University of Karlsruhe (TH)
Department of Mathematics
Institute of Stochastics

Script

Stochastic Methods in Industry

This script was written by Ana Aleksieva in the authority of the University of Karlsruhe according to the lecture prof. P. R. Parthasaraty, Ph. D. from the IIT Mandras (India), did in the winter semester 2007/08 as a visiting professor at the University of Karlsruhe.

Contents

Chapter 1

Background

1.1 Basic Elements of a Queueing Model

The principal actors in a queuing situation are the customer and the server. The interaction between the customer and the server is related to the period of time the customer needs to complete his service. Thus, from the standpoint of the customers-arrivals, we are interested in the time intervals that separate the successive arrivals. Also, in the case of service, it is the service time per customer that counts in the analysed arrivals and service time distributions. These distributions may represent situations where customers arrive and are served individually(e.g, banks or supermarkets). In other situations, customers may arrive and/or be served in groups(e.g, restaurants). The latter case is normally referred as bulk queues.

The manner of choosing customers from the waiting line to start service defines the service discipline. The most common, and apparently fair, discipline is the FCFS rule(first come, first served). In the queues LCFS(last come, first served) and SIRO(service in random order), customers, arriving at a facility may be put in priority queues, such tha,t those with a higher priority will receive preference to start service first.

The facility may include more than one server, and then it could allow as many customers as the number of servers to be served simultaneously. The facility may comprise a number of series stations through which the customer may pass before the service is completed(e.g, processing of a product on a sequence of machines). In this case, waiting lines may not be allowed between the stations. The resulting situations are normally known as queues in series or tandem queues. The most general design of a service facility includes both series and parallel processing stations. This is what we call network queues.

In certain situations, only a limited number of customers may be allowed, possibly because of space limitations(e.g, cars allowed in a drive-in bank). Once the queue fills to capacity, the service for newly arriving customers is denied and they may not join the queue. The calling source may be capable of generating a finite number of customers or(theoretically) infinitely many customers. In a machine shop with a total of M machines, the calling source before any machine breaks down consists of M potential customers. Once a machine is broken, it becomes a customer and hence incapable of generating new calls until it is repaired.

A ”human” customer may jockey from one waiting line to another hoping to reducing his or her waiting time. Some ”human” customers also may balk from joining a waiting line all together because they anticipate a long delay, or they may renege after being in the queue for a while because the wait has been too long.

The basic elements of a queueing model depend on the following factors:

1. Arrivals distribution(single or bulk arrivals).
2. Service-time distribution(single or bulk service).
3. Design of service facility(series, parallel or network stations).
4. Service discipline(FCFS, LCFS, SIRO) and service priority.
5. Queue size(finite or infinite).
6. Calling source(finite or infinite).
7. Human behavior(jockeying, balking and reneging).

1.2 Queues with Combined Arrivals and Departures

A notation that is particularly suited for summarizing the main charecteristics of parallel queues has been universally standardized in the following form

$$(a/b/c) : (d/e/f)$$

where the symbols a, b, c, d, e and f stand for basic elements of the model as follows:

$a \equiv$ arrivals distribution

$b \equiv$ service time(or departures) distribution

$c \equiv$ number of parallel servers(c=1, 2, ..., ∞)

The standard notation replaces the symbols a and b for arrivals and departures by the following codes:

$M \equiv$ Poisson(or Markovian) arrival or departure distribution (or equivalently exponential interarrival or service-time distribution)

$D \equiv$ constant or deterministic interarrival or service time

$E_k \equiv$ Erlangian or Gamma distribution of interarrival or service time distribution with parameter k.

$GI \equiv$ general independent distribution of arrivals(or interarrival time)

$G \equiv$ general distribution of departures(or service time)

Under steady-state conditions we shall be interested in determining the following basic measures of performance.

$p_n \equiv$ (steady-state) probability of n customers in a system

$L_s \equiv$ expected number of customers in a system

$L_q \equiv$ expected number of customers in a queue

$W_s \equiv$ expected waiting time in a system (time in the queue + service time)

$W_q \equiv$ expected waiting time in a queue

By definition, we have

$$L_s = \sum_{n=0}^{\infty} np_n,$$

$$L_q = \sum_{n=c}^{\infty}(n-c)p_n$$

and

$$L_s = \lambda W_s,$$

$$L_q = \lambda W_q.$$

When customers arrive with rate λ but not all arrivals can join the system (this can happen, for example, when there is a limit on the maximum number of members in the system), the equations must be modified by redefining λ to include only those customers that actually join the system. Thus letting

$$\lambda_{eff} = \{\text{effective arrival rate for those, who join the system}\}$$

we have

$$L_s = \lambda_{eff} W_s,$$
$$L_q = \lambda_{eff} W_q,$$
$$W_s = W_q + \frac{1}{\mu},$$
$$L_s = L_q + \frac{1}{\mu},$$
$$\lambda_{eff} = \mu + (L_s - L_q).$$

We can determine all the basic measures of performance in the following order:

$$p_n \to L_s = \sum_{n=0}^{\infty} np_n \to W_s = \frac{L_s}{\lambda} \to W_q = W_s - \frac{1}{\mu} \to L_q = \lambda W_q.$$

Resource Utilization and Traffic Intensity .

Resource Utilization represents the fraction of time a server is engaged in providing service as defined below:

$$\text{Utilization}(\rho) = \frac{\text{time a server is occupied}}{\text{time a server is available}}.$$

For example, consider a queueing system with m servers ($m \geq 1$). Let λ denote the average arrival rate and μ denote the average service rate of each server. If we observe the system in an arbitary interval $(t, t+T)$, then each server on the average will serve

$$\frac{\text{average number of arrivals in (t, t+T)}}{\text{average number of customers, served by m servers}} = \frac{\lambda T}{m\mu}.$$

Therefore,

$$\rho = \frac{\frac{\lambda T}{m\mu}}{T} = \frac{\lambda}{m\mu}.$$

It is clear from the definition that ρ is dimensionless and should be less than unity in order for a server to cope with the service demand, or in other words for the system to be stable.

Flow Conversation Law .

For a stable queueing system, the rate of customers entering the system should equal to the rate of customers leaving the system if we observe it for a sufficient long period of time; that is $\lambda_{out} = \lambda_{in}$. This is beacuse if $\lambda_{out} > \lambda_{in}$, then there will be a steady build-up of customers and the system will eventually become unstable. On the other hand, if $\lambda_{out} < \lambda_{in}$, then customers are created within the system. This notion of flow conversation is useful when we wish to calculate throughput in queueing networks.

Little's Formula .

Before we start examining the stochastic behaviour of a queueing system, let us first state a very simple and yet powerful result that governs its steady-state performance measures. $L = \lambda W$ where L is the average number of customers in the queueing system, W is the average waiting time of a customer in the queueing system, and λ is the average arrival rate of customers to the queueing system. This result is true for any queueing system with the proper interpretation of L, λ and W.

1.3 Probability Theory

Definition 1.1. σ - field of subsets A of a set Ω.

A non-empty collection of subsets A of a set Ω is called a σ - field of subsets of Ω, if the following two properties hold:

(i) If Λ is in A, then Λ^c is also in A.

(ii) If Λ_n is in A, $n = 1, 2, ...$ then $\cup_{n=1}^{\infty}\Lambda_n$ and $\cup_{n=1}^{\infty}\Lambda_n$ are both in A.

Definition 1.2. Probability Space

A probability measure P on a σ- field of subsets A of a set Ω is a real-valued function with domain A, satisfying the following properties:

(i) $P(\Omega) = 1$

(ii) $P(\Lambda) \geq 0$, for all Λ in A.

(iii) If $\Lambda_n, n = 1, 2, ...$ are mutually disjoint sets in A, i.e., for all i and j, $\Lambda_i \cap \Lambda_j = \emptyset$ if $i \neq j$, then $P(\cup_{n=1}^{\infty}\Lambda_n) = \sum_{n=1}^{\infty} P(\Lambda_n)$.

A probability space, denoted by (Ω, A, P), represents a set Ω, a σ- field of subsets A of a set Ω, and a probability measure P defined on A.

Definition 1.3. Discrete Random Variable A discrete random variable X on a probability space (Ω, A, P) is a function X with domain Ω and a range, a finite or countably infinite subset $x_1, x_2, ...$ of the real numbers R such that $\omega : X(\omega) = x_i$ is an event for all i.

Definition 1.4. Probability Mass Function (pmf):

A real-valued function p depends on R by $p_i = P(X = i)$ is called the pmf of the discrete random variable X which satisfies the following properties:

(i) $p_i \geq 0$ for all $i\epsilon R$.

(ii) $\{i\epsilon R : p_i \neq 0\}$ is a finite or countably infinite subset of R. Let $S = \{i_1, i_2, ...\}$ denote this set.

Then

(iii) $\sum_{k\epsilon S} p_k = 1$.

Example Poisson random variable

A discrete random variable X is said to be a Poisson random variable with parameter $\lambda > 0$ if its pmf is given by

$$p_i = \begin{cases} \frac{\lambda^i e^{-\lambda}}{i!} & i = 0, 1, 2... \\ 0 & \text{otherwise} \end{cases} \tag{1.1}$$

clearly, $p_i \geq 0$ for all i and $\sum_{i=0}^{\infty} p_i = 1$.

Definition 1.5. Distribution Function

A function $F(t)$, $-\infty < t < \infty$, defined by $F(t) = P(X \leq t) = \sum_{i \leq t} p_i$ is called the distribution function of the random variable X.

Example Let X be a discrete random varible with pmf: $p_i = P(X = i) = \frac{1}{\pi^2}\frac{6}{i^2}$, $i = 1, 2, ...$ Then $F(x) = \frac{6}{\pi^2}\sum_{i=1}^{\infty} \frac{1}{\pi^2}\varepsilon(x - k)$, where

$$\varepsilon(\xi) = \begin{cases} 1 & \xi \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Definition 1.6. Continuous Random Variable

Let X be a random variable defined on (Ω, A, P) with a distribution function F, that is, $F(x) = P(X \leq x)$. Then X is said to be a continuous random variable if F is absolutely continuous, i.e., if there exists a nonnegative function $f(x)$, such that for every real number x we have $F(x) = \int_{-\infty}^{x} f(t)dt$.

The function f is called the probability density function (pdf) of the random variable X.

Properties of F and f:

(i) $0 \leq F(x) \leq 1$ for all x;

(ii) $F(x)$ is a non-decreasing function of x;
(iii) $F(-\infty) = 0$ and $F(\infty) = 1$;
(iv) $f(x) \geq 0$ and satisfies

$$lim_{x\to\infty}F(x) = F(\infty) = \int_{-\infty}^{\infty} f(t)dt = 1.$$

Let $a, b\epsilon R$ with $a < b$, then

$$P(a < x \leq b) = F(b) - F(a) = \int_{a}^{b} f(t)dt.$$

Example Let X be a continuous random variable which *pdf* is given by

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

Clearly, $f \geq 0$. Also, $\int_{-\infty}^{\infty} f(x)dx = \int_0^1 xdx + \int_1^2 (2-x)dx = 1$. The distribution function F of X is given by

$$F(x) = \begin{cases} 0 & x \leq 0 \\ \int_0^x udu = \frac{x^2}{2} & 0 < x \leq 1 \\ \int_0^1 udu + \int_1^x (2-u)du = 2x - \frac{x^2}{2} - 1 & 1 < x \leq 2 \\ 1 & x > 2 \end{cases}$$

Definition 1.7. Mean or Expected value of a discrete random variable

Let X be a discrete random variable with pmf $p_i = P(X = x_i), i = 1, 2, ...$ If $\sum_{i=1}^{\infty} |x_i| p_i < \infty$ we say that the expected value of X exists and write $E[X] = \sum_{i=1}^{\infty} x_i p_i$. If the sum is not finite, then we say that $E[X]$ does not exist.

Example Let X be a discrete random variable with pmf

$$p_j = P(X = (-1)^{j+1}\frac{3^j}{j}) = \frac{2}{3^j}, j = 1, 2, ...$$

Clearly, $p_j > 0$ and $\sum_{j=1}^{\infty} p_j = 1$. $\sum_{i=1}^{\infty} |x_i| p_i = 2\sum_{i=1}^{\infty} \frac{1}{i} = \infty$.
Therefore, $E[X]$ does not exist.

Example Let X be a Poisson random variable with parameter λ. Then its pmf is given by $p_j = e^{-\lambda}\frac{\lambda^j}{j!}, j = 0, 1, 2...$

Since X takes only non-negative integer values, $\sum_{i=1}^{\infty} |x_i| p_i = \sum_{i=1}^{\infty} x_i p_i = \lambda$.

Result If X is a nonnegative integer-valued random variable, then

$$E[X] = \sum_{i=1}^{\infty} P(X \geq i),$$

if it is provided that the series on the right hand side converges.

Definition 1.8. Probability Generating Function (pgf) or z-transform

Let X be a discrete random variable and let $p_k = P(X = k)$, $k = 0, 1, 2, ...$ Then the function defined by

$$P_X(z) = \sum_{k=0}^{\infty} p_k z^k, |z| \leq 1, \text{ i.e. } P_X(z) = E[z^X]$$

is called the pgf of X.

Example Let X be a Poisson random variable with parameter λ. The pgf of X is given by

$$P_X(z) = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} z^k = e^{-\lambda(1-z)}$$

Result If X and Y are two independent discrete random variables, then

$$P_{X+Y}(z) = P_X(z) + P_Y(z).$$

Result Superposition of Poisson random variables is again a Poisson random variable.

Definition 1.9. Mean or Expected value of a continuous random variable

Let X be a random variable of the continuous type and has *pdf* f. If

$$(*) \int_{-\infty}^{\infty} |x| f(x) dx < \infty$$

we say that the expected value of X exists and write $E[X] = \int x f(x) dx$. If (*) is not satisfied, then we say that $E[X]$ does not exist.

Example Let the *pdf* of X be given by $f(x) = \frac{1}{\pi} \frac{1}{1+x^2}$ (Cauchy pdf). Then,

$$\frac{1}{\pi} \int_{-\infty}^{\infty} |x| f(x) dx = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{|x|}{1+x^2} = \infty.$$

Therefore, $E[X]$ does not exist.

Example Let X be an exponentially distributed random variable with parameter $\mu > 0$. Then its *pdf* is given by

$$f(x) = \mu e^{-\mu x}, \mu > 0, x > 0,$$
$$E[X] = \int_0^\infty x\mu e^{-\mu x} dx = \frac{1}{\mu}.$$

Properties of the Expectation Let X and Y be any two random variables. Then
(i) $E[aX + b] = aE[X] + b$, a, b are constants;
(ii) $E[X + Y] = E[X] + E[Y]$.

In fact, if X_i, $i = 1, 2, ...n$ are *iid* random variables, then $E[\sum_{i=0}^n X_i] = \sum_{i=0}^n E[X_i]$, if $E[X_i]$ exists, for $i = 1, 2, ...n$.

Moment If $E[X^n]$ exists for some positive integer n, we call $E[X^n]$ the n-th moment of X about the origin and it is given by

$$E[X^n] = \begin{cases} \int_{-\infty}^\infty x^n f(x) dx, & f \text{ is continuous} \\ \sum_i x_i^n P(X = x_i), & f \text{ is discrete} \end{cases}$$

Example Let X be a random variable with *pdf*

$$f(x) = \begin{cases} \frac{2}{x^3}, & x \geq 1 \\ 0, & x < 1 \end{cases},$$

then

$$E[X] = \int_1^\infty \frac{2}{x^2} dx = 2$$

but $E[X^2] = \int_1^\infty \frac{2}{x} dx$ does not exist.

Example Let X have the uniform distribution on the first N natural numbers, that is, $P(X = k) = \frac{1}{N}, k = 1, 2, ..., N$. Clearly, moments of all orders exist. In particular,

$$E[X] = \sum_{k=1}^N \frac{k}{N} = \frac{N+1}{2}$$
$$E[X^2] = \sum_{k=1}^N \frac{k^2}{N} = \frac{(N+1)(2N+1)}{6}.$$

Variance Let X be a discrete or continuous random variable. If $E[X^2]$ exists, we call $E[X - E[X]]^2$ the variance of X, and we write $\sigma^2 = Var[X] = E[X - E[X]]^2$. The quantity σ is called the standard deviation of X. The other way of expressing the variance of X is $Var[X] = E[X^2] - (E[X])^2$ which is obtained by the properties of the expectation.

Example Let X be a Poisson random variable with parameter $\lambda > 0$. Then

$$E[X] = \sum_{k=0}^{\infty} k e^{-\lambda} \frac{\lambda^k}{k!} = \lambda$$

and

$$E[X^2] = \sum_{k=0}^{\infty} k^2 e^{-\lambda} \frac{\lambda^k}{k!} = \lambda(\lambda + 1).$$

Therefore, $Var[X] = E[X^2] - (E[X])^2 = \lambda$.

Property of the variance Function Let X and Y be any two random variables. Then $Var[aX + bY] = a^2 Var[X] + b^2 Var[Y]$, where a, b are constant.

From now on we restrict our study to continuous random variables.

Joint Distribution The joint distribution $F(x, y)$ of a two-dimensional random variable (X, Y) is defined as

$$F(X, Y) = P(X \leq x, Y \leq y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v) dv du, \tag{1.2}$$

$x, y \epsilon R^2$ if f exists. The function f is called the joint *pdf* of (X, Y) satisfying

$$F(\infty, \infty) = \lim_{x \to \infty, y \to \infty} \int_{-\infty}^{x} \int_{-\infty}^{y} dv du = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dv du = 1$$

and

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y}$$

Example Let (X, Y) be a two dimensional random variable which *pdf* is given by

$$f(x, y) = \begin{cases} e^{-(x+y)} & 0 < x, y < \infty \\ 0 & \text{otherwise} \end{cases}$$

Then

$$F(x, y) = \int_{0}^{x} \int_{0}^{y} f(u, v) dv du = (1 - e^{-x})(1 - e^{-y}).$$

Therefore

$$F(x, y) = \begin{cases} (1 - e^{-x})(1 - e^{-y}) & 0 < x, y < \infty \\ 0 & \text{otherwise} \end{cases}$$

Definition 1.10. Marginal *pdfs*

If (X, Y) is a random variable with *pdf* f, then

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy \text{ and } f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx$$

are called the marginal pdf's of X and Y respectively. Clearly, $f_X(x) \geq 0$ and $f_Y(y) \geq 0$ and $\int_{-\infty}^{\infty} f_X(x) dx = 1$ and $\int_{-\infty}^{\infty} f_Y(y) dy = 1$.

Example Let (X, Y) be a two dimensional random variable which *pdf* is given by

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

Then the marginal *pdf* $f_X(x)$ of X is given by

$$f_X(x) = \int_{y=x}^{1} 2dy = \begin{cases} 2(1-x) & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

and the marginal *pdf* $f_Y(y)$ of Y is given by

$$f_Y(y) = \int_{x=1}^{0} 2dx = \begin{cases} 2y & 0 < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

Definition 1.11. Conditional pdf's

Let f be the *pdf* of (X, Y). Then at every point (x, y), at which f is continuous, $f_Y(y) > 0$ and is continuous, the conditional *pdf* of X given $Y = y$, is defined by

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}.$$

Similarly, the conditional *pdf* of Y given $X = x$, is defined by

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)},$$

if $f_X(x) > 0$.

Example Let (X, Y) be a two dimensional random variable which *pdf* is the given above $f*$. Then, the conditional *pdfs* are given by

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} = \frac{1}{y}, 0 < x < y.$$

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{1}{1-x}, x < y < 1.$$

Conditional Expectations The conditional expectation of X, given $Y = y$, is defined by

$$E[X|Y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)dx,$$

if $f_Y(y) > 0$.

Similarly, the conditional expectation of Y, given $X = x$, is defined by

$$E[Y|X] = \int_{-\infty}^{\infty} y f_{Y|X}(y|x)dy,$$

if $f_X(x) > 0$.

Example Let (X, Y) be a two dimensional random variable which *pdf* is again $f*$.
Then the conditional expectations $E[X|Y]$ and $E[Y|X]$ are given by

$$E[X|Y] = \int_0^y x f_{X|Y}(x|y)dx = \frac{y}{2}, 0 < y < 1$$

and

$$E[Y|X] = \int_x^1 y f_{Y|X}(y|x)dy = \frac{1+x}{2}, 0 < x < 1$$

Result $E[X] = E[E[X|Y]]$, if $E[X]$ exists.

