



HAL
open science

Robust estimation with latin hypercube sampling: a central limit theorem for Z-estimators

Faouzi Hakimi

► **To cite this version:**

Faouzi Hakimi. Robust estimation with latin hypercube sampling: a central limit theorem for Z-estimators. 2024. hal-04462415

HAL Id: hal-04462415

<https://hal.science/hal-04462415>

Preprint submitted on 16 Feb 2024

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.

Robust estimation with latin hypercube sampling: a central limit theorem for Z-estimators

Faouzi Hakimi^{1,2}

¹CEA, DES, IRESNE, , Cadarache F-13108
Saint-Paul-Lez-Durance, France

²Institut Mathématique de Toulouse,,
31062 Toulouse, France

Correspondence

Faouzi Hakimi, Marseille, France
Email: faouzi.hakimi@protonmail.com

Funding information

Latin hypercube sampling (LHS) is a stratified sampling method widely used in computer experiments. In this work, we extend convergence results on the sample mean with Latin hypercube sampling to the class of Z -estimators, gathering all estimators that can be written as zeros of a sample mean function. In particular, the asymptotic variance of this estimate is obtained. This asymptotic variance is shown to be lower using LHS than using classic independent and identically distributed sampling. A Central Limit theorem for Z -estimators under LHS is also given.

KEYWORDS

Design of Experiments, Statistical Computing, Robust Estimation, Variance Reduction, Latin Hypercube Sampling

1 | INTRODUCTION

Latin Hypercube Sampling (LHS), proposed in [1], is an interesting alternative to *independent and identically distributed (i.i.d)* random sampling to explore the behavior of complex systems through computer experiments [2, 3, 4]. To generate an LHS of size n , the idea is first to divide the range of variation of each variable into n equally probable intervals. To obtain the design of experiments, the n sample points are then placed in order to have only one sample in each row and each column. Figure 1 gives schematic examples of LHS of dimension $d = 2$ and size $n = 4$. The process is similar in higher dimensions.

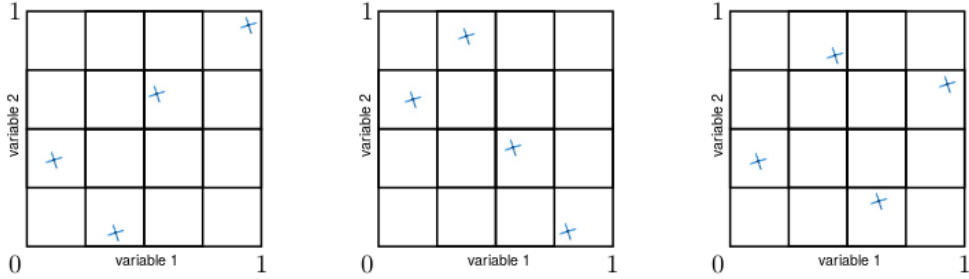


Figure 1 Three schematic LHS examples of dimension $d = 2$ and size $n = 4$.

Several results have been established for the convergence of estimators under LHS. Most of them concern the empirical mean of a measurable function with finite second order moment functions. Indeed, it has been shown in [5] that the asymptotic variance of the mean statistic is lower under LHS than in classic *i.i.d* random sampling for this kind of statistics. A central limit theorem has also been shown for the empirical mean of bounded functions in [6] and generalized to finite third order moment functions in [7]. The aim of this paper is to extend the results on the convergence of empirical mean estimators under LHS to the class of Z -estimators. This class gathers all estimators that can be written as zeros of an empirical mean.

The subject and most results addressed here is covered in Chapter 2 of the thesis manuscript [8]. The paper is divided in five sections. In Section 2, a formal definition and main convergence properties of Latin Hypercube Sampling are given. Definitions and useful properties on Z -estimators are then presented in Section 3. Original results regarding the asymptotic normality of this class of estimators under LHS are provided in Section 4. Finally, an example of application is proposed in Section 5.

Let us now introduce some useful notations regarding this work. We first denote by $\mathbf{X} = (X_1, \dots, X_d)$ the vector of d independent random variables evolving in a measurable space $\mathcal{X} \subset \mathbb{R}^d$. For simplicity and without loss of generality, we assume that the d inputs vary uniformly in $[0, 1]$ so we have that, for j in $\llbracket 1, d \rrbracket$, $X_j \sim U_{[0,1]}$. Indeed, one can always work under uniformity and then use the inverse transformation method [9] to place the support back on the original scale and retrieve the original distribution, as long as the sampling distribution of interest is a product measure (see for instance [6] p543 for details).

A size n ($n \in \mathbb{N}^*$) sample of \mathbf{X} generated using a sampling method, generically denoted "*METHOD*" is written as follows:

- $\mathbf{X}^{METHOD} = (\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)})^T \in M_{n,d}([0, 1])$. We also recall that $M_{n,d}([0, 1])$ denotes the space of matrices of size $n \times d$ with coefficients in $[0, 1]$.
- $\mathbf{x}_j^{METHOD} = (x_j^{(1)}, \dots, x_j^{(n)})^T$ with $j \in \llbracket 1, d \rrbracket$ is the j th column of \mathbf{X}^{METHOD} corresponding to the effective generated sample of the input X_j .

A sample generated by classic *i.i.d* random sampling will be denoted \mathbf{X}^{IID} and a sample generated by Latin Hypercube sampling will be denoted \mathbf{X}^{LHS} . Similar notations will be used for any quantities estimated with either of these two

sampling methods. If no sampling method is mentioned, it means that the results presented do not depend on the sampling method.

We also define the measurable function $g : \mathcal{X} \rightarrow \mathbb{R}^q$ with $q \in \mathbb{N}^*$ such that $g(\mathbf{X}) = \mathbf{Y}$. This function represents in practice the studied simulation code. $\mathbf{Y}^{METHOD} = g(\mathbf{X}^{METHOD}) = (\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(n)})^T$ is the matrix of output samples corresponding to \mathbf{X}^{METHOD} .

For $\mathbf{a} = (a_1, \dots, a_q) \in \mathbb{R}^q$ with $q \in \mathbb{N}^*$, we denote by $\|\mathbf{a}\|$ the Euclidean norm of \mathbf{a} such that $\|\mathbf{a}\|^2 = \sum_{i=1}^q a_i^2$. Similarly, for any matrix \mathbf{A} in $M_{q,q}(\mathbb{R})$, we denote by $\|\mathbf{A}\|$ the pseudo Euclidean norm (Frobenius norm) such that $\|\mathbf{A}\|^2 = \sum_{1 \leq i,j \leq q} A_{i,j}^2$. Here, $A_{i,j}$ with $i,j \in \llbracket 1, q \rrbracket$ are the components of the matrix \mathbf{A} . A pseudo Euclidean norm $\|\cdot\|$ is finally associated to the tensor space $T_{q,q,q}(\mathbb{R})$. This norm is defined, for all \mathbf{T} in $T_{q,q,q}(\mathbb{R})$, by $\|\mathbf{T}\|^2 = \sum_{1 \leq i,j,k \leq q} T_{i,j,k}^2$. Here, $T_{i,j,k}$ with $i,j,k \in \llbracket 1, q \rrbracket$ are the components of the tensor \mathbf{T} .

We denote by $o(1)$ ("small oh-one") a deterministic sequence that converges to 0 and $O(1)$ ("big oh-one") a deterministic sequence that is bounded. We denote by $o_p(1)$ ("small oh-P-one") a sequence of random variables that converges in probability to 0. The expression $O_p(1)$ ("big oh-P-one") denotes a sequence of random variables that is bounded in probability. We recall that a sequence of random variables $(\mathbf{W}_n)_{n \in \mathbb{N}}$ is bounded in probability if, for any scalar $\epsilon > 0$, there exist M and N such that, for all $n > N$, $\mathbb{P}(\|\mathbf{W}_n\| > M) < \epsilon$ (note that this definition holds in the general case where the norm $\|\cdot\|$ is not Euclidean).

