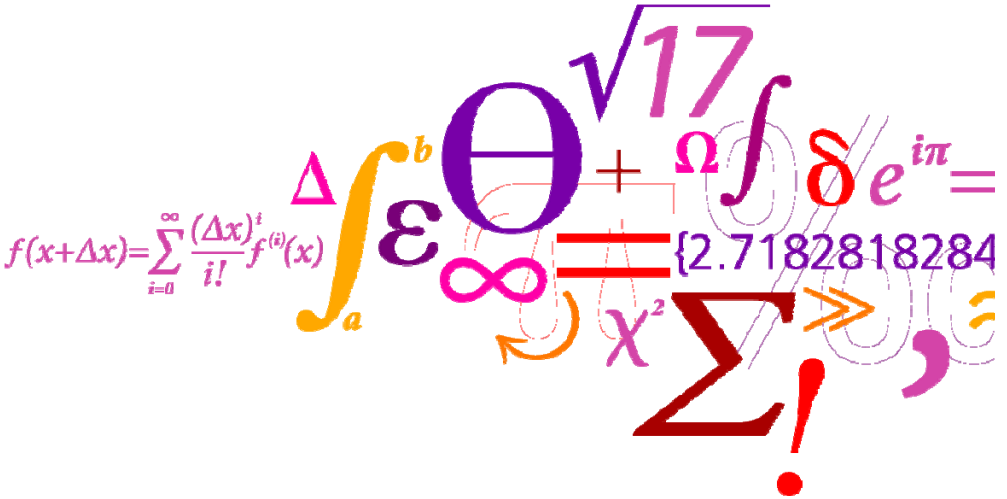
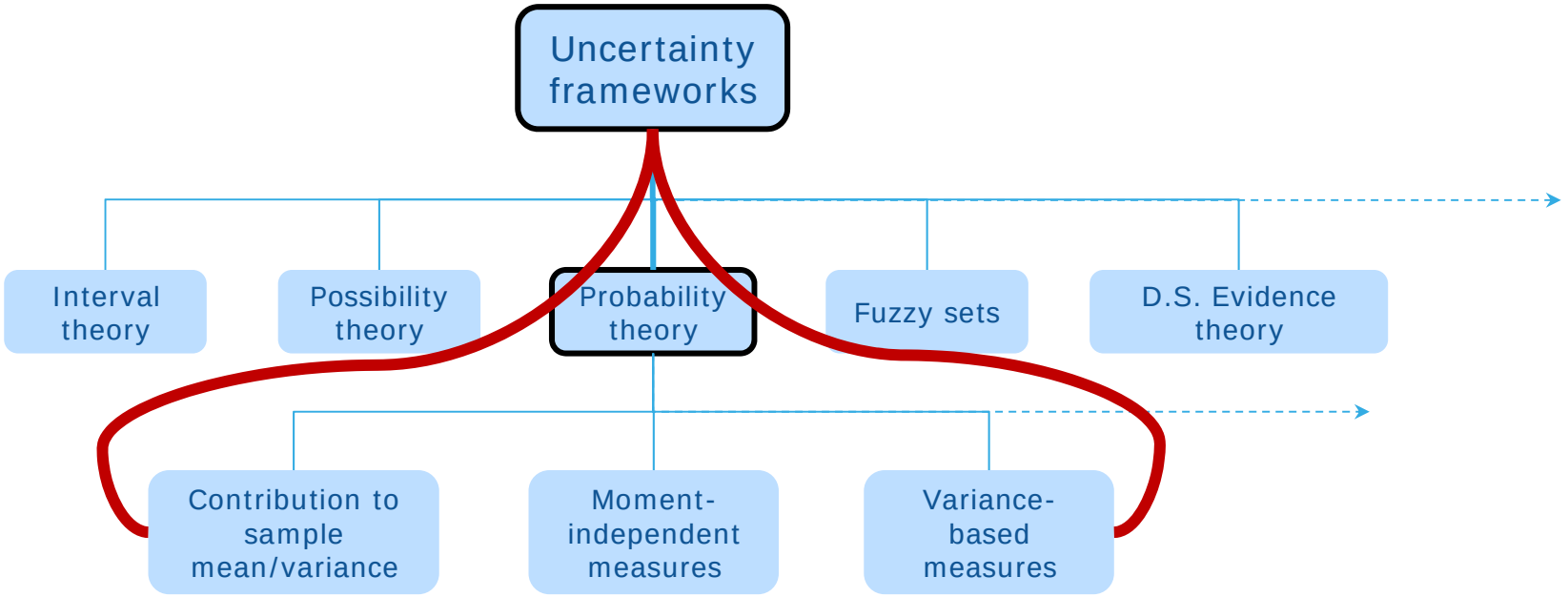
Sampling strategies for uncertainty & sensitivity analysis

