Comparison of Latin Hypercube and

Quasi Monte Carlo Sampling Techniques

Sergei Kucherenko[a] *, Daniel Albrecht*[b]*, Andrea Saltelli*[c]

[a]*Imperial College London, London, SW7 2AZ, UK, s.kucherenko@ic.ac.uk*

[b]*The European Commission, Joint Research Centre, TP 361, 21027 ISPRA(VA), ITALY*

[c]*European Centre for Governance in Complexity, Universitat Autonoma de Barcelona, Catalonia, Spain*

Monte Carlo (MC) simulation employing Latin Hypercube Sampling (LHS) is one of the most popular modelling tools. While its application in areas like experimental design is well justified the efficiency of LHS in other areas such as high dimensional integration can be no better than the standard MC method based on random numbers. To provide a high efficiency of high dimensional integration high uniformity of sampling is required. LHS - being well stratified in one dimension by design, does not provide good uniformity properties in high dimensions. It is known that for high dimensional integrals the convergence rate of the MC estimates based on random sampling is *O*(1/sqrt(*N*)), where *N* is the number of sampled points. A higher rate of convergence can be obtained by using Quasi Monte Carlo (QMC) methods based on low-discrepancy sequences. Asymptotically, QMC can provide the rate of convergence *O*(1/*N*). We compare efficiencies of three sampling methods: the MC method with both random and LHS sampling, and the QMC method with sampling based on Sobol' sequences. We apply the high-dimensional Sobol' sequence generator with advanced uniformity properties (technically these are known as property **A** for all dimensions and property **A'** for adjacent dimensions). Firstly we compare L_2 discrepancies and show that the QMC method has the lowest discrepancy up to dimension 20. Secondly, we use a number of test functions of various complexities for high dimensional integration. Using global sensitivity analysis functions are classified with respect to their dependence on the input variables: functions with not equally important variables (type A), functions with equally important variables and dominant low order terms (type B) and functions with equally important variables and with dominant interaction terms (type C). Comparison shows that for types A and B functions convergence of the QMC method is close to *O*(1/*N*), while the MC method has a convergence close to *O*(1/sqrt(*N*)). For types C functions convergence of the QMC method significantly drops; however it still remains the most efficient method among three sampling techniques. The ANOVA decomposition in a general case can be presented as $f(x) = f_0 + \sum_i f_i(x_i) + r(x)$, where $r(x)$ are the ANOVA terms corresponding to higher order interactions. Variance computed with the LHS design is $E(\varepsilon_{LHS}^2) = \frac{1}{N}\int_{H^n}[r(x)]^2 dx + O(\frac{1}{N})$, and $E(\varepsilon_{MC}^2) = \frac{1}{N}\sum_i \int_{H^n}[f_i(x_i)]^2 dx + \frac{1}{N}\int_{H^n}[r(x)]^2 dx + O(\frac{1}{N})$ for MC. Here ε_{LHS}^2 and ε_{MC}^2 are the convergence rates.

In the ANOVA decomposition of type B functions, the effective dimension d_S is small, hence $r(x)$ is also small comparing to the main effects. In the extreme case d_S is equal to 1, and a function *f*(x) can be presented as a sum of one-dimensional functions $f(x) = f_0 + \sum_i f_i(x_i)$. This means that only one-dimensional projections of the sampled points play a role in the function approximation. For type B functions LHS can achieve a much higher convergence rate than that of the standard MC. The results are summarized in Table 1

Table 1 Classification of functions based on the effective dimensions. Two complementary subsets of variables y and z are considered: $x = (y, z)$.

Type	Description	Relationship between S_i and S_i^{tot}	d_T	d_S	QMC is more efficient than MC	LHS is more efficient than MC
A	A few dominant variables	$S_y^{tot}/n_y >> S_z^{tot}/n_z$	$<< n$	$<< n$	Yes	No
B	No unimportant subsets; only low-order interaction terms are present	$S_i \approx S_j, \forall\, i, j$ $S_i / S_i^{tot} \approx 1, \forall\, i$	$\approx n$	$<< n$	Yes	Yes
C	No unimportant subsets; high-order interaction terms are present	$S_i \approx S_j, \forall\, i, j$ $S_i / S_i^{tot} << 1, \forall\, i$	$\approx n$	$\approx n$	No	No

To test the classification presented above MC, LHS and QMC integration methods were compared considering a set of test functions. Computed root mean square error (RMSE) was approximated by the formula $cN^{-\alpha}$, $0 < \alpha < 1$. For a function presented in Fig. 1 at *n*=360 the exponent for algebraic decay in the case of QMC integration α_{QMC} = 0.94. The LHS method shows the same convergence rate $\alpha \approx 0.5$ as the MC method.

In summary QMC appears preferable to LHS overall, consideration given to the different typologies of functions. The objection that LHS can be made better by optimization (searching an optimum LHS by brute force methods) suffers from both computational hurdle and poor elegance.

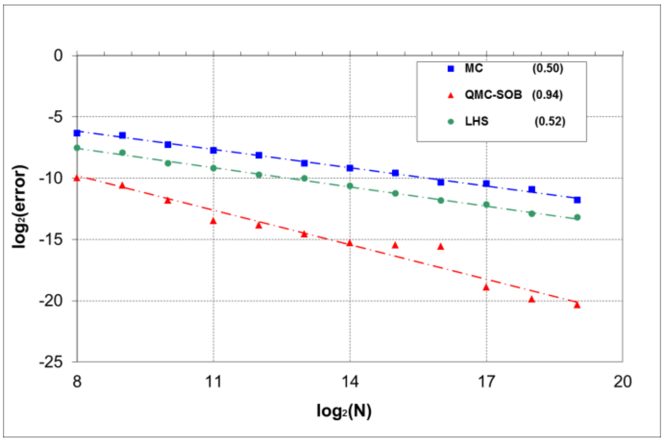

Fig. 1. RMSE versus the number of sampled points for type A model $\sum_{i=1}^{n}(-1)^i \prod_{j=1}^{i} x_j$. n = 360 .