UNIT-3 STATISTICAL MODELS IN SIMULATION

Purpose & Overview

- The world the model-builder sees is probabilistic rather than deterministic.
  - Some statistical model might well describe the variations.
- An appropriate model can be developed by sampling the phenomenon of interest:
  - Select a known distribution through educated guesses
  - Make estimate of the parameter(s)
  - Test for goodness of fit
- In this chapter:
  - Review several important probability distributions
  - Present some typical application of these models

3.1 Review of Terminology and Concepts

- In this section, we will review the following concepts:
  - Discrete random variables
  - Continuous random variables
  - Cumulative distribution function
  - Expectation

Discrete Random Variables [Probability Review]

- X is a discrete random variable if the number of possible values of X is finite, or countably infinite.
- Example: Consider jobs arriving at a job shop.
  - Let X be the number of jobs arriving each week at a job shop.
  - R_x = possible values of X (range space of X) = $\{0,1,2,\ldots\}$
  - $p(x_i)$ = probability the random variable is $x_i = P(X = x_i)$
- $p(x_i)$, $i = 1,2, \ldots$ must satisfy:

  1. $p(x_i) \geq 0, \text{ for all } i$
  2. $\sum_{i=1}^{\infty} p(x_i) = 1$

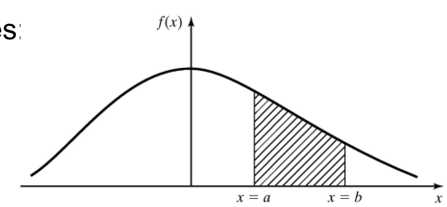
- The collection of pairs $[x_i, p(x_i)]$, $i = 1,2,\ldots$, is called the probability distribution of X, and $p(x_i)$ is called the probability mass function (pmf) of X.

Continuous Random Variables [Probability Review]

- X is a continuous random variable if its range space R_x is an interval or a collection of intervals.
- The probability that X lies in the interval $[a,b]$ is given by:

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

- $f(x)$, denoted as the pdf of X, satisfies:
  1. $f(x) \geq 0$, for all x in R_X
  2. $\int_{R_X} f(x)dx = 1$
  3. $f(x) = 0$, if x is not in R_X

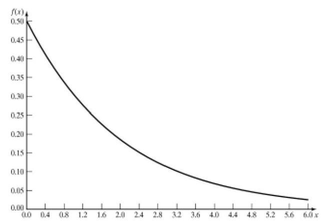


- Properties
  1. $P(X = x_0) = 0$, because $\int_{x_0}^{x_0} f(x)dx = 0$
  2. $P(a \leq X \leq b) = P(a \prec X \leq b) = P(a \leq X \prec b) = P(a \prec X \prec b)$

Continuous Random Variables [Probability Review]

- Example: Life of an inspection device is given by *X*, a continuous random variable with pdf:

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



  - *X* has an exponential distribution with mean 2 years
  - Probability that the device's life is between 2 and 3 years is:

$$P(2 \leq x \leq 3) = \frac{1}{2} \int_2^3 e^{-x/2} dx = 0.14$$

Cumulative Distribution Function [Probability Review]

- Cumulative Distribution Function (cdf) is denoted by *F(x)*, where *F(x) = P(X <= x)*
  - If *X* is discrete, then $F(x) = \sum_{\substack{\text{all} \\ x_i \leq x}} p(x_i)$
  - If *X* is continuous, then $F(x) = \int_{-\infty}^{x} f(t)\,dt$
- Properties

1. F is nondecreasing function. If $a \prec b$, then $F(a) \leq F(b)$
2. $\lim_{x \to \infty} F(x) = 1$
3. $\lim_{x \to -\infty} F(x) = 0$

- All probability question about *X* can be answered in terms of the cdf, e.g.:

$$P(a \prec X \leq b) = F(b) - F(a), \text{ for all } a \prec b$$

Cumulative Distribution Function [Probability Review]

- Example: An inspection device has cdf:

$$F(x) = \frac{1}{2}\int_0^x e^{-t/2}dt = 1 - e^{-x/2}$$

  - The probability that the device lasts for less than 2 years:

$$P(0 \le X \le 2) = F(2) - F(0) = F(2) = 1 - e^{-1} = 0.632$$

  - The probability that it lasts between 2 and 3 years:

$$P(2 \le X \le 3) = F(3) - F(2) = (1 - e^{-(3/2)}) - (1 - e^{-1}) = 0.145$$

Expectation [Probability Review]

- The expected value of *X* is denoted by *E(X)*
  - If *X* is discrete $E(x) = \sum_{\text{all } i} x_i p(x_i)$
  - If *X* is continuous $E(x) = \int_{-\infty}^{\infty} x f(x) dx$
  - a.k.a the mean, m, or the 1st moment of *X*
  - A measure of the central tendency
- The variance of *X* is denoted by *V(X)* or *var(X)* or σ^2
  - Definition: $V(X) = E[(X - E[X]^2]$
  - Also, $V(X) = E(X^2) - [E(x)]^2$
  - A measure of the spread or variation of the possible values of X around the mean
- The standard deviation of *X* is denoted by σ
  - Definition: square root of *V(X)*
  - Expressed in the same units as the mean

Expectations [Probability Review]

- Example: The mean of life of the previous inspection device is:

$$E(X) = \frac{1}{2}\int_0^\infty xe^{-x/2}dx = -xe^{-x/2}\Big|_0^\infty + \int_0^\infty e^{-x/2}dx = 2$$

- To compute variance of *X*, we first compute *E(X²)*:

$$E(X^2) = \frac{1}{2}\int_0^\infty x^2 e^{-x/2}dx = -x^2 e^{-x/2}\Big|_0^\infty + \int_0^\infty e^{-x/2}dx = 8$$

- Hence, the variance and standard deviation of the device's life are:

$$V(X) = 8 - 2^2 = 4$$
$$\sigma = \sqrt{V(X)} = 2$$

3.2 Useful Statistical Models

- In this section, statistical models appropriate to some application areas are presented. The areas include:
  - Queueing systems
  - Inventory and supply-chain systems
  - Reliability and maintainability
  - Limited data

Queueing Systems [Useful Models]

- In a queueing system, interarrival and service-time patterns can be probablistic (for more queueing examples, see Chapter 2).
- Sample statistical models for interarrival or service time distribution:
  - Exponential distribution: if service times are completely random
  - Normal distribution: fairly constant but with some random variability (either positive or negative)
  - Truncated normal distribution: similar to normal distribution but with restricted value.
  - Gamma and Weibull distribution: more general than exponential (involving location of the modes of pdf's and the shapes of tails.)

