

Initial Evaluation of Pure and “Latinized” Centroidal Voronoi Tessellation for Non-Uniform Statistical Sampling*

Vicente J. Romero[§]
Sandia National Laboratories[†]
Albuquerque, NM

John V. Burkardt, Max D. Gunzburger[◇], Janet S. Peterson
School of Computational Science and Information Technology
Florida State University, Tallahassee, FL

Abstract

A recently developed Centroidal Voronoi Tessellation (CVT) sampling method is investigated here to assess its suitability for use in statistical sampling applications. CVT efficiently generates a highly uniform distribution of sample points over arbitrarily shaped M-Dimensional parameter spaces. On several 2-D test problems CVT has recently been found to provide exceedingly effective and efficient point distributions for response surface generation. Additionally, for statistical function integration and estimation of response statistics associated with uniformly distributed random-variable inputs (uncorrelated), CVT has been found in initial investigations to provide superior points sets when compared against Latin-Hypercube and Simple-Random Monte Carlo methods and Halton and Hammersley quasi-Monte-Carlo sequence methods. In this paper, the performance of all these sampling methods and a new variant (“Latinized” CVT) are further compared for non-uniform input distributions. Specifically, given uncorrelated normal inputs in a 2-D test problem, statistical sampling efficiencies are compared for resolving various statistics of response: mean, variance, and exceedence probabilities.

Keywords: Centroidal Voronoi tessellation, statistical sampling methods, uncertainty propagation

1. INTRODUCTION AND BACKGROUND

It is often beneficial in statistical sampling and function integration to sample "uniformly" over the applicable parameter space. Such uniformity, while conceptually simple and intuitive on a qualitative level, is on a quantitative level somewhat complicated to describe and characterize mathematically. Quantitative aspects of uniformity involve: 1) the equality with which points are spaced relative to one another in the parameter space (are they all nominally the same distance from one another?); 2) uniformity of point density over the entire domain of the parameter space (i.e., uniform "coverage" of the whole domain by the set of points, and not just good uniformity within certain regions of the space); and 3) isotropy in the point placement pattern. Each of these aspects of uniformity can be quantified by several mathematical measures as described in reference [2]. We will not discuss these measures further here, but mention them to indicate that quantitative measures do exist for the notion of uniformity.

*This paper is declared a work of the United States Government and is not subject to copyright protection in the U.S.

[§]corresponding author email contact: vjromer@sandia.gov

[†]Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the U.S. Department of Energy's National Nuclear Security Administration under Contract DE-AC04-94AL85000.

[◇]M. Gunzburger also supported in part by the Computer Science Research Institute, Sandia National Laboratories, under contract 18407.

We find that for 2-D data sets the eye is an excellent integrator of the different aspects of uniformity listed above. The intuitive sense of uniformity obtained from viewing sample sets in a unit square (2-D hypercube) usually correlates very strongly with the quantitative measures. Thus, for 2-D data sets like the ones we present later, fairly accurate visual judgements can be made about whether one particular layout of sample points is mathematically more uniform than another, or whether the uniformity varies significantly over the parameter space.

Achieving high sampling uniformity over generic domains is an area of active research. Much effort has been applied to the problem of achieving uniform placement of N samples over M -dimensional hypercubes, where M and N are both arbitrary. It is well recognized that Simple-Random sampling (SRS) Monte Carlo does not do a particularly good job of uniformly spreading out the sample points. The popular Latin Hypercube Sampling (LHS) method ([5]) generally does a much better job of uniformly spreading out the points. This is due to the greater sampling regularity over each individual parameter dimension before the individually generated parameter values are randomly combined into parameter sets which define the coordinates of the sampling points.

Recent efforts to modify LHS to get an even more uniform distribution of points over the parameter space have included Distributed Hypercube Sampling (DHS, [12]) and Improved [Distributed] Hypercube Sampling (IHS, [1]). The fundamentals and history of these are reviewed briefly in [18]. Though the quantitative measure of uniformity used for comparisons in [1] and [12] was somewhat flawed, it does appear that DHS gives better sampling uniformity than LHS, and IHS gives better sampling uniformity than DHS (but is increasingly more computationally expensive as the dimensionality of the parameter space increases). We have recently become aware of another LHS variant, "Optimal Symmetric LHS" (OSLHS, [21]) which also seems to improve the spatial uniformity of LHS samples. Its computational cost and performance relative to DHS and IHS are not yet known, however.

A number of other potential approaches for achieving uniform point placement that are not evolved from an LHS basis are reviewed (and some new ones are presented) in [7]. There, some quantitative metrics related to visual/sensory perception of point uniformity in 2-D are reviewed and some new ones are presented. Many of these non-LHS-based approaches appear to work very well in 2-D, but it is said in [7] that some of the methods may not be applicable or may not perform well in more than two dimensions, and some clearly will not scale up to high dimensions affordably. Others seem more promising for high dimensions, but have not yet been investigated enough.

The so-called "Quasi- Monte Carlo" (QMC, see *e.g.* [14]) sub-random low-discrepancy sequence methods can often achieve reasonably uniform sample placement in hypercubes. The strength of these sequence methods (Halton, Hammersley, Sobol, etc.), is that they can produce fairly uniform point distributions even though samples are added one at a time to the parameter space. The one-at-a-time incremental sampling of QMC (and SRS) enables these methods to have better efficiency prospects than CVT and LHS-type methods in the area of error estimation and control. Not only this, the results achieved are often quite good. For resolving the mean and standard deviation of response measures, Hammersley sequences were found in [11] to converge to within 1% of exact results 3 to 100 times faster than LHS over a large range of test problems. For resolving response probabilities, Hammersley and modified-Halton were found in [15] to perform roughly the same as LHS on balance over several test problems.

However, when the hyperspace dimension becomes moderate to large and/or the sampling density becomes high, some (perhaps all?) sequences suffer from spurious correlation of the samples. This is shown for standard Halton sequences in 16-D (ref. [12]) and 40-D (ref. [15]). Sometimes a modification can be found to suppress or delay the onset of spurious correlation – as a fix from the literature implemented in [15] shows for Halton sequences.

Recently, a long-recognized approach for achieving uniformity of point placement in M-dimensional volumes, called “**Centroidal Voronoi Tessellation**” (CVT), has been made computationally efficient ([10]) for implementing the principles of Centroidal Voronoi diagrams ([6],[13]). These diagrams subdivide arbitrarily shaped domains in arbitrary-dimensional space into arbitrary numbers of nearly uniform subvolumes, or Voronoi cells/regions. Given a set of N points $\{z_i\}$ ($i=1,\dots,N$) in an M-dimensional hypercube, the Voronoi region or Voronoi cell V_j ($j=1,\dots,N$) corresponding to z_j is defined to be all points in the hypercube that are closer to z_j than to any of the other z_i 's. The set $\{V_i\}$ ($i=1,\dots,N$) is called a Voronoi tessellation or Voronoi diagram of the hypercube, the set $\{z_i\}$ ($i=1,\dots,N$) being the generating points or generators. A *centroidal* Voronoi tessellation (CVT) is a special Voronoi tessellation with the property that each generating point z_i is itself the mass centroid of the corresponding Voronoi region V_i .

Although CVTs are deterministic, they can be converged to with probabilistic sampling methods. In [10], new probabilistic CVT construction algorithms were introduced, implemented, and tested. These methods are generally much more computationally efficient than previous deterministic and probabilistic methods for constructing CVTs.

The CVT concept and the algorithms in [10] for their construction can be generalized in many ways (see [3] for details). For example, instead of a hypercube, general regions in M-dimensional space can be treated. This feature has been exploited with great success (see [6]) for discretizing arbitrary 2-D and 3-D domain volumes for computational mechanics analysis with meshless analogues of finite element methods. Furthermore, points can be distributed non-uniformly according to a prescribed density function over the space (like the bi-normal density function that Figure 7 corresponds to).

In initial investigations ([2]) for 2-D, 7-D, and 20-D test cases, CVT has provided greater sampling uniformity than Halton, Hammersley, Sobol, SRS, LHS, DHS, and IHS according to a meaningful subset of non-flawed quantitative quality measures. Additionally, no degradation of sampling uniformity has been detected in higher dimensions (*i.e.*, for the 20-D case).

It is therefore natural to ask whether CVT can be applied for: A) statistical sampling over arbitrary-dimensional spaces of input random variables to calculate various statistics of output response behavior; B) function integration over arbitrarily shaped domains; and C) whether it can serve as a method for generating favorable point distributions for improved response-surface accuracy.

