ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

ISSN 2454-9940
INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
www.ijasem.org
Vol 18, Issue 1, 2024

Exploring Model Populations and Experimental Design: Coverage of Parameter Space Using Latin Hypercube and Orthogonal Sampling

Mrs.V R R Lakshmi Kumari
Assistant Professor, Department of CSE,
Malla Reddy College of Engineering for Women.,
Maisammaguda., Medchal., TS, India

Abstract

In this research, we utilize simulations to construct a guess about how many times Latin Hypercube sampling will need to be run in order to cover a t-dimensional subspace of a d-dimensional parameter space of size n. The formula looks like this: P(k, n, d, t)=1 ek/nt1. We propose that this coverage formula holds true regardless of d, and use this to draw parallels between the two processes (population modelling and experimentation). We further demonstrate that, at the sub-block size level, Orthogonal sampling provides more consistent coverage of the t-dimensional subspace than Latin Hypercube sampling. Particularly applicable to efforts at uncertainty quantification and sensitivity studies, these concepts are worth considering.

Key words :

Model Population, Latin Hypercube Sampling, and Orthogonal Sampling are some relevant terms to keep in mind.

Introduction

In mathematics, it is common practice to specify model parameters to multiple significant digits. The intrinsic variability in the underlying dynamical processes is sometimes lost when these parameters are fitted to a collection of mean observational/experimental data. The idea of a population of models (POMs) [10] is a relatively new method for capturing this important and intrinsic variability, and it involves the construction of a mathematical model with multiple points in parameter space rather than a single point, all of which are chosen to fit a given set of experimental/observational data. Following its first proposal in the field of neuroscience modelling, the POM method has found application in the field of cardiac electrophysiology [1, 15]. In this context, models are calibrated by extracting biomarkers from time course profiles, such as Action Potential Duration and beat-to-beat variability. To ensure variability estimates are within biological ranges for any model, the ranges of extremes for each biomarker seen in the experimental data are employed. If the data cannot be characterized by a collection of biomarkers, then time course profiles may be used to calibrate the population by comparing the data values to the simulation values at a series of time intervals using a normalized root-mean-square (NRMS) comparison. This method may point to a fresh way of doing scientific research. To begin, the approaches arrived at using the POM methodology are fundamentally probabilistic. Second, it prioritizes the feedback paradigm based on testing, modelling, and simulation [5].

Variability in the underlying structure may be represented by permitting changes in the parameter values via the use of experiments based on a population of models, as opposed to experiments based on a single model. Decisions on whether to employ "best" or "mean" data, as well as the challenges associated with selecting such data, are sidestepped by this method. A large number of sample parameter sets, drawn from a potentially high-dimensional parameter space, must be generated before a model population can be constructed. Recent developments in processing capacity have made it feasible to build vast numbers of such models, which in turn leads to a deeper understanding of the systems under study. There are a variety of approaches of sampling the parameter space, each of which is bound by different budgetary and computational considerations. When the number of samples is fixed and independent of the dimension of the space, random sampling, Latin Hypercube sampling (LHS), and Orthogonal sampling (OS) will give increasingly improved coverage of the parameter space compared to a parameter sweep.

POMs may be built in a variety of methods, each tailored to the underlying use case. POMs are created from LHS, for instance in [1], [15], which is helpful since it elucidates the source of variability in cardiac electrophysiology. In this scenario, unlike when POMs are used for parameter fitting, it is less of a concern whether or not all of the parameter space is covered. Similarities between POMs and ABC [7] may be seen in this context. As opposed to randomly sampling the

ISSN 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

www.ijasem.org

Vol 18, Issue 1, 2024

parameter space, ABC often performs adaptive sampling to converge on subregions of parameter space where the calibrated models sit. So, in certain cases, knowing how much of the parameter space k d-dimensional Latin Hypercube trials should cover is crucial.

The authors of article [4] set out to determine how well a 2-dimensional

parameter zone for k LH trials where each trial is n-in-size. In specifically, the estimated coverage of points in the parameter space after k trials was predicted using counting arguments. Numerical results from 100 simulations implemented in MATLAB were compared to these estimations. Based on the simulation findings, the authors hypothesized that with k trials, the probability of covering the whole 2-dimensional parameter space for values 1, 2, n was around 1 ek/n. One benefit of LHS is that it stratifies all univariate means concurrently, as stated by McKay, Beckman, and Conover [12]. Stratifying the bivariate margins is another something that has been recommended, most notably by Tang [14]. In an experiment, for instance, many factors may be considered, but usually just a few of them make a significant difference. To address this issue, one strategy involves mapping the components into a subspace occupied by the effective variables. However, this may cause the sample points to be duplicated in the effective subspace. As Tang points out, even in the case of bivariate margins, there is no assurance that this projection is uniformly distributed, despite the fact that Welch et al. [16] propose LHS as a way for screening for effective variables.