Inventory and supply chain [Useful Models]

- In realistic inventory and supply-chain systems, there are at least three random variables:
  - The number of units demanded per order or per time period
  - The time between demands
  - The lead time
- Sample statistical models for lead time distribution:
  - Gamma
- Sample statistical models for demand distribution:
  - Poisson: simple and extensively tabulated.
  - Negative binomial distribution: longer tail than Poisson (more large demands).
  - Geometric: special case of negative binomial given at least one demand has occurred.

Reliability and maintainability [Useful Models]

- Time to failure (TTF)
  - Exponential: failures are random
  - Gamma: for standby redundancy where each component has an exponential TTF
  - Weibull: failure is due to the most serious of a large number of defects in a system of components
  - Normal: failures are due to wear
  - For cases with limited data, some useful distributions are:
    - Uniform, triangular and beta
  - Other distribution: Bernoulli, binomial and hyper exponential.

3.3 Discrete Distributions

- Discrete random variables are used to describe random phenomena in which only integer values can occur.
- In this section, we will learn about:
  - Bernoulli trials and Bernoulli distribution
  - Binomial distribution
  - Geometric and negative binomial distribution
  - Poisson distribution

Bernoulli Trials and Bernoulli Distribution [Discrete Dist'n]

- Bernoulli Trials:
  - Consider an experiment consisting of n trials, each can be a success or a failure.
    - Let $X_j = 1$ if the jth experiment is a success
    - and $X_j = 0$ if the jth experiment is a failure
  - The Bernoulli distribution (one trial):

$$p_j(x_j) = p(x_j) = \begin{cases} p, & x_j = 1, j = 1,2,\ldots,n \\ 1 - p = q, & x_j = 0, j = 1,2,\ldots,n \\ 0, & \text{otherwise} \end{cases}$$

  - where $E(X_j) = p$ and $V(X_j) = p(1-p) = pq$
- Bernoulli process:
  - The n Bernoulli trials where trails are independent:

$$p(x_1, x_2, \ldots, x_n) = p_1(x_1)\, p_2(x_2) \ldots p_n(x_n)$$

Binomial Distribution [Discrete Dist'n]

- The number of successes in n Bernoulli trials, X, has a binomial distribution.

$$p(x) = \begin{cases} \binom{n}{x} \underbrace{p^x q^{n-x}}, & x = 0,1,2,\ldots,n \\ 0, & \text{otherwise} \end{cases}$$

The number of outcomes having the required number of successes and failures

Probability that there are x successes and (n-x) failures

  - The mean, $E(x) = p + p + \ldots + p = n*p$
  - The variance, $V(X) = pq + pq + \ldots + pq = n*pq$

Geometric & Negative Binomial Distribution [Discrete Dist'n]

- Geometric distribution
  - The number of Bernoulli trials, X, to achieve the 1st success:

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

  - *E(x) = 1/p, and V(X) = q/p²*

- Negative binomial distribution
  - The number of Bernoulli trials, X, until the k^{th} success
  - If Y is a negative binomial distribution with parameters p and k, then:

$$p(x) = \begin{cases} \begin{pmatrix} y-1 \\ k-1 \end{pmatrix} q^{y-k} p^k, & y = k, k+1, k+2, \ldots \\ 0, & \text{otherwise} \end{cases}$$

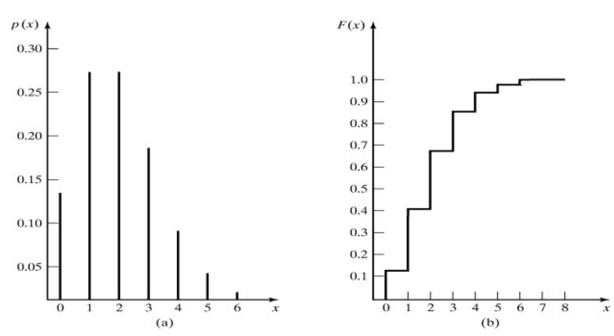
  - *E(Y) = k/p, and V(X) = kq/p²*

Poisson Distribution [Discrete Dist'n]

- Poisson distribution describes many random processes quite well and is mathematically quite simple.
  - where $\alpha > 0$, pdf and cdf are:

$$p(x) = \begin{cases} \dfrac{e^{-\alpha}\alpha^x}{x!}, & x = 0,1,\ldots \\ 0, & \text{otherwise} \end{cases} \qquad F(x) = \sum_{i=0}^{x} \frac{e^{-\alpha}\alpha^i}{i!}$$

  - *E(X) = α = V(X)*



Poisson Distribution [Discrete Dist'n]

- Example: A computer repair person is "beeped" each time there is a call for service. The number of beeps per hour ~ Poisson(α = 2 per hour).
  - The probability of three beeps in the next hour:

    $p(3) = e^{-2}2^3/3! = 0.18$

    also, $p(3) = F(3) - F(2) = 0.857-0.677=0.18$
  - The probability of two or more beeps in a 1-hour period:

    $p(2 \text{ or more}) = 1 - p(0) - p(1)$

    $= 1 - F(1)$

    $= 0.594$

3.4 Continuous Distributions

- Continuous random variables can be used to describe random phenomena in which the variable can take on any value in some interval.
- In this section, the distributions studied are:
  - Uniform
  - Exponential
  - Normal
  - Weibull
  - Lognormal

Uniform Distribution [Continuous Dist'n]

- A random variable *X* is uniformly distributed on the interval (*a,b*), *U*(*a,b*), if its pdf and cdf are:

$$f(x)=\begin{cases}\frac{1}{b-a}, & a \le x \le b\\ 0, & \text{otherwise}\end{cases} \qquad F(x)=\begin{cases}0, & x \prec a\\ \frac{x-a}{b-a}, & a \le x \prec b\\ 1, & x \ge b\end{cases}$$

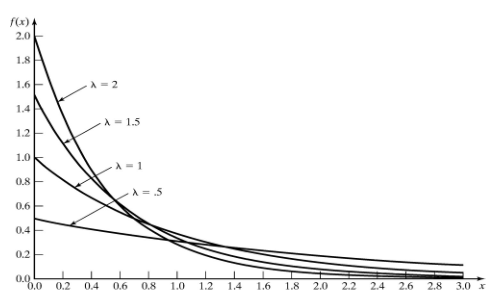
- Properties
  - $P(x_1 < X < x_2)$ is proportional to the length of the interval $[F(x_2) - F(x_1) = (x_2 - x_1)/(b-a)]$
  - $E(X) = (a+b)/2$ $V(X) = (b-a)^2/12$
- U(0,1) provides the means to generate random numbers, from which random variates can be generated.