Gürkan Sin, Associate Professor
PROSYS research center
DTU Chemical Engineering
gsi@kt.dtu.dk



DTU Kemiteknik
Institut for Kemiteknik

Quantifying uncertainty: context



Monte Carlo Method – definition

General definition

"Monte Carlo Method" refers in general to solution techniques that uses random numbers and probability statistics to investigate problems (and hopefully come up with approximate solutions.)

Statistical definition (Halton, 1970)

Representing the solution of a problem as a hypothetical population, and using a random sequence of numbers to construct a sample of the population, from which statistical estimates of the parameter is possible!

RANDOM NUMBERS

Monte carlo methods – used for any purpose – relies on random numbers

It is important and matters how these numbers generated. Many methods developed.

The first step is to generate random numbers. Many algorithms exist (in Matlab etc etc)

Second is to do improvements to optimize a certain criteria: e.g. improve uniform coverage of input space or minimize discprepancies

RANDOM NUMBERS

The following represents relevant methods/class of sampling:

1) Random sampling – relies on the algorithm implemented in your computer. That the sequence generated can not be predicted by any model better than a chance.

2) Stratified sampling: Latin Hypercupe Sampling (LHS)

3) Low discrepancy sampling (Quasi-random sampling): Halton sequence, Sobol sequence, lattices, etc.

Examples of sampling strategies

EXAMPLES OF SAMPLING STRATEGIES

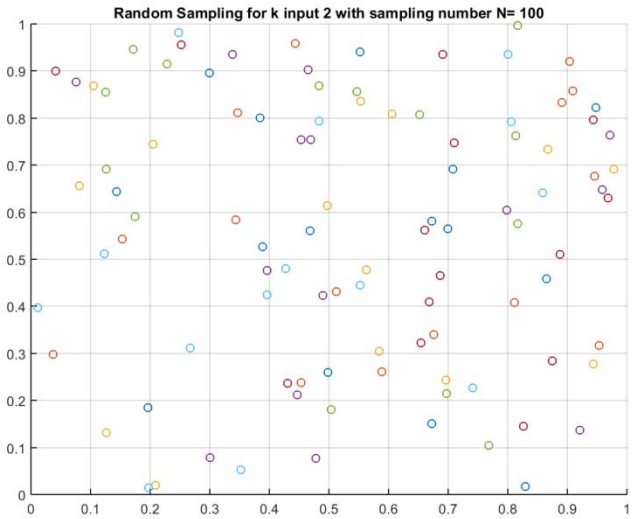
Problem statement

- In this example, we wish to compare three different strategies for generating random samples
- You can easily change sample number N in the matlab script and see the effect for yourself.