A preliminary positive indication regarding item C) for response surface generation is presented in [18]. There, CVT was shown on several 2-D test problems to provide superior point distributions for generating locally-conforming Moving Least Squares response surfaces. Point distributions by CVT, SRS, LHS, and a structured sampling method with deterministically uniform point placement ([17]) were tried in the study.

Reference [19] compared the above sampling methods for sampling performance in 2-D test problems of statistical function integration and estimation of response statistics associated with

uniformly distributed random-variable inputs (uncorrelated). By the same weighted measure of sampling effectiveness defined and used in Section 3.3 of this paper, CVT handily outperformed SRS, LHS, Halton, and Hammersley in resolving various statistics of response: mean, variance, and exceedence probabilities.

In this paper we take a first step toward examining the potential of CVT for improved statistical sampling given *non*-uniform inputs. Specifically, the performance of the above sampling methods and a new CVT variant (“Latinized” CVT) are compared for non-uniform uncorrelated input distributions in a 2-D test problem. Statistical sampling efficiencies are compared for calculating response mean, variance, and exceedence probabilities.

2. UNIFORMLY DISTRIBUTED TEST POINT-SETS AND THEIR MAPPING TO BINORMAL JOINT DENSITIES

Figure 1 shows three LHS and three corresponding CVT point sets for 100 samples in a 2D unit hypercube. The three LHS point sets were generated with the software [9] for different initial seeds (Seed1 = 123456789, Seed2 = 192837465, Seed3 = 987654321) and a uniform joint probability density function (JPDF) over the unit-square parameter space. The three corresponding CVT point sets were generated with the software [4] by using the LHS sets as initial conditions (starting point locations) from which the CVT iterations begin. In all cases each CVT set is much more uniform visually (and quantitatively, see [2]) than its associated LHS set. All three CVT sets are relatively similar visually and quantitatively, even though starting from three very different initial conditions given by the LHS sets.

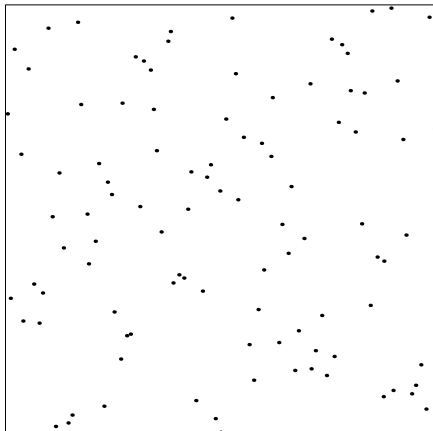
The LHS sets exhibit significantly more clustering and non-uniformity of the points than the CVT sets. For a visual indicator of sampling uniformity, Figure 5 compares a 25-sample LHS set and a 25-sample CVT set started from the LHS set. Non-overlapping circles are drawn in each domain, where each sample point has a circle centered about it having a radius proportional to the distance from the point to its nearest neighbor. The surrounding circles for the CVT set are all fairly uniform in size, whereas the variance in circle size is very large for the LHS set. Thus, the LHS point sets are relatively non-uniform in their “coverage” of the domain.

Besides the three LHS sets and three corresponding CVT sets shown in Figure 1, three SRS sets generated from initial seeds 1, 2, and 3 will also be tested here. These point sets can be seen in reference [18]. They exhibit even less uniformity than the LHS sets in Figure 1. Three CVT sets derived from the three different SRS sets as initial conditions can also be seen in [18]. The different LHS and SRS initial conditions do not have much of an impact on final CVT point uniformity, so the CVT algorithms appear to be robust in this regard.

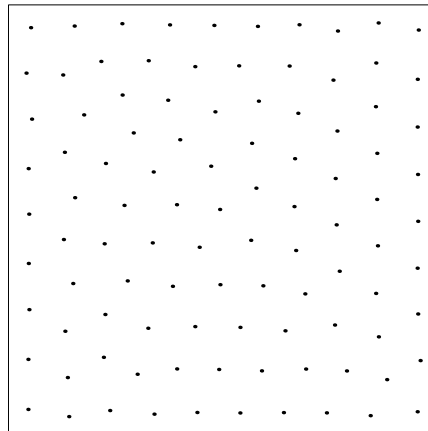
Figures 2 and 3 show Halton and Hammersley point sets and the corresponding CVT sets derived from them. Again, the resulting CVT sets are of essentially equivalent uniformity. The Halton point set is noticeable and quantitatively more uniform than any of the LHS sets; the Hammersley set is even more uniform than the Halton set; and the CVT sets in Figures 1, 2, and 3 are even more uniform than the Hammersley set.

In reference [19] we compared the mentioned point sets for effectiveness in 2-D test problems of statistical function integration and estimation of response statistics for the case of uniformly distributed input random variables (uncorrelated). The CVT point sets performed best, as will be summarized in Section 4 of this paper. In this paper we focus on comparing the per-

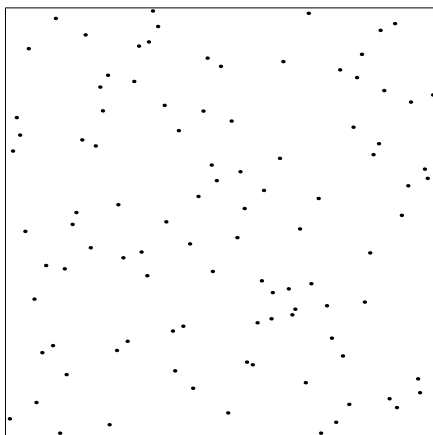
LHS1 point set (from seed 1)



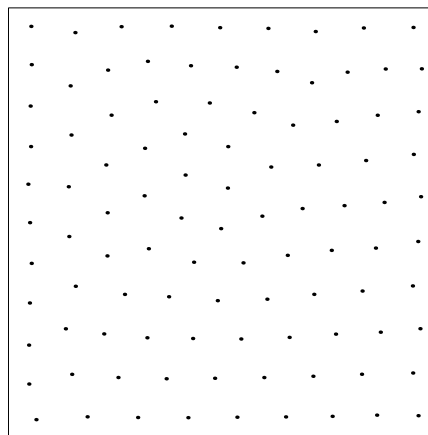
CVT-LHS1 point set (from LHS1)



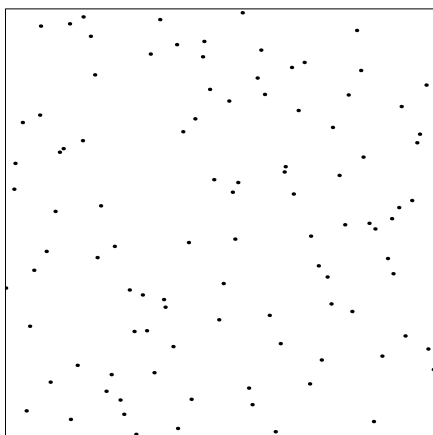
LHS2 point set (from seed 2)



CVT-LHS2 point set (from LHS2)



LHS3 point set (from seed 3)



CVT-LHS3 point set (from LHS3)

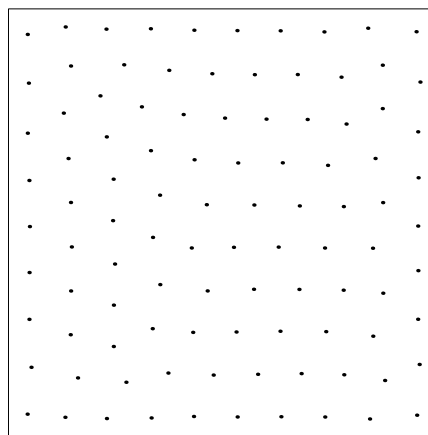


Figure 1. 100-point sample sets on a 2-D unit hypercube for: A) Left Column– uniform JPDP LHS Monte Carlo with three different initial seeds; and B) Right Column– corresponding uniform JPDP CVT sets starting from LHS sets as initial conditions.