Conditional Variance Conditional variance of a random variable X given Y is defined as

$$Var[X|Y] = E[X^2|Y] - (E[X|Y])^2;$$

$$E[X^2|Y] = \int_0^y x^2 f_{X|Y}(x|y)dx = \frac{y^2}{3}, 0 < y < 1;$$

and

$$E[Y^2|X] = \int_x^1 y^2 f_{Y|X}(y|x)dy = \frac{1+x+x^2}{3}, 0 < x < 1;$$

$$Var[X|Y] = E[X^2|Y] - (E[X|Y])^2 = \frac{y^2}{12}, 0 < y < 1;$$

$$Var[Y|X] = E[Y^2|X] - (E[Y|X])^2 = \frac{(x-1)^2}{12}, 0 < x < 1.$$

Theorem $Var[X] = E[Var[X|Y]] + Var[E[X|Y]]$

Independence Two random variables X and Y are said to be independent if

$$f(x, y) = f_X(x)f_Y(y), (x, y)\epsilon R^2$$

Example Let (X, Y) be a two dimensional random variable which *pdf* is given by

$$f(x, y) = \begin{cases} e^{-(x+y)} & 0 < x, y < \infty \\ 0 & \text{otherwise} \end{cases}$$

Clearly, $f_X(x) = e^{-x}$, $f_Y(y) = e^{-y}$ and hence $f(x, y) = f_X(x)f_Y(y)$. Thus, X and Y are independent.

Example Let (X,Y) be a two dimensional random variable which *pdf* is $f*$. Then

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

and

$$f_Y(y) = \begin{cases} 2y & 0 < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

and hence $f(x, y) \neq f_X(x)f_Y(y)$. Thus, X and Y are not independent.

Result If X and Y are independent, then $f_{X|Y}(x|y) = f_X(x)$ and $f_{Y|X}(y|x) = f_Y(y)$.

Definition 1.12. Independent and Identically Distributed (iid) Random Variables
A sequence $\{X_n\}$, $n = 1, 2, ...$, of random variables are said to be *iid* random variables with common probability law of X, if $\{X_n\}$ is an independent sequence of random variables and the distribution of X_n, $n = 1, 2, ...$, is the same as the distribution of X.

Let X and Y be independent random variables with *pdf's* $f_X(x)$ and $f_Y(y)$, and distribution functions $F(x)$ and $G(x)$, respectively. Then the *pdf* $f_Z(z)$ of the random variable $Z = X + Y$ is

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(z-\xi)f_Y(\xi)d\xi = \int_{-\infty}^{\infty} f_X(\xi)f_Y(z-\xi)d\xi$$

The distribution function $H(z)$ of $Z = X + Y$ is given by

$$H(z) = \int_{-\infty}^{\infty} f(z-\xi)f_Y(\xi)d\xi = \int_{-\infty}^{\infty} G(z-\xi)f_X(\xi)d\xi.$$

Example A random variable X is said to be uniformly distributed on an interval (a, b) if its *pdf* is given by

$$f_X(x) = \begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & \text{otherwise} \end{cases}$$

Let X and Y be independent and uniformly distributed over $(0, 1)$. Then $f_X(x) = 1, 0 < x < 1$ and $f_Y(y) = 1, 0 < y < 1$. Let $Z = X + Y$. Then $f_Z(z) = \int_0^2 f_X(x)f_Y(z-y)dx$. The possible values of the integrand are 0 and 1. The integrand takes the value 1, if $0 \leq x \leq 1$ and $0 \leq z - x \leq 1$, i.e., $z - 1 \leq x \leq z$. If $0 \leq z \leq 1$, then the integrand has value 1, if $0 \leq x \leq z$ and 0 otherwise.

Therefore, $f_Z(z) = \int_0^x 1dx = z, 0 \leq z \leq 1$. If $1 < z \leq 2$, then the integrand has value 1, if $z - 1 \leq x \leq 1$ and 0 otherwise. Therefore, $f_Z(z) = \int_{z-1}^1 1dx = 2 - z, 1 < z \leq 2$.

Thus

$$f_Z(z) = \begin{cases} z & 0 \leq z \leq 1 \\ 2-z & 1 < z \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

Let $\{X_n\}$ be a sequence of *iid* random variables with distribution function F. Then the distribution function of $X_1+X_2+...+X_m$ is F_{*m},i.e., $P(X_1+X_2+...+X_m \leq x) = F_{*m}$, where F_{*m} is the m-fold convolution of F itself.

Erlang Random Variable A random variable E_n is said to have Erlang distribution with parameters n(positive integer) and $\lambda(> 0)$, if E_n is the sum of n exponentially distirbuted random variables with parameter λ. The *pdf* of E_n is $f(t) = \frac{\lambda^n t^{n-1} e^{-\lambda t}}{(n-1)!}$ and its mean is $\frac{n}{\lambda}$.

Hyper-exponential Random Variable A random variable H_n is said to have Hyper-exponential distribution if $H_n = \alpha_i X\lambda_i$ with probability α_i, such that $\sum_{i=1}^{n} \alpha_i = 1$ and $X\lambda_i$ is an exponentially distirbuted random variable with parameter λ_i, for $i = 1, 2, ..., n$. The *pdf* $f(t)$ of H_n is given by $\alpha_1(\lambda_1 e^{-\lambda_1 t}) + ... + \alpha_1(\lambda_n e^{-\lambda_n t})$ and its mean is given by $\frac{\alpha_1}{\lambda_1} + ... + \frac{\alpha_n}{\lambda_n}$.

Memoryless or Markov Property of the Exponential Distribution Let X be an exponentially distributed random varible with parameter λ. Then,

$$P(X > a+b|X > a) = \frac{P(X > a+b, X > a)}{P(X > a)} = \frac{P(X > a+b)}{P(X > a)} =$$

$$= \frac{e^{-\lambda(a+b)}}{e^{-\lambda a}} = e^{-\lambda b} = P(X > b)$$

If we think of X as the lifetime of some instrument, then the above equation states that the probability that the instrument lives at least $s + t$ hours, given that it has survived t hours, is the same as the initial probability that it lives at least s hours. In other words, if the instrument is alive at time t, then the distribution of its remaining life is the originial lifetime distribution.

Now we give some problems from Probability theory.

1) Find a sharp bound for $P(A \bigcap B)$ and $P(A \bigcup B)$.

2) n numbers are chosen and multiplied. Find the probability, that the last digit of the product is $1, 3, 7, 9$.

3) Let A, B, C be independent uniform (in $(0, 1)$) random variables. Find the probability that the polynomial $Ax^2 + Bx + C = 0$ has real roots.

4) A stick of length a is broken in 3 pieces. What is the probability that they form a triangle?

5) Let X_1 and X_2 are independent r.v.s and $X_1 \epsilon Poi(\lambda_1)$, $X_2 \epsilon Poi(\lambda_2)$. Prove that X_1+X_2 follows Binomial distribution.

6) $X_1, X_2, ..., X_n$ are independent variables and X_i has a cdf $F_i(x)$ for all $i = 1, 2, ...$ Prove that $Y = Max(X_1, ..., X_n)$ has a cdf $\prod_{i=1}^{n} F_i(x)$.

Now we are going to consider briefly 3 of the basic distributions, which we will use in our course.

1) Uniform distribution

$$f(x) = \frac{1}{b-a}, a < x < b,$$
$$F(x) = \frac{x-a}{b-a}, a < x < b.$$

If we take the unitary interval $(0,1)$, then $f(x) = 1$ and $F(x) = x$ and this way we obtain the standard normal distribution $N(0,1)$.

2) Exponential distribution

$$f(x) = \alpha e^{-\alpha x}, 0 < x,$$
$$F(x) = 1 - e^{-\alpha x}, 0 < x.$$

We can easily calculate $E(x) = \frac{1}{\alpha}$ and $Var(x) = \frac{1}{\alpha^2}$. The most improtant property of the Exponential distribution is the so called

Lack of memmory property (Markov property) .

$$P(X > x + y | X > y) = P(X > x)$$

3) Erlang distribution (Gamma distribution)

$$f(x) = \frac{\alpha^n}{(n-1)!} e^{-\alpha x} x^{n-1}, 0 < x.$$

Here we have $E(x) = \frac{n}{\alpha}$ and $Var(x) = \frac{n}{\alpha^2}$.

If $X_1, ..., X_n$ are *iid* exponential variables, than $X_1 + ... + X_n$ follows Erlang distribution.

If $X_1, ..., X_n$ are *iid* standard $N(0,1)$ variables, than $X_1^2 + ... + X_n^2$ follows a Chi-square distribution.

1.4 Stochastic Processes

Stochastic Process

Definition 1.13. A stochastic process $\{X(t), t \in T\}$ is a collection of random variables, defined on a common probability space. For each t of T, $X(t)$ is called *state* of the stochastic process.

Classficication .

It is typical to think of t as time, T as a set of points in time, and $X(t)$ as the value or state of the stochastic process at time t. We classify stochastic processes according to time and say that they are discrete time or continuous time, depending on whether T is discrete (finite or countably infinite) or continuous.

Counting Process

Definition 1.14. A stochastic process is said to be a *Counting Process* if $(X_t)_{t\geq 0}$ represents the total number of events that have occured up to time t. Hence, a counting process $(X_t)_{t\geq 0}$ must satisfy

i) $X(t) \geq 0$;

ii) $X(t)$ is integer-valued;

iii) If $s < t$, then $X(t) \leq X(s)$ for s, t real numbers;

iv) For $s < t$, $X(t) - X(s)$ equals the number of events that have occured in the interval (s, t).

1.5 Markov Chain

Markov Chain

Definition 1.15. A discrete time stochastic process $\{X_n, n = 0, 1, 2...\}$ is said to be a *Markov Chain* if the conditional distribution of any future state X_{n+1}, given the past states X_0, X_1, X_2, ..., X_{n-1} and the present state X_n, is independent of the past states and depends only on the present state. That is

$$P(X_{n+1} = j | X_0 = i_0, X_1 = i_1, ...X_{n-1} = i_{n-1}, X_n = i) =$$
$$= P(X_{n+1} = j | X_n = i),$$

for all states $i_0, i_1, ..., i_{n-1}, i, j$ and all $n \geq 0$.

Stationary or Homogeneous Transition Probability

Definition 1.16. In the above defintion if the transition probability $P(X_{n+1} = j | X_n = i)$ is independent of n then it is called a *stationary* or *homogeneous transition probability* and we will denote this probability by p_{ij}. The value p_{ij} denotes the probability that the process which is in state i will make a transition into state j. Therefore, $p_{ij} \geq 0$, $i, j \geq 0$; $\sum_{j=0}^{\infty} p_{ij} = 1$. The transition matrix P is given by

$$P = ((p_{ij})) = \begin{bmatrix} p_{00} & p_{01} & \cdots & \cdots \\ p_{10} & p_{11} & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{bmatrix}$$

Example 1. Let X_1, X_2, ... X_n are independent identically distributed random variables and let $S_n = X_1 +\ X_2 +\ ... +\ X_n$. S_{n+1}=S_n+X_n depends only on S_n and therefore $\{S_n\}$ is a Markov chain.

Example 2. (Gambler's ruin) There are two players A and B and P(A winning a game) $= p$. Together they have N Euro to play. If A wins, he gets 1 Euro from B and if A loses, he gives 1 Euro to B. We denote

$$X_n = \{\text{the amount, which A has at the end of the } n\text{-th game}\}.$$

If X_n=i, then X_{n+1}=$i+1$ with probability p, and X_{n+1}=$i-1$ with probability q=$1-p$. Then the transition probability matrix is

$$P = ((p_{ij})) = \begin{bmatrix} 1 & 0 & 0 & 0 & \ldots & 0 \\ q & 0 & p & 0 & \ldots & 0 \\ 0 & q & 0 & p & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & q & 0 & p \\ \ldots & \ldots & \ldots & \ldots & \ldots & 1 \end{bmatrix}$$

Example 3. (Branching processes) Let ue have a generation process. Denote

$$X_n = \{\text{number of persons at the } n\text{-th generation}\},$$

$$\xi_n = \{\text{number of offsprings of the } n\text{-th generation}\}.$$

Then X_{n+1}=ξ_1+ξ_2+...+ξ_{X_n} depends only on X_n and therefore $\{X_n\}$ is a Markov chain.

Example 4. (Discrete queue) At every unit of time one customer is served, if available and a random number ξ of customers join the system at one unit of time. Let

$$P(\xi = i) = a_i, i = 0, 1, 2, ...$$

If $X_n = \{$number of waiting customers at the beginning of the n-th period$\}$, then X_{n+1} depends only on X_n. If $\mathrm{X_{n+1}} = \begin{cases} X_n - 1 + \xi & i \geq 1 \\ \xi & i = 0 \end{cases}$. Then the transition probability matrix is

$$P = ((p_{ij})) = \begin{bmatrix} a_0 & a_1 & a_2 & a_3 & \dots \\ a_0 & a_1 & a_2 & a_3 & \dots \\ 0 & a_0 & a_1 & a_2 & \dots \\ 0 & 0 & a_0 & a_1 & \dots \\ \dots & \dots & \dots & \dots & \dots \end{bmatrix}$$

Example 5. There are 6 balls, 3 red and 3 black, distributed in 2 urns equally. At every time instant one ball is taken from each urn and they are interchanged. Let X_n denote the number of red balls at the end of the n-th draw in urn A. Obviously, $\{X_n\}$ is a Markov chain. We will now determine the elements of the transition probability matrix $p_{ij} = P(X_{n+1} = j | X_n = i)$ and j taked values $i-1$, i, $i+1$, where i=0, 1, 2, 3. Easily we obtain the transition probability matrix

$$P = ((p_{ij})) = \begin{bmatrix} 0 & 1 & 0 & 0 \\ \frac{1}{9} & \frac{4}{9} & \frac{4}{9} & 0 \\ 0 & \frac{4}{9} & \frac{4}{9} & \frac{4}{9} \\ 0 & \frac{4}{9} & \frac{4}{9} & \frac{1}{9} \end{bmatrix}$$

Definition 1.17. A transition probability $P = ((p_{ij}))$ is said to be *doubly stochastic*, if the column sum is also 1.

Definition 1.18. State j is said to be *accessible* from state i if for some $n \geq 0$, $p_{ij}^n \geq 0$.

Definition 1.19. Two states i and j are said to *communicate* if i is accesible from j and j is accessible from i, i.e. for some m, $n > 0$, such that $p_{ij}^m \geq 0$, $p_{ji}^n \geq 0$.

Definition 1.20. State i is said to have a *period d* if $p_{ii}^n = 0$ whenever n is not divisible by d, and d is the greatest integer with this property. If $p_{ii}^n = 0$ for all $n > 0$, then the period of i is defined to be infinite.

Definition 1.21. A *state* with period 1 is said to be *aperiodic*.

Definition 1.22. A *chain* is *aperiodic* if all its states are aperiodic.

Definition 1.23. A Markov chain is said to be *irreducible* if all of its states communicate, that is, if there exists an n such that $p_{ij}^n \geq 0$ for every i and j.

Definition 1.24. Define f_{jj}^n as the probability, that a chain starting at state j returns for the first time to j in n transitions. Hence the probability that the chain ever returns to j is

$$f_{jj} = \sum_{n=1}^{\infty} f_{jj}^n.$$

Definition 1.25. If $f_{jj} = 1$, then j is said to be a *recurrent state*.

Definition 1.26. If $f_{jj} < 1$, j is said to be a *transient state.*

Definition 1.27. When $f_{jj} = 1$, m_{jj}=$\sum_{n=1}^{\infty} nf_{jj}^n$ is the *mean recurrence time.*
If $m_{jj} < \infty$, then j is known as a *positive recurrent*, while if m_{jj}=∞, we say that j is a *null recurrent state.*

Definition 1.28. A non-periodic chain which is irreducible and positive recurrent is said to be *ergodic.*

Stationary probabilities

Definition 1.29. Let a Markov chain be irreducible. Then

$$\lim_{n\to\infty} p_{ij}^n = \pi_j, \pi^t P = \pi^t, \pi_0 + \pi_1 + ... = 1.$$

For j=0, 1,..., π_j are called *stationary probabilities.*

1.6 Markov Process

Independent and Stationary Increments

Definition 1.30. We say, that a stochastic process has *independent increments*, if for all $t_1 < t_2 < ... < t_n$ we have that $X(t_2) - X(t_1)$, $X(t_3) - X(t_2)$, ..., $X(t_n) - X(t_{n-1})$ are independent. If further $X(t_i) - X(t_{i-1})$ depends only on $t_i - t_{i-1}$ and not on t_{i-1} or $X(t_{i-1})$, then the process has *stationary independent increments.*

Markov process

Definition 1.31. A process $\{X(t)\}$ is said to be *Markov process*, if

$$P(a < X(t) \leq b | X(t_n) = x_n, X(t_{n-1}) = x_{n-1}, ..., X(t_0) = x_0) =$$

$$= P(a < X(t) \leq b | X(t_n) = x_n)$$

for all n, where $t_0 < t_1 < ... < t_n < t$.

Martingale

Definition 1.32. If $E(X(t_n)) < \infty$ and $E(X(t_n))|X(t_{n-1}), ..., X(t_0)) = X(t_{n-1})$, then $\{X(t)\}$ is a *martingale.*

1.7 Poisson Process

Poisson process

Definition 1.33. A counting process $\{X(t), t \geq 0\}$ is said to be a *Poisson process* with parameter $\lambda(>0)$ if

i) $X(0) = 0$;

ii) The process has stationary and independent increments;

iii) $P(X(h) = 1) = \lambda h + o(h)$ for a time-interval $(t, t+h)$ (then the probability, that no event occurs is $1 - \lambda h + o(h)$);

iv) $P(X(h) \geq 2) = o(h)$(events occur one at a time).

The probability generating function of $\{X(t)\}$ is

$$P(s,t) = E(s^{X(t)}) = P(X(t) = 0) + sP(X(t) = 1) + s^2 P(X(t) = 2) + ...$$

We also have

$$P(s, t+h) = E(s^{X(t+h)}) = E(s^{X(t)})(1 - \lambda h) + o(h) + sE(s^{X(t)})(\lambda h + o(h)) =$$
$$= P(s,t)(1 - \lambda h) + sE(s^{X(t)})(\lambda h + o(h)).$$

Now we get

$$\lim_{h \to 0} \frac{P(s, t+h) - P(s,t)}{h} = -\lambda(1-s)P(s,t)$$

i.e.

$$\frac{\partial P(s,t)}{\partial t} = -\lambda(1-s)P(s,t).$$

We assumed, that $X(0) = 0$ and then $P(s, 0) = 1$. So, the probability generating function has Poisson distribution

$$P(s,t) = e^{-\lambda(1-s)t}.$$

Then $P_n(t) = P(X(t) = n) = e^{-\lambda t}\frac{(\lambda t)^n}{n!}$, for $n = 0, 1, 2, ...$

For a Poisson process $\{X(t)_{t \geq 0}\}$ with parameter $\lambda(>0)$ we get:

1) $E(X(t)) = \lambda t$ and $Var(X(t)) = \lambda t$;

2) Let T_i is the time of occurence of the i-th event. If $t < 0$, then

$$P_0(t) = P(X(t) = 0) = e^{-\lambda t},$$

i.e. the probability, that the first event occurs after time t is $P_0(t) = P(T_1 > t)$ and then $P(T_1 \leq t) = 1 - e^{-\lambda t}$. Therefore T_1 follows an exponential distribution with mean $\frac{1}{\lambda}$. If we shift the origin to T_i, as above, $T_{i+1} - T_i$ follows an exponential distribution with the same mean $\frac{1}{\lambda}$ and $T_1, T_2 - T_1, T_3 - T_2, ...$ are independent identically distributed exponential variables with the same mean $\frac{1}{\lambda}$;

3) $T_i = (T_i - T_{i-1}) + (T_{i-1} - T_{i-2}) + ... + (T_2 - T_1) + T_1$ follows an Erlang distribution with parameters (i, λ);

4) Let $u < t$, $k < n$; for $k = 0, 1, 2, ..., n$ we have

$$P(X(u) = k|X(t) = n) = \frac{P(X(u) = k \cap X(t) = n)}{P(X(t) = n)} =$$

$$= \frac{P(X(u) = k)P(X(t-u) = n-k)}{P(X(t) = n)} = \frac{\frac{e^{-\lambda u}(\lambda u)^k}{k!}\frac{e^{-\lambda(t-u)}(\lambda(t-u))^{n-k}}{(n-k)!}}{\frac{e^{-\lambda t}(\lambda t)^n}{n!}} =$$

$$= \binom{n}{k}\left(\frac{u}{t}\right)^k\left(1 - \frac{u}{t}\right)^{n-k}.$$

$X(u)|X(t)$ follows Binomial distribution, i.e. $B(X(t), \frac{u}{t})$ and $E(X(u)|X(t)) = \frac{X(t)u}{t}$, $E(\frac{X(u)}{u}|\frac{X(t)}{t}) = \frac{X(t)}{t}$.