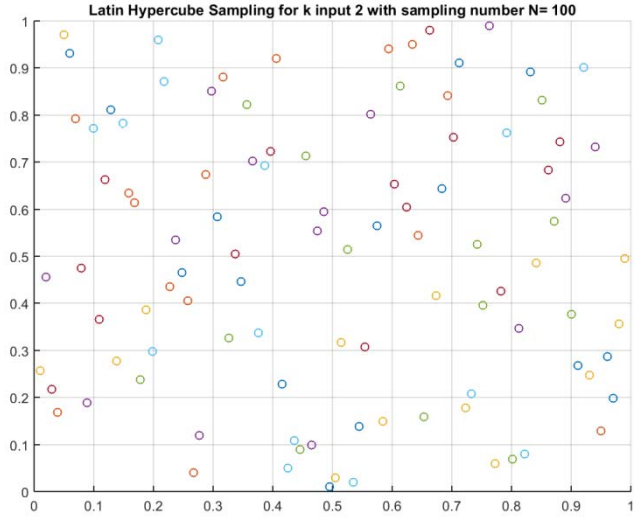
ComparisonOfSamplingMethods.m

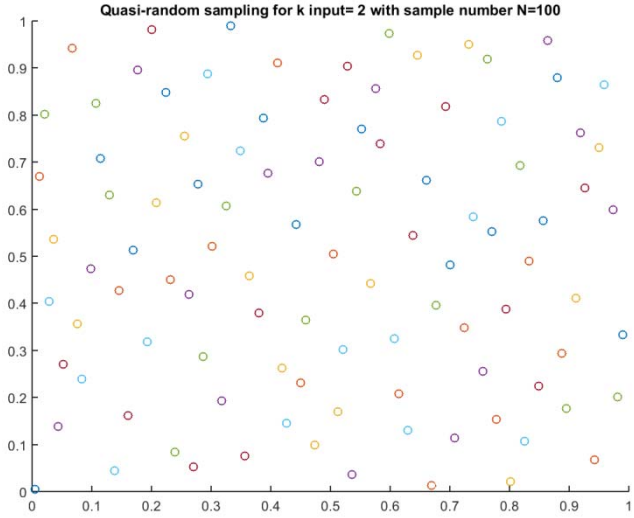
Random sampling

Not represenative coverage



Better coverage





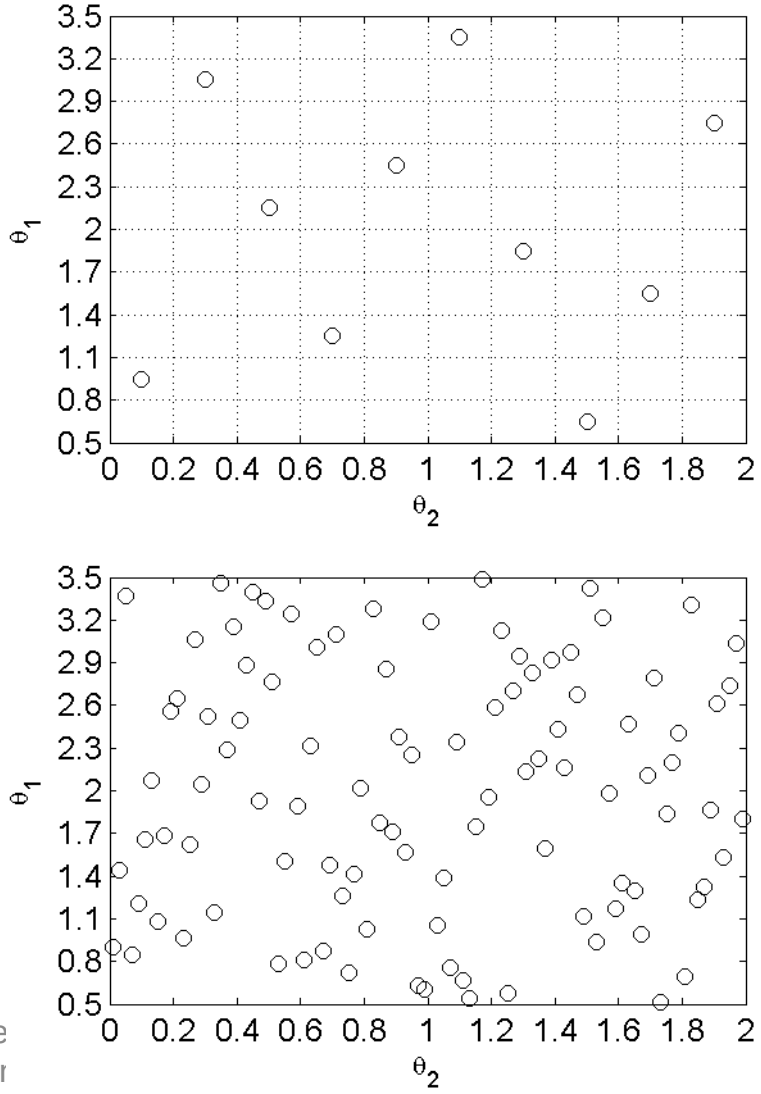
Much Better
Very low discrepancy!

