Application of Voronoi Weights in Monte Carlo Integration with a Given Sampling Plan

M. Vořechovský, V. Sadílek, and J. Eliáš
Faculty of Civil Engineering, Brno University of Technology, Brno, Czech republic

Abstract: The standard way to numerically calculate integrals such as the ones featured in estimation of statistical moments of functions of random variables using Monte Carlo procedure is to: (i) perform some sampling from the random vector, (ii) perform an approximation to the integrals using averages of the functions evaluated at the selected sampling points. If the N_{sim} points are selected with an equal probability (with respect to the joint distribution function) such as in Monte Carlo sampling, the averages use equal weights $1/N_{\text{sim}}$. The problem with Monte Carlo sampling is that the estimated values exhibit a large variance due to the fact that the sampling points are usually not spread uniformly over the domain of random variables. One way to improve the accuracy would be to perform a more advanced sampling.

The paper explores another way to improve the Monte Carlo integration approach: by considering unequal weights. These weights are obtained by transforming the sampling points into sampling probabilities (points within a unit hypercube), and subsequently by associating the sampling points with weights obtained as volumes of regions/cells around the sampling points within a unit hypercube. These cells are constructed by the Voronoi tessellation around each point. Supposedly, this approach could have been considered superior over the naive one because it can suppress inaccuracies stemming from clusters of sampling points.

Keywords: Monte Carlo sampling, estimation of statistical moments, Voronoi tesselation, weighted average, probability weights

1. Introduction

Monte Carlo estimation of statistical integrals is encountered in numerous applications. A typical example is the computer exploration of functions that feature random variables. These random variables form an N_{var}-dimensional vector, where N_{var} is the number of random variables considered. In computer experiments the first step is a selection of optimal sample set, i.e. selection of N_{sim} points from the N_{var} dimensional space. These points then form the sampling plan which is a $N_{\text{sim}} \times N_{\text{var}}$ matrix. The methods used for formulating the plan of experimental points are collectively known as Design of Experiments (DoE). The purpose of DoE is to provide a set of points lying inside a chosen *design domain* that are optimally distributed; the optimality of the sample depends on the nature of the problem. Various authors have suggested intuitive goals for good designs, including "good coverage", the ability to fit complex models, many levels for each factor/variable, and good projection properties. At the same time, a number of different mathematical criteria have been put forth for comparing designs.

The design of experiments is typically performed in a hyper-cubical domain of N_{var} dimensions, where each dimension/variable, U_v, ranges between zero and one ($v = 1, \ldots, N_{\text{var}}$). This *design domain* is to be covered by N_{sim} points as evenly as possible as the points within the design domain represent sampling probabilities. The probability that the i-th experimental point will be located inside some chosen subset of the domain must be equal to V_S/V_D, with V_S being the subset volume and V_D the volume of the whole domain (for unconstrained design $V_D = 1$). Whenever this is valid, the design criterion will be called *uniform*. Even though such uniformity is conceptually simple and intuitive on a qualitative level, it is somewhat complicated to describe and characterize it mathematically. Though some problems do not require this uniformity, it is the crucial assumption in Monte-Carlo integration and its violation may lead to significant errors (Eliáš and Vořechovský, 2016; Vořechovský and Eliáš, 2015).

There exist many other criteria of optimality of the sampling plan: e.g. the Audze-Eglājs (AE) criterion (Audze and Eglājs, 1977) later generalized into the so-called ϕ criterion, the Euclidean MaxiMin and MiniMax distance between points, various measures of discrepancy, criteria based on correlation (orthogonality), designs maximizing entropy and many others. It should also be noted that an experimental design can be also obtained via so-called "quasi-random" low-discrepancy sequences (deterministic versions of MC analysis) that can often achieve reasonably uniform sample placement in hypercubes (Niederreiter, Halton, Sobol', Hammersley, etc.).

As mentioned above, the selection of the sampling points is a crucial step when evaluating approximations to integrals as is performed in Monte Carlo simulations (numerical integration), where equal sampling probabilities inside the design domain are required.

In this article, it is assumed that the sampling points have already been selected and they are not spread uniformly over the design domain. A typical example may be a sample selected using crude Monte Carlo sampling. The article considers the possibility to improve quality of Monte Carlo estimation with such a given sample. The only possibility to improve the estimations of the integrals is to vary the weights associated with individual sampling points. Motivated by the MiniMax criterion of optimality and also by various numerical integrating schemes, we explore the possibility to improve the quality of statistical estimations using Voronoi tessellation, i.e. a particular form of partitioning of the design domain around given sampling points. The *design domain* to be partitioned is the unit hypercube described above and therefore the volumes around individual sampling points represent weights (probabilities) to be used in the weighted averages that estimate the integrals.

