“usepackageamstex12beta.

Stochastic Modeling in Operations Research
(incomplete classnotes)
version 1/17/2007

Contents

Chapter 1

What is Operations Research ?

Contents.

Definitions

Phases of an OR study

Principles of Modeling

Two definitions.

(1) O.R. is concerned with scientifically deciding how best to design and operate *systems*, usually under conditions requiring the allocation of *scarce resources*.

(2) O.R. is a *scientific approach* to *decision making*.

Modern Definition. (suggested but not adopted yet)

Operations research (OR) is the application of scientific methods to improve the effectiveness of operations, decisions and management. By means such as analyzing data, creating mathematical models and proposing innovative approaches, Or professionals develop scientifically based information that gives insight and guides decision making. They also develop related software, systems, services and products.

Clarification.

Or professionals collaborate with clients to design and improve operations, make better decisions, solve problems and advance other managerial functions including policy formulation, planning, forecasting and performance measurement. Clients may be executive, managerial or non-managerial.

These professionals develop information to improve valuable insight and guidance. They apply the most appropriate scientific techniques-selected from mathematics, any of the sciences including social and management sciences, and any branch of engineering. their work normally entails collecting and analyzing data, creating and testing mathematical models, proposing approaches not previously considered, interpreting information, making recommendations, and helping implement the initiatives that result.

Moreover, they develop and help implement software, systems, services and products related to their methods and applications. The systems include strategic decisions-support systems, which play vital role in many organizations. (Reference: Welcome to OR territory by Randy Robinson, ORMS today,pp40-43.)

System. A collection of parts, making up a coherent whole.

Examples. Stoplights, city hospital, telephone switch board, etc.

Problems that are amenable to OR methodologies

(1) Water dams, business decisions like what product to introduce in the market, packaging designs (Markov chains).

(2) Telephone systems, stoplights, communication systems, bank tellers (queueing theory).

(3) Inventory control: How much to stock?

(4) What stocks to buy? When to sell your house? When to perform preventive Maintenance? (Markov decision processes)

(5) How reliable a system is? Examples: car, airplane, manufacturing process.

What is the probability that a system would not fail during a certain period of time?

How to pick a system design that improves reliability?

(6) *Simulation*

∗ Complex systems

∗ *Example.* How to generate random numbers using computers? How to simulate the behavior of a complex communications system?

(7) Linear Programming: How is a long-distance call routed from its origin to its destination?

Phases of an O.R. study

(1) Formulating the problem: Parameters; decisions variables or unknowns; constraints; objective function.

(2) Model construction: Building a mathematical model

(3) Performing the Analysis: (i) solution of the model (analytic, numerical,approximate, simulation,...,etc.) (ii) sensitivity analysis.

(4) Model evaluation: Are the answers realistic?

(5) Implementation of teh findings and updating of the Model

Types of Models

(1) Deterministic (Linear programming, integer programming, network analysis,...,etc)

(2) Probabilistic (Queueing models, systems reliability, simulation, ...etc)

(3) Axiomatic: Pure mathematical fields (measure theory, set theory, probability theory, ... etc)

The Modeling Process.

Real system

Model

Model conclusions

Real conclusions

Principles of Modeling.

(1) All models are approximate; however some are better than others (survival of the fittest).

(2) Do not build a complicated model when a simple one will suffice.

(3) Do not model a problem merely to fit the technique.

(4) The deduction stage must be conducted rigorously.

(5) Models should be validated before implementation.

(6) A model should never be taken too literally (models should not replace reality.

(7) A model cannot be any better than the information that goes into it (GIGO).

Chapter 2

Review of Probability Theory and Random Variables

Contents.

1 Probability Theory

Definitions

Random experiment: involves obtaining observations of some kind

Examples Toss of a coin, throw a die, polling, inspecting an assembly line, counting arrivals at emergency room, etc.

Population: Set of all possible observations. Conceptually, a population could be generated by repeating an experiment indefinitely.

Outcome of an experiment:

Elementary event (simple event): one possible outcome of an experiment

Event (Compound event): One or more possible outcomes of a random experiment

Sample space: the set of all sample points for an experiment is called a sample space; or set of all possible outcomes for an experiment

Notation:

Sample space : Ω

Sample point: ω

Event: A, B, C, D, E *etc.* (any capital letter).

Example. $\Omega = \{w_i, i = 1, \cdots 6\}$, where $w_i = i$. That is $\Omega = \{1, 2, 3, 4, 5, 6\}$. We may think of Ω as representation of possible outcomes of a throw of a die.

More definitions

Union, Intersection and Complementation

Mutually exclusive (disjoint) events

Probability of an event:

Consider a random experiment whose sample space is Ω. For each event E of the sample space Ω define a number $P(E)$ that satisfies the following three axioms (conditions):

(i) $0 \leq P(E) \leq 1$

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

(iii) For any sequence of mutually exclusive (disjoint) events $E_1, E_2, \ldots$,

$$P(\cup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} P(E_i).$$

We refer to $P(E)$ as the probability of the event E.

Examples. Let $\Omega = \{E_1, \ldots, E_{10}\}$. It is known that $P(E_i) = 1/20, i = 1, \ldots, 5$ and $P(E_i) = 1/5, i = 7, \ldots, 9$ and $P(E_{10}) = 3/20$.

Q1: Do these probabilities satisfy the axioms?

A: Yes

Q2: Calculate $P(A)$ where $A = \{E_i, i \geq 6\}$.

A: $P(A) = P(E_6)+P(E_7)+P(E_8)+P(E_9)+P(E_{10}) = 1/20+1/5+1/5+1/5+3/20 = 0.75$

Interpretations of Probability

(i) Relative frequency interpretation: If an experiment in repeated a large number, n, of times and the event E is observed n_E times, the probability of E is

$$P(E) \simeq \frac{n_E}{n}$$

By the $SLLN$, $P(E) = \lim_{n\to\infty} \frac{n_E}{n}$.

(ii) In real world applications one observes (measures) relative frequencies, one cannot measure probabilities. However, one can estimate probabilities.

(iii) At the conceptual level we *assign* probabilities to events. The assignment, however, should make sense. (e.g. $P(H) = .5, p(T) = .5$ in a toss of a fair coin).

(iv) In some cases probabilities can be a measure of belief (subjective probability). This *measure of belief* should however satisfy the axioms.

(v) Typically, we would like to assign probabilities to simple events directly; then use the laws of probability to calculate the probabilities of compound events.

Laws of Probability

(i) Complementation law

$$P(E^c) = 1 - P(E)$$

(ii) Additive law

$$P(E \cup F) = P(E) + P(F) - P(E \cap F)$$

Moreover, if E and F are mutually exclusive

$$P(E \cup F) = P(E) + P(F)$$

Conditional Probability
Definition If $P(B) > 0$, then

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

(iii) **Multiplicative law** (Product rule)

$$P(A \cap B) = P(A|B)P(B)$$

Definition. Any collection of events that is mutually exclusive and collectively exhaustive is said to be a *partition* of the sample space Ω.

(iv) Law of total probability

Let the events $A_1, A_2, \ldots, A_n$ be a partition of the sample space Ω and let B denote an arbitrary event. Then

$$P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i).$$

Theorem 1.1 (Bayes' Theorem) *Let the events $A_1, A_2, \ldots, A_n$ be a partition of the sample space Ω and let B denote an arbitrary event, $P(B) > 0$. Then*

$$P(A_k|B) = \frac{P(B|A_k)P(A_k)}{\sum_{i=1}^{n} P(B|A_i)P(A_i)}.$$

Special case. Let the events A, A^c be a partition of the sample space Ω and let B denote an arbitrary event, $P(B) > 0$. Then

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}.$$

Remarks.

(i) The events of interest here are A_k, $P(A_k)$ are called *prior* probabilities, and $P(A_k|B)$ are called *posterior* probabilities.

(ii) Bayes' Theorem is important in several fields of applications.

Independence

(i) Two events A and B are said to be *independent* if

$$P(A \cap B) = P(A)P(B).$$

(ii) Two events A and B that are not independent are said to be *dependent*.

Random Sampling

Definition. A sample of size n is said to be a *random sample* if the n elements are selected in such a way that every possible combination of n elements has an equal probability of being selected.

In this case the sampling process is called *simple random sampling.*

Remarks. (i) If n is large, we say the random sample provides an honest representation of the population.

(i) Tables of random numbers may be used to select random samples.

2 Discrete Random Variables

A *random variable* (r.v.) X is a real valued function defined on a sample space Ω. That is a random variable is a rule that assigns probabilities to each possible outcome of a random experiment.

Probability mass function (*pmf*)

For a discrete r.v., X, the function $f(x)$ defined by $f(x) = P(X = x)$ for each possible x is said to be a **Probability mass function (*pmf*)**

Probability distribution function (*cdf*)

The *cdf*, $F(x)$, of a discrete r.v., X, is the real valued function defined by the equation

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

Proposition 2.1 *Let X be a discrete r.v. with pmf $f(x)$. Then*

(i) $0 \leq f(x) \leq 1$, *and*

(ii) $\sum_x f(x) = 1$

where the summation is over all possible values of x.

Relations between *pmf* and *cdf*

(i) $F(x) = \sum_{y \leq x} f(y)$ for all x

(ii) $f(x) = F(x) - F(x-)$ for all x .

(iii)

$$P(a < X \leq b) = F(b) - F(a).$$

Properties of Distribution Functions

(i) F is a non-decreasing function; that is if $a < b$, then $F(a) \leq F(b)$.

(ii) $F(\infty) = 1$ and $F(-\infty) = 0$.

(iii) F is right-continuous.

Expected Value and Variance

$$\begin{aligned} E[X] &= \sum_x xP(X=x) \\ &= \sum_x xf(x). \end{aligned}$$

$$E[g(X)] = \sum_x g(x)f(x)$$

$$\mu_k = E[X^k] = \sum_x x^k f(x), k = 1, 2, \ldots .$$

Variance.

$$V(X) = E[(X-\mu)^2] = \sum_x (x-\mu)^2 f(x)$$

A short cut for the variance is

$$V(X) = E[X^2] - (E[X])^2$$

Notation: Sometimes we use $\sigma^2 = V(X)$.

$$\sigma_X = \sqrt{V(X)} = \sqrt{E[(X-\mu)^2]} = \sqrt{\sum_x (x-\mu)^2 f(x)}$$

3 Continuous Random Variables

Probability distribution function (cdf)

A random variable X is said to be *continuous* if its cdf is a continuous function such that

$$\begin{aligned} F_X(x) &= P(X \le x) = \int_{-\infty}^{x} f_X(t)dt \; ; \\ f_X(t) &\ge 0 \; ; \\ \int_{-\infty}^{\infty} f_X(t)dt &= 1 \; . \end{aligned}$$

For a continuous r.v., X, the function $f_X(x)$ is said to be a *probability density function (pdf)*

Proposition 3.1 *Let X be a continuous r.v. with pdf $f(x)$. Then*
(i) $0 \le f(x) \le 1$, and
(ii) $\int_x f(x)dx = 1$
where the integral is over the range of possible values of x.

Useful Relationships

(i)

$$P(a \le X \le b) = F(b) - F(a).$$

(ii)$P(X = x) = F(x) - F(x-) = 0$ for all x .

(iii) $f(x) = F'(x)$ for all x at which f is continuous.

Properties of Distribution Function

(i) F is a non-decreasing function; that is if $a < b$, then $F(a) \le F(b)$.

(ii) $F(\infty) = 1$ and $F(-\infty) = 0$.

(iii) F is continuous.

Expected Value and Variance

$$\mu = E[X] \quad = \quad \int_{-\infty}^{+\infty} xf(x)dx$$

$$E[g(X)] = \int_{-\infty}^{+\infty} g(x)f(x)dx$$

$$\mu_k = E[X^k] = \int_{-\infty}^{+\infty} x^k f(x)dx, k = 1, 2, \ldots.$$

Variance.

$$V(X) = E[(X-\mu)^2] = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)dx$$

Shortcut Formula

$$V(X) = E[X^2] - E[X]^2 = E[X^2] - \mu^2$$

$$\sigma_X = \sqrt{V(X)} = \sqrt{E[(X-\mu)^2]}$$

4 Review of Probability Distributions

Contents.

Poisson distribution

Geometric distribution

Exponential distribution

Poisson.

The Poisson pmf arises when counting the number of events that occur in an interval of time when the events are occurring at a constant rate; examples include number of arrivals at an emergency room, number of items demanded from an inventory; number of items in a batch of a random size.

A rv X is said to have a *Poisson pmf* with parameter $\lambda > 0$ if

$$f(x) = e^{-\lambda}\lambda^x/x!, x = 0, 1, \ldots.$$

Mean: $E[X] = \lambda$
Variance: $V(X) = \lambda$, $\sigma_X = \sqrt{\lambda}$
Example. Suppose the number of typographical errors on a single page of your book has a Poisson distribution with parameter $\lambda = 1/2$. Calculate the probability that there is at least one error on this page.
Solution. Letting X denote the number of errors on a single page, we have

$$P(X \geq 1) = 1 - P(X = 0) = 1 - e^{-0.5} \simeq 0.395$$

Geometric

The geometric distribution arises in situations where one has to wait until the first success. For example, in a sequence of coin tosses (with p =P(head)), the number of *trials*, X, until the first head is thrown is a geometric rv.

A random variable X is said to have a *geometric pmf* with parameter p, $0 < p < 1$, if

$$P(X = n) = q^{n-1}p \ \ (n = 1, 2, \ldots; p > 0, q = 1 - p) \ .$$

Properties.
(i) $\sum_{n=1}^{\infty} P(X = n) = p \sum_{n=1}^{\infty} q^{n-1} = p/(1 - q) = 1.$
(ii) **Mean:** $E[X] = \frac{1}{p}$
(iii) **Second Moment:** $E[X^2] = \frac{2}{p^2} - \frac{1}{p}$
(iv) **Variance:** $V(X) = \frac{q}{p^2}$
(v) **CDF Complement:** $P(X \geq k) = q^{k-1}$
(iv) **Memoryless Property:** $P(X = n + k|X > n) = P(X = k)$; k=1,2,....

Modified Geometric Distribution

For example, in a sequence of coin tosses (with p =P(head)), the number of *tails*, X, until the first head is thrown is a geometric rv. A random variable X is said to have a *geometric pmf* with parameter p, $0 < p < 1$, if

$$P(X = n) = q^n p \ \ (n = 0, 1, \ldots; p > 0, q = 1 - p) \ .$$

Properties.
(i) $\sum_{n=0}^{\infty} P(X = n) = p \sum_{n=0}^{\infty} q^n = p/(1 - q) = 1.$
(ii) **Mean.** $E[X] = \frac{q}{p}$
(iii) **Second Moment.** $E[X^2] = \frac{q}{p^2} + \frac{q^2}{p^2}$

(iv) **Variance.** $V(X) = \frac{q}{p^2}$
(v) **CDF Complement.** $P(X \geq k) = q^k$
(iv) **Memoryless Property.** $P(X = n + k | X > n) = P(X = k)$.

Exponential.

The exponential pdf often arises, in practice, as being the distribution of the amount of time until some specific event occurs. Examples include time until a new car breaks down, time until an arrival at emergency room, ... etc.

A rv X is said to have an *exponential* pdf with parameter $\lambda > 0$ if

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

Example. Suppose that the length of a phone call in minutes is an exponential rv with parameter $\lambda = 1/10$. If someone arrives immediately ahead of you at a public telephone booth, find the probability that you will have to wait (i) more than 10 minutes, and (ii) between 10 and 20 minutes.

Solution Let X be the be the length of a phone call in minutes by the person ahead of you.

(i)

$$P(X > 10) = \overline{F}(10) = e^{-\lambda x} = e^{-1} \simeq 0.368$$

(ii)

$$P(10 < X < 20) = \overline{F}(10) - \overline{F}(20) = e^{-1} - e^{-2} \simeq 0.233$$

Properties
(i) Mean: $E[X] = 1/\lambda$
(ii) Variance: $V(X) = 1/\lambda^2$, $\sigma = 1/\lambda$
(iii) CDF: $F(x) = 1 - e^{-\lambda x}$.
(iv) Memoryless Property

Definition 4.1 *A non-negative random variable is said to be memoryless if*

$$P(X > h + t | X > t) = P(X > h) \ \text{ for all } \quad h, t \geq 0.$$

Proposition 4.2 *The exponential rv has the memoryless property*

Proof. The memoryless property is equivalent to

$$\frac{P(X > h + t; X > t)}{P(X > t)} = P(X > h)$$

or

$$P(X > h + t) = P(X > h)P(X > t)$$

or

$$\overline{F}(h + t) = \overline{F}(h)\overline{F}(t)$$

For the exponential distribution,

$$\overline{F}(h + t) = e^{-\lambda(h+t)} = e^{-\lambda h}e^{-\lambda t} = \overline{F}(h)\overline{F}(t) \ .$$

Converse The exponential distribution is the only continuous distribution with the memoryless property.

Proof. Omitted.