Exponential Distribution [Continuous Dist'n]

- A random variable *X* is exponentially distributed with parameter λ > 0 if its pdf and cdf are:

$$f(x)=\begin{cases}\lambda e^{-\lambda x}, & x \ge 0\\ 0, & \text{elsewhere}\end{cases} \qquad F(x)=\begin{cases}0, & x \prec 0\\ \int_0^x \lambda e^{-\lambda t}dt = 1-e^{-\lambda x}, & x \ge 0\end{cases}$$

  - $E(X) = 1/\lambda$ $V(X) = 1/\lambda^2$
  - Used to model interarrival times when arrivals are completely random, and to model service times that are highly variable
  - For several different exponential pdf's (see figure), the value of intercept on the vertical axis is λ, and all pdf's eventually intersect.



- Memoryless property
  - For all s and t greater or equal to 0:

    P(X > s+t | X > s) = P(X > t)

  - Example: A lamp ~ exp(l = 1/3 per hour), hence, on average, 1 failure per 3 hours.
    - ✓ The probability that the lamp lasts longer than its mean life is: *P(X > 3) = 1-(1-e-3/3) = e-1 = 0.368*

- ✓ The probability that the lamp lasts between 2 to 3 hours is:

P(2 <= X <= 3) = F(3) – F(2) = 0.145

- ✓ The probability that it lasts for another hour given it is operating for 2.5 hours:

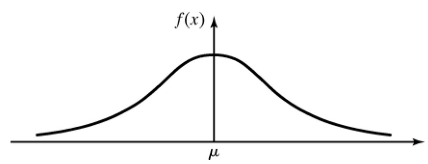
P(X > 3.5 | X > 2.5) = P(X > 1) = e-1/3 = 0.717

Normal Distribution [Continuous Dist'n]

- A random variable *X* is normally distributed has the pdf:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right], \quad -\infty \prec x \prec \infty$$

- Mean: $-\infty \prec \mu \prec \infty$
- Variance: $\sigma^2 \succ 0$
- Denoted as $X \sim N(\mu,\sigma^2)$



- Special properties:
  - $\lim_{x\to-\infty} f(x) = 0$, and $\lim_{x\to\infty} f(x) = 0$.
  - $f(\mu-x)=f(\mu+x)$; the pdf is symmetric about μ.
  - The maximum value of the pdf occurs at $x = \mu$; the mean and mode are equal.

- **Evaluating the distribution:**
  - Use numerical methods (no closed form)
  - Independent of m and s, using the standard normal distribution:

$Z \sim N(0,1)$

  - Transformation of variables: let Z = (X - *m) / s*,

$$F(x) = P(X \le x) = P\left(Z \le \frac{x-\mu}{\sigma}\right)$$

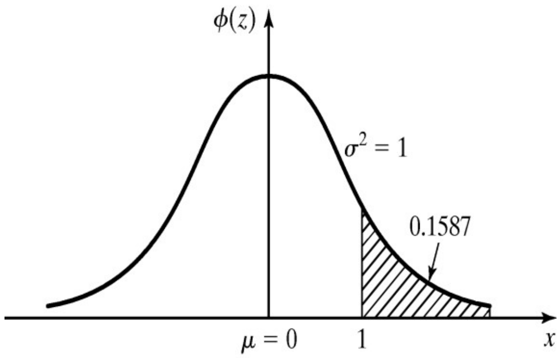
$$= \int_{-\infty}^{(x-\mu)/\sigma} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz \qquad \text{, where } \Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

$$= \int_{-\infty}^{(x-\mu)/\sigma} \phi(z) dz = \Phi\left(\tfrac{x-\mu}{\sigma}\right)$$

Example: The time required to load an oceangoing vessel, *X*, is distributed as *N(12,4)*

- The probability that the vessel is loaded in less than 10 hours:

- Using the symmetry property, F(1) is the complement of F (-1)



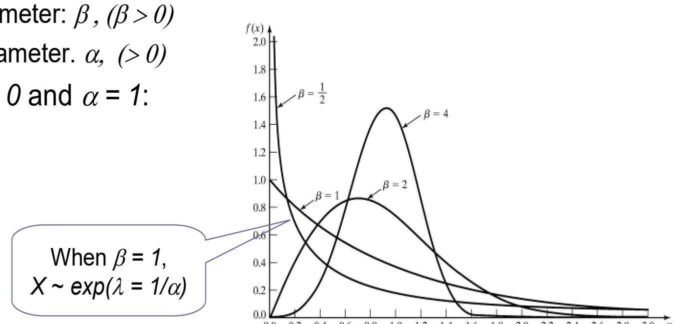
(a)

Weibull Distribution [Continuous Dist'n]

- A random variable X has a Weibull distribution if its pdf has the form:

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

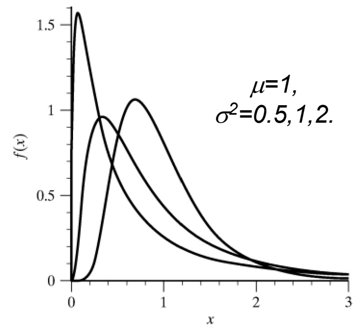
- 3 parameters:
  - Location parameter: υ, $(-\infty \prec \nu \prec \infty)$
  - Scale parameter: β , $(\beta > 0)$
  - Shape parameter. α, (> 0)
- Example: $\upsilon = 0$ and $\alpha = 1$:



Lognormal Distribution [Continuous Dist'n]

- A random variable *X* has a lognormal distribution if its pdf has the form:

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



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

- Relationship with normal distribution
  - When $Y \sim N(\mu, \sigma^2)$, then $X = e^Y \sim \text{lognormal}(\mu, \sigma^2)$
  - Parameters μ and σ^2 are not the mean and variance of the lognormal

Poisson Distribution

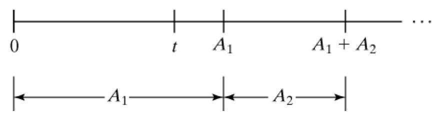
- Definition: *N(t)* is a counting function that represents the number of events occurred in *[0,t]*.
- A counting process {*N(t), t>=0*} is a Poisson process with mean rate λ if:
  - Arrivals occur one at a time
  - {*N(t), t>=0*} has stationary increments
  - {*N(t), t>=0*} has independent increments
- Properties

$$P[N(t) = n] = \frac{e^{-\lambda t}(\lambda t)^n}{n!}, \qquad \text{for } t \geq 0 \text{ and } n = 0,1,2,...$$

  - Equal mean and variance: $E[N(t)] = V[N(t)] = \lambda t$
  - Stationary increment: The number of arrivals in time *s* to *t* is also Poisson-distributed with mean $\lambda(t-s)$

Interarrival Times [Poisson Dist'n]