Finally, a multivariate normal distribution of dimension q ($q \in \mathbb{N}^*$) with a mean equal to $\boldsymbol{\mu} \in \mathbb{R}^q$ and a covariance matrix equal to $\boldsymbol{\Sigma}$ in $M_{q,q}(\mathbb{R})$ is denoted $\mathcal{N}_q(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

2 | DEFINITION AND MAIN PROPERTIES ON LHS

As previously stated, Latin Hypercube Sampling is a statistical method used to generate a near-random sample of parameter values from a multidimensional distribution. To define it formally, we denote, for $d, n \in \mathbb{N}^*$:

1. $\pi_j = (\pi_j(1) \dots \pi_j(n))^T$, $j \in \llbracket 1, d \rrbracket$ as a random permutation of $\llbracket 1, n \rrbracket$, according to the uniform distribution on the set of all possible permutations of $\llbracket 1, n \rrbracket$. The random permutations $(\pi_j)_{j \in \llbracket 1, d \rrbracket}$ are assumed to be independent.
2. $\mathbf{u}_j = (u_j^{(1)}, \dots, u_j^{(n)})^T$, $j \in \llbracket 1, d \rrbracket$ as an *i.i.d.* sample of the uniform distribution $U_{[0,1]}$. The samples $(\mathbf{u}_j)_{j \in \llbracket 1, d \rrbracket}$ are assumed to be independent.

The random permutations $(\pi_j)_{j \in \llbracket 1, d \rrbracket}$ and the samples $(\mathbf{u}_j)_{j \in \llbracket 1, d \rrbracket}$ are also assumed to be independent. The n -sized sampling \mathbf{x}_j^{LHS} of the input X_j , $j \in \llbracket 1, d \rrbracket$, is then defined as follows:

$$\mathbf{x}_j^{LHS} = (x_j^{(1)}, \dots, x_j^{(n)})^T = \left(\frac{1}{n}(\pi_j(1) - u_j^{(1)}), \dots, \frac{1}{n}(\pi_j(n) - u_j^{(n)}) \right)^T. \quad (1)$$

The corresponding LHS design of dimension d and size n is then $\mathbf{X}^{LHS} = (\mathbf{x}_1^{LHS}, \dots, \mathbf{x}_d^{LHS})$.

The LHS method leads to a good point repartition in the sub-projections of dimension 1. Indeed, a LHS verifies these two properties by definition:

- $\forall i \in \llbracket 1, n \rrbracket, \min_{1 \leq i' \leq n} (|x_j^{(i)} - x_j^{(i')}|) \leq \frac{2}{n}$.
- $\max_{1 \leq i, i' \leq n} (|x_j^{(i)} - x_j^{(i')}|) \geq \frac{n-2}{n}$.

As a result of its stratified nature, the realizations of the LHS design are not *i.i.d*. However, several results have been indeed established for the convergence of estimators under LHS. Most of them concern the sample mean (first order U-statistics) of measurable functions. For instance, it has been shown in [1] that, for any measurable function, this estimator is unbiased:

Proposition 1 Let $g : [0, 1]^d \rightarrow \mathbb{R}^q$ with $d, q \in \mathbb{N}^*$ be a measurable function such that $\mathbb{E}(\|g(\mathbf{X})\|) < +\infty$. Denote

$$G_n^{LHS} = \frac{1}{n} \sum_{i=1}^n g(\mathbf{x}^{(i)}),$$

where $\mathbf{x}^{(i)}, i \in \llbracket 1, n \rrbracket$ is such that $\mathbf{X}^{LHS} = (\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)})^T$ with \mathbf{X}^{LHS} being defined using Equation (1). Then, G_n^{LHS} is an unbiased estimator of $G = \mathbb{E}(g(\mathbf{X}))$.

A second interesting characteristic of mean value estimators under LHS is their variance. Indeed, Stein [5] showed that if g is a real-valued function such that $\mathbb{E}(g^2(\mathbf{X})) < +\infty$, then $\mathbb{V}\text{ar}(G_n^{LHS})$ is always asymptotically smaller than $\mathbb{V}\text{ar}(G_n^{IID})$. This result is generalized to multidimensional functions by Loh in [7]. Property 2 summarizes the main results regarding the covariance matrix of G_n^{LHS} :

Proposition 2 Let $g : [0, 1]^d \rightarrow \mathbb{R}^q, d, q \in \mathbb{N}^*$ be a measurable function with $\mathbb{E}(\|g(\mathbf{X})\|^2) < +\infty$. Let $\frac{1}{n} \Sigma_g^{IID}$ and $\frac{1}{n} \Sigma_g^{LHS}$ be the covariance matrices of G_n^{IID} and G_n^{LHS} respectively, with $\Sigma_g^{IID} = \mathbb{E} \left((g(\mathbf{X}) - G)(g(\mathbf{X}) - G)^T \right)$.

We also define, for $\mathbf{x} = (x_1, \dots, x_d) \in [0, 1]^d$:

- $g_{-j}(x_j) = \int_{[0,1]^{d-1}} [g(\mathbf{x}) - G] \prod_{1 \leq k \leq d, k \neq j} dx_k = \mathbb{E}(g(\mathbf{X}) - G | X_j)$ with $j \in \llbracket 1, d \rrbracket$.
- $g_{rem}(\mathbf{x}) = g(\mathbf{x}) - G - \sum_{j=1}^d g_{-j}(x_j)$.

Then we have, when $n \rightarrow +\infty$:

- $\Sigma_g^{LHS} = \int_{[0,1]^d} g_{rem}(\mathbf{x}) g_{rem}(\mathbf{x})^T d\mathbf{x} + o(1)$.
- $\Sigma_g^{IID} = \int_{[0,1]^d} g_{rem}(\mathbf{x}) g_{rem}(\mathbf{x})^T d\mathbf{x} + \sum_{j=1}^d \int_{[0,1]} g_{-j}(x_j) g_{-j}(x_j)^T dx_j$.

We therefore have that $\Sigma_g^{IID} - \Sigma_g^{LHS}$ is asymptotically positive semidefinite, that is,

$$\forall \xi \in \mathbb{R}^d, \lim_{n \rightarrow +\infty} \xi^T (\Sigma_g^{IID} - \Sigma_g^{LHS}) \xi \geq \sum_{j=1}^d \int_{[0,1]} \xi^T g_{-j}(x_j) g_{-j}(x_j)^T \xi dx_j \geq 0.$$

Since G_n^{IID} converges in quadratic mean to G , we can therefore conclude that G_n^{LHS} converges in quadratic mean to G : $\lim_{n \rightarrow +\infty} \mathbb{E}(\|G_n^{LHS} - G\|^2) = 0$. Thus, G_n^{LHS} converges in probability to G .

In addition, Owen [6] showed a Central Limit Theorem (CLT) for this class of estimators under LHS when the model function g is bounded. This was generalized to any function with finite third moment in [7]. Let us introduce, in the

framework of Property 2, $\mathbf{R}_g = \int_{[0,1]^d} g_{rem}(\mathbf{x})g_{rem}(\mathbf{x})^T d\mathbf{x}$. Notice that asymptotically, we have $\Sigma_g^{LHS} = \mathbf{R}_g + o(1)$. The CLT can be then expressed as follows:

Theorem 1 Let $g : [0, 1]^d \rightarrow \mathbb{R}^q$ ($d, q \in \mathbb{N}$) be a measurable function with $\mathbb{E}(\|g(\mathbf{X})\|^3) < +\infty$. Then, assuming that \mathbf{R}_g is non-singular, we have $\sqrt{n}(G_n^{LHS} - G) \xrightarrow{n \rightarrow +\infty} \mathcal{N}_q(0, \mathbf{R}_g)$.

3 | M-ESTIMATORS AND Z-ESTIMATORS

The objective of this work being to extend the convergence results under LHS to the class of Z -estimators, it is important to define this class and its link with the widely known class of M -estimators. M -estimators are a broad class of estimators, generally obtained by optimizing an empirical mean. The definition of M -estimators was motivated by robust statistics. We refer to [10] for a more general discussion on the subject. The Z -estimator class is directly related to the M -estimator class. The brief overview given here is mainly inspired by the fifth chapter of [11].

