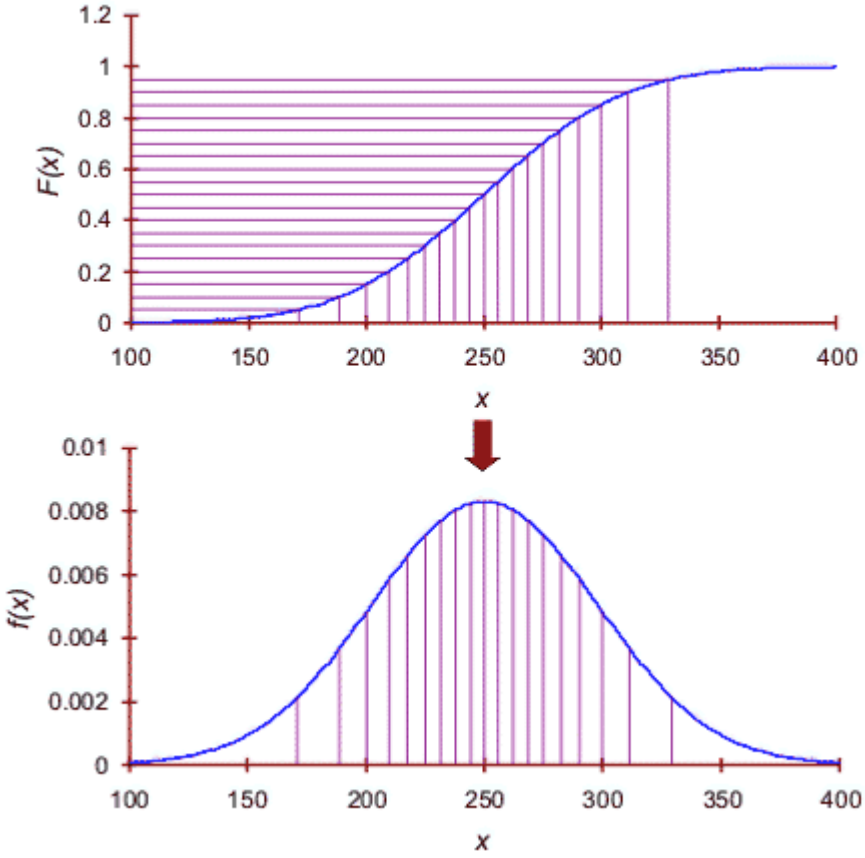
Latin Hypercube sampling

Latin Hypercube sampling, or LHS, is an option that is now available for most risk analysis simulation software programs. In fact, we would say that it is one of the features that is essential in any risk analysis software package. It uses a technique known as "stratified sampling without replacement' (Iman et al., 1980) and proceeds as follows:

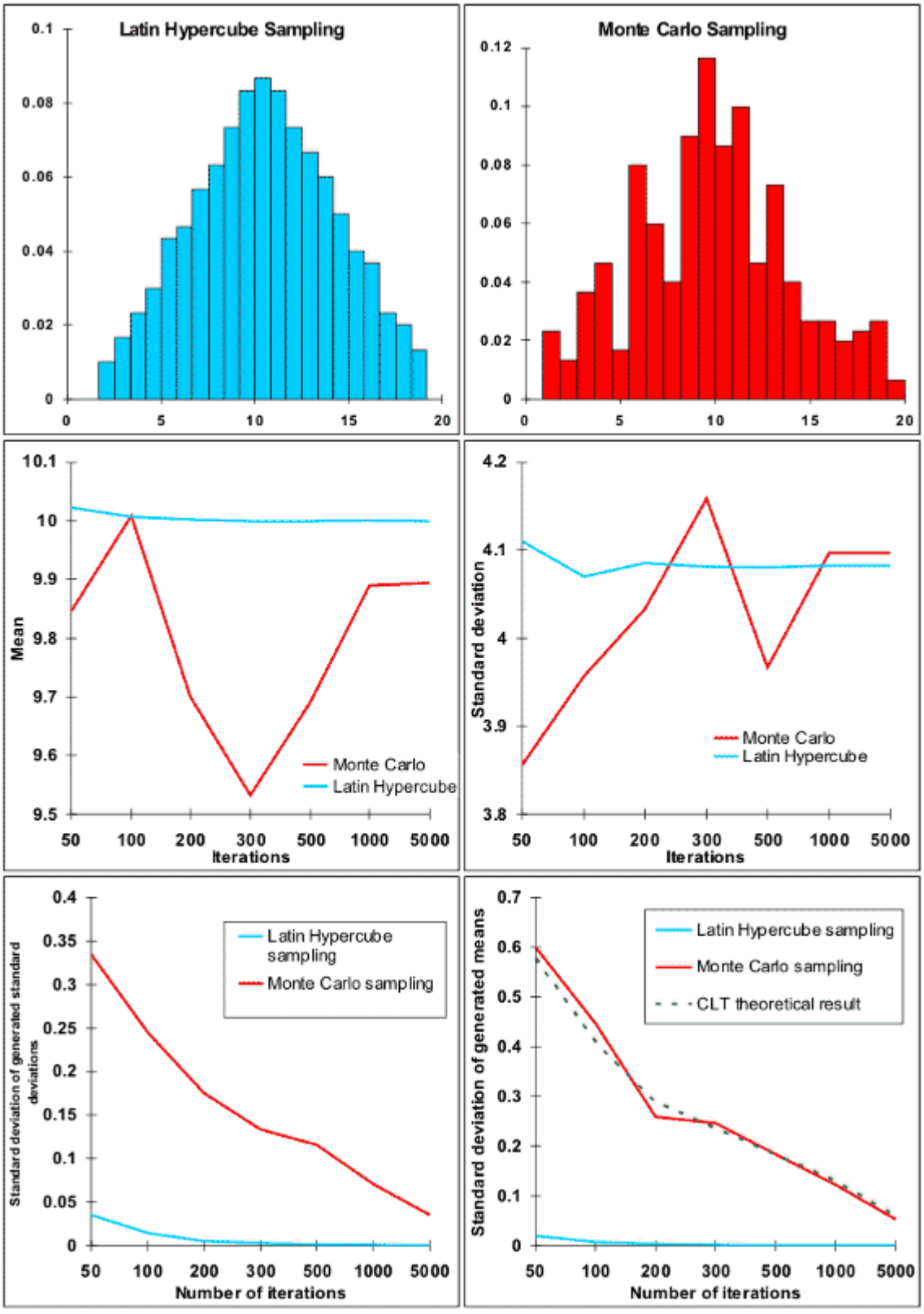
- The probability distribution is split into *n* intervals of equal probability, where *n* is the number of iterations that are to be performed on the model. The Figure 1 below illustrates an example of the stratification that is produced for 20 iterations of a Normal distribution. The bands can be seen to get progressively wider towards the tails as the probability density drops away.



- In the first iteration, one of these intervals is selected using a random number.
- A second random number is then generated to determine where, within that interval, *F(x)* should lie. In practice, the second half of the first random number can be used for this purpose, reducing simulation time.
- *x = G(F(x))* is calculated for that value of *F(x)*.
- The process is repeated for the second iteration but the interval used in the first iteration is marked as having already been used and therefore will not be selected again.
- This process is repeated for all of the iterations. Since the number of iterations *n* is also the number of intervals, each interval will only have been sampled once and the distribution will have been reproduced with predictable uniformity over the *F(x)* range.

The improvement offered by LHS over Monte Carlo can be easily demonstrated.

The Figure 2 below compares the results obtained by sampling from a Triangular(0,10,20) distribution with LHS and Monte Carlo sampling.