5) Let $X_1(t)$ and $X_2(t)$ be independent Poisson processes with parameters λ_1 and λ_2 respectively. Knowing, that $E(s^{X_1(t)}) = e^{-\lambda_1(1-s)t}, E(s^{X_2(t)}) = e^{-\lambda_2(1-s)t}$, we get

$$E(s^{X_1(t)+X_2(t)}) = e^{-\lambda_1(1-s)t}e^{-\lambda_2(1-s)t} = e^{-(\lambda_1+\lambda_2)(1-s)t}.$$

Therefore $X_1(t) + X_2(t)$ is also a Poisson process with parameter $\lambda_1 + \lambda_2$. Now let us calculate

$$P(X_1(t) = n_1|X_1(t) + X_2(t) = n) = \frac{P(X_1(t) = n_1 \cap X_1(t) + X_2(t) = n)}{P(X_1(t) + X_2(t) = n)} =$$

$$= \frac{P(X_1(t) = n_1)P(X_2(t) = n - n_1)}{P(X_1(t) + X_2(t) = n)} = \frac{\frac{e^{-\lambda_1 t}(\lambda_1 t)_1^n}{n_1!}\frac{e^{-\lambda_2 t}(\lambda(n-n_1))^{n-k}}{(n-n_1)!}}{\frac{e^{-(\lambda_1+\lambda_2)t}((\lambda_1+\lambda_2)t)^n}{n!}} =$$

$$= \binom{n}{n_1}\left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^{n_1}\left(\frac{\lambda_2}{\lambda_1 + \lambda_2}\right)^{n-n_1}.$$

So, $X_1(t)|X_1(t) + X_2(t)$ follows Binomial distribution, i.e. $B(X_1(t) + X_2(t), \frac{\lambda_1}{\lambda_1+\lambda_2})$ and $E(X_1(t)|X_1(t) + X_2(t)) = \frac{\lambda_1}{\lambda_1+\lambda_2}(X_1(t) + X_2(t))$.

6) Denote by $N(t)$, the number of events up to time t, T_1 is the time of occurrence of the first event. Then

$$P(u < T_1 < u + \triangle\, u|N(t)) = \frac{P(u < T_1 < u + \triangle\, u, N(t) = 1)}{P(N(t) = 1)} =$$

$$= \frac{(e^{-\lambda u}\lambda \triangle\, u e^{-\lambda(t-u)}}{e^{-\lambda t}\lambda t} = \frac{\triangle\, u}{t},$$

where $0 < u < t$. So, given that $N(t) = 1$, T_1 follows uniform distribution in the interval $(0, t)$. Generalizing this, if we have n eventsin $(0, t)$, the probability distribution function of $T_1, T_2, ..., T_n$ is $\frac{n!}{t^n}$.

Example 1. Customers arrive at a supermarket at the rate of $20/hr$(20 customers per hour). Given, that 100 customers have arrived before 10 a.m., find:

a) the probability, that 2 customers will arrive in the next 5 min.;

b) the probability, that the next customer will arrive before 1 min.;

c) the expected number of customers, which will arrive in the next 2 hr.?

Substituting in the formulas, we get:

a) $P(X(5) = 2) = \frac{e^{-\frac{5}{3}}(\frac{5}{3})^2}{2!}$;

b) $P(T_1 \leq 1) = 1 - e^{-\frac{1}{3}}$;

c) E(number of customers arriving in the next 2 hr.)$= \frac{1}{3}.120 = 40$ customers.

1.8 Birth and Death Processes

Here we can consider birth as an arrival, and death as a departure.

Birth and Death Processes (BDP) and Kolmogorov Equations

Definition 1.34. Let $\{X(t)\}$ be a Markov process with a parameter $\lambda \geq 0$ and with stationary transition probabilities, i.e. $P(X(u) = m|X(t) = n)$ depends only on $t - u$ and does not depend on $X(u)$ and u, where $m, n \geq 0$, and $0 < u < t$. Let us denote

$$P_{ij}(h) = P(X(t+h) = j|X(t) = i)$$

and assume:

1) $P_{ii+1}(h) = \lambda_i h + o(h)$ (birth);

2) $P_{ii-1}(h) = \mu_i h + o(h)$ (death);

3) $P_{ii}(h) = 1 - (\lambda_i + \mu_i)h + o(h)$ (no birth and no death occure);

4) P(more than one birth and/or death occurs) $= o(h)$;

5) $\lambda_0 \geq 0, \lambda_1, \lambda_2, ... > 0$, $\mu_0 = 0, \mu_1, \mu_2, ... > 0$.

Then $\{X(t)\}$ is called a Birth and Death Process (BDP).

It is easy to obtain, that

$$P(X(t+h)=n)=P(X(t+h)=n|X(t)=n+1)P(X(t)=n+1)$$
$$+P(X(t+h)=n|X(t)=n-1)P(X(t)=n-1)$$
$$+P(X(t+h)=n|X(t)=n)P(X(t)=n)+o(h)$$

where $\lim_{h\to\infty}\frac{o(h)}{h}=0$.

Noting $P_n(t)=P(X(t)=n)$, we can write the above expression this way:

$$P_n(t+h)=$$

$$=P_{n+1}(t)(\mu_{n+1}h+o(h))+P_{n-1}(t)(\lambda_{n-1}h+o(h))$$
$$+P_n(t)(1-\lambda_n h-\mu_n h+o(h))+o(h).$$

We get

$$\begin{cases} P_n'(t)=\mu_{n+1}P_{n+1}(t)-(\lambda_n+\mu_n)P_n(t)+\lambda_{n-1}P_{n-1}(t), n=1,2,... \\ P_0'(t)=\mu_1 P_1(t)-\lambda_0 P_0(t). \end{cases}$$

called Kolmogorov equations.

Special cases:

1) Poisson process

2) Pure birth process

3) Pure death process

4) Pure(simple) birth and death process

5) $M/M/1$ queue

6) $M/M/1$ queue with a busy period

7) $M/M/\infty$ queue

Solving the Kolmogorov equations, we obtain the time-dependent probabilities $P_n(t)$. First we study time-independent steady state probabilities, taking $P_n'(t)=0$.

Now from the Kolmogorov equations we have

$$\begin{cases} \mu_{n+1}P_{n+1}-\lambda_n P_n=\mu_n P_n-\lambda_{n-1}P_{n-1} \\ \mu_1 P_1-\lambda_0 P_0=0 \end{cases}$$

Iterating here, we obtain:

$$\mu_n P_n - \lambda_{n-1} P_{n-1} = \mu_{n-1} P_{n-1} - \lambda_{n-2} P_{n-2} = ... = \mu_1 P_1 - \lambda_0 P_0 = 0.$$

Therefore

$$\mu_n P_n = \lambda_{n-1} P_{n-1}, P_n = \frac{\lambda_{n-1}}{\mu_n} P_{n-1} = ... = \frac{\lambda_0 \lambda_1 ... \lambda_{n-1}}{\mu_1 \mu_2 ... \mu_n} P_0.$$

But we know, that $P_0 + P_1 + ... = 1$, and then

$$P_0 (1 + \frac{\lambda_0}{\mu_1} + \frac{\lambda_0 \lambda_1}{\mu_1 \mu_2} + ...) = 1, P_0 = \frac{1}{1 + \sum_{n=1 \to \infty} \frac{\lambda_0 \lambda_1 ... \lambda_{n-1}}{\mu_1 \mu_2 ... \mu_n}}.$$

It is obvious, that this process converges when $\sum_{n=1}^{\infty} \frac{\lambda_0 \lambda_1 ... \lambda_{n-1}}{\mu_1 \mu_2 ... \mu_n} < \infty$.

Chapter 2

Birth and Death Queueing Models

2.1 Model $M/M/1$

For this queueing model there is only one server and the arrival and the service times are exponentially distributed with means $\frac{1}{\lambda}$ and $\frac{1}{\mu}$ correspondingly. So, we have

$$\lambda_n = \lambda, n = 0, 1, 2, ...\text{and } \mu_n = \mu, n = 1, 2, 3, ...$$

The steady-state probabilities $P_0, P_1, ...$ are calculated as follows:

$$P_n = \frac{\lambda_0\lambda_1...\lambda_{n-1}}{\mu_1\mu_2...\mu_n}P_0 = \frac{\lambda^n}{\mu^n}P_0 = \rho^n P_0,$$

where $\rho = \frac{\lambda}{\mu}$ is the traffic intensity and P_0 we find this way:

$$P_0 = \frac{1}{1 + \sum_{j=1}^{\infty} \frac{\lambda_0\lambda_1...\lambda_{j-1}}{\mu_1\mu_2...\mu j}\rho^j} = \frac{1}{1 + \rho + \rho^2 + ...} = 1 - \rho,$$

where $\rho < 1$, i.e. $\lambda < \mu$. Therefore, $P_n = (1 - \rho)\rho^n, n = 0, 1, 2, ...$

Let us denote

$$N_s = \text{system size (the number of the people in the whole system)},$$

$$N_q = \text{queue size (the number of the people in the queue)},$$

$$T_s = \text{total time spent by a customer in the system},$$

$$T_s = \text{waiting time spent by a customer in the queue}.$$

We already know, that $P_n = P(N_s = n) = (1-\rho)\rho^n, n = 1, 2, ...$ and $\sum_{n=0}^{\infty} P_n = 1$. Then N_s follows a geometric distribution. The average number of customers, waiting in the system is given by:

$$E(N_s) = L_s = \sum_{n=1}^{\infty} nP_n = (1 - \rho)\rho \sum_{n=1}^{\infty} n\rho^{n-1} = \frac{(1 - \rho)\rho}{(1 - \rho)^2} = \frac{\rho}{1 - \rho}$$

We observe, that

$$P(\text{a person has to wait}) = 1 - P_0 = 1 - (1-\rho) = \rho.$$

$$E(N_s) = E(N_s|N_s > 0)P(N_s > 0) + E(N_s|N_s = 0)P(N_s = 0)$$

$$P(N_s|N_s > 0) = \frac{E(N_s)}{P(N_s > 0)} = \frac{\rho/(1-\rho)}{\rho} = \frac{1}{1-\rho}$$

$$P(N_s \geq n) = (\rho^n + \rho^{n+1} + ...)(1-\rho) = \frac{\rho^n}{1-\rho}(1-\rho) = \rho^n, n = 1, 2, ...$$

Let us calculate

$$E(N_s(N_s-1)) = \sum_{n=2}^{\infty} n(n-1)P_n = (1-\rho)\rho^2 \sum_{n=2}^{\infty} n(n-1)\rho^{n-2} =$$

$$= \frac{(1-\rho)\rho^2 2}{(1-\rho)^3} = \frac{2\rho^2}{(1-\rho)^2}.$$

Now we can find the variance of the system's size

$$Var(N_s) = E(N_s(N_s-1)) + E(N_s) - E(N_s)^2 =$$

$$= \frac{2\rho^2}{(1-\rho)^2} + \frac{\rho}{1-\rho} - \frac{\rho^2}{(1-\rho)^2} = \frac{\rho}{(1-\rho)^2}.$$

The probability, that there is noone in the queue is given by

$$P(N_q = 0) = P(N_s = 0) + P(N_s = 1) = 1 - \rho + \rho(1-\rho) = 1 - \rho^2.$$

The expected number of customers in the queue is

$$E(N_q) = L_q = \sum_{n=1}^{\infty}(n-1)P_n = \sum_{m=0}^{\infty} mP_{m+1} = \sum_{m=0}^{\infty} m\rho^{m+1}(1-\rho) = \frac{\rho^2}{1-\rho}.$$

Hence, knowing L_s and L_q we obtain the well-famous

Little's formula

$$L_s = L_q + q.$$

i.e.

$$L_s = L_q + 1 - P_0.$$

For the *pdf* of T_s we get

$$(\text{pdf of } T_s) = \sum_{n=0}^{\infty}(\text{pdf of} T_s|N_s = n)P(N_s = n) =$$

$$= \sum_{n=0}^{\infty} (\text{pdf of } \xi_1 + ... + \xi_{n+1}) \rho^n (1-\rho) = \sum_{n=0}^{\infty} \mu^{n+1} \frac{e^{-\mu x}}{n!} x^{n+1-1} \rho^n (1-\rho) =$$

$$= e^{-\mu x} \mu (1-\rho) \sum_{n=0}^{\infty} \frac{(\mu x \rho)^n}{n!} = \mu(1-\rho) e^{-\mu(1-\rho)x}, x > 0,$$

where $\xi_1, ..., \xi_{n+1}$ are *iid* exponential variables with mean $1/\mu$. Then the *pdf* of the total time spent in the system by a customer follows an expopential distribution with mean $\frac{1}{\mu(1-\rho)}$.

The expected waiting time in the system is

$$W_s = E(T_s) = \frac{1}{\mu(1-\rho)}$$

and then

$$L_s = \lambda W_s.$$

For $T_s = T_q + \xi$, where ξ is the service time, applying to the both sides of this equation the operator E(expectancy), we obtain

$$W_s = W_q + \frac{1}{\mu}, W_q = \frac{1}{\mu(1-\rho)} - \frac{1}{\mu} = \frac{\rho}{\mu(1-\rho)}$$

Example 1. For $\lambda_0 = 1, \lambda_1 = 2, \lambda_2 = 5$ and $\mu_1 = 3, \mu_1 = 1, \mu_2 = 2$, we calculate

$$P_1 = \frac{1}{3} P_0, P_2 = \frac{1}{3} . \frac{2}{1} P_0, \frac{1}{3} . \frac{2}{1} . \frac{5}{2} P_0,$$

$$P_0 \Big(1 + \frac{1}{3} + \frac{1}{2} + \frac{2}{3} + \frac{5}{3} \Big), P_0 = \frac{1}{\frac{11}{3}} = \frac{3}{11}.$$

Therefore

$$P_1 = \frac{1}{11}, P_2 = \frac{2}{11}, P_3 = \frac{5}{11}.$$

Here the mean is $1 \frac{1}{11} + 2 \frac{2}{11} + 3 \frac{5}{11} = \frac{20}{11}$, and the variance

$$1^2 \frac{1}{11} + 2^2 \frac{2}{11} + 3^2 \frac{5}{11} - (\frac{20}{11})^2.$$

Example 2. Let us take $\lambda_0 = \lambda_1 = ... = 1, \mu_1 = 1.3, \mu_2 = 2.3, ...$ Then

$$P_n = \frac{(1/3)^n / n!}{1 + \sum_{n=1}^{\infty} \frac{(1/3)^n}{n!}} = e^{-\frac{1}{3}} \frac{(1/3)^n}{n!}.$$

Example 3. Arrivals come at a telephone booth in a Piosson process with mean inter-arrival time 10 min. The convesation time is exponential with mean 3 min. Actually, we have that $\lambda = 1/10, \mu = 1/3$ and then $\rho = 3/10 = 0.3$. Now we find, that

a) P(an arrival has to wait)$= 1 - (1 - \rho) = \rho = 0.3$;

b) The average non empty queue length $= \frac{1}{1-\rho} = \frac{10}{7}$;

c) If a customer has to wait for at least 3 min, another booth will be installed. Then

$$W_q = \frac{L_q}{\lambda} = \frac{\rho'^2}{(1-\rho')\lambda'} = \frac{\lambda'}{\mu^2(\mu - \lambda')} = 3, \rho' = \frac{\lambda'}{\mu}.$$

Then $\lambda' = 0.16/min = 9.6/hr$.

d)

$$P(T_q > 10) = \int_{10}^{\infty} \lambda(1-\rho)e^{-\mu x(1-\rho)}dx = 0.03;$$

e) $P(T_s > 10) = \int_{10}^{\infty} \mu(1-\rho)e^{-\mu x(1-\rho)}dx = 0.01$;

f) The fraction of time the system is in use $= 1 - P_0 = \rho = 0.3$.

Example 4. There are 2 repairmen - slow and fast. The machines break down at the rate of $\lambda = 3$/hr. The nonproductive time of one machine costs 50 euro/hr. The service rates for the 2 repairmen are $\mu_s = 4/hr$ and $\mu_f = 6$/hr correspondingly, the costs are $c_s = 30$ euro/hr and $c_f = 50$ euro/hr. Which repairman is preferred?

For the repairmen we obtain $\rho_s = \frac{3}{4}$, $\rho_f = \frac{1}{2}$. For the slow repairman we get: $L_{s,s} = \frac{\rho_s}{1-\rho_s} = 3$. For the fast repairman we get: $L_{s,f} = \frac{\rho_f}{1-\rho_f} = 1$. Let us assume that the working day is 8 hr. Then the total cost for the slow repairmen is $8.3.50 + 30.8 = 1440$ euro. The total cost for the fast repairman is $8.1.50 + 50.8 = 800$ euro. Hence, we prefer the fast repairman.

2.2 Model $M/M/1/N$

This queueing model has the capacity of N customers including the one in the service facility. If an arriving customer finds the system full, he does not wait and leaves the system. The arrival and the service rate for this queueing model are given by

$$\lambda_n = \lambda, n = 0, 1, ..., N-1, \mu_n = \mu, n = 1, 2, ..., N.$$

Here again $\rho = \frac{\lambda}{\mu}$, but ρ need to be less than 1.

$$P_n = \rho^n P_0, n = 1, 2, ..., N.$$

$$P_0 + P_1 + ... + P_N = 1, P_0(1 + \rho + \rho^2 + ... + \rho^N) = 1,$$

$$P_0 \frac{1-\rho^{N+1}}{1-\rho} = 1, P_0 = \frac{1-\rho}{1-\rho^{N+1}} (= \frac{1}{N+1} \text{when } \rho = 1).$$

Remark. If $\rho < 1$ and $N \to \infty$, then $P_0 = 1 - \rho$.

$$P_N = \rho^N \frac{1-\rho}{1-\rho^{N+1}} (= \frac{1}{N+1} \text{when} \rho = 1) = P(\text{the system is full})$$

$$1 - P_N = P(\text{the system is not full})$$

Performance Measures

$$L_s = \sum_{n=1}^{N} n P_n = (1-\rho) \sum_{n=1}^{N} n \rho^n = \frac{\rho(1-\rho^n)}{1-\rho} - N\rho^{N+1}.$$

The effective arrival rate λ' is the mean rate of the customers actually entering the system (because the system is finite and the arriving customers who find the system full leave the system).

This is given by

$$\lambda' = \sum_{n=0}^{N-1} \lambda_n P_n = \lambda \sum_{n=0}^{N-1} P_n = \lambda(1 - P_N) = \lambda_{eff}$$

The average waiting times in the system W_s and in the queue W_q can be calculated using Little's formula and λ', i.e.

$$W_s = \frac{L_s}{\lambda'}, W_q = \frac{L_q}{\lambda'}.$$

The blocking probability is the probability that customers are blocked and not accepted by the queueing system because the system's capacity is full. This situation occurs in queueing systems that have a finite or no waiting queue. In our model it happens when there are N customers in the system and hence the blocking probability is P_N.

