C. R. Acad. Sci. Paris, Ser. I 355 (2017) 10–14

Contents lists available at ScienceDirect

C. R. Acad. Sci. Paris, Ser. I

www.sciencedirect.com

Group theory/Statistics

Iterative construction of replicated designs based on Sobol' sequences

Construction de plans répliqués à partir de séquences de Sobol'

Laurent Gilquin [a], Lluís Antoni Jiménez Rugama [b], Élise Arnaud [c], Fred J. Hickernell [b], Hervé Monod [d], Clémentine Prieur [c]

[a] *Inria Grenoble – Rhône-Alpes, Inovallée, 655, avenue de l'Europe, 38330 Montbonnot, France*
[b] *Illinois Institute of Technology, Rettaliata Engineering Center, 10 W 32st, Chicago, IL 60616, USA*
[c] *Université Grenoble Alpes, INRIA/LJK, CNRS, 38000 Grenoble, France*
[d] *MaIAGE, INRA, Université Paris-Saclay, 78350 Jouy-en-Josas, France*

ARTICLE INFO

Article history:
Received 7 July 2016
Accepted after revision 28 November 2016
Available online 7 December 2016

Presented by the Editorial Board

ABSTRACT

In the perspective of estimating main effects of model inputs, two approaches are studied to iteratively construct replicated designs based on Sobol' sequences. Space-filling properties of the resulting designs are studied based on two criteria.

RÉSUMÉ

Dans l'objectif d'estimer les effets principaux des paramètres d'un modèle, nous proposons d'étudier deux approches pour construire itérativement des plans répliqués à partir de séquences de Sobol'. Les propriétés de remplissement de l'espace des plans construits sont étudiées sur la base de deux critères.

1. Introduction

Mathematical models often involve a substantial number of poorly known parameters. The effect of these parameters on the output of the model can be assessed through sensitivity analysis. Global sensitivity analysis methods are useful tools to identify the parameters having the most influence on the output. A well-known approach is the variance-based method introduced by Sobol' in [11]. This method estimates sensitivity indices called Sobol' indices that summarize the influence of each model input. Among all Sobol' indices, one can distinguish the first-order indices that estimate the main effect of each input.

E-mail addresses: laurent.gilquin@inria.fr (L. Gilquin), ljimene1@hawk.iit.edu (L.A. Jiménez Rugama).

http://dx.doi.org/10.1016/j.crma.2016.11.013

The procedure to estimate first-order Sobol' indices proposed by Sobol' and its improvements (see Saltelli [9] for an exhaustive survey) all suffer from a prohibitive number of model evaluations that grows with respect to the input space dimension. An elegant solution to reduce this number relies on the construction of particular designs of experiments called replicated designs. The notion of replicated designs was first introduced by McKay through its definition of replicated Latin Hypercubes in [4]. Below we provide this definition in a more general framework:

Definition 1.1. Consider $\boldsymbol{x} \in [0,1)^s$, and $\boldsymbol{x}_u \in [0,1)^{|u|}$ the subset of elements of $\boldsymbol{x}$ given by $u \subsetneq \{1,\dots,s\}$, where $|u|$ is the cardinality of u. Let $\mathcal{P} = \{\boldsymbol{x}_i\}_{i=0}^{n-1}$ and $\mathcal{P}' = \{\boldsymbol{x}'_i\}_{i=0}^{n-1}$ be two point sets in $[0,1)^s$, and denote by $\mathcal{P}^u = \{\boldsymbol{x}_{i,u}\}_{i=0}^{n-1}$ (resp. $\mathcal{P}'^u$) the subset of elements of points in $\mathcal{P}$ (resp. $\mathcal{P}'$) indexed by u. We say that $\mathcal{P}$ and $\mathcal{P}'$ are two replicated designs of order $a = 1,\dots,s-1$, if for any $u \subsetneq \{1,\dots,s\}$ with $|u| = a$, $\mathcal{P}^u$ and $\mathcal{P}'^u$ are the same point set in $[0,1)^a$, perhaps in a different order.