(v) Hazard Rate

The hazard rate (sometimes called the failure rate) function is defined by

$$h(t) = \frac{f(t)}{1 - F(t)}$$

For the exponential distribution

$$\begin{aligned} h(t) &= \frac{f(t)}{1 - F(t)} \\ &= \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} \\ &= \lambda \ . \end{aligned}$$

(vi) Transform of the Exponential Distribution

Let $E[e^{-\theta X}]$ be the Laplace-Stieltjes transform of X, $\theta > 0$. Then

$$\begin{aligned} E[e^{-\theta X}] &:= \int_0^\infty e^{-\theta x} dF(x) \\ &= \int_0^\infty e^{-\theta x} f(x) dx \\ &= \frac{\lambda}{\lambda + \theta} \end{aligned}$$

(vii) Increment Property of the Exponential Distribution

Definition 4.3 *The function f is said to be $o(h)$ (written $f = o(h)$) if*

$$\lim_{h \to 0} \frac{f(h)}{h} = 0.$$

Examples

(i) $f(x) = x$ is not $o(h)$, since

$$\lim_{h\to 0}\frac{f(h)}{h} = \lim_{h\to 0}\frac{h}{h} = 1 \neq 0$$

(ii) $f(x) = x^2$ is $o(h)$, since

$$\lim_{h\to 0}\frac{f(h)}{h} = \lim_{h\to 0}\frac{h^2}{h} = 0$$

Recall:

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$$

FACT.

$$P(t < X < t+h | X > t) = \lambda h + o(h)$$

Proof.

$$\begin{aligned}
P(t < X < t+h|X > t) &= P(X < t+h|X > t) \\
&= 1 - P(X > t+h|X > t) \\
&= 1 - e^{-\lambda h} \\
&= 1 - (1 - \lambda h + \frac{(\lambda h)^2}{2!} \\
&\quad - \frac{(\lambda h)^3}{3!} + \cdots) \\
&= \lambda h + o(h).
\end{aligned}$$

(viii) Minimum of exponential r.v.s

FACT. Let $X_1, \ldots X_k$ be independent exp (α_i) rvs. Let $X = \min\{X_1, X_2, \ldots, X_k\}$, and $\alpha = \alpha_1 + \ldots + \alpha_k$. Then X has an exponential distribution with parameter α.

Proof.

$$\begin{aligned}
P(X > t) &= P(X_1 > t, \ldots, X_k > t) \\
&= P(X_1 > t) \cdots P(X_k > t) \\
&= e^{-\alpha_1 t} \cdots e^{-\alpha_k t} \\
&= e^{-(\alpha_1 + \ldots \alpha_k)t} \\
&= e^{-\alpha t}.
\end{aligned}$$

Chapter 3

Markov Chains

Contents.

1 Stochastic Processes

A stochastic process, $\{X_t; t \in T\}$, T an index set, is a collection (family) of random variables; where T is an index set and for each t, $X(t)$ is a random variable.

Let S be the state space of $\{X_t; t \in T\}$. Assume S is countable.

Interpretation. A stochastic process is a representation of a system that evolved over time.

Examples. Weekly inventory levels, demands.

$*$ At this level of generality, stochastic processes are difficult to analyze.

Index set T

$$\begin{array}{rll} & T = [0, \infty) & \text{continuous time} \\ \text{or} & T = [0, 1, 2, 3, \cdots) & \text{discrete time} \end{array}$$

Remarks.

(i) We interpret t as time, and call $X(t)$ the state of the process at time t.

(ii) If the index set T a countable set, we call X a discrete-time stochastic process.

(iii) If T is continuous, we call we call X a continuous-time stochastic process.

(iv) Any realization of X is called a sample path. (i.e. the number of customers waiting to be served).

2 Markov Chains

Definition. A discrete-time stochastic process $\{X_n, n = 0, 1, \ldots\}$ with integer state space is said to be a Markov chain (M.C.) if it satisfies the Markovian property, i.e.

$$P(X_{n+1}=j|X_0=x_0,\dots,X_{n-1}=x_{n-1},X_n=i)$$
$$=P(X_{n+1}=j|X_n=i)$$

for every choice of the non-negative integer n and the numbers $x_0,\dots x_{n-1},i,j$ in $S=I$.
Interpretation. Future is independent of the past, it only depends on the present.
Better statement Future depends on the Past only through the Present (indirectly).
Notation. $p_{ij}(n)=P(X_{n+1}=j|X_n=i)$ are called the one-step transition probabilities.
Definition. A Markov chain $\{X_n,n=0,1,\dots\}$ is said to have stationary (time-homogeneous) transition probabilities if

$$P(X_{n+1}=j|X_n=i)=P(X_1=j|X_0=i)\equiv p_{ij}$$

for all $n=0,1,\dots$.
Remarks.

(i) A rv is characterized by its distribution.
(ii) A stochastic process is, typically, characterized by its finite dimensional distributions.
(iii) A M.C. is characterized by its initial distribution and its one-step transition matrix.

$$P=\begin{pmatrix} p_{00} & p_{01} & p_{02} & \cdots \\ p_{10} & p_{11} & p_{12} & \cdots \\ \vdots & \vdots & \vdots & \end{pmatrix}$$

Note that all entries in P are non-negative, and all rows add up to 1.

(iv) Chapman-Kolmogorov (C-K) equations.

Let $p_{ij}^{(m)}$ be the m-step transition probability, i.e.

$$p_{ij}^{(m)}=P(X_m=j|X_0=i)$$

Then the C-K equations are

$$p_{ij}^{(m)}=\sum_r p_{ir}^{(m-k)}p_{rj}^{(k)}\ ,(0<k<m)$$

OR in matrix notation

$$P^{(m)}=P^{(m-k)}P^{(k)}$$

FACT. $P^{(m)}=P^m$
Exercise. Write the C-K equations both in algebraic and matrix notation for the following cases: (i) $k=1$, (ii) $k=m-1$.

(v) Stationary Solution.

Let $\pi = (\pi_0, \pi_1, \ldots)$. The solution of $\pi = \pi P, \sum \pi_i = 1$ if it exists is called the *stationary distribution* of the M.C. $\{X_n, n = 0, 1, \ldots\}$.

Interpretation: π_j represents the long-run fraction of time the process spends in state j.

(vi) Transient Solution.

Let $\pi_j^n = P(X_n = j)$ be the unconditional probability that the process is in state j at time n.

Let $\pi^n = (\pi_0^n, \pi_1^n, \ldots)$ be the unconditional distribution at time n. (Note that π^0 is called the inital distribution. Then

$$\pi^n = \pi^{n-1} P$$

Also, we have

$$\pi^n = \pi^0 P^n$$

Example 1. On any particular day Rebecca is either cheerful (c) or gloomy (g). If she is cheerful today then she will be cheerful tomorrow with probability 0.7. If she is gloomy today then she will be gloomy tomorrow with probability 0.4.

(i) What is the transition matrix P?

Solution.

$$P = \begin{pmatrix} 0.7 & 0.3 \\ 0.6 & 0.4 \end{pmatrix}$$

(ii) What is the fraction of days Rebecca is cheerful? gloomy?

Solution. The fraction of days Rebecca is cheerful is the probability that on any given day Rebecca is cheerful. This can be obtained by solving $\pi = \pi P$, where $\pi = (\pi_0, \pi_1)$, and $\pi_0 + \pi_1 = 1$.

Exercise. complete this problem.

Example 2.(Brand Switching Problem)

Suppose that a manufacturer of a product (Brand 1) is competing with only one other similar product (Brand 2). Both manufacturers have been engaged in aggressive advertising programs which include offering rebates, etc. A survey is taken to find out the rates at which consumers are switching brands or staying loyal to brands. Responses to the survey are given below. If the manufacturers are competing for a population of $y = 300,000$ buyers, how should they plan for the future (immediate future, and in the long-run)?

So

$$P = \begin{pmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{pmatrix}$$

Brand Switching Data

This week

Last week	Brand 1	Brand 2	Total
Brand 1	90	10	100
Brand 2	40	160	200

	Brand 1	Brand 2
Brand 1	90/100	10/100
Brand 2	40/200	160/200

Question 1. suppose that customer behavior is not changed over time. If 1/3 of all customers purchased B1 this week.

What percentage will purchase B1 next week?

What percentage will purchase B2 next week?

What percentage will purchase B1 two weeks from now?

What percentage will purchase B2 two weeks from now?

Solution. Note that $\pi^0 = (1/3, 2/3)$, then

$$(\pi_1^1, \pi_2^1) = (\pi_1^0, \pi_2^0)P$$

$$(\pi_1^1, \pi_2^1) = (1/3, 2/3)\begin{pmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{pmatrix} = (1.3/3, 1.7/3) = (.43, .57)$$

B1 buyers will be $300,000(1.3/3) = 130,000$

B2 buyers will be $300,000(1.7/3) = 170,000$.

Two weeks from now: exercise.

Question 2. Determine whether each brand will eventually retain a constant share of the market.

Solution.

We need to solve $\pi = \pi P$, and $\sum_i \pi_i = 1$, that is

$$(\pi_1, \pi_2) = (\pi_1, \pi_2)\begin{pmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{pmatrix}$$

and

$$\pi_1 + \pi_2 = 1$$

Matrix multiplication gives

$$\begin{aligned}\pi_1 &= 0.9\pi_1 + 0.2\pi_2 \\ \pi_2 &= 0.1\pi_1 + 0.8\pi_2 \\ \pi_1 + \pi_2 &= 1\end{aligned}$$

One equation is redundant. Choose the first and the third. we get

$$0.1\pi_1 = 0.2\pi_2 \quad \text{and} \quad \pi_1 + \pi_2 = 1$$

which gives

$$(\pi_1, \pi_2) = (2/3, 1/3)$$

Brand 1 will eventually capture two thirds of the market $(200,000)$ customers.

Camera Store Example (textbook)

Scenario. Suppose D_1, D_2, D_3, ... represent demands for weeks 1, 2, 3, ... Let X_0 be the # of cameras on hand at the end of week 0, i.e. beginning of week 1; and X_1, X_2, $X_3 =$ be the # of cameras on hand at the end of week 1, 2, 3, ...

On Saturday night, the store places an order that is delivered on time Monday morning.

Ordering policy: (s,S) policy, i.e. order up to S if inventory level drops below s; otherwise, do not order. In this example we have a $(1, 3)$ policy.

Suppose D_1, D_2, D_3 are *iid*

$$X_{n+1} = \begin{cases} \max\{(3 - D_{n+1}), 0\} & \text{if } X_n < 1, \\ \max\{(X_n - D_{n+1}), 0\}, & \text{if } X_n \geq 1, \end{cases}$$

for all $n = 0, 1, 2, \ldots$.

For $n = 0, 1, 2, \ldots$,

(i) $X_n = 0$ or 1 or 2 or 3.

(ii) Integer state process

(iii) Finite state process

Remarks

(1) X_{n+1} depends on X_n

(2) X_{n+1} does not depend on X_{n-1} directly i.e. it depends on X_{n-1} only through X_n.

(3) If D_{n+1} is known, X_n gives us enough information to determine X_{n+1}. (Markovian Property).

RECALL: Markov Chains

A discrete time stochastic process is said to be a Markov chain if

$$\begin{aligned} &P\{X_{n+1} = j | X_0 = x_0, X_1 = x_1, \ldots, X_{n-1} = x_{n-1}, X_n = i\} \\ &\qquad = P\{X_{n+1} = j | X_n = i\} = P\{X_1 = j | X_0 = i\}, \end{aligned}$$

for $n = 0, 1, \ldots$ and every sequence $i, j, x_0, x_1, \ldots, x_{n-1}$.

Definition. A stochastic process $\{X_n\}$ $(n = 0, 1, \ldots)$ is said to be a *finite* state Markov chain if it has

(1) A finite # of states

(2) The Markovian property

(3) Stationary transition probabilities

(4) A set of initial probabilities

$$\pi_i^{(0)} = P\{X_0 = i\} \text{ for all } i = 0, 1, \ldots, m.$$

Example. (Inventory example)

One-step transition matrix (Stochastic Matrix)

$$P = \begin{bmatrix} p_{00} & p_{01} & p_{02} & p_{03} \\ p_{10} & p_{11} & p_{12} & p_{13} \\ p_{20} & p_{21} & p_{22} & p_{23} \\ p_{30} & p_{31} & p_{32} & p_{33} \end{bmatrix}$$

In general

$$P = \begin{bmatrix} p_{00} & p_{01} & & p_{0m} \\ p_{10} & p_{11} & \cdots & p_{1m} \\ \vdots & & & \\ p_{m0} & p_{m1} & \cdots & p_{mm} \end{bmatrix}$$

$$p_{i,j} \geq 0 \text{ for all } i, j = 0, 1, \ldots, m.$$

$$\sum_{j=0}^{m} p_{i,j} = 1 \text{ for } i = 0, 1, 2, \ldots, m.$$

Remark. The one-step transition matrix **P** and the distribution of X_0 completely describe the underlying *Markov Chain* $\{X_n, n = 0, 1, 2, \ldots\}$ provided X_n is stationary.

n-step transitions.

$$p_{ij}^{(n)} \stackrel{\text{def}}{=} P\{X_{k+n} = j | X_k = i\} \stackrel{\text{stationarity}}{=} P\{X_n = j | X_0 = i\}$$

n-step transition matrix

$$\mathbf{P} = \begin{bmatrix} p_{00}^{(n)} & p_{01}^{(n)} & & p_{0m}^{(n)} \\ p_{10}^{(n)} & \cdots & \cdots & p_{1m}^{(n)} \\ \\ p_{m0}^{(n)} & \cdots & \cdots & p_{mm}^{(n)} \end{bmatrix}$$

s.t.

$$p_{ij}^{(n)} \geq 0 \text{ for all } \; i, j = 0, 1, \ldots, m, n = 0, 1, 2, \ldots$$

and

$$\sum_{j=0}^{m} p_{ij}^{(n)} = 1 \text{ for all } \; i = 0, 1, \ldots, m, n = 0, 1, 2, \ldots$$

Example. In Inventory Example, assume that D_n has a Poisson distribution with parameter $\lambda = 1$. i.e.

$$\begin{aligned} p\{D_n = k\} &= \tfrac{e^{-\lambda}\lambda^k}{k!} = \tfrac{1}{k!e} \\ p\{D_n \geq 3\} &= 1 - \textstyle\sum_{k=0}^{2} \tfrac{1}{k!e} = 0.080 \\ p\{D_n = 0\} &= 0.368 \\ p\{D_n = 1\} &= 0.368 \\ p\{D_n = 2\} &= 0.184 \\ p\{D_n \geq 1\} &= 1 - p(D_n = 0\} = 0.632 \end{aligned}$$

$$P = \begin{bmatrix} 0.080 & 0.184 & 0.368 & 0.368 \\ 0.632 & 0.368 & 0 & 0 \\ 0.264 & 0.368 & 0.368 & 0 \\ 0.080 & 0.184 & 0.368 & 0.368 \end{bmatrix}$$

Details of calculations:

$$
\begin{aligned}
p_{00} &= p(D_n \geq 3) = 0.080 \\
p_{01} &= p(D_n = 2) = 0.184 \\
p_{02} &= p(D_n = 1) = 0.368 \\
p_{03} &= p(D_n = 0) = 0.368 \\
p_{10} &= p(D_n \geq 1) = 0.632 \\
p_{21} &= p(D_n = 1) = 0.368
\end{aligned}
$$

Classification of States

(1) State j is accessible (reachable) from state i if there exists some sequence of possible transitions which would take the process from state i to state j. ($i \rightsquigarrow j$)

(2) Two states *communicate* if each state is reachable from the other. ($i \rightsquigarrow j$ and $j \rightsquigarrow i$)

(3) A Markov chain is said to be *irreducible* if all its states communicate. That is $\exists\ n$ s.t. $P_{ij}^{(n)} > 0$ for all pairs (i, j). (recurrent M. Chain).

(4) The period of state k is the GCD of all integers n s.t. $P_{kk}^{(n)} > 0$.

(5) If $GCD = 1$, the chain is said to be aperiodic.

Example.

$$\begin{bmatrix} .5 & .3 & .2 \\ .6 & .2 & .2 \\ .1 & .8 & .1 \end{bmatrix}$$

irreducible and aperiodic process

(6) Recurrent states vs. transient states

Example 1

$$\begin{bmatrix} + & + & 0 & 0 \\ + & + & 0 & 0 \\ + & + & 0 & + \\ 0 & 0 & + & 0 \end{bmatrix}$$

0 and 1 are recurrent states

2 and 3 are transient states

Example 2.

1 absorbing state

2 and 3 recurrent states

4 transient state

Example 3. (random walk)

Let Y_0, Y_1, Y_2, ... be *iid* r.v.s. with common distribution given by

$$Y = \begin{cases} 1 & \text{w.p.} \quad p \\ -1 & \text{w.p.} \quad 1-p \end{cases}$$

i.e.

$$\begin{aligned} P\{Y = 1\} &= p \quad \text{and} \\ P\{Y = -1\} &= 1 - p \end{aligned}$$

Let $X_n = Y_0 + Y_1 + \cdots + Y_n$, then $\{X_n, n \geq 0\}$ is called a *simple random walk*.

The random walk $Z_n = \max(X_n, 0)$ has the following representation:

Matrix representation

$$\begin{bmatrix} 1-p & p & & \cdots & & \\ 1-p & 0 & p & & \cdots & \\ \vdots & 1-p & 0 & p & & \cdots \\ & \vdots & 1-p & 0 & p & \\ & & \ddots & \ddots & \ddots & \\ & & & & & \end{bmatrix}$$

3 More M.C. Examples

Example 1 (Machine Reliability)

A machine is either up or down. If it is up at the beginning of a day, then it is up at the beginning of the next day with probability .98 (regardless of the history of the machine); or it fails with probability .02. Once the machine goes down, the company sends a repair person to repair it. If the machine is down at the beginning of a day, it is down at the beginning of the next day with probability .03 (regardless of the history of the machine); or the repair is complete and the machine is up with probability .97. A repaired machine is as good as new. Model the evolution of the machine as a DTMC.