More on Latin Hypercube Sampling

Assume a model with 2 parameters. Each has uniform distributions with upper and lower bounds [0.5 3.5] and [0.1 2] respectively.

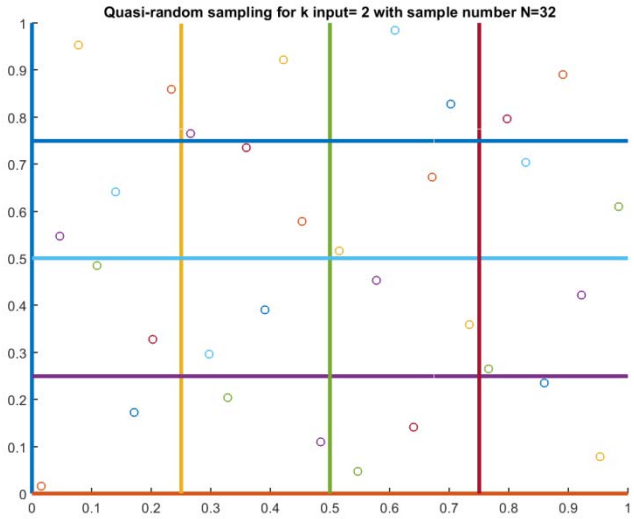
To draw a certain number of samples, say *n*, from the 2X2 parameter space formed by θ_1 and θ_2 :

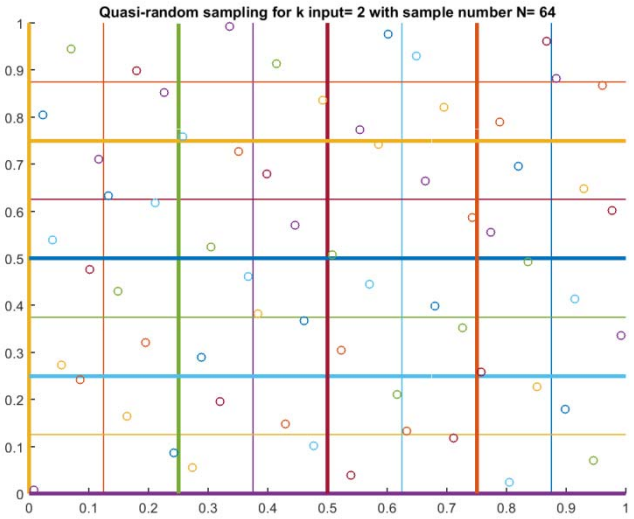
1. The range of each parameter is divided into *n* intervals of equal probability (1/n). Then select one value at random in each interval.
2. The LHS is then completed by randomly pairing the values of θ_1 and θ_2 leading to *n* couples.



