MA 597 Queueing Theory and Applications

(January - May 2021 Semester)

N. Selvaraju

Department of Mathematics
Indian Institute of Technology Guwahati

Queues, Queueing Systems and Queueing Theory

- Stochastic modelling is the application of probability theory to the description and analysis of real world phenomena.
- It is a science with close interaction between theory and practical applications.
- One of the most important domains in stochastic modelling is the field of **queueing theory**[1], and this shall be the topic of this course.
- Many real systems can be reduced to components which can be modelled by the concept of a so-called queue or waiting line. And, queueing theory is the mathematical study of the lines. More generally, queueing theory is concerned with the mathematical modelling and analysis of systems that provide service to random demands.
- Basic idea borrowed from the day-to-day experience of waiting in lines – in the grocery store, on the telephone, at the airport, on the road, etc.
- A queue consists of a system into which there comes a stream of users who demand some capacity of the system over a certain time interval before they leave the system again. It is said that the users are served in the system by one or many servers.

[1] Note: The spelling 'queuing theory' is also used by some.

- A queueing system can thus be described by a (stochastic) specification of the arrival stream and of the system demand for every user as well as a definition of the service mechanism. The former describe the input into a queue, while the latter represents the functioning of the inner mechanisms of a queueing system.
- Because lines form from arrival and service processes that are typically random, queueing theory relies on the mathematical study of stochastic processes.
- Some common types of questions that queueing theory can address are:
  - What is the average time spent waiting in line?
  - How long are the lines on average?
  - How many customers wait more than 2 minutes?
  - How many customers are turned away?
  - How many servers are needed to achieve a target quality of service (QoS)?
  - How fast must the servers work?

  Answers to these questions provide decision makers a way to efficiently manage or design a system.
- Our focus will be dealing with various quantitative models used to analyze queueing systems and their mathematical analysis.
- We assume a prior knowledge of probability theory. Knowledge of continuous-time Markov chains (CTMCs) helpful (key aspects of CTMCs will be presented briefly).

Origin and Applications

- The first real application of queueing theory, in fact the one that engendered the development of the whole field of research, has been the design and analysis of telephone networks.
- At the beginning of the 20th century, telephone calls first went to an operator before they could be connected to the person that was to be reached by the call. Thus an important part of a telephone network could be modelled by a queueing system in which the servers are the operators in a call center who connect the incoming calls (which are modelled by the input stream of users) to their addressees. Here, the time of connecting is represented by the service demand of a user.
- A crucial performance measure of such a system is the probability that a person who wants to get a connection for a call finds all operators busy and thus cannot be served. This value is called the loss probability of the system.
- For a modern-day call centre, where questions are answered instead of cables connected, the service times represent the duration of the call between the user and the operator.

- Agner Krarup Erlang, a Danish mathematician, statistician and engineer, who proposed the formula published his work “The Theory of Probabilities and Telephone Conversations” in 1909. He gained worldwide recognition for this and other works.
- During the 1940s, the *Erlang* became the accepted unit of telecom traffic measurement, and his formula is still used today in the design of modern networks.
- When ARPANET was being considered, the pioneers of this precursor of the Internet used the queueing theory advanced by Erlang and others to show that the system was feasible. One of those involved was Leonard Kleinrock.
- **Queueing theory**, being intricate and yet highly practical field of mathematical study has vast applications in fields like
  - Computer systems and networks (internet, etc)
  - Communication systems and networks (mobile, call center, etc)
  - Manufacturing/service systems and networks (assembly line, supply chain, etc)
  - Healthcare systems (emgergency care, etc)
  - Airport/seaport traffic
  - Military logistics, theme parks, supermarkets, inventories, etc.

The queueing community has a cyberhome:

http://web2.uwindsor.ca/math/hlynka/queue.html

MA597 Queueing Theory and Applications : Course

- Contents: Review of probability, random variables, distributions, generating functions; Poisson, Markov, renewal and semi-Markov processes;
  Characteristics of queueing systems, Little’s law, Markovian and non-Markovian queueing systems, embedded Markov chain applications to M/G/1, G/M/1 and related queueing systems;
  Networks of queues, open and closed queueing networks;
  Queues with vacations, priority queues, queues with modulated arrival process, discrete time queues, introduction to matrix-geometric methods;
  Applications in manufacturing, computer and communication networks.
- Main Text:
  J.F. Shortle, J.M. Thompson, D. Gross and C. Harris, Fundamentals of Queueing Theory, Fifth Edition, Wiley, 2018. (Fourth edition available as Indian edition).
- Other Texts and References:

  - **J. Medhi, Stochastic Models in Queueing Theory, 2nd Edition, Academic Press, 2003.**
  - **L. Kleinrock, Queueing Systems, Vol. 1: Theory, Wiley, 1975.**
  - **J.A. Buzacott and J.G. Shanthikumar, Stochastic Models of Manufacturing Systems, Prentice Hall, 1992.**
  - **R.B. Cooper, Introduction to Queueing Theory, 2nd Edition, North-Holland, 1981.**
  - **R. Nelson, Probability, Stochastic Processes, and Queueing Theory: The Mathematics of Computer Performance Modelling, Springer, 1995.**
  - **E. Gelenbe and G. Pujolle, Introduction to Queueing Networks, 2e, Wiley, 1998.**

Lecture Slots and Evaluation

- The course slot is E-slot (Wednesday and Thursday) and F-slot (Tuesday).
- Interaction sessions and Quizzes/Exams will also be conducted through Microsoft Teams during these slots.
- The mode evaluation pattern is mainly online, and offline mode will also be explored.
- The online quizzes will have all type of question patterns like MCQ with single and multiple correct answers, numerial answer types, answers in one or two sentences, short descriptive answers to be typed online, and also pen-and-paper answers scanned and uploaded online. Other types may be adopted depending on the need.
- If both (online and offline) mode were adopted for a particular evaluation, a suitable normalization will be done.
- The overall evaluation scheme will be dynamic and would change depending on the changing situations. Typically, one quiz in 2-3 weeks will be conducted.

Murphy’s Laws of Queues

- If you change queues, the one you have left will start to move faster than the one you are in now.
- Your queue always goes the slowest.
- Whatever queue you join, no matter how short it looks, will always take the longest for you to get served.

For more, visit the queueing theory webpage.

Beware! Not every queueing like problem can be solved by queueing theory techniques!

Queue for ...!

Probability Theory Refresher

- Every stochastic model has an underlying probability space that consists of:
  1. the sample space Ω – the collection of all possible outcomes of a random experiment
  2. a σ-field (or a σ-algebra) $\mathcal{F}$ over Ω – the collections of 'events' (A non-empty collection of subsets of Ω closed under countable union and complementation)
  3. a set function $P : \mathcal{F} \to \mathbb{R}$ called a probability measure (or probability function or simply probability) that satisfies
     - ★ $P(E) \geq 0$ for all $E \in \mathcal{F}$
     - ★ $P(\Omega) = 1$
     - ★ Let $E_1, E_2, \ldots \in \mathcal{F}$ be a sequence of disjoint events then $P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$

  The triplet $(\Omega, \mathcal{F}, P)$ is called a probability space.
- A simple example: Consider the random experiment of tossing of a coin, where sample space is $\Omega = \{H, T\}$ and $\mathcal{F} = \mathcal{P}(\Omega)$, the power set of Ω. Consider a function $P : \mathcal{F} \to \mathbb{R}$ defined by

$$P(\Omega) = 1,\ P(\{H\}) = 0.6,\ P(\{T\}) = 0.4, \text{ and } P(\emptyset) = 0.$$

- ▶ A singleton event $\omega \in \Omega$ is called an elementary event.
- ▶ If Ω is finite, and $\mathcal{F} = \mathcal{P}(\Omega)$, the power set of Ω, it is sufficient to assign probabilities to each elementary event. Then for any $E \in \mathcal{F}$, $P(E) = \sum_{\omega \in E} P(\{\omega\})$. If the elementary events are equally likely, then we get the classical definition of probability (uniform probability space).
- ▶ If Ω is countable, and $\mathcal{F} = \mathcal{P}(\Omega)$, it is still sufficient to assign probabilities to each elementary event. Then for any $E \in \mathcal{F}$, $P(E) = \sum_{\omega \in E} P(\{\omega\})$. However, in this case we can not assign equal probability to each elementary event.
- ▶ If Ω is uncountable, and $\mathcal{F} = \mathcal{P}(\Omega)$, one can not make an equally likely assignment of probabilities. Indeed, one can not assign positive probability to each elementary event without violating the axiom $P(\Omega) = 1$.
- ▶ Thus, the choice of $\mathcal{F}$ is an important issue. Depending on our objective, we may need to choose a σ-field over Ω (that is smaller than $\mathcal{P}(\Omega)$) on which to define a probability P.

- Conditional Probability: Let B be an event with $P(B) > 0$. For any arbitrary event A, the conditional probability of A given B is defined by $P(A|B) = \frac{P(A \cap B)}{P(B)}$. $P(\cdot|B)$ is also a probability measure (and induces a probability space).
- Theorem of Total Probability:
  Defn: A collection of events $\{E_1, E_2 \ldots\}$ is said to be mutually exclusive if $E_i \cap E_j = \phi, \forall i \neq j$. It is said to be exhaustive if $P(\cup_i E_i) = 1$.
  Result: Let $\{E_1, E_2 \ldots\}$ be a collection of mutually exclusive and exhaustive events with $P(E_i) > 0, \forall i$. Then, for any event A, $P(A) = \sum_i P(A|E_i)P(E_i)$.
- Bayes' Theorem: Let $\{E_1, E_2 \ldots\}$ be a collection of mutually exclusive and exhaustive events with $P(E_i) > 0, \forall i$. Let A be any event with $P(A) > 0$. Then $P(E_i|A) = \frac{P(A|E_i)P(E_i)}{\sum_j P(A|E_j)P(E_j)}, \quad i = 1, 2, \ldots$.
- Independence:
  Defn: A finite collection of events $E_1, E_2, \ldots, E_n$ are said to be independent(or mutually independent) if for any sub-collection $E_{n_1}, \ldots, E_{n_k}$ of $E_1, E_2, \ldots, E_n$, $P\left(\bigcap_{i=1}^{k} E_{n_k}\right) = \prod_{i=1}^{k} P(E_{n_i})$.
  Defn: A countable collection of events $E_1, E_2, \ldots$ are said to be independent if any finite sub-collection are independent.
  Defn: A countable collection of events $E_1, E_2, \ldots$ are said to be pairwise independent if E_i and E_j are independent for $i \neq j$.
  Defn: Given an event C, two events A and B are said to be conditionally independent if $P(A \cap B|C) = P(A|C)P(B|C)$.

