

**NOORUL ISLAM COLLEGE OF ENGINEERING
KUMARACOIL**

SUBJECT NAME : MODELING AND SIMULATION
SUBJECT CODE : XCS 472
CLASS : S7 MSC (SE)

**PREPARED BY
MS. R.I. HEAVEN ROSE
DEPT. OF S/W ENGINEERING**

NOORUL ISLAM COLLEGE OF ENGINEERING, KUMARACOIL
Department of Software Engineering.
Seventh Semester
XCS-472 MODELLING AND SIMULATION

Two marks Question and Answers
UNIT I

1. Define simulation.

A simulation is the imitation of the operation of a real-world process or system over time.

2. What are the advantages of simulation ?

- New policies can be explored without disrupting ongoing operations of the real system.
- New hardware design can be tested without committing resources and their acquisition.
- Hypotheses about how and why certain phenomena occur can be tested for feasibility.
- Insight can be obtained about the interaction of variables.

3. What are the disadvantages of simulation?

- Model building requires special training.
- Simulation results may be difficult to interpret.
- Simulation modeling and analysis can be time consuming and expensive.
- Simulation is used in some cases when an analytical solution is possible.

4. How will we overcome the disadvantages of simulation ?

- Models have the generic the simulators or templates.
- Many simulation software vendors have developed output analysis capabilities.
- Simulation can be performed faster today than the yesterday.
- Closed-form models are not able to analyze most of the complex systems that are encountered in practice

5. What are the areas of application of simulation?

- Institute of electrical and electronic engineering: computer society.
- Institute of electrical and electronic engineering: systems, man and cybernetics society.
- Institute of industrial engineers

6. Define entity activity and attribute?

- An entity is an object of interest in the system.
- An activity represents a time period of specified length.
- An attribute is a property of an entity.

7. Define endogenous and exogenous

- Endogenous is used to describe activities and events occurring within the system.
- Exogenous is used to describe activities and events in the environment that affects the system.

8. Define discrete and continuous system?

- A discrete system is one in which the state variables change only at a discrete set of points in time.
- A continuous system is one in which the state variables change continuously over time.

9. Define model.

A model is defined as a representation of a system for the purpose of studying the system.

10. What are the steps in simulation study.

- Problem formulation
- Setting of objectives and overall project plan.
- Model conceptualization
- Data collection
- Model translation
- verified?
- Validated?
- Experimental design
- Production runs and analysis
- More runs
- Documentation and reporting
- Implementation

11. What are the properties of random numbers?

- The set of random numbers is uniformly distributed between 0 and 1
- Successive random members are independent

12. Define mathematical model?

A mathematical model uses symbolic notation and mathematical equations to represent a system.

13. Define dynamic simulation model.

Dynamic simulation models represent systems as they change over time.

14. Define deterministic model?

Deterministic model have a known set of inputs which will result in a unique set of outputs.

15. Define stochastic simulation model.

A stochastic simulation model has one or more random variables as output.

16. Define analytical methods?

Analytical methods employ the deduction reasoning of mathematical to “solve” the model.

17. How will you implement data into simulation?

It has four phases they are :

- Is a period of discovery or orientation
- Is related to model building and data collection
- Running the model
- Implementation

18. Define simulation clock?

The simulation clock is used to track simulated time.

19. What are the steps involved in simulation?

- Determine the characteristics of each of the simulation.
- Construct a simulation table
- For each repetition I, generate a value of the p inputs, and evaluate the function, calculating a value of the response y_i .

20. Describe the uses of simulation

- Simulation enables the study of the internal interaction of a complex system.
- Informational, organizational and environmental changes can be simulated.
- Simulation can be used to verify analytic solutions.

UNIT II

1. Define discrete random variables?

Let X be a random variable. If the number of possible values of X is finite, or countably infinite, X is called a discrete random variable. The possible values of X may be listed as x_1, x_2, \dots . In the finite case the list terminates. In the countably infinite case the list continues indefinitely.

2. What is continuous random variables?

If the range space R_x of the random variable X is an interval or a collection of intervals, X is called a continuous random variable. For a continuous random variable X the probability that X lies in the interval $[a, b]$ is given by

$$P(a \leq X \leq b) = \int_a^b f(x) dx$$

3. Define probability density function?

The function $f(x)$ is called the probability density function (pdf) of the random variable X . The pdf satisfies the following conditions:

- (a) $f(x) \geq 0$ for all x in R_x .
- (b) $\int_{R_x} f(x) dx = 1$.
- (c) $f(x) = 0$ if x is not in R_x .

4. What is cumulative distribution function?

The cumulative distribution function (cdf), denoted by $F(x)$, measures the probability that the random variable X assumes a value less than or equal to x , that is, $F(x) = P(X \leq x)$.

If X is discrete, then

$$F(x) = \sum_{\substack{\text{all} \\ x_i \leq x}} p(x_i)$$

If X is continuous, then

$$F(x) = \int_{-\infty}^x f(t) dt$$

5. What are the properties of cumulative distribution function?

Properties of cumulative distribution function are listed here:

- (a) F is a nondecreasing function. If $a < b$, then $F(a) \leq F(b)$.
- (b) $\lim_{x \rightarrow \infty} F(x) = 1$.
- (c) $\lim_{x \rightarrow -\infty} F(x) = 0$.

6. What is expectation?

An important concept in probability theory is that of the expectation of a random variable. If X is a random variable, the expected value of X , denoted by $E(x)$, for discrete and continuous variables is defined as follows:

$$E(X) = \sum_{\text{all } i} x_i p(x_i) \quad \text{if } X \text{ is discrete and}$$

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx \quad \text{if } X \text{ is continuous}$$

7. Define mode?

The mode is used in describing several statistical models which appear in this chapter. In the discrete case, the mode is the value of the random variable that occurs most frequently. In the continuous case, the mode is the value at which the pdf is maximized. The mode may not be unique; if the model value occurs at two values of the random variable, the distribution is said to be bimodal.