Let X_n be the state of the machine at the beginning of day n, defined as follows:

$$X_n = \begin{cases} 0 & \text{if the machine is down at the beginning of day } n, \\ 1 & \text{if the machine is up at the beginning of day } n. \end{cases}$$

The description of the system shows that $\{X_n, n \geq 0\}$ is a DTMC with state space $\{0, 1\}$ and the following transition probability matrix

0 1

$$P = \begin{matrix} 0 \\ 1 \end{matrix} \begin{bmatrix} .03 & .97 \\ .02 & .98 \end{bmatrix}$$

Example 2 (Follow-up Example 1)

Now suppose the company maintains two such machines that are identical, behave independently of each other, and each has its own repair person.

Model the new two machine systems as a DTMC.

Let Y_n be the number of machines in the "up" state at the beginning of day n.

Is $\{Y_n, n \geq 0\}$ a DTMC?

First we identify the state space of $\{Y_n, n \geq 0\}$ to be $\{0, 1, 2\}$. Next we check if the Markov property holds, that is, if $P(Y_{n+1} = j \mid Y_n = i, Y_{n-1}, \ldots, Y_0)$ depends only on i and j for

$i, j = 0, 1, 2$. For example, consider the case $Y_n = i = 1$ and $Y_{n+1} = j = 0$. Thus, one machine is up (and one is down) at time n. Since both machines are identical, it does not matter which is up and which is down. In order to move to state 0 at time $n + 1$, the down machine must stay down, and the up machine must go down at the beginning of the next day. Since machines are independent, the probability of this happening is $.03 \times .02$ independent of the history of the two machines. Hence we get $P(Y_{n+1} = 0 \mid Y_n = 1, Y_{n-1}, \ldots, Y_0) = P_{1,0} = .03 \times .02 = .0006$. Proceeding in this fashion we construct the following transition probability matrix:

$$
\begin{array}{c} \\ P = \end{array}
\begin{array}{c} \\ 0 \\ 1 \\ 2 \end{array}
\begin{array}{c}
\begin{array}{ccc} 0 & 1 & 2 \end{array} \\
\left[\begin{array}{ccc}
.0009 & .0582 & .9408 \\
.0006 & .0488 & .9506 \\
.0004 & .0392 & .9604
\end{array}\right]
\end{array}
$$

Example 3 (Telecommunication)

The Switch Corporation manufactures switching equipment for communication networks. Communication networks move data from switch to switch at lightning-fast speed in the form of packets, i.e. strings of zeros and ones (called bits). The switches handle data packets of constant lengths, i.e. the same number of bits in each packet. At a conceptual level, we can think of the switch as a storage device where packets arrive from the network users according to a random process. They are stored in a buffer with capacity to store K packets, and are removed from the buffer one-by-one according to a prespecified protocol. Under one such protocol, time is slotted into intervals of fixed length, say a microsecond. If there is a packet in the buffer at the beginning of a slot, it is removed instantaneously. If there is no packet at the beginning of a slot, no packet is removed during the slot even if more packets arrive during the slot. If a packet arrives during a slot and there is no space for it, it is discarded. Model this system as a DTMC.

Let A_n be the number of packets that arrive at the switch during one slot. (Some of these may be discarded). Let X_n be the number of packets in the buffer at the end of the nth slot. Now, if $X_n = 0$, then there are no packets available for transmission at the beginning of the $(n+1)^{\text{st}}$ slot. Hence all the packets that arrive during that slot, namely A_{n+1} are in the buffer at the end of slot, unless $A_{n+1} > K$, in which case the buffer is full at the end of the $(n+1)^{\text{st}}$ slot. Hence $X_{n+1} = \min\{A_{n+1}, K\}$. In case $X_n > 0$, one packet is removed at the beginning of the $(n + 1)^{\text{st}}$ slot and A_{n+1} packets are added during that slot, subject to the capacity limitation. Combining these cases we get

$$
X_{n+1} = \begin{cases} \min\{A_{n+1}, K\} & \text{if } X_n = 0, \\ \min\{X_n + A_{n+1} - 1, K\} & \text{if } 0 < X_n \le K. \end{cases}
$$

Assume that $\{A_n, n \ge 1\}$ is a sequence of *iid* random variables with common probability mass function (pmf)

$$P(A_n = k) = a_k,\ k \geq 0.$$

Under this assumption $\{X_n, n \geq 0\}$ is a DTMC on state space $\{0, 1, \ldots, K\}$. The transition probabilities can be computed as follows: for $0 \leq j < K$,

$$\begin{aligned} P(X_{n+1} = j \mid X_n = 0) &= P(\min(A_{n+1}, K\} = j \mid X_n = 0) \\ &= P(A_{n+1} = j) \\ &= a_j \end{aligned}$$

$$\begin{aligned} P(X_{n+1} = K \mid X_n = 0) &= P(\min(A_{n+1}, K\} = K \mid X_n = 0) \\ &= P(A_{n+1} \geq K) \\ &= \textstyle\sum_{k=K}^{\infty} a_k \end{aligned}$$

Similarly, for $a \leq i \leq k$, and $i - 1 \leq j < K$,

$$\begin{aligned} P(X_{n+1} = j \mid X_n = i) &= P(\min\{X_n + A_{n+1} - 1, K\} = j \mid X_n = i) \\ &= P(A_{n+1} = j - i + 1) \\ &= a_{j-i+1} \end{aligned}$$

Finally, for $1 \leq i \leq K$,

$$\begin{aligned} P(X_{n+1} = K \mid X_n = i) &= P(\min\{X_n + A_{n+1} - 1, K) = K \mid X_n = i) \\ &= P(A_{n+1} \geq K - i + 1) \\ &= \textstyle\sum_{k=K-i+1}^{\infty} a_k \end{aligned}$$

Combining all these cases, and using the notation

$$b_j = \sum_{k=j}^{\infty} a_k$$

we get the following transition probability matrix:

$$P = \begin{bmatrix} a_0 & a_1 & \dots & a_{K-1} & b_K \\ a_0 & a_1 & \dots & a_{K-1} & b_K \\ 0 & a_0 & \dots & a_{K-2} & b_{K-1} \\ \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & & a_0 & b_1 \end{bmatrix}$$

4 Long-run Properties of M.C.

Suppose the M.C. is irreducible and aperiodic.

A. Steady state probabilities

Let

$$\pi_j = \lim_{n \longrightarrow \infty} P_{ij}^{(n)},\ j \in S,$$

and assume the limits exist for all $j \in S$. Then

$$\pi = (\pi_1, \pi_2, \dots)$$

is called the *steady state* distribution of $\{X_n, n \geq 0\}$.

FACT 1. π satisfies the following conditions:

$$(i) \quad \pi_j > 0 \text{ for all } \ j \in S$$

$$(ii) \quad \pi_j = \sum_{j \in S} \pi_i P_{ij},\ j \in S\ (\pi = \pi P).$$

$$(iii) \quad \sum_{j \in S} \pi_j = 1.$$

FACT 2. Any set of numbers that satisfy conditions (i), (ii), and (iii) is said to be a stationary dist. of $\{X_n, n \geq 0\}$.

FACT 3. Let μ_{jj} = expected recurrence time. Then

$\mu_{jj} = \frac{1}{\pi_j}$ or $(\pi_j = \frac{1}{\mu_{jj}})$.

Camera Example

$$\pi = \pi P$$

$$\sum \pi_j = 1$$

$$(\pi_0, \pi_1, \pi_2, \pi_3) = (\pi_0, \pi_1, \pi_2, \pi_3) \begin{bmatrix} 0.08 & 0.184 & 0.368 & 0.368 \\ 0.632 & 0.368 & 0 & 0 \\ 0.264 & 0.368 & 0.368 & 0 \\ 0.08 & 0.184 & 0.368 & 0.368 \end{bmatrix}$$

$$\begin{array}{lllllllll} \pi_0 & = & 0.08\pi_0 & + & 0.632\pi_1 & + & 0.264\pi_2 & + & 0.08\pi_3 \\ \pi_1 & = & 0.184\pi_0 & + & 0.368\pi_1 & + & 0.368\pi_2 & + & 0.184\pi_3 \\ \pi_2 & = & 0.368\pi_0 & + & & + & 0.368\pi_2 & + & 0.368\pi_3 \\ \pi_3 & = & 0.368\pi_0 & + & & & & + & 0.368\pi_3 \end{array}$$

$$\pi_0 + \pi_1 + \pi_2 + \pi_3 = 1$$

Solve $\Longrightarrow$

$$\begin{array}{lll} \pi_0 = 0.285 & \Longrightarrow & \mu_{00} = \frac{1}{0.285} = 3.51 \text{ weeks} \\ \pi_1 = 0.285 & \Longrightarrow & \mu_{11} = \frac{1}{0.285} = 3.51 \\ \pi_2 = 0.264 & \Longrightarrow & \mu_{22} = \frac{1}{\pi_2} = \frac{1}{0.264} = 3.79 \\ \pi_3 = 0.1644 & \Longrightarrow & \mu_{33} = \frac{1}{\pi_3} = \frac{1}{0.1644} = 6.02 \text{ weeks} \end{array}$$

Example Consider a M.C. with

$$P = \begin{array}{c} 0 \\ 1 \\ 2 \end{array} \begin{bmatrix} .6 & .4 & 0 \\ .4 & .6 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Then

$$\begin{aligned} \pi &= \pi P \quad \text{gives} \\ \pi_0 &= .6\pi_0 + .4\pi_1 \Rightarrow \pi_0 = \pi_1 \\ \pi_1 &= .4\pi_0 + .6\pi_1 \Rightarrow \pi_0 = \pi_1 \\ \pi_3 &= \pi_3 \end{aligned}$$

Substitute in

$$\begin{aligned} \pi_0 + \pi_1 + \pi_2 &= 1 \text{ to obtain} \\ 2\pi_0 + \pi_2 &= 1 \end{aligned}$$

$$\begin{aligned} \text{Let } \pi_2 &= \alpha \\ 2\pi_0 &= 1-\alpha \\ \pi_0 &= \frac{1-\alpha}{2} \end{aligned}$$

Then

$$\pi = \left(\frac{1-\alpha}{2}, \frac{1-\alpha}{2}, \alpha\right) \qquad \text{(valid for any } 0 \le \alpha \le 1.)$$

Therefore, the stationary distribution is not unique.

FACT If the steady state distribution *exists*, then it is unique and it satisfies

$$\begin{aligned} \pi &= \pi P \\ \sum \pi_i &= 1. \end{aligned}$$

4.1 Gambler's Ruin Problem (absorbing states)

Consider a gambler who at each play of the game has probability p of winning one dollar and probability $q = 1 - p$ of losing one dollar. Assuming successive plays of the game are independent, what is the probability that, starting with i dollars,$(0 < i < N)$, the gambler's fortune will reach N before reaching 0.

0 1 2 3 N

$$\underline{P} = \begin{matrix} 0 \\ 1 \\ 2 \\ 3 \\ \\ N \end{matrix} \begin{bmatrix} 1 & & & & & N \\ q & & p & & & \\ & q & & p & & \\ & & \ddots & & \ddots & \\ & & & & & \\ N & & & & & 1 \end{bmatrix}$$

Let $\{X_n, n \ge 0\}$ represents gambler's fortune after nth play.

Let $f_{iN} = p\{$ gambler's fortune will eventually reach $N \mid X_0 = i\}$.

FACT

$$f_{iN} = \begin{cases} \frac{1-\left(\frac{q}{p}\right)^i}{1-\left(\frac{q}{p}\right)^N} & p \neq \frac{1}{2} \\ \\ \frac{i}{N} & p = \frac{1}{2}. \end{cases}$$

Letting $N \to \infty$

$$f_{iN} \to \begin{cases} 1 - \left(\frac{q}{p}\right)^i & p > \frac{1}{2} \\ \\ 0 & p \leq \frac{1}{2}. \end{cases}$$

Exercise. Find f_{iN} if $p = .6$, $N = 10$, $i = 5, 6$.
Find f_{i0} (probability of absorption)

Definition.

$$\begin{aligned} \text{Let } f_{ij}^n &= p\{X_n = j, X_k \neq j, k = 1, \ldots, n-1 \mid X_0 = i\}, i.e. \\ f_{ij}^n &= P \text{ (starting in state } i\text{, the first transition into state } j \\ &\quad \text{occurs at time } n) \\ &= P \text{ (the first passage from } i \text{ to } j \text{ is } n). \\ \textit{Note. } f_{ij}^0 &= 0 \end{aligned}$$

Remark.

Let $Y_{ij} \overset{r.v.}{=} \#$ of transitions made by the M.C. in going from state i to state j for the first time

$f_{ij}^{n} =$ P(starting in state i, the first transition into state j occurs at time n)
(= first passage time from i to j)

Y_{ii} =recurrence time of state i .

Then $f_{ij}^{n} = P\{Y_{ij} = n\}$.

Definition.

$$\text{Let } f_{ij} \;=\; \sum_{n=1}^{\infty} f_{ij}^{n}$$

be the P (of ever making a transition into state j | given that the M.C. starts in i)

FACT.

$$f_{ij} > 0 \quad (i \neq j) \quad \text{iff} \quad i \rightarrow j.$$

Definition.

- state j is said to be *recurrent* if $f_{jj} = \sum_{n=1}^{\infty} f_{jj}^n = 1$.
- state j is said to be *transient* if $f_{jj} = \sum_{n=1}^{\infty} f_{jj}^n < 1$.
- state j is said to be *absorbing* if $f_{jj}^n = 1$.

Remarks.

1. Suppose the M.C. starts in state i and i is recurrent. Then the M.C. will eventually re-enter state i with prob. 1. It can be shown that *state i will be visited infinitely often* ($f_{ij} = 1$).

2. Suppose state i is transient. Then

$$\begin{aligned} 1 - f_{ii} &= \text{Prob. state } i \text{ will never be visited again.} \\ f_{ii} &= \text{Prob. state } i \text{ will be visited again.} \end{aligned}$$

$\Longrightarrow$ Starting in state i, the # of visits to state i is a random variable with a geometric distribution, $f(n) = f_{ii}^n(1 - f_{ii})$ with a *finite mean* $= \frac{1}{1-f_{ii}}$.

3. Let

$$I_n = \begin{cases} 1 & \text{if} \quad X_n = i \\ 0 & \text{if} \quad X_n \neq i, \end{cases} \quad \text{then}$$

$$\sum_{n=0}^{\infty} I_n = \# \text{ of periods the M.C. is in state } i.$$

$$\begin{aligned} E\left[\sum_{n=0}^{\infty} I_n \mid X_0 = i\right] &= \sum_{n=0}^{\infty} E\left[I_n \mid X_0 = i\right] \\ &= \sum_{n=0}^{\infty} P(X_n = i \mid X_0 = i) \\ &= \sum_{n=0}^{\infty} p_{ii}^n \end{aligned}$$

Proposition

- state j is recurrent iff $\sum_{n=0}^{\infty} p_{jj}^{n} = \infty$.
- state j is transient iff $\sum_{n=1}^{\infty} p_{jj}^{n} < \infty$.
- state j is absorbing iff $p_{jj}^{n} = 1$.

Corollary

1. A transient state can be visited only a finite number of times.
2. A finite state M.C. cannot have all its states transients.

Calculating $f_{ij}^{(n)}$ recursively

Note that

$$
\begin{aligned}
p_{ij}^{(n)} &= \sum_{k=1}^{n} f_{ij}^{(k)} p_{jj}^{(n-k)} \\
&= \sum_{k=1}^{n-1} f_{ij}^{(k)} p_{jj}^{(n-k)} + f_{ij}^{(n)} p_{jj}^{(0)} \qquad (\text{note: } p_{jj}^{(0)} = 1 \ )
\end{aligned}
$$

FACT

$$
f_{ij}^{(n)} = p_{ij}^{(n)} - \sum_{k=1}^{n-1} f_{ij}^{(k)} p_{jj}^{(n-k)}
$$

Example 1

$$
\begin{aligned}
f_{ij}^{(1)} &= p_{ij}^{(1)} = p_{ij} \\
f_{ij}^{(2)} &= p_{ij}^{(2)} - f_{ij}^{(1)} p_{jj} \\
&\vdots \\
f_{ij}^{(n)} &= p_{ij}^{(n)} - f_{ij}^{(1)} p_{jj}^{(n-1)} - f_{ij}^{(2)} p_{jj}^{(n-2)} \cdots - f_{ij}^{(n-1)} p_{jj}
\end{aligned}
$$

Example 2 for $i = j$

$$\begin{aligned}
f_{jj}^{(1)} &= p_{jj} \\
f_{jj}^{(2)} &= p_{jj}^{(2)} - f_{jj}^{(1)} p_{jj} \\
&\vdots \\
f_{jj}^{(n)} &= p_{jj}^{(n)} - f_{jj}^{(1)} p_{jj}^{(n-1)} - f_{jj}^{(2)} p_{jj}^{(n-2)} - \cdots - f_{jj}^{(n-1)} p_{jj}
\end{aligned}$$

Corollary

If i is recurrent, and $i \Longleftrightarrow j$, then state j is recurrent.

