var}(Q(\mathbf{X}))} = 1 - S_{\sim u}.$$

The Sobol' index S_u (and hence T_u) can be expressed in terms of a function g linking the components of some parameter vector θ . Let $\tilde{\mathbf{X}}$ be an independent copy of \mathbf{X} , independent of Z . Denote by $\tilde{\mathbf{X}}_{\sim u}$ the subvector of $\tilde{\mathbf{X}}$ whose components are those of $\tilde{\mathbf{X}}$ not indexed by u . (For instance, if $p = 4$ and $u = \{1, 4\}$ then $\tilde{\mathbf{X}}_{\sim u} = (\tilde{X}_2, \tilde{X}_3)$.) If $\theta_u = (\theta_{u1}, \theta_{u2}, \theta_{u3})$ with $\theta_{u1} = \mathbb{E}(Q(\mathbf{X})^2) =: \theta_1$, $\theta_{u2} = \mathbb{E}(Q(\mathbf{X})) =: \theta_2$ and $\theta_{u3} = \mathbb{E}(\mathbb{E}[Q(\mathbf{X}) | \mathbf{X}_u]^2) = \mathbb{E}(Q(\mathbf{X})Q(\tilde{\mathbf{X}}_{\sim u}, \mathbf{X}_u)) = \mathbb{E}(Q(\mathbf{X}_{\sim u}, \mathbf{X}_u)Q(\tilde{\mathbf{X}}_{\sim u}, \mathbf{X}_u))$ then

$$S_u = g(\theta_u) = \frac{\theta_{u3} - \theta_2^2}{\theta_1 - \theta_2^2}. \quad (1)$$

Thus, the construction of an estimator of S_u boils down to the construction of estimators of the three quantities θ_1 , θ_2 and θ_{u3} . This is done in the next section.

3 Nested Monte Carlo estimation of Sobol' indices

An estimator of the Sobol' index (1) based on nested Monte Carlo sampling is constructed as follows. Let $\{\mathbf{X}^{(i)} : i = 1, \dots, n\}$ and $\{\tilde{\mathbf{X}}^{(i)} : i = 1, \dots, n\}$ be independent Monte Carlo samples from the law of \mathbf{X} . For each i , denote by $\mathbf{X}_u^{(i)}$ the subvector of $\mathbf{X}^{(i)}$ whose components are those of $\mathbf{X}^{(i)}$ indexed by u . Likewise, denote by $\mathbf{X}_{\sim u}^{(i)}$ the subvector of $\mathbf{X}^{(i)}$ whose components are those of $\mathbf{X}^{(i)}$ not indexed by u , and denote by $\tilde{\mathbf{X}}_{\sim u}^{(i)}$ the subvector of $\tilde{\mathbf{X}}^{(i)}$ whose components are those of $\tilde{\mathbf{X}}^{(i)}$ not indexed by u . An estimator of S_u in (1) is given by

$$\hat{S}_{u;n,m} = g(\hat{\theta}_{1;n,m}, \hat{\theta}_{2;n,m}, \hat{\theta}_{3;n,m}), \quad (2)$$

where

$$\left. \begin{aligned} \hat{\theta}_{1;n,m} &= \frac{1}{n} \sum_{i=1}^n \hat{Q}_m(\mathbf{X}^{(i)})^2 \\ \hat{\theta}_{2;n,m} &= \frac{1}{n} \sum_{i=1}^n \hat{Q}_m(\mathbf{X}^{(i)}) \\ \hat{\theta}_{3;n,m} &= \frac{1}{n} \sum_{i=1}^n \hat{Q}_m(\mathbf{X}^{(i)}) \tilde{Q}_m(\tilde{\mathbf{X}}_{\sim u}^{(i)}, \mathbf{X}_u^{(i)}) \end{aligned} \right\} \quad (3)$$

and

$$\hat{Q}_m(\mathbf{X}^{(i)}) = \frac{1}{m} \sum_{k=1}^m \varphi(\mathbf{X}^{(i)}, Z^{(i,k)})$$

$$\tilde{Q}_m(\tilde{\mathbf{X}}_{\sim u}^{(i)}, \mathbf{X}_u^{(i)}) = \frac{1}{m} \sum_{k=1}^m \varphi(\tilde{\mathbf{X}}_{\sim u}^{(i)}, \mathbf{X}_u^{(i)}, \tilde{Z}^{(i,k)});$$

here the objects $\{Z^{(i,k)}, \tilde{Z}^{(i,k)} : k = 1, \dots, m; i = 1, \dots, n\}$, are independent and identically distributed random variables, independent of $\{\mathbf{X}^{(i)}, \tilde{\mathbf{X}}^{(i)} : i = 1, \dots, n\}$, representing the randomness of the user's model. (The dependence on u is not represented in the notation $\hat{\theta}_{3;n,m}$ for clarity.) If m is fixed and $n \rightarrow \infty$, then

$$\sqrt{n} \left(\hat{S}_{u;n,m} - S_u \left[1 - \frac{\mathbb{E} \text{Var}(\varphi(\mathbf{X}, Z) | \mathbf{X})}{\mathbb{E} \text{Var}(\varphi(\mathbf{X}, Z) | \mathbf{X}) + m \text{Var} \mathbb{E}(\varphi(\mathbf{X}, Z) | \mathbf{X})} \right] \right) \quad (4)$$

converges to a centered normal distribution with some variance $\sigma_{u;m}^2$ depending on m . If, moreover, $m \rightarrow \infty$ such that $\sqrt{n}/m \rightarrow 0$ then $\sqrt{n}(\widehat{S}_{u;n,m} - S_u)$ goes to a centered normal distribution with variance $\lim_{m \rightarrow \infty} \sigma_{u;m}^2$. For more details, see Mazo (2021).

Since it is based on two nested Monte Carlo loops of sizes n and m , the estimator $\widehat{S}_{u;n,m}$ is called a nested Monte Carlo estimator. See Section E for some more details about nested Monte Carlo estimators. The sizes of the outer and inner loops, n and m , are called the numbers of explorations and repetitions, respectively. Of course, we would like that both n and m go to infinity as fast as possible. But remember that increasing n or m induces an increase of the number of model runs. Hence the following question: if we were allowed to increase the number of model runs by a fixed amount, should we increase n or m ? This raises the question of resource allocation between the number of explorations and the number of repetitions.

Mathematically, this problem can be formulated as follows. The total number of runs of the stochastic model needed to compute the Sobol index estimates corresponding to the p singletons, $u = \{1\}, \dots, \{p\}$, is given by $(p+1)nm =: (p+1)T$. By an abuse of language T is referred to as the computational budget, or budget for short. Let $m = T^\eta$, $0 \leq \eta < 1$, and $n = T^{1-\eta}$ and suppose that $T \rightarrow +\infty$, so that the numbers of explorations and repetitions are allowed to grow with the budget with different rates, both controlled by the parameter η . What value of η should we choose? A theoretically-guided choice for m and n that penalizes bad rankings of the sensitivity index estimates $\widehat{S}_{\{1\};n,m}, \dots, \widehat{S}_{\{p\};n,m}$ was given in Mazo (2021). The next section proposes a theoretically guided choice based on the MSE.

4 Bounding the mean-squared error through regularization

Let $\widehat{\theta} = (\widehat{\theta}_1, \widehat{\theta}_2, \widehat{\theta}_3)$ be the vector with components given in (3) and let $\widehat{S} = \widehat{g}(\widehat{\theta})$ be the Sobol' index estimator in (2). (For simplicity of notation, we simply write $\widehat{\theta}$ instead of $\widehat{\theta}_{n,m}$ and more generally omit the subscripts n , m and u in the notation in the absence of ambiguity.) The MSE of the Sobol' index estimator is given by $\mathbb{E}(g(\widehat{\theta}) - g(\theta))^2$.

Bounding the MSE of an estimator of the form $g(\widehat{\theta})$ depends on properties of the function g . For instance, if g is Lipschitz continuous then an analysis of the MSE similar to Section E can be done. But here g is not Lipschitz continuous and bounding the MSE is more difficult. Let μ_m denote the expectation of $\widehat{\theta}$. It can be checked that μ_m does not depend on n . Let $\nabla^2 g(\theta)$ denote the Hessian matrix of the function g at θ , that is, the matrix with element at the i th row and j th column given by $(\nabla^2 g)_{ij}(\theta) = (\partial^2 g)/(\partial \theta_i \partial \theta_j)(\theta)$. One idea to bound the MSE consists in splitting the difference $g(\widehat{\theta}) - g(\theta)$ into a variance term $g(\widehat{\theta}) - g(\mu_m)$ and a bias term $g(\mu_m) - g(\theta)$ and calculate Taylor expansions. For instance, $g(\widehat{\theta}) - g(\mu_m) = (\widehat{\theta} - \mu_m)^\top \nabla g(\mu_m) + \frac{1}{2}(\widehat{\theta} - \mu_m)^\top \nabla^2 g(\lambda \widehat{\theta} + (1-\lambda)\mu_m)(\widehat{\theta} - \mu_m)$, where $0 < \lambda < 1$. Taking the square and then the expectation in both sides, it follows that

$$\begin{aligned} \mathbb{E}(g(\widehat{\theta}) - g(\mu_m))^2 &\leq \\ 2\|\nabla g(\mu_m)\|_2^2 \mathbb{E}\|\widehat{\theta} - \mu_m\|_2^2 &+ \frac{1}{2} \mathbb{E}\|\widehat{\theta} - \mu_m\|_2^4 \|\nabla^2 g(\lambda \widehat{\theta} + (1-\lambda)\mu_m)\|_F^2, \end{aligned} \quad (5)$$

where $\|\cdot\|_F$ denotes the Fröbenius norm. However, without further condition, it is difficult to guarantee that the second term in the right-hand side is finite.

One idea to overcome this problem is to regularize the estimator. More precisely, we shall add a constant $h > 0$ in the denominator of (1). That is, we shall consider the estimator $\widehat{S}(h) = g_h(\widehat{\theta})$, where $g_h(\theta_1, \theta_2, \theta_3) = (\theta_3 - \theta_2^2)/(\theta_1 - \theta_2^2 + h)$. This allows to control the second-order term in the Taylor expansion as shown by Lemma 1.

