Computing and Statistical Data Analysis
Stat 3: The Monte Carlo Method

London Postgraduate Lectures on Particle Physics;

University of London MSci course PH4515

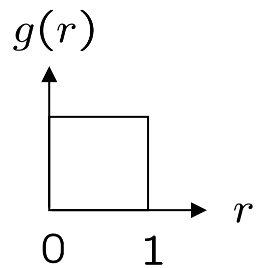
Glen Cowan
Physics Department
Royal Holloway, University of London
g.cowan@rhul.ac.uk
www.pp.rhul.ac.uk/~cowan

Course web page:

www.pp.rhul.ac.uk/~cowan/stat_course.html

The Monte Carlo method

What it is: a numerical technique for calculating probabilities and related quantities using sequences of random numbers.



The usual steps:

(1) Generate sequence $r_1, r_2, \ldots, r_m$ uniform in [0, 1].

(2) Use this to produce another sequence $x_1, x_2, \ldots, x_n$ distributed according to some pdf $f(x)$ in which we're interested (x can be a vector).

(3) Use the x values to estimate some property of $f(x)$, e.g., fraction of x values with $a < x < b$ gives $\int_a^b f(x)\,dx$.

→ MC calculation = integration (at least formally)

MC generated values = 'simulated data'

→ use for testing statistical procedures

Random number generators

Goal: generate uniformly distributed values in [0, 1].

Toss coin for e.g. 32 bit number... (too tiring).

→ 'random number generator'

= computer algorithm to generate $r_1, r_2, \ldots, r_n$.

Example: multiplicative linear congruential generator (MLCG)

$n_{i+1} = (a\, n_i) \bmod m$, where

n_i = integer

a = multiplier

m = modulus

n_0 = seed (initial value)

N.B. mod = modulus (remainder), e.g. 27 mod 5 = 2.

This rule produces a sequence of numbers $n_0, n_1, \ldots$

Random number generators (2)

The sequence is (unfortunately) periodic!

Example (see Brandt Ch 4): $a = 3$, $m = 7$, $n_0 = 1$

$$
\begin{aligned}
n_1 &= (3 \cdot 1) \bmod 7 = 3 \\
n_2 &= (3 \cdot 3) \bmod 7 = 2 \\
n_3 &= (3 \cdot 2) \bmod 7 = 6 \\
n_4 &= (3 \cdot 6) \bmod 7 = 4 \\
n_5 &= (3 \cdot 4) \bmod 7 = 5 \\
n_6 &= (3 \cdot 5) \bmod 7 = 1 \qquad \leftarrow \text{sequence repeats}
\end{aligned}
$$

Choose a, m to obtain long period (maximum = $m - 1$); m usually close to the largest integer that can represented in the computer.

Only use a subset of a single period of the sequence.

Random number generators (3)

$r_i = n_i / m$ are in [0, 1] but are they ‘random’?

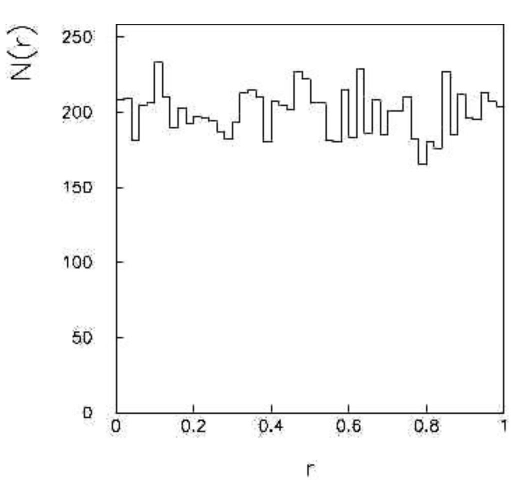
Choose a, m so that the r_i pass various tests of randomness:

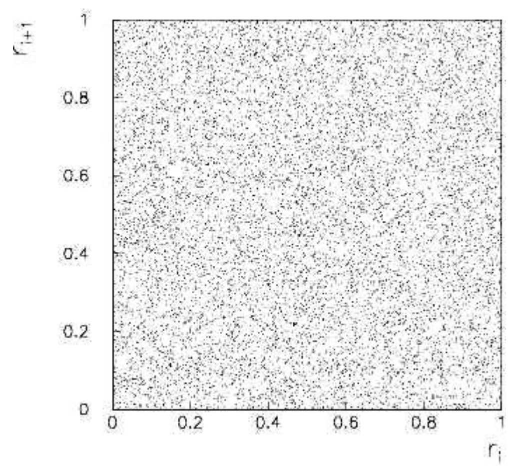
uniform distribution in [0, 1],

all values independent (no correlations between pairs),

e.g. L’Ecuyer, Commun. ACM **31** (1988) 742 suggests

$a = 40692$
$m = 2147483399$





Far better generators available, e.g. `TRandom3`, based on Mersenne twister algorithm, period = $2^{19937} - 1$ (a “Mersenne prime”).

See F. James, Comp. Phys. Comm. 60 (1990) 111; Brandt Ch. 4