Random Variables and Distributions

- Random Variable: Let $(\Omega, \mathcal{F}, P)$ be a probability space. A function $X : \Omega \to \mathbb{R}$ is called a random variable if $X^{-1}((-\infty, a]) \in \mathcal{F}$ for all $a \in \mathbb{R}$.
  - ▶ The random variable X is used to 'push-forward' the measure P on Ω to a measure on $\mathbb{R}$.
- Distribution Function: The distribution function (DF) of a random variable X is a function $F_X : \mathbb{R} \to [0, \infty)$ defined by $F_X(x) = P(X \leq x) = P(X^{-1}((-\infty, x]))$.
  - ▶ Recall properties of F_X.
- Random variable is just a function and does not depend on the probability. But the distribution of the random variable depends on the probability. So keeping the function same if we change the probability then the random variable will remain the same but its distribution will change.
- Discrete Random Variable: A random variable is said to have a discrete distribution if there exists an at most countable set $S_X \subseteq \mathbb{R}$ such that $P(X = x) > 0$ for all $x \in S_X$ and $\sum_{x \in S_X} P(X = x) = 1$.
  Here, S_X is called the support of X. The function $p_X : \mathbb{R} \to [0, 1]$ defined by $p_X(x) = \begin{cases} P(X = x) & \text{if } x \in S_X \\ 0 & \text{otherwise} \end{cases}$ is called the probability mass function of X.
  And, $F_X(x) = \sum_{\substack{y \in S_X \\ y \leq x}} p_X(y)$ and $p_X(x) = F_X(x) - F_X(x-)$.

- Continuous Random Variable: A random variable is said to have a continuous distribution if there exists a non-negative integrable function $f_X : \mathbb{R} \to [0, \infty)$ such that $F_X(x) = \int_{-\infty}^{x} f_X(t)dt$ for all $x \in \mathbb{R}$.
  The function f_X is called the probability density function. The set $S_X = \{x \in \mathbb{R} : f_X(x) > 0\}$ is called support of X.
- Recall properties of discrete/continuous random variables, and that of pmf/pdf.
- Examples:
  - (Binomial Distribution: $Bin(n\,, p)$) $S_X = \{0, 1, \ldots, n\}$, $p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$.
  - (Geometric Distribution: $Geo(p)$) $S_X = \{0, 1, \ldots\}$, $p_X(k) = p(1-p)^k$.
  - (Poisson Distribution: $Poi(\lambda)$ $(\lambda > 0)$) $S_X = \{0, 1, \ldots\}$, $p_X(k) = \frac{e^{-\lambda}\lambda^k}{k!}$.
  - (Exponential Distribution: $Exp(\lambda)$) $F_X(x) = \begin{cases} 1 - e^{-\lambda x} & \text{if } x > 0 \\ 0 & \text{otherwise.} \end{cases}$
  - (Uniform Distribution: $U(a, b)$) $F_X(x) = \begin{cases} 0 & \text{if } x < a \\ \frac{x-a}{b-a} & \text{if } a \le x < b \\ 1 & \text{if } x \ge b. \end{cases}$
  - (Normal Distribution: $\mathcal{N}(\mu, \sigma^2)$) $F_X(x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt \quad$ if $x \in \mathbb{R}$.
  - Recall other well-known distributions as well, like Gamma, etc.

Expectations

- Let X be a discrete RV with PMF $p_X(\cdot)$ and support S_X. The expectation or mean of X is defined by $E(X) = \sum_{x \in S_X} x p_X(x)$ provided $\sum_{x \in S_X} |x| p_X(x) < \infty$.
  Let X be a continuous RV with PDF $f_X(.)$. The expectation of X is defined by $E(X) = \int_{-\infty}^{\infty} x f_X(x)dx$ provided $\int_{-\infty}^{\infty} |x| f_X(x)dx < \infty$.
- Both the above can be written using Stieltjes integral as $E(X) = \int_{-\infty}^{\infty} x dF_X(x)$ provided $\int_{-\infty}^{\infty} |x| dF_X(x) < \infty$.
- If X is a random variable then $Y = g(X)$ is a random variable, where $g : \mathbb{R} \to \mathbb{R}$ is a 'nice' function. Then, an important result is that
  $E\left[g(X)\right] = \int_{-\infty}^{\infty} g(x)dF_X(x)$ provided $\int_{-\infty}^{\infty} |g(x)|dF_X(x) < \infty$.
- Recall properties of expectations.

Special Expectations

- For $r = 1, 2, \ldots$, $\mu_r = E(X^r)$ is called rth raw moment of X, if the expectation exists. In particular, μ_1 is called the mean and denoted by μ.
- $\mu_r' = E\left[(X - E(X))^r\right]$ is called rth central moment of X, if the expectations exist. In particular, $\mu_2' = E\left[(X-\mu)^2\right]$ is called the variance of X and denoted by σ^2 or $Var(X)$.
- Coefficient of Variation: $C_X = \sigma/\mu$.
- (Moment Generating Function or MGF) The moment generating function of a random variable X is defined by $M_X(t) = E\left(e^{tX}\right)$ provided the expectation exists in a neighbourhood of the origin.
  If the MGF $M_X(t)$ exist for $t \in (-a, a)$ for some $a > 0$, the derivatives of all order exist at $t = 0$ and $E\left(X^k\right) = \frac{d^k}{dt^k} M_X(t)\Big|_{t=0}$ for all positive integer k.
- (Moment Inequality) Let X be a RV and $g : [0, \infty) \to [0, \infty)$ be a non-decreasing function such that $E\left(g(|X|)\right)$ is finite. Then for any $c > 0$ with $g(c) > 0$, then $P(|X| \ge c) \le \frac{E(g(|X|))}{g(c)}$.
  Markov Inequality: Let X be a RV with $E\left(|X|^r\right) < \infty$ for some $r > 0$. Then for any $c > 0$, $P(|X| \ge c) \le \frac{E(|X|^r)}{c^r}$.
  Chebyshev Inequality: Let X be a RV with $E(X^2) < \infty$. Let us denote $\mu = E(X)$ and $\sigma^2 = Var(X)$. Then for any $k > 0$, $P(|X - \mu| \ge k) \le \frac{\sigma^2}{k^2}$.

Jointly Distributed Random Variables

- A multivariate random variable or a random vector is a vector $\mathbf{X} = (X_1, ..., X_n)$ whose components are scalar-valued random variables on the same probability space as each other, $(\Omega, \mathcal{F}, P)$.
- For any random vector $\mathbf{X} = (X_1, X_2, \ldots, X_n)$, the joint distribution function is defined by

$$F_{\mathbf{X}}(x_1, \ldots, x_n) = P\left(X_1 \le x_1, \ldots, X_n \le x_n\right), \quad (x_1, \ldots, x_n) \in \mathbb{R}^n.$$

- Marginal DFs: $F_X(x) = \lim_{y\to\infty} F_{X,Y}(x, y)$ and $F_Y(y) = \lim_{x\to\infty} F_{X,Y}(x, y)$.
- Recall properties of joint DFs.

- A random vector (X, Y) is said to have a discrete distribution if there exists an atmost countable set $S_{X,Y} \in \mathbb{R}^2$ such that $P((X, Y) = (x, y)) > 0$ for all $(x, y) \in S_{X,Y}$ and $P((X, Y) \in S_{X,Y}) = 1$.
  $S_{X,Y}$ is called the support of (X, Y). And, the joint probability mass function of (X, Y) is given by a function $p_{X,Y} : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$p_{X,Y}(x, y) = \begin{cases} P(X = x, Y = y) & \text{if } (x, y) \in S_{X,Y} \\ 0 & \text{otherwise.} \end{cases}$$

  $p_X(x) = \sum_{(x,y) \in S_{X,Y}} p_{X,Y}(x, y)$, for fixed $x \in \mathbb{R}$, is called the marginal PMF of X.
- A random vector (X, Y) is said to have a continuous distribution if there exists a non-negative integrable function $f_{X,Y} : \mathbb{R}^2 \to \mathbb{R}$ such that

$$F_{X,Y}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(t, s) ds dt, \quad (x, y) \in \mathbb{R}^2,$$

  where $f_{X,Y}$ is called the joint probability density function of (X, Y) and the set $S_{X,Y} = \left\{(x, y) \in \mathbb{R}^2 : f_{X,Y}(x, y) > 0\right\}$ is called the support of (X, Y).
  $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy$, for fixed $x \in \mathbb{R}$, is called the marginal PDF of X.
- Recall properties of joint/marginal PMFs/PDFs, and the definition and properties $E(g(\mathbf{X}))$.

Independent Random Variables

- The random variables $X_1, X_2, \ldots, X_n$ are said to be independent if

$$F_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} F_{X_i}(x_i), \quad \forall (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n.$$

  In particular, X and Y are independent iff the events $A_x = \{X \le x\}$ and $B_y = \{Y \le y\}$ are independent for all $(x, y) \in \mathbb{R}^2$.
  For discrete/continuous (X, Y), the condition of independence can be written equivalently in terms of their PMFs/PDFs.

