LCG Ex.

- Use X_0 *= 27, a = 17, c = 43, and m = 100.*

$$X_{i+1} = (a X_i + c) \bmod m$$

The X_i *and* R_i *values are:*

- $X_1 = (17*27+43) \bmod 100$

  $= 502 \bmod 100 = 2,$ $R_1 = 0.02;$
- $X_2 = (17*2+32) \bmod 100 = 77,$ $R_2 = 0.77;$
- $X_3 = (17*77+32) \bmod 100 = 52,$ $R_3 = 0.52;$
- *…*

LCG

Remarks:

- $c = 0$, the generator is called Multiplicative LCG.
- If $c \neq 0$, the generator is called Mixed LCG.
- The length of the cycle is called its Period, can be at most?
- m should be chosen to be big
- Choose m of the form 2^k for efficient computation

LCG

Remarks:

- $c = 0$, the generator is called Multiplicative LCG.
- If $c \neq 0$, the generator is called Mixed LCG.
- The length of the cycle is called its Period, can be at most m-1
- m should be chosen to be big
- Choose m of the form 2^k for efficient computation

LCG

Theorem:

If $c \neq 0$, LCG has full period

iff

- Integers *m* and *c* are relatively prime (the only positive integer that divides both *m* and c is 1)
- Every prime number that is a factor of *m* is also a factor of a-1
- If integer *m* is a multiple of 4, a-1 is also a multiple of 4

Multiplicative Congruential Generator

- If c= 0, $X_{i+1}= (a X_i) \bmod m$ i =0,1,2,…
- Break the first condition of the thm. i.e. Not full period
- Its max period is 2^{b-1} where b is the # of digits (size of ward)
- Advantages:

Faster, Simpler, easy to implement

Seed Selection

- Often we need random numbers for more than one variable in a simulation
- E.g., inter-arrival times, service times
- Then, we need to use multiple random number streams such that we do not introduce correlations between the two random variables owing to our choice of random numbers

Seed Selection

For a good selection of seeds:

- Do not use zero
- Avoid even values
- Use a separate seed for a separate stream such that random numbers do not overlap
- Do not use random seeds because they are hard to replicate

Need to test

- The above only tells us how to create RNs from RNG that will have a large period
- It does not guarantee the RNG output will be "random" (e.g., $x_{i+1} = (x_i + 1) \ mod \ m$ not random!)
- Need to apply statistical tests to validate that the RNG gives acceptably random results

Testing Random Number Generators

- Two categories of test
- Test for uniformity …
- Test for independence
- Passing a test is only a necessary condition and not a sufficient condition
- i.e., if a generator fails a test it implies it is bad
- but if a generator passes a test it does not

Testing Distributions

- Comparing Distributions: Tests for Goodness-of-Fit

Know how to compare between two distributions

- Chi-Square Distribution (for discrete models)
- Kolmogorov-Smirnov (K-S) Test (for continuous models)

Goodness-of-fit

- Statistical Tests enable us to compare between two distributions, also known as **Goodness-of-Fit**.
- The **goodness-of-fit** of a statistical model describes how well it fits a set of observations.
- Measures of goodness of fit typically summarize the discrepancy between observed values and the values expected under the model in question
- Goodness-of-fit means how well a statistical

(PEARSON'S) CHI-SQUARE TESTS FOR DISCRETE MODELS

The Pearson's chi-square test is to compare two probability mass functions of two distribution.

If the difference value (Error) is **greater** than the critical value, the two distribution are said to be different or the first distribution does not fit (well) the second distribution.

If the difference if **smaller** than the critical value, the first distribution fits well the second distribution

(PEARSON'S) CHI-SQUARE TEST

- Pearson's chi-square is used to assess two types of comparison:
  - tests of goodness of fit: it establishes whether or not an observed frequency distribution differs from a theoretical distribution.
  - tests of independence. it assesses whether paired observations on two variables are independent of each other.