2. Statistical Moment Estimation Using Monte Carlo Sampling

As mentioned in the introduction, one of the frequent uses of DoE is *statistical sampling* for Monte Carlo integration. We present the application of statistical sampling to the problem of estimating statistical moments of a function of random variables. In particular, a deterministic function, $Z = g(\boldsymbol{X})$, is considered, which can be a computational model or a physical experiment. Z is the uncertain response variable (or generally a vector of the outputs). The vector $\boldsymbol{X} \in \mathbb{R}^{N_{\text{var}}}$ is considered to be a random vector of N_{var} continuous marginals (input random variables describing uncertainties/randomness) with a given joint probability density function (PDF).

Estimation of the statistical moments of variable $Z = g(\boldsymbol{X})$ is, in fact, an estimation of integrals over domains of random variables weighted by a given joint PDF of the input random vector, $f_{\boldsymbol{X}}(\boldsymbol{x})$. We seek the statistical parameters of $Z = g(\boldsymbol{X})$ in the form of the following integral:

$$\mathrm{E}[S[g(\boldsymbol{X})]] = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} S[g(\boldsymbol{x})] \, \mathrm{d}F_{\boldsymbol{X}}(\boldsymbol{x}) \tag{1}$$

where $\mathrm{d}F_{\boldsymbol{X}}(\boldsymbol{x}) = f_{\boldsymbol{X}}(\boldsymbol{x}) \cdot \mathrm{d}x_1 \, \mathrm{d}x_2 \, \cdots \, \mathrm{d}x_{N_{\mathrm{var}}}$ is the infinitesimal probability ($F_{\boldsymbol{X}}$ denotes the joint cumulative density function) and where the particular form of the function $S[g(\cdot)]$ depends on the statistical parameter of interest. For example, to gain the mean value of $g(\cdot)$, $S[g(\cdot)] = g(\cdot)$; higher statistical moments of Z can be obtained by integrating polynomials of $g(\cdot)$. The probability of failure (an event defined as $g(\cdot) < 0$) is obtained in a similar manner: $S[\cdot]$ is replaced by the Heaviside function (or indicator function) $H[-g(\boldsymbol{X})]$, which equals one for a failure event ($g < 0$) and zero otherwise. In this way, the domain of integration of the PDF is limited to the failure domain.

In Monte Carlo sampling, which is the most prevalent statistical sampling technique, the above integrals are numerically estimated using the following procedure: (i) draw N_{sim} realizations of $\boldsymbol{X}$ that share the same probability of occurrence $1/N_{\mathrm{sim}}$ by using its joint distribution $f_{\boldsymbol{X}}(\boldsymbol{x})$; (ii) compute the same number of output realizations of $S[g(\cdot)]$; and (iii) estimate the desired parameters as arithmetical averages. We now limit ourselves to independent random variables in vector $\boldsymbol{X}$. The aspect of the correct representation of the target joint PDF of the inputs mentioned in item (i) is absolutely crucial. Practically, this can be achieved by reproducing a *uniform distribution* in the design space (unit hypercube) that represents the space of sampling probabilities.

Assume now a random vector $\boldsymbol{U}$ that is selected from a multivariate uniform distribution in such a way that its independent marginal variables U_v, $v = 1, \dots, N_{\mathrm{var}}$, are uniform over intervals $(0; 1)$. A vector with such a multivariate distribution is said to have an "independence copula" (Nelsen, 2006)

$$C(u_1, \dots, u_{N_{\mathrm{var}}}) = \mathrm{P}(U_1 \leq u_1, \dots, U_{N_{\mathrm{var}}} \leq u_{N_{\mathrm{var}}}) = \prod_{v=1}^{N_{\mathrm{var}}} u_v \tag{2}$$

These uniform variables can be seen as sampling probabilities: $F_{X_v} = U_v$. The joint cumulative distribution function then reads $F_{\boldsymbol{X}}(\boldsymbol{x}) = \prod_v F_{X_v} = \prod_v U_v$, and $\mathrm{d}F_{\boldsymbol{X}}(\boldsymbol{x}) = \prod_v \mathrm{d}U_v$. The individual random variables can be obtained by inverse transformations

$$\{X_1, \dots, X_{N_{\mathrm{var}}}\} = \{F_1^{-1}(U_1), \dots, F_{N_{\mathrm{var}}}^{-1}(U_{N_{\mathrm{var}}})\} \tag{3}$$

and similarly the realizations of the original random variables are obtained by the component-wise inverse distribution function of a point $\boldsymbol{u}$ (a realization of $\boldsymbol{U}$) representing a sampling probability

$$\boldsymbol{x} = \{x_1, \dots, x_{N_{\mathrm{var}}}\} = \{F_1^{-1}(u_1), \dots, F_{N_{\mathrm{var}}}^{-1}(u_{N_{\mathrm{var}}})\} \tag{4}$$

With the help of this transformation from the original to the uniform joint PDF, the above integral in Eq. (1) can be rewritten as

$$\begin{aligned} \mathrm{E}[S\left[g\left(\boldsymbol{X}\right)\right]] &= \int_0^1 \dots \int_0^1 S\left[g\left(\boldsymbol{x}\right)\right] \, \mathrm{d}C(u_1, \dots, u_{N_{\mathrm{var}}}) \\ &= \int_{[0,1]^{N_{\mathrm{var}}}} S\left[g\left(\boldsymbol{x}\right)\right] \prod_{v=1}^{N_{\mathrm{var}}} \mathrm{d}U_v \end{aligned} \tag{5}$$

so that the integration is performed over a unit hypercube with uniform unit density.

