Lecture 10

Random Number Sequences

Hai-Qing Lin

Beijing Computational Science Research Center

This PowerPoint Notes Is Based on the Textbook '***An Introduction to Computer Simulation Methods : Applications to Physical Systems***', 2nd Edition, Harvey Gould and Jan Tobochnik, Addison-Wesley(1996); "A First Course in Computational Physics"; "Numerical Recipes"; "Elementary Numerical Analysis"; "Computational Methods in Physics and Engineering".

Random Number Sequence

- Random numbers could be generated from any random physical process.
- However, in practice we may use a digital computer, a deterministic machine, to generate sequences of random numbers.
- So what we use is **pseudo-random number**.

Linear Congruential Generator

Most system-supplied random number generator are ***linear congruential generator*** (LCG), which generates a sequence of integers I_1, I_2, I_3, … each between 0 and $m - 1$ by the recurrence relation:

$$I_{j+1} = aI_j + c \ (mod\ m).$$

- ***m*** is called the ***modulus***;
- ***a*** is a positive integer, called the ***multiplier***;
- ***c*** is a positive integer, called the ***increment***.

Linear Congruential Generator

- The maximum possible period is m.
- In general, the period depends on **all three** parameters $\boldsymbol{m}$**,** $\boldsymbol{a}$**,** $\boldsymbol{c}$. They must be chosen carefully to achieve optimum results.
- Random number are usually referred to: $r = I_n / m$, $0 \leq r < 1$ or $r \in [0,1]$.
- To get random number distributed between $[a,b]$, simply by $x = a + (b - a)r$.

Important Features of RNG

- Its sequence satisfies the known statistical tests for randomness (see books on statistics for more).
- The probability distribution is uniform.
- The sequence has long period.
- The method is efficient.
- The sequence is reproducible.
- The algorithm is machine independent.

Choices of Parameters *m*, *a*, *c*

- $\boldsymbol{c} = \mathbf{0}$: *multiplicative congruential method*. The number generation process is a little faster but it cuts down the length of the period of the sequences. Still, it is possible to make the period reasonably long.
- $\boldsymbol{m}$: it should not be larger than the computer's word size w, namely, 2^e on an e-bit binary computer. We also want to pick a value so that the computation of $aI_j + c \pmod{m}$ is fast. Usually, $m = w$ leads to much less random sequences than $m = w - 1$. Another alternative is to let m be the largest prime number less than w.
- $\boldsymbol{a}$: very critical.

Linear Congruential Generator

a	m	c	period
7^5	$2^{31} - 1$	0	$2^{31} - 2$
1664525	2^{32}	1013904223	2^{32}
69069	2^{32}	0	2^{30}
6364136223846793005	2^{64}	1	2^{64}

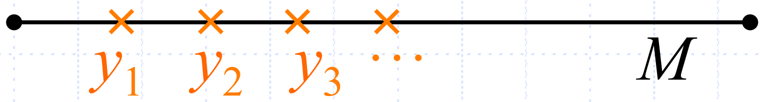
Test of Random Number Generator

- Function of time, any noticeable periodicity?
- Any noticeable pattern? (x,y)-plot, etc.
- Average $\langle x \rangle$ and variance $\langle x^2 \rangle$.
- Correlation? $\langle x_i x_{i+k} \rangle$, $k = 1, 2, \ldots,$ etc.
- Autocorrelation

$$C(k) = \frac{\langle x_{i+k} x_i \rangle - \langle x_i \rangle^2}{\langle x_i^2 \rangle - \langle x_i \rangle^2}.$$

Test of Random Number Generator

- Chi-square test.



- $y(i)$ is the number of data in the ith region

$$\chi^2 = \sum_{i=1}^{M} \frac{(y_i - E_i)^2}{E_i}$$

- ...

Test of Random Number Generator

- There is **no** necessary and sufficient test for the randomness of a finite sequence of numbers.
- The most that can be said is that it is "**apparently**" random.
- Improvement:
  use more than one generator, e.g., Shuffle.

Non-uniform *Discrete* Distribution

- A random integer i has value j with probability p_j, sum of them, $p_1, p_2, \ldots, p_n$, $\Sigma p_j = 1$.
- ξ is the uniformly distributed random no. in the interval (0,1). The random variable i distributed according to probabilities $p_1, p_2, \ldots, p_n$ can be generated by taking a random number ξ and deciding a value for i:

  if $\xi \le p_1$, then $i = 1$;
  if $p_1 \le \xi \le p_1 + p_2$, then $i = 2$;
  $\ldots$
  if $p_1 + \cdots + p_{m-1} \le \xi \le p_1 + \cdots + p_m$, then $i = m$.
- The most common case is $n = 2$.

A special case $p_1 = p_2 \cdots = p_n = 1/n$ can be obtained with a simple operation: $i = \lfloor n\xi \rfloor + 1$ where $\lfloor\ \rfloor$ means floor or truncation to integer.