Example. Cars arrive at a car-wash facility with a rate 5 cars per hour. The washing time is exponential with mean 10 min. Then $\lambda = 5/hr, \mu = 6/hr, \rho = 5/6$. We obtain

$$L_s = \frac{\rho}{1-\rho} = \frac{5/6}{1-5/6} = 5; L_q = \frac{\rho^2}{1-\rho} = \frac{25}{6} = 4.17.$$

2.3 Model $M/M/1$ with Finite Source

One person is in charge with k machines. The operating time has rate λ and the repairtime has rate μ. Let X denote the number of machines with the repairman. If $X = n$, $k - n$ are operating, so that, the arrival rate is $\lambda(k - n)$. We have, that

$$\mu_n = \mu, n = 1, 2, ..., k, \text{because we have only one person} \lambda_n = \lambda(k-n), n = 0, 1, ..., k-1$$

$$P_n = \frac{\lambda_0 \lambda_1 ... \lambda_{n-1}}{\mu_1 \mu_2 ... \mu_n} P_0 = k(k-1)...(k-n+1)\rho^n P_0 = k_{(n)} \rho^n P_0$$

$$P_0 = \frac{1}{1 + k\rho + k(k-1)\rho^2 + ... + k(k-1)...(k-n+1)\rho^n}$$

$$L_s = \sum_{n=1}^{k} n P_n = \sum_{n=1}^{k} (k-(k-n))P_n = k(1-P_0) - \sum_{n=1}^{k}(k-n)k(k-1)...(k-n+1)\rho^n P_0 =$$

$$= k(1-P_0) - \frac{1}{\rho}(1-P_0-P_1) = k(1-P_0) - \frac{1}{\rho}(1-P_0-k\rho P_0) = k - \frac{1-P_0}{\rho}$$

$$\lambda_{eff} = \sum_{n=0}^{k} \lambda_n P_n = \lambda \sum_{n=0} k(k-n)P_n = \lambda(k-L_s)$$

Example. One person is in charge with 3 machnines. It is given, that λ=2/day, μ=4/day. Then we obtain:

$$\rho = 1/2, P_1 = \frac{3}{2}P_0, P_2 = 3.2.\frac{1}{4}P_0, P_3 = 3.2.1.\frac{1}{8}P_0$$

$$P_0(1 + \frac{3}{2} + \frac{3}{2} + \frac{3}{4}) = 1$$

$$P_0 = \frac{4}{19}, P_1 = \frac{6}{19}, P_2 = \frac{6}{19}, P_3 = \frac{3}{19}$$

$$L_s = \frac{27}{19}, \lambda_{eff} = \frac{60}{19}, W_s = \frac{9}{20}$$

$$L_q = L_s - (1-P_0) = \frac{12}{19}, W_q = \frac{1}{5}$$

2.4 Model $M/M/c$

In this model the arrivals form a Poisson process (the interarrival times are exponential) and the service time follows an exponential distribution. There are c servers. So, we have:

$$\lambda_n = \lambda, n = 0, 1, 2, ...$$

$$\mu_n = \begin{cases} n\mu & \text{if } n = 0, 1, ..., c-1 \\ c\mu & \text{if } n = c, c+1, ... \end{cases}$$

and it is easy to obtain

$$P_n = \begin{cases} \frac{\rho^n}{n!} P_0 & n = 0, 1, ..., c-1 \\ \frac{\rho^n}{c! c^{n-c}} P_0 & n = c, c+1, ... \end{cases}$$

$$P_0 = \frac{1}{1 + \frac{\rho}{1!} + \frac{\rho^2}{2!} + ... + \frac{\rho^{c-1}}{(c-1)!} + \frac{\rho^c/c!}{1-\rho/c}}$$

$$P(\text{a person has to wait}) = P(\text{number of customers} \geq c) = \frac{\rho^c/c!P_0}{1-\rho/c},$$

$$L_q = \sum_{n=c}^{\infty}(n-c)P_n = \sum_{n=c}^{\infty}(n-c)\frac{\rho^{n-c+c}}{c!c^{n-c}}P_0 = \frac{\rho^c}{c!}P_0\sum_{m=0}^{\infty}m(\rho/c)^m = \frac{\rho^c}{c!}\frac{\rho/c}{1-(\rho/c)^2}P_0$$

If we put $c = 2$, then $P_1 = \rho P_0, P_2 = \frac{\rho^2}{2!}P_0$ and

$$P(\text{a person has to wait}) = \frac{\rho^2/2!P_0}{1-\rho/2}, P_0 = \frac{2-\rho}{2+\rho}, L_q = \frac{\rho^3}{4-\rho^2}, \frac{\rho}{2} \leq 1.$$

2.5 Model $M/M/c/N$ (Loss and Delay System)

This model is finite.

$$\lambda_n = \lambda, n = 0, 1, ..., N-1$$

$$\mu_n = \begin{cases} n\mu & \text{if } n = 0, 1, ..., c-1 \\ c\mu & \text{if } n = c, c+1, ..., N \end{cases}$$

We consider the cases:

$$P_0 = \begin{cases} \left(\sum_{n=0}^{c-1}\frac{\rho^n}{n!} + \frac{\rho^c(1-\rho/c)^{N-c+1}}{c!(1-\rho/c)}\right)^{-1} & \rho/c \neq 1 \\ \left(\sum_{n=0}^{c-1}\frac{\rho^n}{n!} + \frac{\rho^c N-c+1}{c!}\right)^{-1} & \rho/c = 1 \end{cases}$$

$$\lambda_{eff} = \lambda(1-P_N), L_s = L_q + \frac{\lambda_{eff}}{\mu}$$

2.6 Model $M/M/c$ with a Limited Source k (Repairman Problem)

Here

$$\lambda_n = (k-n)\lambda, n = 0, 1, ..., k-1$$

$$\mu_{\text{n}} = \begin{cases} n\mu & \text{if } n = 0, 1, ..., c-1 \\ c\mu & \text{if } n = c, c+1, ... \end{cases}$$

$$P_n = \begin{cases} \binom{k}{n}\rho^n P_0 & n = 0, 1, ..., c-1 \\ \binom{k}{n}\frac{\rho^n n!}{c!c^{n-c}}P_0 & n = c, c+1, ... \end{cases}$$

2.7 Model $M/M/c/c$ (Loss System)

In comparisson to Model 5, here $c < N$ and there is no queue ($L_q = 0$). For this model

$$\lambda_n = \lambda, n = 0, 1, ..., c-1; \mu_n = n\mu, n = 1, 2, ..., c$$

$$P_n = \frac{\rho^n}{n!} P_0, n = 0, 1, ..., c \text{ this is a finite model}$$

$$P_0 = \frac{1}{1 + \frac{\rho}{1!} + \frac{\rho^2}{2!} + ... + \frac{\rho^c}{c!}}$$

$$\lambda_{eff} = \lambda(1 - P_0); L_q = 0, W_q = 0, W_s = 1/\mu.$$

Actually, above we obtained the Erlang's loss formula.

2.8 Model $M/M/\infty$ (Infinite Server Queue)

$$\lambda_n = \lambda, n = 0, 1, ...; \mu_n = n\mu, n = 1, 2, ...$$

$$P_n = \frac{\rho^n}{n!} P_0, n = 0, 1, ...$$

$$P_0 = e^{-\rho} = \frac{1}{1 + \frac{\rho}{1!} + \frac{\rho^2}{2!} + ...} \text{ and then} P_n = \frac{e^{-\rho}\rho^n}{n!}, L_s = \rho$$

2.9 Self Repair

There are N nodes either in 'on', or in 'off' position. If a node is in 'on' position, it receives messages, otherwise it sends messages. The receiving rate is λ and the sending rate is μ.

$$\lambda_n = (N - n)\lambda; \mu_n = n\mu$$

$$P_n = \frac{\lambda_0 \lambda_1 ... \lambda_{n-1}}{\mu_1 \mu_2 ... \mu_n} P_0 = \binom{N}{n} \rho^n P_0.$$

Therefore

$$P_0 = \frac{1}{(1+\rho)^N}, P_n = \binom{N}{n} \left(\frac{\rho}{1+\rho}\right)^n \left(\frac{1}{1+\rho}\right)^{N-n}, L_s = \frac{N\rho}{1+\rho}$$

2.10 Impatient Customers (Limited Waiting Time)

A customer joins an $M/M/c$ queue, but after a random time following an exponential distribution with parameter ν he loses patience and leaves the system without getting service, if his service has not still started by then.

$$\lambda_n = \lambda, \mu_n = \begin{cases} n\mu & \text{if } n = 1, ..., c \\ c\mu + (n-c)\nu & \text{if } n = c+1, ... \end{cases}$$

The probabilities P_n are to be found numerically.

2.11 Density-Dependent Service

$$\lambda_n = \lambda, n = 0, 1, ...; \mu_n = n^\alpha \mu, n = 1, 2, ...$$

Here $\alpha > 0$ is a pressure coefficient.

Example. The mean time a doctor spends with a patient is 24 min, if no other patient is waiting. This becomes 12 min, if he has other patients. Then

$$\mu_1 = 24, \mu_6 = 12 = 6^\alpha \mu_1, 6^\alpha = 2, \alpha = 0.4,$$

$$P_n = \frac{1}{(n!)^\alpha}\rho^n P_0, P_0 = \frac{1}{1 + \frac{1}{(1!)^\alpha} + \frac{1}{(2!)^\alpha} + ...}.$$

Chapter 3

Time-Dependent Analysis

3.1 Model $M/M/1/1$

The Kolmogorov equations for this model are

$$\begin{cases} P_0'(t) = \mu P_1 - \lambda P_0 \\ P_1'(t) = \lambda P_0 - \mu P_1 \end{cases}$$

with initial conditions

$$P_0(0) = a, P_1(0) = b, a + b = 1.$$

3.1.1 Method 1

Here we calculate

$$P_0'(t) = \mu(1 - P_0) - \lambda P_0$$

$$P_0'(t) + (\lambda + \mu)P_0 = \mu$$

$$(e^{(\lambda+\mu)tP_0})' = \mu e^{(\lambda+\mu)t}$$

$$e^{(\lambda+\mu)t}P_0 = \frac{\mu}{\lambda + \mu}(e^{(\lambda+\mu)t} - 1) + c$$

and then $a = c$.

$$P_0(t) = \frac{\mu}{\lambda + \mu} + \left(a - \frac{\mu}{\lambda + \mu}\right)e^{-(\lambda+\mu)t}$$

$$P_1(t) = \frac{\lambda}{\lambda + \mu} - \left(a - \frac{\mu}{\lambda + \mu}\right)e^{-(\lambda+\mu)t} = \frac{\lambda}{\lambda + \mu} + \left(\nu - \frac{\lambda}{\lambda + \mu}\right)e^{-(\lambda+\mu)t}$$

If $a = \frac{\mu}{\lambda+\mu}$, then $P_0(t) = \frac{\mu}{\lambda+\mu}$ and it is independent of t. Also $P_0(t) \to \frac{\mu}{\lambda+\mu}$, as $t \to \infty$.

3.1.2 Method 2

For this method we take the Laplace transform

$$\begin{cases} s\hat{P}_0 - a = \mu\hat{P}_1 - \lambda\hat{P}_0 \\ s\hat{P}_1 - b = \lambda\hat{P}_0 - \mu\hat{P}_1 \end{cases}$$

$$\begin{cases} (s+\lambda)\hat{P}_0 - \mu\hat{P}_1 = a \\ -\lambda\hat{P}_0 + (s+\mu)\hat{P}_1 = b \end{cases}$$

From this system of two equations, we obtain

$$\begin{pmatrix} s+\lambda & -\mu \\ -\lambda & s+\mu \end{pmatrix} \begin{pmatrix} \hat{P}_0 \\ \hat{P}_1 \end{pmatrix} = \begin{pmatrix} a \\ b \end{pmatrix}$$

$$\hat{P}_0 = \frac{\begin{vmatrix} a & -\mu \\ b & s+\lambda \end{vmatrix}}{\begin{vmatrix} s+\lambda & -\mu \\ -\lambda & s+\mu \end{vmatrix}} = \frac{A}{s} + \frac{B}{s+\lambda+\mu},$$

$$A = \frac{\begin{vmatrix} a & -\mu \\ c & \mu \end{vmatrix}}{\lambda+\mu} = \frac{\mu(a+b)}{\lambda+\mu} = \frac{\mu}{\lambda+\mu},$$

$$B = \frac{\begin{vmatrix} a & -\mu \\ c & -\lambda \end{vmatrix}}{-(\lambda+\mu)} = \frac{a\lambda - b\mu}{\lambda+\mu} = a - \frac{\mu}{\lambda+\mu}.$$

Inverting, we get the solution

1)

$$\begin{vmatrix} s+\lambda_0 & \lambda & \dots & \dots & \dots & \dots \\ \mu_1 & s+\lambda_1+\mu_1 & \lambda_1 & \dots & \dots & \dots \\ \dots & \mu_2 & s+\lambda_2+\mu_2 & \lambda_2 & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & s+\mu_N \end{vmatrix} =$$

$$= s \begin{vmatrix} s+\lambda_0+\mu_1 & \lambda_1 & \dots & \dots & \dots & \dots \\ \mu_1 & s+\lambda_1+\mu_2 & \lambda_2 & \dots & \dots & \dots \\ \dots & \mu_2 & s+\lambda_2+\mu_3 & \lambda_3 & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \mu_{N-1} & s+\lambda_{N-1}+\mu_N \end{vmatrix};$$

2)

$$\begin{vmatrix} 2cos\theta & 1 & ... & ... & ... & ... \\ 1 & 2cos\theta & 1 & ... & ... & \\ 0 & 1 & 2cos\theta & 1 & ... & ... \\ ... & ... & ... & ... & ... & ... \\ ... & ... & ... & ... & 1 & 2cos\theta \end{vmatrix} = \frac{\sin(N+1)\theta}{\sin\theta} = \prod_{k=1}^{N}(2\cos\theta - 2\cos\frac{k\theta}{N+1}).$$

3.2 Model $M/M/1/N$

The Kolmogorov equations for this model are:

$$\begin{cases} P_0'(t) = \mu P_1 - \lambda P_0 \\ P_1'(t) = \mu P_2 - (\lambda+\mu)P_1 + \lambda P_0 \\ ... \\ P_N'(t) = \lambda P_{N-1} - \mu P_N \end{cases}$$

Assume, that $P_0(0) = a_0$, $P_1(0) = a_1$, ..., $P_N(0) = a_N$, $\sum_{k=1}^{N} a_k = 1$. Here again we apply the Laplace transform

$$\begin{cases} s\hat{P_0(t)} - a_0 = \mu\hat{P_1} - \lambda\hat{P_0} \\ s\hat{P_1(t)} - a_1 = \mu\hat{P_2} - (\lambda+\mu)\hat{P_1} + \lambda\hat{P_0} \\ ... \\ s\hat{P_N(t)} - a_N = \lambda\hat{P_{N-1}} - \mu\hat{P_N} \end{cases}$$

and then the matrix equation is

$$\begin{pmatrix} s+\lambda & -\mu & ... & ... & ... \\ -\lambda & s+\lambda+\mu & \mu & ... & ... \\ ... & ... & ... & ... & ... \\ ... & ... & ... & \lambda & s+\mu \end{pmatrix} \begin{pmatrix} \hat{P_0(t)} \\ \hat{P_1(t)} \\ ... \\ \hat{P_N(t)} \end{pmatrix} = \begin{pmatrix} \hat{a_0(t)} \\ \hat{a_1(t)} \\ ... \\ \hat{a_N(t)} \end{pmatrix}.$$

Using the Cramer's rule, we obtain

$$P_0 = \frac{\begin{vmatrix} a_0 & \mu & ... & ... & ... \\ a_1 & s+\lambda+\mu & \mu & ... & ... \\ ... & ... & ... & ... & ... \\ a_N & ... & ... & \lambda & s+\mu \end{vmatrix}}{\begin{vmatrix} s+\lambda & \mu & ... & ... & ... \\ \lambda & s+\lambda+\mu & \mu & ... & ... \\ ... & ... & ... & ... & ... \\ ... & ... & ... & \lambda & s+\mu \end{vmatrix}} = \frac{\begin{vmatrix} a_0 & \mu & ... & ... & ... \\ a_1 & s+\lambda+\mu & \mu & ... & ... \\ ... & ... & ... & ... & ... \\ a_N & ... & ... & \lambda & s+\mu \end{vmatrix}}{s\prod_{k=1}^{N}(s+\lambda+\mu-2\sqrt{\lambda\mu}\cos\frac{k\pi}{N+1})} = \sum_{k=1}^{N} H_k e^{-s_k t}$$

where $s_0 = 0$, $s_j = \lambda + \mu - 2\sqrt{\lambda\mu}\frac{k\pi}{N+1}$. It can be proved by induction, that

$$\begin{vmatrix} \lambda_0 + \mu_1 & \mu_1 & ... & ... & ... \\ \lambda_1 & \lambda_1 + \mu_2 & \mu_2 & ... & ... \\ ... & ... & ... & ... & ... \\ ... & ... & ... & ... & \lambda_{N-1} + \mu_N \end{vmatrix} = \lambda_0\lambda_1...\lambda_{N-1} + \lambda_0...\lambda_{N-2}\mu_N + ... + \mu_1\mu_2...\mu_N$$

and

$$\begin{vmatrix} \lambda_1 + \mu_1 & \lambda_1 & ... & ... & ... \\ \mu_2 & \lambda_2 + \mu_2 & \lambda_2 & ... & ... \\ ... & ... & ... & ... & ... \\ ... & ... & ... & \mu_N & \mu_N \end{vmatrix} = \mu_1\mu_2...\mu_N$$

Thus,

$$P_0 = \lim_{s \to 0} s\hat{P}_0 = \frac{1}{1 + \frac{\lambda_0}{\mu_1} + \frac{\lambda_0\lambda_1}{\mu_1\mu_2} + ... + \frac{\lambda_0\lambda_1...\lambda_{N-1}}{\mu_1\mu_2...\mu_N}}.$$

3.3 State Dependent Finite Model

Here

$$\lambda_n, n = 0, 1, ..., N-1, \mu_n, n = 1, 2, ..., N$$

$$\begin{cases} P_0'(t) = \mu_1 P_1 - \lambda_0 P_0 \\ P_n'(t) = \mu_{n+1} P_{n+1} - (\lambda_n + \mu_n) P_n + \lambda_{n-1} P_{n-1}, n = 1, 2, ..., N-1 \\ P_N'(t) = \lambda_{N-1} P_{N-1} - \mu_N P_N \end{cases}$$

Assume, that $P(X(0) = n) = P_n(0) = a_n$, $\sum_{k=1}^{N} a_k = 1$. The above system could be written in the form

$$P'(t) = AP(t), P(0) = \begin{pmatrix} a_0 \\ a_1 \\ ... \\ a_N \end{pmatrix} \text{ and then } P(t) = e^{At} P(0).$$

Now we apply the Laplace trnasform

$$\begin{cases} s\hat{P_0(t)} - a_0 = \mu_1 \hat{P}_1 - \lambda_0 \hat{P}_0 \\ s\hat{P_n(t)} - a_n = \mu_{n+1} \hat{P_{n+1}} - (\lambda_n + \mu_n) \hat{P}_n + \lambda_{n-1} \hat{P_{n-1}} \\ s\hat{P_N(t)} - a_N = \lambda_{N-1} \hat{P_{N-1}} - \mu_N \hat{P}_N \end{cases}.$$

From here we obtain

$$\begin{cases} (s + \lambda_0)\hat{P}_0 - \mu_1 \hat{P}_1 = a_0 \\ (s + \lambda_n + \mu_n)\hat{P}_n - \lambda_{n-1} \hat{P_{n-1}} - \mu_{n+1} \hat{P_{n+1}} = a_n \\ (s + \mu_N)\hat{P_N} - \lambda_{N-1} \hat{P_{N-1}} = a_N \end{cases}$$

Again using the Cramer's rule, we can find $\hat{P}_n(s)$ for all n, replacing the n-th column in the determinant of the system by the column with the free coefficients. Simplifiing this expression, we get

$$\hat{P}_i(s) = \frac{...}{s(s-1)...(s-s_N)}$$

and therefore

$$P_n(t) = \sum_{j=0}^{N} H_j e^{s_j t} (\text{by splitting into partial fractions}),$$

where

$$0 = s_0 > s_1 > ... > s_N,$$

$$P_i(t) \to H_i \text{ as } \to \infty.$$

The following equality holds

$$H_i(0) = \frac{\frac{\lambda_0 \lambda_1 ... \lambda_{i-1}}{\mu_1 \mu_2 ... \mu_i}}{1 + \frac{\lambda_0}{\mu_1} + \frac{\lambda_0 \lambda_1}{\mu_1 \mu_2} + ... + \frac{\lambda_0 \lambda_1 ... \lambda_{N-1}}{\mu_1 \mu_2 ... \mu_N}}.$$

3.4 Pure Birth Process

$$\lambda_n = n\lambda, n = 1, 2, ...(\mu_n = 0)$$

$$P_n'(t) = -n\lambda P_n(t) + (n-1)\lambda P_{n-1}(t), P_a(0) = 1 (\text{at time 0 we have a persons}),$$

$$(e^{n\lambda t} P_n(t))' = (n-1)\lambda e^{\lambda t} e^{(n-1)\lambda t} P_{n-1}(t), n = a, a+1, ...$$

$$q_n(t) = e^{n\lambda t} P_n(t), q_a(0) = 1,$$

$$q_n'(t) = (n-1)\lambda e^{\lambda t} q_{n-1}(t), n = a, a+1, ...,$$

$$q_a(t) = e^{a\lambda t} P_a(t) = e^{a\lambda t} e^{-a\lambda t} = 1,$$

$$e^{-a\lambda t} = (e^{-\lambda t})^a = (P(\tau > t))^a = (1 - (1 - e^{-\lambda t}))^a \text{ is the survival function},$$

$$q_n(t) = (n-1) \int_0^t q_{n-1}(t) d e^{\lambda t},$$

$$q_{a+1}(t) = a \int_0^t d e^{\lambda t} = a(e^{\lambda t} - 1),$$

$$q_{a+2}(t) = a(a+1) \int_0^t (e^{\lambda n} - 1) d e^{\lambda n} = a(a+1) \frac{(e^{\lambda t} - 1)^2}{2} = \binom{a+1}{2} (e^{\lambda t} - 1)^2,$$

$$q_{a+3}(t) = \binom{a+2}{3} (e^{\lambda t} - 1)^3,$$

$$q_{a+n}(t) = \binom{a+n-1}{n}(e^{\lambda t}-1)^n, n = 0, 1, ...$$

$$P_{a+n}(t) = e^{-(n+a)\lambda t}\binom{a+n-1}{n}(e^{\lambda t}-1)^n,$$

and then

$$P_{a+n}(t) = \binom{a+n-1}{n}e^{-a\lambda t}(1-e^{-\lambda t})^n, n = 0, 1, ...$$

This follows a negative binomial distribution with $Mean = a/\rho$ and $Variance = aq/p^2$

$$\sum_{n=0}^{\infty} P_{a+n}(t) = \frac{e^{-a\lambda t}}{(1-e^{-\lambda t})^a}.$$

Therefore $Mean = \frac{a}{e^{-\lambda t}} = ae^{\lambda t}$, $Variance = \frac{a(1-e^{-\lambda t})}{e^{-2\lambda t}} = ae^{\lambda t}(e^{\lambda t}-1)$.

