Computer system modeling and simulation

3- Random number generation

Sosina M.
Addis Ababa institute of technology (AAiT)
2012 E.C.

Random numbers

- Basic ingredient of discrete system simulation
  - Random numbers are used to generate event times and other random variables

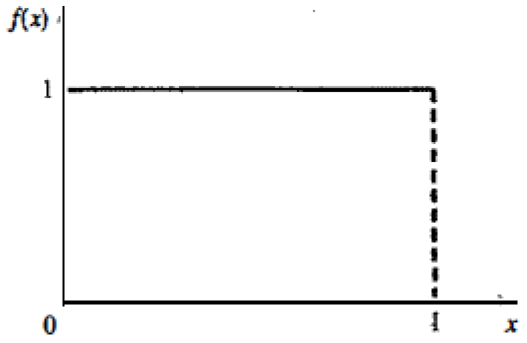
- ***Properties of random numbers***
  - Two important properties of a sequence of random numbers, R1, R2,…,
    - Uniformity
    - Independence
  - Each random number Ri must be an ***independent sample*** drawn from a continuous ***uniform distribution*** [0, 1]

Random numbers

- Probability density function (pdf)

$$f(x) = \begin{cases} 1, & 0 \leq x \leq 1 \\ 0, & otherwise \end{cases}$$

$$E(R) = \frac{1}{2} \qquad var(R) = \frac{1}{12}$$



Pseudo-random numbers

- "Pseudo" is used to imply that the very act of generating random number by a known method removes the potential for true randomness
  - If the method is known, the set of random numbers can be replicated

- The goal of any generation scheme, however, is to produce a sequence of numbers between [0, 1] and that imitates
  - the ideal properties of uniform distribution
  - and independence

Pseudo-random numbers

- Numerous methods can be used to generate a random numbers

- *Important considerations on random number generators*
  - Fast (computationally efficient)
  - Portable to different computers
  - Sufficiently long cycle – refers to the length of the random number sequence
  - Replicable
  - Approximate the ideal statistical properties of uniformity and independence

Techniques for generating random numbers

Linear congruential method

- Produces a sequence of integer numbers Xi in the range (0, m-1) by following a recursive relationships

$$X_{i+1} = (aX_i + c) mod\ m, \qquad i = 0,1,2,\ldots$$

  - The initial value X_0 is called ***the seed***
  - a – multiplier
  - c- is the increment
    - ✓ C=0 →multiplicative congruential generator
    - ✓ C!=0 →mixed congruential generator
- The selection of the values for a, c, m and X_0 drastically affects the statistical properties and the cycle length

Linear congruential method

- Given a sequences of X integer numbers in [0, m), random number between [0, 1) can be generated from

$$Z_i = \frac{X_i}{m}$$

- Example – x0=27, a=17, c=43 and m=100
  - X1=2 → R1=2/100=0.02
  - X2=77 → R2=77/100=0.77
  - X3=52 → R3=52/100=0.52

- How closely the generated numbers R1, R2, … approximate uniformity and independence?
- Other properties – maximum density and maximum period

Linear congruential method

- I=(0, 1/m, 2/m,…,(m-1)/m)
  - Each Xi is an integer in the set{0,1,2,…,m-1} → each Ri is discrete on I
- If m is a very large integer, the values assumed by Ri leave no large gaps on [0, 1] (maximum density)
  - m=$2^{31}-1$ *and* $m=2^{48}$ are commonly used
- Maximum period can be achieved by the proper choice of a, c, m and x0

Linear congruential method

- when m=2^b, c $\neq$ 0
  - The maximum period P= m, if
    - C is relatively prime to m (the common factor is 1)
    - a=1+4k, where k is an integer
- When m=2^b, $c = 0$
  - The longest possible period is P=m/4, if
    - X0 is odd
    - a=3+8k or a=5+8k, k=0, 1,…
- When m is a prime number and c=0
  - The longest possible period P=m-1, if
    - The smallest integer k such that $a^k - 1$ is divisible by m is k=m-1

Linear congruential method

- Example: *using multiplicative congruential method find the period of the generator for a=13, m=2^6 and x0=1, 2, 3, and 4*

i	X_i	X_i	X_i	X_i
0	1	2	3	4
1	13	26	39	52
2	41	18	59	36
3	21	42	63	20
4	17	34	51	4
5	29	58	23	
6	57	50	43	
7	37	10	47	
8	33	2	35	
9	45		7	
10	9		27	
11	53		31	
12	49		19	
13	61		55	
14	25		11	
15	5		15	
16	1		3	