8. What is Poisson distribution?

The Poisson distribution describes many random processes quite well and is mathematically quite simple. It was introduced in 1837 by S.D. Poisson in a book concerning criminal and civil justice matters. The title of this rather old text is *Recherches sur la probabilité des jugements en matière criminelle et en matière civile*. Evidently, the rumor handed down through generations of probability-theory professors concerning the origin of the Poisson distribution is just not true. Rumor has it that the Poisson distribution was first used to model deaths from the kicks of horses in the Prussian Army.

9. Define geometric distribution?

The geometric distribution is related to a sequence of Bernoulli trials; the random variable of interest, X , is defined to be the number of trials to achieve the first success. The distribution of X is given by

$$P(x) = \begin{cases} q^{x-1}, & x=1,2,\dots \\ 0 & \text{otherwise} \end{cases}$$

10. What is gamma distribution?

A function used in defining the gamma distribution is the gamma function, which is defined for all $\beta > 0$ as

$$\Gamma(\beta) = \int_0^{\infty} x^{\beta-1} e^{-x} dx$$

11. Give the mean and variance of Erlang distribution?

The expected value of the sum of random variables is the sum of the expected value of each random variable. Thus,

$$E(X) = E(X_1) + E(X_2) + \dots + E(X_k)$$

The expected values of the exponentially distributed X_j are each given by $1/k\theta$.

Thus,

$$E(X) = \frac{1}{k\theta} + \frac{1}{k\theta} + \dots + \frac{1}{k\theta} = \frac{1}{\theta}$$

If the random variables X_j are independent, the variance of their sum is the sum of the variances, or

$$V(X) = \frac{1}{(k\theta)^2} + \frac{1}{(k\theta)^2} + \dots + \frac{1}{(k\theta)^2} = \frac{1}{k\theta^2}$$

12. What are the special properties of normal distribution?

Some of the special properties of the normal distribution are listed here:

1. $\lim_{x \rightarrow -\infty} f(x) = 0$ and $\lim_{x \rightarrow \infty} f(x) = 0$; the value of $f(x)$ approaches zero as x approaches negative infinity and, similarly, as x approaches positive infinity.
2. $f(\mu - x) = f(\mu + x)$; the pdf is symmetric about μ .
3. The maximum value of the pdf occurs at $x = \mu$. (Thus, the mean and mode are equal.)

13. Define Empirical distribution?

An empirical distributions may be either continuous or discrete in form. It is used when it is impossible or unnecessary to establish that a random variable has any particular known distribution. One advantage of using a known distribution in simulation is the facility with which parameters can be modified to conduct a sensitivity analysis.

14. What are the characteristics of queueing system?

The key elements of a queueing system are the customers and servers. The term "customer" can refer to people, machines, trucks, mechanics, patients, pallets, airplanes, e-mail, cases, orders, or dirty clothes—anything that arrives at a facility and requires service.

15. Give the queueing notation for parallel to server system?

The queueing notation for parallel to server system is based on the format A/B/c/N/K. These letters represent the following system characteristics:

A represents the interarrival-time distribution.

B represents the service-time distribution.

c represents the number of parallel servers.

N represents the system capacity.
 K represents the size of the calling population.

16. What is lognormal distribution?

A random variable X has a lognormal distribution if its pdf is given by

$$f(x) = \begin{cases} \frac{1}{\sqrt{2\pi\sigma x}} \exp\left[-\frac{(\ln x - \mu)^2}{2\sigma^2}\right], & x > 0 \\ 0, & \text{otherwise} \end{cases}$$

where $\sigma^2 > 0$. The mean and variance of a lognormal random variable are

$$E(X) = e^{\mu + \sigma^2/2}$$

$$V(X) = e^{2\mu + \sigma^2} (e^{\sigma^2} - 1)$$

17. What is triangular distribution?

A random variable X has a triangular distribution if its pdf is given by

$$f(x) = \begin{cases} \frac{2(x-a)}{(b-a)(c-a)}, & a \leq x \leq b \\ \frac{2(c-x)}{(c-b)(c-a)}, & b < x \leq c \\ 0, & \text{elsewhere} \end{cases}$$

where $a \leq b \leq c$.

18. What is Weibull distribution?

The random variable X has a Weibull distribution if its pdf has the form

$$f(x) = \begin{cases} \frac{\beta}{\alpha} \left(\frac{x-v}{\alpha}\right)^{\beta-1} \exp\left[-\left(\frac{x-v}{\alpha}\right)^\beta\right], & x \geq v \\ 0, & \text{otherwise} \end{cases}$$

19. Give the mean and variance for uniform distribution?

A random variable X is uniformly distributed on the interval (a,b) if its pdf is given by

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \end{cases}$$

0, otherwise

The mean and variance of the distribution are given by

$$E(X) = \frac{a + b}{2}$$

and

$$V(X) = \frac{(b - a)^2}{12}$$

20. Define exponential distribution, give the mean and variance.

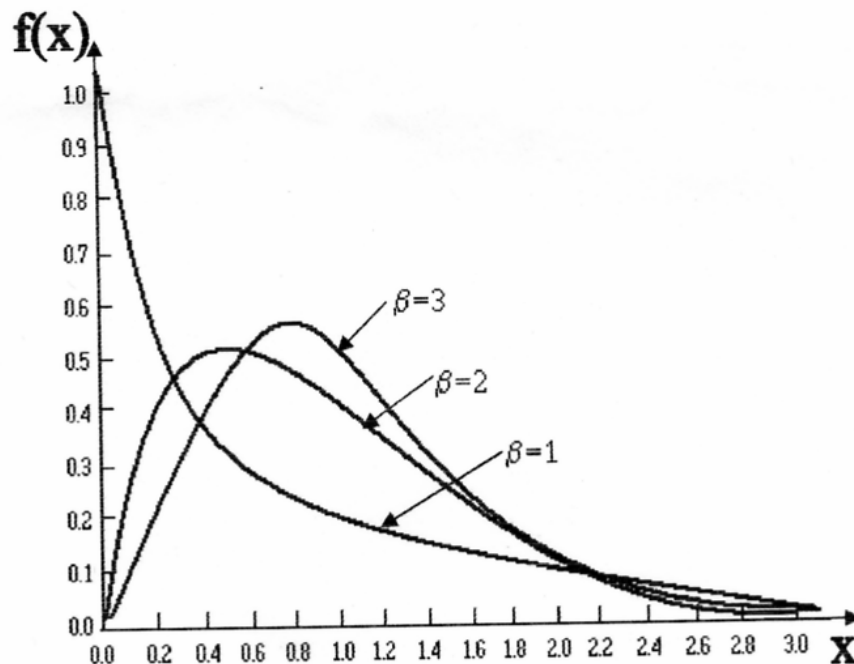
A random variable X is said to be exponentially distributed with parameter $\lambda > 0$ if its pdf is given by