We now assume an estimate of this integral by the following statistic (the average computed using N_{sim} realizations of $\boldsymbol{U}$, namely the sampling points $\boldsymbol{u}_j$ $(j = 1, \dots, N_{\mathrm{sim}})$)

$$\mathrm{E}[S\left[g\left(\boldsymbol{X}\right)\right]] \approx \frac{1}{N_{\mathrm{sim}}} \sum_{i=1}^{N_{\mathrm{sim}}} S\left[g(\boldsymbol{x}_i)\right] \tag{6}$$

where the sampling points $\boldsymbol{x}_i = \{x_{i,1}, \dots, x_{i,v}, \dots, x_{i,N_{\mathrm{var}}}\}$ are selected using the transformation in Eq. (4), i.e. $x_{i,v} = F_v^{-1}(u_{i,v})$, in which we assume that each of the N_{sim} sampling points $\boldsymbol{u}_i$ $(i = 1, \dots, N_{\mathrm{sim}})$ were selected with the same probability of $1/N_{\mathrm{sim}}$. Violation of the uniformity of the distribution of points $\boldsymbol{u}_i$ in the unit hypercube may lead to erroneous estimations of the integrals.

If the sampling points are not selected carefully with respect to equal probabilities in the design domain, the possibility to improve the accuracy in Eq. (6) is to use weights different from $1/N_{\mathrm{sim}}$. These weights reflect the probability content of the cells around individual sampling points

$$\mathrm{E}[S\left[g\left(\boldsymbol{X}\right)\right]] \approx \sum_{i=1}^{N_{\mathrm{sim}}} S\left[g(\boldsymbol{x}_i)\right] \cdot w_i \tag{7}$$

The proposed approach aims at finding appropriate weights that are calculated considering the spatial distribution of the points. Obviously, unvisited regions of the design domain can not be explored by a nonuniform design. Partitioning the space into cells around the given sampling points may help to reduce probabilities associated with points that are participating in clusters of points. Voronoi tessellation has been selected for partitioning of the design space into volumes that are used as the weights w_i, $i = 1, \dots, N_{\mathrm{sim}}$. The following section describes the Voronoi tessellation procedures.

3. Weights Obtained as Volumes of Voronoi Regions

The weights associated with the design points are considered as volumes of Voronoi regions (Aurenhammer, 1991) computed on the sampling points. The Voronoi tessellation in N_{var}-dimensional space results in N_{sim} convex polyhedrons $\mathcal{V}_i$ that encloses all the points that are closer to i-th

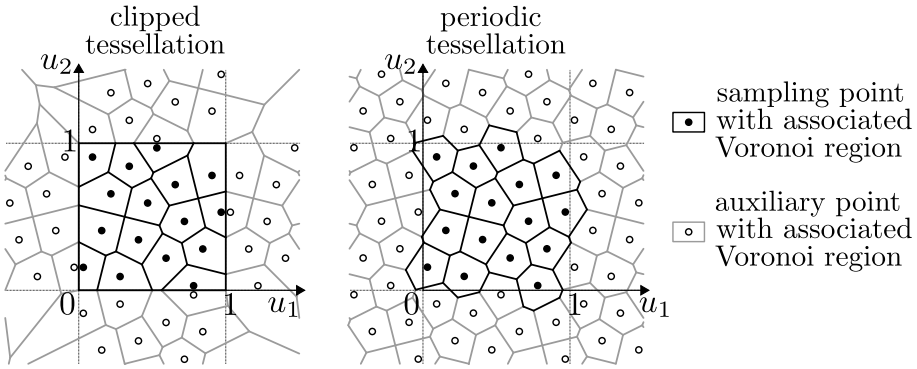

Figure 1. Example of clipped and periodic tessellation for $N_{\text{var}} = 2$ and $N_{\text{sim}} = 16$ with help of reflected and periodically repeated auxiliary points, respectively.

sampling point than any other. Defining the distance of point $\boldsymbol{u}$ from sampling point $\boldsymbol{u}_i$ as $d_i(\boldsymbol{u})$, the Voronoi region associated with i-th sampling point can be formally defined as

$$\mathcal{V}_i = \left\{ \boldsymbol{u} \in \mathbb{R}^{N_{\text{var}}} \mid \forall\, j \neq i : d_i(\boldsymbol{u}) \leq d_j(\boldsymbol{u}) \right\} \tag{8}$$

Two alternatives of Voronoi tessellation that differ in the boundary regions are investigated:

– *clipped* Voronoi tessellation that is limited to the unit hypercube only

$$\mathcal{V}_i = \left\{ \boldsymbol{u} \in \langle 0,\, 1 \rangle^{N_{\text{var}}} \mid \forall\, j \neq i : d_i(\boldsymbol{u}) \leq d_j(\boldsymbol{u}) \right\} \tag{9}$$

– *periodic* Voronoi tessellation which assumes that every sampling point is periodically repeated in the space along all the dimensions.

