Unit 2: System Simulation

Er.Nipun Thapa

The technique of simulation monte-carlo method

- The Monte Carlo simulation is a mathematical technique that predicts possible outcomes of an uncertain event.
- Computer programs use this method to analyze past data and predict a range of future outcomes based on a choice of action.
  - For example, if you want to estimate the first month’s sales of a new product, you can give the Monte Carlo simulation program your historical sales data. The program will estimate different sales values based on factors such as general market conditions, product price, and advertising budget.

Monte Carlo

- Monte Carlo simulations are used to model the probability of different outcomes in a process that cannot easily be predicted due to the intervention of random variables. It is a technique used to understand the impact of risk and uncertainty in prediction and forecasting models.
- Monte Carlo simulation can be used to tackle a range of problems in virtually every field such as finance, engineering, supply chain, and science.
- Monte Carlo simulation is also referred to as multiple probability simulation.

Monte Carlo: Estimating the value of Pi using Monte Carlo

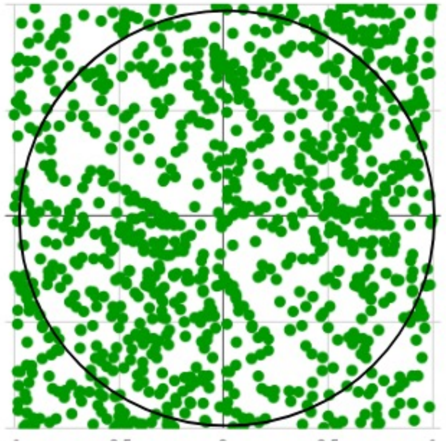
Monte Carlo estimation :

- Monte Carlo methods are a broad class of computational algorithms that rely on repeated random sampling to obtain numerical results. One of the basic examples of getting started with the Monte Carlo algorithm is the estimation of Pi.

Monte Carlo: Estimating the value of Pi using Monte Carlo

Estimation of Pi

The idea is to simulate random (x, y) points in a 2-D plane with domain as a square of side 1 unit. Imagine a circle inside the same domain with same diameter and inscribed into the square. We then calculate the ratio of number points that lied inside the circle and total number of generated points. Refer to the image below:



Monte Carlo: Estimating the value of Pi using Monte Carlo

We know that area of the square is 1 unit sq while that of circle is $\pi * \left(\frac{1}{2}\right)^2 = \frac{\pi}{4}$.

Now for a very large number of generated points,

$$\frac{\text{area of the circle}}{\text{area of the square}} = \frac{\text{no. of points generated inside the circle}}{\text{total no. of points generated or no. of points generated inside the square}}$$

that is,

$$\pi = 4 * \frac{\text{no. of points generated inside the circle}}{\text{no. of points generated inside the square}}$$

Monte Carlo: Estimating the value of Pi using Monte Carlo

Estimation of Pi

The beauty of this algorithm is that we don't need any graphics or simulation to display the generated points. We simply generate random (x, y) pairs and then check if $x^2+y^2<=1$.

If yes, we increment the number of points that appears inside the circle. In randomized and simulation algorithms like Monte Carlo, the more the number of iterations, the more accurate the result is.

Thus, the title is "**Estimating** the value of Pi" and not "Calculating the value of Pi". Below is the algorithm for the method:

Monte Carlo: Estimating the value of Pi using Monte Carlo

The Algorithm

1. Initialize circle_points, square_points and interval to 0.
2. Generate random point x.
3. Generate random point y.
4. Calculate d = x*x + y*y.
5. If d <= 1, increment circle_points.
6. Increment square_points.
7. Increment interval.
8. If increment < NO_OF_ITERATIONS, repeat from 2.
9. Calculate pi = 4*(circle_points/square_points).
10. Terminate.

Monte-Carlo

- Some example of problem solveing using Monte-Carlo simulation are:
  - To find the area of irrrgular surface
  - Numerical integration
  - Random walk problem

1.Finding areas using the Monte Carlo Method

- The Monte Carlo Method gets its name from the city of Monte Carlo and the games of chance that are played in the casinos there.
- In mathematics this name is used whenever a problem is solved by a method that uses random numbers.
- The Monte Carlo method has been used in the following situations:
  1. in computer simulations of processes that involve some element of randomness, such as the diffusion of neutrons out of a nuclear reactor or customers arriving at a queue,
  2. in computer games to decide what the computer-generated characters in the game will do next,
  3. in calculating the areas of irregular shapes.

1.Finding areas using the Monte Carlo Method

- **Finding Areas:** It is generally very difficult to find the area of an irregular shape such as the one shown to the right. We will show how it can easily be done using the Monte Carlo Method.

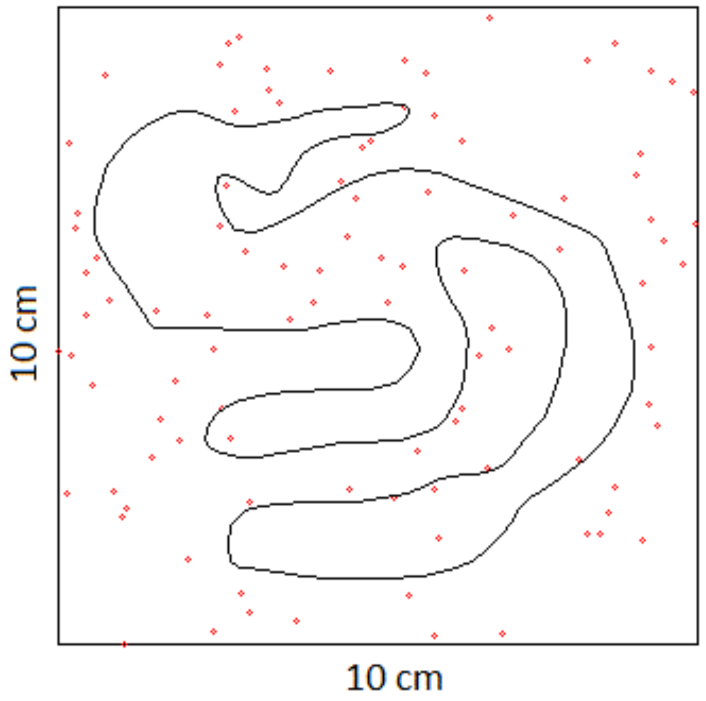
1. Calculus (integration) could be used if the upper and lower boundaries of the shape could be expressed as functions. But for the shape shown that would require breaking up the shape into many smaller shapes and calculating their areas.
2. One could draw the shape on a piece of paper, weigh the paper, cut out the shape, and weigh it. Then the area is proportional to the weight and given by this formula:

$$\text{area of shape} = \text{area of the paper} \times \frac{\text{weight of the shape}}{\text{weight of the paper}}$$

3. One could use an instrument called a planimeter.

1.Finding areas using the Monte Carlo Method

- Here are the steps to the Monte Carlo method:
  1. Put the object inside a rectangle of given area. Suppose that this rectangle has an area of 100 cm^2.
  2. Place a given number of points, say 100, at random locations inside the rectangle.
  3. Count the number of random points that lie inside the object



1.Finding areas using the Monte Carlo Method

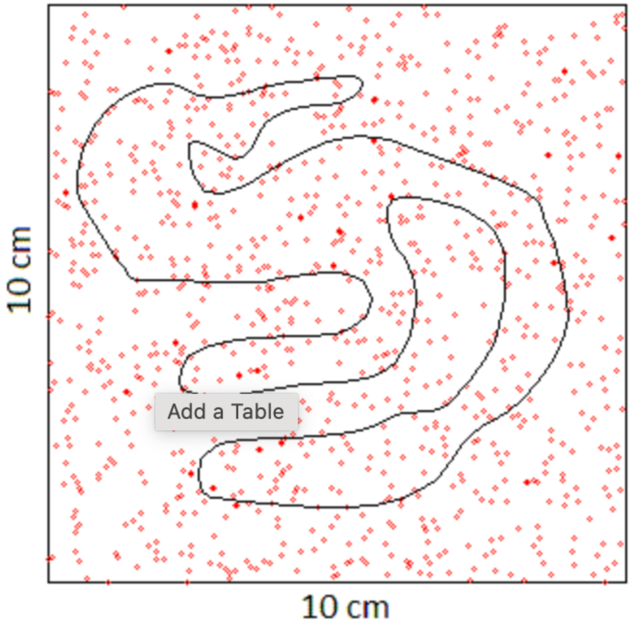
4. The area of the object is proportional to the number of points that lie inside and is given by this formula:

$$\text{area of object} = \text{area of rectangle} \times \frac{\text{number of points inside object}}{\text{total number of points in rectang}}$$

If you count you will find that 22 points lie inside the object. Thus our estimate of the area is:

$$\text{area of object} = 100\ \text{cm}^2 \times \frac{22}{100} = 22\ \text{cm}^2$$

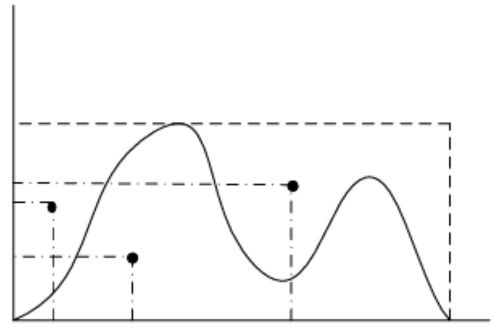
5. If we add another 900 points at random inside the rectangle (for a total of 1000) we get an improved estimate of the area. We now find that 280 points lie inside the object. This puts the area of the object at:



$$\text{area of object} = 100\ \text{cm}^2 \times \frac{280}{1000} = 28\ \text{cm}^2$$

2. Numerical Integration Using Monte Carlo Method

Numerical Integration Using Monte Carlo Method



Given a single valued function f(x) as shown in the figure above, the goal is to determine the integral

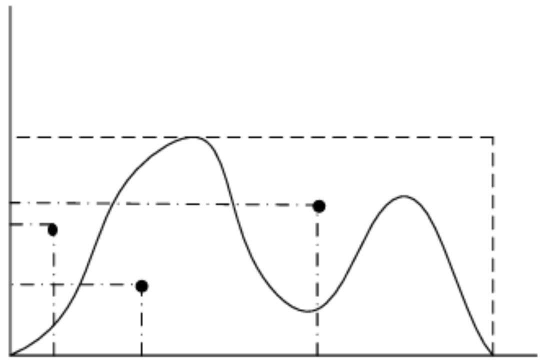
$$I = \int_{x_1=0}^{x_2=x_{max}} f(x)\,dx$$

The above integral is the area under the curve represented by a solid line in the above figure.

2. Numerical Integration Using Monte Carlo Method

Numerical Integration Using Monte Carlo Method

$$I = \int_{x_1=0}^{x_2=x_{max}} f(x)\,dx$$



In order to use the Monte method, we need two parameters:

(I) Range of integration. In the above case it runs from x_1 =0 to x_2= x_{max}. Therefore the full range of integration:

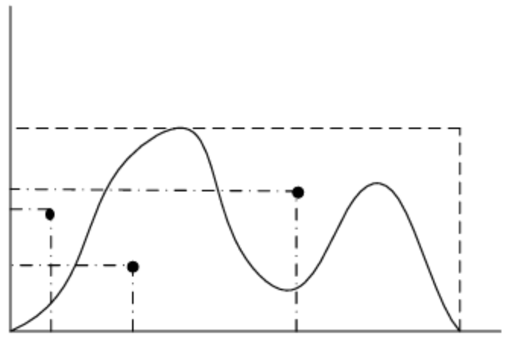
$$x_2 - x_1 = x_{max} - 0 = x_{max}$$

(II) Maximum value of the function f(x) in the range of integration: f_{max}. Values larger than the exact f_{max} are acceptable.