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & \text{elsewhere} \end{cases}$$

The exponential distribution has mean and variance given by

$$E(X) = \frac{1}{\lambda} \quad \text{and} \quad V(X) = \frac{1}{\lambda^2}$$

21. Give the graphical representation of probability density function for the gamma distribution, when $\theta=1$.



UNIT III: (2 MARKS):

1. What are the statistical properties of random numbers?

- Uniformity
- Independence

2. State the errors occurred in generating pseudo random numbers?

- The generated numbers may be uniformly distributed.
- The generated numbers may be discrete valued
- The mean may be too high or too low
- The variance may be too high or too low

3. List some of the errors occurring in random numbers?

- The routine must be fast
- The routine should be portable
- The routine should have sufficient long cycle
- The random numbers should be replicable

4. Explain linear congruential method?

This method explains the relationship as

$$X_{i+1} = (a * X_i + c) \text{ mod } m, \quad i=0,1,2,\dots$$

Where X_i is the seed, a a constant multiplier c is the increment and m is the modulus

5. Define frequency test?

Uses the kolmogorov-simirnov or the chi-square test to compare the distribution of the set of the numbers generated to the uniform distribution.

6. Define runs test?

Tests the runs up and down or the runs above and below the mean by comparing the actual values to expected values the statistic for comparison is the chi-square.

7. Define Autocorrelation test?

Tests the correlation between numbers and compares the sample correlation to the expected correlation of zero.

8. Define Gap test?

Counts the number of digits that appears between repetitions of a expected size of gaps.

9. Define poker test?

Treats numbers grouped together as a poker hand . then the hands obtained are compared to what is expected using the chi-square test.

10. Define Kolmogorov-simirnov test?

This test compares the continuous cdf, $F(x)$, of the uniform distribution to the empirical cdf, $S_n(x)$, of the sample of N observations.

By definition ,

$$F(x) = x, \quad 0 \leq x \leq 1.$$

11. State the probability density function of inverse exponential distribution?

It can be

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0, \\ 0 & \text{otherwise} \end{cases}$$

12. Give the mean for the exponential distribution?

The mean can be

$$E(x) = 1/\lambda.$$

13. Define inverse uniform distributions?

Consider the random variable X that is uniformly distributed on the interval [a,b]. A reasonable guess for generating X is given by

$$X = a + (b-a) R.$$

14. Define the inverse triangular distribution?

It can be given as

$$f(x) = \begin{cases} x & 0 \leq x \leq 1, \\ 2-x & 1 < x \leq 2, \\ 0 & \text{otherwise} \end{cases}$$

15. Define inverse empirical continuous distribution?

If the modeler has been unable to find a theoretical distribution that provides a good model for the input data, then it may be necessary to use the empirical distribution of the data.

16. Give the pdf of random variates?

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

17. Give the cdf of the random variates?

It can be

$$F(x) = \begin{cases} 0, & x < 0 \\ x, & 0 \leq x \leq 1 \\ 1, & x > 1 \end{cases}$$

18. Write some of the consequences of the uniformity and independence?

- If the interval is divided into n classes , or subintervals of equal length the expected number of observations in each intervals in N/n
- The probability of observing variables s independent of previous value drawn.

19. Define chi-square test?

$$X = \sum_{E_i} \frac{(O_i - E_i)^2}{E_i}$$

O_i = observed frequency

E_i =expected frequency.

20. Explain the 3 possibilities for the three digit number in poker test?

- The individual numbers can be different.
- The individual numbers can be same
- Their can be one pair of like digits

UNIT IV
TWO MARKS QUESTIONS AND ANSWERS

1. What are the languages used for simulation?
C++, GPSS/H , Arena etc.
2. How does c++ structure help simulation?
C++ programming language is a structured object oriented language. This language is used more extensively in simulation.
3. What are the components are used in models written in C++?
Clock
Initializing subroutine
Event subroutines
Main program
Report generator
4. What is the main advantage of Gpss language
It is highly structured special purpose simulation language based on the process interaction approach and oriented toward queuing systems
5. What are the main components in GPSS/H used for simulating a model?
Built in file
Screen I/O
Use of arithmetic expression as a block operand.
6. What is the use of GENERATE block?
The GENERATE block represents the arrival event, with the inter arrival times specified by TVEXPO(1,&IAT).RVEXPO stand for “random variable exponential distributed.
7. What is the use of QUEUE block?
The QUEUE block with a queue named systime. It should be noted that the queue block is not needed for queues pr waiting lines to form in GPSS
8. What is the use pf TEST block?
The test block that checks to see if the tome in the system M! is greater than or equal to 4 min.in GPSS/H the maxim is “if true, pass through.
9. What is the use of start control block?
The start control statement controls simulation execution
10. What is the use of QUEUE block with name LINE?
QUEUE block with name LINE begins data collection for waiting line before the cashier.

**NOORUL ISLAM COLLEGE OF ENGINEERING
DEPARTMENT OF SOFTWARE ENGINEERING
FOURTH YEAR M.SC.SOFTWARE ENGG.(5-YEARS)
XCS472-MODELLING AND SIMULATION**

**UNIT-V
TWO MARKS QUESTION AND ANSWERS**

1. What is mean by GIGO?

“GIGO” or “garbage-in, garbage-out,” is a basic concept in computer science and it applies equally in the area of discrete system simulation.

2. What is mean by Histogram and write the construction steps?

A frequency distribution or histogram is useful in identifying the shape of a

you frequency
Distribution

A histogram is constructed as follows:

1. Divide the range of the data into intervals.
2. Label the horizontal axis to conform to the intervals selected.
3. Determine the frequency of occurrences within each interval.
4. Label the vertical axis so that the total occurrences can be plotted for each interval.
5. Plot the frequencies on the vertical axis.