3.5 Simple Death Process

$$\mu_n = n\mu, n = 1, 2, ...(\lambda_n = 0), P_a(0) = 1$$

and analogously as in the previous model we can obtain

$$P_n(t) = \binom{a}{n}e^{-n\mu t}(1-e^{-\mu t})^{a-n},$$

$$\text{Mean} = ae^{-\mu t}, \text{Variance} = ae^{-\mu t}(1-e^{-\lambda t}).$$

3.6 Simple Birth Process

$$\lambda_n = n\lambda, \mu_n = n\mu, n = 0, 1, ...$$

We can show, that

$$\frac{\partial P}{\partial t} = (\lambda s - \mu)(s-1)\frac{\partial P}{\partial s},$$

where $P(s,t) = E(sX(t))$ is the probability generating function

If we differentiate the above equality with respect to s and put $s = 1$, we get

$$\frac{\partial m(t)}{\partial t} = (\lambda - \mu)m(t), m(t) = e^{(\lambda-\mu)t}, X(0) = 1.$$

3.7 Model $M/M/1$ (Busy Period)

At first we will give the definition and some properties of the Modified Bessel function, which we are going to use

Modified Bessel function

Definition 3.1. This is the function $I_n(t) = \sum_{r=0}^{\infty} \frac{(t/2)^{2r+n}}{n!(r+n)!}$.

1) $I_{-n}(t) = I_n(t)$;

2) $I_{n-1}(t) + I_{n+1}(t) = \frac{2nI_n(t)}{t}$;

3) $\sum_{n=-\infty}^{\infty} I_n(t)\zeta^n = e^{t/2}\Big(\rho + \frac{1}{\rho}\Big)$;

4) $L\Big(\frac{nI_n(\alpha t)}{t}\Big) = \Big(\frac{\alpha}{s+\sqrt{s^2-\alpha^2}}\Big)$.

Start with one customer at time $t = 0$. The duration of time a server is busy(i.e. till the queue becomes empty) is called *busy period*. The Kolmogorov equations are

$$\begin{cases} P_0' = \mu P_1 \\ P_1' = \mu P_2 - (\lambda + \mu)P_1 \\ \dots \\ P_n' = \mu P_{n+1} - (\lambda + \mu)P_n + \lambda P_{n-1} \end{cases}$$

$$P_0(t) = P(\text{the system is empty at time t}) = P(\tau \le t),$$

$$P_0'(t) = \text{pdf of the busy period} = \mu P_1(t).$$

Assume, that $P_1(0) = 1$. Taking the Laplace transform, we get

$$s\hat{P}_1 - 1 = \mu\hat{P}_2 - (\lambda + \mu)\hat{P}_1,$$

$$(s + \lambda + \mu)\hat{P}_1 = 1 + \mu\hat{P}_2,$$

$$s + \lambda + \mu = \frac{1}{\hat{P}_1} + \mu\frac{\hat{P}_2}{\hat{P}_1},$$

$$\hat{P}_1 = \frac{1}{s + \lambda + \mu - \mu\frac{\hat{P}_2}{\hat{P}_1}},$$

$$s\hat{P}_2 - 1 = \mu\hat{P}_3 - (\lambda + \mu)\hat{P}_2 + \lambda\hat{P}_1,$$

$$(s + \lambda + \mu)\hat{P}_2 = \mu\hat{P}_3 + \lambda\hat{P}_1,$$

$$s+\lambda+\mu=\mu\frac{\hat{P}_3}{\hat{P}_2}+\lambda\frac{\hat{P}_1}{\hat{P}_2},$$

$$\frac{\hat{P}_2}{\hat{P}_1}=\frac{\lambda}{s+\lambda+\mu-\mu\frac{\hat{P}_3}{\hat{P}_2}}.$$

Let us denote

$$f(s)=\frac{1}{s+\lambda+\mu-\frac{\lambda\mu}{s+\lambda+\mu-\frac{\lambda\mu}{s+\lambda+\mu-...}}},$$

i.e.

$$f(s)=\frac{1}{s+\lambda+\mu-\lambda\mu f(s)},$$

Then

$$\lambda\mu(f(s))^2-(s+\lambda+\mu)f(s)+1=0,$$

$$f(s)=\frac{s+\lambda+\mu-\sqrt{(s+\lambda+\mu)^2-4\lambda\mu}}{2\lambda\mu},$$

$$P_1\hat{(}s)=\frac{2}{s+\lambda+\mu+\sqrt{(s+\lambda+\mu)^2-4\lambda\mu}},$$

$$L(e^{-(\lambda+\mu)t})=\frac{1}{s+\lambda+\mu},$$

$$P_1(t)=2\frac{I_{\alpha t}}{at}e^{-(\lambda+\mu)t}$$

and therefore

$$P_0'(t)=\sqrt{\frac{\mu}{\lambda}}\frac{I_{at}}{t}e^{-(\lambda+\mu)t},t>0.$$

Taking the Laplace transform of $P_0'(t)$ for $s\to 0$, we obtain

$$sP_0=\begin{cases}\frac{2\mu}{\lambda+\mu+\lambda-\mu}=\frac{1}{\rho} & \text{if } \rho\geq 1\\ \frac{2\mu}{\lambda+\mu-\lambda+\mu}=\frac{1}{\rho} & \text{if } \rho<1\end{cases}.$$

3.8 Model $M/M/\infty$ (Infinite Server Queue, Self-Service System)

For this model we have

$$\lambda_n=\lambda,n=0,1,...;\mu_n=n\mu,n=1,2,...$$

$$X(t+h) = \begin{cases} X(t)+1 & \text{with probability } \lambda h + o(h) \\ X(t) & \text{with probability } 1-\lambda h - \mu_{X(t)} h + o(h) \\ X(t)-1 & \text{with probability } \mu_{X(t)} h + o(h) \end{cases}$$

$$P(\zeta, t) = E(\zeta^{X(t)})(\text{pgf});$$

$$P(\zeta, t+h) = E(\zeta^{X(t)+1})(\lambda h + o(h)) +$$

$$+E(\zeta^{X(t)})(1-\lambda h - \mu_{X(t)} h + o(h)) + E(\zeta^{X(t)-1})(\mu_{X(t)} h + o(h)) =$$

$$= \lambda\zeta P(\zeta,t)h + o(h) + P(\zeta,t) - \lambda P(\zeta,t)h - \mu\zeta\frac{\partial P(\zeta,t)}{\partial\zeta} + o(h) + \mu\frac{\partial P(\zeta,t)}{\partial\zeta}h + o(h).$$

Therefore

$$\frac{\partial P(\zeta,t)}{\partial t} = \lim_{h\to 0}\frac{P(\zeta,t+h)-P(\zeta,t)}{h} = \lambda\zeta P(\zeta,t) - \mu\zeta\frac{\partial P(\zeta,t)}{\partial\zeta} + \mu\frac{\partial P(\zeta,t)}{\partial\zeta}$$

and then we have a linear partial differential equation

$$\frac{\partial P(\zeta,t)}{\partial t} = \lambda(\zeta-1)P(\zeta,t) + \mu(1-\zeta)\frac{\partial P(\zeta,t)}{\partial\zeta}$$

with initial conditions(which actually mean, that the system is initially empty)

$$P(\zeta, 0) = 1,$$

$$E(\zeta^{X(0)}) = E(\zeta^0) = 1, (X(0) = 0),$$

$$\frac{\partial P(\zeta,t)}{\partial t} + \mu(\zeta-1)\frac{\partial P(\zeta,t)}{\partial\zeta} = \lambda(\zeta-1)P(\zeta,t).$$

To solve this equation, we use the method of characteristics

$$\frac{dt}{1} = \frac{d\zeta}{\mu(\zeta-1)} = \frac{dP}{\lambda(\zeta-1)P}$$

Now we have to generate 2 independent solutions. The solution of the equation

$$\frac{\mu dt}{1} = \frac{d\zeta}{\zeta-1}$$

is

$$(\zeta-1)e^{-\mu t} = C_1 \tag{3.1}$$

The solution of

$$\frac{\lambda d\zeta}{\mu} = \frac{dP}{P}$$

is

$$Pe^{-\rho\zeta} = C_2 \tag{3.2}$$

and $C_2 = f(C_1)$. Thus

$$Pe^{-\rho\zeta} = f((\zeta - 1)e^{-\mu t}), \tag{3.3}$$

where $f(.)$ is to be determined from the initial condition $P(\zeta, 0) = 1$. Put $t = 0$ in (5.3). Then $e^{-\rho\zeta} = f(\zeta - 1)$ and

$$e^{-\rho(\zeta+1)} = f(\zeta) \tag{3.4}$$

Now we substitute (5.4) in (5.3).

$$Pe^{-\rho\zeta} = e^{-\rho(1+(\zeta-1))e^{-\mu t}}$$

$$P(\zeta, t) = e^{-\rho(1-\zeta)(1-e^{-\mu t})}$$

Note, that $P(1, t) = 1$. Then $X(t)$ follows a Poisson distribution with $Mean = \rho(1-e^{-\mu t})$ and $Variance = \rho(1 - e^{-\mu t})$. If we put $\frac{\partial P}{\partial t} = 0$, then $P(\zeta) = e^{-\rho(1-\zeta)}$.

Chapter 4

Non-Birth-Death Queueing Models

4.1 Model $M^X/M/\infty$ (Bulk Arrival Queues)

Customers arrive in a Poisson fashion with k arrivals with probability $\lambda a_k h + o(h)$, where $k = 1, 2, ...$ and $\sum_{k=1}^{\infty} a_k = 1$. The service times are exponentially distributed with mean $1/\mu$. As in one previous model, we obtain the partial defferential equation

$$\frac{\partial P}{\partial t} = \lambda((A(\zeta) - 1)P(\zeta, t)) + \mu(1 - \zeta)\frac{\partial P}{\partial \zeta}$$

where

$$A(z) = a_1 z + a_2 z^2 + ...$$

Assume that steady state exists and also $P(1) = 1$.

$$\rho(1 - A(\zeta))P = (1 - \zeta)\frac{\partial P}{\partial \zeta}, \frac{1}{P}\frac{\partial P}{\partial \zeta} = \rho\frac{1 - A(\zeta)}{1 - \zeta},$$

$$\ln P = \rho \int_0^{\zeta} \frac{1 - A(\zeta)}{1 - \zeta} d\zeta + C, P(\zeta) = Ce^{\rho \int_0^1 \frac{1-A(u)}{1-u}} du.$$

But we know that $P(1) = 1$ and then

$$1 = Ce^{\rho \int_0^1 \frac{1-A(u)}{1-u}} du,$$

$$P(\zeta) = e^{\rho - \int_{\zeta}^1 \frac{1-A(u)}{1-u}} du.$$

Note, that if $A(\zeta) = \zeta$, $P(\zeta) = e^{-\rho(1-\zeta)}$.

4.2 Model $M^X/M/1$ (Bulk Arrival Queues)

Let customers arrive in batches in a Poisson process with parameter λ. Assume, that the batch size X is a r.v. with $P(X = k) = a_k, k = 1, 2, ...$, i.e. the probability that a batch of k arrivals arrives in an infinite time interval $(t, t + h)$ is $\lambda a_k h + o(h)$.

Let $A(\zeta) = \sum_{k=1}^{\infty} a_k \zeta^k$ is the pgf of the arrivals X.

$$A(\zeta) = A(1) + (\zeta - 1)A'(1) + \frac{(\zeta-1)^2}{2}A''(1) + ...$$

and $A(1) = 1$. The mean is $A'(1)$ and the second moment is $A''(1)$. The variance is given by $A''(1) + A'(1) - (A''(1))^2$. Then

$$\frac{1 - A(\zeta)}{1 - \zeta} = A'(1) - \frac{1-\zeta}{2!}A''(1) + ...$$

Assume, that the steady-state solutions exist. Then we get

$$\mu P_1 - \lambda P_0 = 0,$$

$$\mu P_2 - \lambda P_1 + \lambda a_1 P_0 = 0,$$

$$\mu P_3 - (\lambda + \mu)P_2 + \lambda(a_1 P_1 + a_2 P_0) = 0,$$

$$\mu P_4 - (\lambda + \mu)P_3 + \lambda(a_1 P_2 + a_2 P_1 + a_3 P_0) = 0,$$

$$...$$

Define $P(\zeta) = P_0 + P_1\zeta + P_2\zeta^2 + ...$-the pgf of P. We multiply the equations with $\zeta^0, \zeta^1, ...$ and we add them,

$$\mu(P_1 + P_2\zeta + ...) - \lambda P_0 - (\lambda + \mu)(P_1\zeta + P_2\zeta^2 + ...) +$$

$$+\lambda(a_1 P_0 \zeta + (a_1 P_1 + a_2 P_0)\zeta^2 + (a_1 P_2 + a_2 P_1 + a_3 P_0)\zeta^3 + ...) = 0.$$

Now we operate on the last term

$$\lambda(a_1\zeta(P_0 + P_1\zeta + ...) + a_2\zeta^2(P_0 + P_1\zeta + ...) + a_3\zeta^2(P_0 + P_1\zeta + ...) + ...) = \lambda A(\zeta)P(\zeta),$$

$$\frac{\mu}{\zeta}(P(\zeta) - P_0) - \lambda P_0 - (\lambda + \mu)(P(\zeta) - P_0) + \lambda A(\zeta)P(\zeta) = 0,$$

$$P(\zeta)(\frac{\mu}{\zeta} - (\lambda + \mu) + \lambda A(\zeta)) = \frac{\mu}{\zeta}P_0 + \lambda P_0 - (\lambda + \mu)P_0,$$

$$P(\zeta)(\mu - (\lambda + \mu)\zeta + \lambda\zeta A(\zeta)) = \mu(1 - \zeta)P_0,$$

$$P(\zeta) = \frac{\mu(1-\zeta)P_0}{\mu(1-\zeta) - \lambda\zeta(1 - A(\zeta))} = \frac{P_0}{1 - \rho\zeta\frac{(1-A(\zeta))}{1-\zeta}},$$

$$P(\zeta)(1 - \rho\zeta(A'(1) + \frac{\zeta - 1}{2}A''(1) + ...)) = P_0,$$

$$P_0 = 1 - \rho A'(1)$$

and steady-state exists if $\rho A'(1) < 1$.

$$P'(1)(1-\rho A'(1)) + P(1)(-\rho A'(1) - \frac{A''(1)}{2}) = 0,$$

$$P'(1) = \frac{\rho A'(1) + \frac{A''(1)}{2}}{1-\rho A'(1)}.$$

For the mean system size we obtain

$$L_s = \frac{\rho(2A'(1) + A''(1))}{2(1-\rho A'(1))}.$$

Then stedy state exist if $\rho A'(1) < 1$. Let us consider the following special cases

1) $M/M/1$ **Queue**

$$A(\zeta) = \zeta, A'(1) = 1, A''(1) = 0$$

$$P(\zeta) = \frac{1-\rho}{1-\rho\zeta} \text{ and then } P_n = (1-\rho)\rho^n, n = 1, 2, ...$$

$$L_s = \frac{\rho 2}{2(1-\rho)} = \frac{\rho}{1-\rho}$$

2)

$$A(\zeta) = q\zeta + p\zeta^2, p + q = 1$$

$$a_1 = q, a_2 = p, A'(1) = q, A''(1) = 2p$$

$$1 - A(\zeta) = 1 - q\zeta - p\zeta^2 = (1-\zeta) + p\zeta(1-\zeta)$$

$$\frac{1-A(\zeta)}{1-\zeta} = 1 + p\zeta$$

$$P(\zeta) = \frac{1-\rho q}{1-\rho\zeta(1+p\zeta)}$$

$$L_s = \frac{\rho(2q+2p)}{2(1-\rho q)} = \frac{\rho}{1-\rho q}$$

If in this case $q = 1$, we get case 1).

3)$M^r/M/1$ **Queue** For this model $a_r = 1$, where r is a fixed batch size.

$$A(\zeta) = \zeta^r, A'(1) = r, A''(1) = r(r-1)$$

and therefore

$$P(\zeta) = \frac{1-\rho r}{1-\rho\zeta(1+\zeta+...+\zeta^{n-1})}; L_s = \frac{\rho(2r + r(r-1))}{2(1-\rho r)} = \frac{\rho r(r+1)}{2(1-\rho r)}$$

If $r = 1$, we get case 1). This can be treated as $M/E_r/1$ queue, where E_r=Erlang service=sum of r iid exponential r.vs.

4)

$$a_r = a(1-a)^{r-1}, r = 1, 2, ...$$

This has lack of memory property.

$$A(\zeta) = a\zeta \sum_{r=1}^{\infty} (1-a)^{r-1}\zeta^{r-1} = \frac{a\zeta}{1-(1-a)\zeta}$$

$$A(\zeta)(1-(1-a)\zeta) = a\zeta$$

$$A'(\zeta)(1-(1-a)\zeta) - A(\zeta)(1-a) = a$$

Put $\zeta = 1$.Then $A'(1) = 1/a$, $A''(1) = a/(2(1-a))$ and

$$P(\zeta) = \frac{1-\rho A'(1)}{1-\rho\zeta\frac{a\zeta}{1-(1-a)\zeta}}$$

Example. Consider a multistage machine line process which produces an assembly in quality. Number of defectives per item is 1 or 2. The interarrival times are exponential and $\lambda_1 = 1/hr$, $\lambda_2 = 2/hr$, $\mu = 6/hr$. So, we have $\lambda_1 + \lambda_2 = 3/hr, a_1 = 1/3, a_2 = 2/3$.

$$A(\zeta) = \frac{\zeta + 2\zeta^2}{3}, A'(1) = 5/3, A''(1) = 4/3, \rho = 1/2,$$

$$\rho A'(1) = 1/2.5/3 = 5/6 < 1; L_s = \frac{1/2(10/3 + 4/3)}{2(1-5/6)} = 7.$$

4.3 Two-Station Series Model with Zero Queue Capacity (Sort of a Network Model)

A customer, arriving for service, must go through station 1 and station 2. An entering customer first will go to station 1; he will go to station 2, if it is empty or will wait on station 1 till station 2 becomes empty. A customer will enter the system as long as station 1 is empty. There is no queue in froint of station 1 or station 2. λ is the arrival rate, μ_1 is the service rate at station 1, μ_2 is the service rate at station 2.

Here we give all possible states and their interpretations

$(0,0)$ – No customer in the system

$(1,0)$ – one customer in station 1 and no customer in station 2

$(0,1)$ – no customer in station 1 and one in station 2

$(1,1)$ – both receive service

$(b,1)$ – one customer in station 1 finished service, but station 2 is busy

Therefore

$$\begin{cases} \lambda P_{00} = \mu_2 P_{01} \\ \mu_1 P_{10} = \lambda P_{00} + \mu_2 P_{11} \\ (\mu_2 + \lambda) P_{01} = \mu_1 P_{10} + \mu_2 P_{b1} \\ (\mu_1 + \mu_2) P_{11} = \lambda P_{01} \\ \mu_2 P_{b1} = \mu_1 P_{11} \end{cases}$$

From these equations we obtain

$$\begin{cases} P_{01} = \frac{\lambda}{\mu_2} P_{00} \\ P_{11} = \frac{\lambda}{\mu_1+\mu_2} P_{01} = \frac{\lambda^2}{\mu_2(\mu_1+\mu_2)} P_{00} \\ P_{b1} = \frac{\mu_1}{\mu_2} P_{11} = \frac{\lambda^2}{\mu_2^2(\mu_1+\mu_2)} P_{00} \\ P_{10} = \left(\frac{\lambda}{\mu_1} + \frac{\mu_2}{\mu_1} \frac{\lambda^2}{\mu_1+\mu_2} \right) P_{00} \end{cases}$$

and then

$$P_{00} = \frac{1}{\left(1 + \frac{\lambda}{\mu_2} + \frac{\lambda}{\mu_1} + \frac{\mu_2\lambda^2}{\mu_1\mu_2(\mu_1+\mu_2)} + \frac{\lambda^2}{\mu_2(\mu_1+\mu_2)} + \frac{\lambda}{\mu_1} + \frac{\mu_2\lambda^2}{\mu_1\mu_2(\mu_1+\mu_2)}\right)}$$

$$L_s = P_{01} + P_{10} + 2(P_{11} + P_{01})$$

4.4 Model M/Hyperexponential/$1/r$

Note If $X_1, X_2, ..., X_n$ are independent exponentially distributed r.vs. with parameters $\mu_i, i = 1, 2, ..., n$, then $X_1 + X_2 + ... + X_n$ follows a hyperexponential distribution. For this model we have

$$\begin{cases} \mu_1 P_1 = \lambda P_0 \\ \mu_2 P_2 = \mu_1 P_1 \\ ... \\ \mu_{r-1} P_{r-1} = \mu_r P_r \end{cases}$$

and then $P_r = \frac{\lambda}{\mu_r} P_0$, $P_0 = \frac{1}{1+\lambda(\frac{1}{\mu_1} + \frac{1}{\mu_2 + ... + \frac{1}{\mu_r}})}$.

4.5 Processor Model with Failures

Consider a single processor with an infinite waiting norm. λ is the arrival rate and μ is the service rate. The processor fails at rate ν and when failed all the customers are lost.