Lemma 1. *If $\mathbb{E}(Q(\mathbf{X})^{16}) < +\infty$ then, for each $h > 0$,*

$$\limsup_{n,m \rightarrow \infty} \mathbb{E} \left(\sup_{\lambda \in [0,1]} \|\nabla^2 g_h(\lambda \widehat{\theta} + (1-\lambda)\mu_m)\|_{\mathbb{F}}^4 \right) \leq c_1 + c_2 h^{c_3} =: C(h),$$

where c_1, c_2, c_3 are some positive constant.

The condition $\mathbb{E}(Q(\mathbf{X})^{16}) < +\infty$ is a technical requirement ensuring that all expectations in the proof are finite. Pursuing the calculations started in (5) yields Theorem 1.

Theorem 1. *Under the conditions of Lemma 1, for each $h > 0$, it holds*

$$\mathbb{E} \left(\widehat{S}(h) - S \right)^2 = O \left(\frac{1}{n} + \frac{1}{m^2} + h^2 \right),$$

as $n, m \rightarrow \infty$.

Corollary 1. *Under the conditions of Lemma 1, we have $\lim_{h \rightarrow 0} \lim_{n,m \rightarrow \infty} \mathbb{E}(\widehat{S}(h) - S)^2 = 0$.*

Remember that we assume that $m = T^\eta, n = T^{1-\eta}, 0 \leq \eta < 1$, and that $T \rightarrow \infty$. With $m = T^\eta$ and $n = T^{1-\eta}$, we have that the MSE is $O(T^{1-\eta} + T^{-2\eta} + h^2)$ as $T \rightarrow \infty$. To balance the trade-off between the number of repetitions and explorations, we proceed with a heuristic. For h small enough, one may consider that h^2 is negligible with respect to $T^{1-\eta} + T^{-2\eta}$ and seek the value of η that maximizes the rate at which $T^{1-\eta} + T^{-2\eta}$ vanishes. As seen in Section E, this value is $\eta = 1/3$ and yields $\mathbb{E}(\widehat{S}(h) - S)^2 = O(T^{-2/3})$.

5 Letting the regularization parameter vanish

In Section 4 the regularization parameter h is arbitrarily fixed and the MSE of the resulting estimator is asymptotically less than some quantity proportional to h^2 . Thus Theorem 1 holds for h positive, but the bound is minimized when h is zero. In practice, it is clear that the more the values of m and n , the less h should be. One way to address this issue is to let h go to zero as m and n go to infinity. If the rate at which h goes to zero can be calculated in comparison to the rate at which m and n go to infinity then some hint of how small h should be in practice could be obtained. Next, the regularization parameter is no longer fixed but is a sequence $(h_n)_{n \geq 1}$ indexed by the outer loop sample size n such that $h_n \rightarrow 0$ as $n \rightarrow +\infty$.

Lemma 2. *If $\mathbb{E}(Q(\mathbf{X})^{16}) < +\infty$ then:*

(i) *there exists an universal constant C' such that:*

$$\limsup_{n,m \rightarrow \infty} h_n^{12} \mathbb{E} \left(\sup_{\lambda \in [0,1]} \|\nabla^2 g_{h_n}(\lambda \widehat{\theta} + (1-\lambda)\mu_m)\|_{\mathbb{F}}^4 \right) \leq C';$$

(ii) in addition:

$$\mathbb{E} \left(g_{h_n}(\hat{\theta}) - g(\theta) \right)^2 = O \left(\frac{1}{n} + \frac{1}{m^2} + \frac{1}{n^2 h_n^6} + h_n^2 \right),$$

as $n, m \rightarrow +\infty$.

From Lemma 2, it follows that $\mathbb{E}(\sup_{\lambda \in [0,1]} \|\nabla^2 g_{h_n}(\lambda \hat{\theta} + (1-\lambda)\mu_m)\|_{\mathbb{F}}^4)$ can diverge but at a rate which does not exceed $O(h_n^{-12})$. Note that the term $1/(n^2 h_n^6)$ is decreasing in h_n while the term h_n^2 is increasing in h_n . By assuming that h_n is a negative power of n , we can determine the optimal compromise between $1/(n^2 h_n^6)$ and h_n^2 which is $h_n^* := n^{-1/4}$. We deduce the optimal regularized estimator $\hat{S}(h_n^*) = g_{h_n^*}(\hat{\theta})$ in Theorem 2.

Theorem 2. *Under the assumption of Lemma 2, it holds $\mathbb{E}(\hat{S}(h_n^*) - S)^2 = O(1/m^2 + 1/\sqrt{n})$.*

The proof of Theorem 2 is straightforward using $h_n = h_n^*$ in Lemma 2. It immediately follows that the bias and the MSE of $\hat{S}(h_n^*)$ converges to 0 as $n, m \rightarrow +\infty$. Let us calculate the bias-variance trade-off as in Section 4. Under the constraint $T = nm$ with $m = T^\eta$ and $n = T^{1-\eta}$, solving the problem of optimal balance between the number of explorations and the number of repetitions results in $n_{opt} = T^{4/5}$ and $m_{opt} = T^{1/5}$ and $h_{opt} = T^{-1/5}$. The convergence rate of the bound of the MSE becomes $T^{2/5}$. This rate is slower than $T^{2/3}$ obtained in Section 4. The decrease in the convergence rate can be viewed as a consequence of allowing $\mathbb{E}(\sup_{\lambda \in [0,1]} \|\nabla^2 g_{h_n}(\lambda \hat{\theta} + (1-\lambda)\mu_m)\|_{\mathbb{F}}^4)$ to diverge.

It is important to note that the rate of the optimal regularization parameter is optimal for the bound, but not necessarily for the MSE itself. This is further discussed in Section 8.

6 Simulation study

The behavior of the nested Monte Carlo Sobol' index estimator of Section 3 and 4 is investigated through a simulation study. Two models are considered: a linear model $f(X_1, X_2, Z) = X_1 + \beta X_2 + \sigma Z$ where X_1, X_2 and Z are standard normal independent random variables, and a nonlinear model $f(X_1, X_2, Z) = \sin(X_1) + \beta \sin(X_2)^2 + \sigma Z^4 \sin(X_1)$. The last model is known as the Ishigami function (Ishigami and Homma, 1990) and commonly used as a benchmark in sensitivity analysis. In both models β and σ represent known coefficients and X_1 and X_2 represent two uncertain parameters. The random variable Z represents the intrinsic randomness of the models. In the linear model $\beta = 1.2$ and in the nonlinear model $\beta = 7$. The coefficient σ was set to 0.1 in the nonlinear model and in the linear one two values were considered: $\sigma = 4$ and $\sigma = 0.9$, representing a high noise and low noise context, respectively. These values were those of Mazo (2021).

The budget was set to $T = 10^2, 5 \times 10^2, 2 \times 10^3, 5 \times 10^3, 10^4, 10^5$. Since there are $p = 2$ inputs, this coincides with running the model $(p+1)T = 3T$ times, for each value of T . For each budget, the four scenarios $m = 5, T^{1/3}, T^{2/5}, T^{1/2}$ were tested. In the first scenario, the number of repetitions was set to an arbitrary constant and in the last three scenarios the number of repetitions increases with the budget, albeit with different rates. The following values for the regularization parameter h were tested:

$h = 0, 0.5, 1$, and $h = n^{-1/4}$ as in Section 5. For each combination (T, m) , an estimate of the vector of Sobol' indices of the two uncertain parameters was and the mean-squared error were computed. The numbers of replications were set to 500, 500, 150, 60, 30 and 3 for the successive values of the budget, respectively. (The decrease in the number of replications as the budget increases takes into account the reduction in variability of the estimator.)

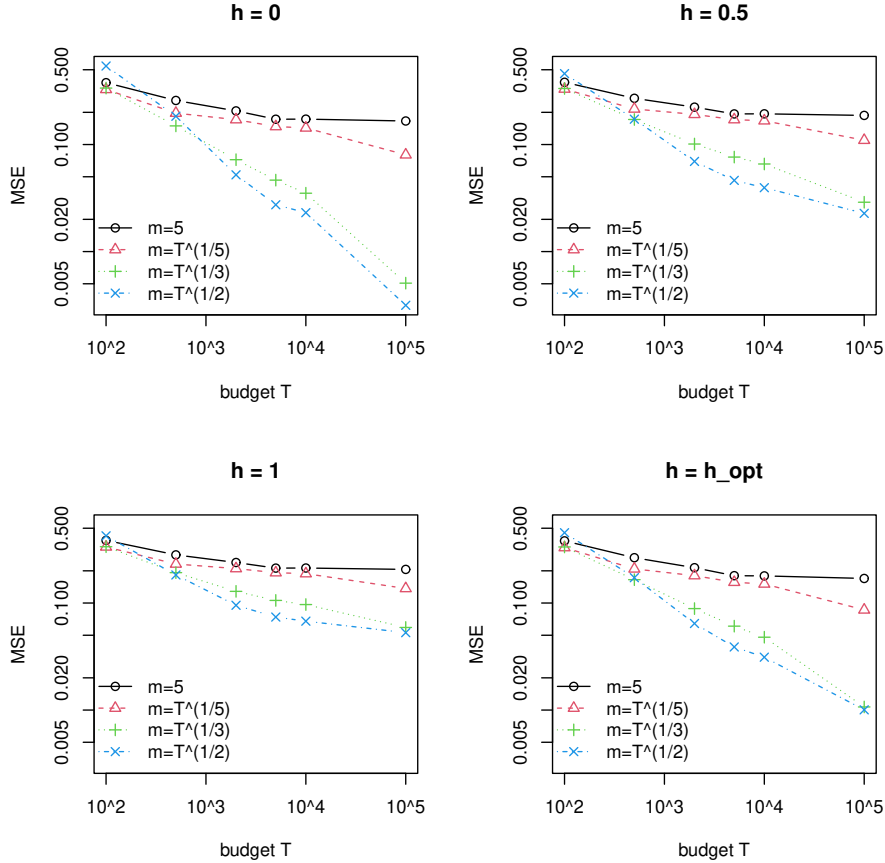


Figure 1: Estimated MSE (logarithmic scale) of the vector of Sobol' indices for the linear model in the high noise context