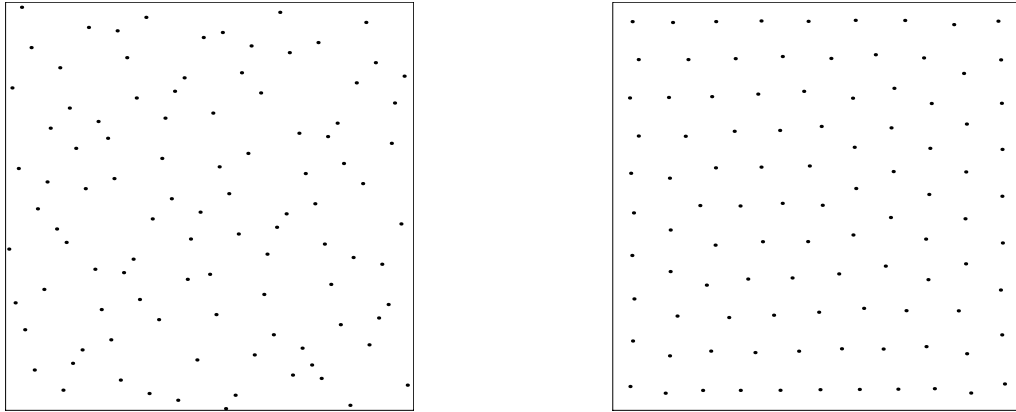


Figure 2. 100-point sample sets on 2-D unit hypercube for:
A) Left plot– Halton QMC sequence;
B) Right plot– corresponding CVT set starting from the Halton set as initial conditions.

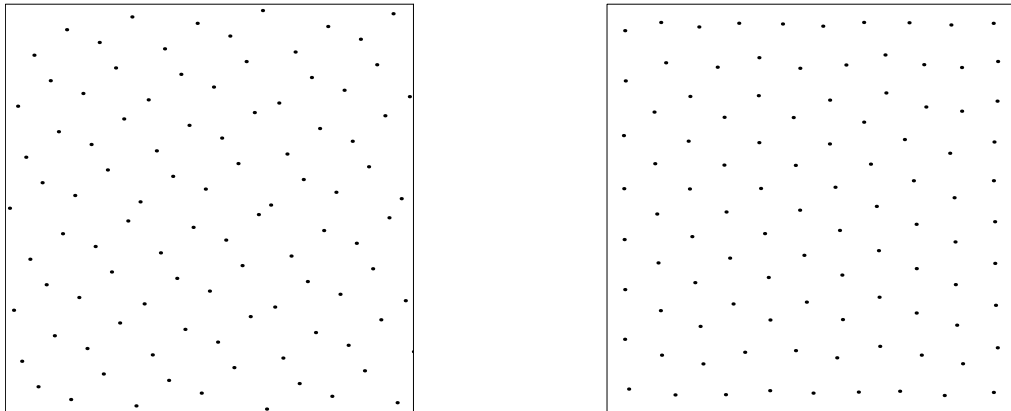


Figure 3. 100-point sample sets on 2-D unit hypercube for:
A) Left plot– Hammersley QMC sequence;
B) Right plot– corresponding CVT set starting from the Hammersley set as initial conditions.

formance of the above sampling methods and a new CVT variant (“Latinized” CVT, see [2]) as starting sets for mapped non-uniform point distributions intended to reflect a JPDF of uncorrelated normal inputs.

Our 2-D test problem has two random inputs $p1$ and $p2$ from independent normal distributions having means 0.5 and standard deviations $\sigma=0.5/3$. The corresponding JPDF is shown in Figure 4 after truncation of the function beyond the unit $p1$ - $p2$ parameter space and renormalization to integrate to one over the space.

The following procedure is used to map a set of uniformly distributed points to a set that reflects the desired non-uniform JPDF. First, for each random variable p in the problem, we consider its cumulative distribution function $CDF(p)$, where

$$\text{CDF}(p) = \int_0^{p \leq 1} \text{PDF}(p') dp' \quad \text{EQ 1}$$

and $\text{PDF}(p)$ is the probability density function of the random input p . We note that the value of the CDF ranges from 0 to 1 as the coordinate p ranges from 0 to 1 over our unit hypercube domain. It can also be shown that realizations $\{p_i\}$ drawn at random from a density function $\text{PDF}(p)$ map through EQ 1 (setting $p=p_i$) into a uniformly distributed set of realizations $\{\text{CDF}(p_i)\}$. This set is therefore distributed uniformly between 0 and 1.

Hence, we recognize that the above properties can be used to inverse-map numbers uniformly distributed between 0 and 1 (produced, e.g., by a random number generator), into realizations $\{p_i\}$ that would appear to be drawn from the density function $\text{PDF}(p)$. In a multidimensional problem, we inverse-map the coordinates of points uniformly distributed in the hypercube into transformed point sets that reflect the individual or “marginal” PDFs of the random inputs contributing to the Joint PDF. Figure 5 helps visualize the multidimensional mapping process. The tick marks on the coordinate axes indicate the projections of the points onto the coordinate axes. The (hopefully) uniformly distributed tick marks ranging from 0 to 1 on each coordinate axis give the random values that are inverse-mapped through the marginal CDFs into transformed tick locations ranging from 0 to 1 on the coordinate axes of the JPDF space. Thus, uniformly distributed points in a unit hypercube are transformed to new locations in the unit hypercube. The transformed coordinate sets define point locations distributed according to the target JPDF. (Our transformation algorithm for mapping uncorrelated uniformly distributed points sets to bivariate normally distributed point sets was verified as described in Section 3.2.) Correlation between random variables can be imparted with the rank correlation procedure described in [8].

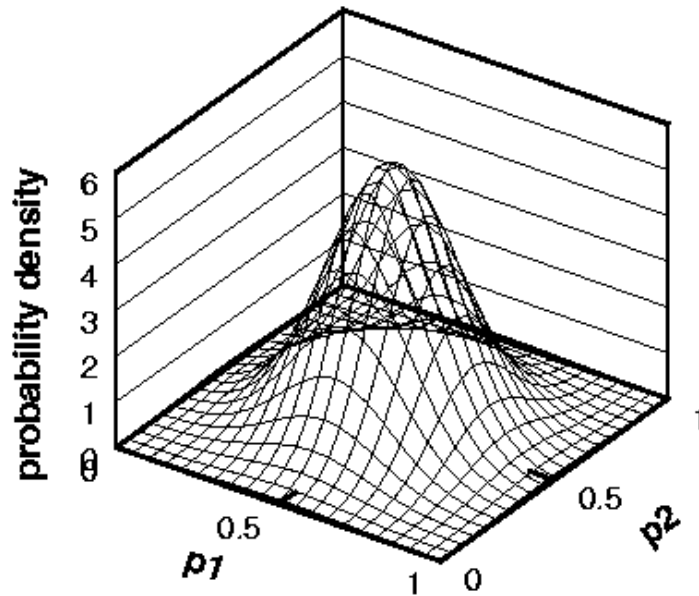
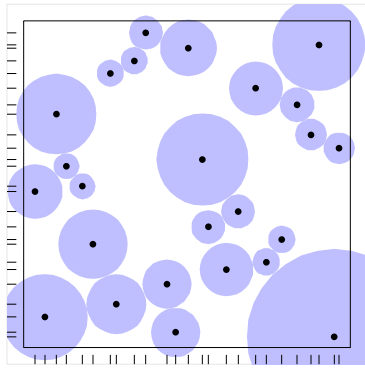
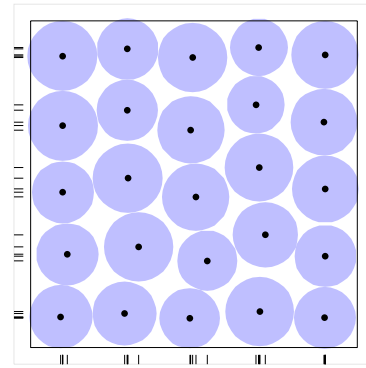


Figure 4. Joint Probability Density Function describing the random variables in the problem: normally distributed parameters p_1 and p_2 with means 0.5, std. deviations $\sigma=0.167$, and truncation of the unit square parameter space at 3σ above and below the mean values.



25 LHS points on Unit Square



25 CVT points on Unit Square

Figure 5. LHS and CVT sample sets showing relative uniformities of point spacing and discrepancies of point projections onto coordinate axes.

The mapping transformation presupposes a point set in an M -dimensional unit hypercube with point locations that project with uniform spacing onto all coordinate axes. However, consider the point sets in Figure 5. Though the CVT set is more uniform *volumetrically* than the LHS set, the LHS points clearly project more uniformly onto the coordinate axes. The projections of the CVT points occur in clusters that portray a “banded” distribution over the 0 to 1 range on each axis, as opposed to the desired uniform distribution. In the limit of a perfectly volumetrically uniform distribution of points over the domain, say a 5×5 rectangular array of points on the unit square, the points would project onto the coordinate axes making 5 uniformly spaced tick marks. These marks would inverse map through the marginal CDFs into only 5 different values or samples of each input variable. Thus, out of twenty-five sampling opportunities, each input variable is sampled at only five values. However, this is not *automatically* bad; the 25 particular *sets* or combinations of the five values of each input variable (when the uniform 5×5 grid of points is mapped to the JPDF space) may pose certain advantages over other point layouts. We are presently striving to understand the particular benefits and disadvantages that arise here.