These two different concepts are demonstrated in Figure 1. The reason for studying the *periodic* tessellation is that the authors have shown recently (Eliáš and Vořechovský, 2016; Vořechovský and Eliáš, 2015) that the presence of boundaries in the hypercubical design domain cause problems. Briefly, one may think of a problem of packing (hyper)balls into a (hyper)cube. It is clear that the boundary is responsible for a kind of wall-effect. It has been shown (Eliáš and Vořechovský, 2016; Vořechovský and Eliáš, 2015) that this problem can be removed by considering periodic extension of the design domain. The balls then permeate through the boundaries without interacting with them, see Figure 1 right.

The *clipped* Voronoi diagrams (Chan et al., 1995; Yan et al., 2013) are used mostly for construction of meshes and therefore available software to compute such tessellation is limited to two and three dimensional space. A similar situation exists for *periodic* Voronoi tessellation (Yan et al., 2011; Rycroft, 2009). In the field of design of experiments more than three variables (factors) can be present and therefore the tessellation must be performed in higher dimensions. In this contribution, Qhull software (Barber et al., 1996) is utilized for both clipped and periodic tessellations because it can compute Voronoi tessellation for arbitrary dimension. On the other hand, it cannot work directly with neither *clipped* nor *periodic* boundary condition and therefore simple tricks are used.

These tricks consist in manipulations of the design domain (together with the sampling points contained) by adding new design domains around it. In order to obtain the *clipped* structure, the design domain is extended by reflecting the original design domain along each dimension. There are two reflections of the original unit interval along each dimension to obtain intervals $\langle -1,\, 0\rangle$ and $\langle 1,\, 2\rangle$. Therefore, the tessellation is performed on $N_{\text{sim}}\left(1+2^{N_{\text{var}}}\right)$ points. The use of reflection automatically provides edges between cells that coincide with the boundary of the original design domain and therefore the volumes outside the design domain can be ignored. The use of reflection to obtain clipped tessellation was proposed in (Pronzato and Müller, 2012).

The *periodic* structure is obtained by periodic extension (replication) of the original design domain along each direction and additionally the replication must be performed to obtain all the "corner" domains to fill a hypercube $\langle -1,\, 0\rangle^{N_{\text{var}}}$. Therefore, $N_{\text{sim}}\cdot 3^{N_{\text{var}}}$ points in total are used for the periodic tessellation.

The computational times needed for the both tessellation types can be substantially reduced if it involves only reflected or periodically repeated points that are close to the original hypercube, because only these points affects the tessellation inside the hypercube. Unfortunately, no effective algorithm has been developed yet to identify such points and therefore the full set of points must be involved for certainty.

In both alternatives, the weights for individual sampling points are the volumes of regions surrounding points. There are three algorithms available for the volume computation: (i) direct integration, (ii) Monte-Carlo integration and (iii) division into simplexes for which analytical formula is available. The first two algorithms are nicely elucidated in (Ong et al., 2003). Here, we perform the third algorithm. Each Voronoi region is (with a help of the Qhull) divided into simplexes. Each simplex has $N_{\text{var}}+1$ vertices denoted $\boldsymbol{v}_i$. The total volume of the region is simply the sum of simplex volumes, that are calculated based on the determinant of coordinate matrix.

$$V_{\text{simplex}} = \left| \frac{1}{N_{\text{var}}!} \begin{pmatrix} \boldsymbol{v}_1 - \boldsymbol{v}_0 \\ \boldsymbol{v}_2 - \boldsymbol{v}_0 \\ \vdots \\ \boldsymbol{v}_{N_{\text{var}}} - \boldsymbol{v}_0 \end{pmatrix} \right| \tag{10}$$

These volumes are used directly as weights of sampling points enclosed withing these cells.

4. Frequency Analysis of Weights

It turns out to be important to see (i) whether the weights are very scattered compared to $1/N_{\text{sim}}$ and, (ii) whether their magnitude tend to depend on the position inside the domain. This is achieved by studying $N_{\text{run}} = 1000$ realizations of samples, each having N_{sim} points within an N_{var}-dimensional hypercube. For each sample, both types of Voronoi tessellation is constructed and the weights are statistically processed.

The results will be presented for two sampling schemes: the classical (crude) Monte Carlo sampling without any optimization (MC-RAND) and LHS (Latin Hypercube Sampling) optimized using the periodic criterion (LHS-PAE). PAE stands for an enhanced version of the Audze-Eglais criterion, see (Eliáš and Vořechovský, 2016; Vořechovský and Eliáš, 2015).