The results are depicted in Figure 1 (linear model, high noise), Figure 2 (linear model, low noise) and Figure 3 (Ishigami). The label h_{opt} corresponds to the case $h = n^{-1/4}$. In all cases the estimated MSE decreases as the budget increases. In the case $m = 5$ in the high noise linear model the MSE stabilizes because the estimator is asymptotically biased unless m grows with the budget. In the low noise linear model the stabilization is not apparent because it occurs for larger budgets. The results for the Ishigami model are similar to the low noise context of the linear model. The curve of the MSE for $m = T^{1/5}$ closely follows $m = 5$ and is only slightly better at $T = 10^5$. The curves of the MSEs under the rates $m = T^{1/3}, T^{1/2}$ stay close to each other: the estimator seems to be robust to the choice of the allocation between the numbers of

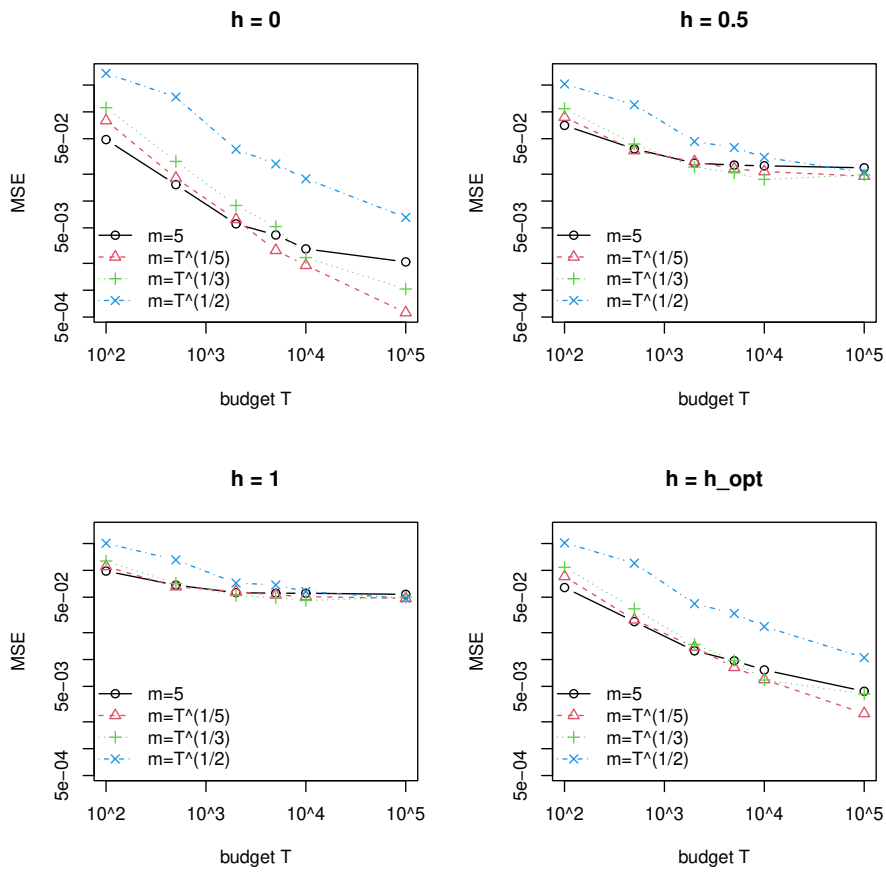


Figure 2: Estimated MSE (logarithmic scale) of the vector of Sobol' indices for the linear model in the low noise context

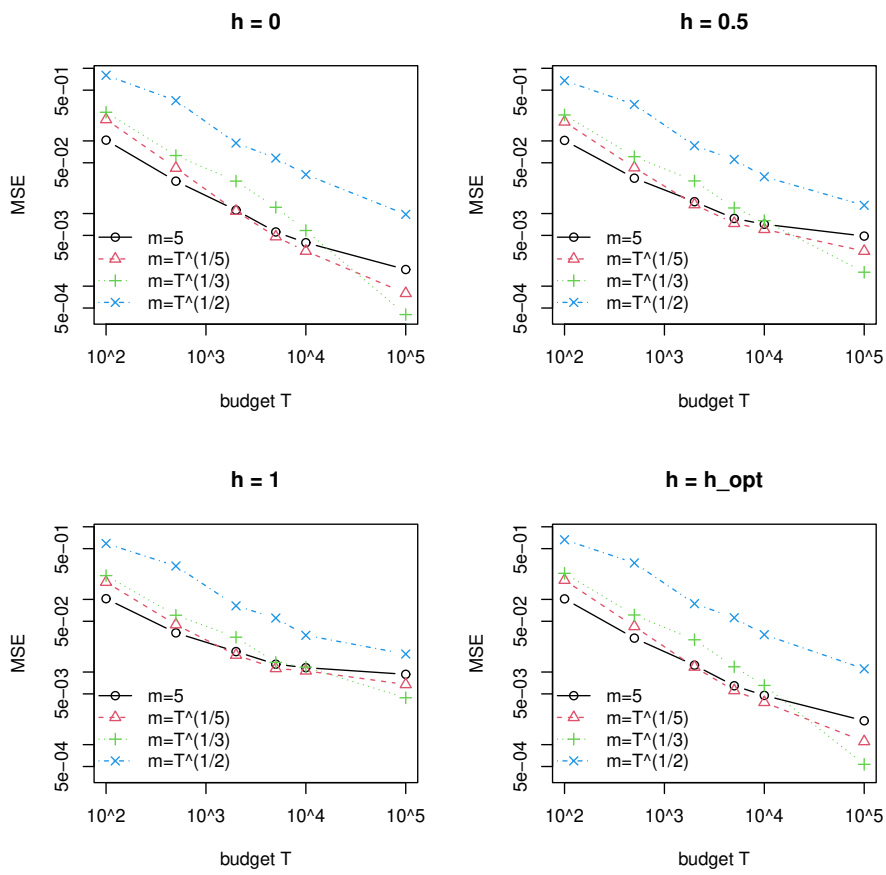
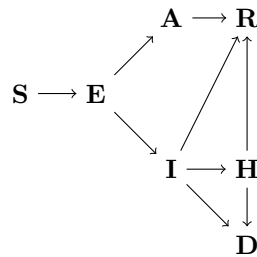


Figure 3: Estimated MSE (logarithmic scale) for the Ishigami model

repetitions and explorations. All these observations agree with those in Mazo (2021). We also observe that stabilization occurs for fixed positive values of h ; the greater the value of h , the higher the value near which the MSE stabilizes. The best performance of the estimator is reached when h is null. The results for $h = h_{\text{opt}} = n^{-1/4}$ and $h = 0$ are close, however. The optimal performance of the case $h = 0$ suggests that the non-regularized Sobol' index nested Monte Carlo estimator may have a MSE converging faster, although a proof of this remains to be discovered. This is further discussed in Section 8.

7 Application to a stochastic COVID-19 epidemic compartmental model

As an illustrative use-case, the sensitivity analysis of a stochastic COVID-19 epidemic compartmental model is performed. The model considered has 7 compartments—Susceptibles (S), Exposed (E), Asymptomatic (A), Infectious (I), Hospitalized (H), Recovered (R), Dead (D)—and each of the 2005 individuals moves to one compartment to another according to the diagram below:



Individuals move at random times according to a continuous time Markov chain which records the number of individuals in each compartment. The stochastic model consists of simulating a trajectory of the Markov chain through Gillespie's First Reaction Method (Gillespie, 1976) and then computing the extinction time of the chain, that is, the time at which there is no more individuals in any of the compartments A, E, and I. The law of the Markov chain is governed by nine parameters, given in Table 1. The QoI is the expected extinction time, conditionally on the parameters. An histogram of the extinction times observed from 10^4 simulated trajectories of a Markov chain with parameters at the nominal values is depicted in Figure 4.

A total budget $T = 216000$ per Sobol' index was set, leading to $216000 \times 9 = 1944000$ model calls. Two settings were considered for the numbers of repetitions and explorations. The first is $m = 60$ and $n = 3600$ and the second is $m = 12$ and $n = 18000$, corresponding to the cases $m = T^{1/3}$ and $m = T^{1/5}$ as found in Section 4 and Section 5, respectively. In view of the results of the simulation study in Section 6, h was set to zero.

Parameter	Description	Nominal value	Range of variation
β	virus transmission rate	2.175	(0.35, 4)
$1/\gamma_E$	mean sojourn duration in E	4.5 days	(2, 7)
$1/\gamma_A$	mean sojourn duration in A	2 days	(1, 3)
$1/\gamma_I$	mean sojourn duration in I	4 days	(3, 5)
$1/\gamma_H$	mean sojourn duration in H	9.5 days	(7, 12)
$p_{(E,A)}$	prob. of becoming A given E	0.5	(0.3, 0.7)
p_C	prob. of becoming H or D given I	0.175	(0.15, 0.2)
$p_{D C}$	prob. of being D given H or D	0.175	(0.15, 0.2)
$p_{(H,D)}$	prob. of becoming D given H	0.0505	(0.001, 0.1)

Table 1: Parameters of the COVID-19 stochastic model

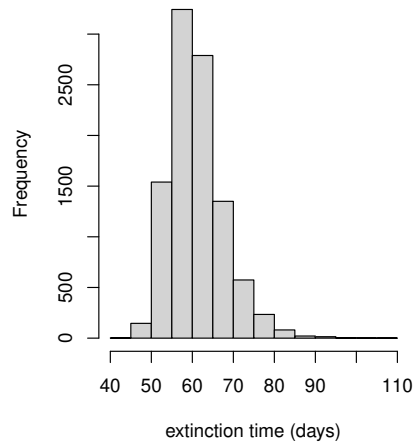


Figure 4: Histogram of the extinction times observed from 10^4 simulated trajectories of a Markov chain with parameters the nominal values given in Table 1