3. What the two types of the parameter estimation?

1. Preliminary statistics: Sample Mean and Sample variance
2. Suggested Estimators.

4. What is mean by Goodness-of-Fit tests?

Goodness-of-fit is a hypothesis test provide helpful guidance for evaluating the suitability of a potential input model.

5. What are types of test used in Goodness-of-fit test?

The two types of test used in Goodness-of-fit test are:

- * Kolmogorov-Smirnov test
- * Chi-square test

6. What is the test statistics for Chi-square test?

The test statistics ios given by

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

Where O_i is the observed frequency in the i th class interval and E_i is the expected frequency in that class interval.

7. What are number of ways to obtain information about a process even if the data not available?

- Engineering data
- Expert option
- Physical or conventional limitations
- The nature of the process.

8. Define Covariance and Correlation.

Let X_1 and X_2 be two random variables and let $\mu_i = E(X_i)$ and $\sigma_i^2 = \text{var}(X_i)$ be the mean and variance of X_i respectively. The Covariance and Correlation are measures of the linear dependence between X_1 and X_2 .

9. What are the steps involved in model building?

-Observing the real system and the interactions among its various components and collecting data on its behavior.

-Construction of a conceptual model—a collection of assumptions on the components and the structure of the system, plus hypothesis on the values of model input parameters.

-The translation of the operational model into a computer recognizable form—the computerized model.

10. Define terminating simulation:

A terminating simulation is one that runs for some duration of time T_E , where E is a specified event which stops the simulation.

11. What are the suggestion given for verification process?

1. Have the computerized representation checked by someone other than its developer.

2. Make a flow diagram which includes each logically possible action a system can take when an event occurs.

3. Closely examine the model output for reasonableness under a variety of settings of the input parameters.

4. Graphical interfaces are recommended for accomplishing verification and validation.

12. What is mean by p-value?

The p-value is the significance level at which one would just reject H_0 for the given value of the test statistics.

13. What is the use of the Kolmogorov-Smirnov test?

The Kolmogorov-Smirnov test is particularly useful when sample sizes are small and when no parameters have been estimated from the data.

14. What is the purpose of Verification of simulation model?

The purpose is to assure that the conceptual model is reflected accurately in the computerized representation.

15. Give the mean and variance for preliminary statistics:

sample mean(\bar{X}) is defined by

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$

and the sample variance, S^2 , is defined by

$$S^2 = \frac{\sum_{i=1}^n X_i^2 - n\bar{X}^2}{n - 1}$$

16. Define Suggested Estimators?

Numerical Estimates of the distribution parameters are needed to reduce the family distributions to a specific distribution and to test the resulting hypothesis.

17. What are steps in the development of a useful model of input data?

1. Collect the data from the real system of interest.
2. Identify a probability distribution to represent the input process.
3. Choose parameter that determine a specific instance of the distribution

family.

4.

Evaluate the chosen distribution and the associated parameters for goodness-of-fit.

18. How will you Select the family of distributions?

A family of distribution is selected on the basis of what might arise in the context being investigated along with the shape of the histogram.

19. Define Quantile-quantile Plots:

A quantile-quantile plot is useful tool for evaluating distribution fit that does not suffer from these problems.

If X is a random variable with cdf F , then the q -quantile of x is that value γ such that $F(\gamma) = P(X \leq \gamma) = q$, for $0 < q < 1$.

20. Give some suggested estimators for distributions often used in simulation:

- Poisson α
- Exponential λ
- Gamma β, θ

**NOORUL ISLAM COLLEGE OF ENGINEERING
DEPARTMENT OF SOFTWARE ENGINEERING
FOURTH YEAR M.SC.SOFTWARE ENGG.(5-YEARS)
XCS472-MODELLING AND SIMULATION**

**UNIT-1
SIXTEEN MARKS QUESTION AND ANSWERS**

1-Describe the advantage and disadvantage of simulation?

Advantages of simulation:-

- New policies can be explored without disrupting ongoing operations of the real system.
- New hardware design can be tested without committing resources and their acquisition.
- Hypotheses about how or why certain phenomena occur can be tested for feasibility.
- Insight can be obtained about the interaction of variables.

Disadvantages of simulation:-

- Model building requires special training.
- Simulation results may be difficult to interpret.
- Simulation modeling and analysis can be time consuming and expensive.
- Simulation is used in some cases when an analytical solution is possible.

Overcome the disadvantages of simulation:-

- Models have the generic the simulators or templates.
- Many simulation software vendors have developed output analysis capabilities.
- Simulation can be performed faster today than the yesterday.
- Closed-form models are not able to analyze most of the complex systems that are encountered in practice.

2-What are the types if models in simulation?

Mathematical model:-

A mathematical model uses symbolic notation and mathematical equations to represent a system.

Static model:-

A static model represents a system at a particular point in time.

Simulation model:-

Dynamic simulation models represent systems as they change over time.

Deterministic model:-

Deterministic model have a known set of inputs which will result in a unique set of outputs.

Stochastic simulation model:-

A stochastic simulation model has one or more random variables as output.

Discrete-Event System Simulation:-

Analytical methods:-

Analytical methods employ the deduction reasoning of mathematical to “solve” the model.

Numerical methods:-

Numerical methods employ computational procedures to “solve” mathematical model.

3- What are the steps in simulation study?

Problem formulation:-

If a problem statement is being developed by the analyst, it is important that the policy makers understand and agree with the formulation.

Setting of objectives and overall project plan:-

The objectives indicate the questions to be answered by the simulation.

Model conceptualization:-

The construction of a model of a system is probably as much art as science.

Data collection:-

There is a constant interplay between the construction of the model and the collection of the needed input data.

Model translation:-

The model must decide whether to program the model in a simulation language such as GPSS/H.

Verified?:-

Verification pertains to the computer program prepared for the simulation model.

Validated?:-

Validation is the determination that a model is an accurate representation of the real system.

Experimental design:-

The alternatives that are to be simulated must be determined.

Production runs and analysis:-

This is used to estimate measures of performance of the system designs that are being simulated.

