CBCS SCHEME

15CS/IS834

Eighth Semester B.E. Degree Examination, Aug./Sept. 2020

System Modeling and Simulation

Time: 3 hrs. Max. Marks: 80

Module-1

1.a. What is simulation? Explain with flowchart, the steps involved in simulation study. **(08 Marks)**

Ans:

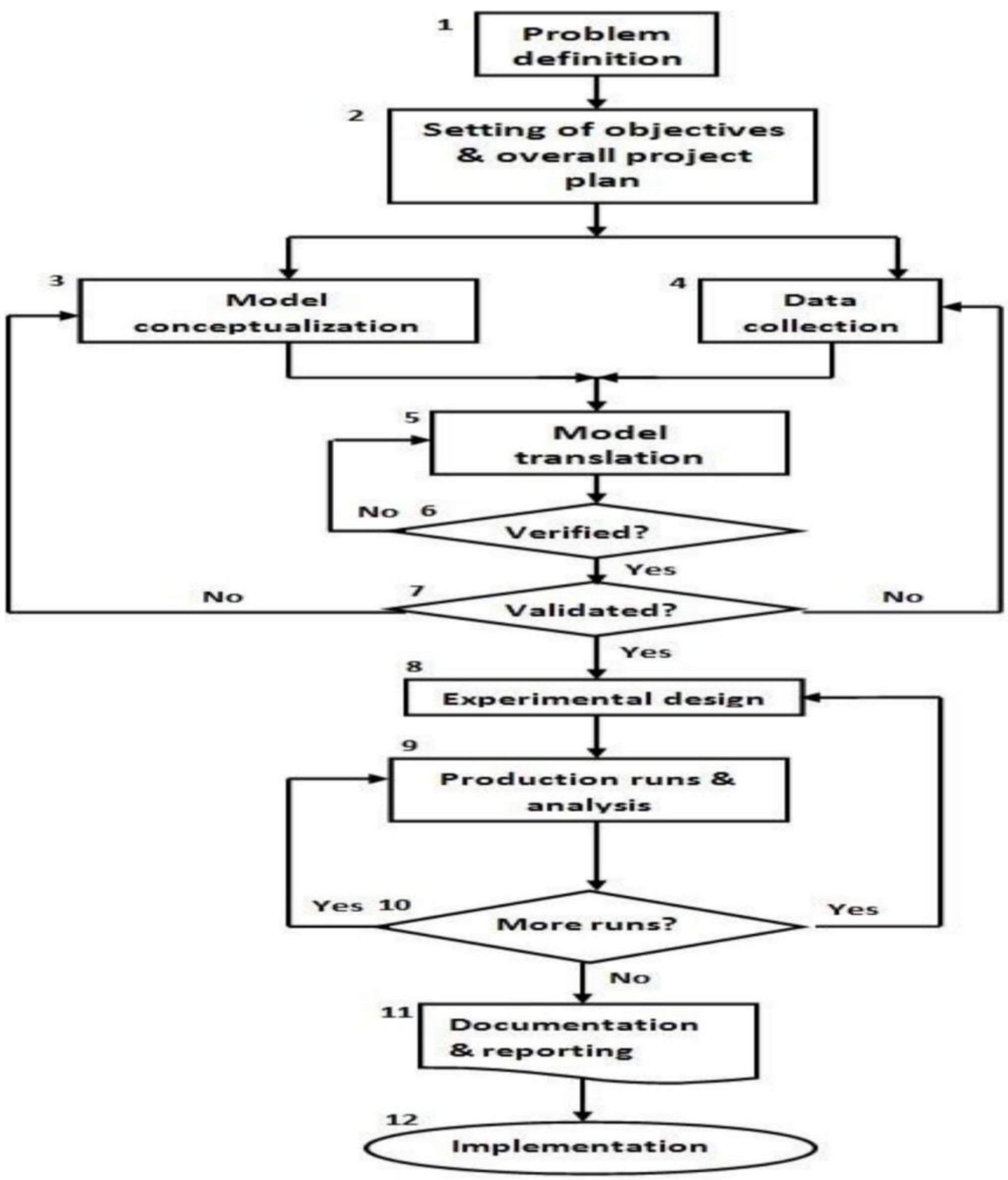


Figure 1.3 Steps in a simulation study

Steps in a Simulation study

1. Problem formulation: Every study begins with a statement of the problem, provided by policy makers. Analyst ensures it's clearly understood. If it is developed by analyst and policy makers should understand and agree with it.

2. Setting of objectives and overall project plan: The objectives indicate the questions to be answered by simulation. At this point a determination should be made concerning whether simulation is the appropriate methodology. Assuming that it is appropriate, the overall project plan should include
   i. A statement of the alternative systems
   ii. A method for evaluating the effectiveness of these alternatives
   iii. Plans for the study in terms of the number of people involved
   iv. Cost of the study
   v. The number of days required to accomplish each phase of the work with the anticipated results.

3. Model conceptualization: The construction of a model of a system is probably as much art as science. The art of modeling is enhanced by ability.
   a. To abstract the essential features of a problem.
   b. To select and modify basic assumptions that characterizes the system.
   c. To enrich and elaborate the model until a useful approximation results.

   Thus, it is best to start with a simple model and build toward greater complexity. Model conceptualizations enhance the quality of the resulting model and increase the confidence of the model user in the application of the model.

4. Data collection: As the complexity of the model changes, the required data elements may also change.

5. Model translation: Real-world systems result in models that require a great deal of information storage and computation. It can be programmed by using simulation languages or special purpose simulation software. Simulation languages are powerful and flexible. Simulation software models development time can be reduced. GPSS/HTM or special-purpose simulation software.

6. Verified: It pertains to the computer program and checking the performance. If the input parameters and logical structure are correctly represented, verification is completed.

7. Validated: It is the determination that a model is an accurate representation of the real system. Achieved through calibration of the model, an iterative process of comparing the model to actual system behavior and the discrepancies between the two.

8. Experimental Design: The alternatives that are to be simulated must be determined. For each system design, decisions need to be made concerning
   a. Length of the initialization period
   b. Length of simulation runs
   c. Number of replication to be made of each run

9. Production runs and analysis: They are used to estimate measures of performance for the system designs that are being simulated.

10. More runs: Based on the analysis of runs that have been completed. The analyst determines if additional runs are needed and what design those additional experiments should follow.

11. Documentation and reporting: Two types of documentation. Program documentation and Process documentation
    1. Program documentation: Can be used again by the same or different analysts to understand how the program operates.
    2. Process documentation: This enable to review the final formulation and alternatives, results of the experiments and the recommended solution to the problem. The final report provides a vehicle of certification.

12. Implementation: Success depends on the previous steps. If the model user has been thoroughly involved and understands the nature of the model and its outputs, likelihood of a vigorous implementation is enhanced.

The simulation model building process can be broken into four phases

I Phase: Consists of steps 1 and 2

- It is period of discovery/orientation
- The analyst may have to restart the process if it is not fine-tuned
- Recalibrations and clarifications may occur in this phase or another phase.

II Phase: Consists of steps 3,4,5,6 and 7

- a model building and data collection
- A continuing interplay is required among the steps

- Exclusion of model user results in implications during implementation.

III Phase: Consists of steps 8,9 and 10

- running the model
- Conceives a thorough plan for experimenting.
- Discrete-event stochastic is a statistical experiment.

IV Phase: Consists of steps 11 and 12

- an implementation
- Successful implementation depends on the involvement of user and every steps successful completion.

1.b. A grocery store has only one checkout counter. Customer arrives at this checkout counter at random from 1 to 5 minutes apart with equal probability. The service time varies from 1 to 6 minutes with probability 0.30, 0.25, 0.05, 0.10, 0.10 and 0.20. Develop a simulation table for 10 customers and find the following:

(i) Average waiting time of customer

(ii) Average service time

(iii) Average time between arrivals

(iv) The probability that server being idle.

Use the following set of random numbers for arrivals 84, 10, 74, 53, 17, 79, 03, 87,27.

Random digit for service time 23, 35, 65, 81, 54, 03, 87, 27, 73, 70. **(08 Marks)**

Ans:

The probability of a time-between-arrival of 1 minute is $= \frac{1}{5} = 0.20$

Table 1.1 Distribution of Time Between Arrivals

Time between Arrivals (Minutes)	Probability	Cumulative Probability	Random Digit Assignment
1	0.20	0.20	01-20
2	0.20	0.40	21-40
3	0.20	0.60	41-60
4	0.20	0.80	61-80
5	0.20	1.00	81-00

Table 1.2 Service-Time Distribution

Service Time (Minutes)	**Probability**	**Cumulative Probability**	**Random Digit Assignment**
1	0.30	0.30	01-30
2	0.25	0.55	31-55
3	0.05	0.60	56-60
4	0.10	0.70	61-70
5	0.10	0.80	71-80
6	0.20	1.00	81-00

Table 1.3 Time-Between-Arrival Determination

Customer	**Random Digits**	**Time between Arrivals (Minutes)**
1	--	--
2	84	5
3	10	1
4	74	4
5	53	3
6	17	1
7	79	4
8	03	1
9	87	5
10	27	2

Table 1.4 Service Times Generated

Customer	**Random Digits**	**Service Time (Minutes)**
1	23	1
2	35	2
3	65	4
4	81	6
5	54	2
6	03	1
7	87	6
8	27	1
9	73	5
10	70	4

Table 1.5 Simulation Table

Customer	Interarrival Time (Minutes)	Arrival Time	Service Time (Minutes)	Time service Begins	Time Service Ends	Waiting Time in Queue (Minutes)	Time Customer Spends in System (Minutes)	Idle Time of Server (Minutes)
1	--	0	1	0	1	0	1	0
2	5	5	2	5	7	0	2	4
3	1	6	4	7	11	1	5	0
4	4	10	6	11	17	1	7	0
5	3	13	2	17	19	4	6	0
6	1	14	1	19	20	5	6	0
7	4	18	6	20	26	2	8	0
8	1	19	1	26	27	7	8	0
9	5	24	5	27	32	3	8	0
10	2	26	4	32	36	6	10	0

(i) Average waiting time of customer in queue:

$$\text{Average waiting time} = \frac{\text{Total time customers wait in queue}}{\text{Total number of customers}}$$

$$= \frac{29}{10}$$

$$= 2.9 \text{ minutes}$$

(ii) Average Service Time:

$$\text{Average service time} = \frac{\text{Total service time}}{\text{Total number of customers}}$$

$$= \frac{32}{10}$$

$$= 3.2 \text{ minutes}$$

(iii) Average time between arrivals

$$\text{Average time between arrivals} = \frac{\text{Sum of all times between arrival}}{\text{Number of arrivals} - 1}$$

$$= \frac{26}{10-1}$$

$$= 2.8889 \text{ minutes}$$

(iv) The probability that server being idle.

$$\text{Probability of idle Server} = \frac{\text{Total idle time of server}}{\text{Total run time of simulation}}$$

$$= \frac{4}{36}$$

$$= 0.1111 \text{ minutes}$$

2.a. Explain major concepts in discrete event simulation. Write the flowchart for arrival and departure events. **(08 Marks)**

Ans:

Major concepts in discrete event simulation

1. System: A collection of entities (e.g., people and machines) that interact together over time to accomplish one or more goals.

2. Model: An abstract representation of a system, usually containing structural, logical, or mathematical relationships that describe a system in terms of state, entities and their attributes, sets, processes, events, activities, and delays.

3. System state: A collection of variables that contain all the information necessary to describe the system at any time.

4. Entity: Any object or component in the system that requires explicit representation in the model (e.g., a server, a customer, a machine).

5. Attributes: The properties of a given entity (e.g., the priority of a waiting customer, the routing of a job through a job shop).

6. List: A collection of (permanently or temporarily) associated entities, ordered in some logical fashion (such as all customers currently in a waiting line, ordered by "first come, first served," or by priority).

7. Event: An instantaneous occurrence that changes the state of a system (such as an arrival of a new customer).

8. Event notice: A record of an event to occur at the current or some future time, along with any associated data necessary to execute the event; at a minimum, the record includes the event type and the event time.

9. Event list: A list of event notices for future events, ordered by time of occurrence; also known as the future event list (FEL).