2. Numerical Integration Using Monte Carlo Method

Numerical Integration Using Monte Carlo Method

$$I = \int_{x_1=0}^{x_2=x_{max}} f(x)\,dx$$



The parameters f_{max} and x_{max} define the sides of a rectangle as shown above. The area of the rectangle is given by:

$$Area\_A = f_{max} * x_{max}$$

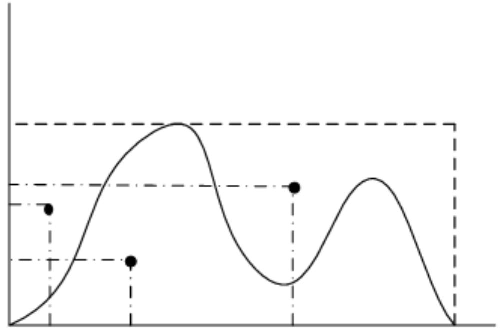
The integral I of the function f(x) is part of the rectangle defined by f_{max} and x_{max}.

Using Monte Carlo to perform the integration amounts to generating a random sequence of points (xr , fr) and checking to see if the points are under the curve defined by f(x) or not.

2. Numerical Integration Using Monte Carlo Method

Numerical Integration Using Monte Carlo Method

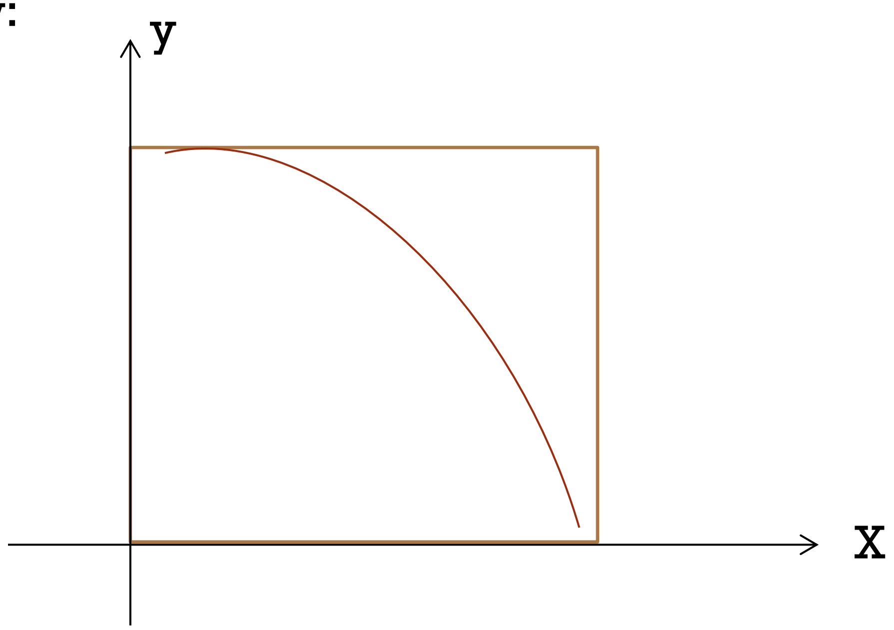
$$I = \int_{x_1=0}^{x_2=x_{max}} f(x)\,dx$$



1. generate a pair of random numbers r_1 and r_2. Note that: $0 \le r_1 \le 1$ and $0 \le r_2 \le 1$
2. Calculate $x_r = r_1 * x_{max}$ and $f_r = r_2 * f_{max}$.
3. Check if the point is under the curve. Check if $f_r \le f(x_r)$
4. If the condition in step (3) is true, then accept the point and update the counter for points under curve (N_accept).
   Note that out of the three points in the above figure only point (3) falls below the curve. For points (1) and (2) $f_r \rangle f(x_r)$
5. Repeat steps (1) through (4) large number of times (N_trials).
   Typical values of N_trials range from 10,000 to 1,000,000.
6. Compute the integral I (=Area under the curve): $I = \frac{N\_accept}{N\_trials} * (f_{max} * x_{max})$

3.Determine problem through random number

- The application of monte-Carlo method for the solution of P(n) requires converting the deterministic model into stocastic model.
- Consider a quadrant of unit circle as shown in fig below:



3.Determine problem through random number

→ All point satisfy the eqn

$x^2+y^2 \le 1,\ x, y \ge 0$ lie inside the circle quadrant.

The eqn can be written as

$$y^2 \le 1-x^2$$

$$\therefore y \le \sqrt{1-x^2}$$

Now if (r_1, r_2) is a pair of random number in range $r(0,1)$, we call this pair acceptable if

$$r_2 \le \sqrt{1-r_1^2}$$

Distributed lag model

- ***Models that have the property of changing only at fixed interval of time.***
- It is used to predict current values of a dependent variable based on both the current values of an explanatory variable (independent variable) and the lagged (past period) values of this explanatory (helpful) variable.
- In economic studies some economic data are collected over uniform time interval such as a month or year. This model consists of linear algebraic equations that represent continuous system but data are available at fixed points in time.

Distributed lag model

- ***Any variable that can be expressed in the form of its current value and one or more previous value is called lagging variable.*** And hence this model is given the name distributed lag model.
- The variable in a previous interval is denoted by attaching –n suffix to the variable. Where –n indicate the nth interval.

Advantages of distributed lag model

- Simple to understand and can be computed by hand, computers are extensively used to run them.
- There is no need for special programming language to organize simulation task.

Distributed lag model

3.16 DISTRIBUTED LAG MODELS [UPTU 2006-07, 2007-08]

If model becomes larger or more complicated then numeric computation technique can be assumed tha record keeping quickly becomes complicated. So addition of a computer and programming language i: important. However, we have some simple technique in simulation that can be applied without difficulty even for large models. If the events all occur synchronously, at fixed intervals of time, the computatio remain simple.

Properties of Distributed Lag Model

1. Model changes only at fixed intervals of times.
2. Model is based on current values of variables on other current values and values that occurred i previous intervals.
3. These are used mainly in econometric studies because some economic data are collected at fixe interval of time, *e.g.*, month or a year.
4. These model consist the rule of linear algebraic equation.
5. They represent a continuous system, but one in which the data is only available at fixed points i time.

What is Lagged Variable?

If current values of variable can be represent in terms of values of one or more previous intervals. Th variable is called as **lagged variable.** Its value in a previous interval is denoted by attaching the suffix n, into the variable.

n = interval.

If $n = 1$, it means previous interval

If $n = 2$, it means One prior to above, and so on.

Consider following equations

$$C = 20 + 0.7\,(Y_R - T_X) \quad \ldots($$

Distributed lag model

$$I = 2 + 0.1\,Y_R \quad \text{...(ii)}$$
$$T_X = 0.2\,Y_R \quad \text{...(iii)}$$
$$Y_R = C + I + G \quad \text{...(iv)}$$

where C is consumption, I is investment, T_X is tax, G is government expenditure, and Y_R is national income. All quantities are expressed in billions of dollars.

Let simplify above equation by putting equation (*iii*) in equation (*i*)

$$C = 20 + 0.7\,(Y_R - 0.2\,Y_R)$$
$$C = 20 + 0.7\,(0.8\,Y_R)$$
$$C = 20 + 0.56\,Y_R \quad \text{...(v)}$$

Put this equation (*v*) in equation (*iv*) and get

$$Y_R = 20 + 0.56\,Y_R + I + G$$
$$Y_R - 0.56\,Y_R = 20 + I + G$$
$$0.44\,Y_R = 20 + I + G$$
$$Y_R = 45.45 + 2.27\,(I + G)$$

The set of equations can then be written in the following form:

$$I = 2 + 0.1\,Y_{R-1} \quad \text{...(i)}$$
$$Y_R = 45.45 + 2.27\,(I + G) \quad \text{...(ii)}$$
$$T_X = 0.2\,Y_R \quad \text{...(iii)}$$
$$C = 20 + 0.7\,(Y_R - T_X) \quad \text{...(iv)}$$

Important

1. There is only one variable Y is taken as lagged variable

Distributed lag model

$$C = 20 + 0.7\,(Y_R - T_X) \qquad ...(iii)$$

...(iv)

Important

1. There is only one variable Y_R is taken as lagged variable.
2. To solve this equation we must have initial value of Y_{R-1}.

Example 4. For the above equation values of G are supplied for all year (time-intervals).

Year	*G*
1991	15
1992	20
1993	25
1994	30
1995	35

Initial value of Y_{R-1} is 75. Then calculate consumption of Ist year.

Solution: Let take Y_R as lagged variable in equation (*i*), then

$$I = 2 + 0.1\,Y_{R-1} \qquad ...(i)$$

$$Y_R = 45.45 + 2.27\,(I + G) \qquad ...(ii)$$

$$T_X = 0.2\,Y_R \qquad ...(iii)$$

$$C = 20 + 0.7\,(Y_R - T_X) \qquad ...(iv)$$

Given $Y_{R-1} = 75$ put this in equation (*i*)

$$I = 2 + 0.1 * 75$$

$$I = 2 + 7.5 = 9.5$$

$$Y_R = 45.45 + 2.27\,(9.5 + 15)$$

Distributed lag model

$$Y_R = 45.45 + 2.27 \times 24.5$$

$$Y_R = 45.45 + 55.615 = 101.065$$

$$T_X = 0.2 * 101.065$$

$$T_X = 20.213$$

$$C = 20 + 0.7\ (101.065 - 20.213)$$

$$C = 20 + 0.7 * 80.852$$

$$C = 20 + 56.5964$$

$$C = 76.596$$

Note: C, I, T_X, G and Y_R are expressed in billions of dollars.

This is the consumption for the 1st year.

Distributed lag model

$$Y_R = 45.45 + 2.27 \times 24.5$$
$$Y_R = 45.45 + 55.615 = 101.065$$
$$T_X = 0.2 * 101.065$$
$$T_X = 20.213$$
$$C = 20 + 0.7\ (101.065 - 20.213)$$
$$C = 20 + 0.7 * 80.852$$
$$C = 20 + 56.5964$$
$$C = 76.596$$

Note: C, I, T_X, G and Y_R are expressed in billions of dollors.

This is the consumption for the 1st year.

Note: Taking the current value of Y_R as the new value of the lagged variable Y_{R-1}, and repeat same calculation for the next interval of year or time.

Cobweb Model

3.17 COBWEB MODEL

The Cobweb model was developed by the Hugorian economist, Nicholas Koldoar. It is an economic model of cyclical supply and demand in which there is a legs in producers response to change of price.

Consider following three points:

- If demand is less than the supply, the fluctuations would increase in magnitude per cycle, so diagram for each cycle would look like an outward spiral. This is called as an unstable Cobweb model.
- If the demand is more than the supply, fluctuations decrease in magnitude per cycle, so diagram for each cycle would look like an inward spiral. This is called as a stable cobweb model.
- Fluctuations may also remain of constant magnitude, so a diagram would produce a simple quadrangle. Such a result would be given from unit price of both supply and demand.

Points (*a*) and (*b*) is the combination of spiral and the supply and demand curves often looks like a cobweb, so we call this theory as cobweb model.

The most frequent "use" for first order linear differential equations supply S and demand D, to market price, P. Generally supply should be dependent upon the price from the previous marketing period. Again assuming market is cleared, the market model in distributed lag form is as follows:

Consider following equation:

$$D = a - b\,P \qquad \ldots(1)$$

$$S = c + d\,P_{-1} \qquad \ldots(2)$$

$$D = S \qquad \ldots(3)$$

Initial value of P_0 is given. This becomes value of P_{-1} used to calculate the value of next interval and again this new value becomes P_1 and again calculate the value of next interval and so on.