Let $\mathbf{X} = (\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)})^T$ be the vector of n realizations of a random vector \mathbf{X} evolving in a measurable space $\mathcal{X} \subset \mathbb{R}^d$, with $d \in \mathbb{N}^*$. Its law is parameterized by a vector $\theta \in \Theta \subset \mathbb{R}^q$, $q \in \mathbb{N}^*$. The statistic $\hat{\theta}_n = \hat{\theta}_n(\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)})$ is a M -estimator of θ if it maximizes a function of the type

$$\theta \rightarrow M_n(\theta) = \frac{1}{n} \sum_{i=1}^n m_\theta(\mathbf{x}^{(i)}). \quad (2)$$

Here, m_θ is a known real-valued function on \mathcal{X} . The maximum value of a function is often found by setting a derivative (or a set of derivative) to 0. In this case, the maximization problem can be reformulated by the following vectorial equations, with ψ_θ corresponding to the vector of derivatives of m_θ :

$$\Psi_n(\theta) = \frac{1}{n} \sum_{i=1}^n \psi_\theta(\mathbf{x}^{(i)}) = 0. \quad (3)$$

More generally, we may consider in Eq. (3) the function ψ_θ as a known measurable and vector-valued map evolving from \mathbb{R}^d to \mathbb{R}^q . ψ_θ is not necessarily a derivative of another function in the general case. The name Z -estimator is used for estimators $\hat{\theta}_n$ satisfying the vectorial equations (3). In this paper, we are focusing on this kind of estimator. Thus, in the following, the notation $\hat{\theta}_n$ will always refer to an estimator defined by Equation (3). This equation may not have an exact solution. Then it is natural to use as estimator a value that is close to zero. Estimators that are sufficiently close to being a zero often have the same asymptotic behavior.

Many known estimators can be defined as Z -estimator. For instance, let \mathbf{X} have a distribution function f_θ with a continuous first derivative in $\theta \in \Theta$. In this case, the maximum likelihood estimator of θ can be written as a Z -estimator as defined by 3 with, for $\mathbf{x} \in \mathbb{R}^d$, $d \in \mathbb{N}^*$, $\psi_\theta(\mathbf{x}) = \left(\frac{\partial \log(f_\theta(\mathbf{x}))}{\partial \theta_1}, \dots, \frac{\partial \log(f_\theta(\mathbf{x}))}{\partial \theta_q} \right)^T$.

The first useful properties regarding Z -estimators concern the link between the consistency of $\Psi_n(\theta)$ and the consistency of $\hat{\theta}_n$. These properties set assumptions on the objective function $\Psi_n(\theta)$ that ensure convergence of the parameter θ we aim to estimate.

For instance, if the parameter θ is a scalar, Proposition 3 presented in [11] gives sufficient conditions for the convergence of $\hat{\theta}_n$:

Proposition 3 *Let Θ be a subset of the real line. Let, for any θ in Θ , $\Psi_n(\theta)$ be random functions and $\Psi(\theta)$ a fixed function such that $\Psi_n(\theta)$ converges to $\Psi(\theta)$ in probability. Assume also that each map $\theta \rightarrow \Psi_n(\theta)$ is continuous and has exactly one zero $\hat{\theta}_n$, or is non-decreasing and converges to 0 in probability. Let θ_0 be a point such that $\Psi(\theta_0 - \epsilon) < 0 < \Psi(\theta_0 + \epsilon)$ for every $\epsilon > 0$.*

Then $\hat{\theta}_n$ is a consistent estimator of θ_0 , meaning that $\hat{\theta}_n \xrightarrow[n \rightarrow +\infty]{P} \theta_0$.

This property has the advantage of holding under simple assumptions. However, as mentioned, it is limited to the scalar case. In [12], one can find assumptions for which the consistency of $\hat{\theta}_n$ is ensured in the multidimensional case:

Proposition 4 *Let Θ be a compact subset of \mathbb{R}^q with $q \in \mathbb{N}^*$. Let also assume that the following hypotheses are true, for any $\theta \in \Theta$:*

- the functions $\Psi_n(\theta)$ and $\Psi(\theta)$ are continuous functions of $\theta \in \Theta$;
- each function $\theta \rightarrow \Psi_n(\theta)$ has exactly one zero $\hat{\theta}_n \in \Theta$;
- $\Psi_n(\theta)$ converges to $\Psi(\theta)$ in probability;
- $\Psi(\theta)$ vanishes only at θ_0 with $\theta_0 \in \Theta$;
- denoting, for $\eta \geq 0$, $w_n(\eta) = \sup\{|\Psi_n(\theta_1) - \Psi_n(\theta_2)|; \|\theta_1 - \theta_2\| \leq \eta, \theta_1, \theta_2 \in \Theta\}$; there exists two sequences (η_k) and (ϵ_k) both decreasing to 0 such that, for all $k \in \mathbb{N}$, $\mathbb{P}(w_n(\eta_k) > \epsilon_k) \xrightarrow[n \rightarrow +\infty]{} 0$.

Then $\hat{\theta}_n$ is a consistent estimator of θ_0 , that is $\hat{\theta}_n \xrightarrow[n \rightarrow +\infty]{P} \theta_0$.

The assumption on $w_n(\eta)$ seems difficult to grasp at first glance. However, as mentioned in [12], if we find a function ϕ from \mathbb{R}_+ to \mathbb{R} such that $\lim_{\eta \rightarrow 0^+} \phi(\eta) = 0$, this assumption on w_n can be obtained through: $\mathbb{P}(w_n(\eta) \geq 2\phi(\eta)) \xrightarrow[n \rightarrow +\infty]{} 0$ for each $\eta \geq 0$. For instance, $w_n(\eta) \xrightarrow[n \rightarrow +\infty]{} \phi(\eta)$, or $\lim_{n \rightarrow +\infty} w_n(\eta) \leq \phi(\eta)$ give both sufficient conditions.

In addition to these convergence properties, several central limit theorems for Z -estimators have been proved. Here we give one of them, proposed in [11]. Theorem 2 relies on the so-called *classic conditions*. They were formulated in the 1930s and 1940s to mathematically tighten the informal derivation of the asymptotic normality of maximum likelihood proposed earlier by [13]. These conditions are stringent, but they are simple. They lead to a simple proof of the central limit theorem. This simplicity will allow us to adapt this theorem to the LHS case.

In particular, a needed assumption for the application of this theorem concerns the existence of a first and a second order derivatives in θ for ψ_θ . Let us introduce these terms, assuming they exist.

We consider that Θ is an open subset of an Euclidean space of dimension q , $q \in \mathbb{N}^*$. The first order partial derivative ψ_θ in $\theta \in \Theta$ is a size $q \times q$ matrix is denoted $\dot{\psi}_\theta$. Its components are such that $\dot{\psi}_{\theta_j, k} = \frac{\partial \psi_{\theta_j}}{\partial \theta_k}$.

The second order partial derivative is a tensor of size $q \times q \times q$ (a q -vector of $q \times q$ matrices) denoted $\ddot{\psi}_\theta$ such that $\ddot{\psi}_{\theta_j, k, l} = \frac{\partial^2 \psi_{\theta_j}}{\partial \theta_k \partial \theta_l}$.

Theorem 2 Let Θ be an open subset of an Euclidean space of dimension q , $q \in \mathbb{N}^*$ and X be a measurable subspace of \mathbb{R}^d , $d \in \mathbb{N}^*$. Assume that, for all θ in Θ and for all x in X , the function $\theta \rightarrow \psi_\theta(x)$ is twice continuously differentiable in θ .

Let $X^{IID} = (x^{(1)}, \dots, x^{(n)})^T$ be the vector of i.i.d realizations of a random variable $\mathbf{X} = (X_1, \dots, X_d)$ evolving in X .

Suppose also that the following assumptions are fulfilled:

1. $\Psi_n^{IID}(\hat{\theta}_n^{IID}) = \frac{1}{n} \sum_{i=1}^n \psi_{\hat{\theta}_n^{IID}}(x^{(i)}) = 0, \forall n \in \mathbb{N}$;
2. there exists a unique θ_0 in Θ such that $\mathbb{E}(\psi_{\theta_0}(\mathbf{X})) = \Psi(\theta_0) = 0$ with θ_0 in Θ ;
3. $\mathbb{E}(\|\psi_{\theta_0}(\mathbf{X})\|^2) < +\infty$;
4. $\mathbb{E}(\psi_{\theta_0}(\mathbf{X}))$ exists and is non-singular;
5. the function $\theta \rightarrow \psi_\theta(x)$ is dominated in norm by a fixed integrable function $\check{\psi}(x)$ for every θ in the neighborhood of θ_0 .