10. Activity: A duration of time of specified length (e.g., a service time or interarrival time), which is known when it begins (although it may be defined in terms of a statistical distribution).
11. Delay: A duration of time of unspecified indefinite length, which is not known until it ends (e.g., a customer's delay in a last-in-first-out waiting line which, when it begins, depends on future arrivals).

12. Clock: A variable representing simulated time, called CLOCK.

Flowchart for Arrival Event

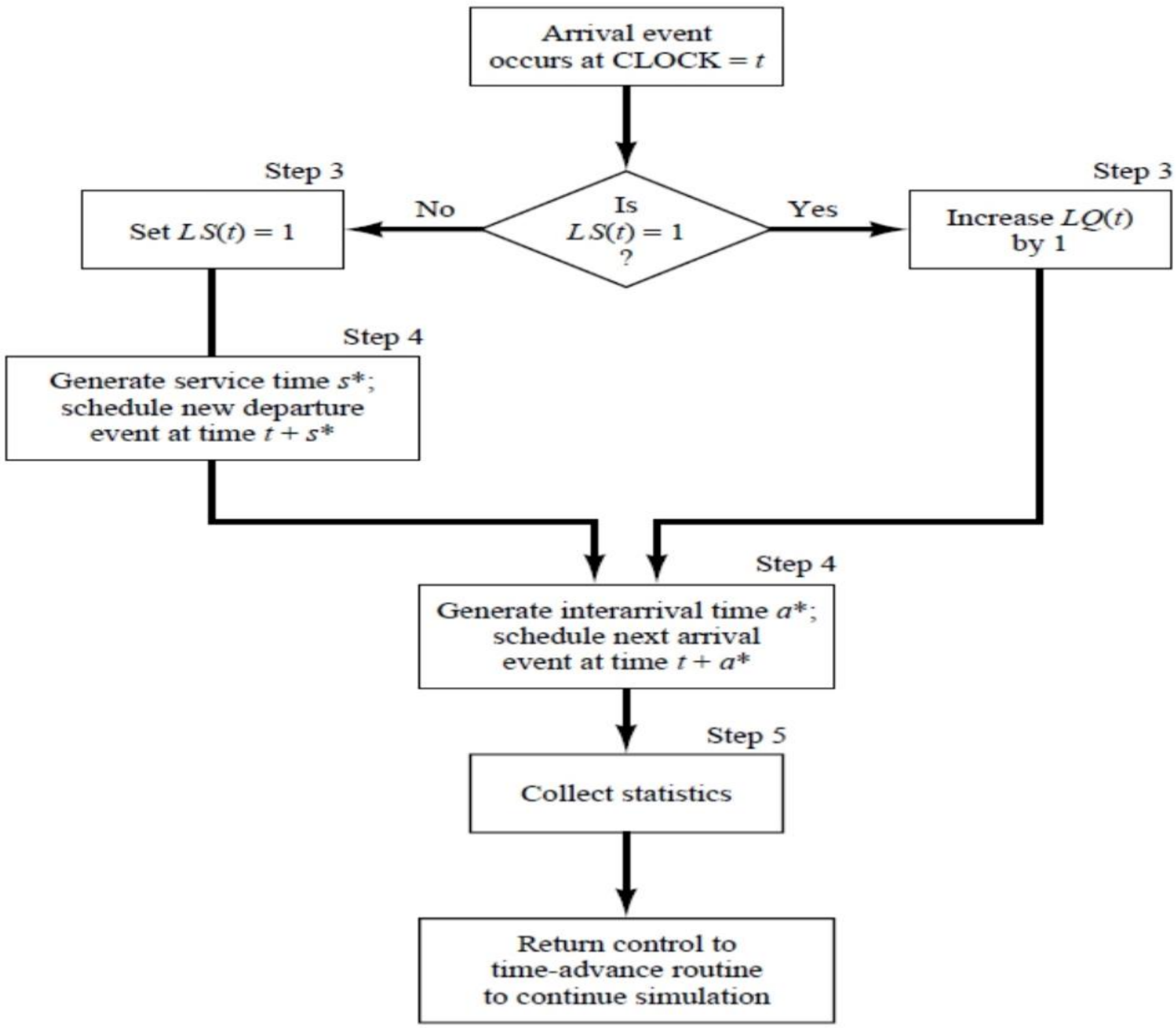


Figure Execution of the arrival event

Flowchart for Departure Event

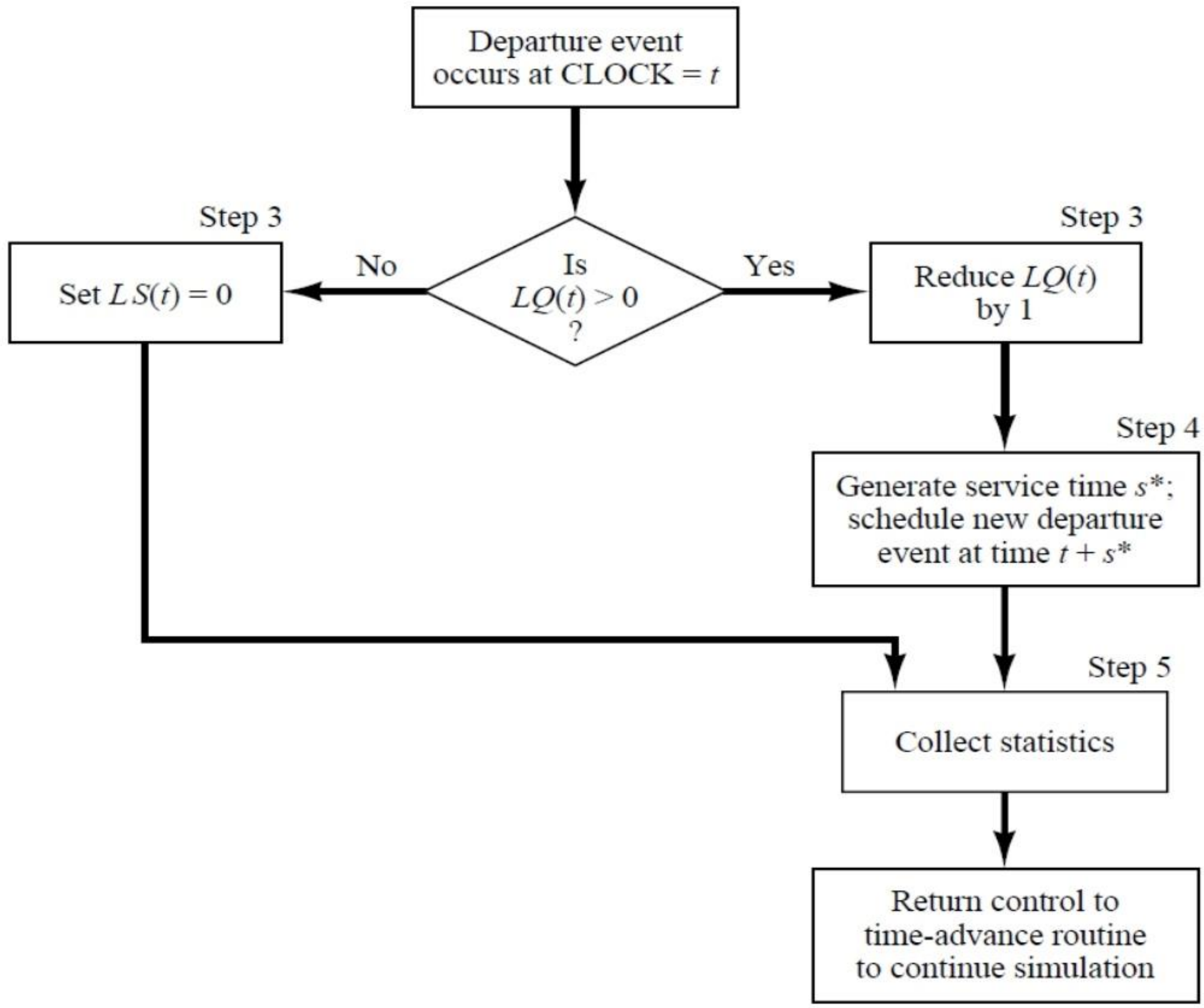


Figure Execution of the departure event.

2.b. Six dump trucks are used to have coal from the entrance of a mine to a rail road. Each truck is loaded by one of two loaders. After loading, a truck immediately moves to the scale to be weighed as soon as possible. Both loaders and scales have first come first serve weighing time for trucks. Travel time from loader to scale is considered negligible. After being weighed, a truck begins a travel time (during which time truck unloads) and then afterwards returns to the loader queue. The activities of loading, weighing and travel time are given in the following table:

Loading time	10	5	5	10	15	10	10
Weighing time	12	12	12	16	12	16	
Travel time	60	100	40	40	80		

End of simulation is completion of two weighing from the scale. Depict simulation table and estimate the loader and scale utilizations. **(08 Marks)**

<u>Ans</u>:

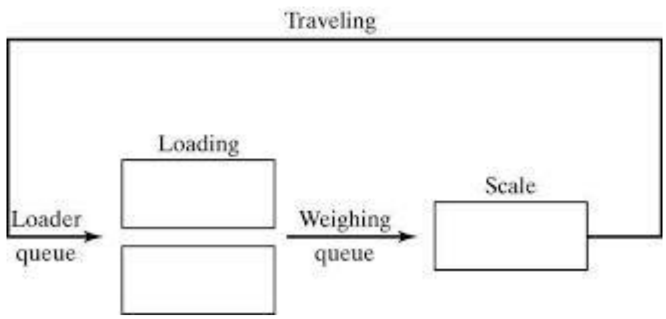


Figure Dump truck problem.

The model has the following components:

<u>System state</u>:

LQ(t) : Number of trucks in loader queue

L(t) : Number of trucks (0,1 or 2) being loaded

WQ(t): Number of trucks in weigh queue

W(t) : Number of trucks (0 or 1) being weighed, all at simulation time t.

<u>Event notices:</u>

(ALQ,t,DT_i) : Dump truck arrives at loader queue (ALQ) at time t.

(EL,t,DT_i) : Dump truck DT_i ends loading (EL) at time t.

(EW,t,DT_i) : Dump truck DT_i ends weighing (EW) at time t.

<u>Entities</u>:

The six dump trucks (DT_1, -----,DT_6)

<u>Lists</u>:

Loader queue : All trucks waiting to begin loading ordered on a first-come, first served basis.

Weigh queue : All trucks waiting to be weighed, ordered on a first-come, first served basis.

Activities:

Loading time, weighing time, and travel time

Delays:

Delay at loader queue, and delay at scale

It has been assumed that five of the trucks are at the loaders and one is at the scale at time 0.

Simulation Table:

Clock t	System State				Lists		Future Event List	Cumulative statistics	
	LQ(t)	L(t)	WQ(t)	W(t)	Loader queue	Weigh queue		B_L	B_S
0	3	2	0	1	DT_4 DT_5 DT_6		$(EL,5,DT_3)$ $(EL,10,DT_2)$ $(EW,12,DT_1)$	0	0
5	2	2	1	1	DT_5 DT_6	DT_3	$(EL,10,DT_2)$ $(EL,5+5,DT_4)$ $(EW,12,DT_1)$	10	5
10	1	2	2	1	DT_6	DT_3 DT_2	$(EL,10,DT_4)$ $(EW,12,DT_1)$ $(EL,10+10,DT_5)$	20	10
10	0	2	3	1		DT_3 DT_2 DT_4	$(EW,12,DT_1)$ $(EL,20,DT_5)$ $(EL,10+15,DT_6)$	20	10
12	0	2	2	1		DT_2 DT_4	$(EL,20,DT_5)$ $(EW,12+12,DT_3)$ $(EL,25,DT_6)$ $(ALQ,12+60,DT_1)$	24	12
20	0	1	3	1		DT_2 DT_4 DT_5	$(EW,24,DT_3)$ $(EL,25,DT_6)$ $(ALQ,72,DT_1)$	40	20
24	0	1	2	1		DT_4 DT_5	$(EL,25,DT_6)$ $(EW,24+12,DT_2)$ $(ALQ,72,DT_1)$ $(ALQ,24+100,DT_3)$	44	24

Total busy time of both loaders = B_L = 44 minutes
Total busy time of scale = B_S = 24 minutes

$$\text{Average loader utilization} = \frac{B_L / \text{Number of loaders}}{\text{Total time}}$$

$$= \frac{44/2}{24} = 0.9166$$

$$\text{Average scale utilization} = \frac{B_S}{\text{Total time}}$$

$$= \frac{24}{24} = 1.00$$

<u>Module-2</u>

3.a. Explain the characteristics of queueing systems. List different queueing notations. **(08 Marks)**

<u>Ans</u>:

Characteristics of Queuing System:

1. <u>Calling population</u>:

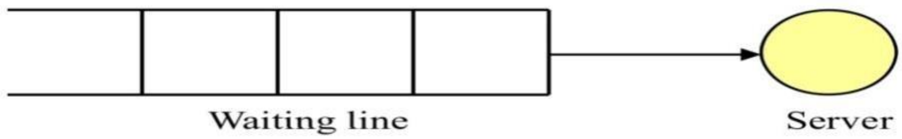
- The population of potential customers is refered to as the *calling population*.
- The calling population can be finite or infinite.
- The key difference between “finite” and “infinite” population model is how the arrival rate is defined.
  - ➢ In an infinite population model, the arrival rate is not affected by the number of customers in the system. Usually the system is viewed as an *open* system, customers come from outside the system and leave the system after finishing the work.
  - ➢ In a finite population model, the arrival rate at a server is affected by the population in the system. Usually the system is viewed as a *closed* system, customers (with a fixed population) don't leave the system, they merely move around system from one server to another, from one queue to another.