Consider following initial values:

$$P_0 = 1.0$$

$$a = 12.4$$

$$b = 1.2$$

$$c = 1.0$$

Cobweb Model

$$d = 0.9$$

Equation can be modified as

$$D = a - b\,P_1$$
$$S = C + d\,P_0$$
$$D = S$$

We will calculate new value of P_1 with the help of P_0

$$a - b\,P_1 = c + d\,P_0$$
$$a - c = d\,P_0 + b\,P_1$$
$$a - c = 0.9 * 1 + 1.2 * P_1$$
$$12.4 - 1.0 = 0.9 + 1.2\,P_1$$
$$11.4 = 0.9 + 1.2\,P_1$$
$$11.4 - 0.9 = 1.2\,P_1 = 10.5/1.2 = 8.75$$

Now again by this P_1, we will calculate value for next interval and let new value is P_2.

$$11.4 = d\,P_1 + bP_2$$
$$11.4 = 0.9 \times 8.75 + 1.2 \times P_2$$
$$11.4 - 7.875 = 1.2\,P_2 \text{ or } P_2 = 2.9375$$

Similarly calculate following value of P.

$P_3 = 7.3$	$P_9 = 5.75$	$\mathbf{P_{13}} = 5.48$
$P_4 = 4.027$	$P_{10} = 5.18$	
$P_5 = 6.5$	$P_{11} = 5.61$	
$P_6 = 4.640$	$P_{12} = 5.29$	$\mathbf{P_{16}}\cdots$
$P_7 = 6.01$	$P_{13} = 5.53$	
$P_8 = 4.99$	$P_{14} = 5.35$	$\mathbf{P_{17}}\cdots$

Cobweb Model

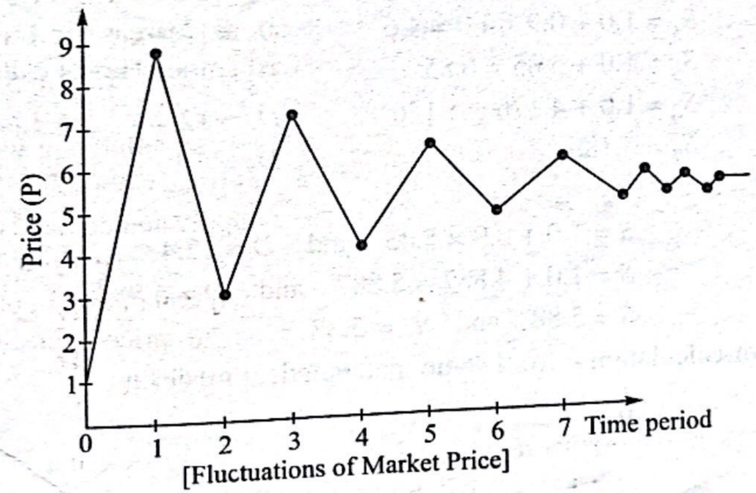


Fig. 3.6. Fluctuations of market price.

This is stable market because amplitude is continuously decreasing and after some intervals it becomes straight line. Cobweb graph is illustrated in Fig. 3.7 for the stable case just considered. Two straight lines plot the linear relationship representing supply and demand as we have plotted before in static mathematical model.

Cobweb Model

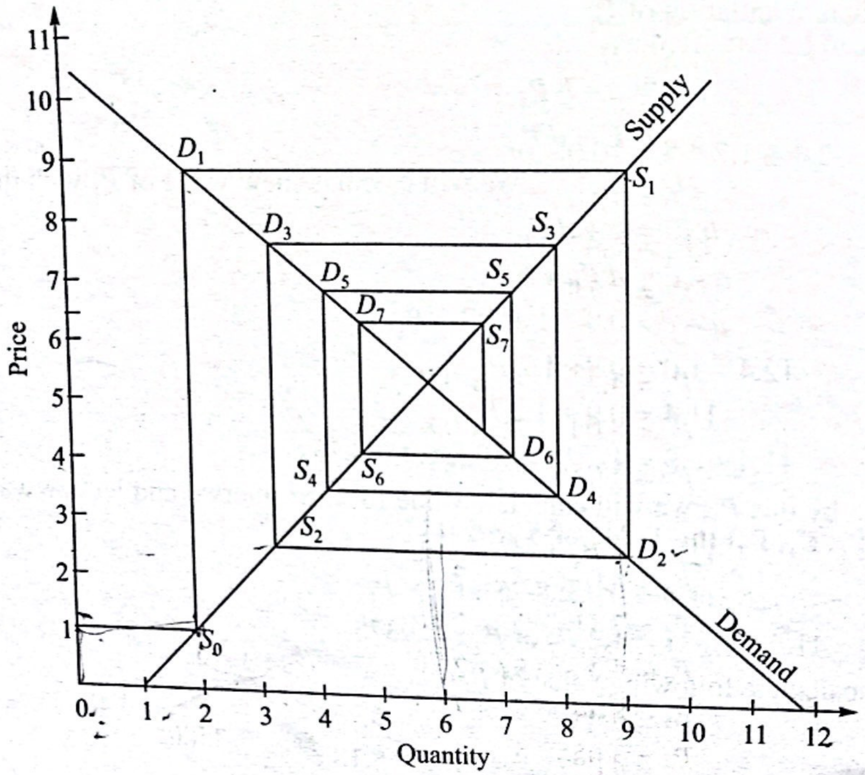


Fig. 3.7. Cobweb model of a market economy.

$P_0 = 1.0$
$P_1 = 8.75$
$P_2 = 2.94$
$P_3 = 7.3$
$P_4 = 4.0$
$P_5 = 6.5$
$P_6 = 4.64$
and $P_7 = 6.01$
Similarly at $P = 5.43$

$S_0 = 1.0 + 0.9 * 1 = 1.9$
$S_1 = 1.0 + 8.75 * 1 = 8.87$
$S_2 = 1.0 + 2.646 = 3.646$
$S_2 = 1.0 + 0.9 * 7.3 = 7.57$
$S_4 = 1.0 + 0.9 * 4.0 = 4.6$
$S_5 = 1.0 + 5.85 = 6.85$
$S_6 = 1.0 + 4.176 = 5.176$
$S_7 = 5.188$

$D_0 = 12.4 - 1.2 * 1.0 = 11.2$
$D_1 = 12.4 - 1.2 * 875 = 1.9$
$D_2 = 12.4 - 1.2 * 2.94 = 8.87$
$D_3 = 12.4 - 1.2 * 7.3 = 3.646$
$D_2 = 12.4 - 1.2 * 4.0 = 7.6$
$D_5 = 12.4 - 1.2 * 6.5 = 4.6$
$D_6 = 12.4 - 1.2 * 4.64 = 6.83$
$D_7 = 5.196$, and so on

Cobweb Model

$$S = 1.0 + 0.9 \times 5.43 \quad \text{and} \quad D = 12.4 - 1.2 * 5.43$$
$$S = 1.0 + 4.887 = 5.887 \quad \text{and} \quad D = 5.884$$
$$S = 5.88 \quad \text{and} \quad D = 5.88$$

Another way of calculation – from static mathematical model

$$P = \frac{a - c}{b + d}$$

$$P = \frac{12.4 - 1.0}{1.2 + 0.9} = \frac{11.4}{2.1} = 5.428 \quad \text{or} \quad 5.43.$$

Example 5. Draw market fluctuations of market price for following values

$P_0 = 5.0$ $\quad a = 10.0$ $\quad b = 0.9$ $\quad c = -2.4$ $\quad d = 1.2$

Solution: Given equation

$$D = a - bP_1$$
$$S = c + d\,P_0$$
$$S = D$$

Cobweb Model

$P_0 = 5.0$ then calculation of P_1

$$S = D$$

$$c + d\,P_0 = a - b\,P_1$$

$$-2.4 + 1.2 * 5 = 10.0 - 0.9\,P_1$$

$$3.6 = 10.0 - 0.9\,P_1$$

or

$$0.9\,P_1 = 6.4$$

$$P_1 = 7.11$$

Similarly

$$P_2 = 4.3 \quad P_5 = 9.7038 \quad P_8 = 3.1$$

$$P_3 = 8.04 \quad P_6 = 0.839 \quad P_9 = 9.6$$

$$P_4 = 3.055 \quad P_7 = 12.65$$

At these Point (P_0, P_1, P_7) the Value of *S* and *D*

$$P_0 = 5.0 \quad S_0 = 3.6 \quad D_0 = 5.5$$

$$P_1 = 7.11 \quad S_1 = 6.13 \quad D_1 = 3.60$$

$$P_2 = 4.3 \quad S_2 = 2.76 \quad D_2 = 6.13$$

$$P_3 = 8.04 \quad S_3 = 7.25 \quad D_3 = 2.76$$

$$P_4 = 3.055 \quad S_4 = 1.26656 \quad D_4 = 7.25005$$

Cobweb Model

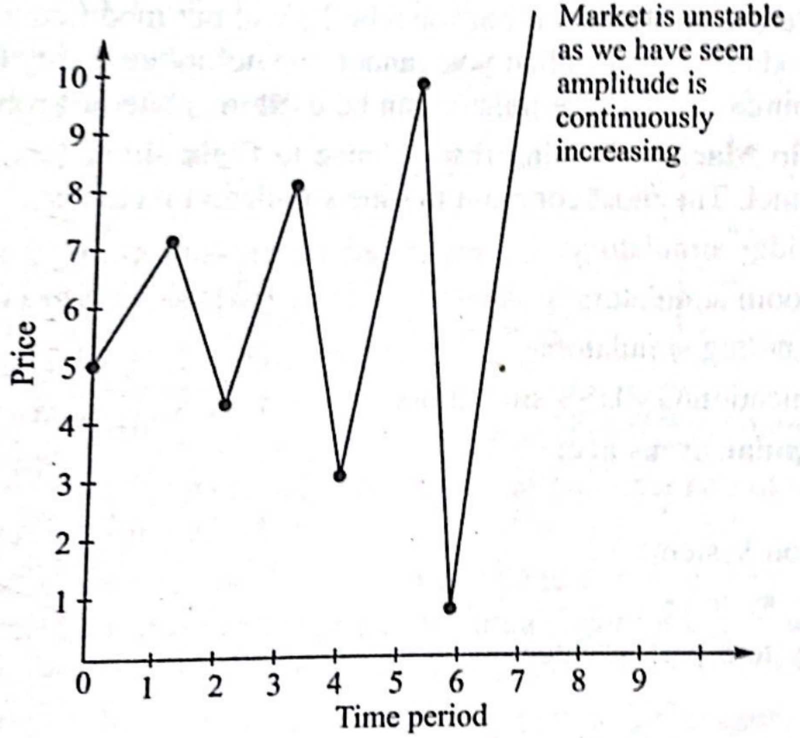


Fig. 3.8. Market fluctuations.

In above Fig. 3.8, market is **unstable,** as we have seen amplitude is continuously increasing.

Steps in a Simulation study

1. **Problem formulation**
2. **Setting of objectives and overall project plan**
3. **Model conceptualization**
4. **Data collection**
5. **Model translation**
6. **Verified**
7. **Validated**
8. **Experimental Design**
9. **Production runs and analysis**
10. **More runs**
11. **Documentation and reporting**
12. **Implementation**

Steps in a Simulation study

1. Problem formulation (Invention)

Every study begins with a *statement of the problem*, provided by policy makers.

Analyst ensures its clearly understood. If it is developed by analyst policy makers should understand and agree with it.

Steps in a Simulation study

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 it is appropriate, the overall project plan should include

- A statement of the ***alternative systems***
- A method for evaluating the effectiveness of these alternatives
- Plans for the study in terms of the number of people involved
- Cost of the study
- The number of days required to accomplish each phase of the work with the anticipated (predicted) results.