Then, if $\hat{\theta}_n^{IID}$ is a consistent estimator of θ_0 , we have:

$$(\hat{\theta}_n^{IID} - \theta_0) = -[\mathbb{E}(\check{\psi}_{\theta_0}(\mathbf{X}))]^{-1} \frac{1}{n} \sum_{i=1}^n \psi_{\theta_0}(x^{(i)}) + \frac{1}{\sqrt{n}} o_p(1). \quad (4)$$

Moreover, we have that the sequence $\sqrt{n}(\hat{\theta}_n^{IID} - \theta_0)$ is asymptotically normal with mean zero and a covariance matrix equal to $[\mathbb{E}(\check{\psi}_{\theta_0}(\mathbf{X}))]^{-1} \mathbb{E}(\psi_{\theta_0}(\mathbf{X})\psi_{\theta_0}(\mathbf{X})^T) [\mathbb{E}(\check{\psi}_{\theta_0}(\mathbf{X}))]^{-T}$.

For the following, it is important to note that we have $\mathbb{E}(\psi_{\theta_0}(\mathbf{X})\psi_{\theta_0}(\mathbf{X})^T) = \Sigma_{\psi_{\theta_0}}^{IID}$ such that $\frac{1}{n} \sum_{i=1}^n \psi_{\theta_0}(x^{(i)})$ is the covariance matrix of $\Psi_n^{IID}(\theta_0)$. It is also important to remark that among the results presented in this section, only Theorem 2 requires the specific use of an i.i.d sample, since its proof relies on the classical Central Limit Theorem (CLT) [14].

4 | Z-ESTIMATORS UNDER LHS

In this section, we extend the convergence properties of Z -estimators to LHS designs. The idea is to combine all the above properties. Indeed, one can first notice that the Z -function $\Psi_n(\theta)$ is the empirical mean of ψ_θ . Now, as mentioned in Section 2, the convergence of this kind of statistic under LHS holds. We use that here to show a central limit theorem for Z -estimators under LHS.

As in Section 2, let $\frac{1}{n} \sum_{i=1}^n \psi_{\theta_0}^{IID}$ and $\frac{1}{n} \sum_{i=1}^n \psi_{\theta_0}^{LHS}$ be the covariance matrices of $\Psi_n^{IID}(\theta)$ and $\Psi_n^{LHS}(\theta)$ respectively. Let us now give some noteworthy convergence properties on $\Psi_n^{LHS}(\theta)$.

Proposition 5 Let Θ be an open bounded subset of \mathbb{R}^q and $X = [0, 1]^d$ (q, d in \mathbb{N}^*). Let $X^{LHS} = (x^{(1)}, \dots, x^{(n)})^T$ be the vector of LHS realizations of a random variable $\mathbf{X} = (X_1, \dots, X_d)$ evolving in X such that $\mathbf{X} \sim U_{[0,1]^d}$. Assume also that, for all $\theta \in \Theta$, ψ_θ is a measurable function from X to \mathbb{R}^q . We then have the following properties on $\Psi_n^{LHS}(\theta) = \frac{1}{n} \sum_{i=1}^n \psi_\theta(x^{(i)})$:

1. If, for all $\theta \in \Theta$, $\mathbb{E}(\|\psi_\theta(\mathbf{X})\|) < +\infty$, $\Psi_n^{LHS}(\theta)$ is an unbiased estimator of $\Psi(\theta) = \mathbb{E}(\psi_\theta(\mathbf{X}))$.
2. If, for all $\theta \in \Theta$, $\mathbb{E}(\|\psi_\theta(\mathbf{X})\|^2) < +\infty$, we also have, when $n \rightarrow +\infty$:

$\Sigma_{\psi_{\theta}}^{LHS} = \int_{[0,1]^d} \psi_{\theta_{rem}}(\mathbf{x}) \psi_{\theta_{rem}}(\mathbf{x})^T d\mathbf{x} + o(1)$, with $\psi_{\theta_{rem}}$ being defined as in Property 2.

Moreover, we have that $\Sigma_{\psi_{\theta}}^{IID} - \Sigma_{\psi_{\theta}}^{LHS}$ is asymptotically positive semi-definite and that $\Psi_n^{LHS}(\theta)$ converges in quadratic mean to $\Psi(\theta)$. In other words, we have $\lim_{n \rightarrow +\infty} \mathbb{E}(\|\Psi_n^{LHS}(\theta) - \Psi(\theta)\|^2) = 0$.

3. If, for all $\theta \in \Theta$, $\mathbb{E}(\|\psi_{\theta}(\mathbf{X})\|^3) < +\infty$ and if $\mathbf{R}_{\psi_{\theta}} = \int_{[0,1]^d} \psi_{\theta_{rem}}(\mathbf{x}) \psi_{\theta_{rem}}(\mathbf{x})^T d\mathbf{x}$ is non-singular, we have that $\sqrt{n}(\Psi_n^{LHS} - \Psi(\theta)) \xrightarrow{n \rightarrow +\infty} \mathcal{N}_q(0, \mathbf{R}_{\psi_{\theta}})$.

Proof. Let us show these properties one by one:

1. Since, for all $\theta \in \Theta$, the function $\mathbf{x} \rightarrow \psi_{\theta}(\mathbf{x})$ with $\mathbf{x} \in \mathcal{X}$ is measurable and $\mathbb{E}(\|\psi_{\theta}(\mathbf{X})\|) < +\infty$, $\Psi_n^{LHS}(\theta)$ is an unbiased estimator of $\Psi(\theta)$ by Property 1.
2. This is a direct consequence of Property 2.
3. This is a direct consequence of Theorem 1. ■

All these properties on $\Psi_n^{LHS}(\theta)$ allow to show that $\hat{\theta}_n^{LHS}$ is a consistent estimator of θ_0 . Indeed, the assertion 2 of Property 5 ensures the convergence in probability of $\Psi_n^{LHS}(\theta)$ to $\Psi(\theta)$. As mentioned before, Properties 3 and 4 do not impose any other conditions on the sampling scheme. We therefore have, under the conditions of application of at least one of these properties, that $\hat{\theta}_n^{LHS}$ converges to θ_0 in probability. Let us now establish a Central Limit Theorem for Z-estimators under LHS.

Theorem 3 Let Θ be an open subset of an Euclidean space of dimension q with $q \in \mathbb{N}^*$ and $\mathcal{X} = [0, 1]^d$ with $d \in \mathbb{N}^*$. Assume that, for all $\theta \in \Theta$ and for all $\mathbf{x} \in \mathcal{X}$, the function $\theta \rightarrow \psi_{\theta}(\mathbf{x})$ is twice continuously differentiable in θ . The first order derivative is a size $q \times q$ matrix denoted $\dot{\psi}_{\theta}$ and the second order partial derivative is a size $q \times q \times q$ tensor denoted $\ddot{\psi}_{\theta}$. Let $\mathbf{X}^{LHS} = (\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)})^T$ be the vector of LHS realizations of a random variable $\mathbf{X} = (X_1, \dots, X_d)$ evolving in \mathcal{X} such that $\mathbf{X} \sim U_{[0,1]^d}$.

Suppose also that the following hypotheses are fulfilled:

1. $\forall n \in \mathbb{N}, \Psi_n(\hat{\theta}_n^{LHS}) = \frac{1}{n} \sum_{i=1}^n \psi_{\hat{\theta}_n^{LHS}}(\mathbf{x}^{(i)}) = 0$;
2. there is $\theta_0 \in \Theta$ such that $\mathbb{E}(\psi_{\theta_0}(\mathbf{X})) = \Psi(\theta_0) = 0$;
3. $\mathbb{E}(\|\psi_{\theta_0}(\mathbf{X})\|^2) < +\infty$;
4. $\mathbb{E}(\dot{\psi}_{\theta_0}(\mathbf{X}))$ is non-singular and such that $\mathbb{E}(\|\ddot{\psi}_{\theta_0}(\mathbf{X})\|^2) < +\infty$;
5. the function $\theta \rightarrow \psi_{\theta}(\mathbf{x})$ is dominated in norm by a fixed integrable function $\tilde{\psi}(\mathbf{x})$ for every θ in the neighborhood of θ_0 .