The LHS point set, on the other hand, would sample each of the input variables at 25 different values. By the nature of LHS ([5]), a sample value would be picked at random from within each of the 25 equal intervals on the 0 to 1 range of each marginal CDF. These would map to 25 points in the JPDF space that each sample a different value of the input variables.

One measure of a point set’s uniformity of projection onto all the coordinate axes is called its *discrepancy*. As uniformity increases, discrepancy decreases. LHS is a lower-discrepancy sampling method than CVT is. Methods specifically designed with low discrepancy in mind are the quasi- or sub- random low-discrepancy sequence methods Halton, Hammersley, Sobol, etc. ([14]). These can have both lower discrepancy than standard LHS *and* higher volumetric uniformity. Though CVT tends to have better volumetric uniformity than the sequence methods, which helps its relative performance in other areas (*cf.* [18], [19]), it also has much higher discrepancy, which hurts its relative performance as a sampling basis for non-uniformly-random

inputs. Therefore, a hybrid of CVT and LHS has recently been formulated ([2]) with appears to have both lower discrepancy than pure CVT and higher volumetric uniformity than pure LHS. In the next section we compare the performance of this hybrid “Latinized” CVT (LCVT) against pure CVT and the other sampling methods.

Figure 6 shows uniformly distributed point sets from SRS, LHS, CVT, LCVT, Halton, and Hammersley methods, and corresponding mapped bi-normal point sets. The SRS, LHS, CVT, and LCVT uniform and mapped sets are typical of the three sets obtained from three different initial seeds described at the start of this Section. The SRS, LHS, CVT, and LCVT sets plotted in Figure 6 correspond to Seed 1. Our mapping process was checked by verifying that our bi-normal results mapped from the Seed 1, 2, and 3 uniform LHS sets in Figure 1 were essentially identical to bi-normal LHS sets generated directly from the LHS code ([8]) that produced the three uniform LHS sets. Thus, our mapping process corresponds almost exactly to the mapping process used in the well-pedigreed code [8].

The effect of high discrepancy in uniform CVT sets is immediately apparent in the mapped set in Figure 6. The mapped CVT set has a rectangular shaped layout of points rather than a circularly oriented layout seemingly more appropriate for the circularly symmetric bi-normal JPDF targeted (Figure 4). Unexpectedly, we find in the next section that this non-intuitive rectangular shaped set of points actually performs relatively well among the six types of mapped sets shown in Figure 6. This rectangular-shaped set performs much better, in fact, than the much more likely looking set shown in Figure 7, which was generated directly with density-weighted CVT. The set mapped from uniform Latinized CVT appears much closer to a bi-normal density than the rectangular mapped CVT set, but actually doesn’t perform quite as well. The performance of the various mapped sets is examined more closely in the next Section.

3. EVALUATION OF STATISTICAL SAMPLING EFFECTIVENESS OF THE METHODS

3.1. 2-D Model Response Function and Statistical Measures of Response in Performance Evaluation

Figure 8 shows an analytic multi-modal function describing system response r as a function of two system inputs p_1 and p_2 :

$$r(p_1,p_2)=\left[0.8\kappa+0.35\sin\left(2.4\pi\frac{\kappa}{\sqrt{2}}\right)\right][1.5\sin(1.3\theta)] \tag{EQ 2}$$

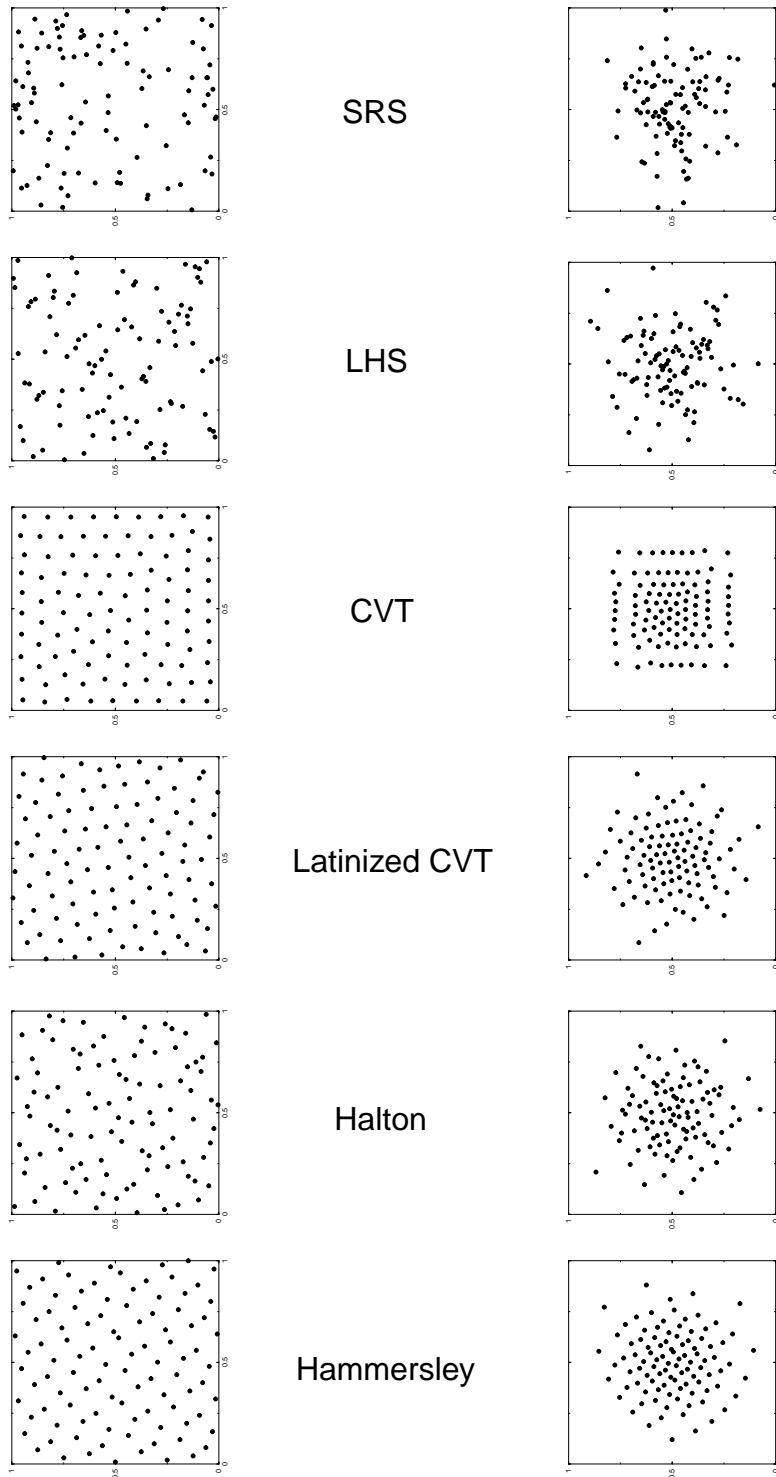
on the domain $0 \leq p_1 \leq 1$ and $0 \leq p_2 \leq 1$, where $\kappa = \sqrt{(p_1)^2 + (p_2)^2}$, $\theta = \text{atan}\left(\frac{p_2}{p_1}\right)$.

A statistical problem arises if p_1 and p_2 are random variables. In that case, any particular realization p_{1i} and p_{2i} of the stochastic variables yields a deterministic response r_i as given by the above functional relationship. An ensemble of responses accompanies the different realizations of p_1 and p_2 as they vary stochastically or randomly according to their individual propensities, or joint propensities if the two variables are correlated.