Steps in a Simulation study

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 an ability

- To abstract the essential features of a problem
- To select and modify basic assumptions that characterize the system
- 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 conceptualization enhance the quality of the resulting model and increase the confidence of the model user in the application of the model.

Steps in a Simulation study

4. Data collection

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

Objective kind of data are to be collected.

Steps in a Simulation study

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.

Steps in a Simulation study

6. Verified

It pertains to be computer program and checking the performance.

If the input parameters and logical structure are correctly represented, verification is completed.

Steps in a Simulation study

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.

Steps in a Simulation study

8. Experimental Design

The alternatives that are to be simulated must be determined. Which alternatives to simulate may be a function of runs.

For each system design, decisions need to be made concerning

- Length of the initialization period
- Length of simulation runs
- Number of replication to be made of each run

Steps in a Simulation study

9. Production runs and analysis

They are used to estimate measures of performance for the system designs that are being simulated.

Steps in a Simulation study

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.

Steps in a Simulation study

11. Documentation and reporting

Two types of documentation.

- Program documentation
- Process documentation

Program documentation

Can be used again by the same or different analysts to understand how the program operates. ***Further modification will be easier.*** Model users can change the input parameters for better performance.

Process documentation

Gives the history of a simulation project. The result of all analysis should be reported clearly and concisely in a final report. 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.

Steps in a Simulation study

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 can be broken into 4 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 continuing interplay is required among the steps
- Exclusion of model user results in implications during implementation

The simulation model building can be broken into 4 phases:

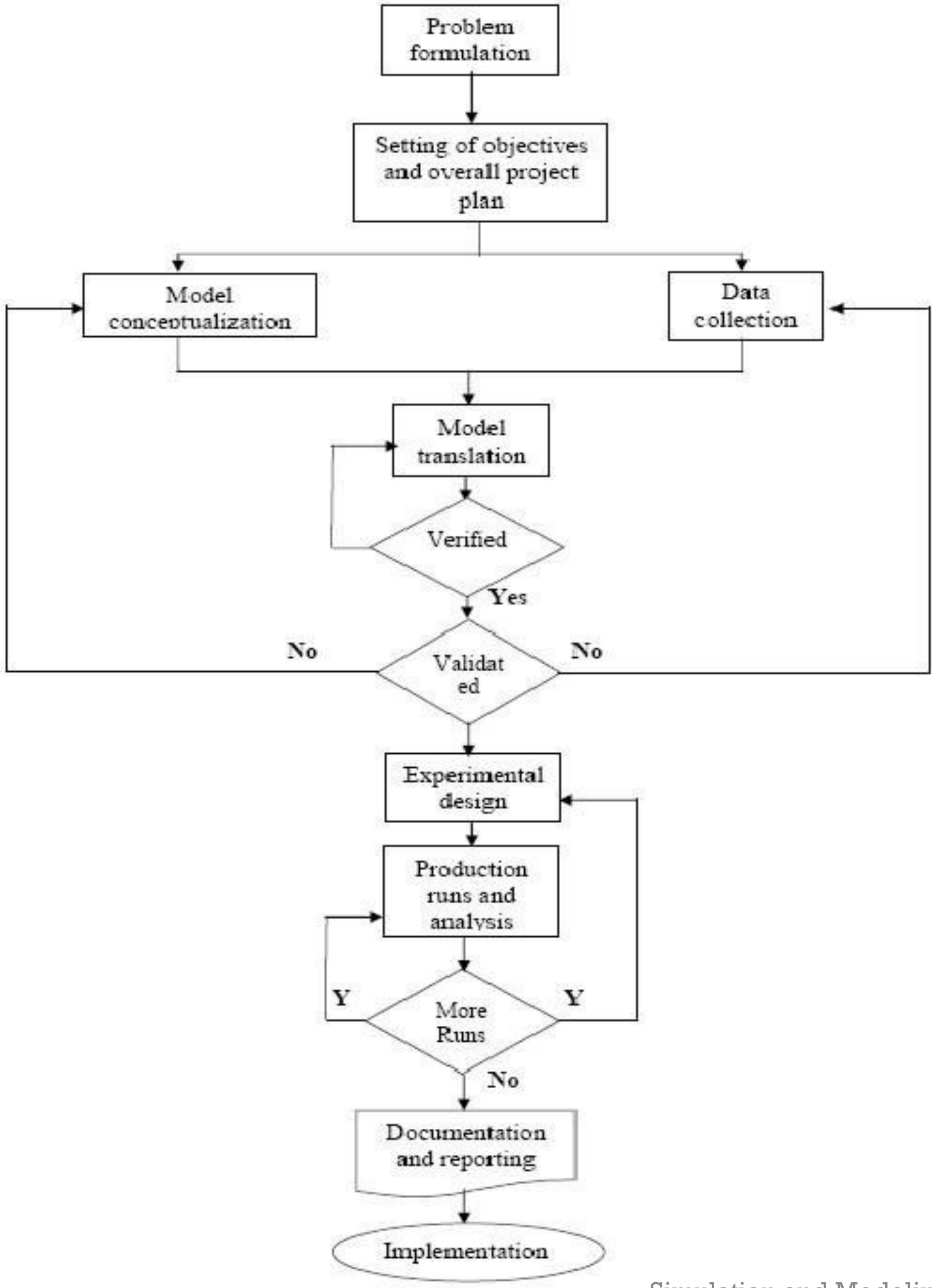
III Phase

- Consists of steps 8,9 and 10
- Conceives a thorough plan for experimenting
- Discrete-event stochastic is a statistical experiment
- The output variables are estimates that contain random error and therefore proper statistical analysis is required.

IV Phase

- Consists of steps 11 and 12
- Successful implementation depends on the involvement of user and every steps successful completion.

Steps in a Simulation study



Queuing system

Queuing system:

- **Queuing system are the waiting lines in which the system attribute are waiting for a service.**
- The queue may be of the customer waiting for the **server** or **server waiting** for customer.
- The waiting line situation arises either there is too much demand on the service facility so that customer have to wait for getting service or there is too less demand in which service facility have to wait for the customer.

Queuing system:

- The line where the entities or customers wait is generally known as queue.
- The combination of all entities in system being served and being waiting for services will be called a queuing system.
- The general diagram of queuing system can be shown as a queuing system involves customers arriving at a constant or variable time rate for service at a service station.
- Customers can be students waiting for registration in college, airplane queuing for landing at airfield, or jobs waiting in machines shop.
- ***They remain in queue till they are provided the service. Sometimes queue being too long, they will leave the queue and go, it results a loss of customer.***

Queuing system:

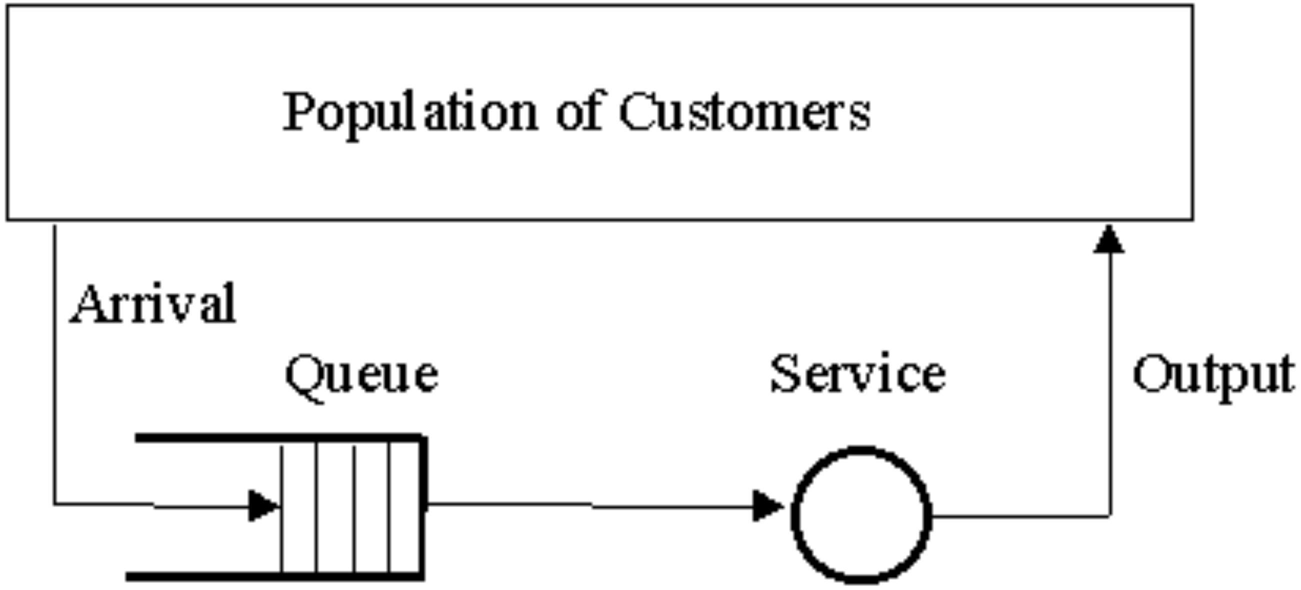


Figure 1

Queuing system:

Queuing system:

- The basic concept of queuing theory is the optimization of **wait time**, **queue length**, and the **service available** to those standing in a queue.
- Cost is one of the important factors in the queuing problem.
- Waiting in queues incur cost, whether human are waiting for services or machines waiting in a machine shop. On the other hand if service counter is waiting for customers that also involves cost.
- **In order to reduce queue length, extra service centers are to be provided but for extra service centers, cost of service becomes higher.**

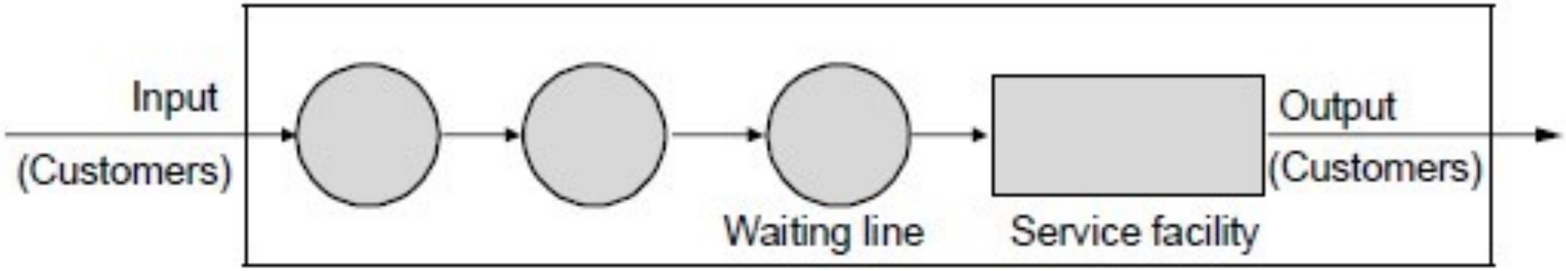
Characteristics of Queuing Systems

- The key elements, of a queuing system are the ***customers*** and ***servers***. The term "***customer***" can refer to people, machines, trucks, mechanics, patients— anything that arrives at a facility and requires service.
- The term "***server***" might refer to receptionists, repairpersons, CPUs in a computer, or washing machines….any resource (person, machine, etc. which provides the requested service.
- Table 1 lists a number of different queuing systems.