Instead, Tang [14] presents a method predicated on the presence of orthogonal arrays, which is a kind of orthogonal sampling. Further, he demonstrates that there is a method of reducing variation by using Orthogonal sampling, which leads to homogeneity on narrow dimensional margins. Tang proposes replacing the elements of an orthogonal array (specified in Section 2) with permutations at random to get an orthogonal sample. In Section 2, we'll provide further context for this approach while also outlining an alternative to Orthogonal sampling. The generation of interaction test suites for the testing of component-based systems has also made use of orthogonal arrays and covering arrays. For big systems, thorough testing may not be practical, thus special suites are created to check for t-way interactions (where t may range from two to six); for more information, see [3], [9]. For the purpose of verifying t > 2 interactions, Bryce and Colbourn provide a density-based greedy approach for generating covering arrays in [2] and [3]. We have looked at the connection between Experimental Design and POM construction based on our own work and that of [5].

Methodology:

building orthogonal sample sets We first formalize the definitions for Orthogonal samples and discuss the well-known techniques used to build Latin Hypercube samples before moving on to actual structures. To create a Latin Hypercube trial, a nd matrix is generated, with each column being a different random permutation of [1, 2, n], and each row being a dtuple. Therefore, given an experiment with d variables, where each parameter value can be one of 1, 2, n, a Latin Hypercube trial is a randomly generated subset of n points from a d-dimensional parameter space satisfying the condition that the projections onto each of the 1-dimensional subspaces are permutations. Below are two examples of Latin Hypercube trials with the parameters d = 3 and n = 8.

LHS1	LHS2	LHS3	OS LHS4
(1, 2, 1)	(1, 3, 2)	((1,1), (1,2), (1,1))	((1,1), (1,3), (1,2))
(2, 3, 3)	(2, 4, 6)	((1,2), (1,3), (1,3))	((1,2), (1,4), (2,2))
(3, 1, 2)	(3, 5, 3)	((1,3), (1,1), (1,2))	((1,3), (2,1), (1,3))
(4, 7, 8)	(4, 7, 8)	((1,4), (2,3), (2,4))	((1,4), (2,3), (2,4))
(5, 8, 5)	(5, 1, 1)	((2,1), (2,4), (2,1))	((2,1), (1,1), (1,1))
(6, 5, 4)	(6, 2, 7)	((2,2), (2,1), (1,4))	((2,2), (1,2), (2,3))
(7, 4, 6)	(7, 8, 4)	((2,3), (1,4), (2,2))	((2,3), (2,4), (1,4))
(8, 6, 7)	(8, 6, 5)	((2,4), (2,2), (2,3))	((2,4), (2,2), (2,1))

Analytical Simulation

In [4], we utilized MATLAB simulations to hypothesize that, for a given n, the number of trials of a Latin Hypercube of size k is proportional to the number of parameters in the system, with d = 2. Here we examine the subspace coverage for t = 2, t = 3, and t = 4 in the parameter space of d = 3, 4, and 5 dimensions. In Fig. (3), we provide the LHS outcomes for d = 3, for both 2-tuples and 3-tuples, at 25%, 50%, 75%, and 100% coverage. Figure (2) displays the LHS outcomes for d = 4, whereas Figure (3) displays the outcomes for d = 5. The graphs show the log10 of the data, which has been averaged across 200 iterations for each quantity. The number of trials needed to achieve a target percentage of coverage is consistent across systems of varying dimensions d, as shown in the 2-dimensional subspaces (t = 2) for d = 3, 4, and 5. For all values of d = 3, 4, and 5, the gradient is 1 at 25%, 50%, and 75% coverage, and about 1.25 at 100% coverage. Similar behaviour is seen for 3-

ISSN 2454-9940
INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
www.ijasem.org
Vol 18, Issue 1, 2024

dimensional subspaces (t = 3) for d = 3, 4, and 5, with the exception that the gradient is 2 for partial coverage and around 2.3 for total coverage. Coverage as a function of trial number (k) and division size (n) was proposed to be 1 ek/n in [4]. Here, we have findings for t = d = 3 and t = d = 4, which seem to indicate that the % coverage at t = d is provided by