Proof

Since $i \Longleftrightarrow j$, there exists k and m s.t.

$$p_{ij}^{k} > 0 \quad \text{and} \quad p_{ij}^{m} > 0. \qquad \text{Now, for any } n \text{ (integer)}$$

$$\begin{aligned}
p_{ij}^{m+n+k} &\geq p_{ji}^{m} p_{ii}^{n} p_{ij}^{k} \\
\sum_{n=1}^{\infty} p_{jj}^{m+n+k} &\geq \sum_{n=1}^{\infty} p_{ji}^{m} p_{ii}^{n} p_{ij}^{k} \\
&= p_{ji}^{m} p_{ij}^{k} \sum_{n=1}^{\infty} p_{ii}^{n} \\
&= \infty
\end{aligned}$$

because i is recurrent, $p_{ji}^{m} > 0$ and $p_{ij}^{k} > 0$.

Therefore j is also recurrent.

Recall.

$$\begin{aligned}
Y_{ij} &= \text{first passage time from } i \text{ to } j \\
Y_{ii} &= \text{recurrence time of state } i
\end{aligned}$$

$$\begin{aligned}
\text{D.F. of } Y_{ij} \text{ is given by } f_{ij}^{n} &= P(Y_{ij} = n) \\
\text{(we also defined } f_{ij} &= \sum_{n=1}^{\infty} f_{ij}^{n}).
\end{aligned}$$

Now, let

$$\begin{aligned}
\mu_{ij} &= E[Y_{ij}] = \sum_{n=1}^{\infty} n f_{ij}^{(n)} \\
&= \text{expected first passage time from } i \text{ to } j \\
\mu_{jj} &= \text{mean recurrence time} \\
&= \text{expected number of transitions to return to state } j
\end{aligned}$$

FACT

$$\mu_{jj} = \begin{cases} \infty & \text{if } j \text{ is transient} \quad (f_{jj} < 1) \\ \sum_{n=1}^{\infty} n f_{jj}^{n} & \text{if } j \text{ is recurrent} \quad (f_{jj} = 1) \end{cases}$$

Definition

Let j be a recurrent state (equiv. $f_{jj} = 1$), then

state j is positive recurrent if $\mu_{jj} < +\infty$, and
state j is null recurrent if $\mu_{jj} = \infty$.

FACT Suppose all states are recurrent. Then
μ_{ij} satisfies the following equations

$$\mu_{ij} = 1 + \sum_{k \neq j} p_{ik} \mu_{kj}$$

Flow Balance Equations

$$\begin{aligned}
\pi &= \pi P \iff \\
\pi_j &= \sum_i \pi_i p_{ij} \iff \\
\pi_j \sum_k p_{jk} &= \sum_i \pi_i p_{ij}
\end{aligned}$$

Prob. flow out = Prob. flow in

$$\pi_j \sum_k p_{jk} = \sum_i \pi_i p_{ij}$$
$$\pi_j = \sum_i \pi_i p_{ij} \quad (*)$$

verification of (*).

Let

$$C(i,j;n) = \text{number of transition from } i \text{ to } j \text{ during [0,n]}$$
$$Y(i,n) = \text{time in state } i \text{ during } [0,n].$$

$$p_{ij} = \lim_{n\to\infty} \frac{C(i,j;n)}{Y(i,n)}$$
$$\pi_i = \lim_{n\to\infty} \frac{Y(i,n)}{n}$$
$$\sum_i \pi_i p_{ij} = \sum_i \lim_{n\to\infty} \frac{Y(i,n)}{n} \frac{C(i,j;n)}{Y(i,n)}$$
$$= \lim_{n\to\infty} \sum_i \frac{C(i,j;n)}{n}$$
$$= \lim_{n\to\infty} \frac{\sum_i C(i,j;n)}{n}$$
$$= \lim_{n\to\infty} \frac{Y(j,n)}{n} = \pi_j$$

Note.

$$\pi_i p_{ij} = \lim_{n\to\infty} \frac{Y(i,n)}{n} \frac{C(i,j;n)}{Y(i,n)} = \lim_{n\to\infty} \frac{C(i,j;n)}{n} = \text{rate from } i \text{ to } j.$$

4.2 Sums of Independent, Identically Distributed, Random Variables

Let Y_i, $i = 1, 2, \ldots$, be *iid* with

$$P\{Y_i = j\} = a_j, \ j \geq 0$$
$$\sum_{j=0}^{\infty} a_j = 1.$$

If we let $X_n = \sum_{i=1}^{n} Y_i$, then $\{X_n,\ n = 1, 2, \ldots\}$ is a M.C. for which

$$p_{ij} = \begin{cases} a_{j-i} & j \geq i \\ 0 & j < i \end{cases}$$

One possible interpretation of X_n is the following: If Y_i represents demand for a commodity during the ith period, then X_n would represent the total demand for the first n periods.

$$0 \quad 1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6 \quad \cdots$$

$$\underline{P} = \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ \vdots \end{array} \begin{bmatrix} a_0 & a_1 & a_2 & a_3 & a_4 & a_5 & a_6 & \cdots \\ 0 & a_0 & a_1 & a_2 & a_3 & a_4 & a_6 & \cdots \\ 0 & 0 & a_0 & a_1 & a_2 & a_3 & a_4 & \cdots \\ 0 & 0 & 0 & a_0 & a_1 & a_2 & a_3 & \cdots \\ 0 & 0 & 0 & 0 & a_0 & a_1 & a_2 & \cdots \\ 0 & 0 & 0 & 0 & 0 & a_0 & a_1 & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & a_0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \cdots \end{bmatrix}$$

B. Expected Average Cost Per Unit Time

(I) Simple Cost functions.

Model

- M.C. with steady state probabilities
- Let $C(X_n) =$ cost incurred when the process is in state X_n at time n, $n = 0, 1, 2, \ldots$.

Remark.

If $\{X_n, n = 0, 1, 2, \ldots\}$ is a finite state M.C. with states $0, 1, 2, \ldots, M$, then $C(X_n)$ takes any of $M + 1$ values, $C(0)$, $C(1)$, ..., $C(M)$, independently of time n.

Definition.

(i) The expected average cost/unit time is defined as

$$E\left\{\frac{1}{N}\sum_{n=1}^{N} C(X_n)\right\}$$

over the first N periods.

(ii) The long-run expected average cost/unit time is defined as

$$\lim_{N \longrightarrow \infty} E\left\{\frac{1}{N}\left[\sum_{n=1}^{N} C(X_n)\right]\right\}$$

(iii) The long-run (actual) average cost per unit time is defined as

$$\lim_{N \longrightarrow \infty} \left\{ \frac{1}{N} \sum_{n=1}^{N} C(X_n) \right\}$$

FACT. For almost all paths of the M.C.

$$(ii) = (iii) = \sum_{j=0}^{M} \pi_j C(j).$$

i.e.

$$\lim_{N \longrightarrow \infty} \left\{ E\left[\frac{1}{N} \sum_{n=1}^{N} C(X_n) \right] \right\} = \lim_{N \longrightarrow \infty} \frac{1}{N} \sum_{n=1}^{N} C(X_n) = \sum_{j=0}^{M} C(j)\pi_j$$

Camera Example

Let $C(X_n) = h(X_n) =$ holding cost, specifically let

state	$C(i) = h(i)$	π_j	$C(j)\pi_j$
0	0	0.285	(0)(0.285)
1	2	0.285	(2)(0.285)
2	8	0.264	(8)(0.264)
3	18	0.166	(18)(0.166)
		1.00	5.67

Expected average cost per week = EC

$$EC = \sum_{j=0}^{3} C(j)\pi_j =$$

$$= (0)(0.285) + \ldots + (18)(0.166) = 5.67.$$

II Complex Cost functions

Model

(i) Finite state M.C. with steady state probabilities

(ii) Associated with this M.C. is a seq. of *iid* r.v.s. $\{D_n\}$

(iii) $C(X_{n-1}, D_n) =$ cost incurred at time n, for $n = 0, 1, 2, \ldots$,

(cost at time n may depend on demand $D_1, D_2, \ldots$).

(iv) The seq. $(X_0, X_1, \ldots, X_{n-1})$ is indep. of D_n

Now, note that

state	$C(j, D_n)$ is a rv.
0	
1	
2	
3	

Definition. Let

$$k(j) = EC(j, D_n)$$

FACT. Suppose $(i) \ldots (iv)$ hold. Then

$$(i) \lim_{N \longrightarrow \infty} \left\{ E\left[\frac{1}{N}\sum_{n=1}^{N} C(X_{n-1}, D_n)\right]\right\} = \sum_{j=0}^{M} k(j)\pi_j$$

$$(ii) \lim_{N \longrightarrow \infty} \left\{ \frac{1}{N}\sum_{n=1}^{N} C(X_{n-1}, D_n)\right\} = \sum_{j=0}^{M} k(j)\pi_j$$

for almost all paths of the M.C.

Camera Example

Possible costs: ordering cost, holding cost, unsatisfied demand (lost sales).

$$\text{-ordering cost} \quad = \begin{cases} 0 & \text{if } z = 0 \\ 10 + 25z & \text{if } z \geq 1 \end{cases}$$

$z =$ amount ordered

-unsatisfied demand= \$50/unit

-holding cost : ignore

Then

$$\begin{aligned} C(X_{n-1}, D_n) &= 10 + (3)(25) + 50 \max\{(D_n - 3), 0\}, \\ & \qquad X_{n-1} = 0 < 1; \\ &= 0 + 50 \max\{(D_n - X_{n-1}, 0\}, X_{n-1} \geq 1; \end{aligned}$$

for $n = 1, 2, \ldots$.

Need to calculate $k(0), k(1), k(2)$ and $k(3)$. Suppose these computed as follows:

state	$k(j)$	π_j	$k(j)\pi_j$
0	86.2	0.285	(86.2)(0.285)
1	18.4	0.285	(18.4)(0.285)
2	5.2	0.264	(5.2)(0.264)
3	1.2	0.166	(1.2)(0.166)
		1.00	31.4

Expected average cost per week = EC

$$
\begin{aligned}
EC &= \sum_{j=0}^{3} k(j)\pi_j \\
&= (86.2)(0.285) + \ldots + (1.2)(0.166) = 31.4\ .
\end{aligned}
$$

Sample Calculations:

$$
\begin{aligned}
k(0) &= E[C(0, D_n)] \\
&= E[85 + 50\max\{(D_n - 3), 0\}] \\
&= 85 + 50E[\max\{(D_n - 3), 0\}],
\end{aligned}
$$

where

$$
\begin{aligned}
& E[\max\{(D_n - 3), 0\}] \\
= \ & 1P(D = 4) + 2P(D = 5) + 3P(D = 6) + \ldots \\
= \ & \sum_{k=4}^{\infty}(k - 3)P(D = k) \\
= \ & \sum_{k=4}^{\infty} kP(D = k) - 3\sum_{k=4}^{\infty} P(D = k) \\
= \ & E(D) - \sum_{k=1}^{3} kP(D = k) - 3[1 - \sum_{k=0}^{3} P(D = k)] \\
= \ & 1 - 3 - \sum_{k=1}^{3} kP(D = k) + 3[\sum_{k=0}^{3} P(D = k)] \\
= \ & .024\ .
\end{aligned}
$$

Thus,

$$k(0) = 85 + (50)(.024) = 86.2$$

Note: The calculations of other $k(j)'s$ are similar.

Chapter 4

Poisson and Markov Processes

Contents.

1 The Poisson Process

A *stochastic process* is a collection of random variables that describes the evolution of some system over time.

A stochastic process $\{N(t), t \geq 0\}$ is said to be a *counting process* if $N(t)$ represents the total number of *events* that have occurred up to time t. A counting process must satisfy:

(i) $N(t) \geq 0$

(ii) $N(t)$ is integer valued.

(iii) If $s < t$, then $N(s) \leq N(t)$.

(iv) For $s < t$, $N(t) - N(s)$ counts the number of events that have occurred in the interval $(s, t]$.

Definition 1.

(i) A counting process is said to have *independent increments* if the number of events that occur in disjoint time intervals are independent. For example, $N(t + s) - N(t)$ and $N(t)$ are independent.

(ii) A counting process is said to have *stationary increments* if the distribution of the number of events that occur in any time interval depends only on the length of that interval. For example, $N(t + s) - N(t)$ and $N(s)$ have the same distribution.

Definition 2. The counting process $\{N(t), t \geq 0\}$ is said to be a *Poisson process* having rate $\lambda, \lambda > 0$, if:

(i) $N(0) = 0$.

(ii) The process has independent increments.

(iii) The number of events in any interval of length t is Poisson distributed with mean λt. That is for all $s, t \geq 0$,

$$P\{N(t+s) - N(s) = n\} = \frac{e^{-\lambda t}(\lambda t)^n}{n!} \quad , \quad n = 0, 1, \ldots.$$

Interevents: Consider a Poisson process. Let X_1 be the time of the first event. For $n \geq 1$, let X_n denote the time between the $(n-1)st$ and nth event. The sequence $\{X_n, n \geq 1\}$ is called the *sequence of interarrival times.*

Proposition 1.1 *The rvs $X_n, n = 1, 2, \ldots$ are iid exponential rvs having mean $1/\lambda$.*

Remark. A process $\{N(t), t \geq 0\}$ is said to be a *renewal process* if the interevent rvs $X_n, n = 1, 2, \ldots$ are iid with some distribution function F.

Definition 3. The counting process $\{N(t), t \geq 0\}$ is said to be a *Poisson process* having rate $\lambda, \lambda > 0$, if:

(i) $N(0) = 0$;

(ii) $\{N(t), t \geq 0\}$ has stationary indepedent increments

(iii) $P(N(h) = 1) = \lambda h + o(h)$

(iv) $P(N(h) \geq 2) = o(h)$

Remark. The above fact implies that $P(N(h) = 0) = 1 - \lambda h + o(h)$

FACT. Definitions 2 and 3 are equivalent.

Proof. Omitted.

Properties

(i) The Poisson process has stationary increments.

(ii) $E[N(t)] = \lambda t$

(iii) For $s \leq t$

$$P\{X_1 < s | N(t) = 1\} = \frac{s}{t} \; .$$

That is the conditional time until the first event is uniformly distributed.

(iv) The Poisson process possesses the *lack of memory property*

(v) For a Poisson Process $\{N(t), t \geq 0\}$ with rate λ,

$$P(N(t) = 0) = e^{-\lambda t}; \text{ and}$$

$$P(N(t) = k) = \frac{e^{-\lambda t}(\lambda t)^k}{k!} \; , \quad k = 0, 1, \ldots.$$

(vi) Merging two *independent* Poisson processes with rates λ_1 and λ_2 results is a Poisson process with rate $\lambda_1 + \lambda_2$.

(vii) Splitting a Poisson process with rate λ where the splitting mechanism is *memoryless* (Bernoulli) with parameter p , results in two *independent* Poisson processes with rates λp and $\lambda(1-p)$ respectively.

Example Customers arrive in a certain store according to a Poisson process with mean rate $\lambda = 4$ per hour. Given that the store opens at $9:00am$,

(i) What is the probability that exactly one customer arrives by $9:30am$?

Solution. Time is measured in hours starting at $9:00am$.

$$P(N(0.5) = 1) = e^{-4(0.5)}4(0.5)/1! = 2e^{-2}$$

(ii) What is the probability that a total of five customers arrive by $11:30am$?

Solution.

$$P(N(2.5) = 5) = e^{-4(2.5)}[4(2.5)]^5/5! = 10^4e^{-10}/12$$

(iii) What is the probability that exactly one customer arrives between $10:30am$ and $11:00am$?

Solution.

$$\begin{aligned} P(N(2) - N(1.5) = 1) &= P(N(0.5) = 1) \\ &= e^{-4(0.5)}4(0.5)/1! = 2e^{-2} \end{aligned}$$

2 Markov Process

Definition. A continuous-time stochastic process $\{X(t), t \geq 0\}$ with integer state space is said to be a Markov process (M.P.) if it satisfies the Markovian property, i.e.

$$\begin{aligned} P(X(t+h) = j|X(t) = i, X(u) = x(u), 0 \leq u < t) \\ = P(X(t+h) = j|X(t) = i) \end{aligned}$$

for all $t, h \geq 0$, and non-negative integers $i, j, x(u) 0 \leq u < t$.

Definition. A Markov process $\{X(t), t \geq 0\}$ is said to have stationary (time-homogeneous) transition probabilities if $P(X(t+h) = j|X(t) = i)$ is independent of t, i.e.

$$P(X(t+h) = j|X(t) = i) = P(X(h) = j|X(0) = i) \equiv p_{ij}(h) .$$

Remarks A MP is a stochastic process that moves from one state to another in accordance with a MC, but the amount of time spent in each state is exponentially distributed.

Example. Suppose a MP enters state i at some time, say 0, and suppose that the process does not leave state i (i.e. a transition does not occur) during the next 10 minutes. What is the probability that the process will not leave state i during the next 5 minutes?

Answer. Since the MP is in state i at time 10, it follows by the Markovian property, that

$$P(T_i > 15|T_i > 10) = P(T_i > 5) = e^{-5\alpha_i} \ ,$$

where α_i is the transition rate out of state i.
FACT. T_i is exponentially distributed with rate, say α_i. That is $P(T_i > t) = e^{-\alpha_i t}$.
Remarks.

(i) $p_{ij}(h)$ are called the transition probabilities for the MP.

(ii) In a MP, times between transitions are exponentially distributed, possibly with different parameters.

(iii) A M.P. is characterized by its initial distribution and its transition matrix.

FACT. Let T_i be the time that the MP stays in state i before making a transition into a different state. Then

$$P(T_i > t + h|T_i > t) = P(T_i > h) \ .$$

Proof. Follows from the Markovian property.

