Uses and caveats of Latin Hypecube Sampling in the parameter space exploration of ecological models

Authors	Andre Chalom
	Paulo Inacio Prado
Date	03/16/2012
Affiliation	Department of Ecology,
	University of Sao Paulo

Abstract

In recent years, both the scientific community and policy makers are gaining confidence in the formulation and use of mathematical models in ecological studies. We are thus seeing a flourishing of more elaborate and complex models, and the questions related to the efficient, systematic and error-proof exploration of parameter spaces are of great importance to better understand, estimate confidences and make use of the output from these models. In this work, we investigate some of the relevant questions related to parameter space exploration, in particular using the technique known as Latin Hypercube Sampling, and assess how are these questions being currently addressed in the literature.

1 Introduction

1.1 Parameter spaces

In order to better pose our questions, we need first to have a precise definition of what is the parameter space of our models, and to distinguish the general use of this expression by the statistical and the modeling community. ** ROSS, 1990 **

Also, it should be mentioned that the parameter space may be constrained (for example, $a_1 + a_2 < 90$ degress) or unconstrained.

...

1.2 Problems in PS exploration

Next, we turn our attention to the kind of problems we might want to address with the exploration of the parameter space. First, the simplest case is asking "is there a region of my parameter space where condition X holds?" This condition might be, for example, the extinction or coexistance of species, some pattern of distribution or abundance of species. We also might be interested in mapping where are these regions. In complex models, where several different regions might exist where the qualitative results of the models are very different, we may ask how many of these regions are there, as well as map the frontiers between them.

Another class of problems arises when the model produces some quantitative response, and we are interessed in determining the dependency of this response to the input parameters. For example, in models where limit cycles arise, we would like to know how the period of the limit cycle varies with each of the parameters. Along with these questions, it may be extremely important to assess the sensitivity of the model to small fluctuations in the parameters. ** KUEHN 2010 ** ** HELTON 2003 **

All these problems may be formulated in a general way, defining some response from the model $\mathbf{Y}$ as a function of the parameter vector $\mathbf{x}$:

$$\mathbf{Y} = \mathbf{f}(\mathbf{x}) \tag{1}$$

In the equation 1, all the quantities are vectors, indicated by the boldface. Here, $\mathbf{x} = [x_1, x_2, \ldots, x_n]$ represent the parameters to the model $\mathbf{f}$, and $\mathbf{Y} = [y_1, y_2, \ldots, y_n]$ represent the some quantitative response from the model.

...

1.3 Stochastic models

When dealing with stochastic models, like several relevant individual based models (IBM), the questions presented become complicated by the fact that running the same model with exactly the same parameters might wield largely different results, both quantitative and qualitatively. In this scenario,

- Is is possible to formulate the stochastic and deterministic problems in a general way?
- What are the conditions for a stochastic model to be subject to parameter space exploration?
- Are there techniques that can be applied to the stochastic models to make them more amenable to PS exploration?

We also give an overview of Marko Chain Monte Carlo methods, and their relation with LHS.

...

2 Latin Hypercube Sampling

2.1 Definition and use

In this section, we describe the Latin Hypercube Sampling, and show how it can be used to efficiently solve the questions posed in section 1. We briefly compare it to some other techniques, like random sampling. We also discuss what are the available implementations of LHS.

Firstly, let us define, in the context of statistical sampling, what is a Latin Square:

Definition If we divide each side in a square in M intervals, and then take samples from the square, the resulting square will be called Latin if and only if there is exactly one sample in each row and each column.

From this definition, it is clear that the number of samples is fixed as M. It should be noted that, in the mathematical literature, it is usual to refer to a somewhat different object as a Latin Square: this would be a square whose sides are divided in M intervals, and is filled with M different symbols, such that for each row and column there is exactly one occurrence of each symbol. The figure 1 shows an example of one "full" Latin Square. If we restrict our attention to just one color, we can visualize a Latin Square in the statistical sense.

A Latin Hypercube is the generalization of the Latin Square to an arbitrary number of dimensions n.

The first step to construct the Latin Hypercube is to divide each parameter dimension in M equally probable intervals. Thus, it is necessary to assign some probability function to each parameter dimension. We will briefly discuss how this choice must be made and how it might affect the result of the sampling.

...

2.2 Algorithms and extensions

As described above, the LH sampling generates an uniform distribution of samples in each parametric dimension. However, there is no guarantee that the correlation between two or more parameters will be zero, and the classical algorithm from McKay [**?**] usually produces correlations as high as 0.3 between pairs of factors, which can difficult or even compromise further analyses. In this section, we will present and compare some algorithms designed to minimize pairwise as well as higher order correlations, or even to produce samples with zero correlation terms.

Figure 1: A stained glass window at the Caius College, Cambridge, showing a full Latin Square. Notice how there is only one occurence of each color in each row and in each column.

** YE ** ** STEINBERG **

...

2.3 Adaptative refinement

Is it possible to adaptatively refine the grid in order to increase the number of points sampled, or if we decide to run the analysis with double the points I have to start from scratch?

...

2.4 Use of Latin Hypercube in ecological studies

Here we present some relevant papers in the ecological literature that made use of LH sampling or similar parameter space exploration techniques. We also try to summarize what are the prerequites that a model must fulfill in order to use it, what are the cautions that must be taken in running the sampling, what are the potential pitfalls interpreting the results.

...

3 Making use of the results

After applying the Latin Hypercube Sampling to an ecological model and gathering the results, what are the next steps? Here, we highlight some statistical analysis (like ANOVA) and numerical interpolation methods that can be used to make sense of the model output.