CS 3.307

Performance Modeling for Computer Systems

Tejas Bodas

Assistant Professor, IIIT Hyderabad

Logistics

- Feel free to contact me anytime at tejas.bodas@iiit.ac.in.
- Office @ A5304.
- Book– Performance modeling and design of computer systems (Cambridge press) by Mor Harchol-Balter (Professor, CMU)
- Other books: 1) Stochastic processes by Sheldon Ross 2) Probabilistic modeling by Isi Mitrani.
- Assignment 1 : 15%. Midsem exam: 30%. Assignment 2: 15% Endsem 40 %.

Course Outline

- Module 1 (2 lectures)
  - Motivation, Probability refresher, Introduction to Stochastic Processes
- Module 2 (4 lectures)
  Poisson Process & Markov Chains
- Module 3 (2 lectures) Elementary Queues
- Module 4 Renewal theorems and Busy period analysis (3 lectures)
- Module 5 (3 lectures) Advanced Queues

Performance modeling for Computer systems

- How do you measure the performance of your computer?
- Speed with which it runs programs. RAM, clock speed, GPU, Cores.
- Storage space ? SSD or not ?
- What is the key word here ? LATENCY!
- Performance metrics?
  - response time (run time, lag, delay, jitter)
  - blocking probability (screen freeze, no disk space, packet loss, buffer full)

Modeling ?

- Design for performance: How many cores or GPU's? which core to use? how to schedule instructions in a core?
- Routing (which core) and scheduling (which program/ instruction to execute)
- How do you know which is a good design? via experimentation?(costly!)
- Performance analysis! via stochastic modeling

Applications Beyond Computers

- Computer systems
  - server farms, cloud computing, distributed storage systems
  - Communication systems, Wifi, Sensor networks.
- Heathcare
  - How many OT? How many Specialists or nurses?
  - Scheduling operations, stocking of medicines, scheduling tests.
- Hospitality industry
  - Designing hotel lobbies for faster checkin
  - Restaurant seating! (How many tables of size 2,4,8?)
- Transportation systems
  - Airline or Railway scheduling
  - Priority scheduling, class differentiation
- Operation Research!
- Henceforth use the term Queueing system!

A single server queue

- One server, one FIFO queue for jobs to wait.
- μ denotes service rate, λ denotes the arrival rate.
- Service requirements S_n and inter-arrival times A_n are typically assumed to be i.i.d.
- In its simplest form, we will assume $S_n \sim Exp(\mu)$ and $A_n \sim Exp(\lambda)$.
- Jobs face queueing delay due to waiting for other jobs.
- This is the most basic $M/M/1$ queue. Modeling this as a Markov chain and solving its stationary distribution gives us mean response time (mean of service time + waiting time).
- $E[T] = \frac{1}{\mu - \lambda}$.

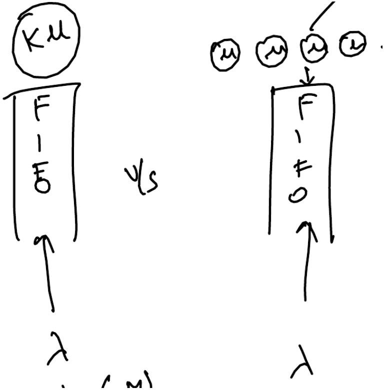
A single server queue

- $E[T] = \frac{1}{\mu - \lambda}$.
- Let N is the number of jobs in the system (Queue + server). Then what is $E[N]$?
- We will see Little's law that says that $E[N] = \lambda E[T]$.
- Mean number of jobs $E[N] = \frac{\lambda}{\mu - \lambda}$.
- This course is about Markov chain analysis to derive such formulas.

Example 1: Doubling the arrival rate

- $E[T] = \frac{1}{\mu - \lambda}$.
- What would happen to $E[T]$ if $\lambda \to 2\lambda$?
- It could blow up if $\mu < 2\lambda$.
- If you want to maintain the same level of response time then do you need to double μ?
- This course is about making such design choices!

Example 2: A fast server versus many slow servers

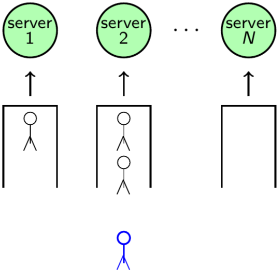


- Which system will have lower $E[T]$?
- Is a fast server $(K\mu)$ better that K normal servers (μ)?
- Does job variability impact this decision? Suppose job sizes were XS, S, M, L, XL.
- In the first model, an S, or M job has to possibly wait behind XL. This is avoided in the second scenario.

Example 3: Central queue or individual

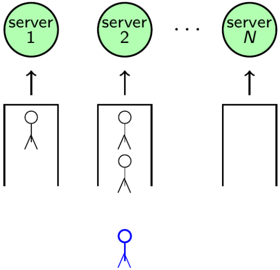
- At Airport immigration, Hotel check-ins you often see central queues.
- But at movie theatres, metro/train ticket counters, you see the second model.
- Which setting has a lower $E[T]$?
- This course will help you answer such performance modeling questions.