- If X and Y are independent and g and h are 'nice' functions, then

$$E(g(X)h(Y)) = E(g(X)) E(h(Y)),$$

  provided all the expectations exist.
- The covariance of two random variables X and Y is defined by

$$Cov(X, Y) = E[(X - E(X))(Y - E(Y))] = E(XY) - E(X)E(Y).$$

- The correlation coefficient of X and Y is defined by

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}}.$$

- Recall properties of joint/marginal moments of functions of random vectors, variances, covariances, and correlation coefficients (with/without the independence condition).

Joint Moment Generating Function

- Let $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ be a RV. The moment generating function (MGF) of $\mathbf{X}$ at $\mathbf{t} = (t_1, t_2, \ldots, t_n)$ is defined by

$$M_{\mathbf{X}}(\mathbf{t}) = E\left(\exp\left(\sum_{i=1}^{n} t_i X_i\right)\right),$$

  provided the expectation exists.
  $E(X_1^{r_1} X_2^{r_2} \cdots X_n^{r_n}) = \frac{\partial^{r_1+r_2+\ldots+r_n}}{\partial t_1^{r_1} \partial t_2^{r_2} \ldots \partial t_n^{r_n}} M_{\mathbf{X}}(\mathbf{t})\Big|_{\mathbf{t}=0}$.
  X and Y are independent iff $M_{X,Y}(t_1, t_2) = M_X(t_1)M_Y(t_2)$.

Conditional Distributions

- Let (X, Y) have a joint PMF $p_{X,Y}(\cdot,\cdot)$ and suppose that the marginal PMF of Y is $p_Y(\cdot)$. The conditional PMF of X, given $Y = y$ is defined by
$$p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}, \text{ provided } p_Y(y) > 0,$$
and the conditional DF of X given $Y = y$ is defined by
$$F_{X|Y}(x|y) = P(X \le x|Y = y) = \sum_{\{u \le x : (u,y) \in S_{X,Y}\}} p_{X|Y}(u|y), \text{ provided } p_Y(y) > 0.$$
- Let (X, Y) be continuous. The conditional DF of X given $Y = y$ is defined as
$$F_{X|Y}(x|y) = \lim_{\epsilon \downarrow 0} P(X \le x|Y \in (y - \epsilon, y + \epsilon]),$$
provided the limit exists and the conditional PDF of X given $Y = y$, $f_{X|Y}(x|y)$, is defined as the non-negative function satisfying
$$F_{X|Y}(x|y) = \int_{-\infty}^{x} f_{X|Y}(t|y)dt, \qquad \forall x \in \mathbb{R}.$$
Result: Let $f_{X,Y}$ be the JPDF of (X, Y) and let f_Y be the marginal PDF of Y. If $f_Y(y) > 0$, then the conditional PDF of X given $Y = y$ exists and is given by
$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}.$$

Conditional Expectations

- (Discrete Case) The conditional expectation of $h(X)$ given $Y = y$ is defined by
$$E(h(X)|Y = y) = \sum_{x:(x,y) \in S_{X,Y}} h(x) f_{X|Y}(x|y),$$
provided it is absolutely summable.
(Continuous Case) The conditional expectation of $h(X)$ given $Y = y$, is defined for all values of y such that $f_Y(y) > 0$, by
$$E(h(X)|Y = y) = \int_{-\infty}^{\infty} h(x) f_{X|Y}(x|y)dx,$$
provided it is absolutely integrable.
- Suppose that (X, Y) is a RV. Define $E(X|Y) = g(Y)$, where $g(y) = E(X|Y = y)$. Thus $E(X|Y)$ is again a random variable.
  - $E(X) = E(E(X|Y))$.
  - $E(X - E(X|Y))^2 \le E(X - f(Y))^2$ for any function f. Thus $E(X|Y)$ is the "best estimate of X given Y".

- (Conditional Variance) Let (X, Y) be a random vector. Then $Var(X|Y) = h(Y)$, where $h(y) = E((X - E(X|Y))^2|Y = y)$
$= E(X^2|Y = y) - (E(X|Y = y))^2$.
  - $Var(X) = E(Var(X|Y)) + Var(E(X|Y))$.
- (Computing Probability by Conditioning)
$$P(E) = \begin{cases} \sum_y P(E|Y = y)P(Y = y) & \text{for } Y \text{ discrete} \\ \int_{-\infty}^{\infty} P(E|Y = y) f_Y(y)dy & \text{for } Y \text{ continuous.} \end{cases}$$
- (Conditional Expectation given an Event)
Let (X, Y) be a random vector. Then
$$E\left[h(X, Y)|(X, Y) \in A\right] = \frac{E[h(X, Y)I_A(X, Y)]}{P((X, Y) \in A))}.$$

- Recall other special expectations like characteristic functions, modes of convergence, limit theorems like LLN and CLT.
- Recall some miscellaneous ideas like handling random sum of random variables, computation of expectation by conditioning, functions of random vectors especially sum of random variables.

Discrete probability distributions, moments and generating functions

Name	Probability Function $p(x)$, discrete	Parameters	Mean $\mathrm{E}[X]$	Variance $\mathrm{E}[X-\mathrm{E}[X]]^2$	Moment Generating Function $\mathrm{E}[e^{tx}]$	Probability Generating Function $\sum_{m=0}^{\infty} p(m)z^m$
Bernoulli	$p(x) = \begin{cases} p & (x=1) \\ 1-p & (x=0) \end{cases}$	$0 \le p \le 1$	p	$p(1-p)$	$pe^t + 1 - p$	$1 - p + pz$
Binomial	$p(x) = \binom{n}{x} p^x (1-p)^{n-x}$	$n = 1, 2, \cdots$ $0 \le p \le 1$	np	$np(1-p)$	$(pe^t + 1 - p)^n$	$(pz + 1 - p)^n$
Poisson	$p(x) = \frac{e^{-\lambda}\lambda^x}{x!}$	$\lambda > 0$	λ	λ	$e^{\lambda(e^t - 1)}$	$e^{-\lambda(1-z)}$
Geometric	$p(x) = p(1-p)^x$	$0 \le p \le 1$	$\frac{1-p}{p}$	$\frac{1-p}{p^2}$	$\frac{p}{1-(1-p)e^t}$	$\frac{p}{1-(1-p)z}$
Negative binomial	$p(x) = \binom{k+x-1}{x} p^k (1-p)^x$	$k > 0$ $0 \le p \le 1$	$\frac{k(1-p)}{p}$	$\frac{k(1-p)}{p^2}$	$\left(\frac{p}{1-(1-p)e^t}\right)^k$	$\left(\frac{p}{1-(1-p)z}\right)^k$

The Characteristic Function, $E[e^{it}]$, can be obtained from the Moment Generating Function by replacing t with it.

Continuous probability distributions, moments and generating functions

Name	Probability Function $f(x)$, continuous	Parameters	Mean $\mathrm{E}[X]$	Variance $\mathrm{E}[X-\mathrm{E}[X]]^2$	Moment Generating Function $\mathrm{E}[e^{tx}]$
Uniform	$f(x) = \frac{1}{b-a}$	$-\infty < a < b < \infty$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{e^{tb} - e^{ta}}{t(b-a)}$
Normal	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}$	$-\infty < \mu < \infty$ $\sigma > 0$	μ	σ^2	$e^{t\mu + (t^2\sigma^2)/2}$
Exponential	$f(x) = \theta e^{-\theta x}$	$\theta > 0$	$\frac{1}{\theta}$	$\frac{1}{\theta^2}$	$\frac{\theta}{\theta - t}$
Gamma	$f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}$	$\alpha, \beta > 0$	$\alpha\beta$	$\alpha\beta^2$	$\left(\frac{1/\beta}{1/\beta - t}\right)^\alpha$
Erlang-k	$f(x) = \frac{(\theta k)^k}{(k-1)!} x^{k-1} e^{-k\theta x}$	$\theta > 0$ $k = 1, 2, \cdots$	$\frac{1}{\theta}$	$\frac{1}{k\theta^2}$	$\left(\frac{k\theta}{k\theta - t}\right)^k$
2-Term hyperexponential	$f(x) = p\theta_1 e^{-\theta_1 x}$ $+(1-p)\theta_2 e^{-\theta_2 x}$	$0 < p < 1$ $\theta_1, \theta_2 > 0$	$\frac{p}{\theta_1} + \frac{1-p}{\theta_2}$	$\frac{p}{\theta_1^2} + \frac{1-p}{\theta_2^2} - \frac{2p(1-p)}{\theta_1\theta_2}$	$\frac{p\theta_1}{\theta_1 - t} + \frac{(1-p)\theta_2}{\theta_2 - t}$
Chi-square	$f(x) = \frac{1}{2^{n/2}\Gamma(n/2)} x^{(n/2)-1} e^{-x/2}$	$n = 1, 2, \cdots$	n	$2n$	$\left(\frac{1/2}{1/2 - t}\right)^{n/2}$
Beta	$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$	$\alpha, \beta > 0$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)} \cdot \sum_{j=0}^{\infty} \frac{\Gamma(\alpha+j)(t)^j}{\Gamma(\alpha+\beta+j)\Gamma(j+1)}$

The Characteristic Function, $E[e^{it}]$, can be obtained from the Moment Generating Function by replacing t with it.