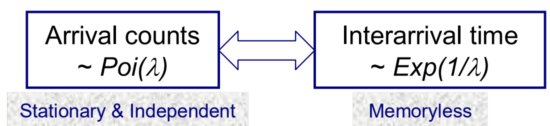
- Consider the interarrival times of a Possion process (A_1, A_2, …), where A_i is the elapsed time between arrival i and arrival $i+1$



  - The 1st arrival occurs after time t iff there are no arrivals in the interval [0,t], hence:

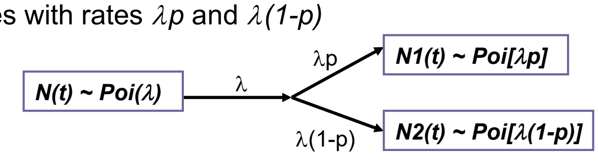
    $P\{A_1 > t\} = P\{N(t) = 0\} = e^{-\lambda t}$

    $P\{A_1 <= t\} = 1 - e^{-\lambda t}$ [cdf of exp(λ)]
  - Interarrival times, A_1, A_2, …, are exponentially distributed and independent with mean $1/\lambda$

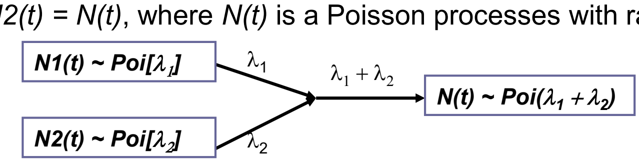


Splitting and Pooling [Poisson Dist'n]

- Splitting:
  - Suppose each event of a Poisson process can be classified as Type I, with probability p and Type II, with probability $1-p$.
  - $N(t) = N1(t) + N2(t)$, where $N1(t)$ and $N2(t)$ are both Poisson processes with rates λp and $\lambda(1-p)$



- Pooling:
  - Suppose two Poisson processes are pooled together
  - $N1(t) + N2(t) = N(t)$, where $N(t)$ is a Poisson processes with rates $\lambda_1 + \lambda_2$



3.5 Poisson process;

Nonstationary Poisson Process (NSPP) [Poisson Dist'n]

- Poisson Process without the stationary increments, characterized by $\lambda(t)$, the arrival rate at time t.
- The expected number of arrivals by time t, $\Lambda(t)$:

$$\Lambda(t) = \int_0^t \lambda(s)ds$$

- Relating stationary Poisson process $n(t)$ *with rate* $\lambda=1$ and NSPP $N(t)$ with rate $\lambda(t)$:
  - Let arrival times of a stationary process with rate $\lambda = 1$ be $t_1, t_2, \ldots$, and arrival times of a NSPP with rate $\lambda(t)$ be $T_1, T_2, \ldots$, we know:

$$t_i = \Lambda(T_i)$$
$$T_i = \Lambda^{-1}(t_i)$$

Nonstationary Poisson Process (NSPP) [Poisson Dist'n]

- Example: Suppose arrivals to a Post Office have rates 2 per minute from 8 am until 12 pm, and then 0.5 per minute until 4 pm.
- Let t = 0 correspond to 8 am, NSPP $N(t)$ has rate function:

$$\lambda(t) = \begin{cases} 2, & 0 \le t \prec 4 \\ 0.5, & 4 \le t \prec 8 \end{cases}$$

Expected number of arrivals by time t:

$$\Lambda(t) = \begin{cases} 2t, & 0 \le t \prec 4 \\ \int_0^4 2ds + \int_4^t 0.5ds = \frac{t}{2} + 6, & 4 \le t \prec 8 \end{cases}$$

- Hence, the probability distribution of the number of arrivals between 11 am and 2 pm.

$$\begin{aligned} P[N(6) - N(3) = k] &= P[N(\Lambda(6)) - N(\Lambda(3)) = k] \\ &= P[N(9) - N(6) = k] \\ &= e^{(9-6)}(9-6)^k/k! \quad = e^3(3)^k/k! \end{aligned}$$

3.6 Empirical Distributions

A distribution whose parameters are the observed values in a sample of data.

- May be used when it is impossible or unnecessary to establish that a random variable has any particular parametric distribution.
- Advantage: no assumption beyond the observed values in the sample.
- Disadvantage: sample might not cover the entire range of possible values.

UNIT 4: QUEUEING MODELS

4.1 Characteristics of Queueing System

- The key element's of queuing system are the **"customer and servers".**
- **Term Customer:** Can refer to people, trucks, mechanics, airplanes or anything that arrives at a facility and requires services.
- **Term Server:** Refer to receptionists, repairperson, medical personal, retrieval machines that provides the requested services.

4.1.1 Calling Population

- The population of potential customers referred to as the **"calling population".**
- The calling population may be assumed to be finite or infinite.
- The calling population is finite and consists
- In system with a large population of potential customers, the calling population is usually assumed to be infinite.
- The main difference between finite and infinite population models is how the arrival rate is defined.
- In an infinite population model, arrival rate is not affected by the number of customer who have left the calling population and joined the queueing.

4.1.2 System Capacity

- In many queueing system , there is a limit to the number of customers that may be in the waiting line or system.
- An arriving customer who finds the system full does not enter but returns immediately to the calling population.

4.1.3 Arrival Process

- The arrival process for "**Infinite population"** models is usually characterized in terms of interarrival time of successive customers.
- Arrivals may occur at scheduled times or at random times.
- When random times , the interarrival times are usually characterized by a probability distribution.
- Customer may arrive one at a time or in batches, the batches may be of constant size or random size.
- The second important class of arrivals is scheduled arrivals such as scheduled airline flight arrivals to an input.
- Third situation occurs when one at customer is assumed to always be present in the queue. So that the server is never idle because of a lack of customer.
- For finite population model, the arrivals process is characterized in a completely different fashion.
- Define customer as pending when that customer is outside the queueing system and a member of the calling population

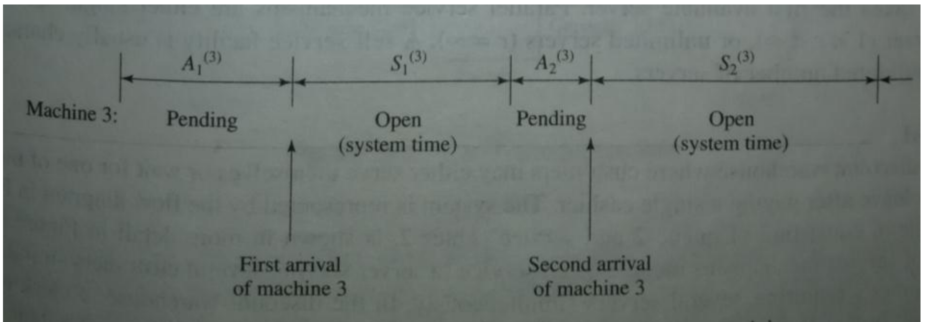


Figure 6.2 Arrival process for a finite-population model.