Example 4: Supermarket queue and load balancing



- Load balancing concerns the questions which queue to join/assign?
- Popular policy is Join shortest Queue (JSQ).
- What should be ideally done is Join smallest work (JSW).
- N is typically large and the overhead in obtaining queue length information is huge ($2N$).

Example 4: Supermarket queue and load balancing



- In that case, sample d servers randomly and join appropriate queue using $JSQ(d)$ or $JSW(d)$.
- Problem with JSW or $JSW(d)$ is that the workload information is typically unknown. How to implement it then?
- How about replicating jobs on d servers and cancelling copies when one copy starts service ?
- This is redundancy-d with cancel on start.
- We do this at super-markets all the time!

Probability Refresher

Random experiments and Sample space

- Random experiment : Experiment involving randomness
  - Coin toss
  - Roll a dice
  - Pick a number at random from $[0, 1]$.
- Sample space Ω: set of all possible outcomes of the random experiment. It could be a finite or infinite set.
  - $\Omega_c = \{H, T\}$
  - $\Omega_d = \{1, 2, \ldots, 6\}$
  - $\Omega_u = [0, 1]$

Events

- A subset $A \subseteq \Omega$ is called an **event**.
- Examples of events
  - Events in the coin experiment: $C_1 = \{T\}$.
  - Events in the dice experiment: $D_1 = 6, D_2 = \{1, 3, 5\}$
  - Events in $U[0, 1]$ experiment: $U_1 = \{0.5\}$, $U_2 = [.25, .75]$.
- Probability of event A is denoted by $\mathbb{P}(A)$.

Probability theory

$\{$Random experiment, Sample space, Events$\}$ are the key ingredients in probability theory.

In probability theory, we are interested in **measuring** the probability of subsets of Ω (events).

Probability measure $\mathbb{P}$ is a **set function**, i.e. it acts on sets and measures the probability of such sets.

sigma-algebra as domain for $\mathbb{P}$

- Event space or *sigma-algebra* $\mathcal{F}$ is a collection of measurable sets that satisfy
  $\bullet\ \emptyset \in \mathcal{F} \quad \bullet\ A \in \mathcal{F} \implies A^c \in \mathcal{F}$
  $\bullet A_1, A_2, \ldots A_n, \ldots \in \Omega \implies \cup_{n=1}^{\infty} A_n \in \Omega$

- The $\sigma-$algebra is said to be closed under formation of compliments and countable unions.

- Is it also closed under the formation of countable intersections?

When Ω is countable and finite, we will consider power-set $\mathcal{P}(\Omega)$ as the domain.

Formal definition of Probability measure $\mathbb{P}$

Definition

A probability measure $\mathbb{P}$ on the *measurable space* $(\Omega, \mathcal{F})$ is a function $\mathbb{P} : \mathcal{F} \to [0, 1]$ s.t.

1. $\mathbb{P}(\emptyset) = 0, \quad \mathbb{P}(\Omega) = 1$
2. For a disjoint collection of event sets $A_1, A_2, \ldots$ from $\mathcal{F}$ we have
$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$$
(countable additivity)

- The trio $(\Omega, \mathcal{F}, \mathbb{P})$ is called as a probability space.

Conditional probability

- Given/If dice rolls odd, what is the probability that the outcome is 1?
- Given/If $\bar{\omega} \in [0, 0.5]$ what is the probability that $\bar{\omega} \in [0, 0.25]$?
- The conditional probability of event B given event A is defined as $\mathbb{P}(B/A) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}$ when $\mathbb{P}(A) > 0$.
- Bayes rule: $P(B/A) = \frac{P(A/B)P(B)}{P(A)}$.

Independence and Mutually exclusive

- Two events A, B are independent iff $P(A/B) = P(A)$ and $P(B/A) = P(B)$.
- Two events A, B are independent iff $P(A \cap B) = P(A)P(B)$.

- If A and B are independent, then so are A^c and B^c.
- What about A and B^c? Are they independent?
- Two events A and B are mutually exclusive if occurrence of one implies that the other event cannot occur. Are they independent?
- If A and B are mutually exclusive, then they are not independent (and vice versa).

Random variable

- Given a random experiment with associated $(\Omega, \mathcal{F}, \mathbb{P})$, it is sometimes difficult to deal directly with $\omega \in \Omega$. eg. rolling a dice ten times.
- Notice that each sample point $\omega \in \Omega$ is not a number but a sequence of numbers.
- Also, we may be interested in functions of these sample points rather than samples themselves. eg: Number of times 6 appears in the 10 rolls.
- In either case, it is often convenient to work in a new *simpler* probability space rather than the original space.
- Random variable is a device which precisely helps us make this mapping from $(\Omega, \mathcal{F}, \mathbb{P})$ to a 'simpler' $(\Omega', \mathcal{F}', P_X)$.
- P_X is called as an induced probability measure on Ω'.

Random variable

- If Ω' is countable, then the random variable is called a discrete random variable.
- In this case it is convenient to use $\mathcal{F}'$ as power-set.
- If $\Omega' \subseteq \mathbb{R}$ or uncountable, then the random variable is a continuous random variable.
- In this case, $\mathcal{F}' = \mathcal{B}(\mathbb{R})$.
- Notation: Random variables denoted by capital letters like X, Y, Z etc. anf their realizations by small letters x, y, z..