2. <u>System capacity</u>:

A limit on the number of customers that may be in the waiting line or system.

- Limited capacity: e.g., an automatic car wash only has room for 10 cars to wait in line to enter the mechanism.
- If system is full no customers are accepted anymore

- Unlimited capacity, e.g., concert ticket sales with no limit on the number of people allowed to wait to purchase tickets.



3. <u>Arrival Process</u>:

infinite-population models:

- The arrival process for infinite-population models is usually characterized in terms of interarrival times of successive customers.
- Arrivals may occur at scheduled times or at random times.

For finite-population models:

- Customer is pending when the customer is outside the queueing system, e.g., machine-repair problem: a machine is "pending" when it is operating, it becomes "not pending" the instant it demands service from the repairman.
- Runtime of a customer is the length of time from departure from the queueing system until that customer's next arrival to the queue, e.g., machine-repair problem, machines are customers and a runtime is time to failure (TTF).
- Let $A_1^{(i)}$, $A_2^{(i)}$, … be the successive runtimes of customer i, and $S_1^{(i)}$,$S_2^{(i)}$ be the corresponding successive system times:

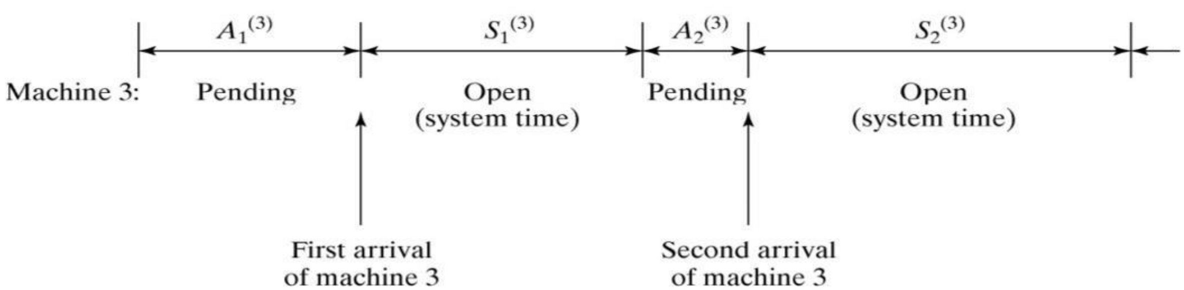


Figure 6.2 Arrival process for a finite-population model.

4. Queue Behavior and Queue Discipline:

Queue behavior: the actions of customers while in a queue waiting for service to begin, for example:

- Balk: leave when they see that the line is too long
- Renege: leave after being in the line when its moving too slowly
- Jockey: move from one line to a shorter line

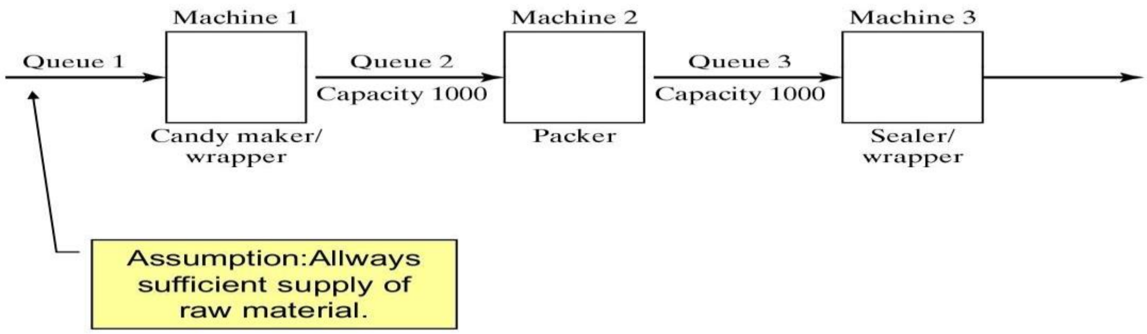
Queue discipline: the logical ordering of customers in a queue that determines which customer is chosen for service when a server becomes free, for example:

- First-in-first-out (FIFO)
- Last-in-first-out (LIFO)
- Service in random order (SIRO)
- Shortest processing time first (SPTF)
- Service according to priority (PR)

5. Service Times and Service Mechanism:

- Service times of successive arrivals are denoted by S_1, S_2, S_3. They May be constant or of random duration.
- $\{S_1, S_2, S_3, \ldots\}$ is usually characterized as a sequence of independent and identically distributed random variables.
- A queueing system consists of a number of service centers and interconnecting queues. Each service center consists of some number of servers, c, working in parallel; that is, upon getting to the head of the line, a customer takes the first available server.
- Parallel service mechanisms are either single server ($c = 1$), multiple server ($1 < c < \infty$), or unlimited servers ($c = \infty$).
- A self-service facility is usually characterized as having an unlimited number of servers.

Example of Queueing System:



Candy production line

- Three machines separated by buffers
- Buffers have capacity of 1000 candies

Queueing Notations:

- A notation system for parallel server queues: A/B/c/N/K
  → 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

  → N, K are usually dropped, if they are infinity
- Common symbols for A and B include
  → M (exponential or Markov)

  → D (constant or deterministic)

  → Ek (Erlang distribution of order k)

  → PH (phase-type)

  → H (Hyperexponential)
  → G (arbitrary or general)
  → GI (general independent)

• P_n	steady-state probability of having n customers in system
• $P_n(t)$	probability of n customers in system at time t
• λ	arrival rate
• λ_e	effective arrival rate
• μ	service rate of one server
• ρ	server utilization
• A_n	interarrival time between customers n-1 and n
• S_n	service time of the n-th arriving customer
• W_n	total time spent in system by the n-th customer
• W_n^Q	total time spent in the waiting line by customer n
• $L(t)$	the number of customers in system at time t
• $L_Q(t)$	the number of customers in queue at time t
• L	long-run time-average number of customers in system
• L_Q	long-run time-average number of customers in queue
• W	long-run average time spent in system per customer
• w_Q	long-run average time spent in queue per customer

Table 6.2 Queueing Notation for Parallel Server Systems

3.b. Define discrete and continuous random variable. Explain the binomial and Poisson distribution. **(08 Marks)**

Ans:

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 $x_1, x_2, \ldots$.
- In the finite case, the list terminates; in the countably infinite case, the list continues indefinitely.
- Example: Consider jobs arriving at a job shop.
    - Let be X number of jobs arriving each week.
    - Rx = range space of X = { 0, 1, 2, … }.
    - Let X be a discrete random variable. With each possible outcome x_i; in Rx, a number $p(x_i) = P(X = x_i)$ gives the probability that the random variable equals the value of x_i.
    - $p(x_i) = 1, 2, \ldots,$ must satisfy the following two conditions:
        1. $p(x_i) \geq 0$, for all i
        2. $\sum_{i=1}^{\infty} p(x_i) = 1$

- The collection of pairs $(x_i, p(x_i))$, i = 1, 2, ... is called the probability distribution of X, and $p(x_i)$ is called the probability mass function (pmf) of X.

Continuous random variables

- If the range space Rx 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

$$P(a \leq X \leq b) = \int_a^b f(x)dx \qquad \text{---- (5.1)}$$

- 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 Rx
    b. $\int_{Rx} f(x)dx = 1$
    c. $f(x) = 0$ if x is not in Rx
- Properties:
    - $P(X = x_0) = 0$, because $\int_{x_0}^{x_0} f(x)\,dx = 0$
    - $P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b)$ ---- (5.2)
- The graphical interpretation of Equation (5.1) is shown in Figure 5.2. The shaded area represents the probability that X lies in the interval [a, b].

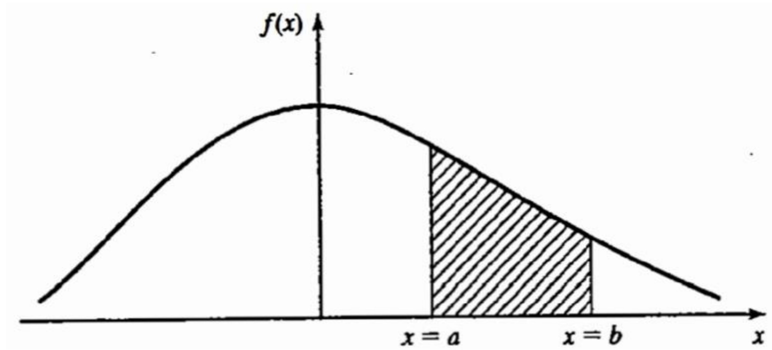


Figure 5.2 Graphical interpretation of P(a <X< b).

Example:

- The life of a device used to inspect cracks in aircraft wings is given by X, a continuous random variable assuming all values in the range $x \geq 0$. The pdf of the lifetime, in years, is as follows:

$$f(x) = \begin{cases} \frac{1}{2} e^{-x/2}, & x \geq 0 \\ 0, & \text{Otherwise} \end{cases}$$

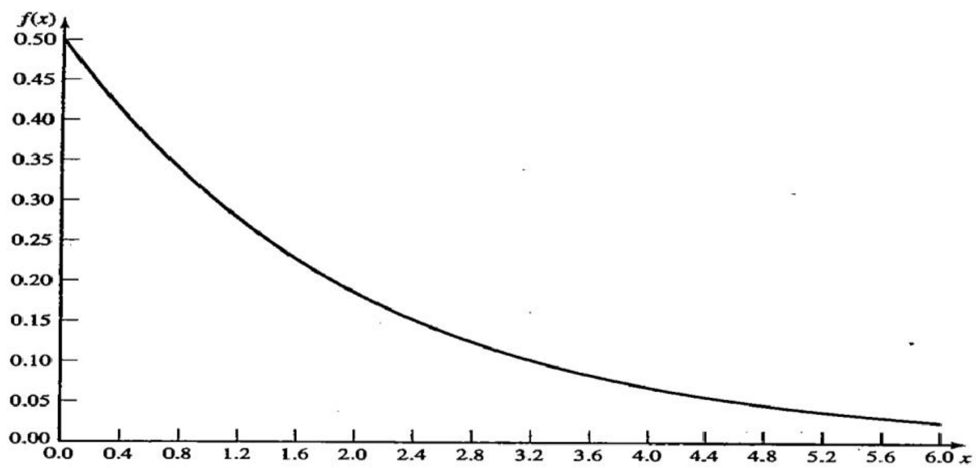


Figure 5.3 pdf for inspection-device life.