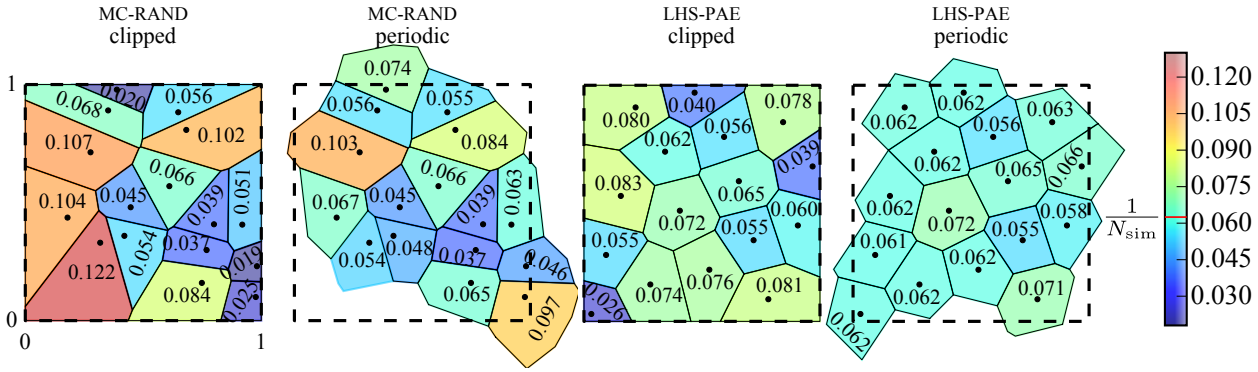

Figure 2. Voronoi weights for MC-RAND and LHS-PAE sampling plans ($N_{\text{var}} = 2$, $N_{\text{sim}} = 16$). Comparison of the *clipped* and *periodic* tessellations.

Figure 2 shows one sample (N_{sim}=16) of a bivariate random vector U_v for both sampling schemes. For the two sampling schemes, both types of Voronoi diagrams (*clipped* and *periodic*) are constructed and visualized with colors depending on the area. The LHS-PAE sampling plans show more uniform distribution of points because the PAE-optimized LH-sampling better avoids clustering and limit the occurrence of empty regions. Therefore, the cells in LHS-PAE have similar volumes and the sampling points are closer to the centers of Voronoi regions. The small differences among weights in LHS-PAE with *periodic* tessellation suggest that weighting will not make much difference in comparison with integrals evaluated using equal weights $1/N_{\text{sim}}$. The MC-RAND sampling plans suffer from point clustering and therefore, high variability in volumes of the Voronoi cells is observed. It should be noticed that the choice of tessellation (*clipped* vs. *periodic*) affects only the boundary regions while the central part of the hypercube is identical.

In order to judge about the spatial distribution of weights within the design domain, the above-mentioned N_{run}=1000 realizations of samples accompanied by Voronoi tessellations were prepared and for each spatial location, the mean value and standard deviation of weights occurring at that location have been calculated. The weights (volumes of Voronoi regions V_{simplex} in a hypercube) depend on the type of tessellation but they are independent of the sampling method (MC vs. LHS). The bivariate histograms in Figure 3 document the dependency of the mean value and the standard deviation of Voronoi region volumes on the position of the sampling point in a square. In the case of *clipped* tessellation, both the mean value and the standard deviation of weights are not uniform in the hypercube. Three zones can be distinguished: (a) the boundary region where the mean value (shown in blue) of weights is underestimated. The boundary strip is followed/balanced by (b) zone parallel to the boundary where the weights are overestimated (see the yellow to red color) and finally, (c) the bulk zone sufficiently far from the boundary, where the weights (volumes) are constant on average. The width of the two boundary zones is decreasing with increasing sample size N_{sim}.

Such a biased representation of different regions in the hypercube partitioned by the *clipped* tessellation must have consequences in Monte Carlo integration. If the points are sampled uniformly, and that is indeed the case of both MC-RAND and LHS-PAE, some error must be introduced due to introduction of nonuniform weighting. If the functions are sensitive to inaccuracies in representation