4.1.4 Queue Behavior and Queue Discipline

- It refers to the actions of customers while in a queue waiting for the service to begin.
- In some situations, there is a possibility that incoming customers will balk(leave when they see that the line is too long) , renege(leave after being in the line when they see that the line is moving slowly) , or jockey(move from one line to another if they think they have chosen a slow line).
- Queue discipline refers to the logical ordering of the customers in a queue and determines which customer will be chosen for service when a server becomes free.
- Common queue disciplines include FIFO, LIFO, service in random order(SIRO), shortest processing time first(SPT) and service according to priority (PR).

4.1.5 Service Times and Service Mechanism

- The service times of successive arrivals are denoted by s1, s2, sn.. They may be constant or of random duration.
- When {s1,s2,sn} is usually characterized as a sequence of independent and identically distributed random variables.
- The exponential, weibull, gamma, lognormal and truncated normal distribution have all been used successively as models of service times in different situations.
- A queueing system consists of a number of service centers and inter connecting queues. Each service center consists of some number of servers c, working in parallel.
- That is upon getting to the head of the line of customer takes the first available server.
- Parallel Service mechanisms are either single server or multiple server($1<c<\infty$) are unlimited servers($c=\infty$).
- A self service facility is usually characterized as having an unlimited number of servers.

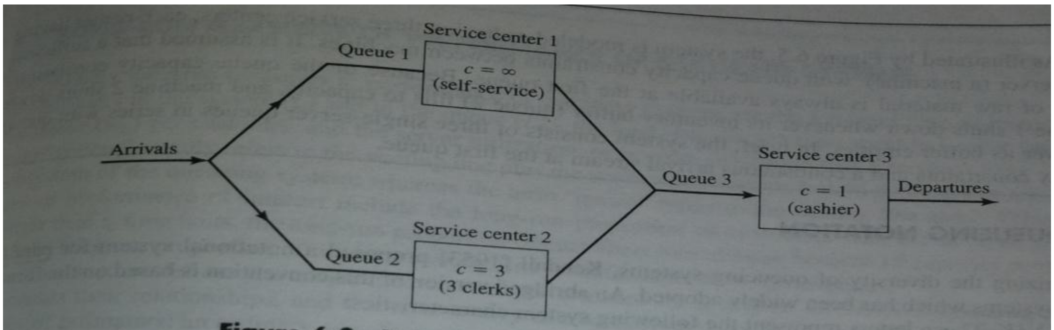


Figure 6.3 Discount warehouse with three service centers.

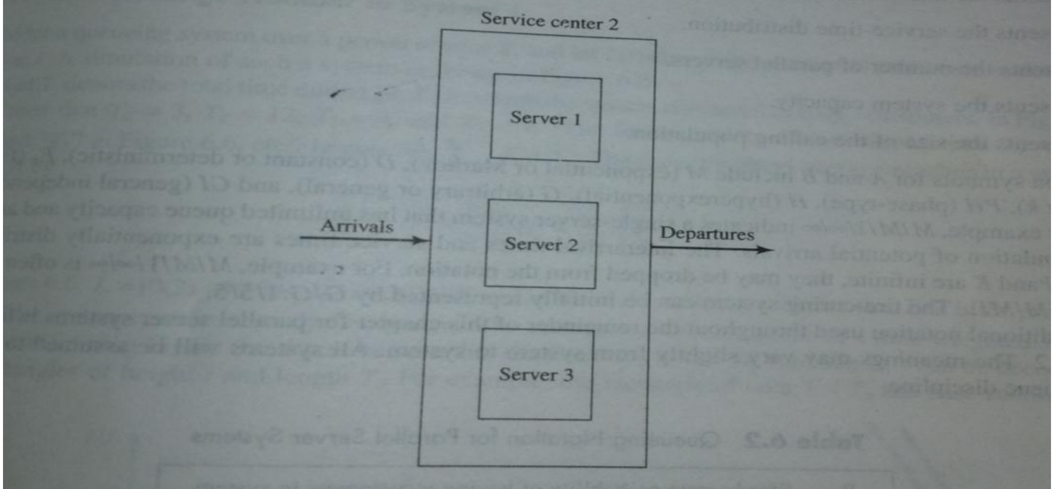


Figure 6.4 Service center 2, with $c = 3$ parallel servers.

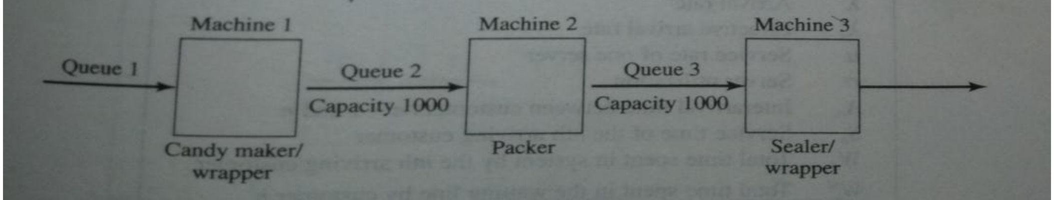


Figure 6.5 Candy-production line.

4.2 <u>Queueing Notation(Kendal's Notation)</u>

- Kendal's proposal a notational s/m for parallel server s/m which has been widely adopted.
- An a bridge version of this convention is based on format A|B|C|N|K
- These letters represent the following s/m characteristics:

A-Represents the InterArrival Time distribution
B-Represents the service time distribution
C-Represents the number of parallel servers
N-Represents the s/m capacity
K-Represents the size of the calling populations

Common symbols for A & B include M(exponential or Markov), D(constant or deterministic), E_k (Erlang of order k), PH (phase-type), H(hyperexponential), G(arbitrary or general), & GI(general independent).

- For eg, M|M|1|∞|∞ indicates a single server s/m that has unlimited queue capacity & an infinite population of potential arrivals
- The interarrival tmes & service times are exponentially distributed when N & K are infinite, they may be dropped from the notation.
- For eg, , M|M|1|∞|∞ is often short ended to M|M|1. The tire-curing s/m can be initially represented by G|G|1|5|5.

- Additional notation used for parallel server queueing s/m are as follows:

Table 6.2 Queueing Notation for Parallel Server Systems

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 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

4.3 Long-run Measures of performance of queueing systems

- The primary long run measures of performance of queueing system are the long run time average number of customer in s/m(L) & queue(L_Q)
- The long run average time spent in s/m(w) & in the queue(w_Q) per customer
- Server utilization or population of time that a server is busy (p).