- X has an exponential distribution with mean 2 years.
- The probability that the life of the device is between 2 and 3 years is calculated as

$$P(2 \leq X \leq 3) = \int_2^3 \frac{1}{2} e^{-x/2} dx$$

$$= - e^{-3/2} + e^{-1} = -0.223 + 0.368 = 0.145$$

Binomial Distribution

- The random variable X that denotes the number of successes in n Bernoulli trials has a binomlal distribution given by p(x), where

$$p(x) = \begin{cases} \binom{n}{x} p^x q^{n-x}, & x = 0,1,2,\text{---},n \\ 0, & \text{Otherwise} \end{cases} \quad (5.12)$$

Where,

$q = 1\text{-}p$

$$\binom{n}{x} = \frac{n!}{x!\,(n-x)!}$$

- An easy approach to calculating the mean and variance of the binomial distribution is to consider X as a sum of n independent Bernoulli random variables, each with mean p and variance p(1 - p) = pq. Then,

  $X = X_1 + X_2 + \text{---} + X_n$

  and the mean, E(X), is given by

  $E(X) = p + p + \ldots + p = np$

  and the variance V(X) is given by

  $V(X) = pq + pq + \cdots + pq = npq$

Poisson Distribution

- The Poisson distribution describes many random processes quite well and is mathematically quite simple.
- The Poisson probability mass function is given by

$$p(x) = \begin{cases} \dfrac{e^{-\alpha}\alpha^x}{x!}, & x = 0,1,2\ldots \\ 0\ , & \text{Otherwise} \end{cases}$$

  Where $\alpha > 0$

- One of the important properties of the Poisson distribution is that the mean and variance are both equal to α,. that is,

  $E(X) = \alpha = V(X)$

- The cumulative distribution function is given by

$$F(x) = \sum_{i=0}^{x} \frac{e^{-\alpha}\alpha^i}{i!}$$

4.a. Explain the following continuous distributions:

(i) Uniform distribution

(ii) Exponential distributions **(08 Marks)**

Ans:

(i) Uniform Distribution

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

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \le x \le b \\ 0, & \text{Otherwise} \end{cases}$$

- The cdf is given by

$$F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \le x < b \\ 1, & x \ge b \end{cases}$$

- Note that

$$P(x_1 < X < x_2) = F(x_2) - F(x_1) = \frac{x_2 - x_1}{b-a}$$

is proportional to the length of the interval, for all x_1 and x_2 satisfying $a \le x_1 < x_2 \le b$.

- The mean and variance of the distribution are given by

$$E(X) = \frac{a+b}{2} \quad \text{and} \quad V(X) = \frac{(b-a)^2}{12}$$

- The pdf and cdf when a = 1 and b = 6 are shown in Figure 5.8.

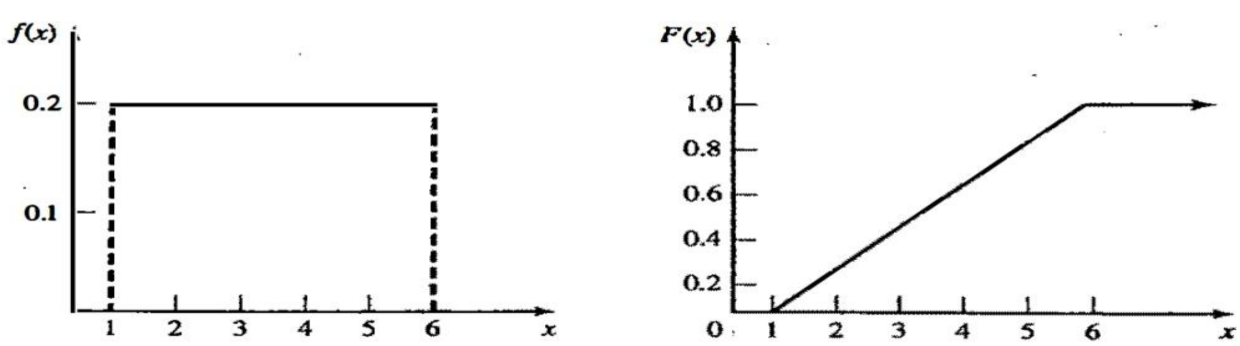


Figure 5.8 pdf and cdf for uniform distribution.

(ii) Exponential distributions

- 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 density function is shown in Figure 5.9. Figure 5.9 also shows the cdf.

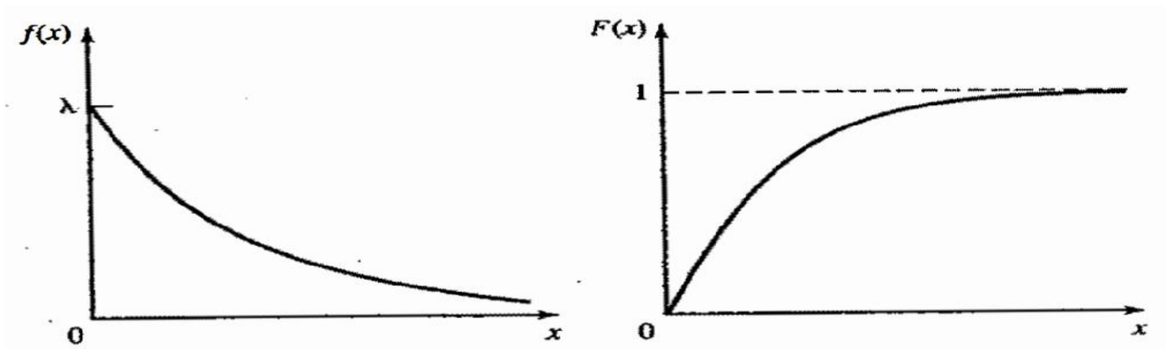


Figure 5.9: Exponential density function and cumulative distribution function

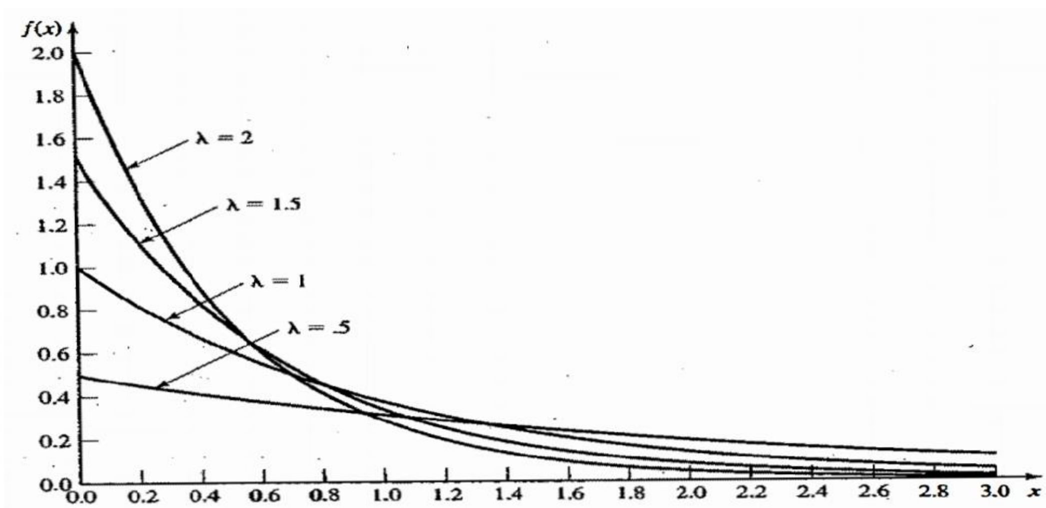


Figure 5.10 pdfs for several exponential distributions.

- The exponential distribution has been used to model interarrival times when arrivals are completely random and to model service times that are highly variable.
- Several different exponential pdf's are shown in Figure 5.10. The value of the intercept on the vertical axis is always equal to the value of λ and all pdf's eventually intersect.
- The exponential distribution has mean and variance given by

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

- cdf of exponential distribution is given by

$$F(x) = \begin{cases} 0 & x<0 \\ 1 - e^{-\lambda x} & x \geq 0 \end{cases}$$

4.b. Explain steady state parameters of M/G/1 queue. **(08 Marks)**

Ans:

Steady State Parameters of M/G/1 Queue

- Suppose that service times have mean $1/\mu$ and variance σ^2 and that there is one server.
- If $\rho = \lambda/\mu < 1$, then the M/G/1 queue has a steady-state probability distribution with steady-state characteristics, as given in Table 6.3.
- In general, there is no simple expression for the steady-state probabilities P_0, P_1, P_2, …. When $\lambda < \mu$, the quantity $\rho = \lambda/\mu$ is the server utilization, or long-run proportion of time the server is busy.
- $1 - P_0 = \rho$ can also be interpreted as the steady-state probability that the system contains one or more customers. $L - L_Q = \rho$ is the time-average number of customers being served.

ρ	$\frac{\lambda}{\mu}$
L	$\rho + \frac{\lambda^2(1/\mu^2 + \sigma^2)}{2(1-\rho)} = \rho + \frac{\rho^2(1+\sigma^2\mu^2)}{2(1-\rho)}$
w	$\frac{1}{\mu} + \frac{\lambda(1/\mu^2 + \sigma^2)}{2(1-\rho)}$
w_Q	$\frac{\lambda(1/\mu^2 + \sigma^2)}{2(1-\rho)}$
L_Q	$\frac{\lambda^2(1/\mu^2 + \sigma^2)}{2(1-\rho)} = \frac{\rho^2(1+\sigma^2\mu^2)}{2(1-\rho)}$
P_0	$1-\rho$

Table 6.3 Steady-State Parameters of the M/G/1 Queue

Where,

L → Long-run time-average number of customers in system

λ → Arrival rate

$\mu \rightarrow$ Service rate

$\rho \rightarrow$ Server utilization

$\sigma \rightarrow$ Mean service time

$w \rightarrow$ w is long-run average time spent in system per customer

$w_Q \rightarrow$ Long run average time spent in Queue per customer

$L_Q \rightarrow$ Long-run time-average number of customers in Queue

$P_o \rightarrow$ Probability of empty system

Example:

- Two workers competing for a job, Able claims to be faster than Baker on average but Baker claims to be more consistent,
- Poisson arrivals at rate $\lambda = 2$ per hour (1/30 per minute)
- **For Able:** $\lambda = 1/30$ per minute, $1/\mu = 24$ minutes, $\rho = \lambda/\mu = 24/30 = 4/5$ & $\sigma^2 = 20^2 = 400$ minutes^2 and the average queue length is computed as

  $$L_Q = \frac{\left(\frac{1}{30}\right)^2 [24^2+400]}{2(1-\frac{4}{5})} = 2.711 \text{ customers}$$

  The proportion of arrivals who find Able idle & thus experience no delay is

  $$P_o = 1 - \rho = 1/5 = 20\ \%$$

- **For Baker:** $\lambda = 1/30$ per minute, $1/\mu = 25$ minutes, $\rho = \lambda/\mu = 25/30 = 5/6$ & $\sigma^2 = 2^2 = 4$ minutes^2 and the average queue length is computed as

  $$L_Q = \frac{\left(\frac{1}{30}\right)^2 [25^2+4]}{2(1-\frac{5}{6})} = 2.079 \text{ customers}$$

  The proportion of arrivals who find Baker idle & thus experience no delay is

  $$P_o = 1 - \rho = 1/6 = 16.7\ \%$$

- Although working faster on average, Able's greater service variability results in an average queue length about 30 % greater than Baker's.

<u>Module-3</u>

5.a. What is the role of maximum density and maximum period in generating random numbers? With given seed 45, constant multiplier 21, increment 49and modulus 40, generate a sequence of five random numbers. **(08 Marks)**

<u>Ans</u>:

Role of maximum density and maximum period in generating random numbers

- The maximum density and maximum period are the secondary properties of random numbers.
- The **maximum density** of random numbers ensures that values assumed by Ri, i= 1,2,… , leave no large gaps on [0,1].
- The **maximum period** refers the length of the sequence of random numbers which are going to repeat after a certain random numbers.
- **Maximum period** is used to achieve maximum density, and to avoid cycling (i.e., recurrence of the same sequence of generated numbers) in practical applications, the generator should have the largest possible period.
- **Maximal period** can be achieved by the proper choice of a, c, m, and Xo.