Laplace Transforms

- A useful transform (which is a mapping from one space to other) in queueing analysis is the LT.
- Let $f : [0, \infty) \to \mathbb{R}$ be a 'suitable' function. The Laplace transform (LT) of f is

$$\mathcal{L}\{f(t)\} = \bar{f}(s) = \int_0^\infty e^{-st} f(t)dt, \quad (t \ge 0)$$

  where s is a complex variable. Under broad conditions, it can be shown that $\bar{f}(s)$ is analytic in the halfplane where $Re(s) > \alpha$, for some constant α.
  ♦ One-to-one correspondence exists between $f(t)$ and $\bar{f}(s)$.
- Some common pairs of the function and its LT are: $(1, \frac{1}{s})$, $(t, \frac{1}{s^2})$, $(t^n, \frac{n!}{s^{n+1}}$ for $n = 0, 1, 2, \dots)$, $(e^{-at}, \frac{1}{s+a})$, $(e^{-at}t^n, \frac{n!}{(s+a)^{n+1}}$ for $n = 0, 1, 2, \dots)$, $(\cos bt, \frac{s}{s^2+b^2})$, $(\sin bt, \frac{b}{s^2+b^2})$
- Properties:
  - $\mathcal{L}\{a_1 f_1(t) + a_2 f_2(t)\} = a_1\bar{f}_1(s) + a_2\bar{f}_2(s)$
  - $\mathcal{L}\{e^{-at}f(t)\} = \bar{f}(s+a)$ and $\mathcal{L}\{f(at)\} = \frac{1}{a}\bar{f}(\frac{s}{a})$
  - $\mathcal{L}\{f^{(n)}(t)\} = s^n\bar{f}(s) - \sum_{i=1}^{n} s^{i-1} f^{(i-1)}(0)$, for derivative $f^{(n)}(t)$ of order $n \ge 1$
  - $\mathcal{L}\{\int_0^t f(x)dx\} = \frac{\bar{f}(s)}{s}$
  - $\mathcal{L}\{\int_0^t g(t-y)f(y)dy\} = \bar{f}(s)\bar{g}(s)$
  - (Limit property) $\lim_{t\to\infty} f(t) = \lim_{s\to 0} s\bar{f}(s)$ and $\lim_{t\to 0} f(t) = \lim_{s\to\infty} s\bar{f}(s)$

Laplace-Stieltjes Transforms

- A related transform is the Laplace-Stieltjes transform (LST) which is defined, for a function $F : [0, \infty) \to \mathbb{R}$, as

$$\mathcal{L}^*\{F(t)\} = F^*(s) = \int_0^\infty e^{-st} dF(t),$$

  where the integral is the Lebesgue-Stieltjes integral.
- For our purposes, we consider F to be the DF of a nonnegative random variable X, so $F^*(s) = E[e^{-sX}] = \int_0^\infty e^{-st}dF(t)$. [And, PDF for f in LT.]
- Examples: (A) For an exponential random variable with mean $\frac{1}{\lambda}$, $F^*(s) = \frac{\lambda}{s+\lambda}$.
  (B) For a discrete random variable X with $P(X = 3) = \frac{3}{10}$, $P(X = 4) = \frac{1}{5}$ and $P(X = 9) = \frac{1}{2}$, $F^*(s) = \frac{3}{10}e^{-3s} + \frac{1}{5}e^{-4s} + \frac{1}{2}e^{-9s}$.
- When a (nonnegative) RV X has a PDF, then the LST of the DF equals the LT of the PDF, i.e., $F^*(s) = \bar{f}(s)$.
  - More generally, $F^*(s) = s\bar{F}(s)$ for any DF F of a nonnegative RV.

- Example: Let $F(t) = 1 - \rho e^{-\mu(1-\rho)}, t \geq 0$
  [Note: This has a point mass $1-\rho$ at $t = 0$ and continuous and differentiable for $t > 0$].
  Using definition, the LST of F is

$$F^*(s) = \int_0^\infty e^{-st} dF(t) = 1 - \rho + \rho \int_0^\infty e^{-st} \mu(1-\rho) e^{-\mu(1-\rho)t} dt = \frac{(s+\mu)(1-\rho)}{s+\mu(1-\rho)}.$$

  ▶ Exercise: Obtain the above using $F^*(s) = s\bar{F}(s)$.
- Note: Similar properties as that of LT. And, $F(t)$ can be uniquely determined from $F^*(s)$.
- (Continuity) Let $X_n, n = 1, 2, \ldots$ be a sequence of RVs with DFs $F_n(t)$ and LSTs $F_n^*(s)$. If, as $n \to \infty$, F_n tends to a distribution function F having LST $F^*(s)$, then as $n \to \infty$, $F_n^*(s) \to F^*(s)$ for $s > 0$, and conversely.
  ♦ Useful in finding out the limiting form of certain distributions.

Moment Generating Functions and Characteristic Functions

- Recall, for nonnegative X, that $M_X(t) = E[e^{tX}] = \int_0^\infty e^{tx} dF(x)$ (if exists).
  ▶ This is similar to LST of F, with $-s$ replacing t, i.e., $M_X(t) = F^*(-t)$.
  ■ For $Exp(\lambda)$ random variable, $M_X(t) = \frac{\lambda}{\lambda-t}$, for $t < \lambda$, and $F^*(s) = \frac{\lambda}{\lambda+s}$.
- Recall, for nonnegative X, that $\phi_X(t) = E[e^{itX}] = \int_0^\infty e^{itx} dF(x)$.
  ▶ This is similar to LST of F, with $-s$ replacing it, i.e., $\phi_X(is) = F^*(s)$.
  ■ For $Exp(\lambda)$ random variable, $\phi_X(t) = \frac{\lambda}{\lambda-it}$, and $F^*(s) = \frac{\lambda}{\lambda+s}$.
- Moments:
  ▼ $E(X^n) = i^{-n}\phi_X^{(n)}(0)$
  ▼ $E(X^n) = M_X^{(n)}(0)$
  ▼ $E(X^n) = (-1)^n F^{*(n)}(0)$
- We will be contend with MGFs and LTs/LSTs for continuous random variables (instead of characteristic functions).
- MGFs too have similar properties: There is one-to-one correspondence betwen MGFs and probability distributions, and the MGF of the sum of independent random variables is the product of the MGFs of the individual random variables.

Probability Generating Functions

- The PGF is an example of a generating function of a sequence and equivalent to z-transform of the PMF.
  ♦ GFs are useful in solving *difference equations*.
- Let X be a RV with $P(X = n) = p_n, n = 0, 1, 2, \ldots$ and $\sum_{n=0}^\infty p_n = 1$. Then

$$P(z) = E[z^X] = \sum_{n=0}^\infty p_n z^n$$

  is the probability generating function (PGF) for the random variable X.
  ▶ This is similar to MGF of X, with e^t replacing z, i.e., $M_X(t) = P(e^t)$.
  ■ For $Poi(\lambda)$ random variable, $M_X(t) = e^{\lambda(e^t-1)}$, and $P(z) = e^{\lambda(z-1)}$.
- $P(1) = \lim_{z\to 1} P(z) = 1$ (limit is from below) and the series $P(z)$ converges absolutely at least for all complex numbers z with $|z| \leq 1$.
- The nth *factorial moment* of X is $E[X(X-1)(X-2)\cdots(X-n+1)] = P^{(n)}(1)$
  ▼ $E(X) = P'(1)$
  ▼ $E(X^2) = P''(1) + P'(1)$

- Example: Let $X_n, n = 1, 2, \ldots$ be independent and identically distributed discrete random variables with $p_k = P(X_n = k)$ and with PGF $P(z) = \sum_k p_k z^k$.
  Further, suppose that N is also a discrete random variable with $P(N = n) = g_n$ and with PGF $G(z) = \sum_n g_n z^n$.
  Assume that N is independent of X_n's.
  Let $S_N = X_1 + X_2 + \cdots + X_N$ and let $H(z)$ be its PGF.
  ▶ Exercise: Show that $H(z) = G(P(z))$.
  ▶ Deduce $E(S_N) = E(X_n)E(N)$.
  ▶ Deduce $Var(S_N) = E(N)Var(X_n) + Var(N)[E(X_n)]^2$.
- For practice: Consider the above with N as $Poi(\lambda)$ for a compound Poisson distribution.
- Extend the above, by considering X_n's as continuous RVs with PDF f and with LT $\bar{f}(s)$ and deduce that the LT of S_N is $G(\bar{f}(s))$.
- For practice: Let the number of patients visiting a doctor is a Poisson RV with mean λ and the time taken by the doctor on a patient is IID uniform over $[0, h]$ (in minutes). Find the mean and variance of the time taken by the doctor to complete the consultation of all the patients.

Stochastic Processes - Overview

- Queueing theory is a part of theory of applied stochastic processes.
- A stochastic process (or a random process) is a collection (or family) of random variables $\{X(t), t \in T\}$, where T is some index set, defined on a common probability space $(\Omega, \mathcal{F}, P)$.
  ▶ Here, T is the parameter space and t is the parameter.
  ▶ The common support for the random variables in a SP is called the state space and denoted by S.
  ⧫ Thus, a SP is a (measurable) mapping from $T \times \Omega$ to $\mathbb{R}$ and hence X is a function of both ω and t, i.e., $X(t, \omega)$.
- For fixed $\omega \in \Omega$, $X(t, \omega)$ is a function of t defined on T, and is called a sample path or realization of the SP. Thus, a SP is a family of these random functions.
  A sample path is a collection of time-ordered data describing what happened to a dynamic process in one instance. A stochastic process is a probability model describing a collection of time-ordered random variables that represent the possible sample paths.
- For fixed $t \in T$, $X(t, \omega)$ is a single random variable. In many SPs, T often represents time and we refer to $X(t)$ (or X_t) as the state of the process at time t.
- Examples:
  ■ Number of heads in the first n tosses in a sequence of coin tosses.
  ■ Number of people affected by COVID in India as a function of time.
  ■ Daily closing prices of a stock on NSE.
  ■ Temperature at Guwahati as a function of time.