Linear congruential method

- Speed and efficiency
  - Most digital computers use a binary representation of numbers
  - The modulo operation can be conducted efficiently when the modulo is a power of 2

Combined linear congruential generators

- Combining two or more multiplicative congruential generators in such a way that the combined generator has ***good statistical properties*** and a ***longer period***
  - Let $X_{i,1}$, $X_{i,1}$, $X_{i,2}, \ldots,$ $X_{i,k}$ be the ith output from k different multiplicative congruential generators

$$X_i = \left(\sum_{j=1}^{k} (-1)^{j-1} X_{i,j}\right) \bmod m_1 - 1$$

$$R_i = \begin{cases} \dfrac{X_i}{m_1} & X_i > 0 \\ \dfrac{m_1 - 1}{m_1} & X_i = 0 \end{cases}$$

$$P = \frac{(m_1-1)(m_2-1)\ldots(m_k-1)}{2^{k-1}}$$

Combined linear congruential generators

- Example:
  - k=2,
  - m1=2,147,483,563, a1=40,014
  - m2=2,147,483,399 and a2=40,692
  - The combined generator has period $=2*10^{18}$

Test for random numbers

- Desirable properties – uniformity and independence
- To check on whether these desirable properties have been achieved, a number of tests can be performed

Frequency test

- For uniformity test
  - **Kolmogorov-smirnov test**
  - **Chi-square test**
- Both methods measure the degree of agreement between
  - The distribution of a sample of generated random numbers
  - And theoretical uniform distribution
- Both tests are based on the null hypothesis of no significant difference between the sample distribution and the theoretical distribution

Kolmogorov-smirnov test

- Compares the continuous CDF, F(x), of the uniform distribution with the empirical CDF, $S_N(x)$, of the sample of N observation

  - F(x)=x, 0≤x ≤1
  - If the samples from the random-number generator are R1,R2,…,RN, then the empirical CDF is

$$S_N(x) = \frac{number\ of\ R1,R2,\ldots,RN\ \ which\ are \leq x}{N}$$

  - As N becomes larger, $S_N(x)$ should become a better approximation to F(x)

Kolmogorov-smirnov test

- The Kolmogorov test is based on the larges absolute deviation between F(x) and $S_N(x)$ over the range of the random variable

$$D=\max |F(x)-S_N(x)|$$

- D is compared with the largest theoretical deviation for N instances generated by an ideal generator
- Critical values for the D distribution are usually tabulated as a function of N and for specific levels of significance