Solution to Problem

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

Where,

Xo → Seed

a → Multiplier

c → Increment

m → Modulus

And

$$R_i = \frac{X_i}{m} \qquad i= 1,2,\ldots.$$

Given $X_0 = 45$, $a = 21$, $m = 40$, $c = 49$

The sequence of X_i and subsequent Ri values is computed as follows:

$X_1 = (aX_0+c) \bmod m$

$= (21\times45+49) \bmod 40$

$= 994 \bmod 40$

$= 34$

$$R_1 = \frac{X_1}{m} = \frac{34}{40} = 0.85$$

$$X_2 = (aX_1 + c) \bmod m$$
$$= (21 \times 34 + 49) \bmod 40$$
$$= 763 \bmod 40$$
$$= 3$$
$$R_2 = \frac{X_2}{m} = \frac{3}{40} = 0.075$$

$$X_3 = (aX_2 + c) \bmod m$$
$$= (21 \times 3 + 49) \bmod 40$$
$$= 112 \bmod 40$$
$$= 32$$
$$R_3 = \frac{X_3}{m} = \frac{32}{40} = 0.8$$

$$X_4 = (aX_3 + c) \bmod m$$
$$= (21 \times 32 + 49) \bmod 40$$
$$= 721 \bmod 40$$
$$= 1$$
$$R_4 = \frac{X_4}{m} = \frac{1}{40} = 0.025$$

$$X_5 = (aX_4 + c) \bmod m$$
$$= (21 \times 1 + 49) \bmod 40$$
$$= 70 \bmod 40$$
$$= 30$$
$$R_5 = \frac{X_5}{m} = \frac{30}{40} = 0.75$$

5.b. The sequence of numbers 0.54, 0.73, 0.98, 0.11, 0.08 has been generated. Use Kolmogorov Simirnov test with α=0.05 to determine if the hypothesis that the numbers are uniformly distributed on the interval [0,1] can be rejected. Compare F(x) and $S_N(x)$ on a graph. $D_{0.05}$ = 0.565. **(08 Marks)**

Ans:

Given

α=0.05, $D_{0.05}$ = 0.565

In testing for uniformity, the hypotheses are as follows:

$H_o : R_i \sim U[0,1]$

$H_1 : R_i \nsim U[0,1]$

Step 1: Rank the data from smallest to largest. Let $R_{(i)}$, denote the ith smallest observation so that

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

$0.08 \leq 0.11 \leq 0.54 \leq 0.73 \leq 0.98$

Step 2: Compute D^+ & D^-

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

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

Table: calculation for Kolmogorov-Smirnov Test

R(i)	0.08	0.11	0.54	0.73	0.98
i/N	0.20	0.40	0.60	0.80	1.00
i/N - R(i)	0.12	0.29	0.06	0.07	0.02
R(i) – (i-1)/N	0.08	-0. 09	0.14	0.13	0.18

$D^+ = \max \{0.12, 0.29, 0.06, 0.07, 0.02\}$

$= 0.29$

$D^- = \max \{0.08, -0.09, 0.14, 0.13, 0.18\}$

$= 0.18$

Step 3: Compute $D = \max\{ D^+, D^-\}$

$D = \max\{ 0.29, 0.18\}$

$D = 0.29$

Step 4: Compute the critical value Dα, for the specified significance level α and the given sample size N.

Here, α=0.05 and N=5

$D_{0.05} = 0.565$

Step 5: Since $D \leq D\alpha$ that is, $0.29 \leq 0.565$

The null hypothesis (Ho) that the data are a sample from a uniform distribution is accepted.

Comparison of F(x) and S_N(x)

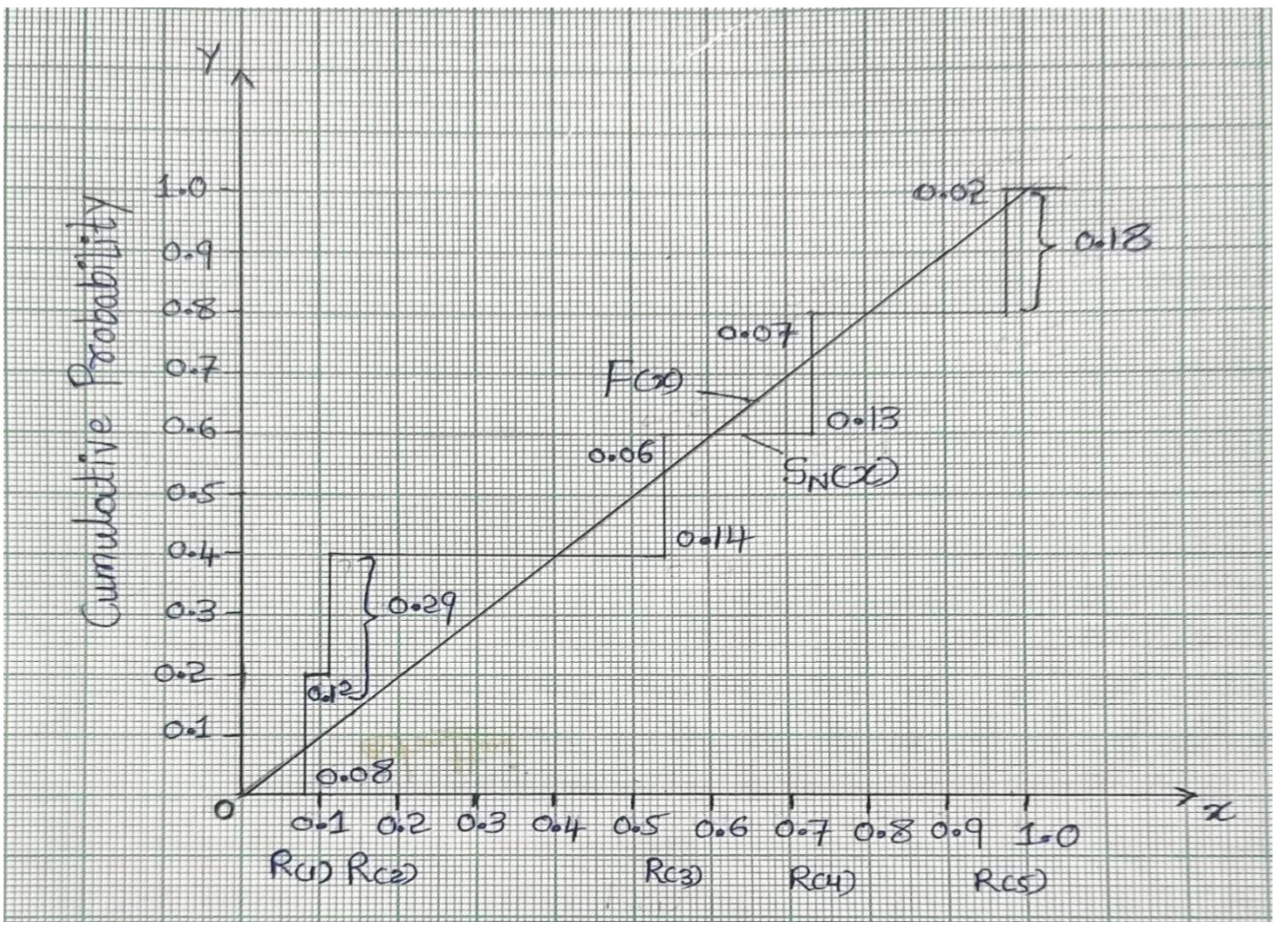


Figure : Comparison of F(x) and S_N(x)

6.a. Explain the inverse transformation technique for exponential distribution. Show the corresponding graphical interpretation. Explain the acceptance rejection technique. **(08 Marks)**

Ans:

Exponential Distribution

- The exponential distribution has the probability density function (pdf)

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

and the cumulative distribution function (cdf)

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

- The parameter λ can be interpreted as the mean number of occurrences per time unit.

$$E(Xi) = \frac{1}{\lambda}$$

- step-by-step procedure for the inverse-transform technique, by the exponential distribution, consists of the following steps:

Step 1: Compute the cdf of the desired random variable X.

For the exponential distribution, the cdf is F(x) = $1 - e^{-\lambda x}$ $x \geq 0$

Step 2: Set F(X) = R on the range of X.

For the exponential distribution, it becomes $1 - e^{-\lambda x}$ =R on the range x≥0.

X is a random variable ,so $1 - e^{-\lambda x}$ is also a random variable, here called R.

Step 3: Solve the equation F(X) = R for X in terms of R.

$$1 - e^{-\lambda x} = R$$

$$e^{-\lambda x} = 1-R$$

$$-\lambda X = \ln(1-R)$$

$$X = -\frac{1}{\lambda}\ln(1-R) \qquad \text{------} \quad (8.1)$$

Equation (8.1) is called a random-variate generator for the exponential distribution. ln general, Equation (8.1) is written as $X= F^{-1}(R)$.

Step 4: Generate (as needed) uniform random numbers R_1, R_2, R_3, ... and compute the desired random variates by $X_i=F^{-1}(R_i)$

For the exponential case, $F^{-1}(R) = (-1/\lambda)\ln(1 - R)$ by Equation (8.1), so

$$X_i = \frac{1}{\lambda}\ln(1-R_i) \qquad \text{-------} \; (8.2)$$

for i= 1, 2, 3, ... One simplification that is usually employed in Equation (8.2) is to replace 1 - Ri by Ri to yield

$$X_i = \frac{1}{\lambda}\ \ln(R_i) \qquad \text{-------} \; (8.3)$$

This alternative is justified by the fact that both Ri and 1- Ri are uniformly distributed on [0, 1].

<u>Graphical interpretation</u>

- Figure 8.2 gives a graphical interpretation of the inverse-transform technique. The cdf shown is $F(x) = 1-e^{-x}$, an exponential distribution with rate $\lambda = 1$.
- To generate a value X_1 with cdf F(x), a random number R_1 between 0 and l is generated, then a horizontal line is drawn from R_1 to the graph of the cdf, then a vertical line is dropped to the x axis to obtain X_1, the desired result. Notice the inverse relation between R_1 and X_1 namely

$$R_1 = 1- e^{-X_1}$$

and

$$X_1 = - \ln(1-R_1)$$

In general, the relation is written as

$$R_1 = F(X_1)$$

and

$$X_1 = F^{-1}(R_1)$$

$$P(X_1 \le x_0) = P(R_1 \le F(x_0)) = F(x_0) \qquad (8.4)$$

- To see the first equality in Equation (8.4), refer to Figure 8.2, where the fixed numbers x_0 and $F(x_0)$ are drawn on their respective axes. It can be seen that $X_1 \le x_0$ when and only when $R_1 \le F(x_0)$.
- Since $0 \le F(x_0) \le 1$, the second equality in Equation (8.4) follows immediately from the fact that R_1 is uniformly distributed on [0, 1].

- Equation 8.4 shows that the cdf of X_1 is F; hence, X_1 has the desired distribution.

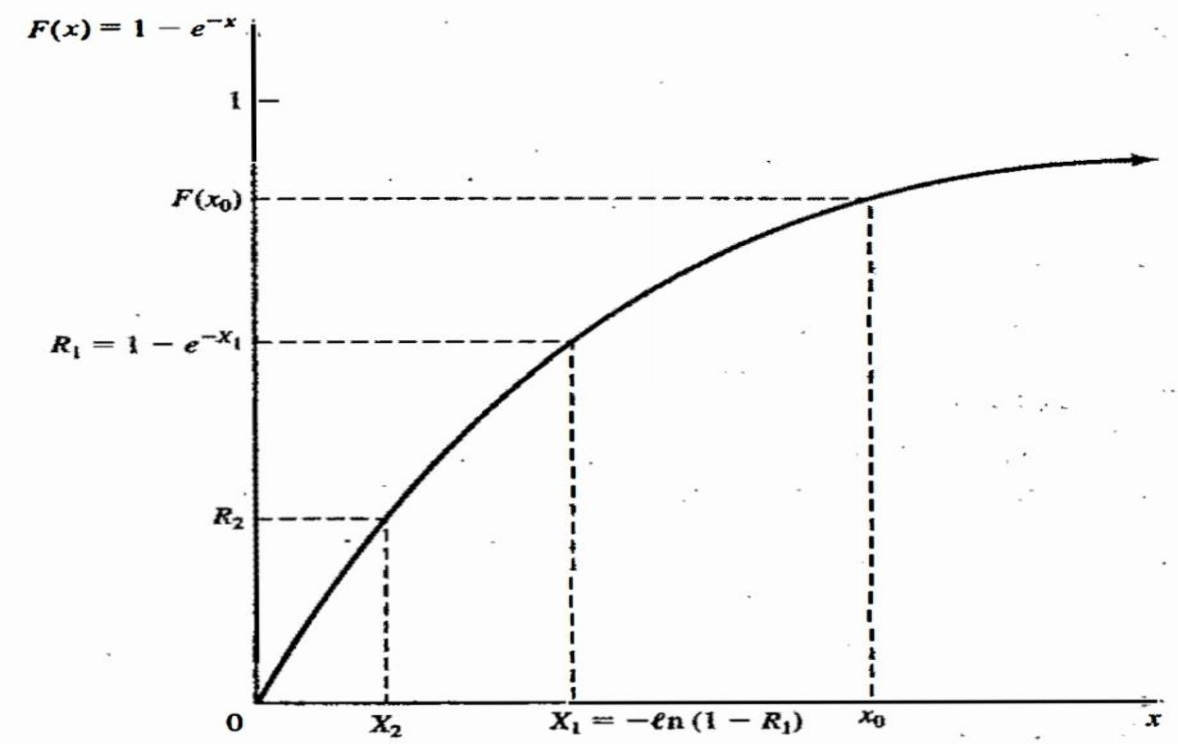


Figure 8.2 Graphical view of the inverse-transform technique

Acceptance Rejection Technique:

Suppose that an analyst needed to devise a method for generating random variates, X, uniformly distributed between 1/4 and 1. One way to proceed would be to follow these steps:

Step 1. Generate a random number R.