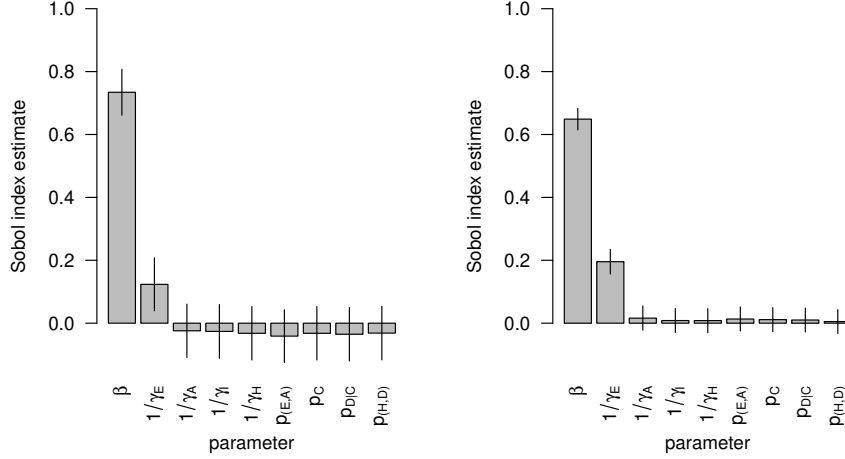


Figure 5: Barplot of the Sobol' index estimates of the nine parameters of the COVID-19 stochastic model. Left: $m = 60$ and $n = 3600$; right: $m = 12$ and $n = 18000$. Error bars are approximate confidence intervals of level 90% based on the asymptotic normality (4) and calculated as the estimate plus/minus (quantile of order 95% of a standard normal distribution) times (square root of the estimated asymptotic variance) divided by \sqrt{n} . The asymptotic variance is calculated from the delta-method, see (Mazo, 2021)

The Sobol' index estimates are depicted in Figure 5. We see that the values of the Sobol' index estimates in the left and right panels are similar: the observed differences are within margin of error. We also see that the error bars are narrower in the right panel. This was expected because the estimates in the right panel are constructed with more explorations than the estimates in the left panel. But notice from (4) that, since less repetitions are used, the estimates in the right panel, in average, tend to underestimate the true values to a greater extent than those in the left panel. In both panels, however, the conclusions are the same: we see that the extinction time of the COVID-19 epidemic is most sensitive to β , the virus transmission rate. Then comes $1/\gamma_E$, the mean sojourn duration in E, albeit with a much smaller effect. All of the other parameters have negligible effects. Note that the Sobol' index estimator constructed in Section 3 can be negative, which indeed may happen when the true Sobol' index is small or when there is insufficient data.

8 Discussion

The bound in Theorem 2 is optimal at a value of the regularization parameter h_n greater than zero, while it was found in the simulation experiments of Section 6 that $h_n = 0$ performed best. This apparent contradiction is discussed next.

The mathematical method used to derive the bound in Theorem 2 is general: it can be applied to every twice differentiable function g and probability law P_n such that Lemma 2 holds with $\hat{\theta} \sim P_n$, where in general $g_h(\theta) := g(\theta + h\mathbf{u})$ for some direction

vector $\mathbf{u} \in \mathbb{R}^p$. (We write h instead of h_n for simplicity.) Note that, with such a wide scope of application, regularization is necessary. Otherwise, the MSE may be infinite, as shows Example 1.

Example 1. Let $a > 0$. Let $\hat{\theta} = B_n X + (1 - B_n) \frac{a+1}{a+2}$ where B_n is a Bernoulli random variable with $\text{pr}(B_n = 1) = \frac{1}{n}$ and X is a continuous random variable with density $f(x) = (a+1)x^a$ over $0 < x < 1$, independent of B_n . (This specifies P_n completely.) Note that $\mathbb{E}(\hat{\theta}) = (a+1)/(a+2)$. Let $g(\theta) = 1/\theta$ for $0 < \theta < 1$. Put $a = 1$. Elementary calculations (see Section D) show that $\mathbb{E}(g_h(\hat{\theta})^2) = (1 - \frac{1}{n})g_h(2/3)^2 + \frac{2}{n}(\log(1 + 1/h) - 1/(h+1))$, and hence $\mathbb{E}((g_0(\hat{\theta}) - g(2/3))^2) = \infty$ for every n .

Remark 1. In sensitivity analysis, the MSE can be infinite, too. An example is given by discrete outputs: in this case there is a nonnull probability that all outputs of the model be the same, leading to a variance estimate equal to zero in the denominator of the Sobol' index estimator. Of course, this happens with decreasing probability as the size of the Monte Carlo sample increases.

The effect of regularization depends on both the function g and the law P_n . In some cases this effect is beneficial, in some others it is detrimental. See Example 2.

Example 2. Let g and $\hat{\theta}$ be as in Example 1. If $a = 1$ then elementary calculations (see Section D) yield that the bias $\mathbb{E}(g_h(\hat{\theta})) - g(2/3)$ is asymptotically equivalent to $-\frac{9}{4}h - \frac{1}{2}n^{-1}$ and that the variance is asymptotically equivalent to $-2n^{-1} \log h$, leading to $\mathbb{E}((g_h(\hat{\theta}) - g(2/3))^2) \sim \frac{81}{16}h^2 + \frac{1}{4}n^{-2} + \frac{9}{4}hn^{-1} - 2n^{-1} \log h$. The optimal regularization parameter is $h = 1/\sqrt{n}$, yielding a MSE asymptotically equivalent to $n^{-1} \log n$. But if $a = 2$ then the bias and the variance are asymptotically equivalent to $-\frac{16}{3}h + \frac{1}{6}n^{-1}$ and $\frac{15}{9}n^{-1}$, respectively. The optimal regularization is then $h = 0$, that is, no regularization.

Theorem 2 applied to the case $a = 2$ in Example 2—that is, doing the calculations in the displayed inequality in the proof of Lemma 2 (ii) with $\mu_m = \theta = 3/4$, $C' = 16$ and $\hat{\theta}$ and g as in Example 2—predicts that $\mathbb{E}((g_h(\hat{\theta}) - g(3/4))^2)$ is bounded by $O(n^{-1/2}h^{-6} + h^2)$, which is optimized at $h = n^{-1/16}$, leading to a bound of order $O(n^{-1/8})$. Thus, the case $a = 2$ in Example 2 demonstrates that there are functions g and laws P_n such that the bound in Theorem 2 can be improved.

The simulation experiments of Section 6 suggest that, in sensitivity analysis with $g(\theta)$ specified as in (1), there is a class of laws P_n such that the MSE is optimized at $h = 0$. Thus, as in Example 2, it is reasonable to conjecture that sharper bounds for the Sobol' index estimator optimized at $h = 0$ could be obtained by imposing some conditions on the model and/or the law under which of the Monte Carlo sample is drawn. This is left as future research.

9 Conclusion

To balance the tradeoff between the number of repetitions m and the number of explorations n in the estimation of Sobol' indices for stochastic models, a regularization parameter h was added to the nested Monte Carlo estimator. The MSE of the regularized estimator was theoretically bounded. Under the asymptotic setting $m = T^\eta$, $n = T^{1-\eta}$ and $T \rightarrow \infty$ the convergence rate of the bound was optimized for h fixed and then for $h_n \rightarrow 0$, yielding for the allocation problem two different solutions $m = T^{1/3}$ and

$m = T^{1/5}$, respectively. A simulation study showed that the first solution with $h = 0$ performs better, suggesting that regularization, while useful for obvious technical reasons, might not improve performance after all (see also Section 8 for a discussion about this observation). The simulation study also showed that the performance is robust with respect to the choice of m and n . As an illustrative use-case, the sensitivity analysis of a COVID-19 epidemic model with nine parameters was performed under the two different solutions with $h = 0$.

As a future work, it could be interesting to work on establishing the convergence rate of the non-regularized estimator. It would be of interest to compare the results with multilevel Monte Carlo methods (Mycek and De Lozzo (2019); Giles and Haji-Ali (2019)) or to adapt the iterative estimation approach of Gilquin et al. (2021). Another estimation approach worth considering is Chatterjee's empirical correlation coefficient (Chatterjee, 2021; Gamboa et al., 2022). Also, since the quantities $\theta_1 = \mathbb{E}(Q(\mathbf{X})^2)$ and $\theta_2 = \mathbb{E}(Q(\mathbf{X}))$ are shared across all Sobol' indices $S_u, u = \{1\}, \dots, \{p\}$, estimators of these quantities using all of the simulations available could be considered. Finally, it would be interesting to extend the present work to a framework where the latent random variables and the random inputs would not be independent.

Acknowledgment

We thank two anonymous reviewers for their detailed and constructive comments which helped to improve the present manuscript. We also thank Clémentine Prieur for her suggestions on an early version.

A Proofs

Proof of Lemma 1

Note that the function g is infinitely differentiable over its domain \mathcal{D} and then its gradient is given by:

$$\nabla g(\theta_1, \theta_2, \theta_3) = \left(-\frac{\theta_3 - \theta_2^2}{(\theta_1 - \theta_2^2)^2}, \frac{2\theta_2(\theta_3 - \theta_1)}{(\theta_1 - \theta_2^2)^2}, \frac{1}{\theta_1 - \theta_2^2} \right)^\top.$$

Furthermore, the hessian matrix of g yields:

$$\nabla^2 g(\theta_1, \theta_2, \theta_3) = \begin{pmatrix} \frac{2(\theta_3 - \theta_2^2)}{(\theta_1 - \theta_2^2)^3} & \frac{2\theta_2(\theta_1 - 2\theta_3 + \theta_2^2)}{(\theta_1 - \theta_2^2)^3} & \frac{-1}{(\theta_1 - \theta_2^2)^2} \\ \frac{2\theta_2(\theta_1 - 2\theta_3 + \theta_2^2)}{(\theta_1 - \theta_2^2)^3} & \frac{2(\theta_3 - \theta_1)(\theta_1 + 3\theta_2^2)}{(\theta_1 - \theta_2^2)^3} & \frac{2\theta_2}{(\theta_1 - \theta_2^2)^2} \\ \frac{-1}{(\theta_1 - \theta_2^2)^2} & \frac{2\theta_2}{(\theta_1 - \theta_2^2)^2} & 0 \end{pmatrix}.$$

For any $(\theta_1, \theta_2, \theta_3) \in \mathcal{D}$, the matrix $\nabla^2 g(\theta_1, \theta_2, \theta_3)$ is under the form $\nabla^2 g(\theta_1, \theta_2, \theta_3) = B(\theta_1, \theta_2, \theta_3)/(\theta_1 - \theta_2^2)^3$ where $B(\theta_1, \theta_2, \theta_3)$ is the matrix:

$$B(\theta_1, \theta_2, \theta_3) = \begin{pmatrix} 2(\theta_3 - \theta_2^2) & 2\theta_2(\theta_1 - 2\theta_3 + \theta_2^2) & -(\theta_1 - \theta_2^2) \\ 2\theta_2(\theta_1 - 2\theta_3 + \theta_2^2) & 2(\theta_3 - \theta_1)(\theta_1 + 3\theta_2^2) & 2\theta_2(\theta_1 - \theta_2^2) \\ -(\theta_1 - \theta_2^2) & 2\theta_2(\theta_1 - \theta_2^2) & 0 \end{pmatrix}.$$

Notice that $B(\theta_1, \theta_2, \theta_3)$ includes only multivariate polynomials of variables θ_1, θ_2 and θ_3 .