Steps in Test of Hypothesis

1. Determine the appropriate test
2. Establish the level of significance: α
3. Formulate the statistical hypothesis
   H_o : The two variables are independent
   H_1 : The two variables are not independent
4. Calculate the test statistic
5. Determine the degree of freedom
6. Compare computed test statistic against the critical value (from the table of the test)

- The critical tabled values are based on sampling distributions of the Pearson chi-square statistic
- If calculated χ^2 is greater than χ^2 table value, reject H_o

Testing

- Testing is not necessary if a well-known simulation package is used or if a well tested generator is used
- we focus on “empirical” tests, that is tests that are applied to an actual sequence of random numbers
- EX:

Chi-Square Test

Chi-Square Test

-

$$\chi^2 = \sum \frac{(obs - \exp)^2}{\exp}$$

-

Observed values

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

Compute terms:

Observed: $O =$

	0
0	8
1	8
2	10
3	9
4	12
5	8
6	10
7	14
8	10
9	11

Expected: $E =$

	0
0	10
1	10
2	10
3	10
4	10
5	10
6	10
7	10
8	10
9	10

Difference: $O - E =$

	0
0	-2
1	-2
2	0
3	-1
4	2
5	-2
6	0
7	4
8	0
9	1

Difference²: $(O - E)^2 =$

	0
0	4
1	4
2	0
3	1
4	4
5	4
6	0
7	16
8	0
9	1

Normalized: $T =$

	0
0	0.4
1	0.4
2	0
3	0.1
4	0.4
5	0.4
6	0
7	1.6
8	0
9	0.1

$$T_i := \frac{(O_i - E_i)^2}{E_i}$$

$$\chi 2 := \sum_{i=0}^{9} \frac{(O_i - E_i)^2}{E_i}$$

$$\chi 2 = 3.4$$

$$\chi^2_{0.05,9} = 16.9$$

Critical Values of the χ^2 Distribution

df \ p	0.995	0.975	0.9	0.5	0.1	0.05	0.025	0.01	0.005	df
1	.000	.000	0.016	0.455	2.706	3.841	5.024	6.635	7.879	1
2	0.010	0.051	0.211	1.386	4.605	5.991	7.378	9.210	10.597	2
3	0.072	0.216	0.584	2.366	6.251	7.815	9.348	11.345	12.838	3
4	0.207	0.484	1.064	3.357	7.779	9.488	11.143	13.277	14.860	4
5	0.412	0.831	1.610	4.351	9.236	11.070	12.832	15.086	16.750	5
6	0.676	1.237	2.204	5.348	10.645	12.592	14.449	16.812	18.548	6
7	0.989	1.690	2.833	6.346	12.017	14.067	16.013	18.475	20.278	7
8	1.344	2.180	3.490	7.344	13.362	15.507	17.535	20.090	21.955	8
9	1.735	2.700	4.168	8.343	14.684	16.919	19.023	21.666	23.589	9
10	2.156	3.247	4.865	9.342	15.987	18.307	20.483	23.209	25.188	10
11	2.603	3.816	5.578	10.341	17.275	19.675	21.920	24.725	26.757	11
12	3.074	4.404	6.304	11.340	18.549	21.026	23.337	26.217	28.300	12
13	3.565	5.009	7.042	12.340	19.812	22.362	24.736	27.688	29.819	13
14	4.075	5.629	7.790	13.339	21.064	23.685	26.119	29.141	31.319	14
15	4.601	6.262	8.547	14.339	22.307	24.996	27.488	30.578	32.801	15

EX.

.34	.9	.25	.89	.87	.44	.12	.21	.46	.67
.83	.76	.79	.64	.7	.81	.94	.74	.22	.74
.96	.99	.77	.67	.56	.41	.52	.73	.99	.02
.47	.3	.17	.82	.56	.05	.45	.31	.78	.05
.79	.71	.23	.19	.82	.93	.65	.37	.39	.42
.99	.17	.99	.46	.05	.66	.1	.42	.18	.49
.37	.51	.54	.01	.81	.28	.69	.34	.75	.49
.72	.43	.56	.97	.3	.94	.96	.58	.73	.05
.06	.39	.84	.24	.4	.64	.4	.19	.79	.62
.18	.26	.97	.88	.64	.47	.6	.11	.29	.78

What is Monte Carlo Simulation ?