The JPDF likelihood function for attaining various input combinations of p_1 and p_2 maps through the response function $r(p_1,p_2)$ into a corresponding likelihood function for response values. Operationally, the resulting response probability density function, $\text{PDF}(r)$, can be approached closer and closer via Monte Carlo sampling as more and more parameter sets or real-

Unmapped, Uniform

Mapped Bi-normal



**Figure 6. 100-point sample sets on a 2-D unit hypercube for:
Left Column– uniformly distributed point sets
Right Column– corresponding bi-normally distributed point sets
mapped from uniform sets.**

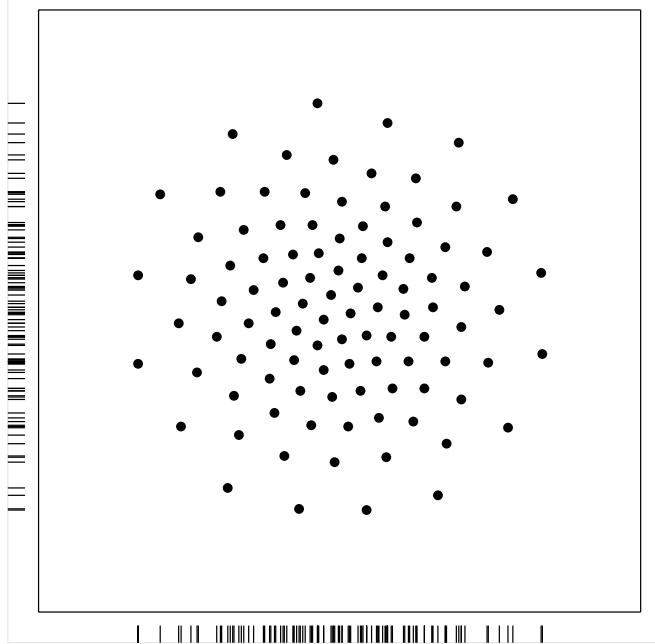


Figure 7. 100-point set in a unit square, generated directly with density-weighted CVT to model the bi-normal joint probability density function shown in Figure 4.

izations $(p_1, p_2)_i$ are randomly generated from the governing input JPDF and are propagated through the response function $r(p_1, p_2)$ into response realizations r_i . The response realizations are distributed in the response space (*i.e.*, along the response coordinate axis r) with a density that, as more and more samples are added, trends toward the exact PDF of response.

Very often, only certain statistical measures of the PDF of response are desired or can be reasonably estimated. Response mean, μ_r , and standard deviation, σ_r , can be estimated directly from the mean $\hat{\mu}_r$ and standard deviation $\hat{\sigma}_r$ of the population or set $\{r_i\}$ of realizations. We have the following definitions:

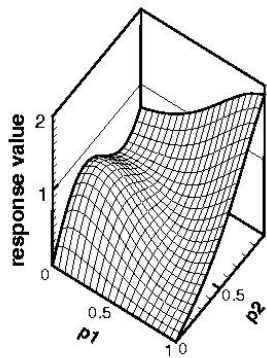


Figure 8. 2-D model function for system response as a function of input parameters p_1 and p_2 .

$$\hat{\mu}_r = \frac{1}{N} \sum_{i=1}^N r_i \quad \text{EQ 3}$$

$$\hat{\sigma}_r = \left[\frac{1}{N-1} \sum_{i=1}^N (r_i - \hat{\mu}_r)^2 \right]^{\frac{1}{2}} \quad \text{EQ 4}$$

where N is the number of realizations or samples of response.

Also of interest is the probability of response exceeding (or not exceeding) some particular threshold value r_T . The former is equivalent to the volume integral of the joint probability density function JPDF(p_1, p_2) integrated over the region of the p_1 - p_2 domain where response exceeds the stipulated threshold r_T . Three such regions corresponding to thresholds $r_T=1.0, 0.5,$ and 0.2 are shown in Figure 9 (as shaded areas on the cutting planes $r_T=1.0, 0.5, 0.2$) for our model function EQ 2.

Exceedence probability is very simply estimated as the ratio of the number of calculated response values at or above the given threshold value, to the total number of samples N drawn from the JPDF. As the number of response realizations increases, the estimate (quotient) trends toward greater accuracy, *i.e.*, toward the actual exceedence probability. This is of course also true for the estimates $\hat{\mu}_r$ and $\hat{\sigma}_r$ of response mean and standard deviation.

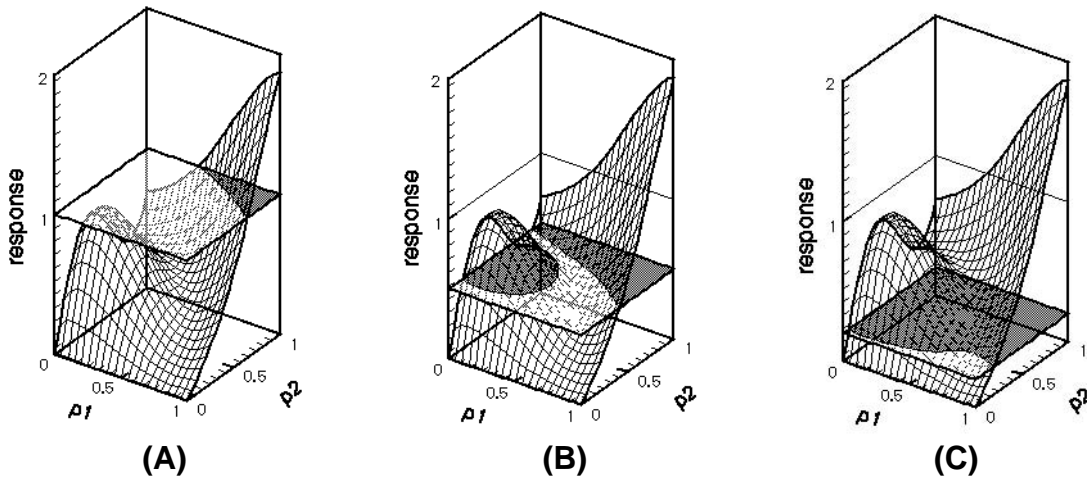


Figure 9. Cutting planes through exact function showing associated exceedence (shaded) and complement (unshaded) regions of the p_1 - p_2 parameter space for response threshold values of 1.0, 0.5, and 0.2, respectively.

3.2. Comparison of Response Statistics from Various Sampling Methods

Here we compare estimates of response mean, standard deviation, and exceedence probabilities as obtained from the various sample sets represented by the right column in Figure 6. We map these $\{(p_1, p_2)_i\}$ sets through our response function EQ 2 to obtain corresponding response sets, and then calculate the aforementioned statistics of the response populations. We then compare the calculated statistics of each response set to reference values obtained from using three million SRS samples at parameter values generated by the sampling code [9]. The reference val-

ues are actually averages of three results, each obtained from one million samples generated from random initial seeds “X”, “Y”, and “Z” (different from seeds 1, 2, and 3 used to generate the 100-sample sets).

Three “replicate” sets of one million samples each are used in preference to one set of three million samples so that *empirical* confidence intervals (CI) on the calculated averages could be compared against their classical CI to reaffirm or caveat them. (Recent research ([16], [20]) has shown that for SRS, empirical CI appear to be somewhat more accurate than classical CI.) Empirical CI are formed by assuming the calculated statistic (response mean, standard deviation, or exceedence probability) is a random realization from a normal or nearly normal distribution about the exact result. Hence a T-distribution with $3 - 1 = 2$ degrees of freedom can be used to get confidence intervals about the small-sample average of the three replicates. Thus, for 95% empirical CI the following formula is used:

$$95\% \text{ confidence half-interval} = 4.303 \frac{\hat{\sigma}_{est}}{\sqrt{3}} \tag{EQ 5}$$

where $\hat{\sigma}_{est}$ is the sample standard deviation (*cf.* EQ 4) of the three estimates.

Table 1 shows various estimates of response mean, standard deviation, and exceedence probabilities calculated from the three one-million-sample SRS sets. The average and standard deviation of the estimates is also shown in the table.

Table 1. Calculated response statistics for reference values (10⁶ samples, Bi-normal JPDF, SRS)

		<u>response statistic</u>			
		$\hat{\mu}_r$	$\hat{\sigma}_r$	$\hat{P}_{0.2}$	$\hat{P}_{0.5}$
REALIZATION	1	0.511872	0.162834	0.984429	0.448457
	2	0.511824	0.162733	0.984585	0.447915
	3	0.511940	0.162737	0.984511	0.449029
	average	0.511879	0.162768	0.984508	0.448467
	std. dev.	5.829E-05	5.720E-05	7.803E-05	0.000557

3.2.1. Mean of Response

The average of the three 10⁶-sample estimates of mean response is taken as the reference value, $\hat{\mu}_{ref} = 0.511879$ from Table 1. Empirical confidence intervals on this reference mean are obtained by substituting the standard deviation of the estimates, $\hat{\sigma}_{est} = 5.829E-05$ from Table 1, into EQ 5. Thus, empirical 95% half-CI are 0.000145. When the reference mean is calculated based on the entire population of $N = 3 \times 10^6$ samples, the value doesn’t change from the averaged value based on three separate 10⁶-sample sets, but the classical CI can be computed. The classical 95% half-CI from standard statistical formulas is somewhat larger, at 0.000185. Using the larger (classical) CI here to be conservative, we say that with at least 95% certainty the true re-

sponse mean μ lies within the range $\hat{\mu}_{ref} \pm 0.000185 = (0.511694, 0.512064)$. The CI range ± 0.000185 is typically very small compared to the *nominal* differences listed in Tables 2 and 3 between $\hat{\mu}_{ref}$ and the estimates of mean response from the 100-sample sets.