Let $\lambda \in [0, 1]$ and $h \in (0, 1)$. Without loss of generality we write $g_h(\theta) = g(\theta + h\mathbf{u})$ with $\mathbf{u} = (1, 0, 0)$. We have

$$\nabla^2 g(\lambda\hat{\theta} + (1-\lambda)\mu_m + h\mathbf{u}) = \frac{B(\lambda\hat{\theta} + (1-\lambda)\mu_m + h\mathbf{u})}{\left(h + \lambda\hat{\theta}_1 + (1-\lambda)\mu_{m1} - (\lambda\hat{\theta}_2 + (1-\lambda)\mu_{m2})^2\right)^3}.$$

Thus:

$$\begin{aligned} & \mathbb{E} \left(\sup_{\lambda} \|\nabla^2 g(\lambda\hat{\theta} + (1-\lambda)\mu_m + h\mathbf{u})\|_F^4 \right) \\ &= \mathbb{E} \left(\sup_{\lambda} \frac{\|B(\lambda\hat{\theta} + (1-\lambda)\mu_m + h\mathbf{u})\|_F^4}{\left(h + \lambda\hat{\theta}_1 + (1-\lambda)\mu_{m1} - (\lambda\hat{\theta}_2 + (1-\lambda)\mu_{m2})^2\right)^{12}} \right) \\ &\leq \sqrt{\mathbb{E} \left(\sup_{\lambda} \frac{1}{\left(h + \lambda\hat{\theta}_1 + (1-\lambda)\mu_{m1} - (\lambda\hat{\theta}_2 + (1-\lambda)\mu_{m2})^2\right)^{24}} \right)} \\ &\quad \times \sqrt{\mathbb{E} \left(\sup_{\lambda} \|B(\lambda\hat{\theta} + (1-\lambda)\mu_m + h\mathbf{u})\|_F^8 \right)}. \end{aligned} \quad (6)$$

Since the function $x \mapsto x^2$ is convex, it holds that $(\lambda\hat{\theta}_2 + (1-\lambda)\mu_{m2})^2 \leq \lambda\hat{\theta}_2^2 + (1-\lambda)\mu_{m2}^2$ and hence $h + \lambda\hat{\theta}_1 + (1-\lambda)\mu_{m1} - (\lambda\hat{\theta}_2 + (1-\lambda)\mu_{m2})^2 \geq h + \lambda(\hat{\theta}_1 - \hat{\theta}_2^2) + (1-\lambda)(\mu_{m1} - \mu_{m2}^2)$. Since $\hat{\theta}_1 - \hat{\theta}_2^2$ and $\mu_{m1} - \mu_{m2}^2$ are nonnegative, it follows that the first factor in the right-hand side of (6) is less than

$$\sqrt{\mathbb{E} \left(\sup_{\lambda} \frac{1}{\left(h + \lambda(\hat{\theta}_1 - \hat{\theta}_2^2) + (1-\lambda)(\mu_{m1} - \mu_{m2}^2)\right)^{24}} \right)}.$$

Since $h + \lambda(\hat{\theta}_1 - \hat{\theta}_2^2) + (1-\lambda)(\mu_{m1} - \mu_{m2}^2) = \lambda(h + \hat{\theta}_1 - \hat{\theta}_2^2) + (1-\lambda)(h + \mu_{m1} - \mu_{m2}^2)$ and since the function $x \mapsto 1/x^{24}$ is convex, a similar argument yields the bound

$$\begin{aligned} & \sqrt{\mathbb{E} \left(\sup_{\lambda} \frac{\lambda}{\left(h + (\hat{\theta}_1 - \hat{\theta}_2^2)\right)^{24}} \right) + \sup_{\lambda} \frac{1-\lambda}{\left(h + \mu_{m1} - \mu_{m2}^2\right)^{24}}} \\ &\leq \sqrt{\mathbb{E} \left(\frac{1}{\left(h + (\hat{\theta}_1 - \hat{\theta}_2^2)\right)^{24}} \right) + \frac{1}{\left(h + \mu_{m1} - \mu_{m2}^2\right)^{24}}} \\ &\leq \sqrt{\mathbb{E} \left(\frac{1}{\left(h + (\hat{\theta}_1 - \hat{\theta}_2^2)\right)^{24}} \right) + \sup_{h \geq 0} \frac{1}{\left(h + \mu_{m1} - \mu_{m2}^2\right)^{24}}}. \end{aligned}$$

One should remark that $\sup_{h \geq 0} \frac{1}{(h + \mu_{m1} - \mu_{m2}^2)^{24}} \leq \frac{1}{(\text{Var}(\mathbb{E}[\varphi(\mathbf{X}, Z) | \mathbf{X}]))^{24}} < +\infty$. Moreover, by relying on Lemma B.1, $\mathbb{E}(h + \hat{\theta}_1 - \hat{\theta}_2^2)^{-24}$ is a bounded by $\frac{1}{(\text{Var}(\mathbb{E}[\varphi(\mathbf{X}, Z) | \mathbf{X}]))^{24}}$ as $n, m \rightarrow +\infty$. Finally, let us deal with the second factor in the right-hand side of (6). Since the matrix B comprises only polynomials and since $\mathbb{E}(Q(\mathbf{X}))^{16} < +\infty$, we have by Lemma B.2 that

$$\limsup_{n, m \rightarrow \infty} \mathbb{E} \left(\sup_{\lambda \in [0, 1]} \|B(\lambda \hat{\theta} + (1 - \lambda)\mu_m + h\mathbf{u})\|_F^8 \right)$$

is bounded by some polynomial in h . The proof is complete.

Proof of Theorem 1

Let $\nabla g_h(\theta)$ denote the gradient of the function g_h at θ . We have that $\mathbb{E}(g_h(\hat{\theta}) - g(\theta))^2 \leq 2\mathbb{E}(g_h(\hat{\theta}) - g_h(\mu_m))^2 + 2(g_h(\mu_m) - g(\theta))^2$. Let us deal with the first term in the right-hand side first. A Taylor expansion leads to

$$\begin{aligned} & \mathbb{E}(g_h(\hat{\theta}) - g_h(\mu_m))^2 \\ & \leq 2\mathbb{E}\{\|\hat{\theta} - \mu_m\|_2^2 \|\nabla g_h(\mu_m)\|_2^2\} + \frac{1}{2}\mathbb{E}\{\|\hat{\theta} - \mu_m\|_2^4 \|\nabla^2 g_h(\lambda \hat{\theta} + (1 - \lambda)\mu_m)\|_F^2\} \\ & \leq 2\sqrt{\mathbb{E}\|\hat{\theta} - \mu_m\|_2^4} \|\nabla g_h(\mu_m)\|_2^2 \\ & \quad + \frac{1}{2}\sqrt{\mathbb{E}\|\hat{\theta} - \mu_m\|_2^8} \sqrt{\mathbb{E}\|\nabla^2 g_h(\lambda \hat{\theta} + (1 - \lambda)\mu_m)\|_F^4} \end{aligned}$$

for some $0 < \lambda < 1$. Since ∇g_h is continuous and $\mu_m \rightarrow \mu$ as $m \rightarrow \infty$, we have by the Marcinkiewicz-Zygmund inequality that the first term in the right-hand side is $O(1/n)$. By Lemma 1 the second factor in the second term is less than $C(h)^2 = O(1)$. Thus, again by the Marcinkiewicz-Zygmund inequality, the second term is $O(1/n^2) = o(1/n)$. Thus $\mathbb{E}(g_h(\hat{\theta}) - g_h(\mu_m))^2 = O(1/n)$.

Next we have

$$(g_h(\mu_m) - g(\theta))^2 \leq 2(g_h(\mu_m) - g_h(\theta))^2 + 2(g_h(\theta) - g(\theta))^2.$$

The first term in the right-hand side is less than $2\|\mu_m - \theta\|_2^2 \|\nabla g_h(\lambda \mu_m + (1 - \lambda)\theta)\|_2^2$ for some $0 < \lambda < 1$. Elementary calculations show that $\|\mu_m - \theta\|_2^2$ is $O(1/m^2)$. The continuity of ∇g_h yields $\|\nabla g_h(\lambda \mu_m + (1 - \lambda)\theta)\|_2^2 = O(1)$. Elementary calculations show that $(g_h(\theta) - g(\theta))^2 = O(h^2)$. The proof is complete.

Proof of Lemma 2

Assume $\mathbb{E}(Q(\mathbf{X}))^{16} < +\infty$. As in in the proof of Lemma 1 in Section A, we let $\mathbf{u} = (1, 0, 0)$ so that $g_{h_n}(\theta) = g(\theta + h_n \mathbf{u})$ for all h and θ .