The replication procedure described in [2,12] allows the estimation of all first-order Sobol' indices with only two replicated designs of order 1. This procedure has the major advantage of reducing the number of model evaluations, evaluating only on designs $\mathcal{P}$ and $\mathcal{P}'$ regardless of the input space dimension. However, Sobol' indices estimates may still not be accurate enough if designs $\mathcal{P}$ and $\mathcal{P}'$ do not explore the input space properly.

In this note, we propose two different constructions of replicated designs of order 1 based on Sobol' sequences. Both constructions ensure that the input space is properly explored and can be used within the replication procedure to estimate all first-order Sobol' indices. The definition of these constructions is recursive, therefore one can iteratively refine each replicated design by adding the corresponding new set of points. We first provide a brief introduction on digital sequences and then present two iterative approaches to construct the two replicated point sets. We end this note by analyzing the space-filling properties of the two designs constructed.

2. Digital sequences background

2.1. Preliminaries

Digital nets and sequences were first introduced by Niederreiter [6] in the numerical integration framework to define good uniformly distributed points in $[0,1)^s$. They can also appear in the literature as digital (t,m,s)-nets and digital (t,s)-sequences, or simply (t,m,s)-nets and (t,s)-sequences. Sobol' and Niederreiter–Xing sequences are two examples of digital sequences detailed in [10] and [7].

Definition 2.1. Let $\mathcal{A}$ be the set of all elementary intervals $A \subset [0,1)^s$ where $A = \prod_{j=1}^{s}[\alpha_j b^{-\gamma_j}, (\alpha_j+1)b^{-\gamma_j})$, with integers $s \ge 1$, $b \ge 2$, $\gamma_j \ge 0$, and $b^{\gamma_j} > \alpha_j \ge 0$. For $m \ge t \ge 0$, the point set $\mathcal{P} \in [0,1)^s$ with b^m points is a (t,m,s)-*net* in base b if every A with volume b^{t-m} contains b^t points of $\mathcal{P}$.

Thus, a (t,m,s)-net is defined such that all elementary intervals of volume b^{t-m} will enclose the same proportion of points of $\mathcal{P}$, namely $b^{t-m}|\mathcal{P}|$ points. The most evenly spread nets are $(0,m,s)$-nets, since each elementary interval of the smallest volume possible, b^{-m}, contains exactly one point. The quality of any (t,m,s)-net is therefore measured by the parameter t, called t-value.

By increasing m, we increase the number of points of the (t,m,s)-net. In the limiting case where $m \to \infty$, we can define the (t,s)-sequence as:

Definition 2.2. For integers $s \ge 1$, $b \ge 2$, and $t \ge 0$, the sequence $\{\boldsymbol{x}_i\}_{i\in\mathbb{N}_0}$ is a (t,s)-sequence in base b, if for every set $\mathcal{P}_{\ell,m} = \{\boldsymbol{x}_i\}_{i=\ell b^m}^{(\ell+1)b^m-1}$ with $\ell \ge 0$ and $m \ge t$, $\mathcal{P}_{\ell,m}$ is a (t,m,s)-net in base b.

The replicated design properties can also apply to digital sequences. Hence, we introduce the following definition,

Definition 2.3. Two digital sequences $\{\boldsymbol{x}_i\}_{i\in\mathbb{N}_0}$ and $\{\boldsymbol{x}'_i\}_{i\in\mathbb{N}_0}$ are *digitally replicated* of order a if for all $m \ge 0$, $\{\boldsymbol{x}_i\}_{i=0}^{b^m-1}$ and $\{\boldsymbol{x}'_i\}_{i=0}^{b^m-1}$ are two replicated designs of order a.

2.2. Sobol' sequences

Sobol' sequences in dimension s are digital sequences in base $b = 2$ that can be computed using the *generating matrices*, a set of s full-rank infinite dimensional upper triangular matrices over the Galois field $\mathbb{F}_2 := \{0,1\}$. These generating matrices are recursively constructed given some primitive polynomials and initial directional numbers. In [1], Kuo and Joe detail this construction and also suggest a particular choice for these matrices that optimize the 2-dimensional projection t-values.