7th International Conference on Sensitivity Analysis of Model Output, July 1–4 2013, Nice

Revisiting Morris Method: a polynomial algebra for design definition with increased efficiency and observability

Jean-Marc, Fédou
Gilles, Menez
Luc, Pronzato
Maria-João, Rendas
Laboratoire I3S, CNRS-UNS, Nice, France

Morris scheme for One At a Time (OAT) designs for sensitivity analysis [1] is widely used for rapid identification of the groups of important (further classified into linear or mixed/non-linear) and unimportant inputs of a multivariate function $f(x), x \in A \subset \mathbb{R}^k$, and is particularly relevant for models whose execution is computationally expensive and time consuming [2].

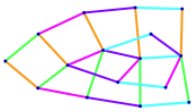
Morris designs are restricted to points in a finite k-dimensional grid $\mathcal{G} \subset A$ covering the domain of f. Stated in simple terms, the method starts by randomly evaluating $f(\cdot)$ at r "initial points" $\{x^{(n)}\}_{n=1}^r$ in $\mathcal{G}$. Starting at each of these r points, say $x^{(n)}$, k successive evaluations of $f(\cdot)$ are made, each two consecutive points enabling the determination of an elementary effect $EE_i^{(n)}$ along a distinct direction $i \in \{1, \ldots, k\}$. Morris designs are thus composed of r paths in $\mathcal{G}$ of size $k+1$, which do not have two segments along the same direction. Each input factor x_i is then classified as irrelevant, linear or other (non-linear or involved in cross-effects), depending on the first and second order statistics of the set of elementary effects $\{EE_i^{(n)}\}_{n=1}^r$ observed. The attractiveness of Morris Elementary Effects method relies on the fact that the size of the designs required to detect the important input factors of $f(\cdot)$ is linear in the number of input factors (being equal to $r(k+1)$) irrespective of the resolution of the grid $\mathcal{G}$, providing an efficient initial screening of the sensitivity of $f(\cdot)$ with respect to each input factor.

Given its higher efficiency, the clustered version of Morris OAT designs (see [1], Section 5), that computes $m > 1$ Elementary Effects $\{EE_i^{(j)}(x)\}_{j=1}^m$ along all directions $i = 1, \ldots, k$ of the input space in the neighbourhood of each point $x^{(n)}$, is especially appealing. Surprisingly, these more complex designs seem to have attracted much less interest than the original ($m = 1$) version. Possible reasons are the lack of a constructive method for finding these designs, and, presumably, concerns about the impact of residual correlation amongst the resulting set of elementary effects along each direction, see [4].

In this communication we complete the original presentation of Morris [1], giving a formal specification of a family of balanced clustered designs for arbitrary values of k and $m \leq 2^{k-1}$ (Morris construction is valid only for pairs (k, m) where k is not prime and m is a divisor of k). Our construction is supported on the definition of an isometry between sub-graphs of the unit cube Q_k equipped of the Manhattan metric, and a set of polynomials in $(x_1, \ldots, x_k)$ on which a convenient inner product is defined. This isometry, based on the association $(s_i)_{i=1}^d \in Q_d \hookrightarrow X_1^{s_1} \cdots X_d^{s_d}$, enables explicit symbolic representation and manipulation of designs, as well as the formal demonstration of their properties. We define (k, m)-edge balanced designs as those that enable the determination of exactly m elementary effects for each direction. Using our polynomial representation this is equivalent to finding the polynomial solutions to a set of equations. The computation of the set of elementary effects provided by any given subgraph is then immediate using our polynomial representation, even if the designs are no longer OAT.

A natural but more complex algebra over the set of polynomials also enables the extension of Morris concept of clustered designs to the estimation of two-factor interaction effects, $SEE_{ij}(x), i \neq j \in \{1, \ldots, k\}$, that can detect the presence of products $x_i x_j$ of pairs of input factors. We define (k, c)-cycle balanced designs as the subgraphs of the unit hypercube Q_k that are (k, m)-edge balanced for some m and contain exactly c 4-cycles in direction (i, j) all $k(k-1)/2$ possible pairs. In the paper we present the system of equations that define these designs. For $c = 1, 2$ and 3 we present families of (k, c)-cycle balanced designs, that are recursively defined in terms of their polynomial representations. The figure below shows a $(5, 1)$-cycle balanced solution, together with the recursive

equations that define the family for $c = 1$. In this Figure, edges are coloured according to the elementary effect that is computed from their end points, there are thus 5 distinct colors. The graph represents a $(5, 1)$-cycle balanced design because there is exactly one cycle of dimension 4 involving all 10 possible pairs of colors.



$$G_1^1 = 1 + X_1, \qquad G_2^1 = G_1^1 + X_2 G_1^1,$$
$$G_3^1 = G_2^1 + X_3\,(1 + X_1 + X_2)\,, \qquad G_4^1 = G_3^1 + X_4\,(1 + X_1 + X_2 + X_3)$$
$$G_5^1 = G_3^1 + X_5\,(1 + X_1 + X_2 + X_3 + X_4) \qquad \cdots$$

Our work is related to a previously proposed "New Morris method", see [6,7]. However, these references only consider the case $c = 1$, i.e., computation of a single mixed effect for each pair of input factors and do not impose that the resulting designs are edge balanced.

We show how our approach can formally be extended to sampling of higher order derivatives by imposing the presence of certain patterns in the designs graphs.

Explicit formulas for the size of our edge and cycle balanced designs are provided, and we compare their economy (defined as in [1], as the ratio of the number of effects computed over the size of the design) to the economy of the original and New Morris methods. The performance of cycle balanced designs is demonstrated by considering both the analytical function on Morris' original paper [1] and a true simulation model of a complex system with a large number of inputs.

Finally, we discuss the problem of finding designs able to detect n-th order terms in each input factor, which requires definition of a suitable isometry between $\{0, \ldots, n\}^k$ and polynomials.

References:

[1] Max D. Morris, "Factorial Sampling Plans for Preliminary Computational Experiments", *Technometrics.*, Vol 32 (2), pp. 161–174, 1991.

[2] F. Campolongo, A. Saltelli, "Design of experiments", in: A. Saltelli, K. Chan, M. Scott (Eds.), *Sensitivity Analysis*, *Probability and Statistics Series*, John Wiley, pp. 5163, 2000.

[3] F Campolongo, J Cariboni, A. Saltelli, "An effective screening design for sensitivity analysis of large models", *Environmental Modelling & Software*, Elsevier, Vol. 22 (10), pp. 1509–1518, 2007.

[4] A. Saltelli, P. Annoni, I. Azzini, F. Campolongo, M. Ratto, S. Tarantola, "Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index", *Comput. Phys. Commun.*, Vol 181, pp. 259–270, 2010.

[5] F. Campolongo, A. Saltelli, J. Cariboni, "From screening to quantitative sensitivity analysis. A unified approach", *Comput. Phys. Commun.*, Vol. 182, pp. 978–988, 2011.

[6] F. Campolongo, R.D. Braddock, "The use of graph theory in the sensitivity analysis of the model output: a second order screening method," *Reliability Engineering and System Safety*, Vol. 64 (1), pp. 1–12, 1999.

[7] R.A. Cropp, R.D. Braddock, "The New Morris Method: an efficient second-order screening method," *Reliability Engineering & System Safety*, Elsevier, Vol 78 (1), pp. 77–83, 2002.

[Jean-Marc Fédou; Laboratoire I3S, CNRS-UNS, 2000 rte des Lucioles,B.P. 121, 06903 Sophia Antipolis cedex, France]
[`fedou@unice.fr` – `www.i3s.unice.fr/~fedou`]