Proof of (i). We have:

$$\begin{aligned}
& \mathbb{E} \left(\sup_{\lambda} \|\nabla^2 g(\lambda\widehat{\theta} + (1-\lambda)\mu_m + h_n \mathbf{u})\|_{\mathbb{F}}^4 \right) \\
&= \mathbb{E} \left(\sup_{\lambda} \frac{\|B(\lambda\widehat{\theta} + (1-\lambda)\mu_m + h_n \mathbf{u})\|_{\mathbb{F}}^4}{\left(h_n + \lambda\widehat{\theta}_1 + (1-\lambda)\mu_{m1} - (\lambda\widehat{\theta}_2 + (1-\lambda)\mu_{m2}) \right)^2} \right)^{12} \\
&\leq \frac{\mathbb{E} \left(\sup_{\lambda} \|B(\lambda\widehat{\theta} + (1-\lambda)\mu_m + h_n \mathbf{u})\|_{\mathbb{F}}^4 \right)}{h_n^{12}},
\end{aligned}$$

because $\lambda\widehat{\theta}_1 + (1-\lambda)\mu_{m1} - (\lambda\widehat{\theta}_2 + (1-\lambda)\mu_{m2})^2 \geq 0$ for all n, m and λ . Since $\|B(\lambda\widehat{\theta} + (1-\lambda)\mu_m + h_n \mathbf{u})\|_{\mathbb{F}}^4$ is a polynomial function of $\widehat{\theta}_1, \widehat{\theta}_2, \widehat{\theta}_3, \mu_m$ and h_n , and since $h_n \rightarrow 0$, $\lambda \in (0, 1)$ and $\mathbb{E}(Q(\mathbf{X})^{16}) < +\infty$, the numerator converges as n and m go to infinity. The proof of (i) is complete.

Proof of (ii). We have already established that (see Proof of Theorem 1):

$$\begin{aligned}
& \mathbb{E}(g_{h_n}(\widehat{\theta}) - g(\theta))^2 \\
&\leq 4\sqrt{\mathbb{E}\|\widehat{\theta} - \mu_m\|_2^4} \|\nabla g_{h_n}(\mu_m)\|_2^2 \\
&\quad + \sqrt{\mathbb{E}\|\widehat{\theta} - \mu_m\|_2^8} \sqrt{\mathbb{E}\|\nabla^2 g_{h_n}(\lambda\widehat{\theta} + (1-\lambda)\mu_m)\|_{\mathbb{F}}^4} \\
&\quad + 4(g_{h_n}(\mu_m) - g_{h_n}(\theta))^2 + 4(g_{h_n}(\theta) - g(\theta))^2.
\end{aligned}$$

By using Marcinkiewicz-Zygmund inequality and the fact that $\|\nabla g_{h_n}(\mu_m)\|_2^2$ is bounded, we have that the first term in the right hand side is $O(1/n)$. The second term is $O(1/(n^2 h_n^6))$ since $\mathbb{E}\|\widehat{\theta} - \mu_m\|_2^8 = O(1/n^2)$ and $\sqrt{\mathbb{E}\|\nabla^2 g_{h_n}(\lambda\widehat{\theta} + (1-\lambda)\mu_m)\|_{\mathbb{F}}^4} = O(1/h_n^6)$ by (i). Finally, by establishing that $(g_{h_n}(\mu_m) - g_{h_n}(\theta))^2 = O(1/m^2)$ and $(g_{h_n}(\theta) - g(\theta))^2 = O(h_n^2)$, we get (ii). The proof of the lemma is complete.

B Supplementary lemmas

Lemma B.1. For all $\alpha > 0$ and all $h \in (0, 1)$,

$$\lim_{n, m \rightarrow +\infty} \mathbb{E} \left[\frac{1}{(h + \widehat{\theta}_1 - \widehat{\theta}_2)^\alpha} \right] = \frac{1}{(h + \text{Var}(Q(\mathbf{X})))^\alpha} \leq \text{Var}(Q(\mathbf{X}))^{-\alpha}.$$

Proof. Let $h \in (0, 1)$ be fixed. The function $\beta_h : x \mapsto 1/(h+x)^\alpha$ is continuously differentiable such that its first derivative is uniformly bounded on \mathbb{R}_+ by $1/h^{\alpha+1}$ then it is Lipschitz. Therefore:

$$\begin{aligned}
\mathbb{E} \left(\beta_h(\widehat{\theta}_1 - \widehat{\theta}_2) - \beta_h(\theta_1 - \theta_2) \right)^2 &\leq \frac{1}{h^{2\alpha+2}} \mathbb{E} \left(\widehat{\theta}_1 - \widehat{\theta}_2 - \theta_1 + \theta_2 \right)^2 \\
&\leq \frac{2}{h^{2\alpha+2}} \left(\text{Var}(\widehat{\theta}_1) + \text{Var}(\widehat{\theta}_2) + \left(\mathbb{E}(\widehat{\theta}_2) - \theta_2 \right)^2 \right),
\end{aligned}$$

using convexity inequality. Based on Marcinkiewicz-Zygmund inequality (see Theorem C.1) and Lemma B.2, it follows that

$$\lim_{n,m \rightarrow \infty} \left((\text{Var}(\widehat{\theta}_1) + \text{Var}(\widehat{\theta}_2^2) + (\mathbb{E}(\widehat{\theta}_2^2) - \theta_2^2)^2) \right) = 0.$$

Straightforwardly:

$$\lim_{n,m \rightarrow \infty} \mathbb{E} \left(\frac{1}{(h + (\widehat{\theta}_1 - \widehat{\theta}_2^2)^\alpha)} \right) = \frac{1}{(h + \theta_1 - \theta_2^2)^\alpha} \leq \frac{1}{\text{Var}(Q(\mathbf{X}))^\alpha}.$$

□

Lemma B.2. *Let $\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(n)}$ be n i.i.d. copies of \mathbf{X} and $Z^{(1,1)}, \dots, Z^{(n,m)}$ be $n \times m$ i.i.d. copies of Z such that $(\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(n)})$ and $(Z^{(1,1)}, \dots, Z^{(n,m)})$ are independent. Then, for all $q \in \mathbb{N}$: $m^{-q} \mathbb{E} \left(\sum_{k=1}^m \varphi(\mathbf{X}^{(1)}, Z^{(1,k)}) \right)^q$ is polynomial in m^{-1} of degree $q - 1$ with constant $\mathbb{E}(\mathbb{E}[\varphi(\mathbf{X}, Z) | \mathbf{X}])^q$.*

Proof. It holds that:

$$\begin{aligned} \mathbb{E} \left(\frac{1}{m} \sum_{k=1}^m \varphi(\mathbf{X}^{(1)}, Z^{(1,k)}) \right)^q &= \frac{1}{m^q} \mathbb{E} \left(\sum_{k_1=1}^m \dots \sum_{k_q=1}^m \varphi(\mathbf{X}^{(1)}, Z^{(1,k_1)}) \dots \varphi(\mathbf{X}^{(1)}, Z^{(1,k_q)}) \right) \\ &= \frac{1}{m^q} \sum_{k_1=1}^m \dots \sum_{k_q=1}^m \mathbb{E} \left(\varphi(\mathbf{X}^{(1)}, Z^{(1,k_1)}) \dots \varphi(\mathbf{X}^{(1)}, Z^{(1,k_q)}) \right). \end{aligned}$$

Denote by $\lambda : \{1, \dots, m\}^q \rightarrow \mathbb{N}$ the map which with each $\mathbf{k} := (k_1, \dots, k_q)$ associates the number of distinct indices among k_1, \dots, k_q . If $1 \leq l \leq q$ then denote by $\rho_l : \lambda^{-1}(l) \rightarrow \{1, \dots, q\}^l$ the map which with each $\mathbf{k} \in \lambda^{-1}(l)$ associates (r_1, \dots, r_l) , where $r_i = |\{j : k_j = k_{j_i}\}|$ for every $i = 1, \dots, l$ and k_{j_1}, \dots, k_{j_l} are the distinct indices found among k_1, \dots, k_q . Obviously, $r_1 + \dots + r_l = q$. We have

$$\begin{aligned} &\sum_{k_1=1}^m \dots \sum_{k_q=1}^m \mathbb{E} \left(\varphi(\mathbf{X}^{(1)}, Z^{(1,k_1)}) \dots \varphi(\mathbf{X}^{(1)}, Z^{(1,k_q)}) \right) \\ &= \sum_{k_1=1}^m \dots \sum_{k_q=1}^m f(\mathbf{k}) \\ &= \sum_{l=1}^q \left(\sum_{(r_1, \dots, r_l) \in \{1, \dots, q\}^l : r_1 + \dots + r_l = q} \left(\sum_{\mathbf{k} \in \lambda^{-1}(l) : \rho_l(\mathbf{k}) = (r_1, \dots, r_l)} f(\mathbf{k}) \right) \right). \quad (7) \end{aligned}$$

Now, since

$$\begin{aligned} &\mathbb{E} \left(\varphi(\mathbf{X}^{(1)}, Z^{(1,k_1)}) \dots \varphi(\mathbf{X}^{(1)}, Z^{(1,k_q)}) \right) \\ &= \mathbb{E} \left(\varphi(\mathbf{X}^{(1)}, Z^{(1,k_{j_1})})^{r_1} \dots \varphi(\mathbf{X}^{(1)}, Z^{(1,k_{j_l})})^{r_l} \right) \\ &= \mathbb{E} \left(\prod_{s=1}^l \mathbb{E} \left[\varphi(\mathbf{X}^{(1)}, Z^{(1,k_{j_s})})^{r_s} \mid \mathbf{X}^{(1)} \right] \right) \end{aligned}$$

is symmetric in r_1, \dots, r_l , it holds that

$$\begin{aligned} & \sum_{(r_1, \dots, r_l) \in \{1, \dots, q\}^l: r_1 + \dots + r_l = q} \left(\sum_{\mathbf{k} \in \lambda^{-1}(l); \rho_l(\mathbf{k}) = (r_1, \dots, r_l)} f(\mathbf{k}) \right) \\ &= c(l, (r_1, \dots, r_l), m) \mathbb{E} \left(\prod_{s=1}^l \mathbb{E} \left[\varphi(\mathbf{X}^{(1)}, Z^{(1, k_{j_s})})^{r_s} \mid \mathbf{X}^{(1)} \right] \right) \end{aligned}$$

where

$$c(l, (r_1, \dots, r_l), m) = \binom{q}{r_1} \binom{q-r_1}{r_2} \dots \binom{q-r_1-\dots-r_{l-1}}{r_l} m(m-1) \dots (m-l+1). \quad (8)$$

Notice that the expression in the right-hand side of (8) is invariant by permutation of r_1, \dots, r_l . Therefore, the sum (7) is a polynomial in m of degree q with constant zero and hence $\mathbb{E} \left(\frac{1}{m} \sum_{k=1}^m \varphi(\mathbf{X}^{(1)}, Z^{(1, k)})^q \right)$ is a polynomial in $\frac{1}{m}$ of degree $q-1$ with constant $\lim_{m \rightarrow +\infty} \mathbb{E} \left(\frac{1}{m} \sum_{k=1}^m \varphi(\mathbf{X}^{(1)}, Z^{(1, k)})^q \right) = \mathbb{E}(\mathbb{E}[\varphi(\mathbf{X}, Z) \mid \mathbf{X}])^q$. \square

C The Marcinkiewicz-Zygmund inequality

Theorem C.1 (Marcinkiewicz and Zygmund (1937)). *Let U_1, \dots, U_n be i.i.d. random variables such that $\mathbb{E}(U_1) = 0$ and $\mathbb{E}|U_1|^q < +\infty$, where $1 \leq q < +\infty$. There exist A_q and B_q depending only on q such that:*

$$A_q \mathbb{E} \left(\left(\sum_{i=1}^n |U_i|^2 \right)^{\frac{q}{2}} \right) \leq \mathbb{E} \left(\left| \sum_{i=1}^n U_i \right|^q \right) \leq B_q \mathbb{E} \left(\left(\sum_{i=1}^n |U_i|^2 \right)^{\frac{q}{2}} \right)$$

Furthermore, there exists C_q independent from n such that:

$$\mathbb{E} \left(\left| \frac{1}{n} \sum_{i=1}^n U_i \right|^q \right) \leq \frac{C_q}{n^{\frac{q}{2}}}. \quad (9)$$

D Supplementary calculations

Some details about the calculations of the bias and the variance of the estimators in Section 8 are given. We start by recalling an elementary formula in Lemma 3.