Then, if $\hat{\theta}_n^{LHS}$ is a consistent estimator of θ_0 , we have:

$$(\hat{\theta}_n^{LHS} - \theta_0) = -[\mathbb{E}(\dot{\psi}_{\theta_0}(\mathbf{X}))]^{-1} \frac{1}{n} \sum_{i=1}^n \psi_{\theta_0}(\mathbf{x}^{(i)}) + \frac{1}{\sqrt{n}} o_p(1). \quad (5)$$

In particular, letting $\frac{1}{n} \Sigma_{\theta_0}^{LHS}$ be the covariance matrix of $\hat{\theta}_n^{LHS}$, we have

$\Sigma_{\theta_0}^{LHS} = [\mathbb{E}(\dot{\psi}_{\theta_0}(\mathbf{X}))]^{-1} \Sigma_{\dot{\psi}_{\theta_0}}^{LHS} [\mathbb{E}(\dot{\psi}_{\theta_0}(\mathbf{X}))]^{-T} + o(1)$, with $\Sigma_{\dot{\psi}_{\theta_0}}^{LHS} = \int_{[0,1]^d} \psi_{\theta_{0,rem}}(\mathbf{x}) \psi_{\theta_{0,rem}}(\mathbf{x})^T d\mathbf{x} + o(1)$ asymptotically. Moreover, we have that $\Sigma_{\theta_0}^{LHS} - \Sigma_{\theta_0}^{IID}$ is asymptotically positive semi-definite.

In addition, if the function ψ_{θ_0} is such that $\mathbb{E}(\|\psi_{\theta_0}(\mathbf{X})\|^3) < +\infty$ and if $\mathbf{R}_{\psi_{\theta_0}} = \int_{[0,1]^d} \psi_{\theta_{0,rem}}(\mathbf{x}) \psi_{\theta_{0,rem}}(\mathbf{x})^T d\mathbf{x}$ is non-singular, we have that the sequence $\sqrt{n}(\hat{\theta}_n^{LHS} - \theta_0)$ is asymptotically normal with mean zero and a covariance matrix equal to $[\mathbb{E}(\dot{\psi}_{\theta_0}(\mathbf{X}))]^{-1} \mathbf{R}_{\psi_{\theta_0}} [\mathbb{E}(\dot{\psi}_{\theta_0}(\mathbf{X}))]^{-T}$.

Proof. The reasoning of this proof is adapted from the one given in [11] to demonstrate Theorem 2.

By Taylor's theorem, as $\Psi_n(\theta)$ is continuous and twice differentiable, $\exists \tilde{\theta}_n^{LHS}$ between θ_0 and $\hat{\theta}_n^{LHS}$ such that:

$$\Psi_n^{LHS}(\hat{\theta}_n^{LHS}) = 0 = \Psi_n^{LHS}(\theta_0) + \dot{\Psi}_n^{LHS}(\theta_0)(\hat{\theta}_n^{LHS} - \theta_0) + \frac{1}{2}(\hat{\theta}_n^{LHS} - \theta_0)^T \ddot{\Psi}_n^{LHS}(\tilde{\theta}_n^{LHS})(\hat{\theta}_n^{LHS} - \theta_0).$$

We have, since $\mathbb{E}(\|\psi_{\theta_0}(\mathbf{X})\|^2) < +\infty$ that $\Psi_n^{LHS}(\theta_0) = \frac{1}{n} \sum_{i=1}^n \psi_{\theta_0}(\mathbf{x}^{(i)})$ converges in probability to $\mathbb{E}(\psi_{\theta_0}(\mathbf{X})) = 0$, thanks to Property 5: $\Psi_n^{LHS}(\theta_0) \xrightarrow[n \rightarrow +\infty]{p} \mathbb{E}(\psi_{\theta_0}(\mathbf{X})) = 0$.

Similarly, since $\dot{\Psi}_n^{LHS}(\theta)$ is the empirical mean of the matrix $\dot{\psi}_{\theta}$ over $\mathcal{X}^{LHS} = (\mathbf{x}^{(1)} \dots \mathbf{x}^{(n)})^T$, then $\dot{\Psi}_n^{LHS} \xrightarrow[n \rightarrow +\infty]{p} \mathbb{E}(\dot{\psi}_{\theta_0}(\mathbf{X}))$ (with $\mathbb{E}(\|\dot{\psi}_{\theta_0}(\mathbf{X})\|^2) < +\infty$). In addition, thanks to assumption 4, we also have that $\mathbb{E}(\dot{\psi}_{\theta_0}(\mathbf{X}))$ is non-singular.

The term $\ddot{\Psi}_n^{LHS}(\theta)$ corresponds to the empirical mean of the $q \times q \times q$ tensor $\ddot{\psi}_{\theta}$ over $\mathcal{X}^{LHS} = (\mathbf{x}^{(1)} \dots \mathbf{x}^{(n)})^T$. By assumption 5, there is a ball \mathcal{B} around θ_0 such that $\ddot{\psi}_{\theta}$ is dominated in norm by $\|\ddot{\psi}\|$. Since we have $\hat{\theta}_n^{LHS} \xrightarrow[n \rightarrow +\infty]{p} \theta_0$, we also have $\mathbb{P}(\tilde{\theta}_n^{LHS} \in \mathcal{B}) \xrightarrow[n \rightarrow +\infty]{} 1$. Moreover, if $\tilde{\theta}_n^{LHS} \in \mathcal{B}$, we have:

$$\|\ddot{\Psi}_n^{LHS}(\tilde{\theta}_n^{LHS})\| = \|\frac{1}{n} \sum_{i=1}^n \ddot{\psi}_{\tilde{\theta}_n^{LHS}}(\mathbf{x}^{(i)})\| \leq \frac{1}{n} \sum_{i=1}^n \|\ddot{\psi}(\mathbf{x}^{(i)})\|.$$

Since $\|\ddot{\psi}\|$ is integrable and the right term is an empirical mean, it converges to a finite value thanks to Property 2. Hence, it is the case for the left term.

So that, we can rewrite the Taylor's expansion as follows:

$$-\Psi_n^{LHS}(\theta_0) = \left(\mathbb{E}(\dot{\psi}_{\theta_0}(\mathbf{X})) + o_p(1) + \frac{1}{2}(\hat{\theta}_n^{LHS} - \theta_0)^T O_p(1) \right) (\hat{\theta}_n^{LHS} - \theta_0).$$

Since $\hat{\theta}_n$ converges in probability to θ_0 , we have that $(\hat{\theta}_n - \theta_0)O_p(1)$ converges to 0 and thus:

$$-\Psi_n^{LHS}(\theta_0) = (\mathbb{E}(\dot{\psi}_{\theta_0}(\mathbf{X})) + o_p(1))(\hat{\theta}_n^{LHS} - \theta_0).$$

This leads to the equation given in 3, considering that $\mathbb{E}(\dot{\psi}_{\theta_0})$ is non-singular and that we have asymptotically

$$\Psi_n^{LHS}(\theta_0) = \frac{1}{\sqrt{n}} O_p(1):$$

$$(\hat{\theta}_n^{LHS} - \theta_0) = -[\mathbb{E}(\dot{\psi}_{\theta_0}(\mathbf{X}))]^{-1} \frac{1}{n} \sum_{i=1}^n \psi_{\theta_0}(\mathbf{x}^{(i)}) + \frac{1}{\sqrt{n}} o_p(1). \quad (6)$$

Using this result, we directly have $\Sigma_{\theta_0}^{LHS} = [\mathbb{E}(\dot{\psi}_{\theta_0}(\mathbf{X}))]^{-1} \Sigma_{\psi_{\theta_0}}^{LHS} [\mathbb{E}(\dot{\psi}_{\theta_0}(\mathbf{X}))]^{-T} + o(1)$. Moreover, we have that $\Sigma_{\psi_{\theta_0}}^{IID} - \Sigma_{\psi_{\theta_0}}^{LHS}$ is asymptotically positive semi-definite thanks to Property 5. Thus, this is also the case for $\Sigma_{\theta_0}^{IID} - \Sigma_{\theta_0}^{LHS}$.