Step 2a. If R ≥ l/4, accept X= R, then go to Step 3.

Step 2b. If R < l/4, reject R, and return to Step 1.

Step 3. If another uniform random variate on [l/4, 1] is needed, repeat the procedure beginning at Step 1. If not, stop.

'Each time Step 1 is executed, a new random number R must be generated. Step 2a is an "acceptance" and Step 2b is a "rejection" in this acceptance-rejection technique.

6.b. Use the Chi-Square test with α=0.05 to test for whether the data shown are uniformly distributed. The test uses n=10 intervals of equal length. $\chi^2_{0.05,9} = 16.9$.

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	0.90	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

(08 Marks)

Ans:

Given,

$\alpha = 0.05$

n = number of classes = 10

N = Sample size = 50

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

In testing for uniformity, the hypotheses are as follows:

$H_o : R_i \sim U[0,1]$

$H_1 : R_i \nsim U[0,1]$

All the given random numbers lie with in [0, 1].

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

$$E_i = \frac{N}{n} = \frac{50}{10} = 5$$

Table 6.b: Computations for Chi-Square Test

Interval	Range	O_i	E_i	$O_i - E_i$	$(O_i - E_i)^2$	$(O_i - E_i)^2/E_i$
1	0.01-0.10	3	5	-2	4	0.8
2	0.11-0.20	3	5	-2	4	0.8
3	0.21-0.30	5	5	0	0	0
4	0.31-0.40	3	5	-2	4	0.8
5	0.41-0.50	6	5	1	1	0.2
6	0.51-0.60	3	5	-2	4	0.8
7	0.61-0.70	5	5	0	0	0
8	0.71-0.80	9	5	4	16	3.2
9	0.81-0.90	7	5	2	4	0.8
10	0.91-1.00	6	5	1	1	0.2
		50	50			7.6

The chi-square test uses the sample statistic

$$\chi^2_0 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

Where,

O_i is the observed number in the ith class

E_i is the expected number in the ith class

n is the number of classes

From Table 6.b:

$\chi_0^2 = 7.6$

Since $\chi_0^2 < \chi^2_{\alpha,n-1}$ that is, $\chi_0^2 < \chi^2_{0.05,10-1}$

$$7.6 < 16.9$$

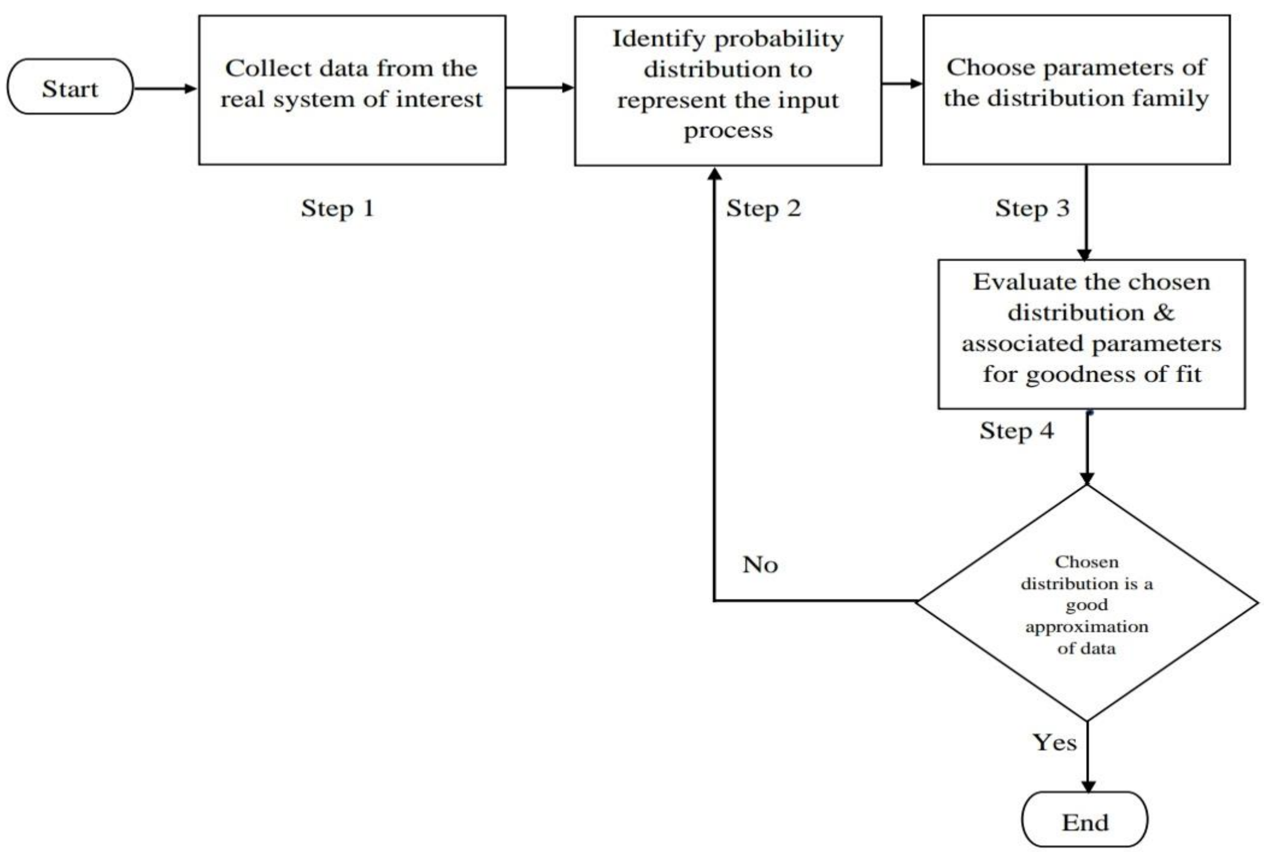
Null hypothesis (Ho) that the data are sample from a uniform distribution is accepted.

Module-4

7.a. List the steps involved in development of a useful model of input data and explain. **(08 Marks)**

Ans:

Flow Chart

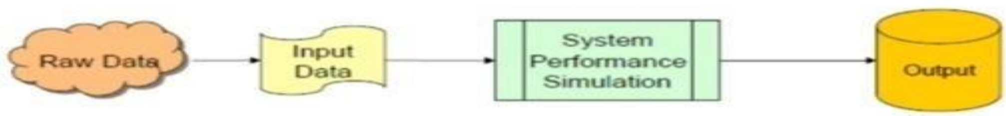


Steps involved in development of a useful model of input data:

1. Data Collection
2. Identifying the Distribution with Data
3. Parameter Estimation
4. Goodness of fit tests

1. Data Collection

- Data collection is One of the biggest task in solving a real problem.
- GIGO: Garbage-In-Garbage-Out



- Even when model structure is valid simulation results can be misleading, if the input data is
  →Inaccurately collected
  →Inappropriately analyzed
  →Not representative of the environment

2. Identifying the Distribution with Data:

<u>Histograms</u>

A frequency distribution or histogram is useful in identifying the shape of a distribution. A histogram is constructed as follows:

1. Divide the range of the data into intervals. (Intervals are usually of equal width; however, unequal widths may be used if the heights of the frequencies are adjusted.)
2. Label the horizontal axis to conform to the intervals selected.
3. Find 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.

<u>Selecting the Family of Distributions</u>

- A family of distribution is selected based on:

  →Context of the input variable

  →Shape of the histogram

- Frequently encountered distributions:
   i. Easier to analyze: Exponential, Normal and poisson
   ii. Difficult to analyze: Beta, Gamma and Weibull

<u>Quantile-Quantile Plots</u>

- A quantile-quantile (q - q) plot is a useful tool for evaluating distribution fit.
- If X is a random variable with cdf F, then the q-quantile of X is that value such that $F(\gamma)=P(X\leq \gamma)=q$, for $0 < q < 1$. When F has an inverse, we write $\gamma= F^{-1}(q)$.
- Now let $\{x_i, i = 1, 2, \ldots, n\}$ be a sample of data from X. Order the observations from the smallest to the largest, and denote these as $\{y_j, j=1, 2, \ldots, n\}$, where $y_1\leq y_2\leq \ldots\leq y_n$. Let j denote the ranking or order number. Therefore, j = 1 for the smallest and j = n for the largest. The q - q plot is based on the fact that y_j is an estimate of the $(j - 1/2)/n$ quantile of X. In other words,

  y_j is approximately $F^{-1}\left(\frac{j-\frac{1}{2}}{n}\right)$

3. Parameter Estimation:

<u>Preliminary statistics: Sample mean and sample variance</u>

i. If observations in a sample size of n are $X_1,X_2,X_3,\ldots,X_n$ (continuous or discrete), the sample mean and variance are:

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

ii. If the data are discrete and have been grouped in a frequency distribution:

$$\bar{X}=\frac{\sum_{j=1}^{k} f_jX_j}{n} \qquad S^2=\frac{\sum_{j=1}^{k} f_jX_j^2-n\bar{X}^2}{n-1}$$

k→ number of distinct values of X

fj → Observed frequency of value X_j

iii. When raw data are unavailable (data are grouped in to class intervals), the approximate sample mean and variance are:

$$\bar{X}=\frac{\sum_{j=1}^{c} f_jm_j}{n} \qquad S^2=\frac{\sum_{j=1}^{c} f_jm_j^2-n\bar{X}^2}{n-1}$$

$f_j \rightarrow$ observed frequency in the jth class interval

$m_j \rightarrow$ midpoint of the jth interval

$c \rightarrow$ number of class intervals

Suggested Estimators:

Distribution	Parameter	Estimator
Poisson	α	$\hat{\alpha} = \bar{X}$
Exponential	λ	$\hat{\lambda} = \frac{1}{\bar{X}}$
Gamma	β, θ	$\hat{\theta} = \frac{1}{\bar{X}}$
Normal	μ, σ^2	$\hat{\mu} = \bar{X}, \hat{\sigma}^2 = S^2$
Lognormal	μ, σ^2	$\hat{\mu} = \bar{X}, \hat{\sigma}^2 = S^2$

Table: Suggested Estimators for Distributions Often Used in Simulation

4. Goodness-of-Fit Tests:

- Conduct hypothesis testing on input data distribution using

  →Kolmogorov-Smirnov test

  →Chi-square test

- No single correct distribution in a real application exists

  →If very little data are available, it is unlikely to reject any candidate distributions

  →If a lot of data are available, it is likely to reject all candidate distributions

7.b. Explain how the method of histograms can be used to identify the shape of a distribution. With an example also mention drawbacks of histogram and advantages of Q-Q plot. **(08 Marks)**

Ans:

Histograms

A frequency distribution or histogram is useful in identifying the shape of a distribution. A histogram is constructed as follows:

1. Divide the range of the data into intervals. (Intervals are usually of equal width; however, unequal widths may be used if the heights of the frequencies are adjusted.)
2. Label the horizontal axis to conform to the intervals selected.
3. Find 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.