2.1 Rate Properties of Markov Processes

Recall

$$p_{ij}(h) = P(X(t+h) = j \mid X(t) = i)$$

Lemma

The transition rates (intensities) are given by

$$(i) \ \ q_{ij} = \lim_{h \longrightarrow 0} \frac{p_{ij}(h)}{h}, \ i \neq j$$

$$(ii) \ \ q_i = \lim_{h \longrightarrow 0} \frac{1 - p_{ij}(h)}{h}, \ i \in S.$$

Remarks

$$(i) \ \ q_i = \sum_{i \neq j} q_{ij}$$

$$(ii) \ \ p_{ij} = \frac{q_{ij}}{\sum_{j \neq i} q_{ij}}$$

Interpretations

$$(i) \quad q_{ij} \quad = \quad \lim_{t \longrightarrow \infty} \frac{C(i,j;t)}{Y(i,t)} \qquad \text{transition rate from } i \text{ to } j.$$

Example

$$(ii) \quad p_i \quad = \quad \lim_{t \longrightarrow \infty} \frac{Y(i,t)}{t} \qquad \text{fraction of time in state } i$$

Flow Balance Equations

$$\text{flow out} \quad = \quad \text{flow in}$$

$$\sum_j p_i q_{ij} \quad = \quad \sum_j p_j q_{ji} \qquad\qquad i \neq j$$

$$p_i \sum_j q_{ij} \quad = \quad \sum_j p_j q_{ji}$$

Example for state *1*

$$q_{10} p_1 = q_{01} p_0 + q_{21} p_2$$

Birth-Death Process

$$\begin{aligned} q_{i,i+1} &= \lambda_i \\ q_{i,i-1} &= \mu_i \\ q_{i,j} &= 0 \quad \text{for } |i-j| > 1 \end{aligned}$$

Example

Flow Balance Equations

state		
0	$\lambda_0 p_0$	$= \mu_1 p_1$
1	$(\lambda_1 + \mu_1) p_1$	$= \lambda_0 p_0 + \mu_2 p_2$
2	$\mu_2 p_2$	$= \lambda_1 p_1$

Another Definition of Birth-Death Processes

Definition. A Markov process $\{X(t), t \geq 0\}$ is said to be a B-D process if

$$\begin{aligned} P(X(t+h) = j | X(t) = i) &= \lambda_i h + o(h); \;\; j = i+1 \\ &= \mu_i h + o(h); \;\; j = i-1 \\ &= 1 - \lambda_i h - \mu_i h + o(h); \;\; j = i. \end{aligned}$$

Transition diagram

Chapter 5

Queueing Models I

Contents.

1 Terminology

Calling population. Total number of distinct potential arrivals (The size of the calling population may be assumed to be finite (limited) or infinite (unlimited).

Arrivals. Let

$A(t) :=$ # of arrivals during $[0, t]$

$\lambda := \lim_{t\to\infty} \frac{A(t)}{t}$ (Mean arrival rate)

$\frac{1}{\lambda} =$ Mean time between arrivals

Service times. Time it takes to process a job. Let distribution of service times has mean $1/\mu$, i.e. μ is the mean service rate.

Queue discipline. FCFS, LCFS, Processor sharing (PS), Round robin (RR), SIRO, SPT, priority rules, etc.

Number of servers. single or multiple servers ($c = \#$ of servers)

Waiting room. Finite vs infinite (Use K for finite waiting room)

Notation.

$A/B/c/K/N$ (Kendall's notation)

A: describes the arrival process

B: describes the service time distribution

c: number of servers

K: buffer size

N: size of the calling population

* A and B may be replaced by

M: Markovian or memoryless

D: Deterministic

E_k : Erlang distribution

G: General (usually the mean and variance are known)

Examples.

$M/M/1$;

$M/M/2/5/20$;

$M/E_3/1$;

$G/G/1$.

M/M/1 queue.

(i) Time between arrivals are iid and exponential, i.e.

$$\begin{aligned} f(t) &= \lambda e^{-\lambda t} \quad \lambda, t > 0; \\ &= 0 \text{ otherwise.} \end{aligned}$$

(ii) Service times are iid and exponential, i.e.

$$\begin{aligned} g(t) &= \mu e^{-\mu t} \quad \mu, t > 0 \ ; \\ &= 0 \text{ otherwise.} \end{aligned}$$

(iii) There is one server, infinite waiting room, and infinite calling population.

(iv) System is in statistical equilibrium (steady state) and stable

(v) Stability condition: $\rho = \frac{\lambda}{\mu} < 1$.

Steady State Conditions.

Let $X(t) = \#$ of customers in system at time t. Then $\{X(t), t \geq 0\}$ is a stochastic process. We are interested in $\{X(t), t \geq 0\}$ when

(i) it is a birth-death process

(ii) it has reached steady state (or stationarity) (i.e. the process has been evolving for a long time)

Performance measures.

L = expected (mean) number of customers in the system

L_q = expected (mean) queue length (excluding customers in service)

W = expected (mean) time in system per arrival

W_q = expected (mean) time in queue per arrival

I = expected (mean) idle time per server

B = expected (mean) busy time per server

$\{P_n, n \geq 0\}$ = distribution of number of customers in system

T = waiting time is system (including service time) for each arrival (random variable)

T_q = waiting time is queue , excluding service time, (delay) for each arrival (random variable)
$P(T \geq t)$ = distribution of waiting times
Percentiles

General Relations.

(i) Little's formula:

$$L = \lambda W$$

$$L_q = \lambda W_q$$

(ii)

$$W = W_q + \frac{1}{\mu}$$

(iii) For systems that obey Little's law

$$L = L_q + \frac{\lambda}{\mu}$$

(iv) Single server

$$L = L_q + (1 - P_0)$$

Example. If λ is the arrival rate in a transmission line, N_q is the average number of packets waiting in queue (but not under transmission), and W_q is the average time spent by a packet waiting in queue (not including transmission time), Little's formula gives

$$N_q = \lambda W_q \ .$$

Moreover, if S is the average transmission time, then Little's formula gives the average number of packets under transmission as

$$\rho = \lambda S \ .$$

Since at most one packet can be under transmission, ρ is also the line's *utilization factor*, i.e., the proportion of time that the line is busy transmitting packets.

2 The Birth-Death Process

Flow balance diagram

Flow Balance Equations.

First we write the global balance equations using the principle of flow balance.

$$\text{Probability Flow out} = \text{Probability Flow in}$$

$$
\begin{aligned}
\text{Flow out} &= \text{Flow in} \\
\lambda_0 P_0 &= \mu_1 P_1 \\
\lambda_1 P_1 + \mu_1 P_1 &= \lambda_0 P_0 + \mu_2 P_2 \\
\lambda_2 P_2 + \mu_2 P_2 &= \lambda_1 P_1 + \mu_3 P_3 \\
&\vdots \\
\lambda_k P_k + \mu_k P_k &= \lambda_{k-1} P_{k-1} + \mu_{k+1} P_{k+1} \\
&\vdots
\end{aligned}
$$

Rewrite as

$$
\begin{aligned}
\lambda_0 P_0 &= \mu_1 P_1 \\
\lambda_1 P_1 &= \mu_2 P_2 \\
\lambda_2 P_2 &= \mu_3 P_3 \\
&\vdots \\
\lambda_k P_k &= \mu_{k+1} P_{k+1} \\
&\vdots
\end{aligned}
$$

More compactly,

$$\lambda_n P_n = \mu_{n+1} P_{n+1} \ , n = 0, 1, 2 \ldots . \tag{5.1}$$

Equations (5.1) are called local (detailed) balance equations.
Solve (5.1) recursively to obtain

$$
\begin{aligned}
P_1 &= \frac{\lambda_0}{\mu_1} P_0 \\
P_2 &= \frac{\lambda_1}{\mu_2} P_1 = \frac{\lambda_0 \lambda_1}{\mu_1 \mu_2} P_0 \\
P_3 &= \frac{\lambda_2}{\mu_3} P_2 = \frac{\lambda_0 \lambda_1 \lambda_2}{\mu_1 \mu_2 \mu_3} P_0 \\
&\vdots
\end{aligned}
$$

So that

$$P_n = \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n} P_0 \ , \tag{5.2}$$

for all $n = 0, 1, 2 \ldots$.
Assumption. $\{P_n, n = 0, 1, 2 \ldots\}$ exist and $\sum_{n=0}^{\infty} P_n = 1$.

Let $C_n = \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n}$, then (5.2) may be written as

$$P_n = C_n P_0 \ , n = 0, 1, 2 \ldots \ .$$

Therefore, $P_0 + P_1 + P_2 + \ldots = 1$ imply

$$[1 + C_1 + C_2 + \ldots] P_0 = 1$$

$$\begin{aligned} P_0 &= \frac{1}{1 + \sum_{k=1}^{\infty} C_k} \\ &= \frac{1}{1 + \sum_{k=1}^{\infty} \prod_{j=1}^{k} \frac{\lambda_{j-1}}{\mu_j}} \end{aligned}$$

Stability. A queueing system (B-D process) is stable if $P_0 > 0$ (What happens if $P_0 = 0$?) or

$$1 + \sum_{k=1}^{\infty} \prod_{j=1}^{k} \frac{\lambda_{j-1}}{\mu_j} < \infty$$

Assuming stability,

$$P_0 = \frac{1}{1 + \sum_{k=1}^{\infty} \prod_{j=1}^{k} \frac{\lambda_{j-1}}{\mu_j}} \tag{5.3}$$

$$P_n = \left(\prod_{j=1}^{n} \frac{\lambda_{j-1}}{\mu_j}\right) P_0 \ , n = 1, 2 \ldots \ . \tag{5.4}$$

3 Models Based on the B-D Process

4 M/M/1 Model

Here we consider an $M/M/1$ single server Markovian model.

$$\begin{aligned} \lambda_j &= \lambda \ , \quad j = 0, 1, 2 \ldots \\ \mu_j &= \mu \ , \quad j = 1, 2 \ldots \end{aligned}$$

Stability. Let $\rho = \frac{\lambda}{\mu}$. Then

$$\begin{aligned} 1 + \sum_{k=1}^{\infty} \rho^k &= \sum_{k=0}^{\infty} \rho^k \\ &= \frac{1}{1 - \rho} \text{ if } \rho < 1 \ . \end{aligned}$$

Therefore $P_0 = 1 - \rho$ and $P_n = \rho^n(1-\rho)$. That is

$$P_n = \rho^n(1-\rho)\ , \quad n = 0, 1, 2\ldots \tag{5.5}$$

Remark. (i) ρ is called the traffic intensity or utilization factor.

(ii) $\rho < 1$ is called the stability condition.

Measures of Performance.

$$\begin{aligned} L &:= \sum_{n=0}^{\infty} nP_n = \frac{\rho}{1-\rho} = \frac{\lambda}{\mu-\lambda} \\ L_q &:= \sum_{n=1}^{\infty} (n-1)P_n = \frac{\rho^2}{1-\rho} = \frac{\lambda^2}{\mu(\mu-\lambda)} \\ W &= \frac{1}{\mu-\lambda} \\ W_q &= \frac{\lambda}{\mu(\mu-\lambda)} \\ U &= 1 - P_0 = \rho \\ I &= 1-\rho \\ P(X \geq k) &= \rho^k \end{aligned}$$

Barbershop Example.

- One barber, no appointments, FCFS
- Busy hours or Saturdays
- Arrivals follow a Poisson process with mean arrival rate of 5.1 customers per hour
- Service times are exponential with mean service time of 10 minutes.

(i) Can you model this system as an $M/M/1$ queue?

Solution. Yes, with

$\lambda = 5.1$ arrivals per hour

$\frac{1}{\mu} = 10$ min, i.e $\mu = 1/10$ per min $= 6$ per hour

$\rho = \frac{\lambda}{\mu} = 5.1/6 = 0.85 < 1$. Therefore system is stable

$$\begin{aligned}
L &= \frac{\rho}{1-\rho} = \frac{.85}{1-.85} = 5.67 \text{ customers} \\
L_q &= \frac{\rho^2}{1-\rho} = \frac{.85^2}{1-.85} = 4.82 \text{ customers} \\
W &= \frac{1}{\mu-\lambda} = \frac{1}{6-5.1} = 1.11 \text{ hrs} \\
W_q &= \frac{\lambda}{\mu(\mu-\lambda)} = \frac{5.1}{6(6-5.1)} = 0.94 \text{ hrs} \\
U &= 1 - P_0 = \rho = 0.85 \\
I &= 1 - \rho = 0.15
\end{aligned}$$

(ii) What is the probability that the number of customers in the shop exceeds 3 (i.e. ≥ 4)?

Solution.

$$P(X \geq 4) = \rho^4 = 0.85^4 = .52$$

That is 52% of the time there will be 4 or more customers waiting.

(iii) What percentage of customers go directly into service? What percentage would have to wait?

Solution.

$P_0 = 1 - \rho = 0.15$. That is the 15% of all customers go directly into service.

$U = 0.85$. That is the 85% of all customers would have to wait.

(iv) What percentage of customers would have to wait at least half an hour before being serviced?

Hint: We need the distribution of waiting times.

Waiting Time Distribution.

Assume $FCFS$ discipline.

Recall : T_q is the time spent in queue per customer and T is the time spent in system per customer.

FACT. For an $M/M/1 - FCFS$ queue,

(i)

$$P(T > t) = e^{-\mu(1-\rho)t} \ , \ \ t \geq 0$$

(ii)

$$\begin{aligned}
P(T_q > t) &= \rho e^{-\mu(1-\rho)t} \ , \ \ t \geq 0 \\
P(T_q = 0) &= 1 - \rho \ .
\end{aligned}$$

Remark. The cdf is given by

$$\begin{aligned} P(T_q \le 0) &= 1-\rho \;\; t=0 \\ &= 1-\rho e^{-\mu(1-\rho)t} \;, \;\; t>0 \end{aligned}$$

FACT. For an $M/M/1-FCFS$ queue, the pdf of T (waiting time in system) and T_q (waiting time in queue) are

$$w(t) = (\mu-\lambda)e^{-(\mu-\lambda)t} \;, \;\; t>0$$

and

$$\begin{aligned} w_q(t) &= 1-\rho \;, \;\; t=0 \\ &= \mu\rho(1-\rho)e^{-\mu\rho(1-\rho)t} \;, \;\; t>0 \end{aligned}$$

Remark. One could obtain W from the pdf of T by

$$W = \int_0^\infty tw(t)dt = \frac{1}{\mu-\lambda}$$

IMP Example. Consider an Interface Message Processor (IMP). Let the pdf for packet size in bits be $f(t) = \mu e^{-\mu t}$ with a mean of $1/\mu$ bits/packet. The capacity of the communication channel is C bits/sec. Packets arrive at random (i.e exponential inter arrival times) with arrival rate λ packets/sec. Find the queueing delay for packets at the IMP.

Solution. We need to calculate W. First note that λ and μ do not have the same units.

Arrival rate: λ packets/sec.

Service rate:μC (packet/bit)(bits/sec)= μC packets/sec.

Therefore, $W = \frac{1}{\mu C-\lambda}$ sec.

Example. (Message Switching) Traffic to a message switching center for a corporation arrives in a random pattern (i.e. exponential inter arrival times) at an average rate of 240 messages per minute. The line has a transmission rate of 800 characters per second. The message length distribution (including control characters) is approximately exponential with an average length of 176 characters.

(i) Calculate the principal measures of system performance assuming that a very large number of message buffers is provided.

Solution.

We can model this system as an $M/M/1$ queue with

$\lambda = 240$ message/minute= 4 messages/second

$\frac{1}{\mu} = \frac{176\; char}{800\; char/sec} = 0.22$ seconds.

$\rho = \frac{\lambda}{\mu} = (4)(0.22) = 0.88 < 1$. Therefore system is stable

$$\begin{aligned}
L &= \frac{\rho}{1-\rho} = \frac{.88}{1-.88} = 7.33 \text{ messages} \\
L_q &= \frac{\rho^2}{1-\rho} = \frac{.88^2}{1-.88} = 6.45 \text{ messages} \\
W &= \frac{1}{\mu-\lambda} = \frac{1}{1/.22-4} = 1.83 \text{ seconds} \\
W_q &= \frac{\lambda}{\mu(\mu-\lambda)} = \frac{4}{1/.22(1/.22-4)} = 1.61 \text{ seconds} \\
U &= 1 - P_0 = \rho = 0.88 \\
I &= 1 - \rho = 0.12
\end{aligned}$$

(ii) Find the 90^{th} percentile time in the system.

Solution.

$$P(T \geq t) = e^{-\mu(1-\rho)t} = p$$

Then find t such that $p = 0.10$. Take ln of both sides and simplify

$$t = \frac{-\ln(p)}{\mu(1-\rho)} = \frac{-0.22\ln(0.10)}{(1-.88)} = 4.21\text{seconds}$$

(iii) Find the 90^{th} percentile time in the queue.

Solution. Exercise

Exercise. Suppose that we have two $M/M/1$ models with parameters (λ_1, μ_1) and (λ_2, μ_2) respectively. Show that if $\rho_1 = \rho_1$ and $W_1 = W_2$, then the two models are identical, in the sense that $\lambda_1 = \lambda_2$ and $\mu_1 = \mu_2$.

Example.(Token Rings)

Assume that frames are generated according to a Poisson process, and that when a station receives permission to send, it empties itself of all queued frames, with the mean queue length being L_q frames. The total arrival rate of all N stations combined is λ frames/sec. Each station contributes λ/N. The service rate (the number of frames/sec that a station can transmit) is μ. The time it takes for a bit to go all the way around an idle ring, or *walk time*, consisting of both the one bit per station delays and the signal propagation delay, plays a key role in mean delay. Denote the walk time by w. Calculate the *scan time*,D, the mean interval between token arrivals at a given station.