Finally, if we suppose that $\mathbb{E}(\|\psi_{\theta_0}(\mathbf{X})\|^3) < +\infty$ and that $\mathbf{R}_{\psi_{\theta_0}}$ is non-singular, we have thanks to the assertion 3 of Property 5 that $\sqrt{n}(\hat{\theta}_n^{LHS} - \theta_0)$ is asymptotically normal with mean zero and a covariance matrix equal to $[\mathbb{E}(\dot{\psi}_{\theta_0}(\mathbf{X}))]^{-1} \mathbf{R}_{\psi_{\theta_0}} [\mathbb{E}(\dot{\psi}_{\theta_0}(\mathbf{X}))]^{-1}$. ■

These results give an asymptotic convergence for $\hat{\theta}_n^{LHS}$ with, in the univariate case, a lower asymptotic variance of estimation than $\hat{\theta}_n^{IID}$ (corresponding to $n(\Sigma_{\hat{\theta}_n^{IID}} - \Sigma_{\hat{\theta}_n^{LHS}})$ being asymptotically positive semi-definite in the multivariate case). Moreover, it gives a central limit theorem for Z -estimators under LHS. While strong regularity conditions on ψ_{θ} are needed for these results to be valid, it remains very useful in many practical cases (eg. for estimation by maximum likelihood). In the next section, we give an example of application.

5 | APPLICATION: PARAMETERS ESTIMATION OF GENERALIZED LINEAR MODELS (GLM)

In the context of the statistical analysis of a computational code, it is common that one wants to approximate its outputs with a regression model. If the estimation of the modeling parameters can be expressed as a Z -estimator and the other conditions of use are satisfied, Theorem 3 ensures that the estimation variance of these parameters is asymptotically lower under LHS than under IID sampling. It also provides a central limit theorem under LHS.

Consider for instance the case of Generalized Linear Models (GLM), proposed in [15]. They were formulated as a way of unifying various statistical models, including linear regression, logistic regression and Poisson regression. To estimate the parameters of a GLM, one generally uses a Maximum Likelihood Estimator (MLE). It is therefore a special case of Z -estimation supposing that the likelihood can be differentiated. Thus, the results presented above can be applied to parameters estimation of a GLM.

5.1 | Definitions and main properties on GLM

Before entering in more details, let us first define GLM more formally. For simplicity and without loss of generality, we focus here on the canonical case. A generalized linear model aims to explain a random variable Z defined in a measurable space $\mathcal{Z} \subset \mathbb{R}$ with a vector of covariables $\mathbf{X} = (X_1, \dots, X_d)$, $d \in \mathbb{N}^*$ evolving in a measurable space $\mathcal{X} \subset \mathbb{R}^d$. GLM are characterized by three components: a probability distribution that models Z , a linear predictor and a link function.

1. The probability distribution is the random component of the model. In the GLM framework, the probability distribution associated with Z belongs to the exponential family. This density function is therefore expressed, for $z \in \mathcal{Z}$,

$\alpha \in \mathbb{R}$ and $\phi > 0$, as follow (canonical form):

$$f(z, \alpha, \phi) = a(\alpha)b(z) \exp\left(z \frac{\alpha}{\phi}\right). \quad (7)$$

Here,

- ϕ is the dispersion parameter (assumed to be known);
 - $a(\alpha) = \exp\left(-\frac{v(\alpha)}{\phi}\right)$, with v being a known twice continuously differentiable function;
 - $b(z) = \exp(w(z, \phi))$ with w being a known function of z and ϕ .
2. The linear predictor is the quantity which incorporates the information about the covariables \mathbf{X} into the model. Let $\theta = (\theta_1, \dots, \theta_d)^T \in \Theta$ be the modeling parameter, with Θ being an open-bounded subset of \mathbb{R}^d . The linear predictor of the GLM η is thus defined as follow, for $\mathbf{x} = (x_1, \dots, x_d) \in \mathcal{X}$:

$$\eta = \sum_{j=1}^d x_j \theta_j = \mathbf{x}\theta. \quad (8)$$

3. Finally, the link function expresses the functional relationship between the random component and the linear predictor. Let h be a function that is monotone and differentiable on $\mathcal{H} \subset \mathbb{R}$. Note that this hypothesis on h implies the existence of the inverse function h^{-1} so that, for any $a \in \mathcal{H}$, $h^{-1} \circ h(a) = a$. Let us also denotes by μ the expected value of Z such that $\mu = \mathbb{E}(Z)$. The canonical link function is then defined by the following equation:

$$h(\mu) = \alpha = \eta. \quad (9)$$

In addition to this definition, GLMs are characterized by properties concerning their first and second-order moments. Indeed, we have the following equalities:

- $\mathbb{E}(Z) = \mu = \dot{v}(\alpha)$;
- $\text{Var}(Z) = \ddot{v}(\alpha)\phi$.

Here, \dot{v} and \ddot{v} are respectively the first and the second order derivatives of v in α .

Let $\mathbf{X} = (\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)})^T$ be the n realizations of \mathbf{X} and $\mathbf{Z} = (z^{(1)}, \dots, z^{(n)})^T$ be the n realizations modeled by a GLM. We also denote by $\boldsymbol{\alpha} = (\alpha^{(1)}, \dots, \alpha^{(n)})^T$ the vector linked to θ and the observations \mathbf{X} by Equation 9: for all $i \in \llbracket 1, n \rrbracket$, $\alpha^{(i)} = h^{-1}(\mathbf{x}^{(i)}\theta)$.

Suppose that one wants to estimate the optimal value $\theta_0 = (\theta_{0,1}, \dots, \theta_{0,d})^T$ of the parameter θ regarding the observations of \mathbf{x} and \mathbf{Z} . The likelihood l of an observation $z^{(i)}, \alpha^{(i)}$ with $i \in \llbracket 1, n \rrbracket$ is given by the following equation:

$$l(z^{(i)}, \mathbf{x}^{(i)}, \theta, \phi) = \log(f(z^{(i)}, \alpha^{(i)}, \phi)) = \frac{1}{\phi} [z^{(i)}\alpha^{(i)} - v(\alpha^{(i)})] + w(z^{(i)}, \phi). \quad (10)$$

If the n realizations are independent, the global likelihood of the observations corresponds to

$L(\mathbf{Z}, \mathbf{X}, \theta, \phi) = \prod_{i=1}^n l(z^{(i)}, \mathbf{x}^{(i)}, \theta, \phi)$. The maximum likelihood estimator corresponds then to $\hat{\theta}_n = \underset{\theta \in \Theta}{\operatorname{argmax}}(L(\mathbf{Z}, \mathbf{X}, \theta, \phi))$. Since the log function is strictly increasing on $]0, +\infty[$, we also have that $\underset{\theta \in \Theta}{\operatorname{argmax}}((\log [L(\mathbf{Z}, \mathbf{X}, \theta, \phi)])$ with $\log [L(\mathbf{Z}, \mathbf{X}, \theta, \phi)] = \sum_{i=1}^n \log [l(z^{(i)}, \mathbf{x}^{(i)}, \theta, \phi)]$.

Since $\log[l(z^{(i)}, \mathbf{x}^{(i)}, \theta, \phi)]$ is differentiable in θ , the maximum likelihood estimator $\hat{\theta}_n$ is a solution of the vectorial equation:

$$\sum_{i=1}^n \nabla \log[l(z^{(i)}, \mathbf{x}^{(i)}, \theta, \phi)] = 0, \quad (11)$$

with $\nabla \log[l(z^{(i)}, \mathbf{x}^{(i)}, \theta, \phi)] = \left(\frac{\partial \log[l(z^{(i)}, \mathbf{x}^{(i)}, \theta, \phi)]}{\partial \theta_1}, \dots, \frac{\partial \log[l(z^{(i)}, \mathbf{x}^{(i)}, \theta, \phi)]}{\partial \theta_d} \right)^T$.

The quantity $\hat{\theta}_n$ is therefore a Z -estimator of θ_0 as defined in Equation (3), with $\psi_{\theta}(\mathbf{x}^{(i)}) = (\psi_{\theta_1}(\mathbf{x}^{(i)}), \dots, \psi_{\theta_d}(\mathbf{x}^{(i)}))^T = \nabla \log[l(z^{(i)}, \mathbf{x}^{(i)}, \theta, \phi)]$.

Moreover, in the canonical case, we have for $j \in \llbracket 1, d \rrbracket$ and $\mathbf{x} = (x_1, \dots, x_d) \in \mathcal{X}$:

$$\psi_{\theta_j}(\mathbf{x}) = \frac{\partial \log[l(z, \mathbf{x}, \theta, \phi)]}{\partial \theta_j} = \frac{z - h^{-1}(\mathbf{x}\theta)}{\phi} x_j. \quad (12)$$

We can see that the estimation of the parameters of a GLM by maximum likelihood fits into the framework of Z -estimation. Thus, let us suppose that the observations of \mathbf{X} are obtained by a LHS. We consider the Z -estimator defined by the equation 11, even though in this case, the realization are not *i.i.d* anymore. Let us discuss the convergence of this estimator under LHS.