Table 1: Examples of Queuing Systems

System	*Customers*	*Server(s)*
Reception desk	People	Receptionist
Repair facility	Machines	Repairperson
Garage	Trucks	Mechanic
Tool crib	Mechanics	Tool-crib clerk
Hospital	Patients	Nurses
Warehouse	Pallets	Crane
Airport	Airplanes	Runway
Production line	Cases	Case packer
Warehouse	Orders	Order picker
Road network	Cars	Traffic light
Grocery	Shoppers	Checkout station
Laundry	Dirty linen	Washing machines/dryers
Job shop	Jobs	Machines/workers
Lumberyard	Trucks	Overhead crane
Saw mill	Logs	Saws
Computer	Jobs	CPU, disk, tapes
Telephone	Calls	Exchange
Ticket office	Football fans	Clerk
Mass transit	Riders	Buses, trains

Elements of queuing system:



Elements of queuing system:

1 . Population of customer:

- **Customer are the entities who wants service from the server. It can be considered either limited (closed system) or unlimited (open system).**
- In systems with a large population of potential customers, the calling population is usually assumed to be finite or infinite. Examples of infinite populations include the potential customers of a restaurant, bank, etc.
- **The main difference between finite and infinite population models is how the arrival rate is defined.**
- In an infinite-population model, the arrival rate is not affected by the number of customers who have left the calling population and joined the queuing system.
- On the other hand, for finite calling population models, the arrival rate to the queuing system does depend on the number of customers being served and waiting.

Elements of queuing system:

2 . Arrival:

- It defines the way that customer enter the system.
- **Mostly arrivals are random with random intervals between two adjacent parameters.**
- Typically the arrival is described by random distribution of intervals also called arrival pattern.
- Arrival process for infinite-population models is usually characterized in terms of inter arrival times of successive customers. Arrivals may occur at scheduled times or at random times. When at random times, the inter arrival times are usually characterized by a probability distribution. The most important model for random arrivals is the Poisson arrival process.

Elements of queuing system:

3 . Queue or waiting line:

- It especially represents a certain number of customers waiting for service. Two important properties of queue are:
  - Maximum size
  - Queuing discipline
  - Maximum size is the maximum number of customers that may be waiting in a queue.
  - Common queue disciplines include first-in, first-out (FIFO); last-in first out (LIFO); service in random order (SIRO); shortest processing time first |(SPT) and service according to priority (PR).

Elements of queuing system:

4 . The service times and service mechanism

- **It represents some activity that takes time and that the customers are waiting for.**
- It may be not only be real service carried on person or machines but it may also be CPU time slice, connection created for telephone calls.

Elements of queuing system:

4 . The service times and service mechanism….

Theoretical models are based on random distribution of service duration also called service patterns.

- System with one server is single channel system and with more servers in multichannel servers.
- The service times of successive arrivals are denoted by S1, S2, S3, . . . They may be constant or of random duration.
- Sometimes services may be identically distributed for all customers of a given type or class or priority, while customers of different types may have completely different service-time distributions. In addition, in some systems, service times depend upon the time of day or the length of the waiting line. For example, servers may work faster than usual when the waiting line is long, thus effectively reducing the service times.
- Each service center consists of some number of servers, c, and 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 ($I < c < \infty$)**, or **unlimited servers ($c = \infty$)**. A self-service facility is usually characterized as having an unlimited number of servers.

Elements of queuing system:

5 . Output:

- Output represents the way customers leave the system.
- Output is mostly ignored by theoretical models but sometimes the customers leaving the server enter the queue again.

Application of queuing system:

- Telecommunication
- Traffic control
- Computer process
- Manufacturing process

Queuing discipline:

- It explains how the customer is solved by the server or the way in which queue is organized. It is the rule by which customer enters and exits the queue. Some queuing discipline are
  - FIFO – First In First Out
  - LIFO – Last In First Out
  - SIRO – Serial In Random Out
  - SPTF – Shortest Processing Time First
  - PR – Service According to Priority

Queuing Notation

- Recognizing the diversity of queuing systems, Kendall [1953] proposed a notational system for parallel server systems which has been widely adopted.
- Kendall classify a queuing notation system as

$$A \,/\, B \,/\, s \,/\, q \,/\, c \,/\, P$$

Where,

A is the Arrival pattern

B is the Service pattern

s is No. of server

q is Queuing discipline

c is System capacity

P is Population size

Queuing Notation..

Arrival and service pattern uses the following notations.

$D - Deterministic\ or\ constant$

$G - General\ Distribution$

$GI - General\ Distribution\ with\ Independent\ Random\ Values$

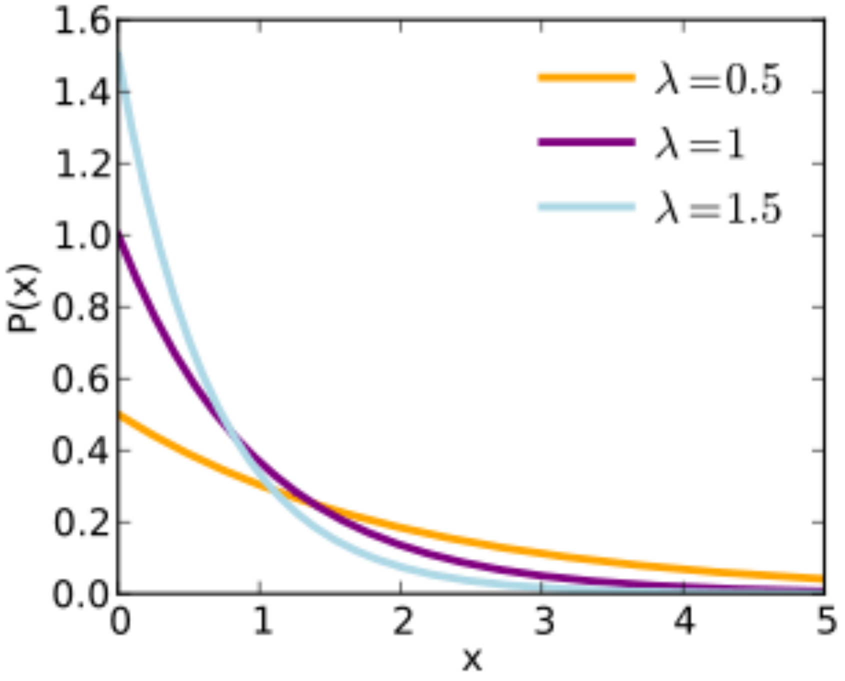
$M - Poisson\ (Markovian)\ process\ or\ Exponential\ Distribution$

$Em - Erlang\ Distribution$

$H - Hyper\ exponential\ distribution$

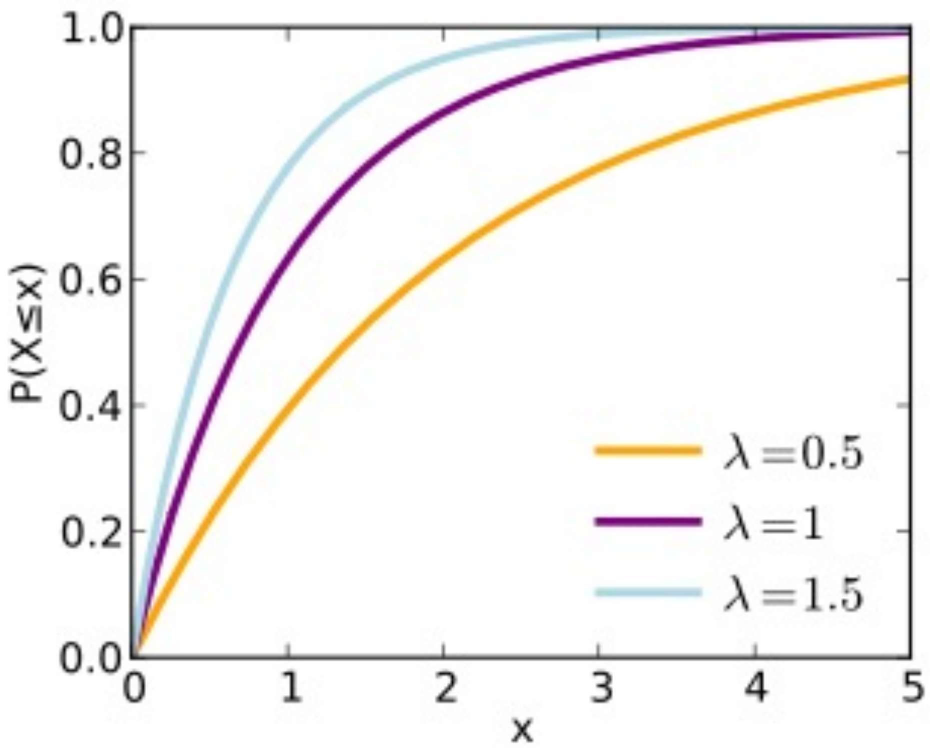
Exponential Distribution

Probability density function



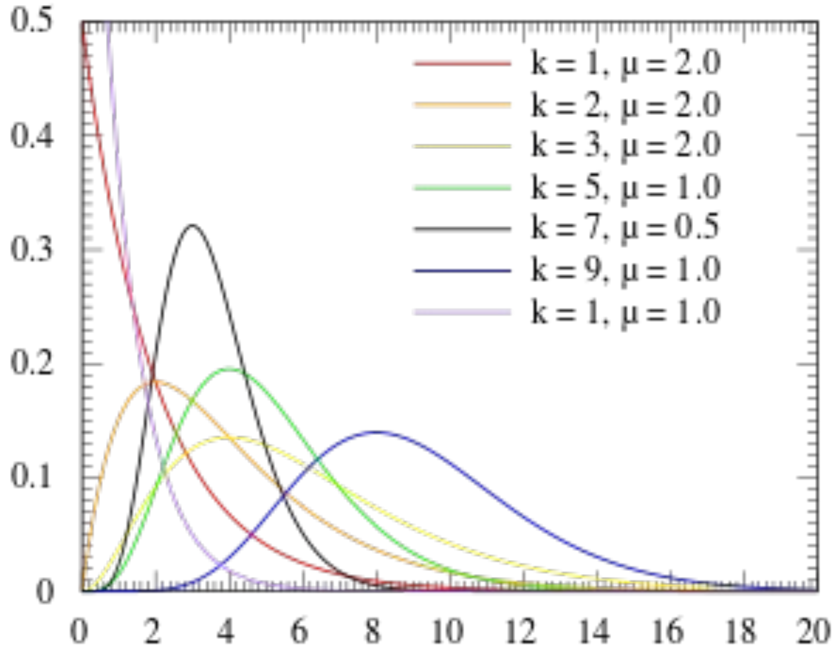
Exponential Distribution

Cumulative distribution function



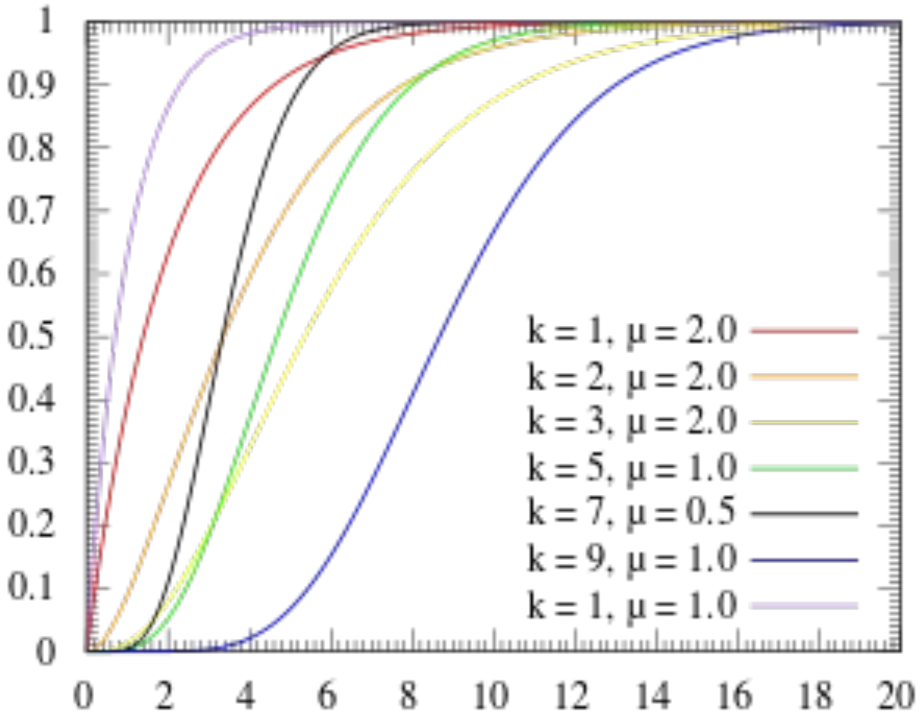
Erlang Distribution

Probability density function



Erlang Distribution

Cumulative distribution function



Queuing Notation..