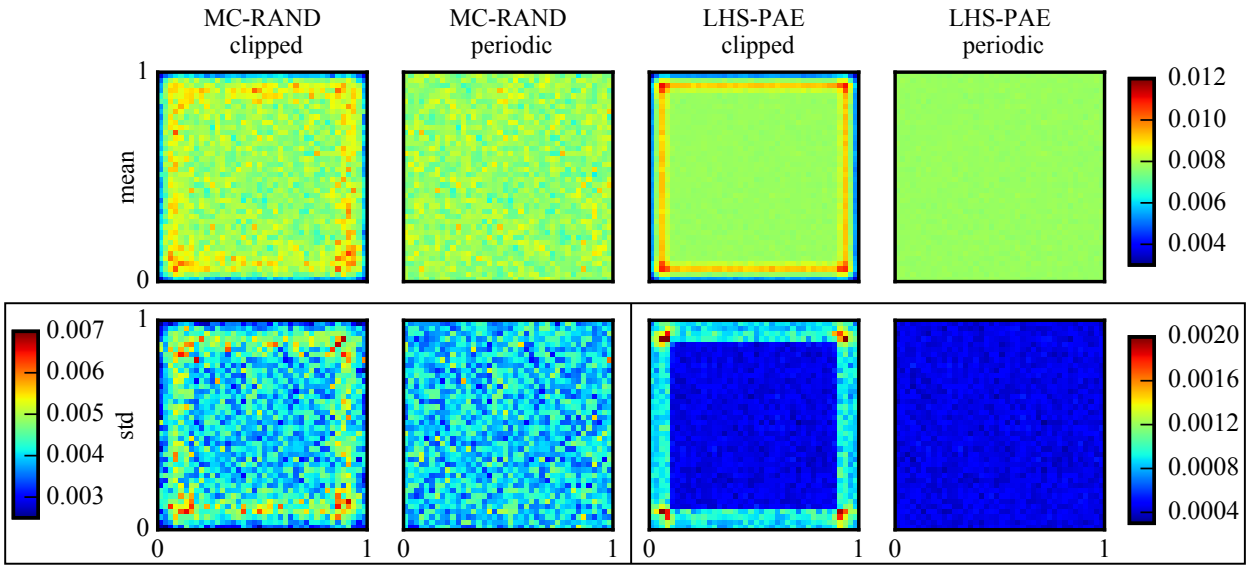

Figure 3. Bivariate histograms of the mean value and the standard deviation of cell volumes for both sampling plans and both tessellation alternatives ($N_{\text{var}} = 2$, $N_{\text{sim}} = 16$ and $N_{\text{run}} = 1000$ realizations).

of the boundary regions, their weighted MC integration may yield biased results. Therefore we conclude that the *clipped* tessellation generally should not be used for weighting in MC integration.

The *periodic* tessellation provides more promising bivariate histograms: no bias around the boundaries is visible for both MC-RAND and LHS-PAE sampling schemes. The statistics of the weights do not depend systematically on the position in the hypercube.

5. Numerical Examples of MC Integration & Discussion

This section studies whether weighting in MC integrals based on the Voronoi tessellation improves the quality of the estimates. Three basic transformations g of standard independent Gaussian random variables X_v, $v = 1, \ldots, N_{\text{var}}$ have been selected for the numerical study. The following array presents formulas of the three functions (first column), the analytical solutions for the mean values (second column) and standard deviations (third column):

$$Z_{\text{sum}} = g_{\text{sum}}(\boldsymbol{X}) = \sum_{v=1}^{N_{\text{var}}} X_v^2 \qquad \mu_{\text{sum}} = N_{\text{var}} \qquad \sigma_{\text{sum}} = \sqrt{2N_{\text{var}}} \tag{11}$$

$$Z_{\text{exp}} = g_{\text{exp}}(\boldsymbol{X}) = \sum_{v=1}^{N_{\text{var}}} \exp(-X_v^2) \qquad \mu_{\text{exp}} = \frac{\sqrt{3}}{3} N_{\text{var}} \qquad \sigma_{\text{exp}} = \sqrt{N_{\text{var}}} \sqrt{\frac{\sqrt{5}}{5} - \frac{1}{3}} \tag{12}$$

$$Z_{\text{prod}} = g_{\text{prod}}(\boldsymbol{X}) = \prod_{v=1}^{N_{\text{var}}} X_v \qquad \mu_{\text{prod}} = 0 \qquad \sigma_{\text{prod}} = 1 \tag{13}$$

The two sampling schemes studied above (MC-RAND and LHS-PAE) have been used to prepare $N_{\text{run}} = 1000$ sampling plans for various sample sizes N_{sim} and dimensions N_{var}. Each sample set is accompanied by both types of tessellations – *periodic* and *clipped*.

The performance of the approaches to estimate the integrals will be demonstrated by showing their ability to estimate the mean value and standard deviation of the transformed variable $Z = g(\boldsymbol{X})$. The estimated mean value and standard deviation are denoted as $\bar{\mu}_Z$ and $\bar{\sigma}_Z$, respectively:

$$\bar{\mu}_Z = \sum_{i=1}^{N_{\text{sim}}} g(\boldsymbol{x}_i) \cdot w_i \tag{14}$$

$$(\bar{\sigma}_Z)^2 = \frac{N_{\text{sim}}}{N_{\text{sim}} - 1} \sum_{i=1}^{N_{\text{sim}}} (g(\boldsymbol{x}_i) - \bar{\mu}_Z)^2 \cdot w_i \tag{15}$$

The term $N_{\text{sim}}/(N_{\text{sim}} - 1)$ is a standard adjustment that makes the sample variance unbiased. Analogical adjustment terms are known for higher statistical moments of random variables, see e.g. (Cramér, 1945). When the individual sampling points have unequal weights, similar correction terms have recently been derived (Rimoldini, 2014). The corrections use terms calculated as sums of pth powers of weights: $V_p = \sum_{i=1}^{N_{\text{sim}}} w_i^p$. For example, in the case of estimation of variance $(\bar{\sigma}_Z)^2$, the term $N_{\text{sim}}/(N_{\text{sim}} - 1)$ is replaced by $V_1^2/\left(V_1^2 - V_2\right)$. The corresponding unweighted form can be achieved by direct substitution of $w_i = 1/N_{\text{sim}}$. This results in $V_p = 1/{N_{\text{sim}}}^{p-1}$ (or simply $V_p = N_{\text{sim}}$) for all p, leading to the known formulas for sample-size unbiased moments (Cramér, 1945). We note that in the case of functions studied in the present paper, the enhancement using V_p makes almost no change compared to the results obtained with considering all $V_p = N_{\text{sim}}$.