5.2 | Z -estimation of GLM parameters under LHS

Let $X^{LHS} = (\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)})^T$ be the realizations of \mathbf{X} generated by a LHS. As before, we suppose for simplicity that we have $\mathcal{X} = [0, 1]^d$ and $\mathbf{X} \sim U_{[0,1]^d}$. We also assume that Θ is an open bounded subset of \mathbb{R}^d . Moreover, we suppose that h is defined, and is twice continuously derivable on \mathcal{H} . We also suppose that h and its first derivative \dot{h} have no zero on \mathcal{H} . Since h is monotone by construction, note that h^{-1} is also defined and twice continuously derivable for any $\theta \in \Theta$ and $x \in \mathcal{X}$ thanks to the inverse function theorem (see for instance [16] for more details).

Since we suppose that \mathbf{X} and θ are bounded, we have that $\Psi_n^{LHS} = \sum_{i=1}^n \psi_{\theta}(\mathbf{x}^{(i)})$ converges in probability to $\mathbb{E}(\psi_{\theta}(\mathbf{X})) = \Psi(\theta)$. As we have seen, the other conditions concerning the convergence of $\hat{\theta}_n$ to θ_0 are not specific to the sampling scheme. The conditions of application of Property 4 (resp 3) are verified both in the case of an IID or a LHS design. We can thus conclude, thanks to Property 4 that $\hat{\theta}_n$ converges in probability in θ_0 .

Let us now verify that the conditions of application of Theorem 3 are fulfilled. First, we see that Ψ_n is continuous and infinitely differentiable in θ . Plus, $\mathbb{E}(\psi_{\theta_0}) = 0$ by construction.

We also have, for $j, k \in \llbracket 1, d \rrbracket$ and $\mathbf{x} = (x_1, \dots, x_d)^T \in \mathcal{X}$:

$$\frac{\partial \psi_{\theta_j}(\mathbf{x})}{\partial \theta_k} = \frac{-1}{\phi \dot{h}(h^{-1}(\mathbf{x}\boldsymbol{\theta}))} x_j x_k. \quad (13)$$

Thus, we have that the matrix of partial derivatives $\dot{\psi}_{\theta_0}$ is such that $\mathbb{E}(\dot{\psi}_{\theta_0}(\mathbf{X}))$ is defined and non-singular since h has no zero on \mathcal{H} . Since the values of \mathbf{X} and $\boldsymbol{\theta}$ are bounded in norm, the function ψ_{θ_0} is bounded and thus $\mathbb{E}(\|\dot{\psi}_{\theta_0}(\mathbf{X})\|^3) < +\infty$.

Finally, we have that the elements of the tensor $\ddot{\psi}_{\theta_0}(\mathbf{x})$ are, for $j, k, l \in \llbracket 1, d \rrbracket$ and $\mathbf{x} = (x_1, \dots, x_d)^T \in \mathcal{X}$, as follows:

$$\frac{\partial^2 \psi_{\theta_j}(\mathbf{x})}{\partial \theta_k \partial \theta_l} = \frac{\ddot{h}(h^{-1}(\mathbf{x}\boldsymbol{\theta}))}{\phi (\dot{h}(h^{-1}(\mathbf{x}\boldsymbol{\theta})))^3} x_j x_k x_l. \quad (14)$$

Thus, $\|\ddot{\psi}_{\theta_0}(\mathbf{x})\|$ can be bounded by an integrable function since we assume that the values of $\boldsymbol{\theta}$ and \mathbf{X} are bounded.

All of these allows us to apply Theorem 3. We therefore have that the covariance matrix of estimation of $\hat{\theta}_n^{LHS}$ is equal to $\frac{1}{n} \boldsymbol{\Sigma}_{\theta_0}^{LHS} = \frac{1}{n} [\mathbb{E}(\dot{\psi}_{\theta_0}(\mathbf{X}))]^{-1} \boldsymbol{\Sigma}_{\dot{\psi}_{\theta_0}}^{LHS} [\mathbb{E}(\dot{\psi}_{\theta_0}(\mathbf{X}))]^{-T} + \frac{1}{n} o(1)$ and that $\boldsymbol{\Sigma}_{\theta_0}^{IID} - \boldsymbol{\Sigma}_{\theta_0}^{LHS}$ is asymptotically positive semidefinite. Note that we have, with the previously introduced notations, $\boldsymbol{\Sigma}_{\dot{\psi}_{\theta_0}}^{LHS} = \int_{[0,1]^d} \psi_{\theta_{0,rem}}(\mathbf{x}) \psi_{\theta_{0,rem}}(\mathbf{x})^T d\mathbf{x} + o(1)$ asymptotically.

Finally, if we suppose that $\mathbb{E}(\|\dot{\psi}_{\theta_0}(\mathbf{X})\|^3) < +\infty$, we have that $\sqrt{n}(\hat{\theta}_n^{LHS} - \theta_0)$ is asymptotically normal with mean zero and a covariance matrix equal to $[\mathbb{E}(\dot{\psi}_{\theta_0}(\mathbf{X}))]^{-1} \mathbf{R}_{\dot{\psi}_{\theta_0}} [\mathbb{E}(\dot{\psi}_{\theta_0}(\mathbf{X}))]^{-1}$, assuming that

$\mathbf{R}_{\dot{\psi}_{\theta_0}} = \int_{[0,1]^d} \psi_{\theta_{0,rem}}(\mathbf{x}) \psi_{\theta_{0,rem}}(\mathbf{x})^T d\mathbf{x}$ is non-singular.

5.3 | Numerical example: a Poisson regression under LHS

To illustrate this result, let us consider a numerical example with a count random variable Z and a vector of covariables $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_9)^T$ ($\mathbf{X} \sim U_{[0,1]^9}$). In an industrial context, Z could represent for instance the number of operating problems evaluated by a simulation code of an industrial facility. In this example, we define Z by the following Poisson density function, for $z \in \mathbb{N}^*$:

$$f(z, \lambda_0) = \exp(-\lambda_0) \frac{1}{z!} \exp(z \log(\lambda_0)), \quad (15)$$

with $\log(\lambda_0) = \mathbf{x}\boldsymbol{\theta}_0$, $\mathbf{x} \in [0, 1]^9$ and $\boldsymbol{\theta}_0 = (\theta_{0,1}, \dots, \theta_{0,9})^T = (10, -\sqrt{2}, 1/2, -1/3, \sqrt{5}, -10, \sqrt{2}, -1/2, -\sqrt{5})^T$.

One can notice Z fits in the framework of Equations 7, 8 and 9.

Let us compare numerically the performances of the maximum likelihood estimation of θ_0 regarding the sampling method (IID or LHS) in this example. To do so, we compare the variance of estimation of each parameters $(\theta_1, \dots, \theta_{10})^T$ regarding the sampling scheme and size. We also verify that there is no significant differences concerning the square

bias of estimation $[\mathbb{E}(\hat{\theta}_j) - \theta_{0,j}]^2, j \in \llbracket 1, 9 \rrbracket$. For each sampling method, the average value of these two metrics is computed over $L = 1000$ independent LHS and IID designs with sample sizes n being from 40 to 100 (by steps of 10).

Figure 2 and 3 show respectively the evolution of the variance and the square bias of estimation of the nine estimated parameters $(\theta_{0,1}, \dots, \theta_{0,9})^T$. As expected, we observe that for the nine estimated parameters, the average variance of estimation is overall lower for the classic LHS design compared to IID. No significant differences between LHS and IID designs are observed in terms of the square bias of estimation. As shown previously, classic LHS designs allow better estimation performances than IID ones, regardless of the theoretical value of the estimated parameters.