Non-uniform *Continuous* Distribution

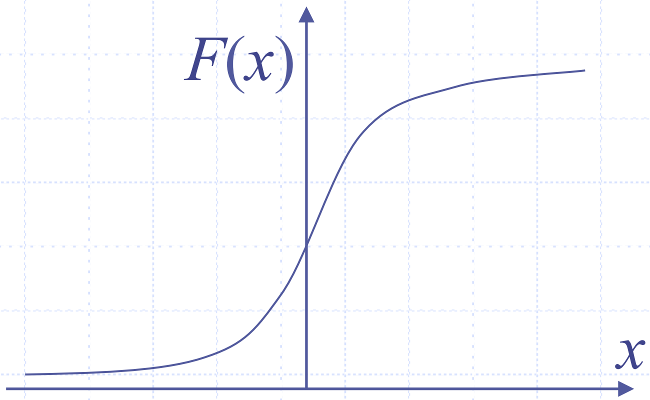
- A random variable x takes real values in some specified domain.
- The probability for x taking values between x and $x + dx$ is $p(x)dx$, where $p(x)$ is probability density.
- The distribution function $F(x)$ is defined by the probability that x is less than or equal to a given value x_0,

$$P(x \le x_0) = F(x_0) = \int_{-\infty}^{x_0} p(x)dx$$

- Since $F(x)$ is a probability, $0 \le F(x) \le 1$

Non-uniform Continuous Distribution

- $F(x)$ is a non-decreasing function of its argument.



- x with probability density $p(x)$ can be generated with

  $$x = F^{-1}(\xi),$$

  where $F^{-1}(\xi)$ is the inverse function of $F(x)$,
  and ξ is a uniformly distributed random number.

Example 1

Generate *x* according to exponential distribution

$$p(x)=\begin{cases} e^{-x}, & x \geq 0; \\ 0, & x < 0. \end{cases}$$

The distribution function is

$$F(x)=\int_{-\infty}^{x} p(x)dx=\int_{0}^{x} e^{-x}dx=1-e^{-x}, x \geq 0.$$

The inverse funciton is

$$x=-\ln(1-\xi).$$

Example 2

Generate x according to Gaussian distribution

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad -\infty < x < \infty.$$

Since the inverse of the Gaussian distribution function cannot be found analytically, it is helpful to generate a *pair* of Gassian random numbers, x and y.

The joint distribution of x and y is

$$p(x, y) = \frac{1}{2\pi} e^{-\left(x^2+y^2\right)/2}, \quad -\infty < x, y < \infty.$$

Introduce polar coordinates $r^2 = x^2 + y^2$ and $\theta = \tan^{-1}(y/x)$, the probability is rewritten as

$$p(x, y)\,dx\,dy = \frac{1}{2\pi} e^{-r^2/2} r\,dr\,d\theta.$$

Example 2

Generate *x* according to Gaussian distribution

Thus θ is distributed uniformly between 0 and 2π;
and r is distributed according to $r\exp(-r^2/2)$.

The distribution function for r is

$$F_r(r) = \int_0^r re^{-r^2/2}dr = 1 - e^{-r^2/2}\xi_1, \quad r = \sqrt{-2\ln(1-\xi_1)}.$$

We can replace $1 - \xi_1$ by ξ_1 since it does not change the probability distribution. The random variable θ can be generated by $\theta = 2\pi\xi_2$. And finally, x and y can be generated by

$$x = r\cos\theta = \sqrt{-2\ln\xi_1}\cos 2\theta\xi_2;$$

$$y = r\sin\theta = \sqrt{-2\ln\xi_1}\sin 2\theta\xi_2.$$

Theorem A:

- The linear congruential sequence defined by a, m, c, and I_0 has period length m if and only if
  - c is relatively prime to m;
  - $b = a - 1$ is a multiple of p, for every prime p dividing m;
  - b is a multiple of 4, if m is a multiple of 4.
- This theorem shows that the maximum period length cannot be achieved when $c = 0$.

Theorem A:

- In general, if d is any divisor of m and if I_n is a multiple of d, all succeeding elements $I_{n+1}, I_{n+2}, \ldots$ of the multiplicative sequence will be multiples of d.
- So we will want I_n to be relatively prime to m for all n. However, it is still possible to achieve an acceptably long period.
- Let $\lambda(m)$ denote the order of a primitive element, i.e., the maximum possible order, modulo m
- We find $\lambda(2) = 1$, $\lambda(4) = 2$, $\lambda(2^e) = 2^{e-2}$ if $e \geq 3$.
  $\lambda(p^e) = p^{e-1}(p-1)$ if $p > 2$.

Theorem B:

- The maximum period possible when $c = 0$ is $\lambda(m)$. This period is achieved if
  - I_0 is relatively prime to m.
  - a is a primitive element modulo m.
- Note that we can obtain a period of length $m - 1$ when m is a prime number.

GFSR

- Generalised feedback shift register(GFSR) is another popular random number generator.
- $x_n = x_{n-p} \oplus x_{n-p}$ where

  $\oplus$ is exclusive or operator, $p > q$ and x_n are integers.
- The first p random numbers must be supplied by another random number generator.
- NOT all values of p and q lead to good results.

Choice of *a*

- Most common one is 16807. Earlier choices were 65539 and 65549. You may mix them.
- Experiments showed that there seems to have some correlations between random numbers (say, 16807) separated by a power of two, e.g., L=32.
- Read more …