- Recall: We know completely about a random variable, if we know the probability distribution of the random variable.
  i.e., X is completely specified by its DF F_X.
- A SP is completely specified (at least their distributional properties) if we know its finite-dimensional distributions (FDDs).
  For each $n \in \mathbb{N}$, for each $t_1, t_2, \ldots, t_n \in T$, the FDD is specified through the joint DF

$$F_{\mathbf{X}}(\mathbf{x}; \mathbf{t}) = F_{X(t_1),\ldots,X(t_n)}(x_1, \ldots, x_n; t_1, \ldots, t_n)$$
$$= P\{X(t_1) \le x_1, \ldots, X(t_n) \le x_n\}, \quad \forall x_1, \ldots, x_n \in \mathbb{R}.$$

- Question: Is this possible for all SPs?
  For many interesting SPs, it is possible to provide the above specification in simple terms.

Classification of SPs

- SPs are distinguished based on:
  - Parameter space T
  - State space S
  - Dependence relationship among the family of RVs $\{X(t)\}$
- Parameter Space (index-set):
  - If T is a countable set (or a subset of integers) representing specific time points, the SP is said to be a discrete-time SP.
    For instance, $T = \{0, 1, 2, \ldots\}$, and in such a case, the discrete-time SP $\{X(t), t \in T\}$ is usually denoted by $\{X_n, n \ge 0\}$, indexed by the nonnegative integers.
  - If T is an interval of the real line, the SP is said to be a continuous-time SP.
    For instance, $T = [0, \infty)$, and in such a case, the continuous-time SP $\{X(t), t \in T\}$ is often written as $\{X(t), t \ge 0\}$ or even simply as $\{X_t, t \ge 0\}$.
- State Space (the space in which the possible values of each $X(t)$ lie):
  - If S is a countable set (or a subset of integers), the SP is said to be a discrete-state SP.
    For instance, $S = \{0, 1, 2, \ldots\}$, and in such cases, the discrete-state SP $\{X(t), t \in T\}$ is often called as a 'chain'.
  - If S is an interval of the real line, the SP is said to be a continuous-state SP.
    For instance, $S = (-\infty, \infty)$, and in such cases, the continuous-state SP $\{X(t), t \in T\}$ is called a real-valued SP (we consider only these).
    ▶ If S is Euclidean-n space, then it is a n-vector SP.

- Examples: (recall)
  ■ Number of heads in the first n tosses in a sequence of coin tosses.
  ■ Number of people affected by COVID in India as a function of time.
  ■ Daily closing prices of a stock on NSE.
  ■ Temperature at Guwahati as a function of time.
- More examples:
  ▶ In a brand-switching model for consumer behaviour, the number of people (observed on a monthly basis) who buy a certain brand of an item (say, coffee).
  ▶ Number of people waiting at a bus stop at any time of the day.
  ▶ Size of a population at a given time.
  ▶ Waiting time of the 10th person of a day who arrive at a bus stop.
  ▶ In a time-sharing computer system or in a production system, the number of jobs waiting at any time and the time a job has to spend in the system.
- Classical (and more useful) classification of SPs is by the nature of dependency relationships among the random variables in a SP.
  - Properties of SPs can be studied in detail.
  - Helpful in stochastic modelling, with immediate application of already developed tools.

Different Classes of SPs

- Independent Processes: The simples and most trivial SP is an independent process $\{X_t\}$, where X_t's are independent.
  ▶ Actually, no dependence among the random variables.
  ▶ $F_{\mathbf{X}}(\mathbf{x};\mathbf{t})$ factors into the product of marginal DFs.
  Example: White noise process (used in many applications like signal processing, etc)
- Stationary Processes: An SP $\{X_t\}$ is said to be **(strictly) stationary** if all the joint DFs $F_{\mathbf{X}}(\mathbf{x};\mathbf{t})$ are invariant to shifts in time, i.e., for any given constant δ the following holds: $F_{\mathbf{X}}(\mathbf{x};\mathbf{t}+\delta) = F_{\mathbf{X}}(\mathbf{x};\mathbf{t})$.
  ▶ The process is in probabilistic equilibrium and that the particular times of the process being examined are of no relevance.
  An SP $\{X_t\}$ is said to be **wide sense stationary** or **covariance stationary** if it possesses finite second moments, $E(X_t) = E(X_{t+\delta})$ and $Cov(X_t, X_{t+\delta}) = E(X_t X_{t+\delta}) - E(X_t)E(X_{t+\delta})$ depends only on δ for all $t \in T$.
  Applications: Time series, Signal processing

- Processes with Stationary Independent Increments: An SP $\{X_t\}$ is called a **process with independent increments** if the increments of the process $X_{t_1} - X_{t_0}, X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}}$ are independent for any $t_0 < t_1 < \cdots < t_n$ and for all values n.
  If the distribution of the increments $X_{t_1+h} - X_{t_1}$ depends only on the length of the interval h and not on the time t_1, then the process is said to have **stationary increments**.
  Examples: Poisson process, Brownian motion
- Martingales: An SP $\{X_t\}$ is called a **martingale** if $E(|X_t|) < \infty$ for all t and $E(X_{t_{n+1}} | X_{t_0} = a_0, X_{t_1} = a_1, \ldots, X_{t_n} = a_n) = a_n$ for any $t_0 < t_1 < \cdots < t_{n+1}$ and for all values $a_0, a_1, \ldots, a_n$.
  ▶ This states that the expected value of $X_{t_{n+1}}$ given the past and present values of $X_{t_0}, X_{t_1}, \ldots, X_{t_n}$ equals the present value of X_{t_n}.
  Applications: Financial mathematics, Insurance, Queueing theory

- Markov Processes: An SP $\{X_t\}$ is called a **Markov process** if it satisfies the Markov property given by
  $P\{X_{t_{n+1}} \le x | X_{t_0} = x_0, X_{t_1} = x_1, \ldots, X_{t_n} = x_n\} = P\{X_{t_{n+1}} \le x | X_{t_n} = x_n\}$
  for any $t_0 < t_1 < \cdots < t_{n+1}$, for any n and for all values $x_0, x_1, \ldots, x_n, x$.
  ▶ A Markov process describes a simple and highly useful form of dependency among the random variables (dependency extending only to the last known state of the process).
  ▶ A Markov process with a discrete state space is referred to as a Markov chain. The discrete-time Markov chain (DTMC) is the easiest to conceptualize and understand (we will start with this later on).
  ▶ The duration of time the process stays in a state is distributed as **geometric** (in case of discrete-time) or **exponential** (in case of continuous-time).
  ▶ Markov processes are central to the study of queueing systems.
  Applications: Wide applications across the spectrum including queueing theory, statistics, economics,finance, population dynamics, physics, information theory, artificial intelligence, etc.
- Birth-Death Processes: A birth-death process (BDP) is a very important special class of Markov chains, with the defining condition that the process moves only to the neighbouring states only (whenever it makes a transition from its current state).
  ▶ Our first queueing models will be based on BDPs.

- Semi-Markov Processes: This class is a generalization of Markov processes wherein the we permit an arbitrary distribution for the duration of time the process stays in a particular state.
  At the instants of state transitions, the process behaves just like an ordinary Markov chain (at those instants we say that we have an imbedded Markov chain)
  ▶ Our non-Markovian queueing models will be based on these processes.
- Random Walks: A sequence of random variables $\{S_n, n \ge 0\}$ is referred to as a (discrete-time) **random walk** (starting at the origin) if $S_n = X_1 + X_2 + \cdots + X_n, n = 1, 2, 3, \ldots,$ where $S_0 = 0$ and X_i's are IID random variables. [If the index is from a continuum, then we have a continuous-time random walk; eg. Brownian motion].
  ▶ The emphasis is on the position of the process after n transitions.
  ▶ A popular random walk model is a simple random walk model wherein the process makes a transition only to neighbouring states.

- Renewal Processes: A renewal process $\{N_t, t \geq 0\}$ is a SP for which $N_t = \max\{n : S_n \leq t\}$, where $S_0 = 0, S_n = X_1 + X_2 + \cdots + X_n, n \geq 1$ for IID non-negative random variables X_i, for $i \geq 1$.
  ► A renewal process is related to a random walk. However, the interest is in counting transitions that take place as a function of time.
  ► In a random walk, the interpretation of S_n is that of the state of the process in random walk, and the time of nth renewal (or transition) in a renewal process.
- Gaussian Processes: A **Gaussian process** is a SP such that every finite collection of those random variables has a multivariate normal distribution.
  ► Very useful in statistical modelling and machine learning
- Wiener Process or Brownian Motion: The **Wiener process** or **Brownian motion** $\{W_t\}$ is characterized by the following properties:
  (a) $W_0 = 0$
  (b) $\{W_t\}$ has independent increments (for every $t > 0$, the future increments $W_{t+u} - W_t, u \geq 0$, are independent of the past values $W_s, s \leq t$)
  (c) $\{W_t\}$ has Gaussian increments ($W_{t+u} - W_t$ is normally distributed with mean 0 and variance u)
  (d) $\{W_t\}$ has continuous paths: W_t is continuous in t.
  ► Wide applications across the spectrum including finance, physics, statistics, etc.
- There exist many other classes of SPs and the interested ones may look for them elsewhere.

Discrete-Time Markov Chains (DTMCs)

Definition

A stochastic process $\{X_n, n \geq 0\}$ is said to be a Discrete-Time Markov Chain (DTMC) or simply a Markov Chain (MC) if

$$P\{X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0\} = P\{X_{n+1} = j | X_n = i\}$$

for all states $i_0, i_1, \ldots, i_{n-1}, i$ and j and all $n \geq 0$ (provided the conditional probabilities are defined).

► For a MC, both parameter and state spaces are discrete. The property given above is referred to as **Markov property**.
► For a MC, the conditional distribution of X_{n+1} (future) given the past $X_0, X_1, \ldots, X_{n-1}$ and present X_n depends only on the present X_n and not on the past $X_0, X_1, \ldots, X_{n-1}$.
► An MC describes a simple and highly useful form of dependency among the random variables (dependency extending only to the last known state of the process).