Kolmogorov-smirnov test

- The test procedure follows these steps

1. Rank the data from smallest to largest. Let R(i) denote the ith smallest observation, so that

$$R_{(1)} \leq R_{(2)} \leq \ldots \leq R_{(N)}$$

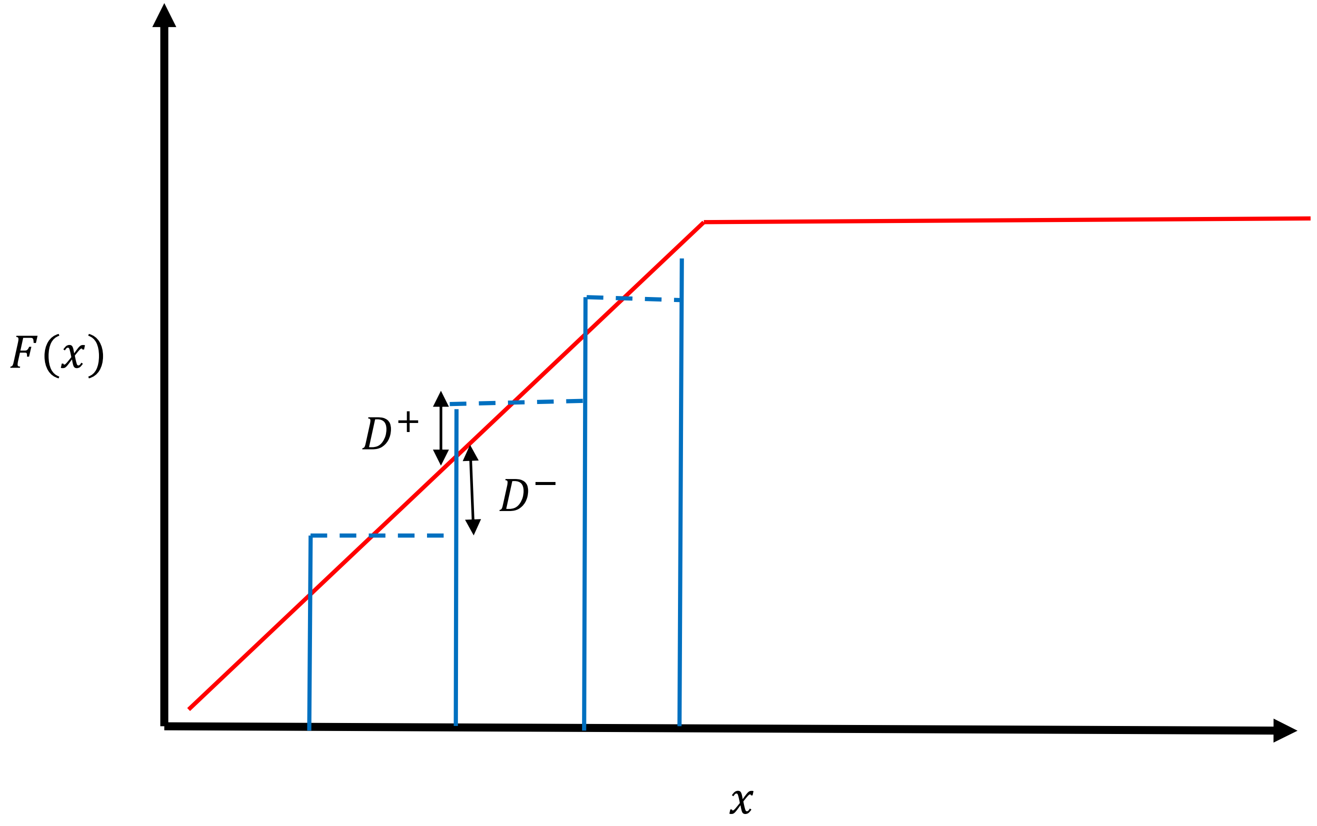
2. Compute

$$D^{+} = \max_{1 \leq i \leq N} \left\{ \frac{i}{N} - R_{(i)} \right\}$$

$$D^{-} = \max_{1 \leq i \leq N} \left\{ R_{(i)} - \frac{i-1}{N} \right\}$$

3. Compute D=max(D^{+}, D^{-})

Kolmogorov-smirnov test



Kolmogorov-smirnov test

4. Locate the value D_α for the specified significance level α and the given sample size N
5. If D>D_α $\rightarrow$ the null hypothesis that the data are a sample from a uniform distribution is rejected. If D ≤D_α , concludes that no difference has been detected

Kolmogorov-smirnov test

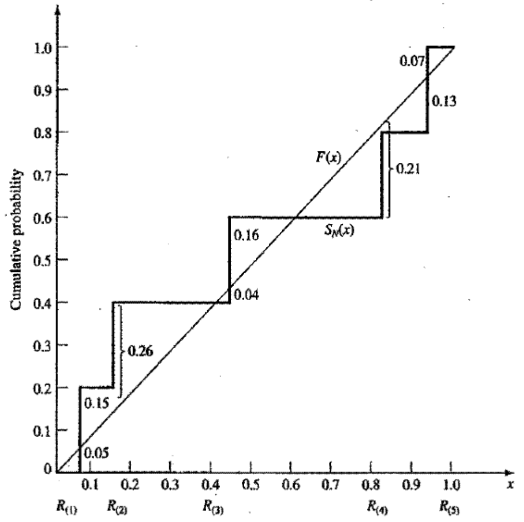
- Kolmogorov test critical values

Degrees of Freedom (N)	$D_{0.10}$	$D_{0.05}$	$D_{0.01}$
1	0.950	0.975	0.995
2	0.776	0.842	0.929
3	0.642	0.708	0.828
4	0.564	0.624	0.733
5	0.510	0.565	0.669
6	0.470	0.521	0.618
7	0.438	0.486	0.577
8	0.411	0.457	0.543
9	0.388	0.432	0.514
10	0.368	0.410	0.490
11	0.352	0.391	0.468
12	0.338	0.375	0.450
13	0.325	0.361	0.433
14	0.314	0.349	0.418
15	0.304	0.338	0.404
16	0.295	0.328	0.392
17	0.286	0.318	0.381
18	0.278	0.309	0.371
19	0.272	0.301	0.363
20	0.264	0.294	0.356
25	0.24	0.27	0.32
30	0.22	0.24	0.29
35	0.21	0.23	0.27