The Kolmogorov equations for this system are

$$\begin{cases} \lambda P_0 = \mu P_1 + \nu(P_1 + P_2 + ...) = \mu P_1 + \nu(1 - P_0) \\ (\lambda + \nu)P_0 = \mu P_1 + \nu \\ (\lambda + \nu + \mu)P_1 = \mu P_2 + \lambda P_0 \\ (\rho + \theta)P_0 = P_1 + \theta, \rho = \lambda/\mu, \theta = \nu/\mu \\ (\rho + \theta + 1)P_1 = \rho P_0 + P_2 \\ (\rho + \theta + 1)P_n = P_{n+1} + lP_{n-1}, n = 1, 2, ... \end{cases}$$

Assume, that the solution is in this form:

$$P_n = \beta^n P_0, n = 1, 2, ...$$

Therefore

$$\beta^{n+1} - (\rho + \theta + 1)\beta^n + \rho\beta^{n-1} = 0$$

$$\beta^2 - (\rho + \theta + 1)\beta + \rho = 0$$

$$\beta = \frac{\rho + \theta + 1 - \sqrt{(\rho + \theta + 1)^2 - 4\rho}}{2}$$

(The other root is > 1.)

$$P_0 = 1 - \beta, L_s = \frac{\beta}{1 - \beta}.$$

4.6 Bulk Service

4.6.1 $M/M(1,2)/1$

Customers arrive at a single server queue with a Poisson rate λ. The server serves 2 customers at a time, if available, otherwise - 1 at a time. Denote the states

$$\begin{cases} 0' = \text{the system is empty} \\ 0 = \text{the system is busy, but the queue is empty} \\ n = \text{there are n customers in the queue} \end{cases}$$

The Kolmogorov equations for the system are

$$\begin{cases} \lambda P_{0'} = \mu P_0 \\ (\lambda + \mu)P_0 = \mu P_1 + \mu P_2 + \lambda P_{0'} \\ (\lambda + \mu)P_n = \mu P_{n+2} + \lambda P_{n-1}, n = 1, 2, ... \end{cases}$$

Assume, that $P_n = \beta^n P_0, n = 1, 2, ...$ Then we get:

$$\mu\beta^{n+2} - (\lambda + \mu)\beta^n + \lambda\beta^{n-1} = 0$$

$$\beta^3 - (\rho+1)\beta + \rho = 0, \rho = \frac{\lambda}{\mu}$$

$$\beta^3 - \beta - \rho(\beta - 1) = 0$$

$$(\beta - 1)(\beta^2 + \beta) - \rho(\beta - 1) = 0$$

$$\beta^2 + \beta = \rho$$

$$\beta = \frac{\sqrt{1+4\rho} - 1}{2}$$

Note, that $\beta < 1$ if and only if $\rho < 2$. Therefore $P_{0'} = P_0/\rho$, $P_n = \beta^n P_0$, $P_0 = \frac{1}{1/\rho + 1/(1-\beta)}$, $L_q = \frac{2\rho(2-\rho)}{4+\rho+\rho\sqrt{1+4\rho}}$.

4.6.2 $M/M(1,k)/1$

Rouche's Theorem If $f(\zeta)$ and $g(\zeta)$ are analytic inside and on a simple closed curve c and $|g(\zeta)| < |f(\zeta)|$ on c, then $f(\zeta)$ and $f(\zeta) + g(\zeta)$ have the same number of roots inside c. For this system the Kolmogorov equations are

$$\begin{cases} \lambda P_{0'} = \mu P_0 \\ (\lambda+\mu)P_0 = \lambda P_{0'} + \mu(P_1 + \mu P_2 + ... + P_k) \\ (\lambda+\mu)P_n = \lambda P_{n-1} + \mu P_{k+n}, n = 1, 2, ... \end{cases}$$

From the last equation above, we obtain

$$\mu\zeta^{k+1} - (\lambda+\mu)\zeta + \lambda = 0.$$

Here we put $f(\zeta) = (\lambda+\mu)\zeta$, $g(\zeta) = \mu\zeta^{k+1} + \lambda$. $f(\zeta)$ and $g(\zeta)$ are analytic inside and on a unit circle c and $|g(\zeta)| < \mu + \lambda = |f(\zeta)|$ on $|\zeta| = 1$. $f(\zeta) + g(\zeta)$ has only one root inside c. Let α be this root. Then $P_n = \alpha^n P_0, n = 1, 2, ...$, $P_{0'} = P_0/\rho$, $P_0 = \frac{1}{1/\rho + 1/(1-\alpha)}$, $L_q = \frac{\rho\alpha}{(1-\alpha)(1-\alpha+\rho)}$.

Note, that $\alpha^{k+1} - (\rho+1)\alpha + \rho = 0$, $\alpha^{k+1} - \alpha - \rho(\alpha - 1) = 0$.

And then $\alpha^k + \alpha^{k-1} + ... + \alpha = \rho$, $\alpha < 1$ is real and $\rho < k$.

Some interesting problems Let $f(t)$ be the *pdf* of a positive r.v. X. Its Laplace transform is

$$E(e^{-sX}) = \int_0^\infty e^{-st} f(t)dt = \varphi(s)$$

and

$$-\varphi'(0) = \int_0^\infty t f(t)dt = \alpha_1 = ''\mu'', \alpha_1 = -\varphi'(0).$$

Example.

$$f(t) = \alpha e^{-\alpha t}, t > 0, \varphi(s) = \frac{\alpha}{\alpha + s}, \varphi'(0) = -\frac{\alpha}{\alpha^2} = -\frac{1}{\alpha}.$$

The mean is $\frac{1}{\alpha}$. We know, that the busy period has density $P_{0'} = \mu P_1$. Using the Laplace transform we get

$$\hat{P_1(s)} = \frac{s + \lambda + \mu - \sqrt{(s + \lambda + \mu)^2 - 4\lambda\mu}}{2\lambda\mu}.$$

Then the *pdf* of the busy period has Laplace transform

$$\varphi(s) = \frac{1}{2\lambda}(s + \lambda + \mu - \sqrt{(s + \lambda + \mu)^2 - 4\lambda\mu})$$

$$\varphi'(0) = \frac{1}{2\lambda}(1 - \frac{2(s + \lambda + \mu)}{2\sqrt{(s + \lambda + \mu)^2 - 4\lambda\mu}}) = \frac{1}{2\lambda}(1 - \frac{\lambda + \mu}{\mu - \lambda}) = \frac{1}{\lambda - \mu}.$$

Hence, the mean busy period is $1/(\mu - \lambda) = 1/(\mu(1 - \rho))$.

Consider an $M/M/1$ queue. Let $E[N]$ be the average number of customers during a busy period. Only one of these will see the system empty.

$$P_0 = \frac{1}{E[N]}, E[N] = \frac{1}{P_0} = \frac{1}{1 - \rho}$$

Suppose T is the duration of the busy period, μ is the service rate. Then

$$\mu E[T] = E[N], E[T] = \frac{1}{\mu(1 - \rho)}.$$

Departure Process of an $M/M/1$ Queue(output process) Interdeparture time = =service time(if there are customers)+arrival time(if the system is empty). The *pdf* of the interdeparture time is

$$\frac{\mu}{s + \mu}\rho + (\mu e^{-\mu t} * \lambda e^{-\lambda t})(1 - \rho)$$

where by * we denote convolution. The Laplace transform of the interdeparture time is

$$\frac{\mu}{s + \mu}\rho + \frac{\lambda}{s + \lambda}\frac{\mu}{s + \mu}(1 - \rho) = \frac{\lambda}{s + \lambda}.$$

So, the interdeparture time has the same Laplace transform as the interarrival time. This result is true for general queueing models. It can be proved by time reversability, i.e. using that $X(t)$ and $X(-t)$ are identically distributed.

Residual lifetime Denote T_n= time of the n-th arrival in a Poisson process with parameter λ. We have, that the Excess lifetime=$E[t] = T_{N(t)+1} - t$. If $E[t] > x$, no events happen between t and $t + x$.

$$P(E[t] > x) = e^{-\lambda x}; P(E[t] \leq x) = 1 - e^{-\lambda x}$$

and it means, that $E[t]$ also follows exponential distribution.

Example. Consider a $M/M/1$ queue. If n customers are in the system, an arriving customer joins the queue with probability $\lambda\frac{n+1}{n+2}$. The service rate is μ again.

$$\mu P_{n+1} = \lambda\frac{n+1}{n+2}P_n,$$

$$P_{n+1} = \rho\frac{n+1}{n+2}P_n = \rho^2\frac{n+1}{n+2}\frac{n}{n+1}P_{n-1},$$

$$P_n = \frac{\rho^n}{n+1}P_0.$$

Then

$$P_0 = \frac{1}{-\ln(1-\rho)}, \rho < 1$$

$$L_s = \sum_{n=1}^{\infty} nP_n = \sum_{n=1}^{\infty} n\frac{\rho^n}{n+1}P_0 = \sum_{n=1}^{\infty}(1 - \frac{1}{n+1})\rho^n P_0 = \frac{\rho P_0}{1-\rho} - (1 - P_0) = \frac{1}{1-\rho}P_0 - 1.$$

We show that W_s is smaller in a $M/M/1$ model with service rate 2μ, than in a $M/M/2$ model with service rate μ.

We know, that $W_s = \frac{1}{\mu(1-\rho)}$ in $M/M/1$ model with service rate μ. $W_s = \frac{4}{\mu(4-\rho^2)}$ in $M/M/2$ model with rate μ. $W_s = \frac{1}{2\mu(1-\frac{\lambda}{2\mu})} = \frac{1}{2\mu-\lambda}$ in $M/M/1$ model with service rate 2μ. Now it is easy to observe, that $\frac{1}{2\mu-\lambda} < \frac{4}{\mu(4-\rho^2)}$ if and only if $\lambda < 2\mu$, which is valid.

Example. Consider a $M/M/\infty$ queue with servers, numbered $1, 2, \ldots.$ When arrives, a customer chooses the lowest number that is free. Let us find the duraction P_n of the time, that a server c is free.

We know, that for a $M/M/c/c$ (loss) system

$$P(\text{all servers are busy}) = \pi_c = \frac{\rho^c/c!}{1 + \rho + \ldots + \rho^c/c!}$$

$$r_c = \text{arrival rate to servers } c+1, c+2, \ldots$$

$$\lambda_c = \text{arrival rate to server } c$$

$$r_c = \lambda P_c$$

$$\lambda_c = r_{c-1} - r_c = \lambda(P_{c-1} - P_c)$$

$$\text{Fraction of time that server c is busy} = \frac{\lambda_c}{\mu}$$

$$\text{Fraction of time that server c is free} = 1 - \frac{\lambda_c}{\mu}$$

Example. Suppose, that two streams of customers arrive in a Poisson fashion with rates λ_1 and λ_2 correspondingly.

$$E[z^{N_i(t)}] = e^{-\lambda_i t(1-z)}, i = 1, 2,$$

N_1 and N_2 are independent and then

$$E[z^{N_1(t)+N_2(t)}] = e^{-(\lambda_1+\lambda_2)t(1-z)},$$

which means, that $N_1(t) + N_2(t) \sim Poi(\lambda_1 + \lambda_2)$.

Example. A system consists of 2 subsystems. With probability p we route a customer to the first system, and with probability $1 - p$—to the second one.

$$P(N_1(t) = j) = \sum_{n=j}^{\infty} P(N_1(t) = j, N(t) = n) =$$

$$= \sum_{n=j}^{\infty} P(N_1(t) = j | N(t) = n) \frac{e^{-\lambda t}(\lambda t)^n}{n!} = \sum_{n=j}^{\infty} \binom{n}{j} p^j (1-p)^{n-j} \frac{e^{-\lambda t}(\lambda t)^n}{n!} =$$

$$= e^{-\lambda t} \sum_{n=j}^{\infty} \frac{(\lambda tp)^j}{j!} \frac{(\lambda t(1-p))^{n-j}}{(n-j)!} = e^{-\lambda tp} \frac{(\lambda tp)^j}{j!}$$

The last equality means, that $N(t)$ is also a Poisson process with rate λp.

Chapter 5

Cost Models

The objective of a queueing cost model is to determine the level of service (either the service rate or the number of servers) that balances the two conflicting costs: the cost of offering the service and the cost of delay in offering the service. The two types of costs are in conflict because an increase in one automatically causes reduction in the other.

Let x= *service level*=service rate(μ) or number of servers(c). Then the total cost can be represented as

$$ETC(x) = EOC(x) + EWC(x)$$

where

$ETC(x)$=expected total cost per unit time

$EOC(x)$=expected cost of operating per unit time

$EWC(x)$=expected cost of waiting per unit time

The simplest forms for EOC and EWC are the following linear equations:

$EOC(x) = C_1 x, EWC(x) = C_2 L_s$,

where

C_1=Marginal cost per unit of x per unit time

C_2=Cost of waiting per unit time per(waiting) customer

The following two examples illustrate the use of the cost model. The first example assumes x to equal the service rate μ, and the second assumes x to equal the number of parallel servers c.

5.1 Model 1: Optimal Service Rate μ

Consider a $M/M/1$ queue with arrival rate λ and service rate μ. Assume, μ is controllable.

C_1=cost per unit increase in μ per unit time

C_2=cost of waiting customer

$$ETC(\mu) = \text{Expected cost of waiting and service per unit time, given } \mu =$$

$$= C_1\mu + C_2 L_s = C_1\mu + C_2\frac{\lambda}{\mu - \lambda}.$$

Now we have that

$$\frac{\partial ETC(\mu)}{\partial \mu} = 0 \rightarrow C_1 - C_2 \frac{\lambda}{(\mu - \lambda)^2} = 0 \rightarrow \mu = \lambda + \sqrt{\frac{C_2 \lambda}{C_1}}.$$

5.2 Model 2: Optimal Number of Servers

Consider a $M/M/c$ model

C_1=Cost of an additional server per unit time

$L_s(c)$=Expected number of customers in the system given c servers

C_2=cost of waiting customer

$$ETC(c) = cC_1 + C_2 L_s(c)$$

Here c is discrete. We have to find c, such that

$$\begin{cases} ETC(c-1) \geq ETC(c) \\ ETC(c+1) \geq ETC(c) \end{cases}$$

i.e.

$$L_s(c) - L_s(c+1) \leq \frac{C_1}{C_2} \leq L_s(c-1) - L_s(c).$$

Chapter 6

Network Models: Communication Network, Manufacturing Network

We dealt with a single isolated queueing system. It is a natural extension for us now to look at collection of interactive queueing systems, networks of queues, where the departures of some queues form the arrivals of others. The analysis of a queueing network is much more complicated and involved due to the interactions among various queues and we have to examine them as a whole. The state of one queue is generally dependent of the others because of feedback loops. From the network topology point of view, queueing networks can be categorized into two generic classes, namely, open queueing networks and closed queueing networks. These queueing networks can be further classified into two major classes, namely networks with single-class customers known as Jackson networks and networks with multi-class customers known as BCMP (Baskett Chandy Muntz Palacios) Networks. Yet another classification of queueing networks with multi-class customers is called mixed networks, which are closed with respect to some customers classes and open with respect to others. In this chapter we will concentrate only on open queueing networks and closed queueing networks. In particular, we are interested in applying the Network-Pade approximation technique to calculate the normalizing constants arising in closed queueing networks with single-class and multi-class customers.

Open Queueing Networks In an open queueing network, customers arrive from external sources outside the domain of interest, go through several queues or even revisit a particular queue more than once and finally leave the system. The total sum of arrival rates is equal to the total departure rate under steady state conditions. These networks are good models for analyzing circuit-switching and packet-switching data networks.

6.1 Two-Server Queue in Tandem

Recall, that in a $M/M/1$ queue the departure process follows a Poisson distribution with the same parameter λ like the arrival process. Using *reversibility* we can prove that the probability distribution of the number of departures in $(0,t)$ in steady state is the same as the probability distribution of the number of arrivals in $(0, t)$ and these two are independent. This result is called Burke's theorem. Let us denote

n_1=number of customers with server 1

n_1=number of customers with server 2

The Burke's theorem gives, that the probability distributions of n_1 and n_2 are independent. The Kolmogorov equations for this model are:

$$\begin{cases} \lambda P_{00} = \mu_2 P_{01} \\ (\lambda+\mu_1)P_{10} = \lambda P_{00} + \mu_2 P_{11} \\ (\lambda+\mu_2)P_{01} = \mu_1 P_{00} + \mu_2 P_{02} \\ (\lambda+\mu_1)P_{n_1 0} = \mu_2 P_{n_1 1} + \lambda P_{n_1-10}, n = 0, 1, ... \\ (\lambda+\mu_2)P_{0n_2} = \mu_1 P_{1n_2-1} + \mu_2 P_{0n_2+1}, n = 1, 2, ... \\ (\lambda+\mu_1+\mu_2)P_{n_1 n_2} = \mu_1 P_{n_1+1n_2-1} + \mu_2 P_{nn_2+1} + \lambda P_{n_1-1n_2}, n_1, n_2 = 1, 2, ... \end{cases}$$

For $P(n_1, n_2) = P_1(n_1)P_2(n_2)$, substituting in the above system and denoting $\rho_1 = \lambda/\mu_1$ and $\rho_2 = \lambda/\mu_2$, we get

$$\begin{cases} \rho_2 P_1(0)P_2(0) = P_1(0)P_2(1) \\ (1+\frac{1}{\rho_1})P_1(n_1)P_2(0) = \frac{1}{\rho_2}P_1(n_1)P_2(1) + P_1(n_1-1)P_2(0) \\ (1+\frac{1}{\rho_2})P_1(0)P_2(n_2) = \frac{1}{\rho_1}P_1(1)P_2(n_2-1) + P_1(0)P_2(n_2+1) \\ (1+\frac{1}{\rho_1}+\frac{1}{\rho_2})P_1(n_1)P_2(n_2) = \frac{1}{\rho_1}P_1(n_1+1)P_2(n_2-1) + \frac{1}{\rho_2}P_1(n_1)P_2(n_2+1) + P_1(n_1-1)P_2(n_2) \end{cases}$$

Thus,

$$P_1(n_1) = \rho_1 P_1(n_1-1) = \rho_1^2 P_1(n_1-2) = ... = \rho_1^{n_1} P_1(0).$$

But we know, that $\sum_{n_1=0}^{\infty} P_1(n_1) = 1$ and then $P_1(0) = 1-\rho_1$, which implies, that

$$P_1(n_1) = (1-\rho_1)\rho_1^{n_1}, \rho_1 < 1$$

Analogously we obtain

$$P_2(n_2) = (1-\rho_2)\rho_2^{n_2}, \rho_2 < 1$$

and therefore

$$P(n_1, n_2) = (1-\rho_1)\rho_1^{n_1}(1-\rho_2)\rho_2^{n_2}$$

$$L_s = \sum_{n_1}\sum_{n_2}(n_1+n_2)P(n_1, n_2) = \frac{\rho_1}{1-\rho_1} + \frac{\rho_2}{1-\rho_2}, W_s = \frac{L_s}{\lambda}$$

This can be generalised to c servers:

$$P(n_1, n_2, ..., n_c) = \prod_{i=1}^{c} \rho_i^{n_i}(1-\rho_i), \rho_i < 1.$$

And then we have

Burke's Theorem for c servers Departure from $M/M/c$ model is also Poisson and its rate is the same as the arrival rate.

6.2 Jackson Network

There are k servers. Customers arrive from otside the system to server $i, i = 1, 2, ..., k$ in accordance with independent Poisson process at rate n_i. Once a customer is served by server i, he joins the queue in froint of server j with probability p_{ij}. Hence $\sum_j p_{ij} \leq 1$.

$1 - \sum_j p_{ij} = P$(a customer departs the system after he/she is served by server i);

λ_j=total arrival rate to server j;

r_j=outside arrival rate;

Traffic equations

$$\lambda_j = r_j + \sum_{i=1}^{k} \lambda_i p_{ij}, j = 1, 2, ..., k.$$

Let $\rho_j = \frac{\lambda_j}{\mu_j} < 1$.

$$P(n_j \text{ customers with server } j, j = 1, 2, ..., k) = \prod_{j=1}^{k} \left(\frac{\lambda_j}{\mu_j}\right)_j^n \left(1 - \frac{\lambda_j}{\mu_j}\right)$$

$$L_s = \sum_{j=1}^{k} \frac{\rho_j}{1 - \rho_j}, \rho_j < 1, W_s = \frac{L_s}{\sum r_j}$$

The solution of the system of the traffic equations is unique. We can write the system in this way:

$$\Lambda = R + \Lambda P$$

where

$$\Lambda = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ ... \\ \lambda_k \end{pmatrix}, R = \begin{pmatrix} r_1 \\ r_2 \\ ... \\ r_k \end{pmatrix}, P = ((p_{ij}))$$

The matrix $I - P$ is nonsingular and then

$$\Lambda(I - P) = R, \Lambda = (I - P)^{-1} R$$

Remark The traffic equations have a unique solution.