$$P(k, n, t, t) = 1 - (1 - 1/n^{t-1})^k$$

Figure 1: Coverage for 2-tuples (left) and 3-tuples (right), for LHS, d = 3

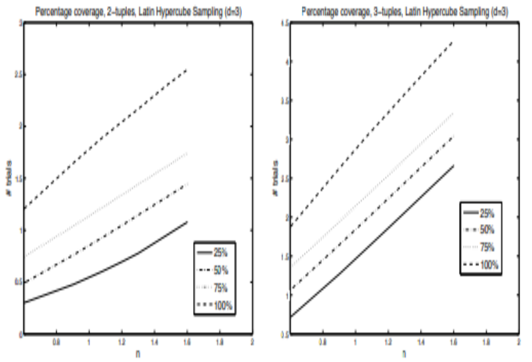
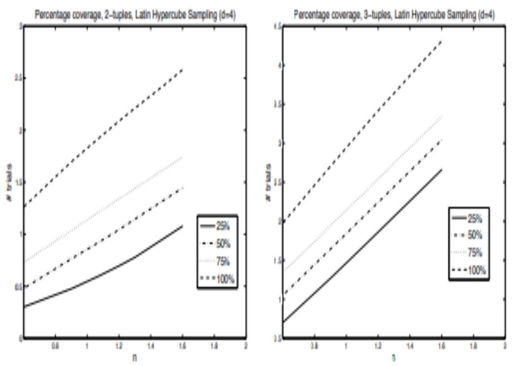


Figure 2: Coverage for 2-tuples (left) and 3-tuples (right), for LHS, d = 4



and, in the asymptotic limit as k becomes large, that it is given by

$$P(k, n, t, t) = 1 - e^{-k/n^{t-1}}$$

More generally we conjecture for any t

$$P(k, n, d, t) = 1 - (1 - 1/n^{t-1})^k$$

and, in the asymptotic limit as k becomes large, that

$$P(k, n, d, t) = 1 - e^{-k/n^{t-1}}.$$

Figure 3: Coverage for 2-tuples (left) and 3-tuples (right), for LHS, d = 5

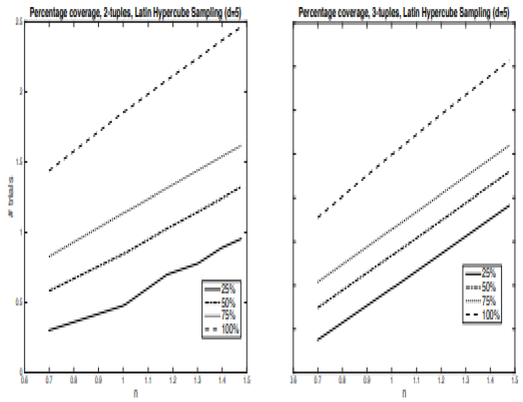
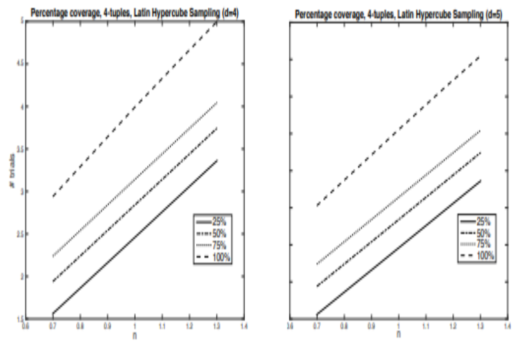


Figure 4: Coverage for 4-tuples for LHS, for d = 4 (left) and d = 5 (right)



This is consistent with the 25%, 50%, 75% coverage in which the gradient of the log data is t − 1. The only question to address is why the gradient is slightly larger than t − 1 for 100% coverage. To see this we see that 100% coverage implies P(k, n, d, t) > 1 − 1/nt−1. Thus, under our conjecture

$$1 - 1/n^{t-1} > 1 - (1 - 1/n^{t-1})^k$$

$$(1 - p)^k > p, \quad p = 1/n^{t-1}.$$

Using the fact that log(1 − p) ≈ −p for p small, then this implies

$$k \approx (t - 1)\log(n)n^{t-1}$$

it is this latter term that gives an apparent gradient slightly larger than t − 1. Thus, we make the following conjecture Conjecture: The coverage of a t dimensional subspace of a d dimensional

ISSN 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

www.ijasem.org

Vol 18, Issue 1, 2024

parameter space of size n when performing k trials of Latin Hypercube sampling is given by P (kn,d,t)

$=1- (1- 1/n^{t-1})$.

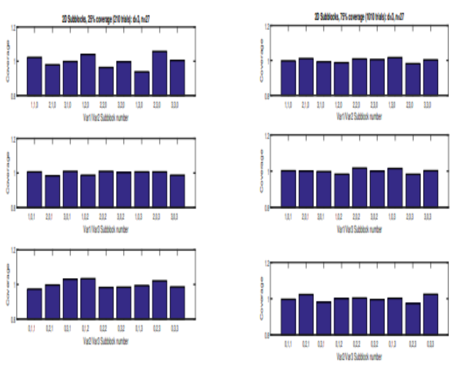
when k is large.

thus, if costs and/or experimental factors influence the size of the sample, we can use this information to direct our experiments. So, this builds confidence in the modelling results. For LHS, where d = 3 and n = 27, we investigate the variability (see Fig (5)) in coverage of the sub-blocks of the 2-dimensional spaces, and compare this with Orthogonal sampling where by design the coverage is uniform over the sub-blocks.