More runs:-

Based on the analysis of runs that have been completed.

Documentation and reporting:-

There are two types of documentations:

- Program
- process

Implementation:-

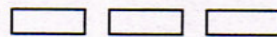
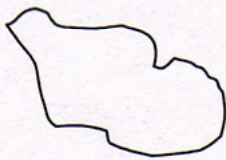
It has four phases they are:

- Is a period of discovery or orientation
- Is related to model building and data collection
- Running the model
- Implementation

4-Write about Simulation of queuing system?

A queuing system is described by its

- Calling population,
- The nature of the arrivals
- The service mechanism
- The system capacity
- The queuing discipline (FIFO)

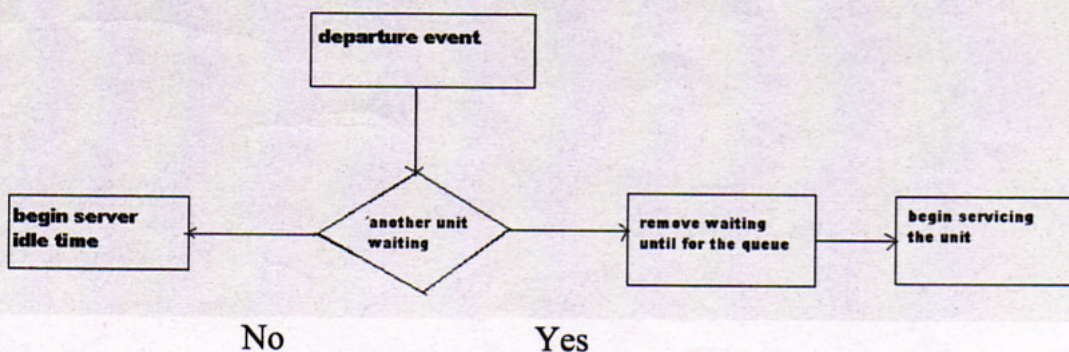


Calling population

wait line

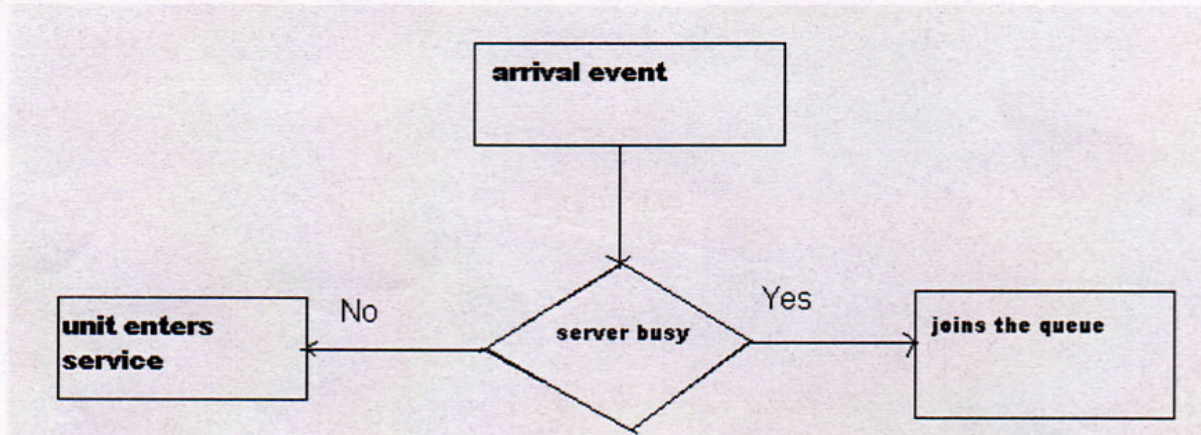
server

- The calling population is infinite
- Complex situation may occur
- The simulation clock is used to track simulated time



Two possible events affect the systems are:

- Arrival event
- Departure event



5-The news paper seller's problem

The paper seller buys the paper for 33 cents and sells them for 50 cents. Newspapers not sold at the end of the day are sold as 5 cents. The newspapers are purchased in bundles of 10 so there are three types they are Good, Fair and Poor with probabilities 0.35, 0.45, 0.20. The problem is determined the optimal number of paper of the newspaper seller should purchase. Simulate it for 20 days.

Distribution of newspaper demand table:-

Demand	Good	Fair	Poor
40	0.03	0.10	0.44
50	0.05	0.18	0.22
60	0.15	0.40	0.16
70	0.20	0.20	0.12
80	0.35	0.08	0.06
90	0.15	0.04	0.00
100	0.07	0.00	0.00

Random digit assessment:-

Type	Probability	Cum. Probability	Rand. Digit
Good	0.35	0.35	01-35
Fair	0.45	0.80	36-80
Poor	0.20	1.00	81-00

The profits are calculated by using the relation:-

$$\text{Profit} = [(\text{revenue from sale}) - (\text{cost of news paper}) - (\text{less profit from excess paper}) + (\text{scale of scrap paper})]$$

Demand	Cum. Probability			Random Digit Ass		
	Good	Fair	Poor	Good	Fair	Poor
40	0.03	0.10	0.44	01-03	01-10	01-44
50	0.08	0.28	0.66	04-08	11-28	45-66
60	0.23	0.68	0.82	09-23	29-68	67-82
70	0.43	0.88	0.94	24-43	69-88	83-94
80	0.78	0.96	1.00	44-78	89-96	95-00
90	0.93	1.00	1.00	79-93	97-00	---
100	1.00	1.00	1.00	94-00	---	---

Simulation table for purchase of 70 newspapers:-

Day	Rand. Dig	Type of day	Rand. dig for demand	Demand	Revenue from sale	Loss profit	Sale of scrap	Daily profit
1	94	Poor	80	60	\$30.00	-----	0.50	7.40
2	77	Fair	20	50	\$25.00	-----	1.00	2.90
3	43	Fair	98	90	\$35.00	\$3.40	-----	8.50
-----	-----	-----	-----	-----	-----	-----	-----	-----
20	78	Fair	96	80	\$35.00	\$1.70	-----	10.20
					\$645.00	\$13.60	5.50	174.90

UNIT:II (16 MARKS)

1. Explain i. Discrete random variable
ii. Continuous random variable
iii .Cumulative distribution function.

Discrete Random Variable:

Let X be a random variable If the number of possible values is infinity or countably infinite,X is called discrete random variable.

