Latin Hyper-Rectangles for Integration in Computer Experiments

David Mease and Derek Bingham

Department of Statistics, University of Michigan, Ann Arbor

Introduction

The goal is evaluation of the integral

$$\mu = \mathrm{E}[g(X)] = \int g(x)f(x)dx$$

The density $f(x)$ is known.

The function $g(x)$ is the (deterministic) output from a complex computer code.

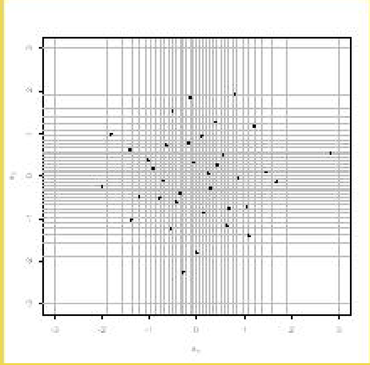
Evaluation of $g(x)$ requires a substantial amount of computation time, making straightforward Monte Carlo integration impractical.

A popular and effective alternative is Latin hypercube sampling (LHS).

Sub-Optimality of Latin Hypercubes

LHS partitions the space into cells of equal probability.

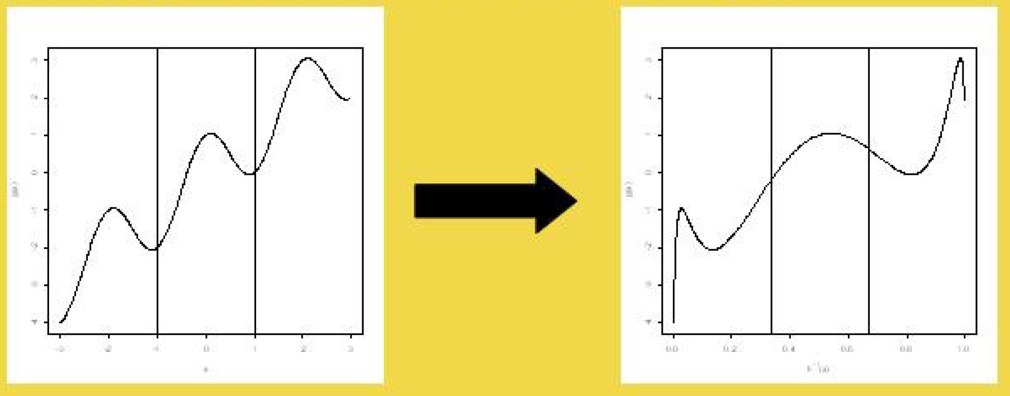
This tends to not fill the space well, but rather puts too many points where the density is large.



A LHS with 36 points for two independent standard normal random variables restricted to the interval [-3,3].

This has gone unnoticed in the literature since the problem of integrating an unknown function with respect to a known density is taken to be the same as the problem of simply integrating an unknown function.

While these two integration problems are mathematically equivalent by way of the inverse CDF mapping, this mapping increases the variability of the original function of interest in places where the density is large.



Solution

We will use unequal cell probabilities which results in hyper-rectangles rather than hypercubes in the probability space.

The estimator must be appropriately weighted to account for the unequal probabilities:

$$\hat{\mu} = \sum_{i=1}^{n} g(X_i)p_i$$

Here, X_i is the random vector X conditioned on the ith cell and p_i is the corresponding probability (weight).

Cell partitions are chosen to minimize

$$\mathrm{Var}[\hat{\mu}] = \sum_{i=1}^{n} \mathrm{Var}[g(X_i)]p_i^2$$

This depends on $\mathrm{Var}[g(X_i)]$ which is unknown.

<u>**Theorem:**</u> In one dimension, if $g(x)$ is Brownian Motion we have

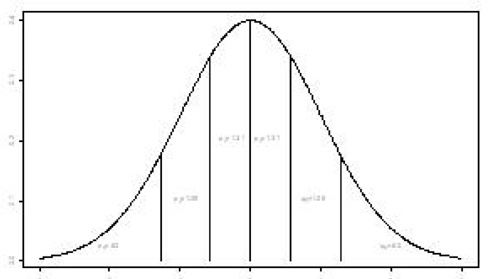
$$\mathrm{E}[\mathrm{Var}[g(X_i)]] = \sigma^2 k_i + \mu^2 \mathrm{Var}(X_i)$$

where μ and σ^2 are the drift and variance parameters respectively.