PMF and CDF of a Discrete r.v.

- Let $X : \Omega \to \Omega'$ be a discrete r.v.
- Let $p_X(x)$ for $x \in \Omega'$ denote the probability that X takes the value x.
- $p_X(x)$ is called as a probability mass function.
- The cumulative distribution function (CDF) $F_X(\cdot)$ is defined as $F_X(x_1) := \sum_{x \leq x_1} p_X(x) = \mathbb{P}\{\omega \in \Omega : X(\omega) \leq x_1\}$.

Expectation, Moments, Variance

- The mean or expectation of a random variable X is denoted by $E[X]$ and is given by $E[X] = \sum_{x \in \Omega'} x p_X(x)$.
- The n^{th} moment of a random variable X is denoted by $E[X^n]$ and is given by $E[X^n] = \sum_{x \in \Omega'} x^n p_X(x)$.
- Functions of random variables are random variables.
- For a function $g(\cdot)$ of a random variable X, its expectation is given by $E[g(X)] := \sum_{x \in \Omega'} g(x) p_X(x)$
- $Var(X) := E[(X - E[X])^2]$
- HW: Prove that $E[(X - E[X])^2] = E[X^2] - E[X]^2$
- For $Y = aX + b$, what is $E[Y]$? $E[Y] = aE[X] + b$. (Linearity of expectation)

Bernoulli random variable

- Bernoulli random variable $X = \begin{cases} 1, & \text{with probability } p \\ 0, & \text{otherwise.} \end{cases}$
- Basic models of Multi-arm bandit problem assume Bernoulli Bandits.
- $E[X] = p, E[X^n] = p.$

Binomial $B(n, p)$ random variable.

- Consider a biased coin (head with probability p) and toss it n times.
- Denote head by 1 and tail by 0.
- Let random variable N denote the number of heads in n tosses.
- PMF of N?. $P_N(k) = \binom{n}{k} p^k (1-p)^{n-k}$.
- HW: What is $E[N], E[N^2], Var(X)$?

Geometric random variable

- Consider a biased coin (head with probability p) and suppose you keep tossing it till head appears the first time.
- Let random variable N denote the number of tosses needed for head to appear first time.
- What is the PMF of N? $p_N(k) = (1-p)^{k-1}p$.
- HW: What is $E[N], E[N^2], Var(N)$?

Poisson random variable

- A Poisson random variable X comes with a parameter λ and has $\Omega' = \mathbb{Z}_{\geq 0}$

- PMF: $p_X(k) = e^{-\lambda}\frac{\lambda^k}{k!}$

- Intuitively its a limiting case of the Binomial distribution with n increasing and p decreasing such that np converges to λ.

- Mean of binomial is np so p should decrease while n increases.

Continuous random variables

- A random variable X is continuous if there exists a non-negative real valued probability density function (PDF) $f_X(\cdot)$ such that $F_X(x) = \int_{u=-\infty}^{x} f_X(u)du$.

- $P_X(a \le X \le b) = \int_a^b f_X(u)du$. (Area under the curve)

$$\frac{dF_X(x)}{dx} = f_X(x) \text{ or } P_X(x < X \le x + h) \simeq f_X(x)h.$$

Mean, Variance, Moments

- $E[X] = \int_{-\infty}^{\infty} u f_X(u) du$
- $E[X^n] = \int_{-\infty}^{\infty} u^n f_X(u) du$
- $E[g(X)] = \int_{-\infty}^{\infty} g(u) f_X(u) du$
- $Var[X] = E[g(X)]$ where $g(x) = (x - E[X])^2$.
- For $Y = aX + b$, $E[Y] = aE[X] + b$.

Exponential random variable ($Exp(\lambda)$)

- This is a non-negative r.v. with parameter λ.
- Its pdf $f_X(x) = \lambda e^{-\lambda x}$ for $x \geq 0$.
- Its CDF is given by $F_X(x) = 1 - e^{-\lambda x}$ for $x \geq 0$.
- $E[X] = \frac{1}{\lambda}$ and $Var(X) = \frac{1}{\lambda^2}$
- $E[X^n] = \frac{n!}{\lambda^n}$

Summary: Multiple random variables

$p_{XY}(x, y) := \mathbb{P}\{\omega \in \Omega : X(w) = x \text{ and } Y(\omega) = y\}$.
$F_{XY}(x, y) := \mathbb{P}\{\omega \in \Omega : X(w) \leq x \text{ and } Y(\omega) \leq y\}$.

The marginal PMF's p_X and p_Y can be obtained from the joint PMF as follows:
$p_X(x) = \sum_y p_{XY}(x, y)$ and $p_Y(y) = \sum_x p_{XY}(x, y)$.

Two random variables, X and Y are independent if the following is true:
$p_{XY}(x, y) = p_X(x)p_Y(y), F_{XY}(x, y) = F_X(x)F_Y(y)$ and $E[XY] = E[X]E[Y]$.

$E[g(X, Y)] = \sum_{xy} g(xy)p_{XY}(xy)$

The rules for continuous random variables are similar.
Also revise conditioning of variables.