- The quantity $P\{X_{n+1} = j | X_n = i\} = p_{ij}(n)$ is referred to as the (one-step) transition probability and gives the conditional probability of making a transition from state i at time n to state j at time $n+1$.
- If $p_{ij}(n) = p_{ij}$ for all $n \geq 1$, then the MC is said to have stationary transition probabilities or the MC is a time-homogeneous MC (and we consider only time-homogeneous MCs).
- The matrix $P = ((p_{ij}))_{i,j \in S}$ is called (one-step) transition probability matrix (TPM). We have that $p_{ij} \geq 0$ for all i and j and $\sum_{j \in S}^{\infty} p_{ij} = 1$.
  ♦ P is a stochastic matrix.

Examples

- ► An MC whose state space is given by set of integers is said to be simple random walk (SRW) if for some $0 < p < 1$, $p_{i,i+1} = p = 1 - p_{i,i-1}$. It is simple symmetric random walk (SSRW) if $p = 1/2$.
- ► In a sequence of coin tosses, the number X_n of heads in the first n tosses.
- ► Consider a communication system that transmit the digits 0 and 1. Each digit, transmitted must pass several stages, at each stage of which there is a probability p that the digit entered will remain unchanged when it leaves.
- ► (A Gambling Model) Consider a gambler who at each play of the game either wins Re. 1 with probability p or losses Re. 1 with probability $1 - p$. Suppose that the gambler quits plays either when he goes broke or he attains a fortune of Rs. N.

- Fact: A random variable is probabilistically specified by its distribution. Likewise, a stochastic process is specified by its finite dimensional distributions (FDDs).
- An MC is specified by its initial distribution and its transition probabilities. Let $P\{X_0 = i\} = \mu_i$ for $i \in S$. Then

$$P\{X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n\} = \left(\prod_{k=0}^{n-1} p_{i_k i_{k+1}}\right) \mu_{i_0}.$$

Chapman-Kolmogorov Equations

Consider an MC having state space S and one-step transition probabilities p_{ij} for $i,\, j \in S$. Let us define

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

These are known as the n-step transition probabilities.

The Chapman-Kolmogorov (CK) equations are given by

$$p_{ij}^{(m+n)} = \sum_{k \in S} p_{ik}^{(m)} p_{kj}^{(n)}$$

for all $m, n \geq 0$ and all $i,\, j \in S$.

If we denote n-step transition probability matrix by $P^{(n)}$, then

$$P^{(n+m)} = P^{(n)} P^{(m)} \implies P^{(n)} = P^n.$$

Note: $P^{(0)} = I$.

Example

Suppose that the chance of rain tomorrow depends on previous weather conditions only through whether or not it is raining today and not on past weather condition. Suppose that if it is raining today, then it will rain tomorrow with probability 0.75. If it is not raining today, then it will rain tomorrow with probability 0.40. Calculate the probability that it will rain four days from today given that it is raining today.

$$P = \begin{bmatrix} 0.75 & 0.25 \\ 0.40 & 0.60 \end{bmatrix}, \quad P^{(4)} = P^4 = \begin{bmatrix} \mathbf{0.6212} & 0.3788 \\ 0.6061 & 0.3938 \end{bmatrix}$$

Example

Suppose that balls are successively distributed among 8 urns, with each ball being equally likely to be put in any of these urns. What is the probability that there will be exactly 3 occupied urns after 9 balls have been distributed?
If X_n is the number of nonempty urns after n balls have been distributed, then $\{X_n\}$ is a MC with states $\{0, 1, 2, \ldots, 8\}$ and with $p_{ii} = i/8 = 1 - p_{i,i+1}, i = 0, 1, 2, \ldots, 8$ and the desired probability is $p_{03}^{(9)}$ which can be computed as 0.00756 using P^9.
But, for our problem, observe that the first transition is deterministic (from 0 to 1) and hence the required probability is equal to $p_{13}^{(8)}$, we can simplify the problem by letting $Y_n = \max\{X_n + 1, 4\}, n \geq 0$ with state space $\{1, 2, 3, 4\}$ and TPM P as given below.

$$P = \begin{bmatrix} 1/8 & 7/8 & 0 & 0 \\ 0 & 2/8 & 6/8 & 0 \\ 0 & 0 & 3/8 & 5/8 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad P^{(4)} = P^4 = \begin{bmatrix} 0.0002 & 0.0256 & 0.2563 & 0.7178 \\ 0 & 0.0039 & 0.0952 & 0.9009 \\ 0 & 0 & 0.0198 & 0.9802 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Now, $p_{13}^{(8)} = \sum_{j=1}^{4} p_{1j}^{(4)} p_{j3}^{(4)} = \mathbf{0.00756}$.

State Probabilities

Consider an MC having state space S and one-step transition probabilities p_{ij} for $i,\, j \in S$. Let us define

$$\pi_j^{(n)} = P\{X_n = j\}$$

as the probability of finding the system in state j at time n (also known as state probabilities). These are known as the n-step transition probabilities.

It can be shown that

$$\pi_j^{(m)} = \sum_{i \in S} \pi_i^{(m-1)} p_{ij}$$

for all $m \geq 1$ and all $i,\, j \in S$. In matrix notation,

$$\boldsymbol{\pi}^{(m)} = \boldsymbol{\pi}^{(m-1)} \mathbf{P}.$$

This means that

$$\boldsymbol{\pi}^{(m)} = \boldsymbol{\pi}^{(m-1)} \mathbf{P} = \boldsymbol{\pi}^{(m-2)} \mathbf{P}^2 = \cdots = \boldsymbol{\pi}^{(0)} \mathbf{P}^m,$$

where $\boldsymbol{\pi}^{(0)}$ is the initial state distribution.

Properties of MCs

- Accessibility: State j is said to be accessible from state i if there exists $n \geq 0$ such that $p_{ij}^{(n)} > 0$, where $p_{ij}^{(0)} = \delta_{ij}$.
  - ▶ If j is not accessible from i, then P (Ever be in j|starting from i) $= 0$.
- Communication: Two states i and j are said to communicate if i and j are accessible from each other, *i.e.*, there exist $m \geq 0$ and $n \geq 0$ such that $p_{ij}^{(n)} > 0$ and $p_{ji}^{(m)} > 0$.

  Notation: $i \to j$: j is accessible from i.
  $i \leftrightarrow j$: i and j communicate.

  ▶ Communication is an equivalence relation (i.e., it satisfies (Reflexivity) $i \leftrightarrow i$, (Symmetry) $i \leftrightarrow j \iff j \leftrightarrow i$, and (Transitivity) $i \leftrightarrow k$ and $k \leftrightarrow j \implies i \leftrightarrow j$).
  ▶ This relation partitions the state space into equivalence classes (known as communicating classes).
- Irreducible/Reducible: An MC is said to be **irreducible** if all states communicate with each other, *i.e.*, there is a single communicating class. A chain is **reducible** otherwise.

- Closed: A subset A of the state space S is said to be **closed** if no one-step transition is possible from any state in A to any state in A^c.
- Absorbing State: If a closed set A contains only a single state, then the state is called as **absorbing** state.
  - ▶ A state i is absorbing if and only if $p_{ii} = 1$.
- If S is closed and does not contain any proper subset which is closed, then we have an *irreducible* MC. If S contains proper subsets that are closed, then the chain is *reducible*.
  - ■ If a closed subset of a reducible MC contains no closed subsets of itself, then it is referred to as an *irreducible sub-MC* (and these may be studied independently of other states).

Examples

$$P_1 = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 1/4 & 1/4 \\ 0 & 1/3 & 2/3 \end{bmatrix} \qquad P_2 = \begin{bmatrix} 1/3 & 2/3 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

- Hitting Time: For any $A \subseteq S$, the hitting time T_A is defined by

$$T_A = \inf\{n \geq 1 : X_n \in A\},$$

with the convention that $\inf \phi = \infty$.

  - ▶ T_A is the first time after 0, when the chain enters A.
  - ▶ T_A is also called **first passage/return time** to A.
  - ▶ $T_{\{i\}}$ will be denoted by T_i, $i \in S$.
- Classification of States:
  - ▸ A state i is called **recurrent** (or persistent) if $P\{T_i < \infty | X_0 = i\} = 1$.
    - ▸ State i is recurrent if and only if $f_{ii} = P\{X_n = i \text{ for some } n \geq 1 | X_0 = i\} = 1$.
  - ▸ A state i is called **transient** if $P(T_i < \infty | X_0 = i) < 1$.
  - ▸ A recurrent state i is called **null recurrent** if $E(T_i|X_0 = i) = \infty$ and **positive recurrent** (or non-null recurrent) if $E(T_i|X_0 = i) < \infty$.
- Let $f_{ii}^{(n)}$ be the probability that a chain starting in state i returns for the first time to i in n transitions. Then $f_{ii} = \sum_{n=1}^{\infty} f_{ii}^{(n)}$. For a recurrent state i, since $f_{ii} = 1$, $\{f_{ii}^{(n)}\}$ defines the first-return time or recurrence time distribution and the mean recurrence time is $M_{ii} = E(T_i|X_0 = i) = \sum_{n=1}^{\infty} n f_{ii}^{(n)}$.

  ▶ Recurrent state i is positive recurrent if $M_{ii} < \infty$ and null recurrent if $M_i = \infty$.

- Periodicity: The period of a state i is defined by the greatest common divisor of all integers $n \geq 1$ for which $p_{ii}^{(n)} > 0$, *i.e.*,

$$d(i) = \begin{cases} \text{gcd}\left\{n \geq 1 : p_{ii}^{(n)} > 0\right\} & \text{if } \left\{n \geq 1 : p_{ii}^{(n)} > 0\right\} \neq \phi \\ 0 & \text{if } \left\{n \geq 1 : p_{ii}^{(n)} > 0\right\} = \phi. \end{cases}$$

If $d(i) = 1$, then the state i is said to be aperiodic.
If $d(i) = \gamma > 1$, then the state i is said to be periodic with period γ.