Kolmogorov-smirnov test

- Example: suppose five number 0.44, 0.81, 0.14, 0.04, 0.93 were generated

$R_{(i)}$	0.05	0.14	0.44	0.81	0.93
i/N	0.20	0.40	0.60	0.80	1.00
$i/N - R_{(i)}$	0.15	0.26	0.16	—	0.07
$R_{(i)} - (i-1)/N$	0.05	—	0.04	0.21	0.13



  - D=max(0.26, 0.21)=0.26
  - Value of D obtained from table for α=0.05 and N=5 is 0.565
  - 0.26<0.565

Chi-square test

- The chi-square test uses the sample statistic

$$X_0^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

  - Where O_i and E_i are the observed number and the expected number, respectively, in the ith class
  - n is the number of classes
- For uniform distribution, the expected number in each class is given by

$$E_i = \frac{N}{n}$$ where N is the total number of observations

- It can be shown that the sampling distribution of X_0^2 is approximately the chi-square distribution with n-1 degrees of freedom

Chi-square test

- Percentage points of the Chi-square distribution with v degree of freedom

v	$\chi^2_{0.005}$	$\chi^2_{0.01}$	$\chi^2_{0.025}$	$\chi^2_{0.05}$	$\chi^2_{0.10}$
1	7.88	6.63	5.02	3.84	2.71
2	10.60	9.21	7.38	5.99	4.61
3	12.84	11.34	9.35	7.81	6.25
4	14.96	13.28	11.14	9.49	7.78
5	16.7	15.1	12.8	11.1	9.2
6	18.5	16.8	14.4	12.6	10.6
7	20.3	18.5	16.0	14.1	12.0
8	22.0	20.1	17.5	15.5	13.4
9	23.6	21.7	19.0	16.9	14.7
10	25.2	23.2	20.5	18.3	16.0
11	26.8	24.7	21.9	19.7	17.3
12	28.3	26.2	23.3	21.0	18.5
13	29.8	27.7	24.7	22.4	19.8
14	31.3	29.1	26.1	23.7	21.1
15	32.8	30.6	27.5	25.0	22.3
16	34.3	32.0	28.8	26.3	23.5
17	35.7	33.4	30.2	27.6	24.8
18	37.2	34.8	31.5	28.9	26.0
19	38.6	36.2	32.9	30.1	27.2
20	40.0	37.6	34.2	31.4	28.4
21	41.4	38.9	35.5	32.7	29.6
22	42.8	40.3	36.8	33.9	30.8
23	44.2	41.6	38.1	35.2	32.0
24	45.6	43.0	39.4	36.4	33.2
25	49.6	44.3	40.6	37.7	34.4
26	48.3	45.6	41.9	38.9	35.6
27	49.6	47.0	43.2	40.1	36.7
28	51.0	48.3	44.5	41.3	37.9
29	52.3	49.6	45.7	42.6	39.1
30	53.7	50.9	47.0	43.8	40.3
40	66.8	63.7	59.3	55.8	51.8
50	79.5	76.2	71.4	67.5	63.2
60	92.0	88.4	83.3	79.1	74.4
70	104.2	100.4	95.0	90.5	85.5
80	116.3	112.3	106.6	101.9	96.6
90	128.3	124.1	118.1	113.1	107.6
100	140.2	135.8	129.6	124.3	118.5

Chi-square test

- Example: use the chi-square test with α=0.005 test
  - n=10 intervals [0, 0.1), [0.1, 0.2), …, [0.9, 1)
  - N=100

0.34	0.90	0.25	0.89	0.87	0.44	0.12	0.21	0.46	0.67
0.83	0.76	0.79	0.64	0.70	0.81	0.94	0.74	0.22	0.74
0.96	0.99	0.77	0.67	0.56	0.41	0.52	0.73	0.99	0.02
0.47	0.30	0.17	0.82	0.56	0.05	0.45	0.31	0.78	0.05
0.79	0.71	0.23	0.19	0.82	0.93	0.65	0.37	0.39	0.42
0.99	0.17	0.99	0.46	0.05	0.66	0.10	0.42	0.18	0.49
0.37	0.51	0.54	0.01	0.81	0.28	0.69	0.34	0.75	0.49
0.72	0.43	0.56	0.97	0.30	0.94	0.96	0.58	0.73	0.05
0.06	0.39	0.84	0.24	0.40	0.64	0.40	0.19	0.79	0.62
0.18	0.26	0.97	0.88	0.64	0.47	0.60	0.11	0.29	0.78