Sums of independent random variable

- Consider $Z = X + Y$. What is the pdf of Z when X and Y?
- What is $p_Z(z)$ or $f_Z(z)$?
- $p_Z(z) = \sum_{(x,y):x+y=z} p_{X,Y}(x,y)$
- $f_Z(z) = \int_{(x,y):x+y=z} f_{X,Y}(x,y)dxdy$.
- Since X and Y are independent $p_{X,Y}(x,y) = p_X(x)p_Y(y)$ and $f_{X,Y}(x,y) = f_X(x)f_Y(y)$. This gives us

Convolution formula

$$p_Z(z) = \sum_x p_X(x)p_Y(z-x)$$
$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x)f_Y(z-x)dx$$

HW: What if X and Y are not independent?

MGF of Sums of independent random variable

- Consider $Z = X + Y$. What is the pdf of Z when X and Y?
- Let $M_X(t)$ and $M_Y(t)$ be their MGF's. What is $M_Z(t)$?
- $M_Z(t) = E[e^{Zt}] = E[e^{(X+Y)t}]$.
- $M_Z(t) = E[e^{Xt}.e^{Yt}]$.
- If X and Y are independent, $E[XY] = E[X]E[Y]$ and $E[g(X)h(Y)] = E[g(X)]E[h(Y)]$.
- $M_Z(t) = E[e^{Xt}].E[e^{Yt}]$.

$$\boxed{M_Z(t) = M_X(t)M_Y(t).}$$

MGF of Sums of independent random variable

- Consider $Z = X + Y$. What is the MGF of Z when X and Y?

$$M_Z(t) = M_X(t)M_Y(t).$$

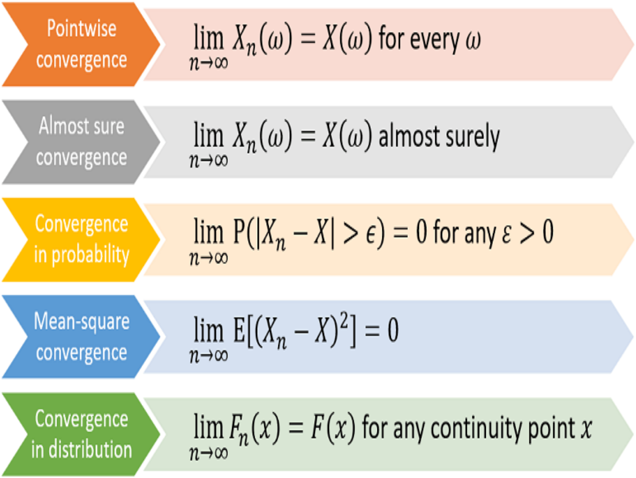
- What about $M_Z(t)$ when $Z = X_1 + X_2 + \ldots X_n$ and X_i are iid.?

- $M_Z(t) = (M_X(t))^n$.

- What about $M_Z(t)$ when $Z = X_1 + X_2 + \ldots X_N$ where N is a positive discrete random variable? section 4.5

Convergence of Random Variables

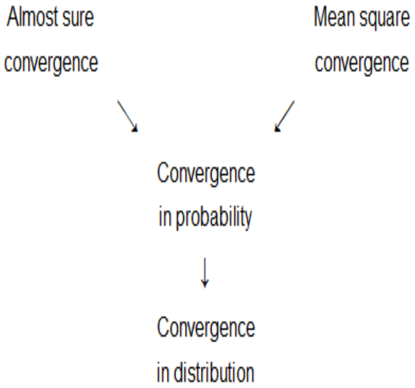
Summary



[1]

[1] Image from probabilitycourse.com

Relation between modes of convergence (no proofs)



https://en.wikipedia.org/wiki/Proofs_of_convergence_of_random_variables

Introduction to Stochastic processes

Introduction to Stochastic processes

- Stochastic process $\{X(t), t \in T\}$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a collection of random variables defined such that for every $t \in T$ we have $X(t) : \Omega \to \mathcal{S}$.
- T is the parameter space (often resembles time) and $\mathcal{S}$ is the state space.
- Random variable $X(t)$ is often denoted by $X(\omega, t)$.
- When t is fixed and ω is the only variable, we have a random variable $X(\cdot, t)$. When ω is fixed and t is the variable, we have a $X(\omega, \cdot)$ as a function of time. This is also called as a realization or sampe path of a stochastic process.

Introduction to Stochastic processes

- When T is countable, we have a discrete time process.
- If T is a subset of real line, we have a continuous time process.
- State space could be integers or real numbers
- State space could be $\mathbb{R}^n$ or $\mathbb{Z}^n$ valued

Elementary Examples

- The process of rolling a dice 6 times.
- You bank balance over a week.
- Temperature fluctuations in a 1hr window.
- Number of customers in IKEA every day.

Introduction to Stochastic processes

A c.t.s.p. is called an *independent increment process* if for any choice of parameters $t_0 < t_1 < \ldots < t_n$, the n increment random variables $X(t_1) - X(t_0), X(t_2) - X(t_1), \ldots, X(t_n) - X(t_{n-1})$ are independent.