Lemma 3 (Partial fraction decomposition). *Let $a \in \mathbb{R}$ and $\tilde{q} \in \mathbb{N}$. For every $x \in \mathbb{R}$ and $h \in \mathbb{R}$ such that $h+x \neq 0$, it holds*

$$\frac{x^a}{h+x} = \sum_{q=0}^{\tilde{q}} x^{a-q-1} h^q (-1)^q + h^{\tilde{q}+1} (-1)^{\tilde{q}+1} \frac{x^{a-\tilde{q}-1}}{h+x}.$$

The proof of Lemma 3 is by mathematical induction. We have

$$\mathbb{E}(g_h(\hat{\theta})) = \left(1 - \frac{1}{n}\right) \mathbb{E} \left(g_h \left(\frac{a+1}{a+2} \right) \right) + \frac{1}{n} \mathbb{E}(g_h(X)) \quad (10)$$

with $\mathbb{E}(g_h(X)) = \int_0^1 g_h(x)f(x) dx = (a+1) \int_0^1 x^a/(h+x) dx$. With $a = 1$, taking $\tilde{q} = 0$ in Lemma 3 yields

$$\mathbb{E}(g_h(X)) = 2 \int_0^1 \left(1 - h \frac{1}{h+x}\right) dx = 1 - 2h \log \left(\frac{1+h}{h}\right). \quad (11)$$

Plugging (11) into (10) and elementary calculations yield the asymptotic expansion of the bias given in Example 2. Noticing that $\mathbb{E}(g_h(X)^2) = 2 \int_0^1 \frac{x}{(h+x)^2} dx = 2 \int_0^1 \frac{1}{h+x} (1 - h \frac{1}{h+x}) dx$, the same reasoning yields the asymptotic expansion of the variance. With $a = 2$, similar calculations with $\tilde{q} = 1$ in Lemma 3 yield the desired results.

E Background of nested Monte Carlo estimation

We say that an estimator $\hat{\theta}_{n,m} = (\hat{\theta}_{1;n,m}, \dots, \hat{\theta}_{d;n,m}) \in \mathbb{R}^d$ is a nested Monte Carlo estimator if each component $\hat{\theta}_{j;n,m}$ has the form

$$\begin{aligned} \hat{\theta}_{j;n,m} = & \\ & \frac{1}{n} \sum_{i=1}^n \phi_{j,1} \left(\frac{1}{m} \sum_{k=1}^m \varphi_{j,1}(\xi^{(i)}, \zeta^{(i,k)}) \right) \times \dots \times \phi_{j,p_j} \left(\frac{1}{m} \sum_{k=1}^m \varphi_{j,p_j}(\xi^{(i)}, \zeta^{(i,k)}) \right), \end{aligned} \quad (12)$$

where $\phi_{j,1}, \varphi_{j,1}, \dots, \phi_{j,p_j}, \varphi_{j,p_j}$ are some measurable functions and $\{\xi^{(i)}, \zeta^{(i,k)}, k = 1, \dots, m\}, i = 1, \dots, n$, are some independent and identically distributed sets of random vectors. The estimator (12) is called a nested Monte Carlo estimator because it is built from two nested Monte Carlo loops: an outer loop is used to simulate the random vectors $\xi^{(1)}, \dots, \xi^{(n)}$, and, for each $\xi^{(i)}$, an inner loop is used to simulate $\zeta^{(i,k)}, k = 1, \dots, m$. The random vectors $\{\zeta^{(i,k)}\}$ and $\{\xi^{(i)}\}$ are assumed to be independent. Let $\theta = (\theta_1, \dots, \theta_d) \in \mathbb{R}^d$ denote the vector with components

$$\begin{aligned} \theta_j := & \\ \mathbb{E} \left(\phi_{j,1} \left(\mathbb{E} \left[\varphi_{j,1}(\xi^{(1)}, \zeta^{(1,1)}) \mid \xi^{(1)} \right] \right) \times \dots \times \phi_{j,p_j} \left(\mathbb{E} \left[\varphi_{j,p_j}(\xi^{(1)}, \zeta^{(1,1)}) \mid \xi^{(1)} \right] \right) \right), \end{aligned} \quad (13)$$

$j = 1, \dots, d$. If, for each $i = 1, \dots, n$ and conditionally on $\xi^{(i)}$, the random variables $\zeta^{(i,k)}$ are independent and identically distributed then the MSE of $\hat{\theta}_{n,m}$ is of the form

$$\mathbb{E} \|\hat{\theta}_{n,m} - \theta\|_2^2 = \frac{1}{n} \text{Trace}(\Sigma_m) + \|b_m\|_2^2, \quad (14)$$

where $b_m = \mathbb{E}\hat{\theta}_{n,m} - \theta$ is the bias of the estimator $\hat{\theta}_{n,m}$ and Σ_m is the variance-covariance matrix of the vector

$$\begin{pmatrix} \phi_{1,1} \left(\frac{1}{m} \sum_{k=1}^m \varphi_{1,1}(\xi^{(1)}, \zeta^{(1,k)}) \right) \times \dots \times \phi_{1,p_1} \left(\frac{1}{m} \sum_{k=1}^m \varphi_{1,p_1}(\xi^{(1)}, \zeta^{(1,k)}) \right) \\ \vdots \\ \phi_{d,1} \left(\frac{1}{m} \sum_{k=1}^m \varphi_{d,1}(\xi^{(1)}, \zeta^{(1,k)}) \right) \times \dots \times \phi_{d,p_d} \left(\frac{1}{m} \sum_{k=1}^m \varphi_{d,p_d}(\xi^{(1)}, \zeta^{(1,k)}) \right) \end{pmatrix}.$$

The MSE in equation (14) is the sum of Σ_m/n , the variance of $\hat{\theta}_{n,m}$, and $\|b_m\|_2^2$, its squared bias.

Assumption 1. As $m \rightarrow +\infty$, it holds $b_m \rightarrow 0$ and $\Sigma_m \rightarrow \Sigma$ for some nonnegative definite matrix Σ .

Assumption 1 is fulfilled provided that the functions $\phi_{j,1}, \dots, \phi_{j,p_j}, j = 1, \dots, d$, are bounded, Lipschitz continuous or have bounded derivatives (Giorgi et al., 2017, 2020). If Assumption 1 is satisfied then letting both m and n go to infinity yields that $\mathbb{E} \|\hat{\theta}_{n,m} - \theta\|_2^2 \rightarrow 0$, that is, $\hat{\theta}_{n,m}$ converges in quadratic mean to θ . Since Σ_m is a constant asymptotically as $m \rightarrow \infty$, the MSE of $\hat{\theta}_{n,m}$ is of order $1/n + \|b_m\|_2^2$. If $p = 1$ and ϕ_1 is a smooth function with uniformly bounded third derivative then Hong and Juneja (2009) showed that $b_m = O(1/m)$ and thus $\mathbb{E} \|\hat{\theta}_{n,m} - \theta\|_2^2 = O(1/n + 1/m^2)$, leading to $O(T^{2/3})$ reached for $n = O(T^{2/3})$ and $m = O(T^{1/3})$, the optimal solution. If ϕ_1 is simply Lipschitz continuous, Rainforth et al. (2018) rather found that $\mathbb{E} \|\hat{\theta}_{n,m} - \theta\|_2^2 = O(1/n + 1/m)$, leading to a MSE of order $O(T^{1/2})$, ensured by the choice of $n = O(T^{1/2})$ and $m = O(T^{1/2})$.

References

- Evan Baker, Pierre Barbillon, Arindam Fadikar, Robert B. Gramacy, Radu Herbei, David Higdon, Jiangeng Huang, Leah R. Johnson, Pulong Ma, Anirban Mondal, Bianica Pires, Jerome Sacks, and Vadim Sokolov. Analyzing Stochastic Computer Models: A Review with Opportunities. *Statistical Science*, 37(1):64–89, February 2022. ISSN 0883-4237, 2168-8745. doi: 10.1214/21-STS822. Publisher: Institute of Mathematical Statistics.
- Mickaël Binois, Robert B. Gramacy, and Mike Ludkovski. Practical Heteroscedastic Gaussian Process Modeling for Large Simulation Experiments. *Journal of Computational and Graphical Statistics*, 27(4):808–821, 2018. doi: 10.1080/10618600.2018.1458625.
- Mickaël Binois, Jiangeng Huang, Robert B. Gramacy, and Mike Ludkovski. Replication or Exploration? Sequential Design for Stochastic Simulation Experiments. *Technometrics*, 61(1):7–23, 2019. doi: 10.1080/00401706.2018.1469433.
- G. Castellán, A. Cousien, and V.C. Tran. Non-parametric adaptive estimation of order 1 Sobol’ indices in stochastic models, with an application to Epidemiology. *Electronic Journal of Statistics*, 14(1):50 – 81, 2020. doi: 10.1214/19-EJS1627.
- Sourav Chatterjee. A New Coefficient of Correlation. *Journal of the American Statistical Association*, 116(536):2009–2022, October 2021. ISSN 0162-1459. doi: 10.1080/01621459.2020.1758115. Publisher: Taylor & Francis _eprint: <https://doi.org/10.1080/01621459.2020.1758115>.
- Xi Chen and Qiang Zhou. Sequential experimental designs for stochastic kriging. In *Proceedings of the Winter Simulation Conference 2014*, pages 3821–3832, 2014. doi: 10.1109/WSC.2014.7020209.
- Xi Chen and Qiang Zhou. Sequential design strategies for mean response surface metamodeling via stochastic kriging with adaptive exploration and exploitation. *European Journal of Operational Research*, 262(2):575–585, 2017. ISSN 0377-2217. doi: <https://doi.org/10.1016/j.ejor.2017.03.042>.