Solution. The scan time, D, (delay) is divided into two parts, the walk time, w, and the time it takes to service Q requests queued up for service (at all stations), each of which requires $1/\mu$ sec. That is

$$D = w + \frac{Q}{\mu}$$

But $Q = \lambda D$, so

$$D = w + \frac{\lambda D}{\mu}$$

Let $\rho = \frac{\lambda}{\mu}$ be the utilization of the entire ring, and solve for D, we find

$$D = \frac{w}{1-\rho}$$

Notice that D is proportional to the walk time both for light and heavy traffic.

Example. (Terminal Concentrators)

Consider a terminal concentrator with four 4800 bps (bits/second) input lines and one 9600 bps output line. The mean packet size is 1000 bits. Each of the four lines delivers Poisson traffic with an average rate of $\lambda_i = 2$ packets per second ($i = 1, 2, 3, 4$).

(i) What is the mean delay experienced by a *packet* from the moment that a bit arrives at the concentrator until the moment that bit is retransmitted on the output line?

(ii) What is the mean number of packets in the concentrator, including the one in service?

(iii) What is the probability that a packet finds 10 packets in the concentrator upon arrival?

Solution.

$\lambda_i = 2$ packets/sec.

$\lambda = 8$ packets/sec.

$\mu C = (1/1000)(9600) = 9.6$ packets/sec. (Service rate)

The model is an $M/M/1$ queue with

$$\rho = \frac{\lambda}{\mu C} = \frac{8}{9.6} = .83 < 1$$

Therefore

(i) $W = \frac{1}{\mu C - 1} = \frac{1}{9.6-8} = .625$ sec.

(ii) $L = \lambda W = 8(.625) = 4.99$ packets.

(iii) $P(X \geq 10) = \rho^{10} = .833^{10} = .16$

Example.(Dedicated Versus Shared Channels)

Two computers are connected by a 64 kbps line. There are eight parallel sessions using the line. Each session generates Poisson traffic with a mean of 2 packets/sec. The packet lengths are exponentially distributed with mean of 2000 bits. The system designers must choose between giving each session a dedicated 8 kpbs piece of bandwidth (via TDM or FDM) or having all packets compete for a single 64 kbps shared channel. Which alternative gives better response time (i.e. W)?

Solution. We need to compare two alternative models.

Alternative 1.

For the TDM or FDM , each 8 kbps operates as an independent $M/M/1$ queue with $\lambda = 2$ packets/sec and $\mu = 4$ packets/sec. Therefore

$$W = \frac{1}{\mu - \lambda} = \frac{1}{4-2} = 0.5 \, sec. = 500 \, msec.$$

Alternative 2.

The single 64 kbps is modeled as an $M/M/1$ queue with $\lambda = 16$ packets/sec and $\mu = 32$ packets/sec. Therefore

$$W = \frac{1}{\mu - \lambda} = \frac{1}{32-16} = 0.0625 \, sec = 62.5 \, msec.$$

Splitting up a single channel into 4 fixed size pieces makes the response time worse. The reason is that it frequently happens that several of the smaller channels are idle, while other ones are processing work at the reduced

Exercise. Suppose that we have two $M/M/1$ models with parameters (λ_1, μ_1) and (λ_2, μ_2) respectively. Show that if $\rho_1 = \rho_2$ and $W_1 = W_2$, then the two models are identical, (i.e. in $\lambda_1 = \lambda_2$ and $\mu_1 = \mu_2$.)

5 M/M/c/∞/∞

Here we consider a Markovian queueing model with Parallel Channels.

Flow balance diagram

Recall:

$$P_n = (\prod_{j=1}^{n} \frac{\lambda_{j-1}}{\mu_j}) P_0 \ , n = 1, 2 \ldots \ .$$

Now,

$$\begin{aligned} \lambda_n &= \lambda \ , \ \ n = 0, 1, 2 \ldots \\ \mu_n &= n\mu \ , \ \ 1 \leq n \leq c \\ &= c\mu \ , \ \ n \geq c \ . \end{aligned}$$

Substituting λ_n and μ_n in the B-D steady state distribution we obtain all the results below.

Stability. Let $\rho = \frac{\lambda}{c\mu}$. Then $\rho < 1$ is called the stability condition. Let $a = \frac{\lambda}{\mu}$ be the offered load.

$$\begin{aligned} P_n &= \frac{a^n}{n!} P_0 \ , \ \ 1 \leq n \leq c \\ &= \frac{a^n}{c! c^{n-c}} P_0 \ , \ \ n \geq c, \end{aligned}$$

where

$$
\begin{aligned}
P_0 &= [\sum_{n=0}^{c-1} \frac{a^n}{n!} + \sum_{n=c}^{\infty} \frac{a^n}{c!c^{n-c}}]^{-1} \\
&= [\sum_{n=0}^{c-1} \frac{a^n}{n!} + \frac{a^c}{c!} \sum_{n=c}^{\infty} \frac{a^{n-c}}{c^{n-c}}]^{-1} \\
&= [\sum_{n=0}^{c-1} \frac{a^n}{n!} + \frac{a^c}{c!(1-\rho)}]^{-1} .
\end{aligned}
$$

Measures of Performance.

$$
\begin{aligned}
L_q &:= \sum_{n=c}^{\infty} (n-c)P_n = \frac{a^c \rho}{c!(1-\rho)^2} P_0 \\
W_q &= \frac{L_q}{\lambda} = [\frac{a^c}{c!(c\mu)(1-\rho)^2}]P_0 \\
W &= W_q + \frac{1}{\mu} = \frac{1}{\mu} + [\frac{a^c}{c!(c\mu)(1-\rho)^2}]P_0 \\
L &= \lambda W = L_q + \frac{\lambda}{\mu} = a + \frac{a^c \rho}{c!(1-\rho)^2} P_0 \\
U &= \rho
\end{aligned}
$$

Also the mean number of busy servers is given by

$$
B = a = \frac{\lambda}{\mu} .
$$

FACT.

$$
\begin{aligned}
P(X \geq c) &= \sum_{n=c}^{\infty} P_n = \frac{a^c}{c!(1-\rho)} P_0 = \frac{P_c}{1-\rho} \\
&= \frac{a^c/c!}{a^c/c! + (1-\rho)\sum_{n=0}^{c-1} a^n/n!} .
\end{aligned}
$$

Remark. The relation $P(X \geq c) = \frac{P_c}{1-\rho}$ is called the Erlang second (delay) formula. It represents the probability that customers would have to wait. (Or percentage of customers that wait)

FACT. For an $M/M/c - FCFS$ queue,

(i)

$$
\begin{aligned}
P(T_q = 0) &= \sum_{n=0}^{c-1} P_n = 1 - \frac{a^c p_0}{c!(1-\rho)} \; ; \\
P(T_q > t) &= (1 - P(T_q = 0))e^{-c\mu(1-\rho)t} \\
&= \frac{a^c p_0}{c!(1-\rho)} e^{-c\mu(1-\rho)t} \; , \quad t \geq 0
\end{aligned}
$$

(ii)

$$
P(T_q > t | T_q > 0) \quad = \quad e^{-c\mu(1-\rho)t} \; , \quad t > 0
$$

(iii)

$$
P(T > t) = e^{-\mu t}[1 + \frac{a^c(1 - e^{-\mu t(c-1-a)}}{c!(1-\rho)(c-1-a)} P_0] \; , \quad t \geq 0
$$

Proof. Let $W_q(t) = P(T_q \leq t)$

Proof.

$$
\begin{aligned}
W_q(0) &= P(T_q = 0) = P(X \leq c-1) \\
&= \sum_{n=c}^{c-1} p_n \\
&= 1 - \sum_{n=0}^{\infty} p_n \\
&= 1 - \sum_{n=c}^{\infty} \frac{a^n}{c!} c^{n-c} p_0 \\
&= 1 - \frac{a^c}{c!} \sum_{n=c}^{\infty} \left(\frac{a}{c}\right)^{n-c} p_0 \\
&= 1 - \frac{a^c}{c!(1-p)} p_0
\end{aligned}
$$

$$\begin{aligned}
W_q(t) &= W_q(0) + \sum_{n=c}^{\infty} p(n-c+1 \text{ completions in } \leq t \mid \text{arrival} \\
&\quad \text{finds } n \text{ in system}) p_n \\
&= W_q(0) + \sum_{n=c}^{\infty} \int_0^t \frac{c\mu(c\mu x)^{n-c}}{(n-c)!} e^{-c\mu x} dx \frac{a^n}{c^{n-c}c!} p_0 \\
&= W_q(0) + \frac{a^n p_0}{c^{n-c}c!} \int_0^t ce^{-c\mu x} \sum_{n=c}^{\infty} \frac{\mu(c\mu x)^{n-c}}{(n-c)!} dx \\
&= W_q(0) + \frac{a^c p_0}{(c-1)!} \int_0^t \mu e^{-c\mu x} \sum_{n=c}^{\infty} \frac{(\mu a x)^{n-c}}{(n-c)!} dx \\
&= W_q(0) + \frac{a^c p_0}{(c-1)!} \int_0^t \mu e^{-c\mu x} e^{\mu a x} dx \\
&= W_q(0) + \frac{a^c p_0}{(c-1)!} \int_0^t \mu e^{-\mu(c-a)x} dx \\
&= W_q(0) + \frac{a^c p_0}{(c-1)!(c-a)} \int_0^t \mu(c-a) e^{-\mu(c-a)x} dx \\
&= W_q(0) + \frac{a^c p_0}{c!(1-\rho)} \left[1 - e^{-\mu(c-a)t}\right]
\end{aligned}$$

OR

$$= W_q(0) + \frac{a^c p_0}{c!(1-\rho)} \left[1 - e^{-(c\mu-\lambda)+t}\right].$$

Proofs of other statements are similar.

Example. An airline is planning a new telephone reservation center. Each agent will have a reservations terminal and can serve a typical caller in 5 minutes, the service time being exponentially distributed. Calls arrive randomly and the system has a large message buffering system to hold calls that arrive when no agent is free. An average of 36 calls per hour is expected during the peak period of the day. The design criterion for the new facility is that the probability a caller will find all agents busy must not exceed 0.1 (10%).

(i) How many agents and terminals should be provided?

(ii) How will this system perform if the number of callers per hour is 10% higher than anticipated?

Solution.

(i) This problem can be modeled as an $M/M/c$ queue with $\lambda = 0.6$ calls per min and $1/\mu = 5min$. Thus $a = \frac{\lambda}{\mu} = (0.6)(5) = 3$ is the offered traffic. For this system to be stable we need a minimum of $c = 4$ agents. Now

c	$P(X \geq c)$
4	0.5094
5	0.2362
6	0.0991

Therefore $c = 6$.

(ii) Exercise.

6 Finite Buffer Models)

I. M/M/c/K (Finite Buffer Multiserver Model) $[(K \geq c)]$

Flow balance diagram

$$\begin{aligned} \lambda_n &= \lambda\ , \quad n = 0, 1, 2 \ldots, K-1 \\ &= 0\ , \quad n \geq K \\ \mu_n &= n\mu\ , \quad 1 \leq n \leq c \\ &= c\mu\ , \quad n \geq c\ . \end{aligned}$$

Stability. Let $\rho = \frac{\lambda}{c\mu}$. This system is stable whether $\rho < 1$ or not. Recall $a = \lambda/\mu$

We need to consider two separate cases: $\rho = 1$, and $\rho \neq 1$.

For $n = 0, 1, 2 \ldots, K$

$$\begin{aligned} P_n &= \frac{a^n}{n!} P_0\ , \quad 1 \leq n \leq c \\ &= \frac{a^n}{c! c^{n-c}} P_0\ , \quad n \geq c, \end{aligned}$$

where

$$\begin{aligned} P_0 &= [\sum_{n=0}^{c-1} \frac{a^n}{n!} + \frac{a^c(1-\rho^{K-c+1})}{c!(1-\rho)}]^{-1} \quad ; \rho \neq 1 \\ &= [\sum_{n=0}^{c-1} \frac{a^n}{n!} + \frac{a^c(K-c+1)}{c!}]^{-1} \quad ; \rho = 1 \end{aligned}$$

Note. When $\rho = 1$, $a = c$.
Proof. of P_0.

For $\rho \neq 1$

$$\begin{aligned} P_0 &= [\sum_{n=0}^{c-1} \frac{a^n}{n!} + \sum_{n=c}^{K} \frac{a^n}{c! c^{n-c}}]^{-1} \\ &= [\sum_{n=0}^{c-1} \frac{a^n}{n!} + \frac{a^c}{c!} \sum_{n=c}^{K} \rho^n]^{-1} \end{aligned}$$

Then the proof follows by noting that for $\rho \neq 1$

$$\sum_{n=c}^{K} \rho^n = \frac{1 - \rho^{K-c+1}}{1 - \rho} \ ;$$

and for $\rho = 1$

$$\sum_{n=c}^{K} \rho^n = K - c + 1 \ .$$

FACT. The effective arrival rate (arrival that join the system) is given by

$$\lambda' = \lambda(1 - P_K),$$

and the overflow rate is λP_K.

Measures of Performance.

$$\begin{aligned} L_q &= \frac{a^c \rho}{c!(1-\rho)^2}[1 - \rho^{K-c+1} - (1-\rho)(K-c+1)\rho^{K-c}]P_0; \rho \neq 1 \\ L_q &= \frac{c^c}{c!} \frac{(K-c)(K-c+1)}{2} P_0; \rho = 1 \\ W_q &= \frac{L_q}{\lambda(1 - P_K)} \\ W &= W_q + \frac{1}{\mu} \\ L &= \lambda' W = \lambda(1 - P_K) W \end{aligned}$$

Proof. For L_q formula for $\rho \neq 1$:

$$
\begin{aligned}
L_q &= \sum_{n=c}^{K}(n-c)P_n \\
&= \frac{P_0}{c!}\sum_{n=c}^{K}(n-c)\frac{a^n}{c^{n-c}} \\
&= \frac{P_0 a^c}{c!}\sum_{n=c}^{K}(n-c)(a/c)^{n-c} \\
&= \frac{P_0 a^c \rho}{c!}\sum_{n=c}^{K}(n-c)\rho^{n-c-1} \\
&= \frac{P_0 a^c \rho}{c!}\sum_{i=0}^{K-c} i\rho^{i-1} \\
&= \frac{P_0 a^c \rho}{c!}\sum_{i=0}^{K-c} \frac{d}{d\rho}\rho^{i} \\
&= \frac{P_0 a^c \rho}{c!}\frac{d}{d\rho}\sum_{i=0}^{K-c} \rho^{i} \\
&= \frac{P_0 a^c \rho}{c!}\frac{d}{d\rho}[\frac{1-\rho^{K-c+1}}{1-\rho}] \\
&= \frac{a^c \rho}{c!(1-\rho)^2}[1-\rho^{K-c+1}-(1-\rho)(K-c+1)\rho^{K-c}]P_0
\end{aligned}
$$

The proof when $\rho = 1$ is similar, and is given below. Here, $a = \frac{\lambda}{\mu} = c$.

$$
\begin{aligned}
P_n &= \frac{a^n}{n!}P_0 = \frac{c^n}{n!}P_0\ , \quad 1 \leq n \leq c \\
&= \frac{a^n}{c!c^{n-c}}P_0 = \frac{c^c}{c!}P_0\ , \quad n \geq c,
\end{aligned}
$$

Now,

$$
\begin{aligned}
L_q &= \sum_{n=c+1}^{K} (n-c)P_n \\
&= \sum_{n=c+1}^{K} (n-c)\frac{c^c}{c!}P_0 \\
&= \frac{c^c}{c!}P_0 \sum_{n=c+1}^{K} (n-c) \\
&= \frac{c^c}{c!}P_0 \sum_{j=1}^{K-c} j \\
&= \frac{c^c}{c!}\frac{(K-c)(K-c+1)}{2}P_0 \,.
\end{aligned}
$$

Waiting Times Distribution.
Again, let $W_q(t) = P(T_q \le t)$

FACT.

$$
W_q(t) = W_q(0) + \sum_{n=c}^{K-1} \pi_n - \sum_{n=c}^{k-1} \pi_n \sum_{i=0}^{n-c} \frac{(\mu c t)^i e^{-\mu c t}}{i!}
$$

Proof.

$$
\begin{aligned}
W_q(t) &= W_q(0) + \sum_{n=c}^{k-1} P(n-c+1 \text{ completions in } \le t \mid \\
&\qquad \text{arrival found } n \text{ in system})\pi_n \\
&= W_q(0) + \sum_{n=c}^{k-1} \pi_n \int_0^t \frac{c\mu(c\mu x)^{n-c}}{(n-c)!} e^{-c\mu x} dx
\end{aligned}
$$

The rest of the proof is similar to the infinite buffer case.