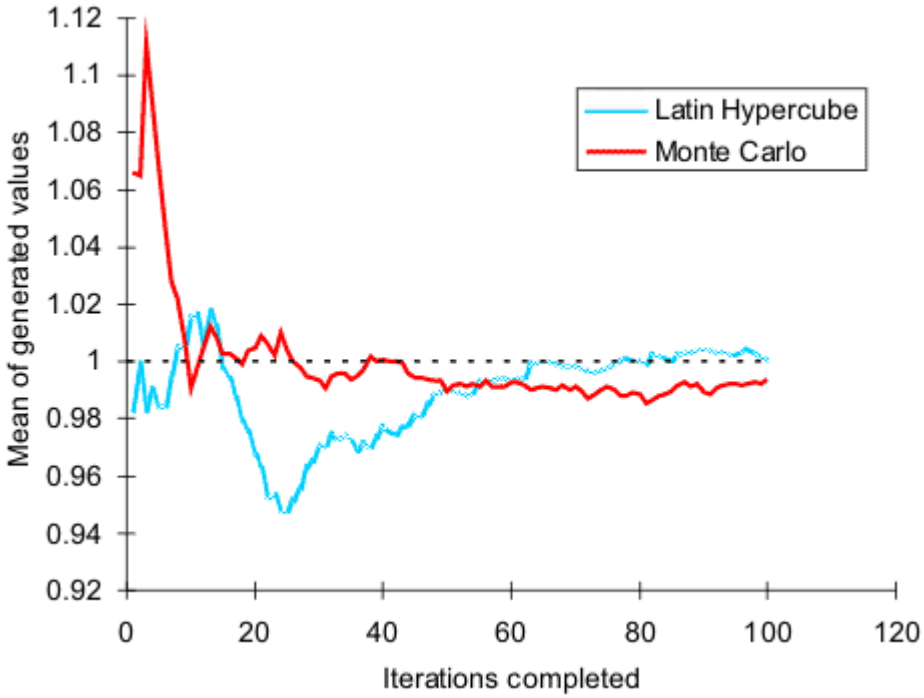
The top panels of Figure 2 show histograms of the Triangular distribution after one simulation of 300 iterations: the LHS clearly reproduces the distribution much better.

The middle panels of Figure 2 show an example of the convergence of the two sampling techniques to the true values of the distribution's mean and standard deviation. In the Monte Carlo test, the distribution was sampled 50 times, then another 50 to make 100, then another 100 to make 200, and so on to give simulations of 50, 100, 200, 300, 500, 1000, 5000 iterations. In the LHS test, seven different simulations were run for the seven different numbers of iterations. The difference between the approaches was taken because the LHS has an "memory" and the Monte Carlo sampling does not. An "memory" is where the sampling algorithm takes account of where it has already sampled from in the distribution. From these two panels, one can get the feel for the consistency provided by LHS.

The bottom two panels provide a more general picture. To produce these diagrams, the Triangular distribution was sampled in seven separate simulations again with the following number of iterations: 50, 100, 200, 300, 500, 1000, 5000 for both LHS and Monte Carlo sampling. This was repeated 100 times and the mean and standard deviation of the results were noted. The standard deviation of these statistics were calculated to give a feel for how much the results might naturally vary from one simulation to another. LHS consistently produces values for the distribution's statistics that are nearer the theoretical values of the input distribution than Monte Carlo sampling. In fact, one can see that the spread in results using just 100 LHS samples is smaller than the spread using 5000 MC samples!

Important: The benefit of LHS is eroded if one does not complete the number of iterations nominated at the beginning, i.e. if one halts the program in mid-simulation.

The Figure 3 below illustrates an example where a Normal(1,0.1) distribution is simulated for 100 iterations with both Monte Carlo sampling and LHS.



The mean of the values generated has roughly the same degree of variance from the true mean of 1 until the number of iterations completed gets close to the prescribed 100, when LHS pulls more sharply in to the desired value.