Continuous random variable:

If the range of X in the interval or a collection of intervals X is called continuous random variable.

Cumulative distribution function:.

It is denoted by $f(x)$ measures the probability the variable X assumes a value less than or equal to x

That is,

$$F(x) = P(X \leq x);$$

2. Explain i)Expectation
ii)Mode

Expectation:

If X is a random variable the expected value of X, denoted by $E(x)$ for discrete and continuous

And also an example program

Mode:

The mode is used in describing several statistical method.In discrete case mode occurs most frequently, in continuous case, the mode is the value at which pdf is maximized.

3. Explain various statistical models?
They are

- Queuing systems
- Inventory model
- Reliability and Maintainability
- Limited Data
- Other distributions

Queuing Systems:

It describes the waiting problem. And various problems were solved by simulation.

Inventory model:

It contains 3 variables

- The number of units
- The time between demands
- The Lead time

Reliability and maintainability:

Time to failure has been modeled with numerous distributions, including exponential gamma and weibull.

Limited Data:

Data collection has been complete.3distribution has been applied.

Other distributions:

The binomial and Bernoulli are two discrete distributions.

4. Explain Discrete Distributions?

Bernoulli trials and Bernoulli distributions:

Let it contains n trails, of which it denotes success or failure.

Binomial Distribution:

The random variable X that denotes the number of success in n Bernoulli trials.

Geometric distribution:

It is related to sequence of Bernoulli trials X is defined to be the number of trials to achieve the first success.

Poisson distribution:

It describes various random processes quite well and simple.

5.Explain the characteristics of queuing system?

The calling population:

The population is infinite.

System capacity:

There is a limit of customers.

The arrival process:

The arrival is specified by means of inter arrival time of successive customers.

Queue Behavior and Queue Discipline:

Queue behavior refers to customer action and queue waiting for service to begin.

Service times and Service Mechanisms:

They may be random or constant duration

**NOORUL ISLAM COLLEGE OF ENGINEERING
DEPARTMENT OF SOFTWARE ENGINEERING
FOURTH YEAR M.SC.SOFTWARE ENGG.(5-YEARS)
XCS472-MODELLING AND SIMULATION**

**UNIT-III
SIXTEEN MARKS QUESTION AND ANSWERS**

1) How will you generate pseudo Random Numbers?

Pseudo means false, so false random numbers are being generated. The goal of any generation scheme, however, is to produce a sequence of numbers between zero and 1 which simulates, or imitates, the ideal properties of uniform distribution and independence as closely as possible. When generating pseudo random numbers, certain problems or errors can occur.

1. The generated numbers may not be uniformly distributed.
2. The generated numbers may be discrete-valued instead of continuous valued.
3. The mean of the generated numbers may be too high or too low.
4. The variance of the generated numbers may be too high or too low.
5. There may be dependence.

Usually, random numbers are generated by a digital computer as part of the simulation. Numerous methods can be used to generate the values. In selecting among these methods, or routines, there are a number of important considerations:

1. The routine should be fast. Individual computations are inexpensive, but simulation may require many hundreds of thousands of random numbers. The total cost can be managed by selecting a computationally efficient method of random number generation.
2. The routine should be portable to different computers, and ideally to different programming languages. This is desirable so that the simulation program produces the same results wherever it is executed.
3. The routine should have a sufficiently long cycle. A special case of cycling is degenerating. A routine degenerates when the same random numbers appear repeatedly.
4. The random numbers should be replicable. Given the starting point, it should be possible to generate the same set of random numbers, completely independent of the system that is being simulated.
5. Most important, and as indicated previously, the generated random numbers should closely approximate the ideal statistical properties of uniformity and independence.

Inventing techniques that seem to generate random numbers is easy; inventing techniques that really do produce sequences that appear to be independent, uniformly distributed random numbers is incredibly difficult.

2) Explain the frequency test:

A basic test that should always be performed to validate a new generator is the test of uniformity. Two different methods of testing are available.

a) **The Kolmogorov-Smirnov test:**

This test compares the continuous cdf, $F(x)$, of the uniform distribution to the empirical cdf, $S_N(x)$, of the sample of N observations. By definition,

$$F(x) = x, 0 \leq x \leq 1$$

If the sample from the random number generator is R_1, R_2, \dots, R_N , then the empirical cdf, $S_N(x)$, is defined by

$$S_N(x) = \frac{\text{Number of } R_1, R_2, \dots, R_N \text{ which are } \leq x}{N}$$

The Kolmogorov-Smirnov test is based on the largest absolute deviation between $F(x)$ and $S_N(x)$ over the range of the random variable. That is, it is based on the statistic

$$D = \max[F(x) - S_N(x)]$$

b) **The chi-square test:**

The chi-square test uses the sample statistic

$$\chi_0^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}$$

Where O_i is the observed number in the i th class, E_i is the expected number in the i th class, and n is the number of the classes.

For the uniform distribution, E_i is the expected number in each class is given by

$$E_i = N/n$$

For equally spaced classes, Where N is the total number of observations

3. Explain Run test:

Runs up and runs down: Consider a generator that provided a set of 40 numbers in the following sequence:

0.08	0.09	0.23	0.29	0.42	0.55	0.58	0.72	0.89	0.91
0.11	0.16	0.18	0.31	0.41	0.53	0.71	0.73	0.74	0.84
0.02	0.09	0.30	0.32	0.45	0.47	0.69	0.74	0.91	0.95
0.12	0.13	0.29	0.36	0.38	0.54	0.68	0.86	0.88	0.91

Both the Kolmogorov-smirnov test and the Chi-Square test would indicate that the numbers are informally distributed.

The run test examine the arrangements of numbers in a sequence to test the hypothesis of independence.

Consider the following sequence generated by tossing a coin 10 times.

H T T H H T T T H T

There are three mutually exclusive outcomes, or events, with respect to the sequence. A run is defined as a succession of similar events preceded and followed by a different event.