Example:

1. $D / M / 1$

$Arrival\ pattern - Deterministic$

$Service\ pattern - Exponential\ Distribution$

$No.\ of\ Server - 1$

$Queuing\ discipline - FIFO$

$System\ capacity - Infinite$

$Population\ size - Infinite$

Note : systems will be assumed to have a FIFO queue discipline.

Queuing Notation..

Example:

2. $M\ /\ D\ /\ 2\ /\ LIFO$

$Arrival\ pattern - Exponential\ Distribution$

$Service\ pattern - Deterministic$

$No.of\ Server - 2$

$Queuing\ discipline - LIFO$

$System\ capacity - Infinite$

$Population\ size - Infinite$

Queuing Notation..

Example:

3. $G\ /\ Em\ /\ 1\ /\ 20$

$Arrival\ pattern - General\ Distribution$

$Service\ pattern - Erlang\ Distribution$

$No.of\ Server - 1$

$Queuing\ discipline - FIFO$

$System\ capacity - 20$

$Population\ size - Infinite$

Queuing Notation..

Example:

4. M / M / 1 / œ /œ

indicates a single-server system that has unlimited queue capacity and an infinite population of potential arrivals. The inter arrival times and service times are exponentially distributed. When C and P are infinite, they may be dropped from the notation. For example, M / M / 1 / œ / œ is often shortened to M/M/1.

Queuing Notation..

Example

a) **M/D/2/5/∞** stands for a queuing system having exponential arrival times, deterministic service time, 2 servers, capacity of 5 customers, and infinite population.

b) If notation is given as **M/D/2** means exponential arrival time, deterministic service time, 2 servers, infinite service capacity, and infinite population.

Queuing Notation..

Examples:

i. **D/M/1** =

Deterministic (known) input, one exponential server, one unlimited FIFO or unspecified queue, unlimited customer population.

i. **M/G/3/20** =

Poisson input, three servers with any distribution, maximum number of customers 20, 32 unlimited customer population.

ii. **D/M/1/LIFO/10/50** =

Deterministic arrivals, one exponential server, queue is a stack of the maximum size 9, total number of customers 50.

Queuing Notation..

Examples:

1. **D / *D* / 2 / *LIFO***
2. **D / M / 1 / 2**
3. **Gi / H / 2 / SIRO / ∞ / 20**
4. **D / G / 3 / 20**
5. **H / Em / 2 / FIFO / 15 / 20**
6. **Gi / G / 4**
7. **D / M / 1 / 2 / 30**
8. **Gi / H / 2 / LIFO / 20**

Queuing Notation..

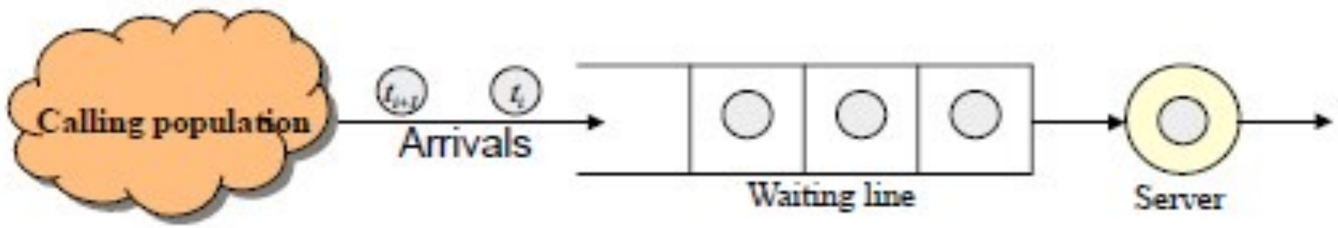
P_n Steady-state probability of having n customers in system
$P_n(t)$ Probability of n customers in system at time t
λ Arrival rate
λ_e Effective arrival rate
μ Service rate of one server
ρ Server utilization
A_n Interarrival time between customers $n - 1$ and n
S_n, Service time of the nth arriving customer
W_n Total time spent in system by the nth arriving customer
W_n^Q Total time spent in the waiting line by customer n .
L(t) The number of customers in system at time /
$L_Q(t)$ 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
ώ Long-run average time spent in system per customer
$ώ_Q$ Long-run average time spent in queue per customer

Table: Queuing Notation for Parallel Server Systems:

Simulation of queuing system

Queuing system state:

- System
  - Server
  - Units (in queue or being served)
  - Clock
- State of the system
  - Number of units in the system
  - Status of server (idle, busy)
- Events
  - Arrival of a unit
  - Departure of a unit



Queuing Model

- **Single Server**
- **Multiple Server**

Within these single server and multiple server there are two types:

- **Finite queue length** : restriction in queue length
- **Infinite queue length** : no restriction in queue length

Queuing Model

Balking:

Balking is a queue behavior wherein people leave as soon as they realize that they will to wait.

So **if an arrival doesn’t join the system and leave is said to be Balking.**

Balking can also be two types

- Forced balking
- Unforced balking

Reneging:

Reneging refer to a queue behavior wherein **people leave a queue after they are tired of waiting .**

Retrial or Jockeying Queue:

As the name suggests, this particular queue behavior refer to **customers’ response to rejoin a queue that they had left earlier due to balking or reneging.**

Queuing Model

Polling:

When there are more than one queue forming (establishing) for same service, the action of sharing service between the queues is called polling.

A bus picking up passengers from different stoppage along its route is an example of polling service.

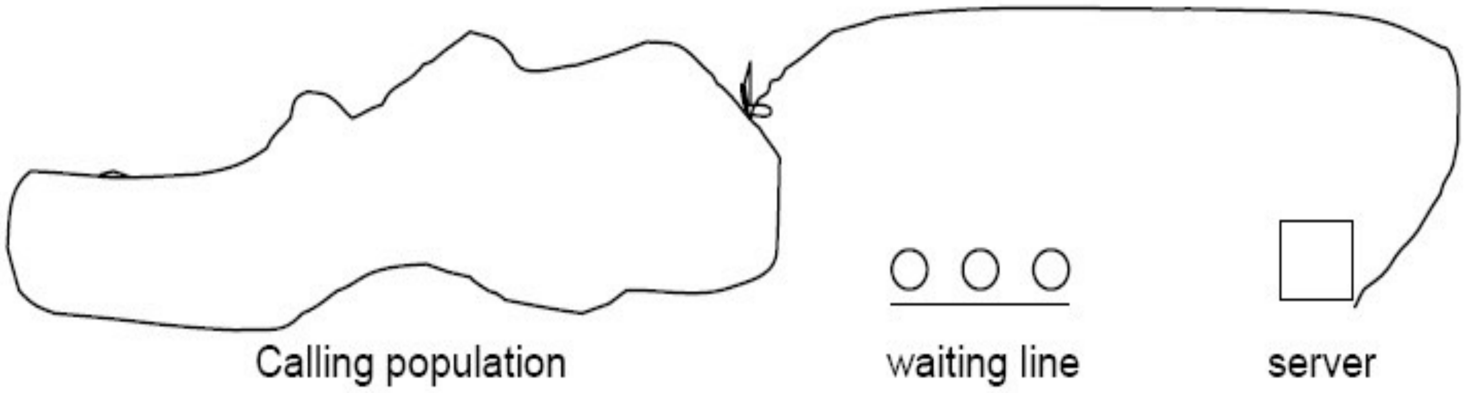
Separate queue for ladies and gents at ticket window, is another example of polling service.

Queuing Model

1) Single server queue:

A queuing system is described by its calling population, the nature of arrivals, the service mechanism, system capacity and the queuing discipline.

A single channel queuing system is portrayed in fig below.



Queuing Model

So in a **single server queue,**

- **Calling population is infinite**
  - Arrival rate does not change
- **Units are served according to FIFO**
- Arrivals are defined by the distribution of time between arrivals
  - Inter-arrival time
- Service time are according to distribution
- **Arrival rate must be less than service rate**
  - Stable system
- Otherwise waiting line will grow unbounded
  - Unstable system

Queuing Model

Arrival event:

- If server idle unit gets service, otherwise unit enters queue.

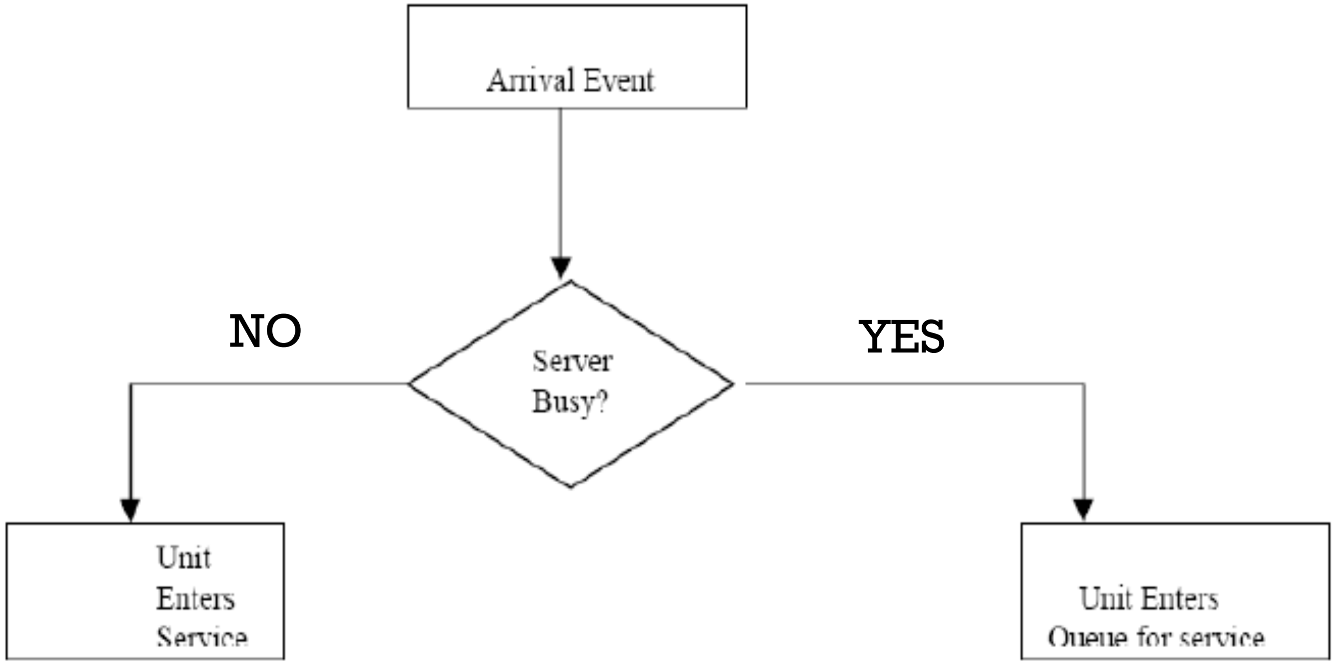


Fig: Unit entering system-flow diagram

Queuing Model

Departure event:

- If queue is not empty begin servicing next unit, otherwise service will be idle.