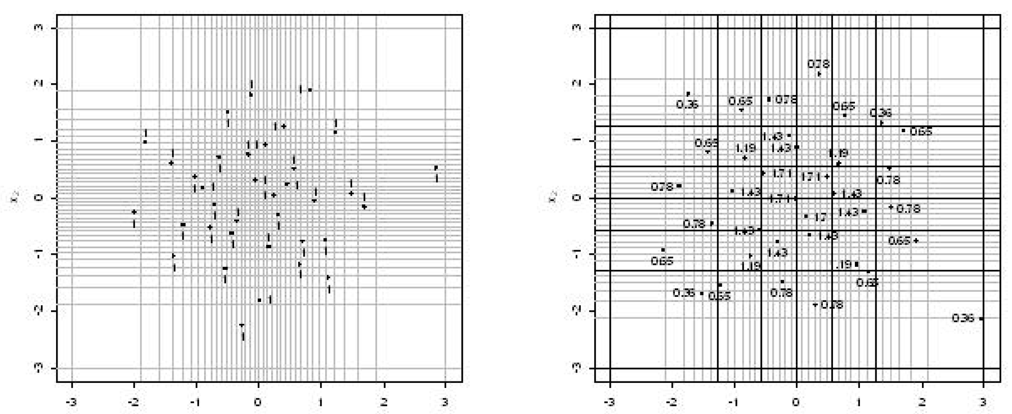
Assuming the drift dominates we minimize $\sum_{i=1}^{n} \mathrm{Var}[X_i]p_i^2$

Optimal Partitions

For n=6 the cell partitions and weights are shown below for the standard normal density restricted to [-3,3]:



In dimensions higher than one, analogous partitioning can be used based on stratified Latin Hypercube sampling.



Weights for the equal probability cells of LHS (left panel) and the weights for optimal cells (right panel) with n=36 for two independent standard normal random variables restricted to the interval [-3,3].

Results

The following table gives the resulting variance of the proposed estimator for various test functions in two dimensions.

These are based on n=36 points.

The distribution is taken to be that of two independent normal random variables restricted to [-3,3].

Comparisons are made to LHS and stratified LHS (SLHS).

Function	LHS Variance	SLHS Variance	Variance of Proposed Estimator
x_1	.00014	.00014	.00021
x_1^2	.00280	.00280	.00180
x_1^4	.34060	.34060	.17190
e^{x_1}	.00670	.00670	.00440
$(x_1+x_2)^2$	.10826	.02340	.01320
x_1x_2	.02917	.00456	.00226
$e^{(x_1+x_2)}$	.45512	.31943	.06623
$x_1/(x_1+x_2+10)$	5.8500×10^{-6}	3.2641×10^{-6}	2.6906×10^{-6}
$\cos(x+y)$	.00639	.00096	.00115

The proposed method generally gives smaller variance.

While the proposed method may give slightly larger variance for some functions, it gives substantially smaller variance for many others.

The method generally gives smaller variance even for additive functions.

All other current methods for improving LHS do not give any improvement for additive functions.

Future Work

Orthogonal array based Latin hypercube designs are generally preferred to stratified Latin hypercube designs.

While the proposed methodology can often be used in conjunction with orthogonal array based designs, in many cases there are problems with randomness in the weights.

Future research is needed to extend the proposed methodology to orthogonal array based Latin hypercube sampling.

Conclusions

Current Latin hypercube sampling methods are based on partitioning the space into cells of equal probability, which can be shown to be sub-optimal.

The proposed methodology improves these current techniques by partitioning the space in an optimal manner using knowledge of the distribution.

Simulation results demonstrate that this optimal partitioning leads to substantial variance reduction for "most" functions.

Acknowledgements

This research was supported by NSF grant DMS-0103886.

The authors are grateful to the DAE 2003 organizers for their support.