Example. There are 2 servers. We know, that

$r_1 = 4, r_2 = 5, \mu_1 = 8, \mu_2 = 10, p_{11} = 0, p_{12} = 1/2, p_{21} = 1/4, p_{22} = 0$. So, given this data, we obtain the system of traffic equations

$$\begin{cases} \lambda_1 = 4 + \frac{1}{4}\lambda_2 \\ \lambda_2 = 4 + \frac{1}{4}\lambda_1 \end{cases}$$

which has solution $\lambda_1 = 6, \lambda_2 = 8$. Therefore $L_s = \frac{6}{8-6} = 7, W_s = 7/9$.

Closed Queueing Networks A closed queueing network is one in which customers neither arrive at nor depart from the system. The existing customers in the network simply circulate through various queues and may revisit a particular queue more than once as in case of open queueing networks. These networks are good models for analyzing window-type network flow controls as well as CPU job scheduling problems. We analyze queueing networks in the following order:

- Single class customers
– Markovian queues in tandem(no feedback)
– Open queueing networks
- Open acyclic networks(no feedback)
- General open cyclic networks(feedback)
– Closed queueing networks(feedback)
- Multi-class customers
–Open queueing networks(feedback)
– Closed queueing networks(feedback)

What we mean by feedback is: customers who are finishing service in one node can join the same node immediately or after visiting some other node based on some probability law, whereas, in networks with no feedback, a customer can visit a node at most once.

6.3 Closed System

There are m customers moving among a system of k servers. The service rates are $\mu_i, i = 1, 2, ..., k$.

p_{ij}=P(a customer, who is in node i will go to node j after completing his service)

$\sum_{j=1}^{k} p_{ij} = 1$ and $P = ((p_{ij}))$ is the transition probability matrix. Assume, that P is irreducible. Then the stationary probabilities $\pi = (\pi_1, ..., \pi_k)$ exist and satisfy the equation

$$\pi P = \pi, \text{i.e.} \pi_j = \sum_{i=1}^{k} \pi_i p_{ij}$$

and $\sum_{i=1}^{k} \pi_j = 1$. If $\lambda_m(j)$ = total arrival rate to node j, $\lambda_m = \sum \lambda_m(j)$, we obtain

$$\frac{\lambda_m(j)}{\lambda_m} = \sum_i \frac{\lambda_m(i)}{\lambda_m} p_{ij}.$$

For this model $r_j = 0$, because no customers come from outside.

From the above equalities we get

$$\pi_j = \frac{\lambda_m(j)}{\lambda_m},$$

which implies, that

$$\lambda_m(j) = \lambda_m \pi_j$$

There is an interesting result:

$$P(n_j \text{ customers at server } j, j = 1, 2, ..., k) = K \prod_{j=1}^{k} \left(\frac{\lambda_m(j)}{\mu_j} \right)^{n_j} =$$

$$= C \prod_{j=1}^{k} \left(\frac{\pi}{\mu_j} \right)^{n_j}, \sum n_j = k,$$

$$C = \frac{1}{\sum_{n_1,...,n_k} \prod \left(\frac{\pi}{\mu_j} \right)^{n_j}}, \sum n_j = m.$$

Chapter 7

Non-Markovian Queues

7.1 Model $M/G/1$

Customers arrive according to a Poisson process with parameter λ to a single server queue. Service times are *iid* variables with *pdf* $\nu(t)$, *cdf* $B(t)$, mean $1/\mu$ and variance σ^2. The process is non-Markovian, since we also need knowledge of the service time, elapsed for the customer getting service. To overcome that, we observe the process only just after service completion times. We study the steady state, which at any time depends on how long the unit in service is getting service. Denote:

T_n+=departure time of the n-th customer

$X_n = X(T_n+)$=system size (number of customers in the system) at time T_n+

Y=the number of customers joining the queue during the service time of the presenting customers

$$q_n = P(n \text{ customers arrive during one service time}) =$$

$$= P(Y=n) = \int_0^\infty e^{-\lambda t}\frac{(\lambda t)^n}{n!}\nu(t)dt = e^{\lambda T}\frac{(\lambda T)^n}{n!} = \int_0^\infty e^{-\lambda t}\frac{(\lambda t)^k}{k!}dB(t)$$
$$= \int_0^\infty P(n \text{ customers arrive}|T=t)P(T=t)dt$$

T is a random variable with *pdf* $h(t)$.

Note, that if $T = 1/\mu$, $q_n = \frac{e^{-\frac{\lambda}{\mu}}(\frac{\lambda}{\mu})^k}{k!}$.

$$X_{n+1} = \begin{cases} X_n - 1 + Y & \text{if } X_n > 0 \\ Y & \text{if } X_n = 0 \end{cases}$$

or in short, $X_{n+1} = [X_n - 1]^+ + Y$. If $X_n = 1$ or $X_n = 0$, then $X_{n+1} = Y$.

The following equality holds:

$$\lim_{t\to\infty} P(X(t) = j) = \lim P(X(T_n) = j).$$

Transition Probabilities

$$p_{ij} = P(X_{n+1} = j | X_n = j) = \begin{cases} P(Y = j - i + 1) = q_{j-i+1} = \pi_j, & \text{if } i > 0 \\ q_j & \text{if } i = 0 \end{cases}.$$

The transition probability matrix is

$$P = ((p_{ij})) = \begin{pmatrix} q_0 & q_1 & q_2 & \dots \\ q_0 & q_1 & q_2 & \dots \\ 0 & q_0 & q_1 & \dots \\ 0 & 0 & q_0 & \dots \\ \dots & \dots & \dots & \dots \end{pmatrix}.$$

Stationary probabilities exist if $\rho = \frac{\lambda}{\mu} < 1$. Also

$$\pi P = \pi$$

$$(\pi_0, \pi_1, \pi_2, \dots) \begin{pmatrix} q_0 & q_1 & q_2 & \dots \\ q_0 & q_1 & q_2 & \dots \\ 0 & q_0 & q_1 & \dots \\ 0 & 0 & q_0 & \dots \\ \dots & \dots & \dots & \dots \end{pmatrix} = (\pi_0, \pi_1, \pi_2, \dots)$$

Now we multiply the terms in the left side to obtain these in the right one. The result is the following system of equations

$$\begin{cases} q_0\pi_0 + q_0\pi_1 = \pi_0 \\ q_1\pi_0 + q_1\pi_1 + q_0\pi_2 = \pi_1 \\ q_2\pi_0 + q_2\pi_1 + q_1\pi_2 + q_0\pi_3 = \pi_2 \\ \dots \end{cases}$$

Now we multiply the first equation by ζ^0, the second by ζ^1, etc., then add the right hand sides of all equations and obtain the function

$$P(\zeta) = \pi_0 + \pi_1\zeta + \pi_2\zeta^2 + \dots$$

Let us put

$$Q(\zeta) = q_0 + q_1\zeta + q_2\zeta^2 + \dots$$

We are going to find some relations between these 2 functions.

$$\pi_0 Q(\zeta) + \pi_1 Q(\zeta) + \pi_2\zeta Q(\zeta) + \pi_3\zeta^2 Q(\zeta) + \dots = P(\zeta)$$

$$Q(\zeta)[\pi_0 + \pi_1 + \pi_2\zeta + \pi_3\zeta^2 + \dots] = P(\zeta)$$

$$Q(\zeta)[\pi_0 + \frac{1}{\zeta}(\pi_1\zeta + \pi_2\zeta^2 + \pi_3\zeta^3 + ...)] = P(\zeta)$$
$$Q(\zeta)[\pi_0 + \frac{1}{\zeta}(P(\zeta) - \pi_0)] = P(\zeta)$$
$$Q(\zeta)[\pi_0(1 - \frac{1}{\zeta}) + \frac{P(\zeta)}{\zeta}] = P(\zeta)$$
$$Q(\zeta)[\pi_0(\zeta - 1) + P(\zeta)] = \zeta P(\zeta)$$
$$P(\zeta)[Q(\zeta) - \zeta] = \pi_0(1 - \zeta)Q(\zeta)$$
$$P(\zeta) = \frac{\pi_0(1 - \zeta)Q(\zeta)}{Q(\zeta) - \zeta} \tag{7.1}$$
$$Q(\zeta) = \sum_{k=0}^{\infty} q_k\zeta = \sum_{k=0}^{\infty}\Big(\int_0^{\infty} e^{-\lambda t}\frac{(\lambda t)^k}{k!}b(t)dt\Big)\zeta^k = \int_0^{\infty} e^{-\lambda t}\sum_{k=0}^{\infty}\Big(\frac{(\lambda t)^k}{k!}\zeta^k\Big)b(t)dt$$

and then

$$Q(\zeta) = \int_0^{\infty} e^{-\lambda t(1-\zeta)}b(t)dt.$$

Now we determine π_0 in (7.1)

$$P(\zeta)[Q(\zeta) - \zeta] = \pi_0(1 - \zeta)Q(\zeta), P(1) = 1, Q(1) = 1$$
$$P'(\zeta)(Q(\zeta) - \zeta) + P(\zeta)(Q'(\zeta) - 1) = \pi_0(-1)Q(\zeta) + \pi_0(1 - \zeta)Q'(\zeta).$$

Here we put $\zeta = 1$ and obtain

$$Q'(1) - 1 = \pi_0(-1), \pi_0 = 1 - Q'(1) = \frac{\lambda}{\mu} = \rho$$

and then

$$P(\zeta) = \frac{(1 - \rho)(1 - \zeta)Q(\zeta)}{Q(\zeta) - \zeta}$$
$$P''(\zeta)(Q(\zeta)-\zeta)+2P'(\zeta)(Q'(\zeta)-1)+P(\zeta)Q''(\zeta) = \pi_0(-1)Q''(\zeta)+\pi_0(1-\zeta)Q''(\zeta)-\pi_0 Q'(\zeta)$$

Put $\zeta = 1$:

$$2P'(1)(Q'(1) - 1) + Q''(1) = -2\pi_0 Q'(1)$$
$$P'(1) = \frac{Q''(1) + 2(1 - \rho)Q'(1)}{2(1 - Q'(1))} = L_s = \rho + \frac{Q''(1)}{2(1 - Q'(1))}.$$

We have already proved, that $L_s = L_q + \rho$. Then we obtain

Pollakzek-Khinchin Formula

$$L_q = \frac{Q''(1)}{2(1 - Q'(1))} = \frac{Q''(1)}{2(1 - \rho)}.$$

Thus,

$$W_q = \frac{L_q}{\lambda} = \frac{Q''(1)}{2\lambda(1 - \rho)}$$

7.1.1 Special Cases

Here we consider some special cases for the service distribution.

1. $M/D/1$ Here the service$(1/\mu)$ is deterministic.

$$Q(\zeta) = \int_0^\infty e^{-\lambda t(1-\zeta)} dB(t) = e^{-\frac{\lambda}{\mu}(1-\zeta)} = e^{-\rho(1-\zeta)}$$

$$P(\zeta) = \frac{(1-\rho)(1-\zeta)}{1-\zeta e^{\rho(1-\zeta)}}$$

$$L_q = \frac{Q''(1)}{2(1-\rho)} = \frac{\rho^2}{2(1-\rho)}.$$

2. $M/M/1$

$$Q(\zeta) = \int_0^\infty e^{-\lambda t(1-\zeta)} \mu e^{-\mu t} dt = \frac{\mu}{\mu + \lambda(1-\zeta)} = \frac{1}{1+\rho(1-\zeta)}$$

$$P(\zeta) = \frac{(1-\rho)(1-\zeta)}{1-\zeta(1+\rho(1-\zeta))} = \frac{1-\rho}{1-\rho\zeta}$$

$$L_q = \frac{\rho^2}{(1-\rho)}$$

Therefore

$$L_q(M/D/1) = \frac{1}{2} L_q(M/M/1).$$

3. $M/E_k/1$ For the Erlang distribution the mean is $\frac{k}{\mu}$ and the variance is $\frac{k}{\mu^2}$. After replacing μ by $k\mu$, the mean is $\frac{1}{\mu}$ and the variance is $\frac{1}{k\mu^2}$.

$$Q(\zeta) = \frac{1}{(1+\frac{\rho}{k}(1-\zeta))^k}$$

$$L_q = \frac{1+k}{2k} \frac{\rho^2}{(1-\rho)}$$

Another improtant result is

$$Q''(1) = \lambda^2\sigma^2 + \rho^2$$

where σ is the standard deviation from the service time. Then the Pollakzek-Khinchin formula takes the form

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

Let us summarize the results:

Model $M/D/1$: $\sigma = 0$, $L_q = \frac{\rho^2}{2(1-\rho)}$.
Model $M/M/1$: $\sigma = \frac{1}{\mu}$, $L_q = \frac{\rho^2}{1-\rho}$.
Model $M/E_k/1$: $\sigma^2 = \frac{1}{k\mu^2}$, $L_q = \frac{k+1}{2k}\frac{\rho^2}{1-\rho}$.
!Note, that

$$L_q(M/E_k/1) \to L_q(M/D/1), k \to \infty$$

!Observe, that

$$L_q(M/D/1) < L_q(M/E_k/1) < L_q(M/M/1).$$

7.2 Model $M/G/\infty$

λ=arrival rate of the customers
$B(t)$=*cdf* of the service time
There is noone waiting. Let us denote:
$A(t)$=number of customers arriving in $(0,t)$
$N(t)$=number of present customers at time t

$$P(A(t) = k) = \frac{e^{-\lambda t (\lambda t)^k}}{k!}$$

$r(t)$=P(a person who arrives in $(0,t)$ is still in the system)

$$P(N(t) = n) = \sum_{k=0}^{\infty} P(N(t) = n|A(t) = k)P(A(t) = k) =$$

$$= \sum_{k=n}^{\infty} \binom{k}{n} (r(t))^n (1 - r(t))^{k-n} \frac{e^{-\lambda t}(\lambda t)^k}{k!}$$

Consider a Poisson process $X(t)$ and assume $X(t) = 1$. The *pdf* of an arrival in $(u, u + \triangle u)$, given that $X(t) = 1$ is

$$\frac{\lambda e^{-\lambda u} e^{-\lambda(t-u)}}{\lambda t e^{-\lambda t}} = \frac{1}{t}$$

where $0 < u < t$ and there is one arrival till time u and no arrival in the interval between the times u and t. If s is the service time, then

$$B(t) = P(s \le t) \to 1 - B(t) = P(s > t) = P(s > t - u)$$

$$r(t) = \frac{1}{t}\int_0^t (1 - B(t-u))du$$

$$P(N(t) = n) = e^{-\lambda t} \sum_{k=n}^{\infty} \binom{k}{n} (\lambda t r(t))^n \frac{(\lambda t(1 - r(t))^{k-n}}{k!} = e^{-\lambda t r(t)} \frac{(\lambda t r(t))^n}{n!}$$

Therefore the mean is:

$$\lambda t r(t) = \lambda \int_0^t (1 - B(t-u))du = \lambda \int_0^t (1 - B(u))du \to \frac{\lambda}{\mu}, \text{ for } t \to \infty$$

Example. If $B(u) = 1 - e^{-\mu u}$. We get $M/M/\infty$ queue.

7.3 Priority Queue

NPRP=nonpreemptive priority, PRP=preemptive priority

In a nonpreemptive rule, a lower priority customer who is getting service will complete his service even if a higher order priority customer arrives during his service time. There are m priority classes: $1, 2, ..., m$. 1 has highest priority, and m has lowest priority.

λ_i=arrival rate; $1/\mu_i$=mean service time. Assume general service and one server.

$\rho_k = \frac{\lambda_k}{\mu_k}$. Assume $\rho_1 + \rho_2 + ... + \rho_m < 1$.

Denote $S_k = \rho_1 + \rho_2 + ... + \rho_k$.

${W_q}^{(k)}$=average waiting time of the k-th priority customer in the queue is

$$\frac{\sum_{i=1}^{m} \lambda_i \Big({\sigma_i}^2 + \frac{1}{\mu_i} \Big)}{2(1 - S_{k-1})(1 - S_k)}, S_0 = 0.$$

One server queue with two priority classes We consider now a one server queue with two priority classes. Again we have Poisson arrivals and exponential service times.

λ_1, λ_2=arrival rates

μ_1, μ_2=service rates

1 has higher priority and assume, that we have nonpreemptive priority.

N_1=number of type 1 customers

N_2=number of type 2 customers

S_0=time required to finish service the item at hand

S_1=total time required to complete the service of all type 1 customers

S_2=total time required to complete the service of all type 2 customers

Note, that $S_0 = 0$ if the system is empty.

${T_q}^{(1)}$=waiting time in queue of a type 1 customer

${T_q}^{(2)}$=waiting time in queue of a type 2 customer

${W_q}^{(1)}$=mean waiting time in queue of a type 1 customer

${W_q}^{(2)}$=mean waiting time in queue of a type 2 customer

The ”combined service distribution” is

$$\frac{\lambda_1}{\lambda_1 + \lambda_2}(1 - e^{-\mu_1 t}) + \frac{\lambda_2}{\lambda_1 + \lambda_2}(1 - e^{-\mu_2 t}).$$

In this formula $\frac{\lambda_i}{\lambda_1 + \lambda_2}, i = 1, 2$ is the probability, that an arrival is type i arrival.

The mean is

$$\frac{\lambda_1}{\lambda_1 + \lambda_2} \frac{1}{\mu_1} + \frac{\lambda_2}{\lambda_1 + \lambda_2} \frac{1}{\mu_2}.$$

The probability, that the system is busy is

$$(\lambda_1 + \lambda_2)\Big(\frac{\lambda_1}{\lambda_1 + \lambda_2} \frac{1}{\mu_1} + \frac{\lambda_2}{\lambda_1 + \lambda_2} \frac{1}{\mu_2} \Big) = \rho_1 + \rho_2,$$

where $\rho_1 = \frac{\lambda_1}{\mu_1}, \rho_2 = \frac{\lambda_2}{\mu_2}$.

$$E(S_0) = E(S_0|\text{the system is busy})P(\text{the system is busy})+$$

$$+E(S_0|\text{the system is idle})P(\text{the system is idle})$$

The second term in this sum is 0.

$$E(S_0|\text{the system is busy}) =$$

$$= E(S_0|\text{the system is busy with type 1 customer})P(\text{the system is busy with type 1 customer})+$$

$$+E(S_0|\text{the system is busy with type 2 customer})P(\text{the system is busy with type 2 customer}) =$$

$$= \frac{1}{\mu_1}\frac{\rho_1}{\rho_1+\rho_2} + \frac{1}{\mu_2}\frac{\rho_2}{\rho_1+\rho_2}$$

$$E(S_0) = \frac{\rho_1}{\mu_1} + \frac{\rho_2}{\mu_2}$$

$$T_q{}^{(1)} = S_0 + S_1$$

$$E(S_1) = E(Y_1 + ... + Y_{N_1}) = \frac{\rho_1}{\mu_1}E(N_1) = \frac{1}{\mu_1}L_q{}^{(1)} = \frac{\lambda_1}{\mu_1}W_q{}^{(1)} = \rho_1 W_q{}^{(1)}$$

$$W_q{}^{(1)} = E(T_q{}^{(1)}) = \frac{\rho_1}{\mu_1} + \frac{\rho_2}{\mu_2} + \rho_1 W_q{}^{(1)}$$

$$W_q{}^{(1)} = \frac{\frac{\rho_1}{\mu_1} + \frac{\rho_2}{\mu_2}}{1-\rho_1}$$

$$T_q{}^{(2)} = S_1' + S_1 + S_2 + S_0$$

Here S_1'=number of customers, who join later, S_1=present customers before from type 1, S_2=present customers before from type 2, S_0=present customers with remaining service.

As before

$$E(S_1) = \rho_1 W_q{}^{(1)}, E(S_2) = \rho_2 W_q{}^{(2)}, E(S_1') = \rho_1 W_q{}^{(2)}$$

$$W_q{}^{(2)} = \rho_1 W_q{}^{(2)} + \rho_2 W_q{}^{(2)} + E(S_0)$$

$$W_q{}^{(2)} = \frac{\rho_1 W_q{}^{(1)} + E(S_0)}{1-\rho_1-\rho_2} = \frac{\frac{\rho_1 E(S_0)}{1-\rho_1} + E(S_0)}{1-\rho_1-\rho_2} = \frac{E(S_0)}{(1-\rho_1)(1-\rho_1-\rho_2)}$$

$$T_q{}^{(3)} = S_1' + S_2' + S_1 + S_2 + S_3 + S_0.$$

7.4 Model $G/M/1$

We consider the process at arrival epochs. Service times are exponentially distributed and interarrival times are independent with *pdf* $a(t)$ and *cdf* $A(t)$. Consider 2 consecutive arrival epochs

$$Y_n, Y_{n+1}, \text{numbers of customers present}$$

$$Y_{n+1} = Y_n + 1 - B_{n+1}, B_{n+1} \leq Y_{n+1}$$

$$E(\zeta^{B_{n+1}}) = \int_0^\infty e^{-\mu t(1-\zeta)} a(t)dt$$

where B_{n+1} is the number of service completions during one arrival time. Y_n is a Markov chain.

$$p_{ij} = P(Y_{n+1} = j | Y_n = i) = P(B_{n+1} = i + 1 - j) = g_{i+1-j} =$$

$$= P(i + 1 - j \text{services are completed during one arrival time}), i + 1 \geq j \geq 1.$$

This number is independent on n.