Example

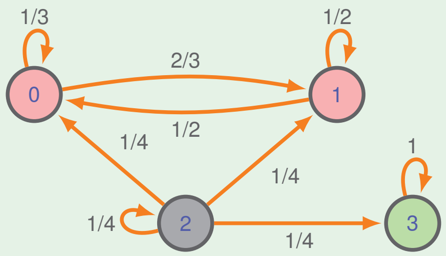
Consider an MC with $S = \{0, \pm 1, \pm 2, \ldots\}$ and with $p_{i,i+1} = a$, $p_{i,i-1} = b$, $p_{ii} = c$, where $a + b + c = 1$, $a > 0$, $b > 0$, $c \geq 0$.
Determine the period of states (different cases).

State Diagram

- A state transition diagram (or simply state diagram) for an MC is a directed graph where the nodes represent the states and the edges represent possible one-step transitions. More precisely, the state diagram contains an edge from node i to node j if and only if $p_{ij} > 0$.
  ⧫ Very useful tool and we will use extensively.

Example

Consider an MC with state space $S = \{0, 1, 2, 3\}$ and with TPM $P = \begin{bmatrix} 1/3 & 2/3 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 0 & 0 & 0 & 1 \end{bmatrix}$.



MC has three classes $\{0, 1\}$, $\{2\}$ and $\{3\}$ and hence reducible.

Some Results

1. (Number of Visits) For any state i, let N_i be the number of visits to state i. Then,
   a i recurrent implies $P\{N_i = \infty | X_0 = i\} = 1$.
   b i transient implies $P\{N_i = n | X_0 = i\} = f_{ii}^n (1 - f_{ii})$ for $n = 0, 1, 2, \ldots$, where $f_{ii} = P\{T_i < \infty | X_0 = i\}$ is the probability of returning to i starting from i. Thus $N_i | X_0 = i \sim Geo(1 - f_{ii})$.

   Corollary: A state i is recurrent iff $\sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty$ and transient iff $\sum_{n=1}^{\infty} p_{ii}^{(n)} < \infty$.

   ▶ $P\{X_n = i$ for infinitely many $n | X_0 = i\} = 1$ or 0 iff recurrent or transient.
2. If the state space S is finite, then at least one state must be recurrent.
3. Positive recurrence, null recurrence and transience are all class properties. Also, all the states in a class have the same period.
4. All states of a finite irreducible MC are positive recurrent.
5. Let i be recurrent and $i \to j$. Then $f_{ji} = P\{T_i < \infty | X_0 = j\} = 1$ and j is recurrent. [Note: Not true if i is transient.]
6. Suppose that $\{X_n\}$ is irreducible and recurrent. Then for all $i \in S$, $P_\mu\{T_i < \infty\} = 1$ for any initial distribution μ.

Long-Term Behaviour of MCs

- Main interest: What is the long-run behaviour of an MC?
  ▶ One way is to look at $\lim_{n\to\infty} P^n$.

  Example: $\boxed{\lim_{n\to\infty} P^n = \lim_{n\to\infty} \begin{bmatrix} 1/5 & 4/5 \\ 2/3 & 1/3 \end{bmatrix}^n = \begin{bmatrix} 5/11 & 6/11 \\ 5/11 & 6/11 \end{bmatrix}}$.
- Long-term behaviour is related to three concepts:
  - Limiting distributions
  - Stationary distributins
  - Ergodicity

Limiting Distribution

- A vector $\{\pi_i\}_{i\in S}$ is called the limiting distribution for a MC with transition probability matrix $P = ((p_{ij}))$ if $\pi_i = \lim_{n\to\infty} p_{ji}^{(n)}$, $i, j \in S$ (provided the limits exist) and $\sum_{i\in S} \pi_i = 1$.
  ▶ Possible that limiting probabilities exist but not the limiting distribution.
- By the above, when a limiting distribution exists, it does not depend on the initial state, so we can write $\pi_i = \lim_{n\to\infty} \pi_i^{(n)}$, $i \in S$.
- π_i is the probability of being in state i a long time from now.
- Question is: When does an MC have a limiting distribution (that does not depend on the initial PMF)?
  And, how to determine it?

Stationary Distribution

- A vector $\{\pi_i\}_{i\in S}$ is called a stationary distribution (or invariant distribution) for a MC with transition probability matrix $P = ((p_{ij}))$ if $\pi_i \geq 0$ for all $i \in S$, $\sum_{i\in S} \pi_i = 1$ and $\sum_{j\in S} \pi_j p_{ji} = \pi_i$ for all $i \in S$.
  ▶ $\sum_{j\in S} \pi_j p_{ji} = \pi_i$ for all $i \in S \implies \boldsymbol{\pi} P = \boldsymbol{\pi}$. Thus a stationary distribution is a left eigen vector corrosponding to eigen value 1 and $\boldsymbol{\pi}\, \boldsymbol{e} = 1$.
  ▶ $P(X_n = i) = \pi_i$ if $P(X_0 = i) = \pi_i$ for all $i \in S$ and MC is stationary.
  ▶ If S is finite, then stationary distribution exists.
- In general, a stationary distribution may not exist. Even if it exist, it may not be unique.
  ▶ SRW does not admit a stationary distribution.
- Why do we care if our Markov chain is stationary?
  ▶ If it were stationary and we knew what the distribution of each X_n was, then we would know a lot because we would know the long run proportion of time that the MC was in any state. Hence, solving for $\boldsymbol{\pi}$ is an important part of MC analysis.

Theorem

For an irreducible MC, a stationary distribution π exists if and only if all states are positive recurrent. In this case, the stationary distribution is unique and $\pi_i = 1/M_{ii}$, where M_{ii} is the mean recurrence time to state i.
Furthermore, if the chain is aperiodic, then the limiting probability distribution exists and is equal to the stationary distribution.

▶ We can't make a transient or null recurrent MC stationary [i.e., if there is no stationary distribution, then the MC is either transient or null recurrent, and $\pi_i = 0$ for all $i \in S$].
▶ If MC is reducible, the stationary distribution may not be unique.
▶ No conditions on the period of the MC for the existence and uniqueness of the stationary distribution (not true with limiting probabilities).
▶ The limiting distribution of an MC is also a stationary distribution.
▶ Existence of a stationary distribution does not imply the existence of a limiting distribution.
▶ It can shown that if the limiting distribution exists, then it is the only stationary distribution.

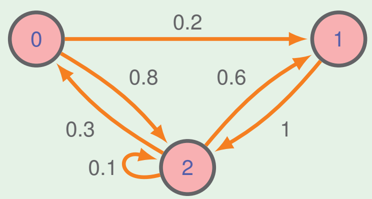
Example

Three out of every four trucks on the road are followed by a car, while only one out of every five cars is followed by a truck. If you see a truck pass by on the road, on average how many vehicles pass before you see another truck?

Let $\{X_n\}$ be a MC with $S = \{0, 1\}$ (0-truck, 1-car) and with $P = \begin{bmatrix} 1/4 & 3/4 \\ 1/5 & 4/5 \end{bmatrix}$. The unique stationary distribution is $\pi_0 = 4/19$ and $\pi_1 = 15/19$.
If you see a truck pass by then the average number of vehicles that pass by before you see another truck corresponds to the mean recurrence time to state 0, given that you are currently in state 0. By the above, the mean recurrence time to state 0 is $M_{00} = 1/\pi_0 = 19/4$, which is roughly 5 vehicles.

Example

$$P = \begin{bmatrix} 0 & 0.2 & 0.8 \\ 0 & 0 & 1 \\ 0.3 & 0.6 & 0.1 \end{bmatrix}$$

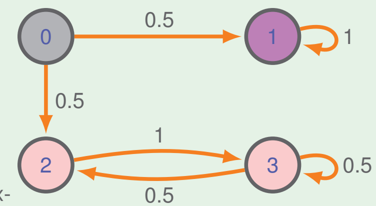


Irreducible, positive recurrent, aperiodic.
$\boldsymbol{\pi} P = \boldsymbol{\pi}, \boldsymbol{\pi}\, \boldsymbol{e} = 1$
$\Rightarrow$ unique $\boldsymbol{\pi} = (0.153, 0.337, 0.510)$.
Limiting distribution equals $\boldsymbol{\pi}$.

Example

$$P = \begin{bmatrix} 0 & 0.5 & 0.5 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0.5 & 0.5 \end{bmatrix}$$

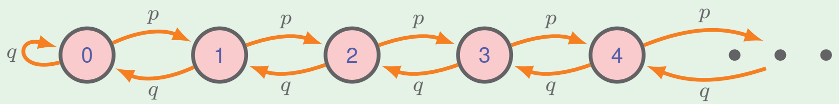
Reducible MC
3 communication classes $\{0\}, \{1\}$ and $\{2, 3\}$
State 0 is transient, states $1, 2, 3$ are positive recurrent
$\boldsymbol{\pi} P = \boldsymbol{\pi}, \boldsymbol{\pi}\, \boldsymbol{e} = 1$
$\Rightarrow$ Solution exists but not unique
$\boldsymbol{\pi} = \alpha(0, 1, 0, 0) + (1-\alpha)(0, 0, 1/3, 2/3), 0 \leq \alpha \leq 1$



No limiting distribution; limiting probabilities $\lim_{n\to\infty} p_{ij}^{(n)}$ exists, but depends on the starting state as observed from

$$\lim_{n\to\infty} P^n = \begin{bmatrix} 0 & 1/2 & 1/6 & 1/3 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1/3 & 2/3 \\ 0 & 0 & 1/3 & 2/3 \end{bmatrix}$$

Example



With $0 < p < 1$ and $q = 1 - p$, MC is irreducible and aperiodic.
If MC is positive recurrent, then a unique stationary distribution exists and equals the limiting distribution.
If MC is null recurrent or transient, then there is no stationary distribution (and hence no limiting distribution).
Try solving $\boldsymbol{\pi} P = \boldsymbol{\pi}, \boldsymbol{\pi}\, \mathbf{e} = 1$
First part gives $\pi_n = \left(\frac{p}{q}\right)^n \pi_0, n \geq 1$ and the normalization condition implies that

$$1 = \sum_{n=0}^{\infty} \pi_n = \pi_0 \sum_{n=0}^{\infty} \left(\frac{p}{q}\right)^n.$$

If $p < q$, then the geometric converges and $\pi_0 > 0$ (and hence $\pi_n > 0$) and we have a solution to the stationary equations.
▶ MC is positive recurrent if $p < q$.
If $p \geq q$, then the geometric does not converge and $\pi_0 = 0$ (and hence $\pi_n = 0$) and there is no stationary distribution.
▶ MC is not positive recurrent if $p \geq q$.

