Randomness in Computing and Simulations for Spherical and Spatial Geoscience Applications

J. A. Rod Blais
Department of Geomatics Engineering
Pacific Institute for the Mathematical Sciences
University of Calgary, Calgary, Alberta T2N 1N4
blais@ucalgary.ca
www.ucalgary.ca/~blais
Introduction to Randomness

- Mathematics: deterministic measures → probabilistic measures
- Physics: classical (deterministic) → modern (quantum physics)
- Computations: deterministic → pseudo random and stochastic
- Simulations: computations → Monte Carlo and stochastic
- Predictions: simulations → stochastic and probabilistic
- Practical considerations:
  - exact rigorous computations are not always possible
  - stochastic simulations often offer practical solutions
  - reproducibility often restricted to confidence levels
- Applications to direct and inverse problems
Distributional Considerations

Testing for 'randomness' in data sequences includes spatial, spectral, phase, autocorrelation, and numerous other characteristics. Randomness in data sequences does not obviously imply, nor is it implied by, low discrepancy or a uniform density as can easily be seen in sample spatial and spectral plots.

Well known procedures can be used to transform a random variate $y$ with a distribution $p(y)$ into another variate $x$ with distribution $p(x)$. The transformation follows the usual approach with Jacobians in integrals:

$$p(y)dy = p(x) \left| \frac{\partial x}{\partial y} \right| dy$$

which can obviously be simplified with uniform distributions.

Depending on the intended applications, strict randomness may or may not be critical. Furthermore, a uniform density or distribution may or may not even be relevant in the application context.
True Random Number Sequences

Basis: physical phenomena theoretically or empirically 'known' to be random

Examples:
- HotBits service based on radioactive decay
  www.fourmilab.ch
- Quantis generated by quantum mechanical process
  www.idquantique.com
- Random generated by atmospheric noise (radio static)
  www.random.org
Pseudo Random Number Sequences

Basis: computational rounding-off or related errors (machine dependent)

Common Methodology:
most often using some linear congruential model applied recursively such as
\[ x_n = c \odot x_{n-1} \mod \rho \]  
(for large prime \( \rho \) and constant \( c \))
or lagged Fibonacci congruential sequence, such as
\[ x_n = x_{n-p} \odot x_{n-q} \mod \rho \]  
(for large prime \( \rho \) and \( p, q \))
in which \( \odot \) usually stands for ordinary multiplication
Chaotic Random Number Sequences

Basis: computational chaotic processes (highly parameter dependent)

Common Methodology:

Using the Logistic equation (with parameter equal to four)

\[ x_n = 4 \times x_{n-1} (1 - x_{n-1}) \quad \text{for } n = 1, 2, 3, \ldots \]

using some random seed \( x_n \) in \((0, 1)\), which exhibits randomness with a density

\[ \rho(x) = \frac{1}{\pi [x (1 - x)]^{1/2}} \quad \text{(not a uniform distribution)} \]

Other similar procedures given in the literature (see e.g. Blais & Zhang, 2011)
**Quasi Random Number Sequences**

Basis: computational sequences of low discrepancy over intervals

Common Methodology:

Van der Corput (binary) sequences:

\[1, 10, 11, 100, 101, 111, \ldots \Rightarrow 0.1, 0.01, 0.11, 0.001, 0.101, 0.111, \ldots\]

Halton (binary) sequences:

\[\frac{1}{2}, \frac{1}{4}, \frac{3}{4}, \frac{1}{8}, \frac{5}{8}, \frac{3}{8}, \frac{7}{8}, \frac{1}{16}, \ldots\]

\(\pi\) digital expansion: 3, 1, 4, 1, 5, 9, 2, 6, 5, 3, 5, 8, 9, 7, 9, 3, 2, 3, 8, ...

\(e\) digital expansion: 2, 7, 1, 8, 2, 8, 1, 8, 2, 8, 4, 5, 9, 0, 4, 5, 2, 3, 5, ...

![Graphs and plots](image)
Expected Errors in Volume Estimates

Numerical Recipes state:

\[ \int_{V} f \, dV \approx V\langle f \rangle \pm \sqrt{\left( \langle f^2 \rangle - \langle f \rangle^2 \right)} / N \]

implying a variance \( O(1/N) \)

In general, then with \( N \) data values, it is well known that

- True Random Numbers \( \Rightarrow O(1/N) \) error variance
- Pseudo Random Numbers \( \Rightarrow O(1/N) \) error variance
- Chaotic Random Numbers \( \Rightarrow O(1/N) \) error variance

but

- Quasi Random Numbers \( \Rightarrow O((\ln N)^2s / N^2) \) error variance

for spatial dimension \( s \), and surprisingly (and yet to be generally replicated in practice) under the so-called 'superefficiency conditions with dynamical correlations for large \( N \')

- Chaotic Random Numbers \( \Rightarrow O(1/N^2) \) error variance

(see e.g. Umeno 2000, 1999, 1998 and Blais & Zhang, 2011)
Variance Reduction Strategies

• Importance Sampling Approach
  Essentially by analyzing the nature of the integrand
  Variable of integration may be transformed for better results
  Significant improvements are possible with complex problems

• Stratified Sampling Approach
  Largely by analyzing the characteristics of the integration domain
  Segmenting the domain may be considered for different sampling
  Small sample means often contribute to better overall results

• Mixed/Adaptive Strategies
  Both importance and stratified sampling can often be combined
  into optimal mixed and/or adaptive implementations, especially
  in high-dimensional applications
General Spatial Applications

Pseudo Random Numbers
Cartesian products of independent identically distributed sequences \( \{u_i\} \sim U(-1, +1) \)
such as using PLFG (see e.g. Tan & Blais, 2000) designed for parallel processing

Quasi Random Numbers
Cartesian products of independent identically distributed sequences \( \{v_j\} \sim U(-1, +1) \)
using Halton’s strategy or something similar

Gibbs Sampler Approach
A Gibbs sampler is a technique for generating random variables indirectly from some (marginal) distribution without calculating the density. For example, in digital image restoration, the Gibbs sampler is often based on immediate pixel neighborhoods for the Markov random field (see e.g. Geman & Geman, 1984).
Spherical Surface Explorations

Marsaglia [1972] with pseudo-random numbers:

Generate \( \{u_i\} \sim U(-1, 1) \) and \( \{v_j\} \sim U(-1, 1) \) independent with \( s_i^2 \equiv u_i^2 + v_i^2 < 1 \)

Estimate spherical points: \( \{2u_i(1-s_i^2)^{1/2}, 2v_i(1-s_i^2)^{1/2}, 2s_i^2 - 1\} \)
Spherical Surface Partitioning

Halton’s strategy for Quasi Random Numbers applied to the spherical surface:
  octahedron and icosahedron triangular faces subdivided as in spherical quadtrees
Concluding Remarks

• Stochastic simulations are very common in geoscience and elsewhere
• Characteristics of random numbers used often taken for granted but
  Spatial, phase and spectral maps of projections are often informative
• Autocorrelations and crosscorrelations should also be scrutinized
  Autocorrelations in Atmospheric Noise appear problematic
  Crosscorrelations can be serious in multidimensional applications
  Possible crosscorrelations need to be avoided in parallel computations
• Randomness may or may not be really critical in applications
  In volume estimations, variance reduction strategies are very useful
  In stochastic simulations, theoretical assumptions need be realized
• Simple Monte Carlo simulations are generally quite straightforward
• Complex Markov Chains of Monte Carlo simulations are more challenging
• In most geoscience processes, all simulation assumptions need be considered!