Example:

- The number of vehicles arriving at the northwest corner of an intersection in a 5-minute period between 7:00 A.M. and 7:05 A.M. was monitored for five workdays over a 20-week period.
- Table 9.1 shows the resulting data. The first entry in the table indicates that there were 12 5-minute periods during which zero vehicles arrived, 10 periods during which one vehicle arrived, and so on.
- The number of automobiles is a discrete variable, and there are sample data, so the histogram may have a cell for each possible value in the range of the data. The resulting histogram is shown in Figure 9 .2.

Arrivals per Period	*Frequency*	*Arrivals per Period*	*Frequency*
0	12	6	7
1	10	7	5
2	19	8	5
3	17	9	3
4	10	10	3
5	8	11	1

Table 9.1 Number of Arrivals in a 5-Minute Period

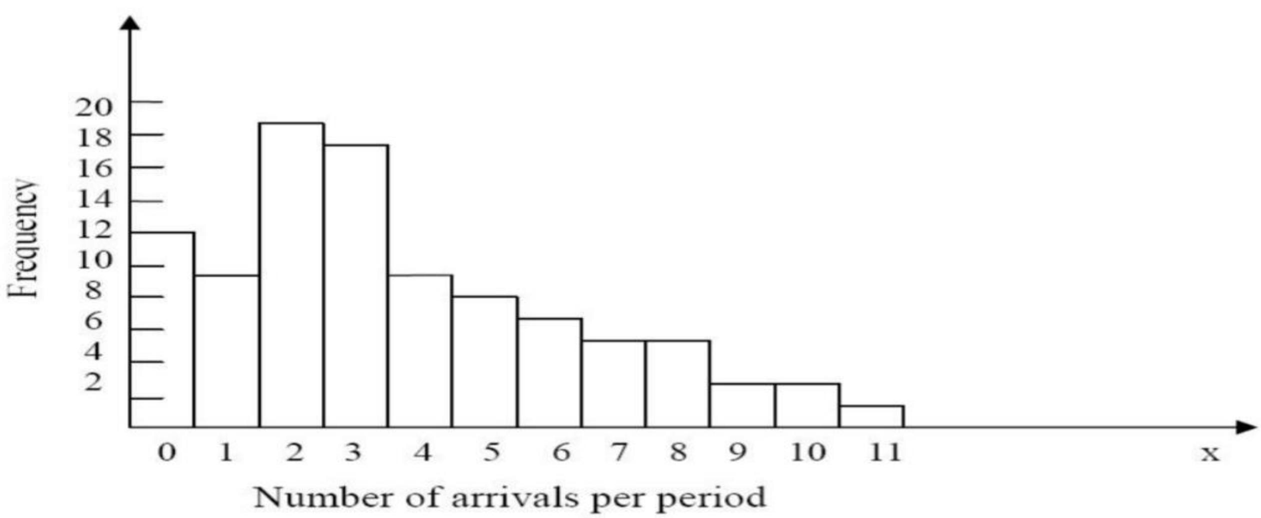


Figure 9.2 Histogram of number of arrivals per period.

Drawbacks of Histogram

- A histogram is not as useful for evaluating the fit of the chosen distribution. When there is a small number of data points, say 30 or fewer, a histogram can be rather ragged.
- Our perception of the fit depends on the widths of the histogram intervals. But, even if the intervals are chosen well, grouping data into cells makes it difficult to compare a histogram to a continuous probability density function.
- It uses only with continuous data.
- In Histogram, it is not easy to compare two data sets.
- The use of intervals in the Histogram prevents the calculation of an exact measure of central tendency.

Advantages of Q-Q plot

- A quantile-quantile (q - q) plot is a useful tool for evaluating distribution fit.
- The q - q plot can also be used to compare two samples of data to see whether they can be represented by the same distribution (that is, that they are homogeneous).
- Since Q-Q plot is like probability plot. So, while comparing two datasets the sample size need not to be equal.
- Since we need to normalize the dataset, so we don't need to care about the dimensions of values.

8.a. Customers arriving at a busy bank counter in a 5 minutes period between 10 to 2 pm was recorded for days given below:

Arrivals/period	0	1	2	3	4	5	6	7	8	9	10
Frequency	15	12	10	10	8	7	5	4	3	2	4

Use Chi-Square test to check whether the data follows Poisson distribution at 5% level of significance. $\chi^2_{0.05,4} = 9.49$.

Ans:

Step 1:

The null hypothesis is Ho : The random variable is Poisson distributed.

The alternative hypothesis is H_1: The random variable is not Poisson distributed.

Step 2: Estimate $\hat{\alpha}$

$$\hat{\alpha} = \overline{X} = \frac{\sum_{j=1}^{k} f_j X_j}{n}$$

Where,

k is the number of distinct values of X

f_j is the observed frequency of the value x_j of X

$$\hat{\alpha} = \frac{(15\times0)+(12\times1)+(10\times2)+(10\times3)+(8\times4)+(7\times5)+(5\times6)+(4\times7)+(3\times8)+(2\times9)+(4\times10)}{80}$$

$$= \frac{269}{80}$$

$$\hat{\alpha} = 3.3625$$

Step 3: Theoretical hypothesized probability for each interval using pmf of poisson distribution is given by

$$p(x) = \begin{cases} \frac{e^{-\alpha}\alpha^x}{x!}, & x = 0,1,2\ldots \\ 0, & \text{Otherwise} \end{cases}$$

$$p(0) = \frac{e^{-3.3625}(3.3625)^0}{0!} = 0.0346$$

$$p(1) = \frac{e^{-3.3625}(3.3625)^1}{1!} = 0.1165$$

$$p(2) = \frac{e^{-3.3625}(3.3625)^2}{2!} = 0.1959$$

$$p(3) = \frac{e^{-3.3625}(3.3625)^3}{3!} = 0.2196$$

$$p(4) = \frac{e^{-3.3625}(3.3625)^4}{4!} = 0.1846$$

$$p(5) = \frac{e^{-3.3625}(3.3625)^5}{5!} = 0.1241$$

$$p(6) = \frac{e^{-3.3625}(3.3625)^6}{6!} = 0.0696$$

$$p(7) = \frac{e^{-3.3625}(3.3625)^7}{7!} = 0.0334$$

$$p(8) = \frac{e^{-3.3625}(3.3625)^8}{8!} = 0.0140$$

$$p(9) = \frac{e^{-3.3625}(3.3625)^9}{9!} = 0.0052$$

$$p(10) = \frac{e^{-3.3625}(3.3625)^{10}}{10!} = 0.0018$$

Step 4: Compute Expected value using $E_i = n \times p_i$,Where n is data size (n=80).

$E_0 = 80 \times 0.0346 = 2.768$

$E_1 = 80 \times 0.1165 = 9.320$

$E_2 = 80 \times 0.1959 = 15.672$

$E_3 = 80 \times 0.2196 = 17.568$

$E_4 = 80 \times 0.1846 = 14.768$

$E_5 = 80 \times 0.1241 = 9.928$

$E_6 = 80 \times 0.0696 = 5.568$

$E_7 = 80 \times 0.0334 = 2.672$

$E_8 = 80 \times 0.0140 = 1.120$

$E_9 = 80 \times 0.0052 = 0.416$

$E_{10} = 80 \times 0.0018 = 0.144$

Step 5: Compute critical value χ^2_0

$$\chi^2_0 = \sum \frac{(Oi - Ei)^2}{Ei}$$

Where,

Oi – Observed frequency

Ei – Expected frequency

Arrivals/Period	Observed Frequency Oi		Expected Frequency Ei		(Oi-Ei)	$(Oi-Ei)^2$	$\frac{(Oi-Ei)^2}{Ei}$
0	15	27	2.768	12.088	14.912	222.368	18.396
1	12		9.320				
2	10		15.672		-5.672	32.172	2.053
3	10		17.568		-7.568	57.275	3.260
4	8		14.768		-6.768	45.806	3.102
5	7		9.928		-2.928	8.573	0.864
6	5		5.568				
7	4		2.672				
8	3	18	1.120	9.920	8.080	65.286	6.581
9	2		0.416				
10	4		0.144				
	80						34.256

$\chi^2_0 = 34.256$

Step 6: The critical value that is degree of freedom

$\chi^2_{\alpha,k-s-1} = \chi^2_{0.05,6-1-1} = \chi^2_{0.05,4} = 9.49$

Where,

k – number of intervals after grouping

s – number of parameter for estimation

Step 7: Since $\chi^2_0 > \chi^2_{0.05,4}$

That is, 34.256 > 9.49

Reject Null hypothesis (Ho).

8.b. The time required for 30 different employs to compute and record the number of hours worked during week days given:

1.88	2.62	1.49	0.35	0.82	2.03	1.54	0.21	0.39	2.03
2.16	0.90	1.90	0.63	0.17	0.03	0.45	0.31	0.15	2.03
4.29	0.04	1.73	0.92	2.81	0.05	5.5	2.16	0.48	0.18

Use the Chi-Square to test the hypothesis that these service times are exponentially distributed at 5% of level of significance. Let the number of intervals be K = 6 and critical value 9.49. **(08 Marks)**

Ans:

Given,

Data size = n = 30

$\alpha = 0.05$

Number of intervals = k = 6

Step 1: From Hypotheses

Ho : The random variable is exponentially distributed.

H_1 : The random variable is not exponentially distributed.

Step 2: Estimate parameter (that is $\hat{\lambda}$) for Exponential distribution

$$\hat{\lambda} = \frac{1}{\bar{X}}$$

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{40.25}{30} = 1.34$$

$$\hat{\lambda} = \frac{1}{1.34} = 0.75$$

Step 3: Find end points for class intervals, in order to perform Chi-Square test with interval of equal probability.

The probability of each interval $p = \frac{1}{k} = \frac{1}{6} = 0.17$

The endpoints for each intervals computed from the cdf for exponential distribution.

$$a_i = -\frac{1}{\lambda} \ln(1-ip) \qquad i = 0,1,\ldots, k$$

where, $a_0 = 0$ and $a_k = a_6 = \infty$

$$a_1 = -\frac{1}{0.75} \ln(1-1\times0.17) = 0.25$$

$a_2 = -\frac{1}{0.75}$ ln(1-2×0.17) = 0.55

$a_3 = -\frac{1}{0.75}$ ln(1-3×0.17) = 0.95

$a_4 = -\frac{1}{0.75}$ ln(1-4×0.17) = 1.52

$a_5 = -\frac{1}{0.75}$ ln(1-5×0.17) = 2.53

$a_6 = \infty$

Step 4: Expected frequency in each class interval is given by

$E_i = \frac{n}{k} = \frac{30}{6} = 5$

Step 5: Construct a table Chi-Square goodness-of-fit Test for exponential distribution.

Class Interval	Range	Observed Frequency O_i	Expected Frequency E_i	$\frac{(O_i-E_i)^2}{E_i}$
[0,0.25)	0-0.25	7	5	0.8
[0.25,0.55)	0.26-0.55	5	5	0.0
[0.55,0.95)	0.56-0.95	4	5	0.2
[0.95,1.52)	0.96-1.52	1	5	3.2
[1.52,2.53)	1.53-2.53	9	5	3.2
[2.53,∞)	2.54-∞	4	5	0.2
		30	30	7.6

$\chi^2_0 = \sum \frac{(Oi-Ei)^2}{Ei} = 7.6$

Step 6: The critical value that is degree of freedom

$\chi^2_{\alpha,k-s-1} = \chi^2_{0.05,6-1-1} = \chi^2_{0.05,4} = 9.49$

Where,

k – number of intervals after grouping

s – number of parameter for estimation

Step 6: Compare χ^2_0 and $\chi^2_{0.05,4}$

Since $\chi^2_0 < \chi^2_{0.05,4}$ that is 7.6 < 9.49

The null hypothesis(Ho) is accepted.

Module-5

9.a. Explain the types of simulation with respect to output analysis. Give atleast two examples. **(08 Marks)**

Ans:

Types of simulation with respect to output Analysis:

i. Terminating or transient simulation
ii. Steady state simulation

i. Terminating or Transient Simulation:

- A terminating simulation is one that runs for some duration of time T_E, where E is specified event (or set of events) that stops the simulation.

Example 1:

Communication system consists of several components plus several backup components as shown in below figure.