Example

$P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$
MC is irreducible and positive recurrent.
Has a unique stationary distribution $\pi_0 = \pi_1 = 1/2$ (Ths system spends half of the time in each state).
What about the limiting distribution?

P^n does not converge as $P^n = \begin{cases} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, n \text{ even}, \\ \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, n \text{ odd}. \end{cases}$

No limiting distribution [Note: MC has period 2].

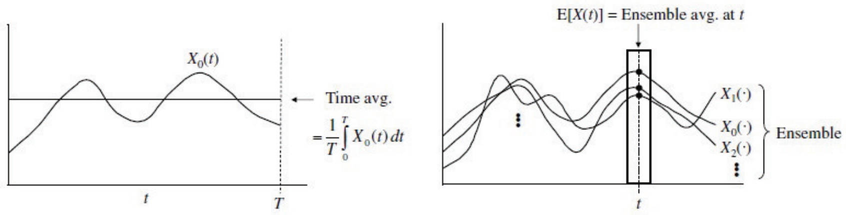
But, if we choose the initial distribution as $\boldsymbol{\pi}^{(0)} = (1/2, 1/2)$, then $\boldsymbol{\pi}^{(n)} = (1/2, 1/2)$ for all n.
i.e., Even though $\lim_{n\to\infty} P^n$ does not exist, it is possible for $\lim_{n\to\infty} \boldsymbol{\pi}^{(n)}$ to exist, but only if the starting state is chosen randomly according to the stationary distribution.

- How to determine the presence of recurrence in a MC?
- Many results are available. One sufficient condition is given below.
  Result: An irreducible, aperiodic chain is positive recurrent if there exists a nonnegative solution of the system

$$\sum_{j=0}^{\infty} p_{ij} x_j \leq x_i - 1 \;\; (i \neq 0) \quad \text{such that} \quad \sum_{j=0}^{\infty} p_{0j} x_j < \infty.$$

Ergodicity

- A closely related idea is **ergodicity**.
  ▶ Deals with the problem of determining measures of a stochastic process $\{X(t)\}$ from a single realization.
  ▶ $\{X(t)\}$ is ergodic if time averages equal ensemble averages.
  ▶ A time average is obtained from one sample realization of the process. Over an infinitely long time horizon, the time average is $\bar{x} = \lim_{T\to\infty} \frac{1}{T} \int_0^T X_0(t)dt$.
  ▶ An ensemble average is obtained from multiple realizations of the process at a fixed point in time t. With an infinite number of realizations, the ensemble average is
  $m(t) = E(X(t)) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} X_i(t)dt.$
  ▶ Our interest in ergodicity involves the convergence of both time and ensemble averages.



- We say that a process is **ergodic** (with respect to its first moment) if $\bar{x} = \lim_{t\to\infty} m(t) < \infty$. That is, the ensemble average $m(t)$ converges to a limit as $t \to \infty$ and this limiting value equals the time average.
- We can similarly define ergodicity with respect to a higher moment n if $\lim_{T\to\infty} \frac{1}{T}\int_0^T [X_0(t)]^n dt = \lim_{t\to\infty} E[X(t)]^n < \infty.$ We say that a process is fully ergodic (ergodic in distribution function) if this property holds for all moments.
  - ▶ In queueing theory, we are typically interested in fully ergodic processes.

Linkage between the three concepts

Table : Long-run behavior concepts for DTMC

Condition	Resulting Properties	Comment
Irreducible, positive recurrent	Unique stationary distribution	π_i is the long-run fraction of time in state i
Irreducible, positive recurrent, $\pi^{(0)} = \pi$	Unique stationary distribution, process is stationary and ergodic	Ensemble averages = Time avarages
Irreducible, positive recurrent, aperiodic	Unique stationary distribution, process is ergodic, limiting distribution exists (equal to stationary distribution)	Process independent of starting state in the limit

- The existence of a limiting distribution is the strongest condition, ergodicity is somewhat weaker, and a unique solution to the stationary equations is the weakest of the three conditions.
- Some authors, when dealing with Markov chains, use a slightly different definition of ergodicity, requiring a state to be aperiodic in order to be ergodic.
  - ▶ Our definition is slightly less restrictive (According to this definition, it is possible for a chain to be ergodic without being aperiodic).

Balance Equations

- The stationary probability π_i, if it exists, gives the long-run proportion of time in state i.
- The product $\pi_i p_{ij}$ is the long-run proportion of transitions that go from state i to state j.
- If we think of a transition from state i to state j as a unit of *flow* from state i to state j, then $\pi_i p_{ij}$ would be the *rate of flow* from state i to state j. Similarly, with this flow interpretation, we have
  $\pi_j =$ "rate of flow out of state j", and
  $\sum_{i\in S} \pi_i p_{ij} =$ "rate of flow into state j".
- The equations $\boldsymbol{\pi} = \boldsymbol{\pi} P$ have the interpretation
  "rate of flow into state j" $=$ "rate of flow out of state j ", for every $j \in S$.
  That is, the stationary distribution is that vector $\boldsymbol{\pi}$ which achieves balance of flow.
- The equations $\boldsymbol{\pi} = \boldsymbol{\pi} P$ are called Balance Equations or Global Balance Equations.
  - ▶ All stationary distributions must create global balance.

- If the stationary probabilities $\boldsymbol{\pi}$ also satisfy $\pi_i p_{ij} = \pi_j p_{ji}$,for every $i, j \in S$, then we say that $\boldsymbol{\pi}$ also creates *local balance*. These equations are called the Local Balance Equations (or *Detailed Balance Equations*), because they specify balance of flow between every pair of states:
  "rate of flow from i to j" $=$ "rate of flow from j to i", for every $i, j \in S$.
- If one can find a vector $\boldsymbol{\pi}$ that satisfies local balance, then $\boldsymbol{\pi}$ also satisfies the global balance equations.
- The local balance equations are typically much simpler to solve than global balance equations.
- We will mostly work with global balance equations.

Reversibility

- Local balance is connected with a property of Markov chains (and stochastic processes in general) called **reversibility**, or **time-reversibility**. Just as not all Markov chains satisfy local balance, not all Markov chains are reversible.
- Local balance and reversibility (and global balance as well) are properties of only stationary Markov chains.
- Start out with a stationary Markov chain and then extend the time index back to $-\infty$, so that now our Markov chain is $\{X_n : n \in \{\dots, -2, -1, 0, 1, 2, \dots\}\}$. Imagine running the chain backwards in time to obtain a new process $\{Y_n = X_{-n} : n \in \{\dots, -1, 0, 1, \dots\}\}$.This process is called the reversed chain.
  ▶ It is also an MC. It is also stationary and has the same stationary distribution as the original chain.
  ▶ However, the reversed chain does not in general have the same transition matrix as the original chain. If q_{ij} is the transition probability in the reversed chain, then it can be shown that $q_{ij} = \frac{p_{ji}\pi_j}{\pi_i}$.
- We say that a stationary MC is reversible, or time-reversible, if the transition matrix of the reversed chain is the same as the transition matrix of the original chain.
  ▶ That is, an MC is reversible if and only if $q_{ij} = p_{ij}$, i.e., iff $\pi_i p_{ij} = \pi_j p_{ji}$, i.e., iff local balance is satisfied in equilibrium.

Memoryless property

- For an MC with $p_{ii} = 0$, the time that the chain spends in state i is equal to 1.
- For an MC with $p_{ii} > 0$, the number of time units that the system spends in the state i (also known as *sojourn time* or *waiting time* or *residence time*) is geometrically distributed.
  - Assume that the MC has just entered the state i. It will remain in i at the next step with probability p_{ii} and it will leave this state at the next step with probability $1 - p_{ii}$.
  - Independent of what happened in one step, similar property holds in the next step as well.
  - Let $\tau_i = \min\{n \geq 1 : X_n \neq i\}$. Then, the distribution of the sojourn time in state i is $P_i\{\tau_i = n\} = (1 - p_{ii})p_{ii}^{n-1}, \quad n \geq 1.$

Birth-Death Chain

- A birth-death chain is a special of a DTMC with $S = \{0, 1, 2, \dots\}$ and with TPM

$$P = \begin{bmatrix} 1-b_0 & b_0 & 0 & 0 & \cdots \\ d_1 & 1-b_1-d_1 & b_1 & 0 & \cdots \\ 0 & d_2 & 1-b_2-d_2 & b_2 & \cdots \\ \vdots & & & & \ddots \end{bmatrix}$$

where $b_i > 0, i \geq 0$ is the probability that a single birth will occur at the next time step, $d_i > 0, i \geq 1$ is the probability that at the next time step a single death will occur, and $1 - b_0, 1 - b_i - d_i, i \geq 1$ is the probability that the state will not change at the next time step.
▶ Transitions to nearest neighbour states only (multiple births and/or deaths not possible).
▶ Very useful model for queues and obtaining of solutions is easier because of the special structure (tridiagonal) of the TPM.
▶ We will see more details later.