More on Quasi random sampling

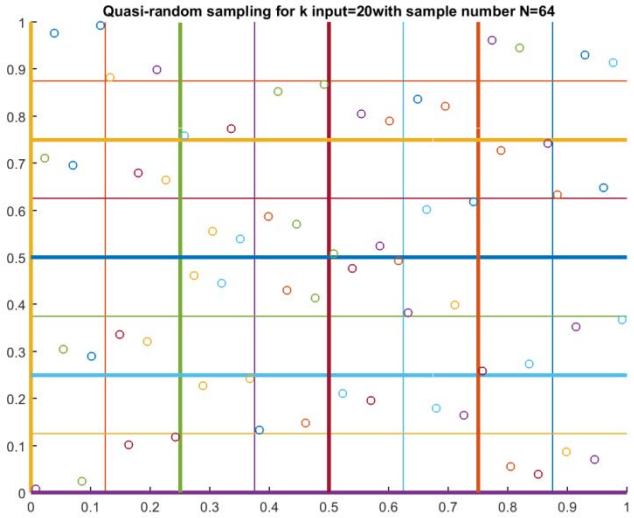
Therese are many sequences that generates quasi random sampling. We focus on Sobol sequences. These sequences use a base of two to form successively finer uniform partitions of the unit interval. Let us some examples for k=2 with N=32, 62 and 128.





Drawbacks of quasi random sampling

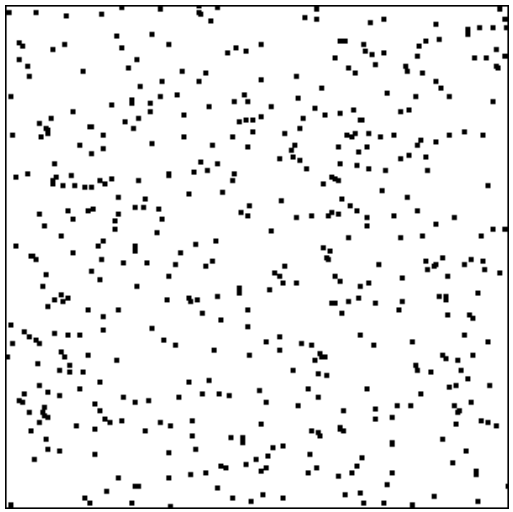
For large k's N has to be very large otherwise the discrepancy is very high. See below.



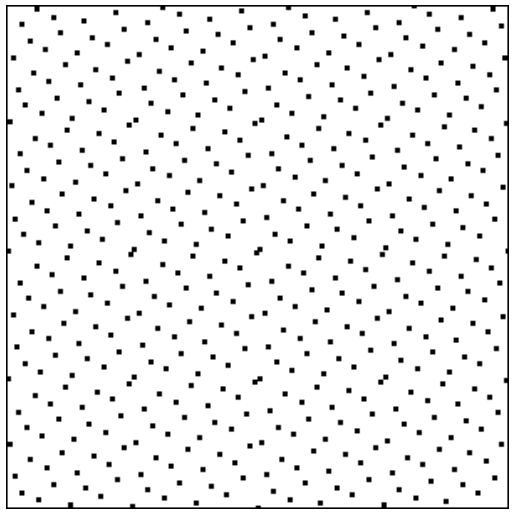
Other quasi random using hammersley

There are many more sampling techniques...

Random sampling (n=500)



Hammersley sampling (n=500)



Bottom line: obtain representative samples to cover parameter space !

For most engineering applications, this is not a critical step.

Double check the reproducibility of the results using different sampling techniques or repeat the same technique.