Three approaches to the weighting in Monte-Carlo type numerical integration are compared:

- *uniform* weights that assign each design point a constant weight $w_i = 1/N_{\text{sim}}$,
- *clipped* Voronoi weights that assign weights according to the "volumes" of Voronoi regions obtained using the *clipped* tessellation,
- *periodic* Voronoi weights that use *periodic* Voronoi tessellation.

Based on the study focused on spatial distribution of weights within the design domain, one might expect that the *periodic* tessellation may or may not deliver improvement while the *clipped* tessellation must lead to biased MC integration when used to perform weighting. The results of numerical study are presented in Figure 4 for all three functions, two estimated statistical moments, two sampling schemes and three alternatives of weighting. Each alternative presents a line of the average with a scatter-band (shaded area) obtained as the mean value ± one standard deviation – both computed using $N_{\text{run}} = 1000$ realizations. A three dimensional domain has been selected: $N_{\text{var}} = 3$.

The standard procedure employing *uniform* weights leads to convergence to the exact value as expected. The estimator variance decreases with increasing N_{sim}.

The *clipped* Voronoi weights provide poor estimates as expected. Although the variance is generally lower than in case of *uniform* weights, the average converges significantly slower. The

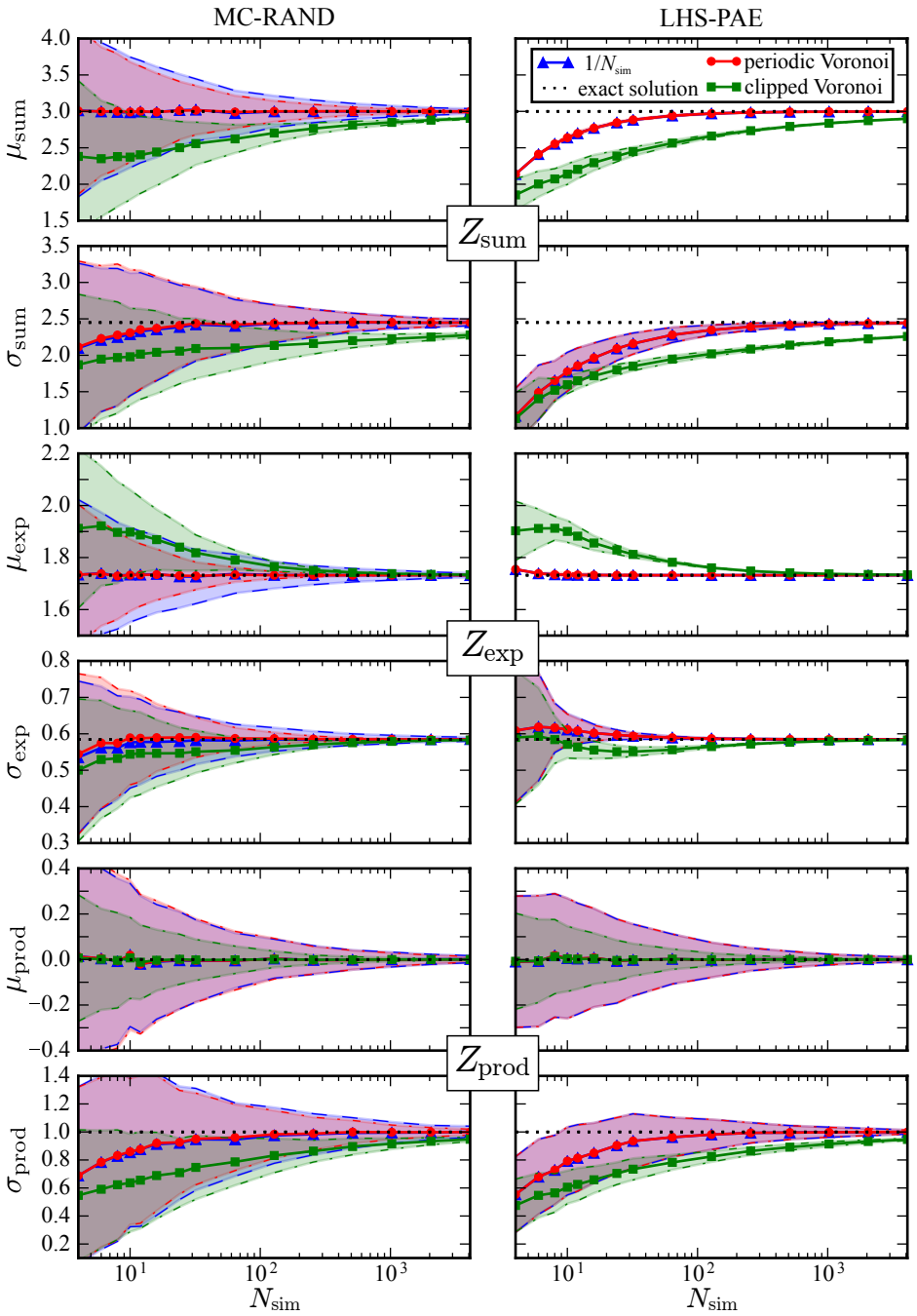

Figure 4. Convergence of the estimated mean values and standard deviations of the three transformed variables g_{sum}, g_{exp} and g_{prod}, computed for $N_{\text{var}} = 3$. Left column: MC-RAND, Right column: LHS-PAE.