4.3.1 Time average Number in s/m (L):

- Consider a queueing s/m over a period of time T & let L(t) denote the number of customer I the s/m at time t.
- Let Ti denote the total time during[0,T] in which the s/m contained exactly I customers.

$$\boxed{\hat{L} = \sum_{i=0}^{\infty} i\left(\frac{T_i}{T}\right)}$$

- where $\hat{L}$ is the time weighted average number in a System. i

- Consider an example of Queueing s/m with line Segment 3, 12, 4, 1. Compute the time weighted - average number in a s/m.

Soln

$$\hat{L} = \sum_{i=0}^{\infty} i\left(\frac{T_i}{T}\right)$$

$$\hat{L} = \left[\, 0(3) + 1(12) + 2(4) + 3(1) \,\right] / 20$$

$$= 23/20$$

$$= 1.15 \text{ customers.}$$

4.3.2 Average Time spent in s/m per customer (w):

- Average s/m time is given as:

$$\boxed{\hat{w} = \frac{1}{N}\sum_{i=1}^{N} w_i} \quad \text{--- (1)}$$

where,

N - is the number of arrivals during [0,T]

w_i - is customer spend in the s/m during [0,T]

- For stable s/m N-> ∞

$$\boxed{\hat{w} \rightarrow w} \quad \text{-- (2)}$$

With probability 1, where w is called the long-run average s/m time.

- Considering the equation 1 & 2 are written as,

$$\boxed{\hat{w}_Q = \frac{1}{N}\sum_{i=1}^{N} w_i^Q \rightarrow w_Q}$$

where,

w_i^Q = is the total time customer i spends waiting in queue.

$\hat{w}_Q$ - is the observed average time spent in queue.

w_Q - is the long run average delay per customer

example :- Consider the Queueing s/m with $N=5$ Customer arrive at $w_1 = 2$ & $w_5 = 20-16=4$ but w_2, w_3 & w_4 cannot be computed unless more is know about the s/m. Arrival occur at times 0, 3, 5, 7 & 16 & departures occur at time 2, 8, 10 & 14.

Sol^n

$$\boxed{\hat{w} = \frac{1}{N}\sum_{i=1}^{N} w_i}$$

$w_1 = 2, \quad w_5 = 4$

$w_2 = 8 - 3 = 5$

$w_3 = 10 - 5 = 5$

$w_4 = 14 - 7 = 7$

$$\hat{w} = \frac{2 + 5 + 5 + 7 + 4}{5} = \frac{23}{5} = 4.6 \text{ time units.}$$

4.3.3 Server utilization:

- Server utilization is defined as the population of time server is busy
- Server utilization is denoted by ꞵ is defined over a specified time interval[01]
- Long run server utilization is denoted by p

Ƥ -> P   as T -> ∞

❖ **Server utilization in G|G|C|∞|∞ queues**

- Consider a queuing s/m with c identical servers in parallel
- If arriving customer finds more than one server idle the customer choose a server without favoring any particular server.
- The average number of busy servers say Ls is given by,

Ls = λ / μ   0<= Ls <= C

- The long run average server utilization is defined by

$$\boxed{p = \frac{L_s}{c} = \frac{\lambda}{c\mu}} \quad \therefore \; 0 \le p \le 1$$

- The utilization P can be interpreted as the proportion of time an arbitrary server is busy in the long run

Example:

Customer arrive at random to a license bureau at a rate of $\lambda = 50$ customer per hour. Currently there are 20 clerks, each serving $\mu = 5$ customers per hour on the average. Compute long-run or steady state average utilization of a server & average number of busy server.

Sol^n

Average utilization of server:

$$\boxed{\rho = \frac{\lambda}{c\mu}}$$

$$\rho = \frac{50}{20(5)} = 0.5$$

Average number of busy servers is:

$$\boxed{L_s = \frac{\lambda}{\mu}}$$

$$L_s = \frac{50}{5} = 10$$

4.4 STEADY-STATE BEHAVIOUR OF INFINITE-POPULATION MARKOVIAN MODLES

- For the infinite population models, the arrivals are assumed to follow a poisson process with rate λ arrivals per time unit
- The interarrival times are assumed to be exponentially distributed with mean 1/λ
- Service times may be exponentially distributed(M) or arbitrary(G)
- The queue discipline will be FIFO because of the exponential distributed assumptions on the arrival process, these model are called “MARKOVIAN MODEL”.
- The steady-state parameter L, the time average number of customers in the s/m can be computed as

$$L = \sum_{n=0}^{\infty} nPn$$

Where Pn are the steady state probability of finding n customers in the s/m

- Other steady state parameters can be computed readily from little equation to whole system & to queue alone

$$w = L/\lambda$$
$$w_Q = w - (1/\mu)$$
$$L_Q = \lambda w_Q$$

Where λ is the arrival rate & μ is the service rate per server

4.4.1 SINGLE-SERVER QUEUE WITH POISSON ARRIVALS & UNLIMITED CAPACITY: M|G|1

- Suppose that service times have mean $1/\mu$ & variance σ^2 & that there is one server
- If $P = \lambda / \mu < 1$, then the M|G|1 queue has a steady state probability distribution with steady state characteristics
- The quantity $P = \lambda / \mu$ is the server utilization or lon run proportion of time the server is busy
- Steady state parameters of the M|G|1 are:

Notation	Description
① $P = \frac{\lambda}{\mu}$	• P is server ultilization • λ is arrival rate • μ is service rate
② $L = P + \frac{P^2(1+\sigma^2\mu^2)}{2(1-P)}$	• L is long run time average number of customers in s/m • σ is the mean service time
③ $w = \frac{1}{\mu} + \frac{\lambda(1/\mu^2 + \sigma^2)}{2(1-P)}$	• w is long run average time spent in s/m per customer
④ $w_Q = \frac{\lambda(1/\mu^2 + \sigma^2)}{2(1-P)}$	• wQ is long run average time spent in queue per customer
⑤ $L_Q = \frac{P^2(1+\sigma^2\mu^2)}{2(1-P)}$	• LQ is long run time avg no. of customer in queue
⑥ $P_0 = 1-P$	• P_0 is steady state probability of customers in s/m

__Example__ : Consider a candy factory for making a candy at rate $\lambda = 1.5$ per hour. Observation over several months has found by the single m/c. It's mean service time $\sigma = 1/2$ hour, service rate is $\mu = 2$. Compute long run time average number of customer in s/m, long run time average number of customer in queue & long run average time spent in queue per customer.