The c.t.m.p. is said to have *stationary and independent increments* if in addition $X(t_2 + s) - X(t_1 + s)$ has the same distribution as $X(t_2) - X(t_1)$ for all $t_1, t_2 \in T$ and any $s > 0$.

Examples

- Sequence of i.i.d random variables.

- General random walk: If $X_1, X_2, \ldots$ is a sequence i.i.d of random variables, then $S_n = \sum_{i=1}^{n} X_i$ is a random walk.

- Weiner process: $\{X(t), t \geq 0\}$ is a Weiner process if
  1. $X(0) = 0$
  2. $\{X(t), t \geq 0\}$ has stationary and independent increments
  3. for every $t > 0$, $X(t)$ is normally distributed with mean 0 and variance t.

- $\{X(t), t \geq 0\}$ is a Markov process if for $t_1 < t_2 < \ldots t_n < t$ we have
  $P(X(t) \leq x | X(t_1) = x_1, \ldots, X(t_n) = x_n) = P(X(t) \leq x | X(t_n) = x_n)$

- Random walk and Weiner process are examples of Markov processes.

CS 3.307: Intro to Stochastic Processes

Tejas Bodas

Assistant Professor, IIIT Hyderabad

Introduction to Stochastic processes

- Stochastic process $\{X(t), t \in T\}$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a collection of random variables defined such that for every $t \in T$ we have $X(t) : \Omega \to \mathcal{S}$.

- T is the parameter space (often resembles time) and $\mathcal{S}$ is the state space.

- Random variable $X(t)$ is often denoted by $X(\omega, t)$.

- When t is fixed and ω is the only variable, we have a random variable $X(\cdot, t)$. When ω is fixed and t is the variable, we have a $X(\omega, \cdot)$ as a function of time. This is also called as a realization or sampe path of a stochastic process.

Introduction to Stochastic processes

- When T is countable, we have a discrete time process.
- If T is a subset of real line, we have a continuous time process.
- State space could be integers or real numbers
- State space could be $\mathbb{R}^n$ or $\mathbb{Z}^n$ valued

Elementary Examples

- The process of rolling a dice 6 times.
- You bank balance over a week.
- Temperature fluctuations in a 1hr window.
- Number of customers in IKEA every day.

Introduction to Stochastic processes

A c.t.s.p. is called an *independent increment process* if for any choice of parameters $t_0 < t_1 < \ldots < t_n$, the n increment random variables $X(t_1) - X(t_0), X(t_2) - X(t_1), \ldots, X(t_n) - X(t_{n-1})$ are independent.

The c.t.m.p. is said to have *stationary increments* if in addition $X(t_2 + s) - X(t_1 + s)$ has the same distribution as $X(t_2) - X(t_1)$ for all $t_1, t_2 \in T$ and any $s > 0$.

Examples

- Sequence of i.i.d random variables.

- General random walk: If $X_1, X_2, \ldots$ is a sequence i.i.d of random variables, then $S_n = \sum_{i=1}^{n} X_i$ is a random walk.

- Weiner process: $\{X(t), t \geq 0\}$ is a Weiner process if
  1. $X(0) = 0$
  2. $\{X(t), t \geq 0\}$ has stationary and independent increments
  3. for every $t > 0$, $X(t)$ is normally distributed with mean 0 and variance t.

- $\{X(t), t \geq 0\}$ is a Markov process if for $t_1 < t_2 < \ldots t_n < t$ we have
  $P(X(t) \leq x | X(t_1) = x_1, \ldots, X(t_n) = x_n) = P(X(t) \leq x | X(t_n) = x_n)$

- Random walk and Weiner process are examples of Markov processes.

Bernoulli/Binomial process

- Bernoulli(p) random variable

- Bernoulli process is a sequence of independent r.v.'s $\{X_i, i = 1, 2, \ldots\}$ where each X_i is a Bernoulli(p) random variable.

- Binomial random variable S_n counts the sum of n independent Bernoulli(p) variables

- let X_i denote the associated Bernoulli variable for toss i, $i = 1, \ldots, n$. Then $S_n = \sum_{i=1}^{n} X_i$ denotes the number of heads/event and $P(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}$.

- $E[S_n]$? $\mathrm{Var}(S_n)$?

Bernoulli/Binomial process

- $\{S_n = \sum_{i=1}^{n} X_i, n = 1, 2, \ldots\}$ is called as a Binomial process.
- Let $T := \{\text{smallest } n : S_n > 0.\}$.
- T is a geometric random variable with parameter p, i.e., $P(T = n_1) = p(1-p)^{(n_1-1)}$.
- Memoryless property: $P(T > m + n / T > n) = P(T > m)$.

Counting process

Stochastic process $\{N(t), t \geq 0\}$ is a counting process if it represents the total number of events upto time t.

It satisfies the following

- $N(t) \geq 0$ and is integer valued
- For $s \leq t$, we have $N(s) \leq N(t)$. $N(t) - N(s)$ denotes the number of events in the interval (t, s)
- $N(t)$ can have independent increments
- $N(t)$ can have stationary increments

Poisson process

A Poisson process with rate $\lambda, \lambda \geq 0$ is a counting process $\{N(t), t \geq 0\}$ with the following properties

- $N(0) = 0$
- $N(t)$ has independent and stationary increments
- Number of events in an interval of length t is a Poisson distribution with mean λt. (Hence stationary increments)
- $E[N(t+s) - N(t)] = \lambda s$

Condition 3 is difficult to verify ! Hence ...