- **Monte Carlo methods** are a widely used class of **computational algorithms** for simulating the behavior of various physical and mathematical systems, and for other computations.
- **Monte Carlo algorithm** is often used to find solutions to mathematical numerical problems (which may have many variables) that cannot easily be solved, (e.g. integral calculus, or other numerical methods)

Monte Carlo Simulation

- A scheme employing random numbers which is used to solve certain stochastic or deterministic problems where the passage of time plays no substantive role.
- Common problem is the estimation of $\int_{\Omega} f(\mathbf{x})d\mathbf{x}$, where *f* is a function, **x** is a vector and Ω is domain of integration.
- Special case: Estimate $\int_a^b f(x)dx$ for scalar *x* and limits of integration *a*, *b*

Monte Carlo Simulation

Let X be a uniform random variable on the interval $[a, b]$ with density

$$p(x) = \frac{1}{b-a}, \quad a \le x \le b$$

and let $x_1, \ldots, x_n$ be a random sample from X. Then

$$\begin{aligned}
\int_a^b f(x)dx &= \int_a^b \frac{f(x)}{p(x)} p(x)dx \\
&= (b-a)\int_a^b f(x)p(x)dx \\
&= (b-a)E[f(X)] \\
&\approx \frac{b-a}{n}\sum_{i=1}^{n} f(x_i)\,.
\end{aligned}$$

Monte Carlo Simulation

Example: Estimate $\int_0^b \sin(x)dx$.

We approximate this by

$$\frac{b}{n}\sum_{i=1}^{n}\sin(x_i) ,$$

where x_1, …, x_n are a sample from a uniform [0, *b*] random variable.

Monte Carlo Simulation

Example: Estimate $\int_0^b \sin(x)dx$.

	n = 10	n = 100	n = 1000	n = 2000
b = 1 (answer = 2)	1.753	2.032	1.994	1.999
b = 2 (answer = 0)	-0.898	-0.013	0.137	0.079

There is considerable variability in the quality of solution; accuracy of numerical integration sensitive to integrand and domain of integration

Case Study
Cake’s shop problem

An owner of a bakery shop would like to determine how many 10-inch birthday cakes he should produce each day in order to maximize his profit. His present method of determining the quantity to bake is based on his best guess.

Cake’s shop problem

- The production costs are $2.00 per cake.
- And the profit for each cake is $2.5.
- However, If over estimates the daily demand, some cakes will be left over at the end of the day. The policy is to sell all leftover cakes to a local store that specializes in day-old items. He is currently receiving $1.50 per cake for the surplus cakes, thus incurring a loss of $0.50 per cake.

Cake's shop problem

- Case 1: The production quantity is less than or equal to demand

If $x <= d$, $z = 2.5x$

- Case 1: The production quantity is greater than the demand

If $x > d$ $z = 2.5d + (x-d)(-0.5)$

$Z = 3.00\,d - 0.5\,x$

Cake's shop problem

- Generalizeing:

p = selling price for each cake

c = cost of each unit

s = day-old price

- If $x <= d$, $z = (p-c)\,x$
- If $x > d$ $z = (p-c)\,d + (x-d)\,(s-c)$

$Z = (p-s)\,d - (s-c)\,x$

Historical day demand for the birthday cakes

Daily deman	Frequency (	Probability Distribution
0	1	0.05
1	2	0.1
2	1	0.05
3	2	0.1
4	3	0.15
5	6	0.3
6	3	0.15
7	1	0.05
8	1	0.05
Total	20	1

$$relative\_frequency = \frac{frequency\_of\_observation}{total\_number\_of\_observations}$$