__Sol^n__

↳ long run time average number of customer in s/m

$$\boxed{L = \rho + \frac{\rho^2 (1 + \sigma^2 \mu^2)}{2(1-\rho)}}$$

$$\rho = \frac{\lambda}{\mu} = 1.5/2 = 0.75$$

$$L = 0.75 + \frac{0.75 (1 + (0.5)^2 (2)^2)}{2(1 - 0.75)}$$

$$= 3.75$$

↳ long run time average number of customer in queue

$$\boxed{L_Q = \frac{\rho^2 (1 + \sigma^2 \mu^2)}{2(1-\rho)}}$$

$$L_Q = \frac{(0.75)^2 (1 + (0.5)^2 (2)^2)}{2(1-0.75)}$$

$$= 2.25$$

↳ __long run average time spent in queue per customer:__

$$w_Q = \frac{\lambda (1/\mu^2 + \sigma^2)}{2(1-\rho)}$$

$$w_Q = \frac{1.5 (1/(2)^2 + (0.5)^2)}{2(1 - 0.75)}$$

$$= 1.5$$

- Steady state parameters of the m/m/1 queue

Notation	Description
$L = \frac{\rho}{1-\rho}$	• L is long run time average number of customer in s/m • ρ is server utilization
$\omega = \frac{1}{\mu(1-\rho)}$	• ω is long run average time spent is s/m per customer • μ is service rate
$\omega_q = \frac{\rho}{\mu(1-\rho)}$	• ω_q is long run average time spent in queue per customer
$L_q = \frac{\rho^2}{1-\rho}$	• L_q is long run time average number of customer in queue
$P_n = (1-\rho)\rho^n$	• P_n is steady state probability of n customer in s/m

4.4 2 MULTISERVER QUEUE: M|M|C|∞|∞

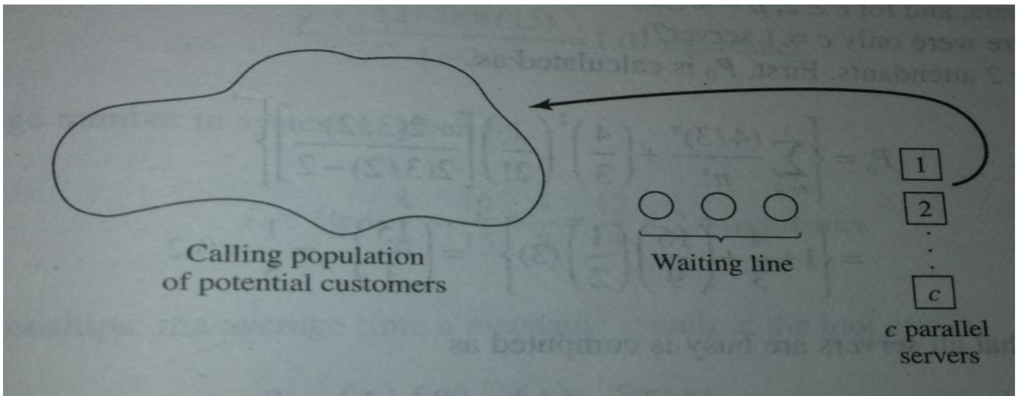


Figure 6.13 Multiserver queueing system.

- Suppose that there are c channels operating in parallel
- Each of these channels has an independent & identical exponential service time distribution with mean $1/\mu$
- The arrival process is poisson with rate λ. Arrival will join a single queue & enter the first available service channel

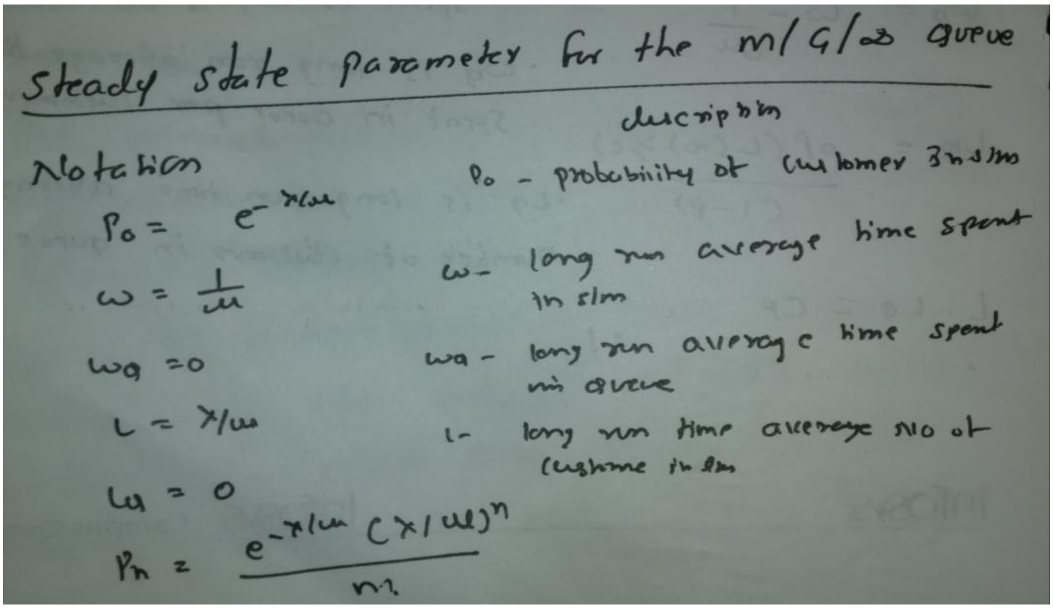
- For the M|M|C queue to have statistical equilibrium the offered load must satisfy λ/μ <c in which case λ/ (cμ) = P the server utilization.

· The steady state parameters for the m/m/c queue

Notation	Description
· $\rho = \frac{\lambda}{\mu}$	· ρ is server utilization · λ arrival rate · μ service rate
· $P_0 = \left\{\left[\sum_{n=0}^{c-1} \frac{(c\rho)^n}{n!}\right] + \left[(c\rho)^c \left(\frac{1}{c!}\right)\left(\frac{1}{1-\rho}\right)\right]\right\}^{-1}$	· is steady state for probability of customer in s/m
· $L = c\rho + \frac{\rho P(L(\infty) \geq c)}{(1-\rho)}$	· L is long run time average number of customer in s/m
$w = \frac{L}{\lambda}$	· w is long run average time spent is s/m per customer
$w_Q = w - \frac{1}{\mu}$	· w_Q is long run average time spent in queue per customer
$L_Q = \frac{\rho P(L(\infty) \geq c)}{(1-\rho)}$	· L_Q is long run time average number of customer in queue
$L - L_Q = c\rho$	· 1

WHEN THE NUMBER OF SERVERS IS INFINITE (M|c|∞|∞)

- There are at least three situations in which it is appropriate to treat the number of server as infinite
  1. When each customer is its own server in other words in a self service s/m
  2. When service capacity far exceeds service demand as in a so called ample server s/m
  3. When wee want to know how many servers are required so that customer will rarely be delayed.