reason is the improper volumes of boundary cells, the boundary zones are underestimated followed by the overestimated zone (see e.g. the bivariate histograms in Figure 3). It eventually also converges to the correct solution as the biased boundary region gets narrower with increasing N_{sim}.

Application of the *periodic* Voronoi diagram to obtain weights for calculation of statistics of the transformed random variable Z provides, on average, similar convergence to the exact value as observed with *uniform* weights. Also the variance of the estimator is similar. In fact, the variance of estimators seems to be slightly improved for MC-RAND sampling scheme, however, considering the relatively high computational cost related to evaluation of periodic Voronoi weights, the improvement is not worth the effort.

6. Conclusions

This paper studied two alternatives of Voronoi tessellation in an attempt to improve Monte-Carlo integration for small N_{sim} by weighting individual sampling points. The weights were obtained as volumes of the Voronoi cells – the regions surrounding the sampling points in the design domain (unit hypercube).

Weighting using the *clipped* Voronoi tessellation (a tessellation limited to the design domain) was found inapplicable due to problems related to the presence of boundaries of the unit hypercube. The tessellation results in systematic appearance of underestimated regions near the boundaries followed by regions with over-weighted regions.

The *periodic* tessellation slightly improves the integration if the location of sampling points is not optimized such as in the case for crude Monte Carlo sampling. However, the minor improvement does not outweigh the additional effort spend on the evaluation of the volumes of the regions and the tessellation.

Acknowledgements

The authors acknowledge financial support provided by projects nos. 16-22230S and GA14-10930 “SPADD” supported by the Czech Science Foundation.

References

Audze, P. and V. Eglājs. New approach for planning out of experiments. In *Problems of Dynamics and Strengths*, 35:104–107, Russian, 1977.

Aurenhammer, F. Voronoi Diagrams - A Survey of a Fundamental Geometric Data Structure. *ACM Computing Surveys*, 23(3):343–405, 1991.

Barber, C.B., D.P. Dobkin and H.T. Huhdanpaa. The Quickhull algorithm for convex hulls. *ACM Transactions on Mathematical Software*, 22(4):469–483, 1996. http://www.qhull.org

Chan, T.M.Y., J. Snoeyink and Ch.K. Yap. Output-sensitive construction of polytopes in four dimensions and clipped Voronoi diagrams in three. *In: Proceedings of the sixth annual ACM–SIAM symposium on discrete algorithms*, 282–291, 2013.

Cramér, H. *Mathematical Methods of Statistics*. Princeton landmarks in mathematics and physics, 1945.

Eliáš, J. and M. Vořechovský. Modification of the Audze–Eglajs criterion to achieve a uniform distribution of sampling points. *Advances in Engineering Software*, under review.

Nelsen, R.B. *An Introduction to Copulas*, volume XIV of *Springer Series in Statistics*. Springer, 2nd edition, 2006.

Ong, H.L., H.C. Huang and W.M. Huin. Finding the exact volume of a polyhedron. *Advances in Engineering Software*, 34(6):351–356, 2003. doi:10.1016/S0965-9978(03)00030-9.

Pronzato, L. and W.G. Müller. Design of computer experiments: space filling and beyond. *Statistics and Computing*, 22(3):681–701, 2012. doi:10.1007/s11222-011-9242-3.

Rimoldini, L. Weighted skewness and kurtosis unbiased by sample size and Gaussian uncertainties. *Astronomy and Computing*, 5:1–8, 2014. doi:10.1016/j.ascom.2014.02.001.

Rycroft, Ch.H. Voro++: a three-dimensional Voronoi cell library in C++. Documentation of the Voro++ software, 2009.

Vořechovský, M. and J. Eliáš. Improved formulation of AudzeEglajs criterion for space-filling designs.In: *Proc. of 12th International Conference on Applications of Statistics and Probability in Civil Engineering*, ICASP12, Vancouver, Canada, July 12–15, 2015.

Yan, D.-M., K. Wang, B. Lévy and L. Alonso. Computing 2D Periodic Centroidal Voronoi Tessellation. *In: 8th International Symposium on Voronoi Diagrams in Science and Engineering*, 2011. doi:10.1109/ISVD.2011.31.

Yan, D.-M., W. Wang, B. Lévy and Y. Liu. Efficient Computation of Clipped Voronoi Diagram for Mesh Generation. *Computer-Aided Design*, 45(4):843–852, 2013. doi:10.1016/j.cad.2011.09.004.