If a is the total number of runs in a truly random sequence, the mean and variance of a are given by

$$\mu_a = 2N - 1/3$$

$$\sigma_a^2 = 16N - 29/90$$

For $N > 20$, the distribution of a is reasonably approximated by a normal distribution, $N(\mu_a, \sigma_a^2)$.

The test statistic is

$$Z_0 = (a - \mu_a) / \sigma_a$$

Runs above and below the mean:

The test for runs up and runs down is not completely adequate to assess the independence of a group of numbers.

Consider the following 40 numbers:

0.63	0.84	0.75	0.18	0.13	0.92	0.57	0.77	0.30	0.71
0.42	0.05	0.78	0.74	0.68	0.03	0.18	0.51	0.10	0.37

The pluses and minuses are as follows:

- ++--+++++---++++--+-

The maximum number of runs is $N = n_1 + n_2$

The mean and variance is given by

$$\mu_b = \frac{2n_1n_2}{N} + \frac{1}{2}$$

$$\Sigma_b^2 = \frac{2n_1n_2(2n_1n_2 - N)}{N^2(N-1)}$$

4) Gap Test

The gap test is used to determine the significance of the interval between the recurrences of the same digit. A gap of length x occurs between the recurrences of some specified digit. The following example illustrates the length of gaps associated with the digit 3:

4,	1,	<u>3</u> ,	5,	1,	7
2,	8,	2,	0,	9	1

3, 5, 2, 7, 9, 4
 1, 6, 3, 3, 9, 3

To facilitate the analysis, the digit 3 has been underlined. There are five 3's in the list. Thus only four gaps can occur. The first gap is of length 10, the second gap is of length 7, and so on. The frequency of the gaps is of interest. The general formula is

$$P(t \text{ followed by exactly } x \text{ non-}t \text{ digits}) = (0.9)^x(0.1), \quad x=0,1,2,\dots$$

In the example above, only the digit 3 was examined. However, to fully analyze a set of numbers for independence using the gap test, every digit, 0,1,2,...,9, must be analyzed.

The theoretical frequency distribution for randomly ordered digits is given by

$$P(\text{gap} \leq x) = F(x) = 0.1 \sum_{n=0}^x (0.9)^n = 1 - 0.9^{x+1} \quad \longrightarrow \quad (1)$$

The procedure for the test follows the steps below.

Step1: Specify the cdf for the theoretical frequency distribution given by equation(1) based on the selected class interval width.

Step2: Arrange the observed sample of gaps in a cumulative distribution with these same classes.

Step3: Find D, the maximum deviation between $F(x)$ and $S_N(x)$.

Step4: Determine the critical value, D_{α} , and the sample size N.

Step5: If the calculated value of D is greater than the tabulated value of D_{α} , the null hypothesis of independence is rejected.

It should be noted that using the Kolmogorov-Smirnov test when the underlying distribution is discrete results in a reduction in the Type I error, α , and an increase in type II error, β .

5) Explain exponential distribution using Inverse Transform Technique:

The exponential distribution, has probability density function (pdf) given by

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

and cumulative distribution function function (cdf) given by

$$F(x) = \int_{-\infty}^x f(t)dt = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

The parameter λ can be interpreted as the mean number of occurrences per time unit. Notice that for any I

$$E(X_i) = 1/\lambda$$

So that $1/\lambda$ is the mean interarrival time. The inverse transform technique can be utilized, at least in principle, for any distribution, but it is most useful when the cdf, $F(x)$, is of such simple form that its inverse, F^{-1} , can be easily computed. A step-by-step procedure for the inverse transform technique, illustrated by the exponential distribution, is as follows:

Step1: Compute the cdf of the desired random variable X. For the exponential distribution, the cdf is $F(x) = 1 - e^{-\lambda x}$, $x \geq 0$.

Step2: Set $F(x) = R$ on the range of X. For the exponential distribution, it becomes $1 - e^{-\lambda x} = R$ on the range $x \geq 0$.

Step3: Solve the equation $F(X) = R$ for X in terms of R. For the exponential distribution, the solution proceeds as follows:

$$\begin{aligned}1 - e^{-\lambda X} &= R \\e^{-\lambda X} &= 1 - R \\-\lambda X &= \ln(1 - R) \\X &= -1/\lambda \ln(1 - R)\end{aligned}$$

Step4: Generate uniform random numbers R_1, R_2, R_3, \dots . And compute the desired random variates by

$$X_i = F^{-1}(R_i)$$

For the exponential case, $F^{-1}(R) = (-1/\lambda)\ln(1-R)$, so that

$$X_i = -1/\lambda \ln(1 - R_i)$$

For $i = 1, 2, 3, \dots$ one simplification that is usually employed in above equation is to replace $1 - R_i$ by R_i yield

$$X_i = -1/\lambda \ln R_i$$

Which is justified since both R_i and $1 - R_i$ are uniformly distributed on $(0,1)$.

UNIT - IV

1> Selection of simulation software

ans: there are many features that are relevant when selecting simulation software. the following advice are used while selecting simulation software:

a) Do not focus on a single issue such as ease of use. consider the accuracy and level of detail obtainable, ease of learning, vendor support and applicability to your problems.

b) Execution speed is important

c) Beware of administration claims and demonstrations.

d) Ask the vendor to solve a small version of your problem.

e) Beware the "checklist" with "yes" and "no" as the entries.

f) Simulation users ask if the simulation model can link to use code or routines written in external languages such as C, C++, FORTRAN.

g) There may be a significant trade-off between the graphical model building environments and ones based on simulation language.

2> Simulation in c++

ans: C++ is a widely used programming language that has been used extensively in simulation. Unlike FORTRAN or C, the object orientedness of C++ does not support modular construction of large models. The following components are common to almost all models written in C++:

Clock: a variable defining simulated time.

Initializing subroutine : a routine to define the system state at time 0.

Min-time event subroutine: a routine that identifies the imminent event ie. the element of the future event list which has the smallest time stamp.

Event subroutines: for each event type, a subroutine to update system state when the event occurs.

Main program: provides overall control of the event-scheduling algorithm

Report generator: a routine that computes summary statistics from cumulative statistics and prints a report at the end of the simulation.