Hand Simulation

- Take a sheet of paper and cut it into twenty equal pieces.
- Follow the historical daily demand frequency in the table,
- write the number zero on one piece.
- On two of the remaining pieces write the number one, which stands for the demand of one unit.
- Check the numbers you have written carefully, because this “deck” of twenty

Hand Simulation

- The first step is the selection of the production quantity, Assume (x=3).
- use the deck of twenty slips of paper to generate a demand by selecting one slip of paper at random.
- Suppose the first slip drawn has a 5 written on it.
- We shall then use a demand of 5 cakes for the first simulated day of bakery shop operation.
- i.e. underproduction of 2 cakes.

Hand Simulation

- Since $x<d$, we can computer our first day's profit using the expression
- 2.5x=25(3)=\$7.5. i.e. Total profit of \$7.5.
- Generate the demand for second day (reshuffle and draw a piece) suppose d=1
- Since $x>d$ use the second case
- z=3(1) - 0.5(3) = \$1.5
- So the total profit is 7.5 + 1.5 = 9

10-day simulation results for production quantity x=3

Day	Generated demand	Daily profit	Total profit
1	5	7.5	7.5
2	1	1.5	9
3	6	7.5	16.5
4	3	7.5	24
5	4	7.5	31.5
6	4	7.5	39
7	3	7.5	46.5
8	0	-1.5	45
9	5	7.5	52.5
10	6	7.5	60

Hand Simulation

- Now we perform the same ten day simulation for another quantity production x =1,2,3…8
- Compare the total profit for each one
- Pick the best profit to be the suggested production quantity
- Of course if we run the simulation for more days we get more accurate estimate.

10-days Simulation Results for various production quantities

Production Size	Ten Day Simulated profit $
1	25
2	44
3	60
4	79
5	90
6	93
7	91
8	89
9	5
10	6

From the table it is clear that the best production quantity that maximizes the profit is at x=6. The results are based on only 10-day simulation.

The role of random numbers in simulation

- Suppose we select random numbers in sets of two digits.
- This will provide us with 100 two-digit random numbers from 00 to 99 with each two-digit random number having a 1/100 chance of being selected
- 0 unites, the relative frequency of 0 is 5% Thus we want 5% of the 100 possible two-digit random numbers to correspond to a demand of 0 units.
- While choosing any five numbers of the 100 numbers will do we may assign a demand of 0 to the first 5 numbers i.e. 00, 01,02,03, and 04

Random number Intervals and the daily demand

Daily Dema	relative Frequen	Interval of Random num
0	0.15	00 to 04
1	0.1	05 to 14
2	0.05	15 to 19
3	0.1	20 to 29
4	0.15	30 to 44
5	0.3	45 to 74
6	0.15	75 to 89
7	0.05	90 to 94
8	0.05	95 to 99

Results of simulating ten daily demands

Random number Simulated daily demand	simulated daily demand
63	5
27	3
15	2
99	8
86	6
71	5
74	5
45	5
11	1
2	0

The role of random numbers

- For any simulation problem in which a relative frequency distribution of a variable can be developed,
- It is easy to apply the above random number based procedure to simulate values of the variable.
- First, develop a table of intervals by associating an interval of random numbers with each possible value of the variable
- Then as each random number is selected, you can simply check the corresponding interval and find the associated value of the variable.

The role of random numbers

- Obviously, for long and complex simulations that require numerous calculations, a high speed computer simulation process is desirable.
- In computer simulation pseudo- random numbers are used in exactly the same way as the random numbers selected from random number tables above.
- It would be very risky to make a decision based on the results of such a short period of simulation.

The role of random numbers

- When we think of performing the simulation calculations for a simulated period as long as 500 days, the problems of carrying out the simulation for even a case as small as the bakery shop problem are significant.
- For example let us consider the 500 days. The mathematical model does not change but the work we have to go through to evaluate the results does change but expands . Now we can create a table similar to the ten-day table to evaluate each order size for 500 days of operation.