Remark. let π_n = long-run fraction of arrivals that find n in system.
Then, $\pi_n \ne p_n$ (i.e., PASTA does not hold)

FACT.

$$
\pi_n = \frac{p_n}{1-p_K}, \quad n = 0, 1, 2, \ldots, K-1.
$$

Proof. For a birth/death process

$$\lambda \pi_j = \lambda_j p_j \quad (*)$$

Now, sum over all $j \Longrightarrow \lambda = \sum \lambda_j p_j$

Therefore (*) $\Longrightarrow \pi_j = \dfrac{\lambda_j p_j}{\sum \lambda_j p_j}$

Recall

$$\begin{aligned} \lambda_j &= \lambda, \quad j \leq k-1 \\ & \quad 0, \quad \text{otherwise} \quad \Longrightarrow \end{aligned}$$

$$\pi_j = \frac{\lambda p_j}{\lambda \sum_{j=0}^{k-1} p_j} = \frac{p_j}{1 - p_K}.$$

II. M/M/1/K (Finite Buffer Single Server Model)

Flow balance diagram

$$\begin{aligned} \lambda_n &= \lambda \, , \quad n = 0, 1, 2 \ldots, K-1 \\ &= 0 \, , \quad n \geq K \\ \mu_n &= \mu \, , \quad n = 1, 2 \ldots, K. \end{aligned}$$

Stability. Let $\rho = \frac{\lambda}{\mu}$. This system is stable whether $\rho < 1$ or not.

We need to consider two separate cases: $\rho = 1$, and $\rho \neq 1$.

For $n = 0, 1, 2 \ldots, K$, $P_n = \rho^n P_0$. That is

$$\begin{aligned} P_n &= \frac{\rho^n (1-\rho)}{1 - \rho^{K+1}} \, , \quad \rho \neq 1 \\ &= \frac{1}{K+1} \, , \quad \rho = 1, \end{aligned}$$

Remarks. (i) The stationary probabilities could be obtained directly from the stationary probabilities of the $M/M/1$ model using truncation when $\rho < 1$. That is

$$P_n = \frac{\rho^n (1-\rho)}{\sum_{n=0}^{K} \rho^n} = \frac{\rho^n (1-\rho)}{1 - \rho^{K+1}}$$

(ii)

$$\begin{aligned} P_0 &= \frac{1-\rho}{1 - \rho^{K+1}} \, , \quad \rho \neq 1 \\ &= \frac{1}{K+1} \, , \quad \rho = 1, \end{aligned}$$

FACT. The effective arrival rate (arrivals that join the system) is given by

$$\lambda' = \lambda(1 - P_K) = \mu(1 - p_0) \ .$$

In particular, if $\rho \neq 1$

$$\lambda' = \frac{\lambda(1 - \rho^K)}{1 - \rho^{K+1}}$$

and if $\rho = 1$

$$\lambda' = \frac{\lambda K}{K + 1} \ .$$

The rate at which customers are blocked (lost) is λP_K.

Measures of Performance.

Case 1:$\rho = 1$ $L = K/2$

Case 2: $\rho \neq 1$

$$\begin{aligned}
L &:= \sum_{n=0}^{K} n P_n = \frac{\rho}{1-\rho} - \frac{(K+1)\rho^{K+1}}{1-\rho^{K+1}} \\
U &= 1 - P_0 = \frac{\lambda'}{\mu} = \frac{\rho(1-\rho^K)}{1-\rho^{K+1}} \\
L_q &= L - (1 - P_0) = L - \frac{\lambda'}{\mu} = L - \frac{\rho(1-\rho^K)}{1-\rho^{K+1}} \\
W &= \frac{L}{\lambda(1 - P_K)} \\
W_q &= \frac{L_q}{\lambda(1 - P_K)} \\
I &= P_0 = \frac{1-\rho}{1-\rho^{K+1}}
\end{aligned}$$

Example (Message Switching) Suppose the Co. has the same arrival pattern, message length distribution, and line speed as described in the example. Suppose, however, that it is desired to provide only a minimum number of message buffers required to guarantee that $P_K < 0.005$.

(i) How many buffers should be provided?

Solution. Need to find K such that

$$\frac{\rho^K(1-\rho)}{1-\rho^{K+1}} = 0.005$$

where $\rho = 0.88$. Therefore $K = 25.142607$, i.e. $K = 26$. Thus we need 25 buffers.

(ii) For this number of buffers calculate L, L_q, W, and W_q.
Solution.

$$\begin{aligned} L &= 6.44 \text{ messages} \\ L_q &= 5.573 \text{ messages} \\ W &= 1.62 \text{ seconds} \\ W_q &= 1.40 \text{ seconds} \end{aligned}$$

Exercise. Check these calculations using the formulas. Then compute $\lambda_0 := \lambda p_k$, the overflow rate.

7 M/M/c/c Erlang Loss Model

Flow balance diagram
For $n = 0, 1, 2 \ldots, c$

$$P_n = \frac{a^n}{n!} P_0 = \frac{a^n/n!}{\sum_{j=0}^{c} a^j/j!}\ .$$

$$P_c = \frac{a^c/c!}{\sum_{j=0}^{c} a^j/j!}\ .$$

Remark.

$$\begin{aligned} B(c,a) &= P_c, \text{ is called the Erlang loss probability} \\ &= \text{P(all servers are busy)} \\ &= \text{P(an arrival will be rejected)} \\ &= \text{overflow probability} \end{aligned}$$

FACT. The effective arrival rate (arrivals that join the system) is given by $\lambda' = \lambda(1 - P_c)$, and the overflow rate is λP_c.
Measures of Performance.

$$\begin{aligned} W &= \frac{1}{\mu} \\ L &= \lambda' W = \lambda(1 - P_c)/\mu = a(1 - P_c) \\ U &= \rho(1 - P_c) \end{aligned}$$

Note that $L_q = W_q = 0$ (why?).

Example. What is the minimal number of servers needed, in an $M/M/c$ Erlang loss system, to handle an offered load $a = \lambda/\mu = 2$ erlangs, with a loss no higher than 2%?
Solution. Need to find c such that $B(c, a = 2) \leq 0.02$.

$$\begin{aligned}
B(0,2) &= 1 \\
B(1,2) &= 2/3 \\
B(2,2) &= 2/5 \\
B(3,2) &= 4/19 \\
B(4,2) &= 2/21 \approx .095 \\
B(5,2) &= 4/109 \approx .037 \\
B(6,2) &= 4/381 \approx .01 < 0.02.
\end{aligned}$$

Therefore $c = 6$.
Applications. In many computer systems, there is a maximum number of connections that can be established at any one time. Companies that subscribe to certain networks may only have a certain number of virtual circuits open at any one instant, the number being determined when the company subscribes to the service. ATM switches allow for a fixed number of outgoing lines, so here too a fixed number of connections can coexist.

In all these and other cases, it is interesting to be able to compute the probability that an attempt to establish a new connection fails because the maximum number of connections already exists. We can model this environment by the Erlang loss model.

8 M/M/∞/∞ Unlimited Service Model

Flow balance diagram
For $n = 0, 1, 2 \ldots,$

$$P_n = \frac{a^n e^{-a}}{n!}$$

Measures of Performance.

$$\begin{aligned}
W &= \frac{1}{\mu} \\
L &= \lambda W = \frac{\lambda}{\mu}
\end{aligned}$$

Note that $L_q = W_q = 0$ (why?).
Example. Calls in a telephone system arrive randomly at an exchange at the rate of 140 per hour. If there is a very large number of lines available to handle the calls, that last an average of 3 minutes,

(i) What is the average number of lines, L, in use? What is the standard deviation?

(ii) Estimate the 90th and 95th percentile of the number of lines in use.

Solution.

(i) This problem can be modeled as an $M/M/\infty$ queue with $\lambda = 140/60$ arrivals per min and $1/\mu = 3min$. Thus $L = \frac{\lambda}{\mu} = (14/6)(3) = 7$

$\sigma = \sqrt{L} = \sqrt{7}$

(ii) Use the normal distribution as an estimate of percentile values.

90th percentile $= 7 + 1.28\sqrt{7} = 10.38$ or 10 lines.

95th percentile $= 7 + 1.645\sqrt{7} = 11.35$ or 11 lines.

9 Finite Population Models

I. M/M/1//N (Finite Population Single Server Model)($N \geq 1$)

Flow balance diagram

$$\begin{aligned}
\lambda_n &= (N-n)\lambda\ , \quad n = 0, 1, 2 \ldots, N \\
&= 0\ , \quad n \geq N \\
\mu_n &= \mu\ , \quad 1 \leq n \leq N\ .
\end{aligned}$$

Stability. This system is always stable. Recall $a = \lambda/\mu$ is called offered load per idle source.

For $n = 0, 1, 2 \ldots, N$

$$P_n = [\frac{N! a^n}{(N-n)!}] P_0\ ,$$

where

$$P_0 = [\sum_{n=0}^{N} \frac{N! a^n}{(N-n)!}]^{-1}$$

Measures of Performance.

$$\begin{aligned}
L_q &:= \sum_{n=0}^{N} (n-1) P_n = N - \frac{\lambda + \mu}{\lambda}(1 - P_0) \\
L &:= \sum_{n=0}^{N} n P_n = L_q + (1 - P_0) = N - \frac{\mu}{\lambda}(1 - P_0) \\
W &= \frac{L}{\lambda(N-L)} = \frac{N}{\mu(1-p_0)} - \frac{1}{\lambda} \\
W_q &= \frac{L_q}{\lambda(N-L)}
\end{aligned}$$

FACT. The effective arrival rate (arrivals that join the system) is given by

$$\lambda' = \sum_{n=0}^{N} \lambda_n P_n = \lambda(N-L) = \mu(1-p_0) \ .$$

Remarks. (i) Equating the effective arrival rate and effective departure rate gives

$$\lambda(N-L) = \mu(1-p_0)$$

which leads to

$$L = N - \frac{\mu}{\lambda}(1-P_0)$$

With this argument, the above formula is valid for non-exponential service times as well.

(ii) Note that $W \geq 1/u$ and $W \geq \frac{N}{\mu} - \frac{1}{\lambda}$ which is the case when $p_0 = 0$. Thus we have the following inequality

$$W \geq \max(\frac{N}{\mu} - \frac{1}{\lambda}, \frac{1}{\mu}) \ .$$

Example. An OR analyst believes he can model the word processing and electronic mail activities in his executive office as an $M/M/1//4$ queueing system. He generates so many letters, memos, and email messages each day that four secretaries ceaselessly type away at their workstations that are connected to a large computer system over an LAN. Each secretary works on average, for 40 seconds before she makes a request for service to the computer system. A request for service is processed in one second on average (processing times being exponentially distributed). The analyst has measured the mean response time using his electronic watch (he gets 1.05 seconds) and estimates the throughput as 350 requests per hour. The OR analyst has decided to hire two additional secretaries to keep up with his productivity and will connect their workstations to the same LAN if the $M/M/1//6$ model indicates a mean response time of less than 1.5 seconds. Should he hire the two secretaries?

Solution.

First examine the current system. Here $1/\lambda = 40$ sec/job, $1/\mu = 1$ sec/job.

With $N = 4$,

$$P_0 = [\sum_{n=0}^{N} \frac{N! a^n}{(N-n)!}]^{-1} = 0.90253$$

$$\begin{aligned} L &:= N - \frac{\mu}{\lambda}(1-P_0) = 4 - 40(1-.90253) = .1012 \\ \lambda' &= \lambda(N-L) = 0.09747 \text{ jobs/sec} = 350.88 \text{ jobs/hr} \\ W &= \frac{L}{\lambda'} = 1.038 \end{aligned}$$

With $N = 6$, $P_0 = 0.85390$, $\lambda' = 525.95$ requests/hr and $W = 1.069$ seconds.

The OR analyst can add two workstations without seriously degrading the performance of his office staff.

Exercise. Verify all calculations.

II. M/M/c//N (Finite Population Multiserver Model) ($N \geq c$)

Flow balance diagram

$$\begin{aligned} \lambda_n &= (N-n)\lambda \ , \quad n = 0, 1, 2 \ldots, N \\ &= 0 \ , \quad n \geq N \\ \mu_n &= n\mu \ , \quad 1 \leq n \leq c \\ &= c\mu \ , \quad n \geq c \ . \end{aligned}$$

Stability. This system is always stable. Recall $a = \lambda/\mu$ is the offered load per idle source.

For $n = 0, 1, 2 \ldots, N$

$$\begin{aligned} P_n &= \frac{N!a^n}{(N-n)!n!} P_0 \ , \quad 1 \leq n \leq c \\ &= \frac{N!a^n}{(N-n)!c!c^{n-c}} P_0 \ , \quad n \geq c, \end{aligned}$$

where

$$P_0 = [\sum_{n=0}^{c-1} \frac{N!a^n}{(N-n)!n!} + \sum_{n=c}^{N} \frac{N!a^n}{(N-n)!c!c^{n-c}}]^{-1}$$

Measures of Performance.

$$\begin{aligned} L &:= \sum_{n=0}^{N} nP_n \\ \lambda' &= \lambda(N-L) \\ L_q &= L - \frac{\lambda'}{\mu} = (a+1)L - aN \\ W &= \frac{L}{\lambda(N-L)} \\ W_q &= \frac{L_q}{\lambda(N-L)} \end{aligned}$$

FACTS.

(i)The effective arrival rate (arrivals that join the system) is given by

$$\lambda' = \sum_{n=0}^{N} \lambda_n P_n = (N-L)\lambda$$

(ii) Arrival point (customer-average) probabilities are equal to the time-average probabilities with one customer removed from the system. That is,

$$\pi_n(N) = p_n(N-1)$$

III. M/M/c/c/N (Finite Population Loss Model) $(N \geq c)$

Flow balance diagram

$$\begin{aligned} \lambda_n &= (N-n)\lambda\ , \quad n = 0, 1, 2 \ldots, c \\ &= 0\ , \quad n \geq c \\ \mu_n &= n\mu\ , \quad 1 \leq n \leq c \\ &= 0\ , \quad n \geq c\ . \end{aligned}$$

Stability. This system is always stable. Recall $a = \lambda/\mu$ is the offered load per idle source.

For $n = 0, 1, 2 \ldots, c$

$$\begin{aligned} P_n &= \frac{N!a^n}{(N-n)!n!} P_0\ , \quad 1 \leq n \leq c \\ &= \frac{\binom{N}{n} a^n}{\sum_{k=0}^{c} \binom{N}{k} a^k}\ , \quad 1 \leq n \leq c\ . \end{aligned}$$

Note P_c is called *Engest loss formula.*

Remark. Let $p = \frac{a}{1+a}$, i.e. $a = \frac{p}{1-p}$. Then $a^n = \frac{p^n}{(1-p)^n}$ implies that for all n

$$a^n(1-p)^N = p^n(1-p)^{N-n}$$

Now, multiplying the denominator and numerator of p_n by $(1-p)^N$, we obtain

$$p_n = \frac{\binom{N}{n} p^n (1-p)^{N-n}}{\sum_{k=0}^{c} \binom{N}{k} p^k (1-p)^{N-k}}\ , \quad 0 \leq n \leq c\ .$$

which is recognized as a truncated binomial probability distribution. Note also that if $c = N$, the distribution is binomial.

Measures of Performance.

$$\begin{aligned} L &:= \sum_{n=0}^{c} nP_n \\ \lambda' &= \lambda(N - L) \\ W &= \frac{L}{\lambda'} \\ L_q &= W_q = 0\ . \end{aligned}$$

FACTS. (i) $\pi_n(N) = p_n(N-1) = \frac{\binom{N-1}{n}a^n}{\sum_{k=0}^{c}\binom{N-1}{k}a^k}$, for all values of n.

(ii) The stationary distribution p_n is the same for non-exponential service times. (Insensitivity)

10 System Availability

Consider a K-out-of-N system with repair.

- system functions if at least K components function
- i.e. system fails if $N - K + 1$ components fail

System states. $0, 1, \ldots, N - K + 1$, *i.e.* # of down components/machines
System Probabilities. $p_0, p_1, \ldots, p_{N-K+1}$

Objective. Compute system availiability, i.e.

$$\begin{aligned} \text{P (system is available)} &= \text{long-run fraction of time the system is available} \\ &= \sum_{n=0}^{N-K} p_n \\ &= 1 - p_{N-K+1} \end{aligned}$$

Note.

$$\begin{aligned} p_{N-K+1} &= \text{unavailability of the system} \\ &= \text{Prob.(system is unavailable)} \end{aligned}$$

Assumptions.

1. Time to failure is exp (λ)
2. Repair times are exp (μ).
3. # of servers, $1 \leq c \leq N - K + 1$.

Notation. Let $a = \dfrac{\lambda}{\mu}$

Flow Balance Diagram. (Birth-Death process)

Note. 1-out-of-N system ($K = 1$) is a usual finite population model.

Flow Balance Equations. (Global balance equations)

$$\begin{aligned}
\lambda_0 p_0 &= \mu_1 p_1 \\
\lambda_n p_n + \mu_n p_n &= \lambda_{n-1} p_{n-1} + \mu_{n+1} p_{n+1}, n = 1, 2, \ldots, N - K + 1
\end{aligned}$$

where

$$\begin{aligned}
\lambda_n &= (N - n)\lambda, \quad n = 0, 1, \ldots, N - K + 1 \\
\mu_n &= n\mu \qquad\qquad n \leq c \\
&= c\mu \qquad\qquad c \leq n \leq N - K + 1.
\end{aligned}$$

These equations lead to the detailed balance equations (DBE)

$$\lambda_n p_n = \mu_{n+1} p_{n+1} \qquad n = 0, 1, \ldots, N-K$$

Solving the DBE, we get

$$\begin{aligned} p_1 &= \frac{N\lambda}{\mu} p_0 = Nap_0 \\ p_2 &= \frac{(N)(N-1)}{1 \times 2} a^2 p_0 \end{aligned}$$

In general

$$\begin{aligned} p_n &= \frac{N!}{(N-n)!n!} a^n p_0, \quad 1 \le n \le c \\ p_n &= \frac{N!}{(N-n)!c!c^{n-c}} a^n p_0, \quad c \le n \le N-K+1 \\ p_0 &= \left[\sum_{n=0}^{c-1} \binom{N}{n} a^n + \sum_{n=c}^{N-K+1} \frac{N!}{(N-n)!c!c^{n-1}} a^n \right]^{-1} \end{aligned}$$

Remark. If each component has its own server, i.e. $c = N - K + 1$, then we get

$$p_n = \binom{N}{n} a^n / \sum_{n=0}^{N-K+1} \binom{N}{n} a^n, \quad n = 0, 1, \ldots, N-K+1$$

In this case, the system is denoted by $M/G//\binom{N}{K}$, and probabilities $\{p_n\}$ are valid for any service time distribution (insensitivity phenomenon).