We take the differences from $\hat{\mu}_{ref}$ in Tables 2 and 3 as nominal measures of the error of the estimates from the 100-sample sets. For SRS, LHS, CVT, and LCVT methods there is no unique 100-sample set. For SRS and LHS the sets depend on the initial seed and the particular pairing of the 0 to 1 random variates on the p1 and p2 axes (of a uniform JPDF set). The CVT, and LCVT sets further depend –fairly insensitively if enough iterations are performed to stabilize certain uniformity measures, see [2]– on the starting sample set (initial condition). We therefore use three instantiations of SRS, LHS, CVT, and LCVT sets to begin to obtain a representative picture of the errors we might expect from a random realization of each of these types of sets. For each of these methods we average the individual errors from the three instantiations to determine an average magnitude of error. This measure reflects contributions from both the average error (bias) in the three estimates, as well as the variance of the three results. (This error measure is zero only if both the average error (bias) is zero and the variance of the estimates is zero.) Furthermore, this error measure applies as well to the Halton and Hammersley results which consist of only one instantiation because they are deterministic sampling methods.

To also obtain a broad picture of the each method’s sampling efficacy across the different types of statistics calculated, we just use a simple ranking scheme for method accuracy for each of the various calculated statistics (response mean, variance, and exceedence probabilities). This allows us to compare method performance across the different types of statistics calculated. This is perhaps somewhat more satisfying than the piecemeal comparisons in, e.g., [15] and [19] that fail to give an explicit impression (quantitative balanced indicator) of the overall performance of the various sampling methods across a matrix of test problems. Hence, the accuracy ranking of each method with respect to average magnitude of error is given on the final lines in Tables 2 and 3. Rank 1 indicates the method was the most accurate and therefore ranked first in performance. Rank 6 indicates the method was the least accurate among the sampling schemes tried.

Table 2. Calculated response means (100 samples, Normal 2D JPDF)

		<u>SRS</u>		<u>LHS</u>		<u>Latinized CVT</u>	
		$\hat{\mu}_r$	$\hat{\mu}_r$ error	$\hat{\mu}_r$	$\hat{\mu}_r$ error	$\hat{\mu}_r$	$\hat{\mu}_r$ error
REALIZATION	1	0.49980	-0.01208	0.51466	+0.00278	0.51238	+0.00050
	2	0.51058	-0.00130	0.51110	-0.00079	0.50780	-0.00408
	3	0.50315	-0.00873	0.50752	-0.00436	0.51182	-5.867E-05
average		0.50451	-0.00737	0.51109	-0.00079	0.51067	-0.00121
std. dev.		0.00552	0.00552	0.00357	0.00357	0.00250	0.00250
avg. error mag.			0.00737 Rank 6		0.00264 Rank 4		0.00155 Rank 1

Table 3. Calculated response means (100 samples, Normal 2D JPDF)

		<u>CVT</u>		<u>Halton</u>		<u>Hammersley</u>	
		$\hat{\mu}_r$	$\hat{\mu}_r$ error	$\hat{\mu}_r$	$\hat{\mu}_r$ error	$\hat{\mu}_r$	$\hat{\mu}_r$ error
REALIZATION	1	0.50868	-0.00320	0.50565	-0.00623	0.51029	-0.00159
	2	0.51148	-0.00040				
	3	0.50840	-0.00348				
average		0.50952	-0.00236	0.50565	-0.00623	0.51029	-0.00159
std. dev.		0.00170	0.00170				
avg. error mag.			0.00236 Rank 3		0.00623 Rank 5		0.00159 Rank 2

3.2.2. Standard Deviation of Response

Tables 4 and 5 show the estimates of the standard deviation of response. Nominal errors from the reference value $\hat{\sigma}_{ref}=0.162768$ are also shown. This value is the average of the three standard deviations in Table 1 calculated from the three 10^6 SRS sets. The standard deviation of these three estimates is $\hat{\sigma}_{est}=5.719E-05$. Empirical 95% half-CI by EQ 5 are 0.000142. Accordingly, we say that with 95% confidence the true response standard deviation σ lies within the range $\hat{\sigma}_{ref} \pm 0.000142 = (0.162626, 0.162910)$. The CI are negligibly small.

Table 4. Calculated response standard deviations (100 samples, Normal 2D JPDF)

		<u>SRS</u>		<u>LHS</u>		<u>Latinized CVT</u>	
		$\hat{\sigma}_r$	$\hat{\sigma}_r$ error	$\hat{\sigma}_r$	$\hat{\sigma}_r$ error	$\hat{\sigma}_r$	$\hat{\sigma}_r$ error
REALIZATION	1	0.16874	+0.00597	0.18054	+0.01777	0.15782	-0.00495
	2	0.16265	-0.00012	0.15570	-0.00707	0.15532	-0.00745
	3	0.15699	-0.00578	0.14191	-0.02086	0.15554	-0.00723
average		0.16279	2.533E-05	0.15938	-0.00339	0.15623	-0.00654
std. dev.		0.00588	0.00588	0.01958	0.01958	0.00138	0.00138
avg. error mag.			0.00396 Rank 2		0.01523 Rank 6		0.00654 Rank 3

Table 5. Calculated response standard deviations (100 samples, Normal 2D JPDF)

		<u>CVT</u>		<u>Halton</u>		<u>Hammersley</u>	
		$\hat{\sigma}_r$	$\hat{\sigma}_r$ error	$\hat{\sigma}_r$	$\hat{\sigma}_r$ error	$\hat{\sigma}_r$	$\hat{\sigma}_r$ error
REALIZATION	1	0.14804	-0.01473	0.15415	-0.00862	0.16077	-0.00200
	2	0.14906	-0.01371				
	3	0.14615	-0.01662				
average		0.14775	-0.01502	0.15415	-0.00862	0.16077	-0.00200
std. dev.		0.00148	0.00148				
avg. error mag.			0.01502 Rank 5		0.00862 Rank 4		0.00200 Rank 1

3.2.3. Response Exceedence Probability for $r_T=0.2$

Tables 6 and 7 show the estimates of the exceedence probability (EP) corresponding to a response threshold level of $r_T=0.2$. Nominal errors from the reference value $\hat{P}_{0.2, ref}=0.984508$ are also shown. This value is the average of the three EPs in Table 1 calculated from the three 10^6 SRS sets. The standard deviation of these three estimates is $\hat{\sigma}_{est}=7.803E-05$. Empirical 95% half-CI by EQ 5 are 0.000194. When the reference EP is calculated based on the entire population of $N=3 \times 10^6$ samples, the value doesn't change from the averaged value based on three separate 10^6 -sample sets, but classical CI can be computed. The classical 95% half-CI from standard statistical formulas is somewhat smaller, at 0.000140. Using the larger (empirical) 95% half-CI for conservatism, we say that to 95% confidence the true probability $P_{0.2}$ of response exceeding the threshold value $r_T=0.2$ lies within the range $\hat{P}_{0.2, ref} \pm 0.000194 = (0.984314, 0.984702)$. The CI are negligibly small. We note that the SRS and LHS results are both ranked at 4.5 because together they occupy the 4th and 5th ranks and both have the same error magnitude.

