Simulation
Sources of Random Numbers

• ‘Man Made’ Systems
  – Dice
  – Cards
  – Drawing Straws

• Natural Systems
  – Noise imposed on a signal
  – Movement of particles in a gas
  – Vibration of molecules in semiconductors

• ‘Random’ Numbers from a Computer
‘Pseudo’ Random

• A collection of numbers that is not random, but for all practical purposes ‘looks’ random is said to be a collection of ‘pseudo’ random deviates.
• There are a number of methods of generating sequences of ‘pseudo’ random numbers.
• We will look at one.
Congruential Generators

• These are defined by the relationship;

\[ X_{n+1} = (a \cdot X_n + c) \mod m \]

• The output from these is a deterministic mapping into the integers 0,\ldots, m-1.

• To convert this to a generator on \([0,1)\),
  divide by \(m\).

• However, note that granularity is retained.
Considerations

• Correlation between successive numbers.
• ‘Filling’ of the space (particularly k-space).
  – Distribution in $[0,1]^k$.
  – Falling on ‘lines’ in 2 dimensions.
  – Vacant ‘chunks’ of k-space.
• Finite aspect - number of possibilities.
Some implementations

<table>
<thead>
<tr>
<th>a</th>
<th>c</th>
<th>m</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>0</td>
<td>13</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>16</td>
</tr>
<tr>
<td>69069</td>
<td>1</td>
<td>$2^{32}-1$</td>
</tr>
</tbody>
</table>
Other Distributions

• It is a fact that a cdf will allow conversion from $[0,1)$ to the distribution of interest.
  – This is called Inverse Transform.
• Approximate methods can often be used to generate from some distributions.
• Exact methods can be more ‘expensive.’
Inverse Transform

• \( G(u) = \min\{x|F(x)\geq u\} \)
• This defines a mapping from \([0,1)\) to \(X\), such that if \(u \sim U[0,1)\) then \(x \sim F(X)\).
• Examine this map for the case of:
  – Continuous distributions
  – Discrete distributions
  – Mixed distributions
Approximate Methods

• \( X = U_1 + \ldots + U_{12} - 6 \)

• Gives an approximation to the normal distribution using a combination of uniforms.

• Method is ‘cheap’ and reasonable, but not exact.
Exact for Normal

• Consider the bivariate standard normal.
• The joint distribution is a smooth hill centred at the origin.
• The following generates two normal deviates;
  – Generate Theta = 2\pi U_1
  – Then let E = -\ln(U_2) and R = \sqrt{2E}
  – Let X = R\cos(\Theta) and Y = R\sin(\Theta)
  – X, Y are a pair of independent Standard Normals.
• This is Box-Muller algorithm.
Rejection

• Given a ‘general’ pdf, f.
• And an ‘easy to sample from’ pdf, g, such that Mg() bounds f().
• Then the following generates samples from f;
  – Generate Y from g.
  – Generate U.
  – When MU<=f(Y)/g(Y) return Y.
+ves and –ves Rejection

• General.
• Is simple to program.
• Can be applied without too much ‘intricate’ knowledge.
• *Can be inefficient.*
• *Bounding constant, M, can be critical.*
• *Ensuring bounding needs to be done with care.*
Composition

• Where \( f \) is a combination of other distributions, that is \( f = \sum \phi_h \), then following will generate from the required distn;

• First select the part of the mixture, by generating \( U \). Comparison of \( u \) with the \( \psi \)s will give the particular \( h \) to sample from.

• Then generate from the conditional, that is \( h \).
Example / Exercises

• Construct a rejection algorithm to sample from the triangle (a, b, c)
• Explicitly write down the triangle as a mixture.
• Construct an inverse transform in order to simulate from the triangle.
Higher Dimensions

• So far we have only seen single dimensions.
• It is possible to simulate for higher dimensional situations.
• One way to do this is to use the idea of conditional distributions.
• Simulate $x$ from the marginal distribution $f(x)$.
• Then simulate $y$ from $f(y|x)$.
• If $f(y|x)=f(y)$ for all $x$, this is particularly simple.
Estimates based on Simulation

• The idea in all of this is to estimate the properties of some function of the random variable of interest.
• Thus we may wish to know, for example, the average ‘profit’, which is a function of order size.
• Let Profit = Profit(OrderSz)
• Let \( f(OrderSz) = \text{Ga}(a,b) \).
Estimates Contd

• Then Profit is a random variable (since it can be considered to be a transformation of Order Size which has density $Ga()$).

• The ‘average’ profit is the expected value of Profit;

• The expectation can be written as an integral of $\text{Profit(OrderSz)} \ast f(\text{OrderSz})$. 
Estimates Contd^2

• This integral can be estimated by a finite sum.
• One way to do the integration would be to choose a ‘grid’ along the density of OrderSz.
• Another method, would be to simulate from the density for OrderSz, and estimate the integral using the values of the curve at the sampled values.
• This latter method is called a Monte-Carlo method.
Monte Carlo Method

• A Monte Carlo method is a method which uses random numbers to solve a deterministic problem.
• In the case outlined, the deterministic problem was to carry out an integral, and the method was to approximate the integral by a finite sum on simulated values.
• The error in the point estimates due to the simulation is termed Monte-Carlo error.
Monte Carlo Error

• In this case, the simulated realisations are uncorrelated, so we can use simple sampling theory to come up with estimates of our Monte Carlo error.

• Sampling theory tells us that the estimate $\hat{I}$ comes from a distribution which is approximately Normal and has mean $I$, with standard deviation $\sigma/\sqrt{n}$, where $\sigma$ is the standard deviation for Profit.
An interval for I

• Thus an interval estimate for expected profit may be constructed.
• Sigma is unknown and so can be estimated from the data. We denote by s, the sample estimate of sigma.
• The interval has width which depends on a parameter, t, which is typically about 2 for a 95%CI when n is large.
• Interval is \([I-t*s/sqrt(n),I+t*s/sqrt(n)]\)
Notes

• This is an interval for **expected** Profit. We may be interested in the probability that profit is less than zero, for example, which is **not** given by this.

• The interval closes as \( n \) increases.

• This interval requires independent realisations from our simulator.

• In general, we may be interested in correlated samples too.
Summary

• Pseudo random numbers are a deterministic sequence. You should know this, and what to look out for.
• Given U[0,1) it is possible to map to other distributions.
• When estimating quantities of interest, be aware that there is monte carlo error as well as variability due to the population uncertainty.
• Many other option – importance sampling, Markov chain Monte Carlo (MCMC)…