Fig 4.1: pg 105.

The simulation begins by setting the simulation clock to zero, initiating cumulative statistics to 0, generating any initial events and placing them on the FutureEventList, and defining the system state at time 0. The simulation program then cycles, repeatedly passing the current least time. The value of the variable clock does not change in an event routine. If the simulation is not over, control passes again to the time advance routine.

The efficiency of a simulation model in terms of computer runtime is determined to a large extent by the techniques used to manipulate the FutureEventList.

Eg: Single Server Queue Simulation in C++.

3> Simulation in GPSS

ans: GPSS is a highly structured, special purpose simulation language based on the process interaction approach and oriented toward queueing systems. There are over 40 standard blocks in GPSS. Blocks represent events, delays and other actions that affect transaction flow. The block diagram is converted to block statements, control statements are added, and the result is a GPSS model.

The 1st version of GPSS was released by IBM about 1961. GPSS is a product of Wolverine Software Corporation, Annandale. It is a flexible yet powerful tool for simulation.

GPSS/H includes built in file and screen I/O, use of an arithmetic expression as a block operand.

Eg: Single Server Queue Simulation in GPSS/H.

Fig 4.10 pg 116

The GENERATE block represents the arrival event, with the interarrival times specified by RVEXPO(1,&IAT). RVEXPO stands for "random

variable,exponential distributed", the 1 indicates the random number stream to use and &IAT indicates the mean time for the exponential distribution comes from a so called ampervariable &IAT.

The next block is a QUEUE with a queue named SYSTIME.It should be noted that the QUEUE block is not needed for queues or waiting lines to form in GPSS.

The QUEUE and DEPART block combination is not necessary for queues to be modeled,but rather is used for statistical data collection.

The next QUEUE block(with name LINE) begins data collection for waiting line before the cashier.

There is a TEST block that checks to see if thev time in the system M!,is greater than or equal to 4min.in GPSS/H the maxim is "if true,pass through".

The START control statement controls simulation execution.It starts the simulation,sets up a "termination to-go" counter the initial value its operand(&LIMIT),and control and the length of the simulation.

**NOORUL ISLAM COLLEGE OF ENGINEERING
DEPARTMENT OF SOFTWARE ENGINEERING
FOURTH YEAR M.SC.SOFTWARE ENGG.(5-YEARS)
XCS472-MODELLING AND SIMULATION**

**UNIT-V
SIXTEEN MARKS QUESTION AND ANSWERS**

I. Identifying the distribution with Data

- *Histograms
- *Selecting the family of Distributions
- *Quantile-Quantile Plots

Histograms:

A frequency distribution or histogram is useful in identifying the shape of a distribution.

It is constructed as follows:

1. Divide the range of the data into intervals
2. Label the horizontal axis to conform to the intervals selected.
3. Determine the frequency of occurrences within each interval.
4. Label the vertical axis so that the total occurrences can be plotted for each interval.
5. Plot the frequencies on the vertical axis

Selecting the family of distributions:

A family of distribution is selected on the basis of what might arise in the context being investigated along with the shape of the histogram.

Examples:

Binomial
Negative binomial
Poisson
Normal
Lognormal
Gamma

Quantile-quantile Plots:

A quantile-quantile plot is useful tool for evaluating distribution fit that does not suffer from these problems.

If X is a random variable with cdf F , then the q -quantile of x is that value γ such that $F(\gamma) = P(X \leq \gamma) = q$, for $0 < q < 1$.

2. Explain Parameters Estimation?

Two types of parameter estimation are

Preliminary statistics

Suggested estimators

Preliminary statistics:

In a number of instances, the sample mean, or the sample mean and sample variance, are used to estimate the parameters of a hypothesized distribution.

If the observations in a sample of size n are X_1, X_2, \dots, X_n , the sample mean (\bar{X}) is defined by

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$

and the sample variance, S^2 , is defined by

$$S^2 = \frac{\sum_{i=1}^n X_i^2 - n\bar{X}^2}{n-1}$$

Suggested Estimators:

Numerical Estimates of the distribution parameters are needed to reduce the family distributions to a specific distribution and to test the resulting hypothesis.

3. Explain the Goodness-Of-fit tests?

Goodness-of-fit tests provide helpful guidance for evaluating the suitability of a potential model.

Chi-Square test:

One procedure for testing the hypothesis that a random sample of size n of the random variable X follows a specific distributional form is the chi-square goodness-of-fit test.

The test statistics is given by

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

Where O_i is the observed frequency in the i th class interval and E_i is the expected frequency in that class interval

Chi-Square test with equal probabilities:

If a continuous distributional assumption is being tested, class intervals that are equal in probability rather than equal in width of interval should be used.

Probability $P_i = 1/K$

$$E_i = nP_i \geq 5$$

Kolmogorov-Smirnov Goodness-of-fit test:

The Kolmogorov-Smirnov test is particularly useful when sample sizes are small and when no parameters have been estimated from the data.

It will not take any special tables when an exponential distribution is assumed.

4.Explain the process of selecting input models without data?

There are number of ways to obtain information about a process even if data are not available:

Engineering data:

Often a product or process has performance rating provided by the manufacturer.

Expert option:

Talk to people who are experienced with the process or similar processes.

Physical or conventional limitations:

Most real processes have physical limits on performance.

The nature of the process:

It is used to specify the particular choice even when no data are available.

When data are not available the uniform,triangular, and beta distributions are often used as input models.

5.Explain Model building ,verification and validation:

Steps in model building:

-Observing the real system and the interactions among its various components and collecting data on its behavior.

-Construction of a conceptual model-a collection of assumptions on the components and the structure of the system, plus hypothesis on the values of model input parameters.

-The translation of the operational model into a computer recognizable form-the computerized model.

Verification of simulation model:

The purpose is to assure that the conceptual model is reflected accurately in the computerized representation.

Many suggestion can be given for use in the verification process.

1.Have the computerized representation checked by someone other than its developer.

2.Make a flow diagram which includes each logically possible action a system can take when an event occurs.

3.Closely examine the model output for reasonableness under a variety of settings of the input parameters.

4Graphical interfaces are recommended for accomplishing verification and validation.