Poisson process – Alternative definition

A function f is said to be $o(h)$ if $\lim_{h\to 0}\frac{f(h)}{h} = 0$.

A Poisson process with rate $\lambda, \lambda \geq 0$ is a counting process $\{N(t), t \geq 0\}$ with the following properties

- $N(0) = 0$
- $N(t)$ has independent and stationary increments
- $P\{N(h) = 1\} = \lambda h + o(h)$
- $P\{N(h) \geq 2\} = o(h)$

Poisson process

Lemma

Definition 1 $\implies$ *Definition 2*

Proof on board.

Lemma

Definition 2 $\implies$ *Definition 1*

Self Study: Refer Sheldon Ross, Stochastic processes, Theorem 2.1.1

Poisson Processes Definition 3

A ctsp $\{N(t), t \geq 0\}$ is a Poisson process with rate $\lambda > 0$ if

- $N(0) = 0$
- $N(t)$ is a counting process with stationary and independent increments
- X_i, the time interval between $i - 1$th and ith event is exponentially distributed with parameter λ.

Lemma

Definition 1/2 $\Longrightarrow$ *Definition 3*

Proof:

- What is $P(X_1 > t) =?$

$$P(X_1 > t) = P(N(0, t) = 0) = e^{-\lambda t}$$

- This implies $F_{X_1}(t) = P(X_1 \leq t) = 1 - e^{-\lambda t}$ and hence X_1 has exponential distribution.
- What is $P(X_2 > t | X_1 = s)$?

$$\begin{aligned} P(X_2 > t | X_1 = s) &= P(N(s, t + s] = 0 | X_1 = s) \\ &= P(N(s, t + s] = 0) \text{ (indep. increments)} \\ &= e^{-\lambda t} \text{ (stat. increments)} \end{aligned}$$

- This implies X_2 is exponential.Repeating the arguments yields the lemma.

Definition 3 $\implies$ Definition 1

Lemma

i.i.d exponential interarrival time implies $N(0,t)$ has Poisson distribution with rate λt.

- Let $S_0 = 0$ and $S_n = \sum_{i=1}^{n} X_i$
- If $S_n = t$, we say that the nth renewal happened at time t.
- $f_{S_n}(t) = \lambda \left[\frac{(\lambda t)^{n-1} e^{-\lambda t}}{n-1!}\right]$ and $F_{S_n}(t) = \int_{x=0}^{t} \lambda \left[\frac{(\lambda x)^{n-1} e^{-\lambda x}}{n-1!}\right] dx$

More on $F_{S_n}(t)$

- $F_{S_n}(t) = \int_0^t \lambda \left[\frac{(\lambda x)^{n-1} e^{-\lambda x}}{n-1!}\right] dx$
- Integration by parts $(u(x) = e^{-\lambda x},\ v'(x) = \lambda \left[\frac{(\lambda x)^{n-1}}{n-1!}\right])$

$$\int_a^b u(x)v'(x)dx = [u(x)v(x)]_a^b - \int_a^b u'(x)v(x)dx$$

- $F_{S_n}(t) = \left[\frac{(\lambda x)^n e^{-\lambda x}}{n!}\right]_0^t - \int_0^t \left[\frac{-\lambda e^{-\lambda x}(\lambda x)^n}{n!}\right] dx$
- $F_{S_n}(t) = \frac{(\lambda t)^n e^{-\lambda t}}{n!} + F_{S_{n+1}}(t)$

$$F_{S_n}(t) - F_{S_{n+1}}(t) = \frac{(\lambda t)^n e^{-\lambda t}}{n!}$$

Relation between S_n and $N(t)$

$$N(t) = sup\{n : S_n \leq t\}$$

$$N(t) \geq n \Leftrightarrow S_n \leq t$$

- $P\{N(t) \geq n\} = P\{S_n \leq t\}$
- $P\{N(t) = n\} = P\{N(t) \geq n\} - P\{N(t) \geq n+1\}.$
- $P\{N(t) = n\} = P\{S_n \leq t\} - P\{S_{n+1} \leq t\}.$
- $P\{N(t) = n\} = Poisson(\lambda t).$

Lemma

Exponential interarrival times imply $N(t)$ has Poisson distribution with rate λt

Properties of Poisson Process (Self Study)

Merging: Merging two independent Poisson processes with rate λ_1 and λ_2 leads to a Poisson process with rate $\lambda_1 + \lambda_2$.

Splitting: If you label each event point of a Poisson(λ) process as type A or type B with probability p or $1 - p$ respectively, then Events of type A form a Poisson $(p\lambda)$ process. Similarly Events of type B form a Poisson $((1-p)\lambda)$ process.

Conditional distribution of Arrival times

Lemma

Given that 1 event of $P.P.(\lambda)$ has happened by time t, it is equally likely to have happened anywhere in $[0, t]$ i.e., $P\{X_1 < s|N(t) = 1\} = \frac{s}{t}$.