Table 6. Calculated response exceedence probabilities, threshold=0.2 (100 samples, Normal 2D JPDF)

		<u>SRS</u>		<u>LHS</u>		<u>Latinized CVT</u>	
		$\hat{P}_{0.2}$	$\hat{P}_{0.2}$ error	$\hat{P}_{0.2}$	$\hat{P}_{0.2}$ error	$\hat{P}_{0.2}$	$\hat{P}_{0.2}$ error
REALIZATION	1	0.97	-0.01451	0.98	-0.00451	0.99	+0.00549
	2	0.99	+0.00549	0.99	+0.00549	0.98	-0.00451
	3	0.99	+0.00549	1.00	+0.01549	0.98	-0.00451
average		0.98333	-0.00117	0.99	+0.00549	0.98333	-0.00118
std. dev.		0.01155	0.01155	0.01	0.01	0.00577	0.00577
avg. error mag.			0.00850 Rank 4.5		0.00850 Rank 4.5		0.00484 Rank 2

Table 7. Calculated response exceedence probabilities, threshold=0.2 (100 samples, Normal 2D JPDF)

		<u>CVT</u>		<u>Halton</u>		<u>Hammersley</u>	
		$\hat{P}_{0.2}$	$\hat{P}_{0.2}$ error	$\hat{P}_{0.2}$	$\hat{P}_{0.2}$ error	$\hat{P}_{0.2}$	$\hat{P}_{0.2}$ error
REALIZATION	1	0.99	+0.00549	0.97	-0.01451	0.98	-0.00451
	2	0.99	+0.00549				
	3	0.99	+0.00549				
average		0.99	+0.00549	0.97	-0.01451	0.98	-0.00451
std. dev.		0.0	0.0				
avg. error mag.			0.00549 Rank 3		0.01451 Rank 6		0.00451 Rank 1

3.2.4. Response Exceedence Probability for $r_T=0.5$

Tables 8 and 9 show the estimates of the exceedence probability (EP) corresponding to a response threshold of $r_T=0.5$. Nominal errors from the reference value $\hat{P}_{0.5, ref}=0.448467$ are also shown. This value is the average of the three EPs in Table 1 calculated from the three 10^6 SRS sets. The standard deviation of these three estimates is $\hat{\sigma}_{est}=0.000557$. Empirical 95% half-CI by EQ 5 are 0.001384. When the reference EP is calculated based on the entire population of $N=3 \times 10^6$ samples, the value doesn't change from the averaged value based on three separate 10^6 -sample sets, but classical CI can be computed. The classical 95% half-CI from standard sta-

tistical formulas is considerably smaller, at 0.000563. Using the larger (empirical) 95% half-CI for conservatism, we say that to 95% confidence the true probability $P_{0.5}$ of response exceeding the threshold value $r_T=0.5$ lies within the range $\hat{P}_{0.5, ref} \pm 0.001384 = (0.447083, 0.449851)$. The CI are negligibly small.

Table 8. Calculated response exceedence probabilities, threshold=0.5 (100 samples, Normal 2D JPFD)

		<u>SRS</u>		<u>LHS</u>		<u>Latinized CVT</u>	
		$\hat{P}_{0.5}$	$\hat{P}_{0.5}$ error	$\hat{P}_{0.5}$	$\hat{P}_{0.5}$ error	$\hat{P}_{0.5}$	$\hat{P}_{0.5}$ error
REALIZATION	1	0.43	-0.01847	0.44	-0.00847	0.46	+0.01153
	2	0.41	-0.03847	0.45	+0.00153	0.43	-0.01847
	3	0.43	-0.01847	0.45	-0.00847	0.47	+0.02153
average		0.42333	-0.02513	0.44667	-0.0018	0.45333	+0.00487
std. dev.		0.01155	0.01155	0.00577	0.00577	0.02082	0.02082
avg. error mag.			0.02513 Rank 6		0.00384 Rank 1		0.01718 Rank 5

Table 9. Calculated response exceedence probabilities, threshold=0.5 (100 samples, Normal 2D JPFD)

		<u>CVT</u>		<u>Halton</u>		<u>Hammersley</u>	
		$\hat{P}_{0.5}$	$\hat{P}_{0.5}$ error	$\hat{P}_{0.5}$	$\hat{P}_{0.5}$ error	$\hat{P}_{0.5}$	$\hat{P}_{0.5}$ error
REALIZATION	1	0.44	-0.00847	0.46	+0.01153	0.44	-0.00847
	2	0.46	+0.01153				
	3	0.46	+0.01153				
average		0.45333	+0.00487	0.46	+0.01153	0.44	-0.00847
std. dev.		0.01155	0.01155				
avg. error mag.			0.01051 Rank 3		0.01153 Rank 4		0.00847 Rank 2

3.3. Weighted Measure of Statistical Sampling Merit

The performance rankings for the sampling schemes and statistical quantities tested are summarized in Table 10. The last column contains a normalized weighted figure of merit which is a broad measure of each method's sampling performance across the different types of statistics calculated. This figure of merit is obtained by first averaging the rankings for the (two) exceedence probabilities calculated, and then averaging this rank for EPs in with the ranks for the mean and standard deviation calculations. These averages are then divided by the number of sampling methods involved. Hence, the normalized ranks in this column add up to unity. This type of normalization allows comparison to other investigations such as those in, e.g., [11], [15] and [19], if their results are also normalized in this manner. The bar chart in Figure 10 helps visually assess the relative performance of the sampling methods according to our normalized figure of merit. The shorter the bar, the better the particular method ranks on balance across all the statistical quantities calculated. We see that Hammersley sampling ranked overall best on this series of test problems, then LCVT, CVT, LHS, SRS, and finally Halton.

Table 10. Sampling Method accuracy rankings for various calculated statistics of response

Sampling Method	response mean	response standard deviation	exceedence probability (0.2 threshold)	exceedence probability (0.5 threshold)	normalized weighted average
SRS	6	2	4.5	6	0.21
LHS	4	6	4.5	1	0.20
CVT	3	5	3	3	0.17
LCVT	1	3	2	5	0.12
Halton	5	4	6	4	0.22
Hammersley	2	1	1	2	0.07

A second bar chart corresponding to the investigation in [19] is plotted in Figure 10. The investigation was similar to the one in this paper, but compared calculated statistics based on a uniform JPDF, and did not include the LCVT sampling method. Since there is no mapping here from uniform sets to nonuniform JPDFs, only volumetric uniformity matters here and discrepancy properties are immaterial. Since pure CVT is more volumetrically uniform than LCVT, and for that matter, more volumetrically uniform than all the other sampling methods we've tested, CVT would be expected to generally rank best. This is the case shown in Figure 10 for the set of test problems investigated in [19].

4. DISCUSSION AND CONCLUSION

According to our weighted figure of merit, Hammersley sampling strongly ranked overall best on the set of bi-normal JPDF test problems in this paper, then LCVT, CVT, LHS, SRS, and

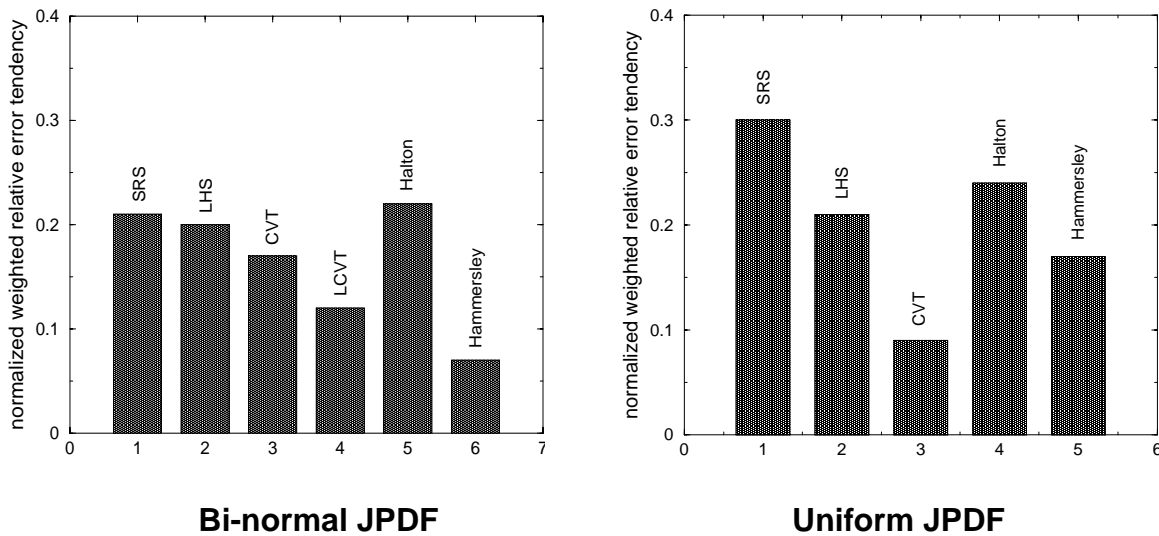


Figure 10. Normalized weighted measure of sampling method relative error tendency in calculated mean, standard deviation, and $r_T = 0.2$ and 0.5 exceedence probabilities for uniform and bi-normal joint probability densities in 2D test problems.

finally Halton. Furthermore, in [11], Hammersley was found to be significantly more efficient than SRS and LHS for resolving mean and standard deviation of response over a large set of test problems. For resolving response probabilities, Hammersley and modified-Halton were found in [15] to perform roughly the same as LHS on balance over several test problems.