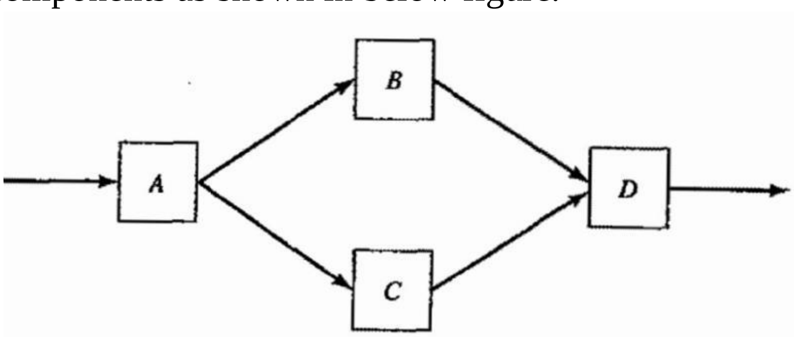


Fig: Example of Communications System

Consider system over a period of time T_E, until the system fails. The stopping event E is defined by E={A fails, or D fails, or (B and C both fail)}.Intial conditions are that all components are new at time 0.

The stopping time T_E is generally unpredictable in advance, in fact T_E is probably the output variable of interest as it represents the total time until the system breaks down. One goal of the simulation might be to estimate E(T_E), the mean time of system failure.

Example 2:

The Shady Grove Bank opens at 8:30 A.M. (time 0) with no customers present and 8 of the 11 tellers working (initial conditions) and closes at 4:30 P.M. (time

T_E= 480 minutes). Here the event E is merely the fact that the bank has been open for 480 minutes. The simulation analyst is interested in modeling the interaction between customers and tellers over the entire day, including the effect of starting up and of closing down at the end of the day.

ii. Steady State Simulation:

- A steady state simulation is a simulation whose objective is to study long run or steady state behavior of a non-terminating simulation.

Example 1:

Consider the widget-manufacturing process , beginning with the second shift when the complete production process is under way. It is desired to estimate long-run production levels and production efficiencies. For the relatively long period of 13 shifts, this may be considered as a steady-state simulation. To obtain sufficiently precise estimates of production efficiency and other response variables, the analyst could decide to simulate for any length of time, T_E (even longer than 13 shifts)-that is, T_E is not determined by the nature of the problem (as it was in terminating simulations); rather, it is set by the analyst as one parameter in the design of the simulation experiment.

Example 2:

HAL Inc., a large computer-service bureau, has many customers worldwide. Thus, its large computer system with many servers, workstations, and peripherals runs continuously, 24 hours per day. To handle an increased work load, HAL is considering additional CPUs, memory, and storage devices in various configurations. Although the load on HAL's computers varies throughout the day, management wants the system to be able to accommodate sustained periods of peak load. Furthermore, the time frame in which HAL's business will change in any substantial way is unknown, so there is no fixed planning horizon. Thus, a steady-state simulation at peak-load conditions is appropriate. HAL systems staff develops a simulation model of the existing system with the current peak work load and then explores several possibilities for expanding capacity. HAL is interested in long-run average throughput and utilization of each computer. The stopping time, T_E, is determined not by the nature of the problem, but rather by the simulation analyst, either arbitrarily or with a certain statistical precision in mind.

9.b. Explain the concepts of point estimation and interval estimation. **(08 Marks)**

Ans:

Point Estimation

Point Estimation for discrete time data:

Point estimator of θ based on the data $\{Y_1,----,Y_n\}$ is defined by

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

Where,

n→Sample size

Point estimator $\hat{\theta}$ is said to be unbiased for θ if its expected value is θ, that is if

$E(\hat{\theta}) = \theta$

In general, however

$E(\hat{\theta}) \neq \theta$

Point Estimation for continuous-time data:

The point estimator of Ø based on the data $\{Y(t), 0 \leq t \leq T_E\}$ is defined by

$$\hat{\phi} = \frac{1}{T_E}\int_0^{T_E} Y(t)\,dt$$

TE→run length of simulation

In general,

$E(\hat{\phi}) \neq \phi$

And $\hat{\phi}$ is said to be biased for Ø.

Usually, system performance measures can be put into the common framework of θ or Ø :

- Example: The proportion of days on which sales are lost through an out-ofstock situation, let:

$$Y_i = \begin{cases} 1, & \text{if out of stock on day } i \\ 0, & \text{Otherwise} \end{cases}$$

- Example: Proportion of time that the queue length is larger than k0

$$Y(t) = \begin{cases} 1, & \text{if } L_Q(t) > k_0 \\ 0, & \text{Otherwise} \end{cases}$$

Interval Estimation

- The interval Estimator of θ is the sample mean of R independent replication.

$$\bar{Y} = \sum_{i=1}^{R} \frac{\bar{Y}i}{R}$$

- $\bar{Y}$ is not θ, it is an estimate, and it has error confidence interval is measure of that error.
- Let $S^2 = \frac{1}{R-1} \sum_{i=1}^{R} (Y_i - \bar{Y}_{..})^2$ be the sample variance across replication.

- Confidence interval, which assumes the Yi are

$$\bar{Y}_{..} \pm t_{\alpha/2,R-1} \frac{S}{\sqrt{R}}$$

$t_{\alpha/2,R-1}$ – Quantile of the t distribution with R-1 degrees of freedom

$t_{\alpha/2,R-1} \frac{S}{\sqrt{R}}$ will tend to get smaller as R increases.

- Normal theory prediction interval is

$$\bar{Y} \pm t_{\alpha/2,R-1} S \sqrt{1 + \frac{1}{R}}$$

The length of this interval will not go to 0 as R increases. In fact in the limit it becomes $\theta \pm Z_{\alpha/2}$

- Prediction interval is a measure of risk.
- Confidence interval is a measure of error.

10.a. Explain in detail about the model building, verifying and validation in the building process through a diagram. **(08 Marks)**

Ans:

Model Building, Verification, and Validation

- The first step in model building consists of observing the real system and the interactions among its various components and collecting data on its behavior.

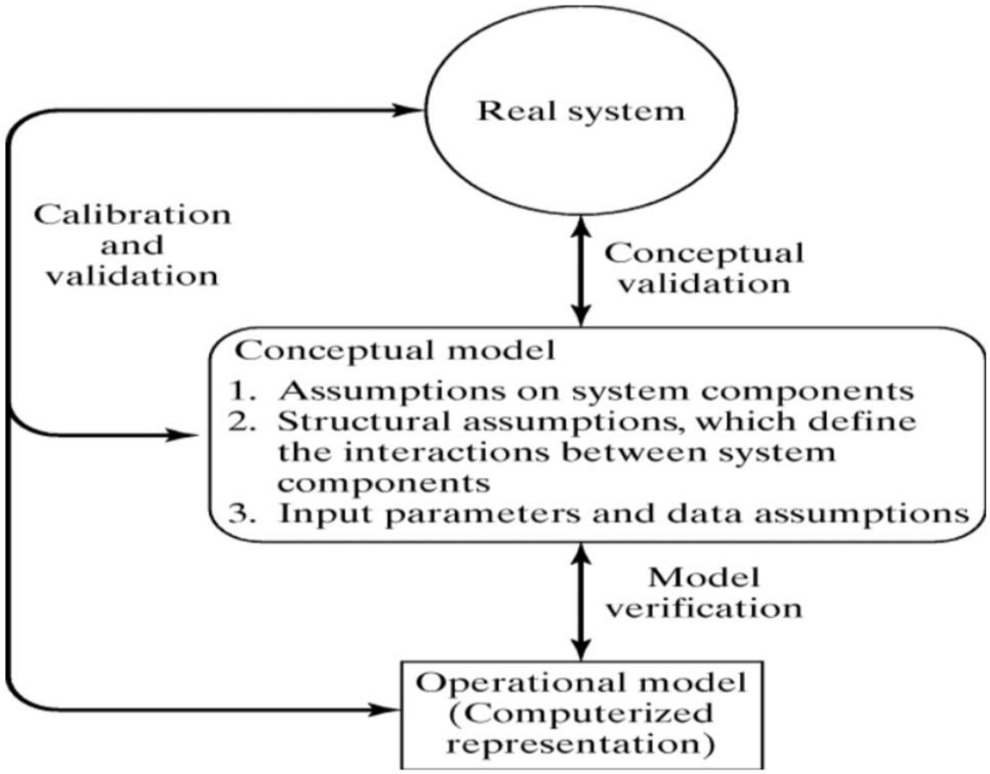


Figure 1 0.1 Model building, verification, and validation.

- Persons familiar with the system,. or any subsystem, should be questioned to take advantage of their special knowledge.
- Operators, technicians, repair and maintenance personnel, engineers, supervisors, and managers understand certain aspects of the system that might be unfamiliar to others.
- As model development proceeds, new questions may arise, and the model developers will return to this step of learning true system structure and behavior.
- The second step in model building is the construction of a conceptual model – a collection of assumptions on the components and the structure of the system, plus hypotheses on the values of model input parameters.
- As is illustrated by Figure 10. 1, conceptual validation is the comparison of the real system to the conceptual model.
- The third step is the implementation of an operational model, usually by using simulation software and incorporating the assumptions of the conceptual model into the worldview and concepts of the simulation software.
- In actuality, model building is not a linear process with three steps. Instead; the model builder will return to each of these steps many times while building, verifying, and validating the model.

- Figure 10.1 depicts the ongoing model building process, in which the need for verification and validation causes continual comparison of the real system to the conceptual model and to the operational model and induces repeated modification of the model to improve its accuracy.

10.b. Explain 3-steps approach to validation of simulation models by Naylor and Finger. **(08 Marks)**

Ans:

As an aid in the validation process, Naylor and Finger formulated a three step approach which has been widely followed:-

1. Build a model that has high face validity.
2. Validate model assumptions.
3. Compare the model input-output transformations to corresponding inputoutput transformations for the real system.

Face Validity:

- The first goal of the simulation modeler is to construct a model that appears reasonable on its face to model users and others who are knowledgeable about the real system being simulated.
- The users of a model should be involved in model construction from its conceptualization to its implementation to ensure that a high degree of realism is built into the model through reasonable assumptions regarding system structure, and reliable data.
- Another advantage of user involvement is the increase in the models perceived validity or credibility without which manager will not be willing to trust simulation results as the basis for decision making.
- Sensitivity analysis can also be used to check model's face validity.
- The model user is asked if the model behaves in the expected way when one or more input variables is changed.

Validation of Model Assumptions:

- Model assumptions fall into two general classes: structural assumptions and data assumptions.
- **Structural assumptions:**
    i. Involves how system operates.

ii. Includes simplifications & abstractions of reality.

Example: consider customer queueing and service facility in a bank. Structural assumptions are customer waiting in one line versus many lines. Customers are served according FCFS versus priority.

- **Data assumptions:**

i. based on the collection of reliable data and correct statistical analysis of the data.

Example: Inter arrival time of customers, service times of commercial accounts.

Validating Input-Output Transformation

- In this phase of validation process the model is viewed as input –output transformation.
- That is, the model accepts the values of input parameters and transforms these inputs into output measure of performance. It is this correspondence that is being validated.
- Instead of validating the model input-output transformation by predicting the future ,the modeler may use past historical data which has been served for validation purposes that is, if one set has been used to develop calibrate the model, its recommended that a separate data test be used as final validation test.

Input-Output Validation: Using Historical Input Data

- To conduct a validation test using historical input data, it is important that all input data and all the system response data, such as average delay, be collected during the same time period.
- If not taken on same time then, comparison of model responses to system responses could be misleading.
- Implementation of this technique could be difficult for a large system because of the need for simultaneous data collection of all input variables and those response variables of primary interest.

Input-Output Validation: Using a Turing Test

- when no statistical test is readily applicable, Persons knowledgeable about system behavior can be used to compare model output to system output.

- For example, suppose that five reports of system performance over five different days are prepared, and simulation outputs are used to produce five "fake" reports.
- The 10 reports should all be in exactly in the same format.
- The ten reports are randomly shuffled and given to the engineers, who are asked to decide which reports are fake and which are real.
- If engineer identifies substantial number of fake reports the model builder questions the engineer and uses the information gained to improve the model. or else the modeler will conclude that this test provides no evidence of model inadequacy.
- This type of validation test is called as TURING TEST.