Performance Measures.

- $p_{N-K+1} = \dfrac{(N-K+1)!}{(K-1)!c!c^{N-K+1-c}}\, a^{N-K+1} p_0$

  is the unavailability of the system.

- System availability $= 1 - p_{N-K+1} = \sum_{n=0}^{N-K} p_n$.

Exercise. Find the system availability when $N = 5$, $\lambda = 3$, $\mu = 5$ and $K = 3$, $K = 2$, $K = 1$, i.e. (3 systems, 3-out-of-5, 2-out-of-5, and 1-out-of-5).

11 Double Ended Queue

This model is also called synchronization queue, fork/join station, or kitting process.

Consider a queueing model consisting of two finite input buffers, B_1 and B_2, fed by arrivals from two finite populations of sizes K_1 and K_2. The first population feeds the first buffer and the second population feeds the second buffer. As soon as there is one part in each buffer, two parts one from each buffer are joined and exit immediately. Therefore, at least one buffer has no parts at all times and parts in the other buffer wait until one part is available in each buffer.

This model can be descibed by the well known taxi-cab problem where taxis and passengers form two different queues, say the taxi queue and the passenger queue respectively. A passenger (taxi) who arrives at the taxi stand without finding taxis(passengers) in the taxi (passenger) queue has to wait in the passenger (taxi) queue and leaves as soon as a taxi (passenger) comes to the stand. This model has many applications in various areas. Examples include parallel processing, database concurrency control, flexible manufacturing systems, communication protocols and so on.

Typically one is interestred in computing the mean number of jobs in each buffer, system throughput, and characterization of the departure process, i.e. the distribution of time between inter-departures.

Let the time until each member in population $K_1(K_2)$ requests service i.e. joins the buffer is exponential with parameter $\lambda_1(\lambda_2)$. In this case the times between requests are exponential with parameters $n_i\lambda_i$, where n_i is the number of active elements in population $(K_i - \#in B_i), i = 1, 2$.

Note that a variation of this model is to consider two Poisson input processes and finite buffers. It can be shown that the system with two Poisson input processes and infinite buffers is unstable.

Model Analysis.

Let $X_i(t)$ be the number of units in buffer $B_i; i = 1, 2$. Note that $X(t) = X_1(t) - X_2(t)$ is a birth-death process with state space $S = \{-K_2, \ldots, 0, \ldots, K_1\}$. The transition rates are given by

$$
\begin{aligned}
q(i, i+1) &= K_1\lambda_1, -K_2 \leq i \leq 0 \\
&= (K_1 - i)\lambda_1, 0 \leq i \leq K_1 - 1 \\
q(i, i-1) &= (K_2 - i)\lambda_2, -K_2 + 1 \leq i \leq 0 \\
&= K_2\lambda_2, 0 \leq i \leq K_1
\end{aligned}
$$

Graph: State: $\{-K_2, \ldots, K_1\}$.

Now, we solve the birth death balance equations to compute p_n, $-K_2 \leq n \leq K_1$.

Performance Measures.

Let L_1 and L_2 be the mean number of units in buffer B_1 and B_2 respectively. Moreover, let λ_e be the effective arrival rate which is equal to the system throughput.

$$\begin{aligned}
L_1 &= \sum_{n=0}^{K_1} np_n \\
L_2 &= \sum_{n=-K_2}^{-1} -np_n = \sum_{n=1}^{K_2} np_{(-n)} \\
\lambda_e &= \sum_{n=-K_2}^{K_1} (q(n, n+1) + q(n, n-1))p_n
\end{aligned}$$

Modified State Description.

In stead of above let $X(t) = X_1(t) - X_2(t) + K_2$ which give a birth death process with state space $S = \{0, \ldots, K_1 + K_2\}$. Using the birth death model, the transition rates are given by $\lambda_i = q(i, i+1)$ and $\mu_i = q(i, i-1)$, where

$$\begin{aligned}
\lambda_n &= K_1\lambda_1, 0 \leq n \leq K_2 \\
&= (K_2 + K_1 - n)\lambda_1, K_2 \leq n \leq K_2 + K_1 - 1 \\
\mu_n &= n\lambda_2, 1 \leq n \leq K_2 \\
&= K_2\lambda_2, K_2 \leq i \leq K_2 + K_1\ .
\end{aligned}$$

Remark. Because we have a finite stae process, this B/D process is stable. Recall that the stationary distribution for any stable birth death process is given by

$$P_0 = \frac{1}{1 + \sum_{k=1}^{\infty} \prod_{j=1}^{k} \frac{\lambda_{j-1}}{\mu_j}} \tag{5.6}$$

$$P_n = (\prod_{j=1}^{n} \frac{\lambda_{j-1}}{\mu_j})P_0\ , n = 1, 2 \ldots\ . \tag{5.7}$$

Substituting λ_n and μ_n in the stationary equation formulas, we obtain

$$
\begin{aligned}
p_n &= \frac{K_1^n}{n!}(\frac{\lambda_1}{\lambda_2})^n p_0, 0 \le n \le K_2 \\
&= \frac{K_1^{K_2}\prod_{i=1}^{n}(K_2+i)}{K_2!K_2^{n-K_2}}(\frac{\lambda_1}{\lambda_2})^n p_0, K_2 \le n \le K_2+K_1
\end{aligned}
$$

$$
P_0 = [\sum_{n=0}^{K_2}\prod_{j=1}^{n}\frac{K_1^j}{j!}(\frac{\lambda_1}{\lambda_2})^n + \sum_{n=K_2+1}^{K_2+K_1+1}\frac{K_1^{K_2}\prod_{i=1}^{n}(K_2+i)}{K_2!K_2^{n-K_2}}(\frac{\lambda_1}{\lambda_2})^n]^{-1} \tag{5.8}
$$

Performance Measures with the Modified Model.

Let L_1 and L_2 be the mean number of units in buffer B_1 and B_2 respectively. Moreover, let λ_e be the effective arrival rate which is equal to the system throughput.

$$
\begin{aligned}
L_1 &= \sum_{n=0}^{K_2}(K_2-n)p_n \\
L_2 &= \sum_{n=0}^{K_1} np_(K_2+n) \\
\lambda_e &= \sum_{n=0}^{K_2+K_1}(\lambda_n+\mu_n)p_n
\end{aligned}
$$

Chapter 6

System Reliability

Contents.

1 Introduction

Definition. Reliability, $R(t)$, is usually defined as the probability that a device (system) performs adequately over the interval $[0, t]$.

Remark. $R(t)$ means adequate performance over $[0, t]$ not just at time t.

A system is a collection of components.

Examples: airplane, engines, brakes, etc.

Components can be independent or dependent (very complex).

Example: Automobile parts: wiring joints, brakes, transmission, exhaust, body, etc.

Structural function of a system.

Consider a system that is made up of n components (subsystems). Let the rv X_i represent the performance of component i, i.e.

$$\begin{aligned} X_i &= 1, \text{ if component } i \text{ performs adequately during } [0, t], \\ &= 0, \text{ if component } i \text{ fails during } [0, t] \end{aligned}$$

Let X be the performance of the system, i.e.

$$\begin{aligned} X &= 1, \text{ if system performs adequately during } [0, t], \\ &= 0, \text{ if system fails during } [0, t] \end{aligned}$$

Remarks.

(i) X is called the structure function of the system.

(ii) One can think of X as the result of a random experiment. (e.g. toss n coins and let $X = 1$ if at least k coins turn up heads, and 0 otherwise)

2 Types of Systems

Series Systems.

System fails if one component fails, equivalently, the system works iff all components work, in which case

$$X = X_1 X_2 \dots X_n = \min(X_1, \dots X_n)$$

Parallel Systems.

System functions if at least one component functions, equivalently, the system fails iff all components fail. The structure function is given by

$$X = \max(X_1, \dots X_n) = 1 - (1 - X_1)(1 - X_2) \dots (1 - X_n)$$

k out of n Systems.

System works if at least k out of n components function properly. The system function is given by

$$\begin{aligned} X &= 1, \; \sum_{i=1}^{n} X_i \geq k \\ &= 0, \; \sum_{i=1}^{n} X_i < k \end{aligned}$$

For a 2 out of 3 system, we have

$$X = X_1 X_2 X_3 + X_1 X_2 (1 - X_3) + X_1 (1 - X_2) X_3 + (1 - X_1) X_2 X_3$$

Remark.

(i) A series system is k out of n system with $k = n$.

(ii) A parallel system is k out of n system with $k = 1$.

3 System Reliability

Consider a system made up of n mutually independent components. Let $P(X_i = 1) = p_i$. Note that $E(X_i) = p_i$.

Definition. The system reliability r is defined as

$$r = P(X = 1) = E(X)$$

Series Systems.

The reliability of the system is given by

$$
\begin{aligned}
r = E(X) &= E(X_1 X_2 \ldots X_n) \\
&= E(X_1)E(X_2)\ldots E(X_n) \\
&= p_1 p_2 \ldots p_n
\end{aligned}
$$

Example. Suppose a system consists of $n = 4$ identical components linked in series. What must be the value of p so that $r = 0.95$?
Solution. We have to solve the following equation

$$p^4 = 0.95$$

which gives

$$p = 0.95^{1/4} = 0.987 \; .$$

Parallel Systems.
The reliability of the system is given by

$$
\begin{aligned}
r &= E(X) = E[1 - (1 - X_1)(1 - X_2)\ldots(1 - X_n)] \\
&= 1 - E[(1 - X_1)(1 - X_2)\ldots(1 - X_n)] \\
&= 1 - (1 - p_1)(1 - p_2)\ldots(1 - p_n)
\end{aligned}
$$

If the system is homogeneous, then

$$r = 1 - (1 - p)^n \; .$$

Example. Suppose $p = 0.95$. How many components must be linked in parallel so that the system reliability $r = 0.99$?
Solution. We must have

$$0.99 = 1 - (1 - 0.95)^n$$

that is

$$0.01 = 0.05^n$$

Taking logarithms

$$n = \ln 0.01 / \ln 0.05 = 1.5372$$

So $n = 2$.

Example. Suppose a system consists of $n = 3$ components linked in parallel. What must be the value of p so that $r = 0.99$?
Solution. We have to solve the following equation

$$1 - (1-p)^3 = 0.99$$

which gives

$$p = 1 - 0.01^{1/3} = 0.7846 .$$

k out of n Systems.

For the 2 out of 3 system, we have

$$r = E[X] = p_1p_2p_3 + p_1p_2(1-p_3) + p_1(1-p_2)p_3 + (1-p_1)p_2p_3 .$$

More generally,

$$r = P(\sum_{i=1}^{n} X_i \geq k)$$

Suppose all components are identical, i.e. $p_1 = p_2 = \ldots = p_n$. Then

$$r = \sum_{i=k}^{n} \binom{n}{i} p^i (1-p)^{n-i}$$

Example. Suppose a system consists of $n = 8$ iid components with $p = 0.95$. The system fails if at least 4 components failed. What is the reliability of this system?
Solution.

$$r = \sum_{i=4}^{8} \binom{8}{i} 0.95^i (0.05)^{8-i} = 0.9999$$

Remark. For systems that are non of the above the system function can be represented as a directed graph.

4 Reliability Using Failure Rates

Consider a single component. Let T be a rv that denotes the time to failure of the component. Suppose T has a pdf $f(t)$ and a df $F(t) = P(T \leq t)$.

The structure function X is then given by

$$\begin{aligned} X &= 1, T \geq t \\ &= 0, T < t. \end{aligned}$$

Therefore,

$$R(t) := P(X = 1) = P(T \geq t) = 1 - F(t).$$

Hazard rate: Let the hazard rate (failure rate) be given by

$$h(t) = \frac{f(t)}{1 - F(t)}$$

Interpretation: $h(t)dt$ is the probability that the system will fail in $(t, t + dt)$ given that it has survived until t.

Types of systems:

(i) systems with increasing failure rate (IFR)

(ii) systems with decreasing failure rate (DFR)

Remark. A system with exponential failure distribution is both IFR and DFR (why?)

FACT. Let T with df $F(t)$ be the time to failure of a system. Then

$$R(t) := 1 - F(t) = e^{-\int_0^t h(u)du}$$

Proof. Note that

$$h(t) = \frac{f(t)}{1 - F(t)} = -\frac{d}{dt} \ln(1 - F(t))$$

That is

$$\frac{d}{dt} \ln(1 - F(t)) = -h(t)$$

That is

$$\ln(R(t)) = -\int_0^t h(u)du$$

Thus

$$R(t) := 1 - F(t) = e^{-\int_0^t h(u)du}$$

Corollary 4.1

$$f(t) = -\frac{dR(t)}{dt} = h(t)e^{-\int_0^t h(u)du}$$

(Called generalized exponential form)

Bounds for IFR Distributions.

FACTS.

(i) For exponential distributions

$$R(t) = e^{-\lambda t}$$

(ii) Lower bound:

$$\begin{aligned} R(t) &\geq e^{-\lambda t}, t < 1/\lambda \\ & 0, t \geq 1/\lambda \end{aligned}$$

(iii) Upper bound

$$\begin{aligned} R(t) &\leq 1, t \leq 1/\lambda \\ & e^{-wt}, t > 1/\lambda, \end{aligned}$$

where w is the solution of $1 - \frac{w}{\lambda} = e^{-wt}$.

Having calculated (estimated) $R(t)$ for each component, we use the results of previous section to calculate the reliability of the whole system.

Sometimes we do not want to calculate a numerical value of reliability, but rather assess the effect of different component designs on total (system) reliability.

FACTS.

(i) Series structures of independent IFR components (not necessarily identical) are also IFR.

(ii) Series structures of independent DFR components (not necessarily identical) are also DFR.

(iii) K out of n systems of independent IFR identical components are also IFR.

(iv) Parallel structures of independent IFR identical components are also IFR.

Example. Find the failure rate and reliability for the Weibull df.

Solution. Recall the Weibull df.

$$F(t) = 1 - exp(-(t/\alpha)^{\beta})$$

Now,

$$h(t) = \lambda t^{\beta-1}$$

where λ is a constant given in terms of α and β.

Exercise.

Find the value of λ in terms of α and β.

$$R(t) = exp(-(t/\alpha)^{\beta})$$

Mean life time $= \int_0^{\infty} t f(t) dt = \int_0^{\infty} R(t) dt$.

Example. An aircraft has four identical engines each of which has a failure rate λ. For a successful flight at least two engines should be operating.

(i) What is the probability (reliability) that the aircraft will make a t hours successful flight?

Solution.

$$p = 1 - F(t) = e^{-\int_0^t \lambda du} = e^{-\lambda t}$$

$$R(t) = \sum_{k=2}^{4} \binom{4}{k} (e^{-\lambda t})^k (1 - e^{-\lambda t})^{4-k}$$

Simplify (exercise),

$$R(t) = 3e^{-4\lambda t} - 8e^{-3\lambda t} + 6e^{-2\lambda t}$$

(ii) What is the expected life-time, L, of the aircraft?

Solution.

$$\begin{aligned} E(L) &= \int_0^\infty R(t)dt \\ &= \frac{3}{4\lambda} - \frac{8}{3\lambda} + \frac{6}{2\lambda} \\ &= \frac{13}{12\lambda} \end{aligned}$$

Example. Suppose the aircraft is such that it needs at least one engine on each side for a successful flight.

(i) Find new $R(t)$.

Graph

Solution.

First, note that

$$\begin{aligned} X &= [1 - (1 - X_1)(1 - X_2)]^2 \\ EX &= [1 - (1 - p)(1 - p)]^2 . \end{aligned}$$

Therefore,

$$\begin{aligned} R(t) &= [1 - (1 - e^{-\lambda t})^2]^2 \\ &= [2e^{-\lambda t} - e^{-2\lambda t}]^2 \\ &= e^{-4\lambda t} - 4e^{-3\lambda t} + 4e^{-2\lambda t} \end{aligned}$$

(ii) Find the new expected life-time.

Solution.

exercise

Remark. For k out of n system with $p = e^{-\lambda t}$

$$R(t) = \sum_{i=k}^{n} \binom{n}{i} (e^{-\lambda t})^i (1 - e^{-\lambda t})^{n-i}$$

Chapter 7

Appedix

Contents.

1 Final Project Guidelines

Abstract (one page)

1. Model description
2. Results, findings
3. Significance

Project Details

1. Introduction
2. Problem statement (show pictorially)
3. Sources of randomness, input data (distributions)
4. Assumptions
5. Resulting model
6. Objective, justification
7. Analysis (show all details)
8. Validation and verification
9. Model results/Conclusions
10. Sensitivity Analysis/ "What" if analysis
11. Real Problem Conclusions
12. Recommendations
13. Possible extensions