Put $h_i = 1 - g_i - g_{i+1} - g_{i+2} - ...$ and then the transition probability matrix takes the form

$$P = ((p_{ij})) = \begin{pmatrix} h_0 & g_0 & 0 & ... & 0 \\ h_1 & g_1 & g_0 & ... & 0 \\ ... & ... & ... & ... & ... \end{pmatrix}$$

The row sums in this matrix are 0.

Let B denote the number of service completions during one interarrival time.

$$P(B = n) = \int_0^\infty e^{-\mu t} \frac{(\mu t)^n}{n!} dA(t), n = 0, 1, 2, ...$$

$$g_n = P(B = n)$$

$$G(g) = g_0 + g_1\zeta + ... = \int_0^\infty e^{-\mu t(1-\zeta)} dA(t)$$

$G(\zeta)$ =pdf of number of service completions during one interarrival time; $G(1) = 1$.

$$G'(1) = \int_0^\infty t dA(t) = \frac{\mu}{\lambda} = \frac{1}{\rho}$$

where $1/\lambda$ is the mean of $A(t)$.

$G'(1) > 1$ iff $\rho < 1$.

$\pi = (\pi_0, \pi_1, ...)$, the stationary distribution, exists.

X_n=number of people in the system just after an arrival "seen" by an arrival

$$(\pi_0, \pi_1, ...) = (\pi_0, \pi_1, ...) \begin{pmatrix} h_0 & g_0 & 0 & ... & 0 \\ h_1 & g_1 & g_0 & ... & 0 \\ ... & ... & ... & ... & ... \end{pmatrix}$$

We omit the first equation and obtain the system

$$\begin{cases} \pi_1 = g_0\pi_0 + g_1\pi_1 + g_2\pi_2 + ... \\ \pi_2 = g_0\pi_1 + g_1\pi_2 + g_2\pi_3 + ... \\ ... \\ \pi_n = g_0\pi_{n-1} + g_1\pi_n + g_2\pi_{n+1} + ... \end{cases}$$

Assume, that $\pi_n = C{\zeta_0}^n$.

$$C{\zeta_0}^n = g_0 C{\zeta_0}^{n-1} + g_1 C{\zeta_0}^n + g_2 C{\zeta_0}^{n+1} + ...$$

$$\zeta_0 = g_0 + g_1\zeta_0 + g_2\zeta_0^2 + ...$$

$$G(\zeta_0) = \zeta_0 \rightarrow \text{a functional equation}$$

$$\zeta_0 < 1 \text{iff} G'(1) > 1 \text{iff} \frac{\mu}{\lambda} > 1 \text{iff} \rho < 1$$

$\pi_n = C\zeta_0^n, n = 1, 2, ...,$ but $\sum_{n=0}^{\infty} \pi_n = 1$ and then

$$\pi_n = (1 - \zeta_0)\zeta_0^n$$

We also have, that

$$L_s = \frac{\zeta_0}{1 - \zeta_0}$$

Example 1: Model $M/M/1$

$$G(\zeta) = \int_0^{\infty} e^{-\mu t(1-\zeta)} dA(t)$$

where $A(t) = 1 - e^{-\lambda t}$.

$$G(\zeta) = \int_0^{\infty} e^{-\mu t(1-\zeta)} \lambda e^{-\lambda t} dt = \frac{\lambda}{\lambda + \mu(1 - \zeta)} = \frac{\rho}{\rho + 1 - \zeta}$$

$$G(\zeta) = \zeta \text{ and then} \frac{\rho}{\rho + 1 - \zeta} = \zeta$$

and the roots of this equation are $1, \rho$.

$$L_s = \frac{\rho}{1 - \rho}$$

Example 2: Model $D/M/1$. For this model $D = 1/\lambda$.

$$G(\zeta) = \int_0^{\infty} e^{-\mu t(1-\zeta)} dA(t) = e^{-\frac{\mu}{\lambda}(1-\zeta)} = e^{-\frac{1-\zeta}{\rho}}.$$

Therefore

$$e^{-\frac{1-\zeta}{\rho}} = \zeta$$

Example 3: Model $E_k/M/1$ The Laplace transform of E_k with k and λ is $\left(\frac{\lambda}{s+\lambda}\right)^k$.

The Laplace transform of E_k with k and $k\lambda$ is $\left(\frac{k\lambda}{s+k\lambda}\right)^k$.

$$G(\zeta) = \left(\frac{k\lambda}{\mu(1-\zeta)+k\lambda}\right)^k = \left(\frac{k\rho}{1-\zeta+2\rho}\right)^k = \zeta$$

and that goes to $e^{\frac{1-\zeta}{\zeta}}$ as $k \to \infty$.

For $k = 2$ we obtain

$$\left(\frac{2\rho}{1-\zeta+2\rho}\right)^2 = \zeta$$

From here we obtain the following cubic equation

$$(\zeta - 1)(\zeta^2 - (1+4\rho)\zeta + 4\rho^2) = 0$$

Its(positive) roots are $\zeta_1 = 1$ and $\zeta_0 = \frac{1+4\rho-\sqrt{1+8\rho}}{2}$.

If $\rho < 1$, then $\zeta_0 < 1$. When $\rho = 3/8$, $\zeta_0 = 1/4$.

Remark.

$$\lim_{t\to\infty}(X(t) = j) \neq \lim_{n\to\infty}(X_n = j)$$

$$\lim_{t\to\infty}(X(t) = j) = \begin{cases} 1-\rho & j = 0 \\ \rho(1-\zeta_0)\zeta_0^{i-1} & j = 1, 2, ... \end{cases}$$

$$L_s = \frac{\rho}{1-\zeta_0}$$

7.5 Model $G/G/1$

W_n=waiting time of the n-th customer

S_n=service time of the n-th customer

X_{n+1}=interarrival time between the n-th and the $n+1$-th customer

Then the Lindley's equation holds:

$$W_{n+1} = max_{0,W_n+S_n-X_{n+1}}$$

Let us denote $F_n(x) = P(W_n \leq x)$. Then

$$F_{n+1}(x) = \begin{cases} 0 & \text{if } x < 0 \\ \int_{-\infty}^{x} F_n(x-y)dG(y) \text{if } x \geq 0 \end{cases}$$

$$G(x) = P(S_n - X_{n+1} \leq x)$$

$$\lim_{n\to\infty} F_n(x) = F(x) \text{ satisfies } F(x) = \int_{-\infty}^{x} F_n(x-y)dG(y),$$

which is called Wiener-Hopf equation.

If $\rho < 1$, F is non-defective.

If $\rho > 1$, $F(x) \equiv 0$.

Assume $W_1 = 0$.

$$W_2 = max\{0, W_1 + U_1\} = max\{W_1 + U_1\}, U_n = S_n - X_{n+1};$$

$$W_3 = max\{0, W_2 + U_2\} = max\{0, U_1, U_1 + U_2\};$$

$$W_{n+1} = max\{0, U_1, U_1 + U_2, U_1 + ... + U_n\}.$$

Chapter 8

Deterministic Inventory Models

Introduction Inventory is a list of goods and materials, or those goods and materials themselves, held available in a stock by a business. Inventory are held in order to manage and hide from the customer the fact that manufacture/supply delay is longer than delivey delay, and also to case the effect of imperfections in the manufacturing process that lower production efficiencies if production capacity stands idle for luck of materials.

Costs Involved in Inventory Models We want to minimize the total Inventory cost.

Total Inventory cost = Purchasing cost + Setup cost + Holding cost + Shortage cost

Purchasing cost is the price per unit of an inventory item. At times the item is offered at a discount if the ordered size exceeds a certain amount, which is a factor deciding how much to order.

Setup cost represents the fixed charge incurred, when an oreder is placed regardless of its size. Increasing the order quantity reduces the setup cost, but increases the average inventory level and hence the cost of tied capital. Reducing the order size increases frequency of ordering and the associated setup cost. An inventory cost model balances the two costs.

Holding cost represents the cost of maintaining inventory in stock. It includes interest on capital and the cost of storage, maintenance and handling.

Shortage cost is the penalty incurred when we run out of stock. It includes potential loss of income and the more subjective cost of loss customer's goodwill.

Demand is deterministic!

y = order quantity(the number of units)

D = demand rate(units/unit time)

t = ordering cycle length(time units)

k = setup cost(independent of y)

h = holding cost(unit/unit time)
p = penalty cost/unit/unit time
a = production rate We want to determine y, which minimizes the total cost.

8.1 EOQ(Economic Order Quantity) Model

In this model we have
- constant rate demand.
- instantaneous replenishment
- no shortages

We want to determine y , which minimizes the total cost.

An order of size y is placed and received instantaneously when the inventory model reaches 0.

$$t = \frac{y}{D}$$

$$\text{Total cost per unit time } = TCU(y) =$$

$$= \frac{k + h\frac{1}{2}ty}{t} = \frac{k}{t} + \frac{h}{2}y = \frac{kD}{y} + \frac{h}{2}y.$$

$$\frac{\partial TCU(y)}{\partial y} = -\frac{kD}{y^2} + \frac{h}{2} = 0$$

$$y* = \sqrt{\frac{2kD}{h}} - \text{ Wilson's formula}$$

$$\frac{\partial^2 TCU(y)}{\partial y^2} = \frac{2kD}{y^3} > 0$$

$$t* = \frac{y*}{D} = \sqrt{\frac{2k}{Dh}}$$

and then the minimal total cost is

$$TCU(y*) = \sqrt{2kDh}$$

As k increases, $y*$ increases too.

As h increases, $y*$ decreases.

8.2 EOQ Model with Finite Replenishment

This could be visualized with a triangle ABC, E is a point of AB, $AE = t_1$, $EC = t_2$, $AB = a - D$, $BC = D$.

$$TCU(y) = \frac{k + h\frac{1}{2}tBE}{t}$$

$$BE = (a - D)t_1 = Dt_2$$

$$(a - D)(t_1 + t_2) = Dt_2 + (a - D)t_2 = at_2 \to t_2 = (1 - \frac{D}{a})t \to BE = D(1 - \frac{D}{a})t$$

$$TCU(y) = \frac{kD}{y} + \frac{D(a - D)yh}{2a} = \frac{kD}{y} + \frac{1}{2}h(1 - \frac{D}{a})y$$

$$\frac{\partial TCU(y)}{\partial y} = -\frac{kD}{y^2} + \frac{h}{2}(1 - \frac{D}{a}) = 0$$

$$y* = \sqrt{\frac{2kD}{h}(1 - \frac{D}{a})}.$$

Therefore

$$TCU(y*) = \sqrt{2kDh(1 - \frac{D}{a})}$$

Remark. If $a = \infty$, we get Model 1.

8.3 EOQ Model with Storages

Shortages are filled as soon as the inventory order is realized.

$$y = Q_1 + Q_2, t = t_1 + t_2,$$

where

$$Q_1 = t_1 D, Q_2 = t_2 D = (t - t_1)D = y - Q_1$$

$$TCU(y, Q_1) = \frac{k + h\frac{1}{2}Q_1 t_1 + p\frac{1}{2}Q_2 t_2}{t} = \frac{kD}{y} + \frac{1}{2}h\frac{Q_1^2}{y} + \frac{\frac{1}{2}P(y - Q_1)^2}{y}$$

We are going to determine y and Q_1, so that $TCU(y)$ is minimal.

$$\frac{\partial TCU(y, Q_1)}{\partial Q_1} = h\frac{Q_1}{y}D - \frac{p(y - Q_1)D}{y} = 0$$

$$Q_1 = \frac{py}{h + p}, y - Q_1 = \frac{hy}{h + p}$$

$$TCU(y, Q_1) = \frac{kD}{y} + \frac{1}{2}h\frac{p^2y^2D}{(h + p)^2} + \frac{1}{2}p\frac{h^2y^2D}{(h + p)^2} =$$

$$= \frac{kD}{y} + \frac{1}{2}\frac{y^2}{(h + p)^2}hp(h + p)$$

$$\frac{\partial TCU(y, Q_1)}{\partial y} = -\frac{kD}{y^2} + \frac{\frac{1}{2}}{\frac{1}{h} + \frac{1}{p}} = 0$$

$$y* = \sqrt{2kD(\frac{1}{h} + \frac{1}{p})} > \sqrt{\frac{2kD}{h}}$$

If $p = \infty$ we get Model 1.

8.4 EOQ Model with Storages and Finite Replenishment

$$TCU(y,w) = \frac{kD}{y} + \frac{h\Big(y(1-\frac{D}{a})-w\Big)^2 + pw^2}{2(1-\frac{D}{a})y}$$

$$y* = \sqrt{\frac{2kD(p+h)(1-\frac{D}{a})}{ph}}, w* = \sqrt{\frac{2kDh(p+h)(1-\frac{D}{a})}{p(p+h)}}.$$

8.5 Price Break

An inventory item may be purchased at a discount if the size of the order exceeds q. In the previous Model 4 we have not taken into account the purchase price. We had setup, holding and storage costs.

cy = purchase cost
$\frac{cy}{t} = cD = const$ = purchase cost/unit time(independent of y)
This vanishes when we differentiate $TCU(y)$ with respect to y.
Let us denote:
c = cost per item
c_1 = the cost of an item, if you purchase $\leq q$ items
c_2 = the cost of an item, if you purchase $> q$ items
Of course, $c_1 > c_2$ and q = price break.
So,

$$c = \begin{cases} c_1 & \text{if } y \leq q \\ c_2 & \text{if } y > q \end{cases}$$

$$TCU(y) = \begin{cases} Dc_1 + \frac{kD}{y} + \frac{h}{2}y & \text{if } y \leq q \\ Dc_2 + \frac{kD}{y} + \frac{h}{2}y & \text{if } y > q \end{cases}$$

If we make a graphic(try alone), then If q is in zone 1: y_m is the order quantity.
If q is in zone 2: q is the order quantity(because q is the minimum).
If q is in zone 3: y_m is the order quantity.
Given this situation, find y_m as before.
If $q \leq y_m$, the order quantity is y_m, otherwise determine Q, such that:
- $TCU_1(y_m) = TCU_1(Q)$
or find out if
- $TCU_1(y_m) > or < TCU_2(q)$.

8.6 Multiitem EOQ with Storage Limitation

(This model is related to Portfolio Optimization in Finance)

There are n items competing for limited space. For $i = 1, 2, ..., n$, let:
y_i = order quantity(the number of units)
D_i = demand rate(units/unit time)
k_i = setup cost(independent of y)
h_i = holding cost(unit/unit time)
a_i = storage area
A = maximum available storage
Our task is to minimaze the total cost

$$TCU(y_1, ..., y_n) = \sum_{i=1}^{n} \left(\frac{k_i D_i}{y_i} + \frac{h_i y_i}{2} \right),$$

such that

$$a_1 y_1 + ... + a_n y_n \leq A.$$

Compute $y*_i = \sqrt{\frac{2k_i D_i}{h_i}}$ (EOQ formula). If $\sum a_i y_i \leq A$, these are the optimal solutions. Otherwise use Lagrange multipliers, i.e. compose the function

$$L(\lambda, y_1, ..., y_n) = TCU(y_1, ..., y_n) - \lambda(\sum a_i y_i - A)$$

and find its minimum. Determine $y_1, ..., y_n, \lambda$:

$$\begin{cases} \frac{\partial L}{\partial y_i} = 0, i = 1, 2, ..., n \\ \frac{\partial L}{\partial \lambda} = 0 \end{cases}$$

Solving this system, we obtain:

$$y*_i = \sqrt{\frac{2k_i D_i}{h_i - 2\lambda * a_i}}$$

This is a minimization problem, so $\lambda < 0$. Successively reduce λ till $\sum a_i y_i \approx A$.

8.7 Dynamic EOQ Model with no Setup Cost

There is a planning horizon with n equal periods. Each period has a limited production capacity that can include several production levels(e.g. regular time, overtime). A current period may produce more than its immediate demand to satisfy demand for later periods in which case a holding cost must be changed(transportation problem).

8.8 Dynamic EOQ Model with Setup Costs

There are n periods. For $i = 1, 2, ..., n$:
z_i = amount ordered
D_i = demand for period i
x_i = inventory at the start of period i
k_i = setup cost

$$C_i(z_i) = \begin{cases} 0 & z_i = 0 \\ k_i + c_i(z_i) & z_i > 0 \end{cases} = \text{Production cost}$$

where $c_i(z_i)$=marginal cost.

$$x_{i+1} = x_i + z_i - D_i, 0 \leq x_{i+1} \leq D_{i+1} + D_{i+2} + ... + D_n$$

$$f_1(x_2) = \min_{z_1 = D_1 + x_2 - x_1} \{C_1(z_1) + h_1(x_2)\}$$

$$f_i(x_{i+1}) = \min_{0 \leq z_i \leq D_i + C_i} \{C_i(z_i) + h_i(x_{i+1}) + f_{i-1}(x_i)\} =$$

$$= \min_{0 \leq z_i \leq D_i + C_i} \{C_i(z_i) + h_i(x_{i+1}) + f_{i-1}(x_{i+1} + D_i - z_i)\}$$

$f_i(x_{i+1})$=minimal Inventory cost for periods $1, 2, ..., i$ given the end-of-period inventory x_{i+1}.

We actually break the problem in several smaller problems, find the optimal solutions and then the general solution.

Chapter 9

Probabilistic Inventory Models

Demand is random(which makes our life more difficult!) with specified distribution.

9.1 "Probabilitized" EOQ Model

L = Lead time = time between placing an order and receiving an order
x_L = random demand during lead time
$\mu_L = E(x_L)$
$\sigma_L = SD(x_L)$
B = Buffer stock size
α = maximum allowed probability of running out of a stock
Determine the optimal buffer size.

$$P(x_L > B + \mu_L) \leq \alpha$$

$$1 - F(B + \mu_L) \leq \alpha.$$

If x_L is $N(DL, \sigma^2 L)$

$$P(\frac{x_L - \mu_L}{\sigma_L} > \frac{B}{\sigma_L}) \leq \alpha.$$

Determine B from normal tables.

9.2 Probabilistic EOQ Model

$f(x)$ = pdf of demand during lead time L
D = expected demand/unit time
h = holding cost/unit
p = storage cost/unit
k = setup cost/order
Determine $R*, y*$, such that $TCU(R, y)$ is minimal(R is reorder point, y is ordered amount).

We make some calculations:)

$$y* = \sqrt{\frac{2D(k+pS)}{R}}, \int_{R*}^{\infty} xf(x)dx = \frac{hy*}{pD},$$

$$S = \int_{R}^{\infty} (x-R)f(x)dx,$$

$$R > \int_{0}^{\infty} xf(x)dx \rightarrow \int_{0}^{\infty} (x-R)f(x)dx > 0.$$

Assume

$$\hat{y} = \sqrt{\frac{2D(k+pEx)}{h}} > y' = \frac{pD}{h}.$$

Note that $R = 0 \rightarrow S = Ex$. If $S_0 = 0$ then $y_1 = \frac{2kD}{h}, R_0 = 0 \rightarrow S_1 = Ex$. Then we determine

$$y_2 \rightarrow R_1 \rightarrow S_2 \rightarrow y_3 \rightarrow R_2 \rightarrow S_3 ...$$

and stop when $y_i \approx y_{i+1}$.

If $f(x)$ is uniform(i.e. constant), then S takes a simple form $S = \int_{R}^{\infty} \frac{x-h}{b-a} dx$.

If $f(x)$ is $exp(\alpha)$, then $S = \int_{R}^{\infty} (x-R)e^{-\alpha x} dx$.

9.3 Single-Period without Setup Cost

$f(D)$ = pdf of demand

D = random demand during the period

h = holding cost/unit during the period

p = storage cost/unit during the period

y = order quantity

x = inventory on hand before an order is placed

$C(y)$ = Total cost = Holding cost + Shortage cost, and then for the expectancy we obtain

$$E[C(y)] = h\int_{0}^{y} (y-D)f(D)dD + p\int_{y}^{\infty} (D-y)f(D)dD$$

$$\frac{\partial E[C(y)]}{\partial y} = h\int_{0}^{y} f(D)dD - p\int_{y}^{\infty})f(D)dD = 0$$

$$h\int_{0}^{y} f(D)dD + p\int_{0}^{y})f(D)dD - p = 0$$

$$\int_{0}^{y} f(D)dD = \frac{p}{p+h}$$

Remark. If D is discrete

$$P(D \leq y*-1) \leq \frac{p}{p+h} \leq P(D \leq y*)$$

9.4 Single-Period with Setup Cost($s - S$ Policy)

$$E[C(\hat{y})] = k + E[C(y)].$$

First calculate $y*$, which minimizes $E[C(y)]$.
Let $y* = S$.
$E[C(\hat{y})]$ is minimal for the same $y*$.
Determine $s < S$ such that $E[C(\hat{s})] = E[C(s)]$.
If $x < s$, order $S - x$.
If $x > s$ do not order.