- A. Courcoul, H. Monod, M. Nielen, D. Klinkenberg, L. Hogerwerf, F. Beaudeau, and E. Vergu. Modelling the effect of heterogeneity of shedding on the within herd *coxiella burnetii* spread and identification of key parameters by sensitivity analysis. *Journal of Theoretical Biology*, 284(1):130–141, 2011. ISSN 0022-5193. doi: <https://doi.org/10.1016/j.jtbi.2011.06.017>.
- S. da Veiga. Kernel-based ANOVA decomposition and Shapley effects – Application to global sensitivity analysis, 2021. arXiv:2101.05487.
- S. da Veiga and F Gamboa. Efficient estimation of sensitivity indices. *Journal of Non-parametric Statistics*, 25(3):573–595, 2013. doi: 10.1080/10485252.2013.784762.
- S. da Veiga, J-M Loubes, and M. Solís. Efficient estimation of conditional covariance matrices for dimension reduction. *Communications in Statistics - Theory and Methods*, 46(9):4403–4424, 2017. doi: 10.1080/03610926.2015.1083109.
- P. Étoré, C. Prieur, D. K. Pham, and L. Li. Global sensitivity analysis for models described by Stochastic Differential Equations. *Methodology and Computing in Applied Probability*, 22(2):803–831, June 2020. ISSN 1387-5841, 1573-7713. doi: 10.1007/s11009-019-09732-6.
- J-C. Fort, T. Klein, A. Lagnoux, and B. Laurent. Estimation of the Sobol’ indices in a linear functional multidimensional model. *Journal of Statistical Planning and Inference*, 143(9):1590–1605, 2013. ISSN 0378-3758. doi: <https://doi.org/10.1016/j.jspi.2013.04.007>.
- J-C. Fort, T. Klein, and A. Lagnoux. Global sensitivity analysis and Wasserstein spaces. *SIAM/ASA Journal on Uncertainty Quantification*, 9(2):880–921, 2021. doi: 10.1137/20M1354957.
- F. Gamboa, T. Klein, A. Lagnoux, and L. Moreno. Sensitivity analysis in general metric spaces. *Reliability Engineering & System Safety*, 212:107611, 2021.
- Fabrice Gamboa, Pierre Gremaud, Thierry Klein, and Agnès Lagnoux. Global sensitivity analysis: A novel generation of mighty estimators based on rank statistics. *Bernoulli*, 28(4):2345–2374, November 2022. ISSN 1350-7265. Publisher: Bernoulli Society for Mathematical Statistics and Probability.
- M. B. Giles and A.-L. Haji-Ali. Multilevel nested simulation for efficient risk estimation. *SIAM/ASA Journal on Uncertainty Quantification*, 7(2):497–525, 2019.
- Daniel T Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22(4):403–434, 1976. ISSN 0021-9991. doi: [https://doi.org/10.1016/0021-9991\(76\)90041-3](https://doi.org/10.1016/0021-9991(76)90041-3).
- L. Gilquin, C. Prieur, E. Arnaud, and H. Monod. Iterative estimation of Sobol’ indices based on replicated designs. *Computational and Applied Mathematics*, 40(1):18, January 2021. ISSN 1807-0302. doi: 10.1007/s40314-020-01402-5.
- D. Giorgi, V. Lemaire, and G. Pagès. Limit theorems for weighted and regular Multi-level estimators. *Monte Carlo Methods and Applications*, 23(1):43–70, 2017. doi: [doi:10.1515/mcma-2017-0102](https://doi.org/10.1515/mcma-2017-0102).

- D. Giorgi, V. Lemaire, and G. Pagès. Weak Error for Nested Multilevel Monte Carlo. *Methodology and Computing in Applied Probability*, 22(3):1325–1348, September 2020. ISSN 1573-7713. doi: 10.1007/s11009-019-09751-3.
- J. L. Hart, A. Alexanderian, and P. A. Gremaud. Efficient computation of Sobol’ indices for stochastic models. *SIAM Journal on Scientific Computing*, 39(4):A1514–A1530, 2017. doi: 10.1137/16M106193X.
- L. J. Hong and S. Juneja. Estimating the mean of a non-linear function of conditional expectation. In *Proceedings of the 2009 Winter Simulation Conference (WSC)*, pages 1223–1236, 2009. doi: 10.1109/WSC.2009.5429428.
- T. Ishigami and T. Homma. An importance quantification technique in uncertainty analysis for computer models. [1990] *Proceedings. First International Symposium on Uncertainty Modeling and Analysis*, pages 398–403, 1990.
- Alexandre Janon, Thierry Klein, Agnes Lagnoux-Renaudie, Maëlle Nodet, and Clémentine Prieur. Asymptotic normality and efficiency of two Sobol’ index estimators. *ESAIM: Probability and Statistics*, 18:342–364, October 2014. doi: 10.1051/ps/2013040.
- Angela M. Jarrett, Yaning Liu, N. G. Cogan, and M. Yousuff Hussaini. Global sensitivity analysis used to interpret biological experimental results. *Journal of Mathematical Biology*, 71(1):151–170, July 2015. ISSN 1432-1416. doi: 10.1007/s00285-014-0818-3.
- M. Navarro Jimenez, O. P. Le Maître, and O. M. Knio. Nonintrusive polynomial chaos expansions for sensitivity analysis in stochastic differential equations. *SIAM/ASA Journal on Uncertainty Quantification*, 5(1):378–402, 2017. doi: 10.1137/16M1061989.
- Henri Mermoz Kouye, Gildas Mazo, Clémentine Prieur, and Elisabeta Vergu. Performing global sensitivity analysis on simulations of a continuous-time markov chain model motivated by epidemiology. *Computational and Applied Mathematics*, 43(7):409, 2024.
- O. P. Le Maître, O. M. Knio, and A. Moraes. Variance decomposition in stochastic simulators. *The Journal of Chemical Physics*, 142(24):244115, 2015. doi: 10.1063/1.4922922.
- Nora Lüthen, Stefano Marelli, and Bruno Sudret. Sparse Polynomial Chaos Expansions: Literature Survey and Benchmark. *SIAM/ASA Journal on Uncertainty Quantification*, 9(2):593–649, January 2021. doi: 10.1137/20M1315774. Publisher: Society for Industrial and Applied Mathematics.
- J. Marcinkiewicz and A. Zygmund. Sur les fonctions indépendantes. *Fundamenta Mathematicae*, 29(1):60–90, 1937.
- G. Mazo. A trade-off between explorations and repetitions for estimators of two global sensitivity indices in stochastic models induced by probability measures. *SIAM/ASA Journal on Uncertainty Quantification*, 9(4):1673–1713, 2021. doi: 10.1137/19M1272706.

- G. Mazo and L. Tournier. An inference method for global sensitivity analysis. *Technometrics*, pages 1–13, 2025. doi: <https://doi.org/10.1080/00401706.2024.2431113>.
- P. Mycek and M. De Lozzo. Multilevel Monte Carlo covariance estimation for the computation of Sobol’ Indices. *SIAM/ASA Journal on Uncertainty Quantification*, 7(4):1323–1348, 2019. doi: 10.1137/18M1216389.
- M. Navarro Jimenez, O. P. Le Maître, and O. M. Knio. Global sensitivity analysis in stochastic simulators of uncertain reaction networks. *The Journal of Chemical Physics*, 145(24):244106, 2016. doi: 10.1063/1.4971797.
- I. Panin. Risk of estimators for Sobol’ sensitivity indices based on metamodels. *Electronic Journal of Statistics*, 15(1):235 – 281, 2021. doi: 10.1214/20-EJS1793.
- T. Rainforth, R. Cornish, H. Yang, A. Warrington, and F. Wood. On Nesting Monte Carlo Estimators. In Jennifer Dy and Andreas Krause, editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 4267–4276. PMLR, 10–15 Jul 2018.
- Bruno Sudret. Global sensitivity analysis using polynomial chaos expansions. *Reliability Engineering and System Safety*, 93(7):964–979, 2008. ISSN 0951-8320. doi: <https://doi.org/10.1016/j.ress.2007.04.002>. Bayesian Networks in Dependability.
- Bruno Sudret, Stefano Marelli, and Joe Wiart. Surrogate models for uncertainty quantification: An overview. In *2017 11th European Conference on Antennas and Propagation (EUCAP)*, pages 793–797, March 2017. doi: 10.23919/EUCAP.2017.7928679.
- Rui Tuo and C. F. Jeff Wu. A Theoretical Framework for Calibration in Computer Models: Parametrization, Estimation and Convergence Properties. *SIAM/ASA Journal on Uncertainty Quantification*, 4(1):767–795, January 2016. doi: 10.1137/151005841. Publisher: Society for Industrial and Applied Mathematics.
- X. Zhu and B. Sudret. Global sensitivity analysis for stochastic simulators based on generalized lambda surrogate models. *Reliability Engineering & System Safety*, 214: 107815, 2021. ISSN 0951-8320. doi: <https://doi.org/10.1016/j.ress.2021.107815>.
- Xujia Zhu and Bruno Sudret. STOCHASTIC POLYNOMIAL CHAOS EXPANSIONS TO EMULATE STOCHASTIC SIMULATORS. *International Journal for Uncertainty Quantification*, 13(2), 2023. ISSN 2152-5080, 2152-5099. doi: 10.1615/Int.J.UncertaintyQuantification.2022042912. Publisher: Begel House Inc.