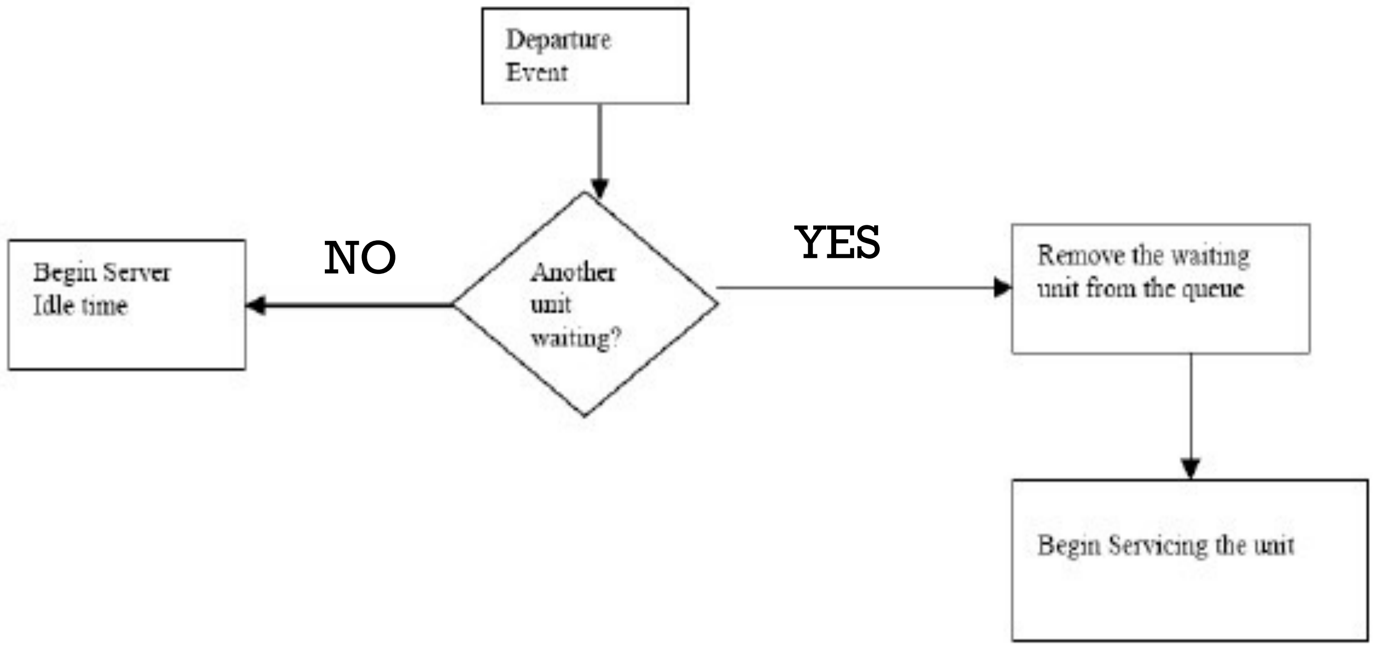
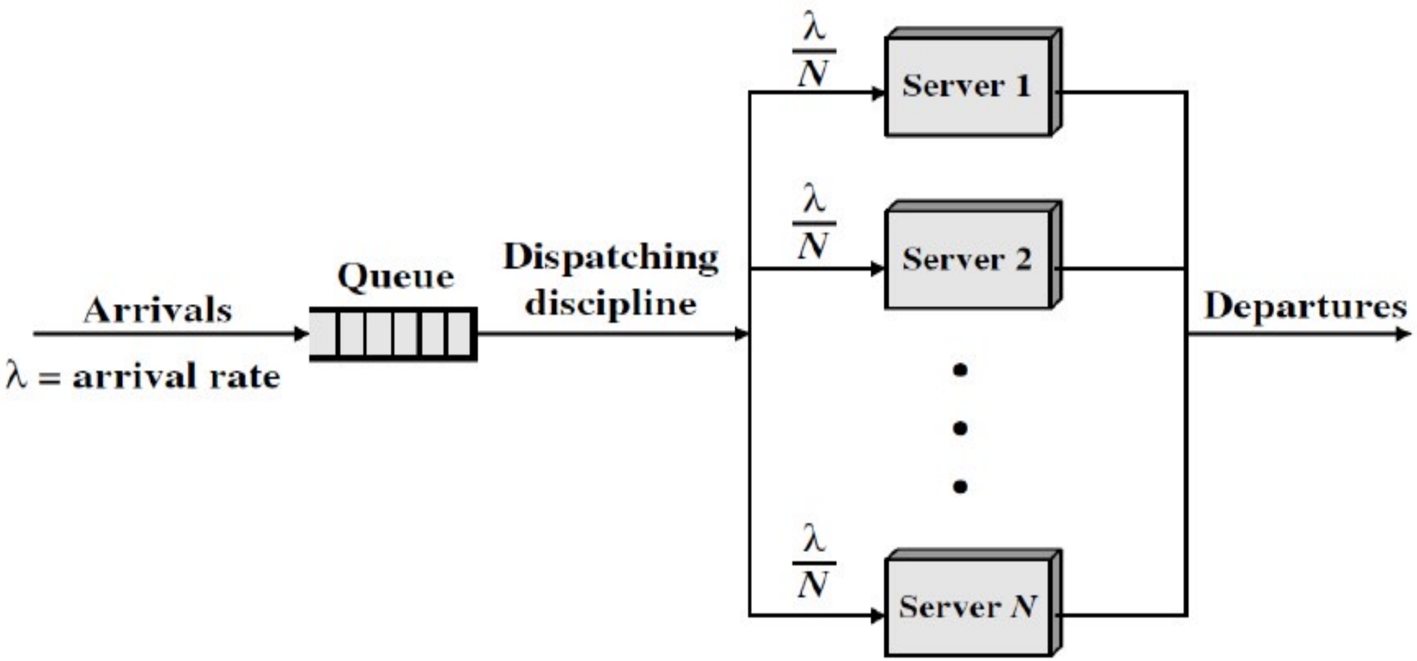


Fig: Service just-completed flow diagram

Queuing Model

2) Multi-server Queue



Queuing Model

Multi-server Queue

Figure shows a generalization of the simple model we have been discussing for multiple servers, all sharing a common queue.

If an item arrives and at least one server is available, then the item is immediately dispatched to that server.

It is assumed that all servers are identical; thus, if more than one server is available, it makes no difference which server is chosen for the item.

If all servers are busy, a queue begins to form. As soon as one server becomes free, an item is dispatched from the queue using the dispatching discipline in force.

Multi-server Queue…

The key characteristics typically chosen for the multi-server queue correspond to those for the single-server queue.

That is, we assume an infinite population and an infinite queue size, with a single infinite queue shared among all servers.

Unless otherwise stated, the dispatching discipline is FIFO.

For the multi-server case, if all servers are assumed identical, the selection of a particular server for a waiting item has no effect on service time.

Queuing Model

Multi-server Queue

The total server utilization in case of Multi-server queue for N server system is

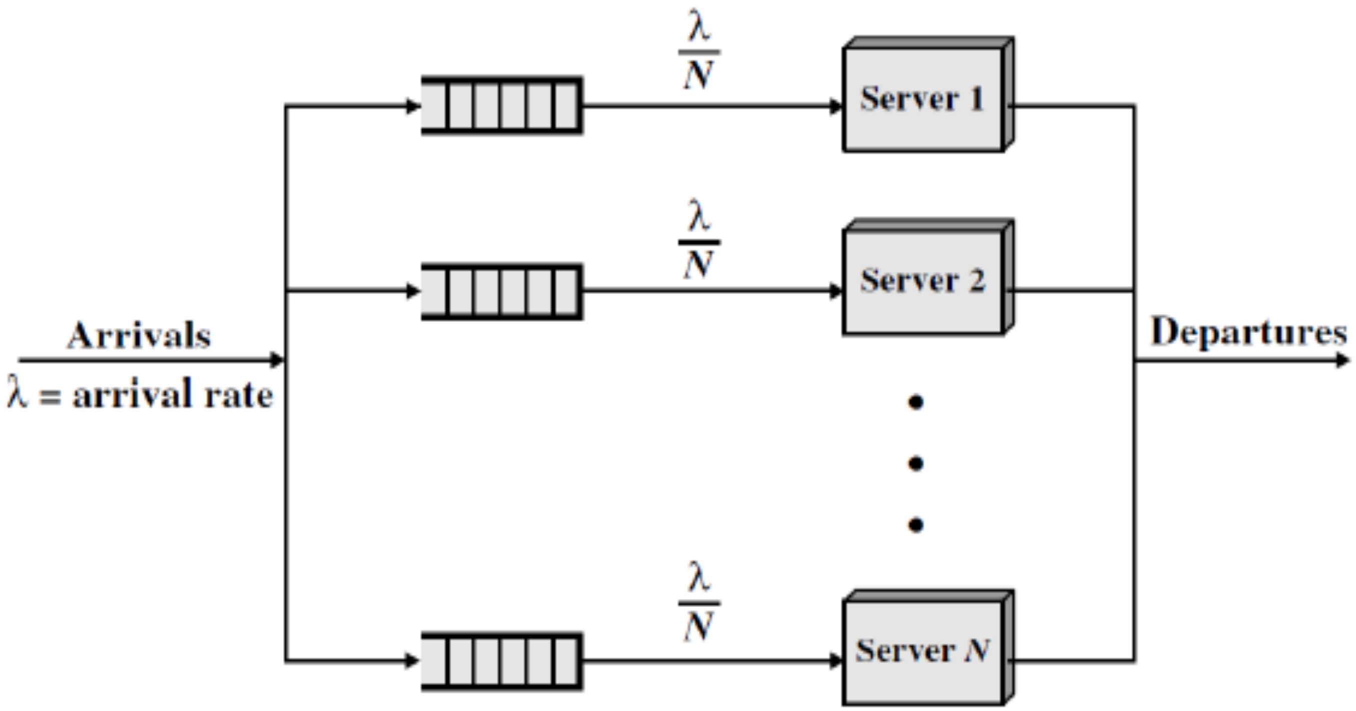
(server utilization) $\rho = \lambda / c\mu$

Where μ is the service rate and λ is the arrival rate.

Queuing Model

Multi-server Queue

There is another concept which is called multiple single server queue system as shown below



Queuing Model

Some notation or Formula used to Measure the different parameter of queue

Two principal measures of queuing system are;

1. The mean **number** of customers waiting and
2. The mean **time** the customer spend waiting

Both these quantities may refer to the total number of entities in the system, those waiting and those being served or they may refer only to customer in the waiting line.

