Sequence construction for integral estimation

S. Varet[1], *S. Lefebvre*[1], *G. Durand*[1], *A. Roblin*[1] *and S. Cohen*[2]

We consider a deterministic numerical code which computes the infrared signature of an aircraft in its surroundings. This code takes as input parameters a large number of variables describing the aircraft and its environement characteristics. However, the code gives a result for a known physical configuration. Its use is then limited. Indeed, the code doesn't allow to know the envelope describing the set of possible values of the signature, this envelope resulting from a partial knowledge of the aircraft and its environement characteristics. More precisely, for a given attack scenario, some of the variables are only known through their probability density function, like meteorological variables, and some other variables through their variation interval. This envelope is essential to estimate the detection performance of infrared sensors. The search of this envelope leads to the estimation of the integral of a function h, with unknown analytic form, defined on a large dimensional space. Typically, the dimension can be of order 30 up to 60.

The first idea could be the use of the standard Monte Carlo method (MC) which consists in estimating the integral with a mean of h function values at points taken randomly and uniformly. It exists also variants of this method which allow to draw the points with a probability density chosen a priori. However, in the case of our numerical code, for cost reasons, the size of the h values sample is limited. Thus, the quality of the integral estimator is not always satisfactory. When the dimension is small, about 10, the precision can be better with the Quasi-Monte Carlo method (QMC). In this case, the points are selected in a determinist way. However, when the dimension is greater than 10, according to the Koksma-Hlawka inequality [Hla61], the QMC estimator quality decreases compared to the MC estimator one. That's why we want to reduce the dimension before applying the QMC method.

To reduce the dimension, we could use design of experiments [DFS02] or functional analysis of variance [ES81]. However, these two approaches require assumptions which are not fulfilled in the case of infrared signature. We can cite as an example the variables independence assumption which is not satisfied for two significant input data : the temperature and the relative humidity. That's why we have proposed a new method where the validity assumptions are less restrictive and which allows to evaluate the variables significance. This method is based on the computation of indexes which represent the proportion of variance of h explained by each variable. The number of variables obtained after the analysis of their significance is similar to the notion of effective dimension

[1]ONERA, DOTA/MPSO, Chemin de la Hunière, 91761 Palaiseau cedex, France.
[2]Université Paul Sabatier, LSP, 31062 Toulouse cedex 9, France.

[MC94]. Thus, we have adapted this notion to the above method.

After dimension reduction, we can apply the QMC method with the no significant variables fixed to a constant value. However, the QMC method presents another defect. Indeed, the points sequences used present irregularities on some low order projections. Such irregularities for important variables could lead to a poor convergence rate. As far as we know, the existing criteria to estimate the quality of the projection repartition are not computable in practice and their approximation requires a lot of function h evaluations. That's why we have introduced a new criterion using the indexes previously defined. Thus, we can evaluate the adequacy of the sequences to our function of interest h. More precisely, we evaluate the adequacy in terms of distribution quality of the projections. In addition, some numerical results on test functions have shown a correlation between the criterion value and the corresponding QMC estimator quality.

However, this criterion only allows to choose the best sequence among a set of sequences. Thus the chosen sequence is not necessarily optimal for our criterion. That's why we have searched to build an optimal sequence of points for our criterion in order to maximize the corresponding QMC estimator quality.

We will briefly present our method to compute the proportion of variance explained by the variables as well as our criterion. Then We will present our method of sequence construction. Finally, We will illustrate these results with numerical simulations.

Références

[DFS02] J.-J. Droesbeke, J. Fine et G. Saporta – *Plans d'expériences. applications à l'entreprise*, Technip, 2002.

[ES81] B. Efron et C. Stein – « The jackknife estimate of variance », *The Annals of Statistics* **9** (1981), no. 3, p. 586–596.

[Hla61] E. Hlawka – « Funktionen von beschränkter variation in der theorie der gleichverteilung », *Annali di Matematica Pura ed Applicata* **54** (1961), p. 325–333.

[MC94] W. J. Morokoff et R. E. Caflisch – « Quasi-random sequences and their discrepancies. », *SIAM Journal on Scientific Computing* **15** (1994), no. 6, p. 1251–1279 (English).