Hence, Hammersley is consistently the best performer or among the top performers in these empirical studies. Hammersley is also the only one of the top contenders in these studies that allows incremental addition of samples to the parameter space (as little as one at a time), which enables it to have better efficiency prospects in both error estimation and control. Given these apparent advantages, it seems that Hammersley sampling might be a superior choice in many circumstances. However, when the number of random inputs grows beyond 10 or so dimensions and/or the sampling density in the hypercube becomes high, Hammersley might suffer from the spurious correlation effects that plague other sub-random sequence methods. This is shown, e.g., for standard Halton sequences in 16-D (ref. [12]) and 40-D (ref. [15]). This is something the authors need to further inquire about; the answer may already exist in the literature.

Furthermore, we cannot yet dismiss the competitive potential of CVT or LCVT based on the single limited investigation conducted in this paper. In particular, more than three instantiations of SRS, LHS, CVT, and LCVT point sets are needed to more reliably reflect the true performance tendencies of these methods on our test problems. Also, sample sets of much larger size than 100 would be valuable particularly to get another significant digit of resolution in the calculated exceedence probabilities in the study. Moreover, our results are somewhat tied to the specific figure of merit employed in this study. This figure of merit has the advantage that it allows comparison of merit across different types of statistics calculated and different problem

sets, but other better measures may exist for our purposes. Certainly, our weighted metric does not reveal method performance in the individual categories of response mean, standard deviation, and exceedence probability (but these can be found in Table 10 for the problems in this paper).

Finally, empirical studies are only point glimpses of the relative accuracy tendencies of one method over another under a very specific set of conditions. Certainly, much more empirical work needs to be performed to assess the performance of CVT and LCVT versus other sampling methods over a diverse problem space, but even more valuable would be more theoretical work to ascertain which method might be expected to perform best under given conditions (the characteristics of the function involved; the number of input random variables/dimensions; character of the JPDF, etc.).

This being said, we have early empirical indications of the promise of CVT in uniform JPDF problems. In [19], CVT strongly ranked overall best as expected, then Hammersley, LHS, Halton, and finally SRS. In particular, for statistical integration of functions, which involves uniform sampling over the integration domain, CVT appears to be the natural best choice theoretically, as corroborated by findings in [19]. Also, in point placement for response-surfaces, CVT appears very promising relative to other structured and unstructured sampling methods (see [18]). Already, for irregular (non-hypercube) interpolation and integration domains, the uniformity of CVT sampling over the domain gives it a well recognized status in the application of 2-D and 3-D meshless finite-element methods.

Hence, when volumetric sampling uniformity is desirable, early indications are that CVT performs very well versus other sampling methods. However, to reiterate, much more empirical and theoretical work remains to be done to broadly assess and characterize the potential of CVT and LCVT for various sampling tasks.

REFERENCES

- [1] Beachkofski, B.K., and R.V. Grandhi, "Improved Distributed Hypercube Sampling," paper AIAA-2002-1274, 43rd Structures, Structural Dynamics, and Materials Conference, April 22-25, 2002, Denver, CO.
- [2] Burkardt, J., M. Gunzburger, H. Nguyen, J. Peterson, and Y. Saka, "A new methodology for point sampling I: Uniform sampling in hypercubes" in preparation
- [3] Burkardt, J., M. Gunzburger, H. Nguyen, J. Peterson, and Y. Saka, "A new methodology for point sampling II: Nonuniform, anisotropic sampling in general regions" in preparation
- [4] Burkardt, J.V., M.D. Gunzburger, J.S. Peterson, research sampling software, http://www.csit.fsu.edu/~burkardt/f_src/cvt_dataset/cvt_dataset.html
- [5] Conover, W.M., "On a Better Method for Selecting Input Variables," 1975 unpublished Los Alamos National Laboratories manuscript, reproduced as Appendix A of "Latin Hypercube Sampling and the Propagation of Uncertainty in Analyses of Complex Systems" by J.C. Helton and F.J. Davis, Sandia National Laboratories report SAND2001-0417, printed November 2002.
- [6] Du, Q., V. Faber, M. Gunzburger, "Centroidal Voronoi tessellations: applications and algorithms," *SIAM Review*. Vol. 41, 1999, pp. 637-676.

- [7] Hanson, K.M., "Quasi-Monte Carlo: halftoning in high dimensions?," to appear in Proceedings of SPIE 5016 (2003), *Computational Imaging*, C.A. Bouman and R.L. Stevenson, eds.
- [8] Iman, R.L., and Conover, W.J., "A Distribution-Free Approach to Inducing Rank Correlation Among Input Variables," *Communications in Statistics*, B11(3) 1982a, pp. 311-334.
- [9] Iman, R.L., and Shortencarier, M.J., "A FORTRAN77 Program and User's Guide for the Generation of Latin Hypercube and Random Samples to Use with Computer Models," Sandia National Laboratories report SAND83-2365 (RG), printed March 1984.
- [10] Ju, L., Q. Du., M. Gunzburger, "Probabilistic algorithms for centroidal Voronoi tessellations and their parallel implementation," *Parallel Computing*, Vol. 28, 2002, pp. 1477-1500.
- [11] Kalagnanam, J.R., and Diwekar, U.M., "An Efficient Sampling Technique for Off-line Quality Control," *Technometrics*, August 1997, Vol.39, No.3.
- [12] Manteufel, R.D., "Distributed Hypercube Sampling Algorithm," paper AIAA-2001-1673, 42nd Structures, Structural Dynamics, and Materials Conference, April 16-19, 2001, Seattle, WA.
- [13] Okabe, A., B. Boots, K. Sugihara, S. Chui, *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*, 2nd Edition, Wiley, Chichester, 2000.
- [14] Press, W.H., S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, *Numerical Recipes in Fortran: The Art of Scientific Computing*, 2nd Edition, Cambridge University Press, 1992.
- [15] Robinson, D., and Atcitty, D., "Comparison of Quasi- and Pseudo- Monte Carlo Sampling for Reliability and Uncertainty Analysis," paper AIAA-1999-1589, 40th Structures, Structural Dynamics, and Materials Conference, April 12-15, St. Louis, MO.
- [16] Romero, V.J., "Effect of Initial Seed and Number of Samples on Simple-Random and Latin-Hypercube Monte Carlo Probabilities –Confidence Interval Considerations," paper PMC2000-176, 8th ASCE Specialty Conference on Probabilistic Mechanics and Structural Reliability, Univ. of Notre Dame, IN, July 24-26, 2000.
- [17] Romero, V.J., and S.D. Bankston, "Progressive-Lattice-Sampling Methodology and Application to Benchmark Probability Quantification Problems," Sandia National Laboratories report SAND98-0567, printed March 1998.
- [18] Romero, V.J., J.S. Burkardt, M.D. Gunzburger, J.S. Peterson, T. Krishnamurthy, "Initial Application and Evaluation of a Promising New Sampling Method for Response Surface Generation: Centroidal Voronoi Tessellation," paper AIAA-2003-2008, 44th Structures, Structural Dynamics, and Materials Conference (5th AIAA Non-Deterministic Approaches Forum), Apr. 7-10, 2003, Norfolk, VA.
- [19] Romero, V.J., J.S. Burkardt, M.D. Gunzburger, J.S. Peterson, "Initial Evaluation of Centroidal Voronoi Tessellation Method for Statistical Sampling and Function Integration," Sandia National Laboratories library archive SAND2003-3672C, an extended version (with comparison of exceedence probabilities) of the paper in the proceedings of the 4th International Symposium on Uncertainty Modeling and Analysis (ISUMA'03), Sept. 21-24, 2003, University of Maryland, College Park, MA.
- [20] Romero, V.J., and A.L. Mellen, 2002, preliminary research findings not yet published.
- [21] Ye, K.Q., W. Li, A. Sudjianto, "Algorithmic Construction of Optimal Symmetric Latin Hypercube Designs," *Journal of Statistical Planning and Inference*, Vol. 90 (2000), pp. 145-159.