Sampling for dependent/correlated inputs

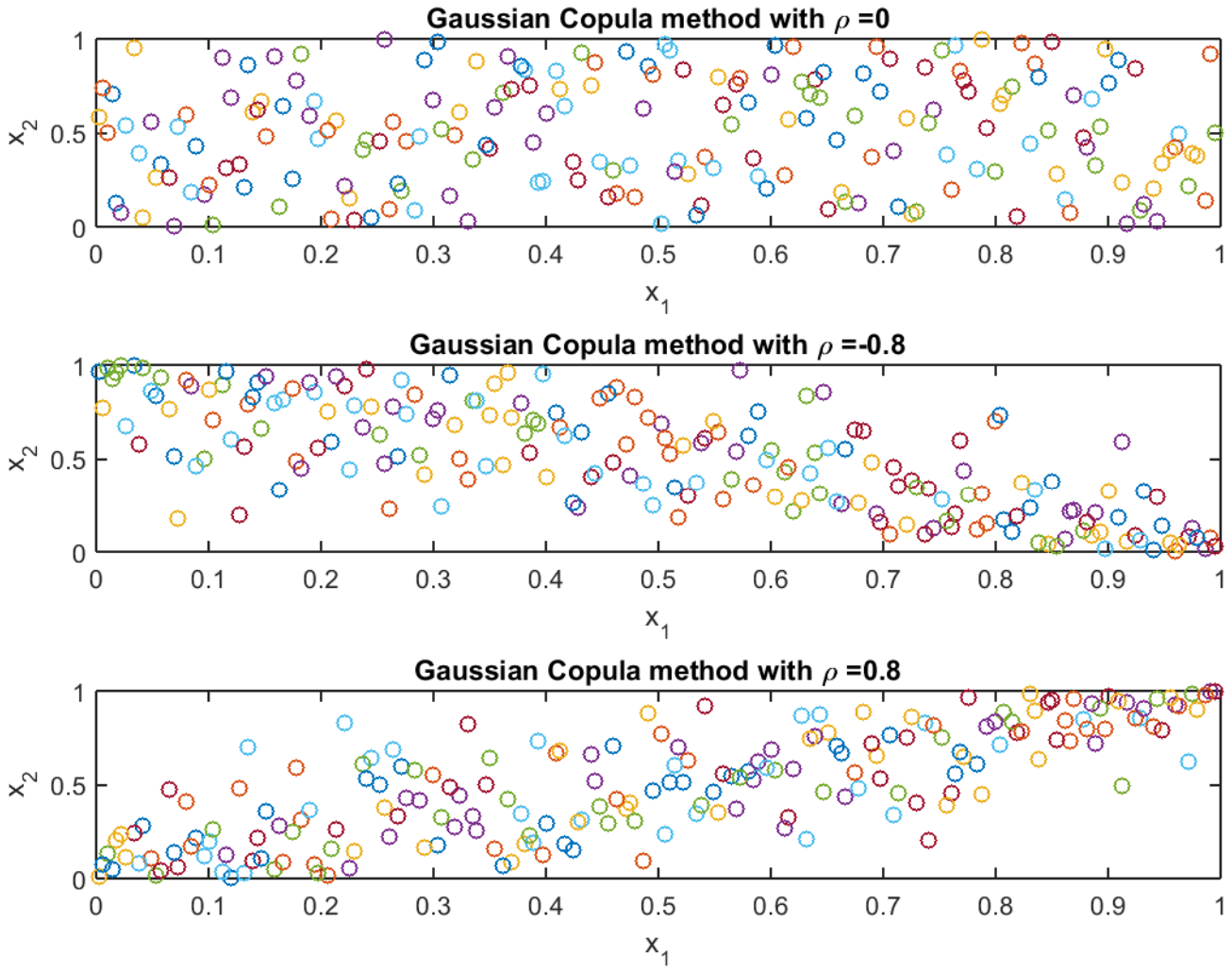
Sometimes (actually most of the time), parameters/inputs are correlated or dependent (depending lexicon used in your community). When sampling correlated parameters, one needs to preserve the original correlation structure.