Proof.

$$
\begin{aligned}
P\{X_1 < s|N(t) = 1\} &= \frac{P\{X_1 < s, N(t) = 1\}}{P(N(t) = 1)} \\
&= \frac{P\{N[0,s) = 1, N[s,t] = 0\}}{P(N(t) = 1)} \\
&= \frac{P\{N[0,s) = 1\}P\{N[s,t] = 0\}}{P(N(t) = 1)} \\
&= \frac{\lambda s e^{-\lambda s} e^{-\lambda(t-s)}}{\lambda t e^{-\lambda t}} = \frac{s}{t}
\end{aligned}
$$

□

First Queueing Example: Infinite server Queues

- Imagine a system with infinite servers and jobs arrive to this system according to $PP(\lambda)$.
- Every arriving job has a independent service requirement with distribution G and is immediately assigned a server for service.
- When the job receives service, he leaves the system.
- Let $N(t)$ denote the number of arrivals till time t.
- Let $X(t)$ denote the number of customers present in this system at time t.
- Example of such systems: Malls, Tourist spots, Gardens, number of active phone calls, etc

First Queueing Example: Infinite server Queues

- What is the pmf of $X(t)$, i.e., $P(X(t) = k)$?
- First condition on $N(t)$. What is $P(X(t) = k|N(t) = n)$?
- Of the n jobs that arrived (uniformly placed in the interval $[0, t]$), k are yet to complete service.
- Let p denote the probability that an arbitrary of these customers is still receiving service at time t.
- Then $P(X(t) = k|N(t) = n) = \binom{n}{k} p^k (1-p)^{n-k}$.
- Now unconditioning on $N(t)$, we get

$$\begin{aligned} P(X(t) = k) &= \sum_{n=k}^{\infty} P(X(t) = k|N(t) = n)P(N(t) = n) \\ &= e^{-\lambda tp} \frac{(\lambda tp)^j}{j!} \end{aligned}$$

where $p = \int_0^t (1 - G(t - x)) \frac{dx}{t}$.

CS 3.307
Performance Modeling for Computer Systems

Tejas Bodas

Assistant Professor, IIIT Hyderabad

Markov Process

- There are two versions of Markov chains- Discrete time and Continuous time.
- A stochastic process $\{X_n, n \in \mathbb{Z}_+\}$ is a discrete time Markov chain if for any $n_1 < n_2 < \ldots < n_k < n$,

$$P(X_n = j | X_{n_1} = x_1, .., X_{n_k} = i) = P(X_n = j | X_{n_k} = i)$$

- $\{X(t), t \geq 0\}$ is a Markov process (ctmc) if for $t_1 < t_2 < \ldots t_n < t$,

$$P(X(t) = j | X(t_1) = x_1, .., X(t_n) = i) = P(X(t) = j | X(t_n) = i)$$

- This is known as the Markov property.
- State space in both cases can be integers or general ($\mathbb{R}^d$)
- We will stick with integer or finite state space

Example: Coin with memory!

- In a Markovian coin with memory, the outcome of the next toss depends on the current toss.
- $X_n = 1$ for heads and $X_n = -1$ otherwise. $\mathcal{S} = \{+1, -1\}$.
- Sticky coin : $P(X_{n+1} = 1 | X_n = 1) = 0.9$ and $P(X_{n+1} = -1 | X_n = -1) = 0.8$ for all n.
- Flippy Coin: $P(X_{n+1} = 1 | X_n = 1) = 0.1$ while $P(X_{n+1} = -1 | X_n = -1) = 0.3$ for all n.
- This can be represented by a transition diagram (see board)
- The transition probability matrix P for the two cases is
  $P_s = \begin{bmatrix} 0.9 & .1 \\ 0.2 & 0.8 \end{bmatrix}$ and $P_f = \begin{bmatrix} 0.1 & 0.9 \\ 0.7 & 0.3 \end{bmatrix}$
- The row corresponds to present state and the column corresponds to next state.

Running example: Dice with memory!

- In a markovian dice with memory, the outcome of the next roll depends on the current roll.
- $X_n = i$ for $i \in \mathcal{S}$ where $\mathcal{S} = \{1, \ldots, 6\}$.
- Example transition probability matrix

$$P = \begin{bmatrix} 0.9 & .1 & 0 & 0 & 0 & 0 \\ 0 & .9 & .1 & 0 & 0 & 0 \\ 0 & 0 & 0.9 & 0.1 & 0 & 0 \\ 0 & 0 & 0 & 0.9 & 0.1 & 0 \\ 0 & 0 & 0 & 0 & 0.9 & 0.1 \\ 0.1 & 0 & 0 & 0 & 0 & 0.9 \end{bmatrix}$$

- In the ctmc counterpart for these examples, imagine the coin tosses itself/ dice rolls itself after waiting in the state for a random time that is exponentially distributed. (more later)

Time-homogenous Markov Process

- A DTMC is said to be time homogeneous if the one step transition probabilities are same at all time.
- $P(X_{n+1} = j | X_n = i) = P(X_{n+1+s} = j | X_{n+s} = i) := p_{ij}$
- One step transition probability matrix $P = [[p_{ij}]]$
- $i, j \in \mathcal{S}$ which is countable and $|\mathcal{S}| \leq \infty$

For a CTMC ...