Queuing Model

Average number of customers in the System $\overline{L_S} = \frac{\rho}{1-\rho} = \frac{\frac{\lambda}{\mu}}{1-\frac{\lambda}{\mu}} = \frac{\lambda}{\mu-\lambda}$

Average number of customers in the Queue $\overline{L_Q}$

= Average number of customers in the System − Server Utilization

$$= \overline{L_S} - \frac{\lambda}{\mu} = \frac{\lambda}{\mu-\lambda} - \frac{\lambda}{\mu} = \frac{\lambda^2}{\mu(\mu-\lambda)}$$

Queuing Model

Average waiting time in the System $\overline{W}_S = \frac{Average\ \ number\ \ of\ customer\ \ in\ the\ system}{Mean\ \ arrival\ \ rate}$

$$= \frac{\overline{L}_S}{\lambda} = \frac{\frac{\lambda}{\mu-\lambda}}{\lambda} = \frac{1}{\mu-\lambda}$$

Average waiting time in the Queue $\overline{W}_Q = \frac{Average\ \ number\ \ of\ customer\ \ in\ the\ Queue}{Mean\ \ arrival\ \ rate}$

$$= \frac{\overline{L}_Q}{\lambda} = \frac{\frac{\lambda^2}{\mu(\mu-\lambda)}}{\lambda} = \frac{\lambda}{\mu(\mu-\lambda)}$$

Queuing Model

Example 1

Q.N > At the ticket counter of football stadium, people come in queue and purchase tickets. Arrival rate of customers is 1/min. It takes at the average 20 seconds to purchase the ticket.

(a) If a sport fan arrives 2 minutes before the game starts and if he takes exactly 1.5 minutes to reach the correct seat after he purchases a ticket, can the sport fan expects to be seated for the kick-off?

Solution:

(a) A minute is used as unit of time. Since ticket is disbursed in 20 seconds, this means, three customers enter the stadium per minute, that is service rate is 3 per minute.

Therefore,

λ = 1 arrival/min

μ = 3 arrivals/min

WS = waiting time in the system = 1/(μ- λ)=0.5 minutes

The average time to get the ticket plus the time to reach the correct seat is 2 minutes exactly, so the sports fan can expect to be seated for the kick-off.

Queuing Model

Example 2

Q.N > At the Bank counter, people come in queue for service. Arrival rate of customers is 2/min. It takes at the average 15 seconds to take service.

(a) If bank will close 2 minutes and if he takes exactly 1 minutes to reach the door, can the customer leave bank in time?

Solution:

(a) A minute is used as unit of time. Therefore,

λ = 2 arrival/min

μ = 4 arrivals/min

WS = waiting time in the system=1/(μ- λ)= 1/(4-2) = 0.5 minutes

- Now total time will be 1+0.5 = 1.5 minutes.

Queuing Model

Example 3

Q.N > At the orchid college, student come in queue for service. Arrival rate of students is 1/min. It takes at the average 22 seconds to take service.

(a) If college will close 3 minutes and if he takes exactly 1.5 minutes to reach the door, can the student leave college in time?

Solution:

(a) A minute is used as unit of time. Therefore,

λ = 1 arrival/min

μ = 2 arrivals/min

WS = waiting time in the system=1/(μ- λ)= 1/(2-1) = 1 minutes

- Now total time will be 1+1.5 = 2.5 minutes.

Queuing Model

Example 4

Q.N > Customers arrive in a bank according to a Poisson's process with mean inter arrival time of 10 minutes. Customers spend an average of 5 minutes on the single available counter, and leave.

(a) What is the probability that a customer will not have to wait at the counter?

(b) What is the expected number of customers in the bank?

(c) How much time can a customer expect to spend in the bank?

***Solution*:**

We will take an hour as the unit of time. Thus, $\lambda = 6$ customers/hour,

$\mu = 12$ customers/hour.

The customer will not have to wait if there are no customers in the bank.

Thus, $P0 = 1 - \lambda/\mu = 1 - 6/12 = 0.5$

Expected numbers of customers in the bank are given by

$LS = \lambda / (\mu - \lambda) = 6/6 = 1$

Expected time to be spent in the bank is given by

$WS = 1/(\mu - \lambda) = 1/(12\text{-}6) = 1/6$ hour = 10 minutes.

Queuing Model

Example 5

Q.N. > At the Banking, people come in queue . Arrival rate of customers is 1/min. It takes at the average 30 seconds to take the token. If a customer arrives 5 minutes before the bank closed and if he takes exactly 4.5 minutes to reach the correct counter after he take a token, can the customer expects to take banking service?

Solution :

A minute is used as unit of time. Since token takes 30 seconds, this means, two customers enter the bank per minute, that is service rate is 2 per minute.

Therefore,

λ = 1 arrival/min

μ = 2 arrivals/min

WS = waiting time in the system=1/(μ- λ)=1 minutes

and if he takes exactly 4.5 minutes to reach the correct counter after he take a token, so bank is closed because 4.5+1=5.5

Queuing Model

A Customers	B Time since last Arrival (Min)	C Arrival Time	D Service Time	E Time Service Begins	F Time customer waits in queue	G Time Service Ends	H Time customer spends in system	I Idle Time of Server
1	–	0	4	0	0	4	4	0
2	8	8	1	8	0	9	1	4
3	6	14	4	14	0	18	4	5
4	1	15	3	18	3	21	6	0
5	8	23	2	23	0	25	2	2
6	3	26	4	26	0	30	4	1
7	8	34	5	34	0	39	5	4
8	7	41	4	41	0	45	4	2
9	2	43	5	45	2	50	7	0
10	3	46	3	50	4	53	7	0
11	1	47	3	53	6	56	9	0
12	1	48	5	56	8	61	13	0
13	5	53	4	61	8	65	12	0
14	6	59	1	65	6	66	7	0
15	3	62	5	66	4	71	9	0
16	8	70	4	71	1	75	5	0
17	1	71	3	75	4	78	7	0
18	2	73	3	78	5	81	8	0
19	4	77	2	81	4	83	6	0
20	5	82	3	83	1	86	4	0
			68		56		124	18

Queuing Model

1. $$\text{Average waiting time} = \frac{\text{total time customers wait in queue}}{\text{total number of customers}} = \frac{56}{20} = 2.8 \text{ min}$$

2. $$\text{Probability of wait} = \frac{\text{Number of customers who wait}}{\text{total number of customers}} = \frac{13}{20} = 0.65$$

3. $$\text{Probability of idle server} = \frac{\text{Total idle time of server(minutes)}}{\text{Total run time of simulation(minutes)}} = \frac{18}{86} = 0.21$$

4. The average service time is 3.4 minutes, determined as follows:

Average service time (minutes) =

$$= \frac{\text{Total Service time(minutes)}}{\text{Total run timTotal Number of Customers)}} = \frac{68}{20} = 3.4 \text{ minutes}$$

Measures of System Performance

The performance of a queuing system can be evaluated in terms of a number of response parameters, however the following four are generally employed.

- ✓ ***Average number of customer in the queue or in the system***
- ✓ ***Average waiting time of the customer in the queue or in the system***
- ✓ ***System utilization (Server utilization)***
- ✓ ***The cost of waiting time and idle time***

Measures of System Performance

- Each of these measures has its own importance.
- The knowledge of average number of customers in the queue or in the system helps to determine the space requirements of the waiting entities.
- Also too long a waiting line may discourage the prospectus customers, while **no queue may suggest that service offered is not good quality to attract customers.**

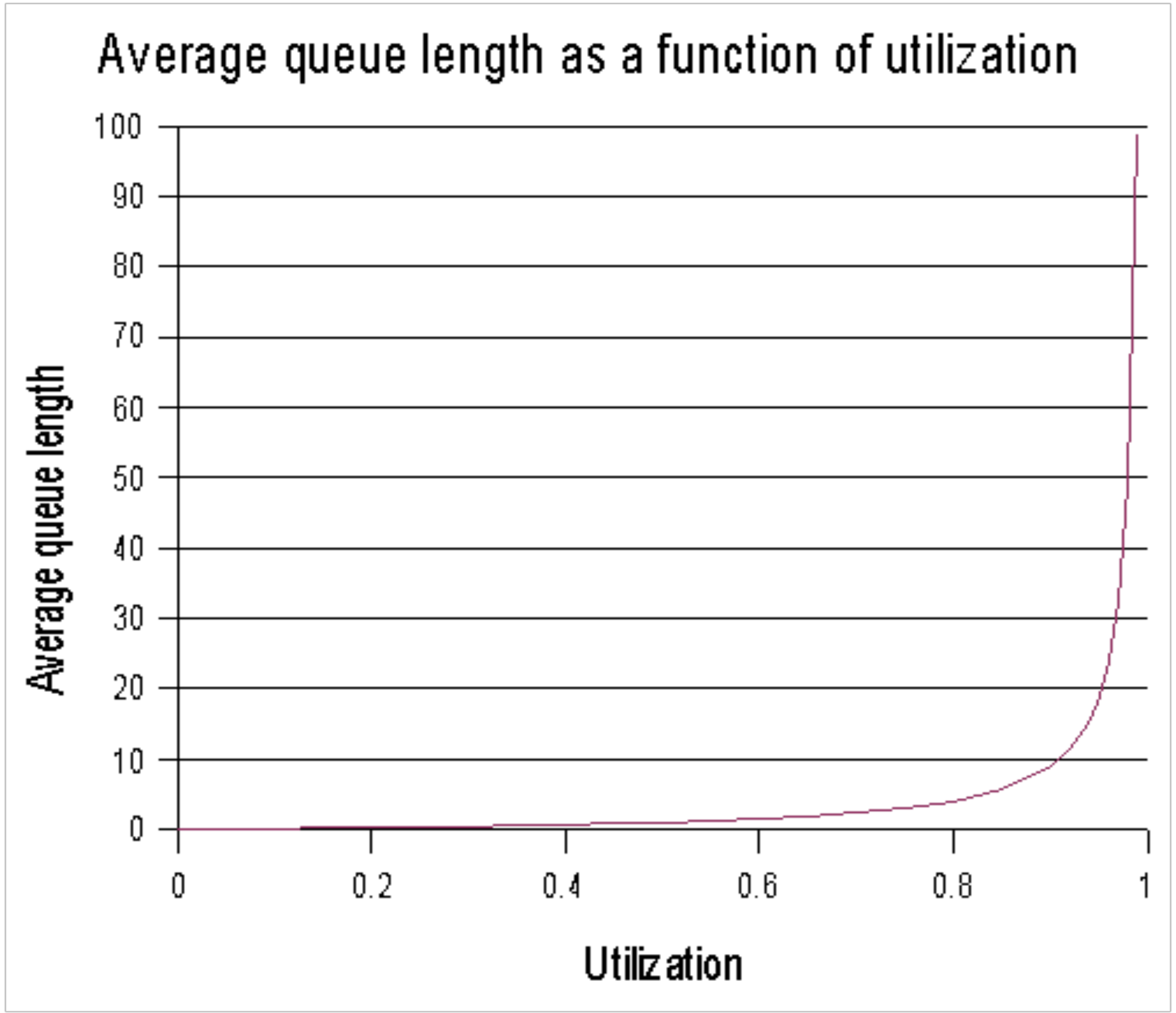
Measures of System Performance

- Let Ta be mean arrival time, Ts be service time, λ be arrival rate and μ be the service rate then, the ratio of mean service rate and mean inter-arrival rate is called the **traffic intensity** (u).

$$u = \frac{T_s}{T_a}$$

- The probability that an entity have to wait more than a given time is known as **delay distribution**.
- The knowledge of average waiting time in the queue is necessary for determining the cost of waiting in the queue.

Measures of System Performance



Measures of System Performance

System utilization, that is, the percentage capacity utilized reflects the extent to which the facility is busy rather than idle. System utilization factor (S) is the ratio of average arrival rate (λ) to the average service rate (μ).

$S = \lambda / \mu$, in case of single server model

$S = \lambda / n\mu$, in case of 'n' server model

The system utilization can be increased by increasing the arrival rate which amounts to increasing the average queue length as well as the average waiting time, as shown is above figure. Under normal circumstances 100% system utilization is not a realistic goal.

Measures of System Performance

Conservation Law

An important law in queuing theory states

$$L = \lambda w$$

Where, L is the long-run in the system, λ is the arrival rate and w is the long-run time in the system.

Often called as **Little's equation**.

Unit 2 Finished