Steady state parameter for the M/G/∞ queue

Notation	description
$P_0 = e^{-\lambda/\mu}$	P_0 - probability of customer in s/m
$w = \frac{1}{\mu}$	w - long run average time spent in s/m
$w_q = 0$	w_q - long run average time spent in queue
$L = \lambda/\mu$	L - long run time average no of customer in s/m
$L_q = 0$	

$$P_n = \frac{e^{-\lambda/\mu}(\lambda/\mu)^n}{n!}$$

4.5 STEADY STATE BEHAVIOR OF FINITE POPULATION MODELS (M|M|C|K|K)

- In many practical problems, the assumption of an infinite calling population leads to invalid results because the calling population is, in fact small.
- When the calling population is small, the presence of one or more customers in the system have a strong effect on the distribution of future arrivals and the use of an infinite population model can be misleading.
- Consider a finite calling population model with k customers. The time between the end of one service visit and the next call for service for each member of the population is assumed to be exponentially distributed with mean 1/ λ time units.
- Service times are also exponentially distributed, with mean 1/ μ time units. There are c parallel servers and system capacity is so that all arrivals remain for service. Such a system is shown in figure.

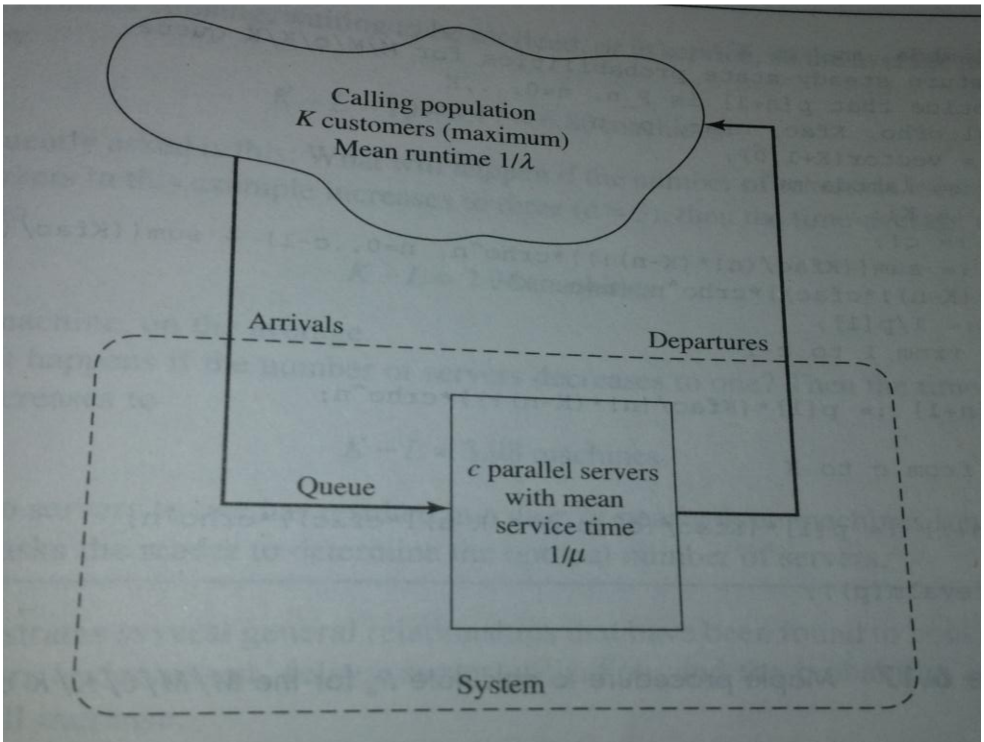


Figure 6.16 Finite-population queueing model.

The effective arrival rate λ_e has several valid interpretations:

Λ_e = long-run effective arrival rate of customers to queue
= long-run effective arrival rate of customers entering service
= long-run rate at which customers exit from service
= long-run rate at which customers enter the calling population
=long-run rate at which customers exit from the calling population.

Table 6.8 Steady-State Parameters for the M/M/c/K/K Queue

P_0	$\left[\sum_{n=0}^{c-1}\binom{K}{n}\left(\frac{\lambda}{\mu}\right)^n + \sum_{n=c}^{K}\frac{K!}{(K-n)!c!c^{n-c}}\left(\frac{\lambda}{\mu}\right)^n\right]^{-1}$
P_n	$\begin{cases}\binom{K}{n}\left(\frac{\lambda}{\mu}\right)^n P_0, & n=0,1,\ldots,c-1\\ \frac{K!}{(K-n)!c!c^{n-c}}\left(\frac{\lambda}{\mu}\right)^n P_0, & n=c,c+1,\ldots,K\end{cases}$
L	$\sum_{n=c}^{K} nP_n$
L_Q	$\sum_{n=c+1}^{K}(n-c)P_n$
λ_e	$\sum_{n=0}^{K}(K-n)\lambda P_n$
w	L/λ_e
w_Q	L_Q/λ_e
ρ	$\frac{L-L_Q}{c}=\frac{\lambda_e}{c\mu}$

4.6 NETWORKS OF QUEUE

- Many systems are naturally modeled as networks of single queues in which customer departing from one queue may be routed to another
- The following results assume a stable system with infinite calling population and no limit on system capacity.

1) Provided that no customers are created or destroyed in the queue,then the departure rate out of a queue is the same as the arrival rate into the queue over the long run.
2) If customers arrive to queue i at rate λi and a fraction $0 \le p_{ij} \le 1$ of them are routed to queue j upon departure, then the arrival rate from queue i to queue j is λ_{ipij} is over long run
3) The overall arrival rate into queue j,λ_i is the sum of the arrival rate from all source.If customers arrive from outside the network at rate a_i then

$$\boxed{\lambda_j = a_j + \sum_{\text{all } i} \lambda_i P_{ij}}$$

4) If queue j has ci<∞ parallel servers, each working at rate μ ,then the long run utilization of each server is

$$\boxed{\rho_j = \frac{\lambda_j}{c_j \mu_j}}$$

& Pj<1 is required for queue to be stable

5) If, for each queue j ,arrivals from outside the network form a poisson process with rate a and if there are ci identical services delivering exponentially distributed service times with mean $1/\mu$ then in steady state queue j behaves like a M|M|C; queue with arrival rate

$$\boxed{\lambda_j = a_j + \sum_{\text{all } i} \lambda_i P_{ij}}$$