Conclusions

In this paper we have used simulations to give a conjecture about the coverage of a t dimensional subspace of a d dimensional parameter space of size n when performing k trials of Latin Hypercube sampling. This coverage takes the form P(k, n, d, t)=1 − (1 − 1/nt−1)k or 1 − e−k/nt−1 when k is large. This extends the work in [4]. We suggest that the coverage is independent of d and this allows us to make connections between building Populations of Models and Experimental Designs. We also show that Orthogonal sampling is superior to Latin Hypercube sampling in terms of giving a more uniform coverage of the t dimensional subspace at the sub-block size level when only attempting partial coverage of this subspace. We will attempt to prove our conjecture analytically in a subsequent paper. Finally, we note that the results described here have direct relevance to uncertainty quantification and sensitivity analyses in terms of the sampling techniques ([6]).

Figure 5: Sub-block coverage in each of the 2-dimensional subspaces for LHS with d=3, n=27, for trials giving 25% and 75% coverage.



References

[1] O.J. Britton, A. Bueno-Orovio, K. Van Ammel, H.R. Luc, R. Towart, D.J. Gallacher, and B. Rodriguez. Experimentally calibrated population of models predicts and explains inter subject variability in cardiac cellular electrophysiology. PNAS, 2014.

[2] R.C. Bryce and C.J. Colbourn. The density algorithm for pairwise interaction testing. Software Testing, Verification and Reliability, 17:159–182, 2007.

[3] R.C. Bryce and C.J. Colbourn. The density-based greedy algorithm for higher strength covering arrays. Software Testing, Verification and Reliability, 19:37–53, 2009.

[4] K. Burrage, P.M. Burrage, D. Donovan, T. McCourt, and H.B. Thompson. Estimates on the coverage of parameter space using populations of models. Modelling and Simulation, IASTED, ACTA Press, pages DOI: 10.2316/P.2014.813–013, 2014.

[5] A. Carusi, K. Burrage, and B. Rodriguez. Bridging Experiments, Models and Simulations: An Integrative Approach to Validation in Computational Cardiac Electrophysiology. Am. J. Physiology, 303(2):H144–55, 2012.

[6] S-K. Choi, R.V. Gandhi, R.A. Canfield, and C.L. Pettit. Polynomial Chaos expansion with Latin Hypercube sampling for estimating response variability. AIAA Journal, 42(6):1191–1198, 2004.

[7] C.C. Drovandi, A.N. Pettitt, and M.J. Faddy. Approximate Bayesian computation using indirect inference. Journal of the Royal Statistical Society: Series C (Applied Statistics), 60(3):317–337, 2011.

[8] J.C. Hilton and F.J. Davis. Latin Hypercube sampling and the propogation of uncertainty inanalyses of complex systems. Reliability Engineering and Systems Safety, 81:23–69, 2003.

[9] D.R. Kuhn, D.R. Wallace, and A.M. Gallo. Software fault interactions and implications for software testing. IEEE Transactions on Software Engineering, 30(6):418–421, 2004.

[10] E. Marder and A.L. Taylor. Multiple models to capture the variability of biological neurons and networks. Computation and Systems, Nature Neuroscience, 14(2):133–138, 2011.

[11] M.D. McKay. Latin Hypercube sampling as a tool in uncertainty analysis of computer models. Proceedings of the 1992 Winter Simulation Conference, ed. J.J. Swain, D. Goldsman, R.C. Crain, J.R. Wilson, pages 557–564, 1992.

[12] M.D. McKay, R.J. Beckman, and W.J. Conover. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics, 21(2):239–245, 1979.

[13] M. Stein. Large sample properties of simulations using Latin Hypercube sampling. Technometrics, 29(2):143–151, 1987.

[14] B. Tang. Orthogonal Array-Based Latin Hypercubes. Journal of the American Statistical Association, 88(424):1392–1397, 1993.

[15] J. Walmsley, J.F. Rodriguez, G.R. Mirams, and K. Burrage; I. R. Efimov; B. Rodriguez. MRNA expression levels in failing human hearts predict cellular electrophysiological remodelling: A population based simulation study. PLoS ONE, 8(2), 2013.

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 18, Issue 1, 2024

[16] W.J. Welch, R.J. Buck, J. Sacks, H.P. Wynn, T.J. Mitchell, and M.D. Morris. Screening, Predicting, and Computer Experiments. Technometrics, 34:15–25, 1992.