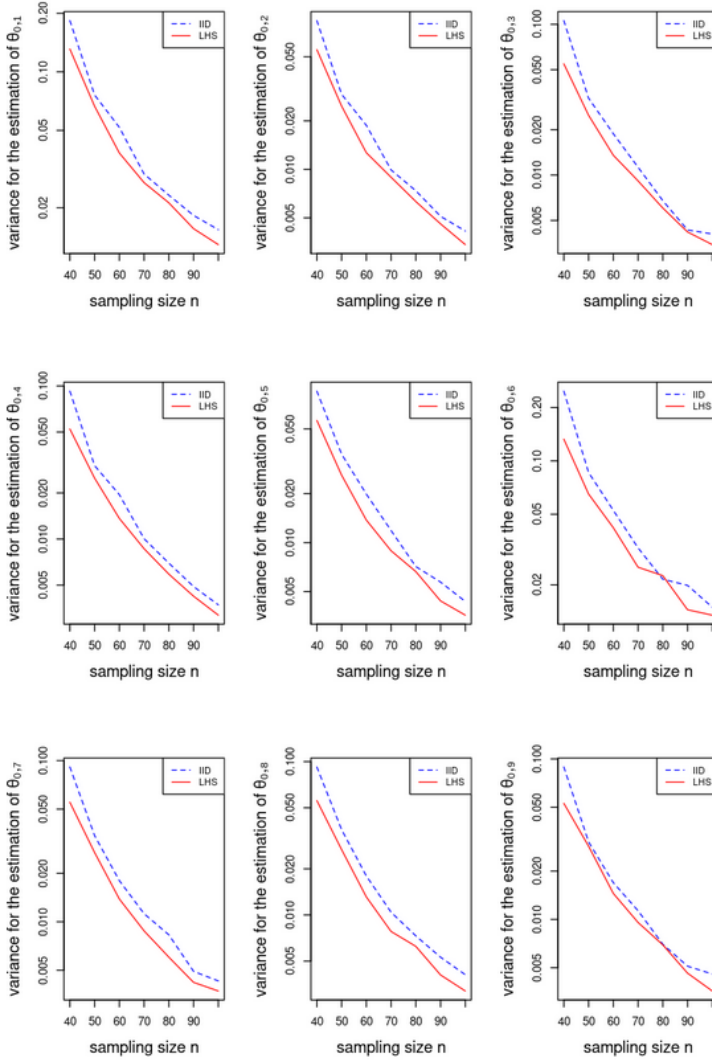


Figure 2 Average estimation variances of $(\theta_{0,1}, \dots, \theta_{0,9})^T$ according to the sampling size n for IID and LHS designs (decimal logarithmic scale).

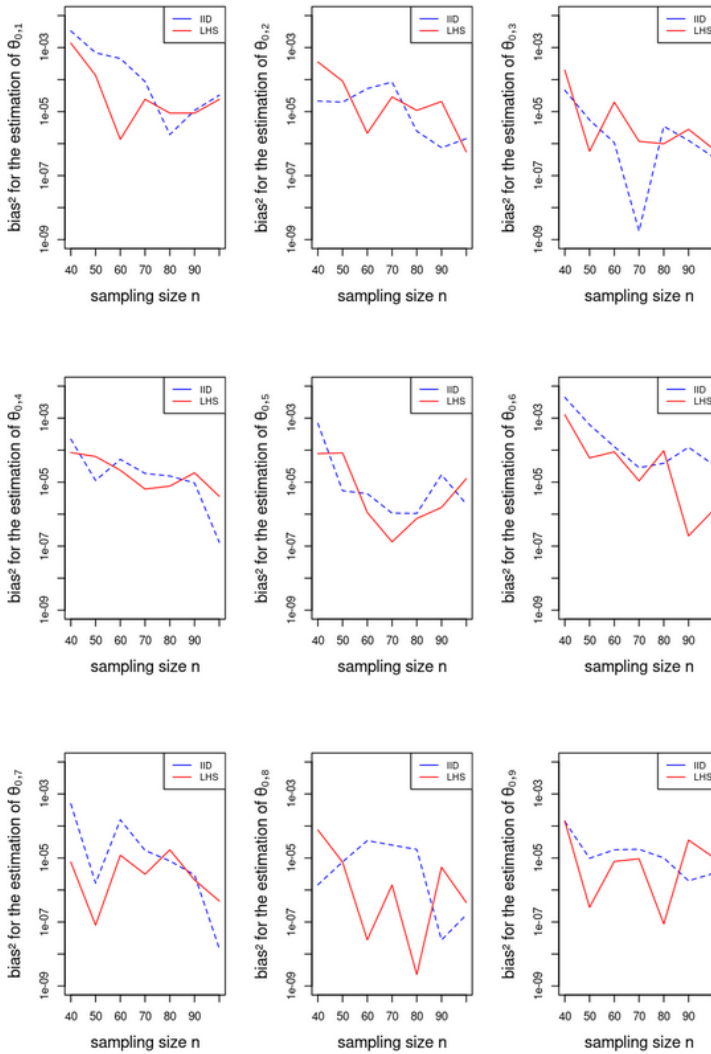


Figure 3 Average estimation square bias of $(\theta_{0,1}, \dots, \theta_{0,9})^T$ according to the sampling size n for IID and LHS designs (decimal logarithmic scale).

6 | CONCLUSION AND PROSPECTS

The work presented here is a continuation of what was previously done on the asymptotic convergence of estimators under Latin Hypercube Sampling (LHS). Convergence results proposed in [5], [6] and [7] concerning the estimation of the empirical mean have been extended to the class of Z-estimators. In particular, a central limit theorem for this class of estimators under LHS with a lower asymptotic variance than in the simple *independent and identically distributed (i.i.d)* random sampling has been proposed. An example of application of these results for the estimation of GLM parameters has also been presented. Nevertheless, some restrictive regularity conditions have been imposed

on these estimators to establish this convergence. A possible perspective on this work could therefore be to lift some of these restrictions, such as the existence of the Z -function second derivative for instance. Indeed, versions of the central limit theorem for Z -estimators that are not assuming the existence of a second derivative of the Z -function are proposed (in [11] for example). This could be used to obtain more general results than what is proposed here.

Acknowledgements

Support from my supervisors (Claude Brayer, Fabrice Gamboa, Benoît Habert and Amandine Marrel) during my PhD thesis, of which this work is a continuation, is gratefully acknowledged. I would also like to thank Anouar Meynaoui for his help in writing this paper.

references

- [1] Mckay M, Beckman R, Conover W. A Comparison of Three Methods for Selecting Vales of Input Variables in the Analysis of Output From a Computer Code 21. *Technometrics* 1979;p. 239–245.
- [2] Helton JC, Davis FJ. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. *Reliability Engineering & System Safety* 81 2003;p. 23–69.
- [3] Helton JC, Davis FJ, Johnson JD. A comparison of uncertainty and sensitivity analysis results obtained with random and Latin hypercube sampling. *Reliability Engineering & System Safety* 89 2005;p. 305–330.
- [4] Viana FA. A tutorial on Latin hypercube design of experiments. *Quality and reliability engineering international* 2016;32(5):1975–1985.
- [5] Stein M. Large Sample Properties of Simulations Using Latin Hypercube Sampling. *Technometrics* 2 1987;p. 143–151.
- [6] Owen AB. A Central Limit Theorem for Latin Hypercube Sampling. *Journal of the Royal Statistical Society 54: Series B (Methodological)* 1992;p. 541–551.
- [7] Loh WL. On latin hypercube sampling. *Annals of Statistics* 24 1996;p. 2058–2080.
- [8] Hakimi F. High-dimensional sensitivity analysis methods for computationally expensive simulators modeling a severe nuclear accident. PhD thesis, Paul Sabatier University, Toulouse, France; 2023.
- [9] Devroye L. 2: General principles in random variate generation. In: *Non-Uniform Random Variate Generation*(originally published with Springer-Verlag; 1986. p. 27–39.
- [10] Rousseeuw PJ, Hampel FR, Ronchetti EM, Stahel WA. *Robust Statistics: The Approach Based on Influence Functions*. Wiley; 1986.
- [11] Van der Vaart AW. *Asymptotic statistics*. Cambridge University Press; 1998.
- [12] Dacunha-Castelle D, Duflo M. *Probability and Statistics Volume II*. Springer-Verlag; 1986.
- [13] Fisher RA. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London, Series A* 222;p. 309–368.
- [14] Fischer H. *A History of the Central Limit Theorem: From Classical to Modern Probability Theory*. New York: Springer; 2011.
- [15] Nelder J, Wedderburn R. Generalized Linear Models. *Journal of the Royal Statistical Society Series A (General)* Blackwell Publishing 135;p. 370–384.
- [16] Allendoerfer CB. *Calculus of Several Variables and Differentiable Manifolds*. New York: Macmillan; 1974.