Several methods exist to generate dependent samples:

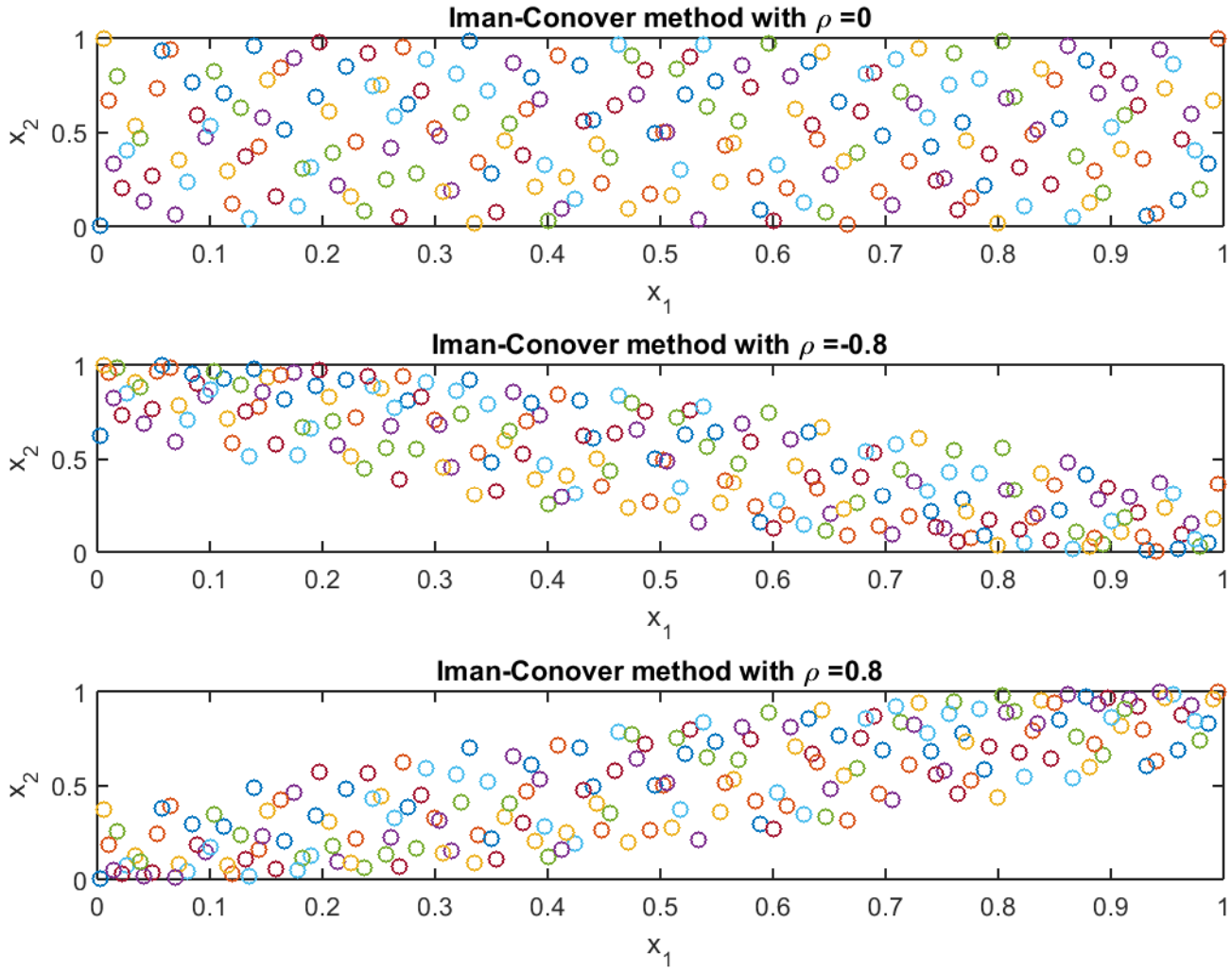
1) Copula family: e.g. Gaussian copula, t-copula, etc
2) Iman – Conover rank correlation method

Important is that linear correlation coefficient are preserved in the generated samples.

Strategies to generate dependent samples



Strategies to generate dependent samples



Many sampling techniques available in Matlab

- Halton sequence, sobol sequence, lhsdesign, rand, etc…
- Generation of dependent samples using Gaussian copula, Iman Conover method, etc
- They are the basis of sampling (Monte Carlo) techniques for UA/SA
- We will explore them in the next lectures.