Chi-square test

- Computation for Chi-square test

Interval	O_i	E_i	$O_i - E_i$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
1	8	10	−2	4	0.4
2	8	10	−2	4	0.4
3	10	10	0	0	0.0
4	9	10	−1	1	0.1
5	12	10	2	4	0.4
6	8	10	−2	4	0.4
7	10	10	0	0	0.0
8	14	10	4	16	1.6
9	10	10	0	0	0.0
10	11	10	1	1	0.1
	100	100	0		3.4

- X_0^2=3.4 is much smaller than $X_{0.05,9}^2$=16.9 => the null hypothesis for a uniform distribution is not rejected

Uniformity test

- Both Kolmogorov and chi-square are acceptable for testing the uniformity of a sample of data – provided that the sample size is large
- Kolmogorov test is more powerful of the two
- Furthermore, Kolmogorov test can be applied to small sample size, chi-square is valid only for large samples (n>50)

Test for autocorrelation

- Test of autocorrelation is concerned about the dependence between numbers in a sequence
- Example: consider the following sequence of numbers

0.12	0.01	0.23	0.28	0.89	0.31	0.64	0.28	0.83	0.93
0.99	0.15	0.33	0.35	0.91	0.41	0.60	0.27	0.75	0.88
0.68	0.49	0.05	0.43	0.95	0.58	0.19	0.36	0.69	0.87

  - The 5^{th}, 10^{th} and so on indicates a very large number in that position – the numbers in the sequence might be related

Test for autocorrelation

- Computing the autocorrelation between every m numbers, starting with the ith number
- The autocorrelation ρ_{im} between the following numbers would be of interest:

$$R_i, R_{i+m}, R_{i+2m}, \dots, R_{i+(M+1)m}$$

  - M is the largest integer such that $i + (M + 1)m \leq$ N, N is the total value in the sequence

Test for autocorrelation

- A nonzero autocorrelation implies a lack of independence
- For large values of M, the distribution estimator of ρ_{im}, denoted $\widehat{\rho_{im}}$, is approximately normal if the values $R_i, R_{i+m}, R_{i+2m}, \ldots, R_{i+(M+1)m}$ are uncorrelated
- The test statistic can be formed as follows

$$Z_0 = \frac{\widehat{\rho_{im}}}{\sigma_{\widehat{\rho_{im}}}}$$

- Which is distributed normally with a mean of zero and variance of 1

Test for autocorrelation

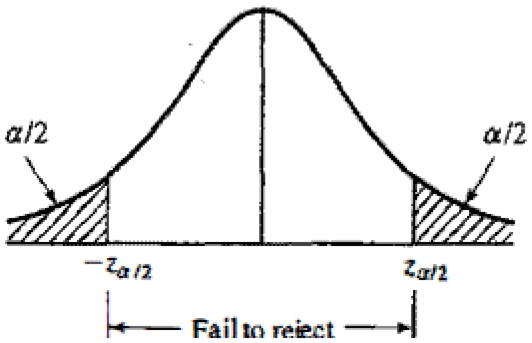
- The formula for $\widehat{\rho_{im}}$ in the slightly different form and the standard deviation of the estimator are given by

$$\widehat{\rho_{im}} = \frac{1}{M+1}\left[\sum_{k=0}^{M} R_i, R_{i+(k+1)m}\right] - 0.25$$

$$\sigma_{\widehat{\rho_{im}}} = \frac{\sqrt{13M+7}}{12(M+1)}$$

- After computing Z_0, do not reject the null hypothesis of independence if

$$-z_{\alpha/2} \leq Z_0 \leq z_{\alpha/2}$$



Test for autocorrelation

- **Example**: test whether the 3rd, 8th, 13th , and son on, numbers in the sequence are autocorrelated using α=0.05
  - i=3, m=5, M=4

0.12	0.01	0.23	0.28	0.89	0.64	0.28	0.83	0.75	0.93
0.99	0.15	0.33	0.35	0.91	0.60	0.27	0.75	0.83	0.88
0.68	0.49	0.05	0.43	0.95	0.19	0.36	0.69	0.69	0.87

$$\widehat{\rho}_{35} = -0.1945$$
$$\widehat{\sigma} = 0.128$$
$$Z_0 = -1.516 < z_{0.025} = 1.96$$

The hypothesis of independence cannot be rejected on the basis of this test

is