- For a time homogeneous CTMC, we have

$$\begin{aligned} P(X(t)=j|X(t_n)=i) &= P(X(t+s)=j|X(t_n+s)=i) \\ &= P(X(t-t_n)=j|X(0)=i). \end{aligned}$$

- We have a transition probability matrix with entries $p_{ij}(t)$, i.e., $P(t) = [[p_{ij}(t)]]$.

DTMC – Time spent in a state

- For a time homogeneous DTMC, we have a transition probability matrix with entries p_{ij}, i.e., $P = [[p_{ij}]]$.
- Let $Y_n = inf\{s > 0 : X_{n+s} \neq X_n\}$
- Y_n is the remaining time that the process spends in whichever state it is in, at time n.
- Consider a Markov coin, its state transition matrix and diagram
- Y_n is geometric random variable.
- What would be the time spent in a state for a continuous time Markov chain ?

CTMC – Time spent in a state

- For a time homogeneous CTMC, we have a transition probability matrix with entries $p_{ij}(t)$, i.e., $P(t) = [[p_{ij}(t)]]$.
- Let $Y_t = inf\{s > 0 : X(t+s) \neq X(t)\}$
- Y_t is the remaining time that the process spends in whichever state it is in, at time t.
- Intuitively, the time spent in a state should depend only on what state it is in, and not on the previous state.

Theorem

$$P(Y_t > u | X(t) = i) := \bar{G}_i(u) = e^{-a_i u}$$

for all $i \in \mathcal{S}$ *and* $t \geq 0, u \geq 0$ *and for some real number* $a_i \in [0, \infty]$.

Proof 1

- $\bar{G}_i(u+v) = P(X(s) = i, s \in [t, t+u+v] | X(t) = i)$
- $\bar{G}_i(u+v) = P(X(s) = i, s \in [t+u, t+u+v]; X(p) = i, p \in [t, t+u] | X(t) = i)$
- $P(AB|C) = P(A|BC)P(B|C)$
- Due to Markov property we have $P(AB|C) = P(A|B)P(B|C)$
- $P(X(s) = i, s \in [t+u, t+u+v] | X(p) = i, p \in [t, t+u]) =$
- $P(X(s) = i, s \in [t+u, t+u+v] | X(t+u) = i) = \bar{G}_i(v)$
- $P(X(p) = i, p \in [t, t+u] | X(t = i)) = \bar{G}_i(u)$
- $\bar{G}_i(u+v) = \bar{G}_i(u)\bar{G}_i(v)$
- Only CCDF function which satisfies this equation is the exponential distribution. This requires a proof. We will skip this part.

Simpler Proof

- Let τ_i denote the time the CTMC spends in state i before moving out. Suppose the CTMC is in state i at time 0.
- What is $P(\tau_i > s + t | \tau_i > s)$?
- Note that $X(s) = i$ and therefore from the Markov property,

$$\begin{aligned} P(\tau_i > s + t | \tau_i > s) &= P(X(u) = i, u \in [s, s + t] | X(t) = i, t \in [0, s]) \\ &= P(X(u) = i, u \in [s, s + t] | X(s) = i) \\ &= P((X(u) = i, u \in [0, t] | X(0) = i) \\ &= P(\tau_i > t). \end{aligned}$$

- Since $P(\tau_i > s + t | \tau_i > s) = P(\tau_i > t)$, this implies the distribution has memoriless property and must be exponential.

Finite dimensional distributions

- Consider a DTMC $\{X_n, n \geq 0\}$ with tpm denoted by P.
- We assume M states and X_0 denotes the initial state.
- You can start in any starting state or may pick your starting state randomly.
- Let $\bar{\mu} = (\mu_1, \ldots, \mu_M)$ denote the initial distribution.
- How does one obtain the finite dimensional distribution $P(X_0 = x_0, X_1 = x_1, \ldots X_k = x_k)$?

Finite dimensional distributions

- Consider a CTMC $\{X_t, t \geq 0\}$ with t-time pm given by $P(t)$.
- We assume M states and X_0 denotes the initial state.
- Let $\bar{\mu} = (\mu_1, \ldots, \mu_M)$ denote the initial distribution.
- How does one obtain the finite dimensional distribution $P(X_0 = x_0, X_{t_1} = x_1, \ldots X_{t_k} = x_k)$?

Chapman Kolmogorov Equations for DTMC

- $P = [[p_{ij}]]$ denotes the one step transition probability matrix.
- Let $P^{(n)}$ denote the n-step transition probability matrix.
- CK equation tells us that $P^{(n+l)} = P^{(n)}P^{(l)}$.
- $p_{ij}^{(n+l)} = P(X_{n+l} = j | X_0 = i) = \sum_k P(X_{n+l} = j, X_n = k | X_0 = i)$
- $p_{ij}^{(n+l)} = \sum_k P(X_{n+l} = j | X_n = k, X_0 = i) P(X_n = k | X_0 = i)$
- $p_{ij}^{(n+l)} = \sum_k P(X_{n+l} = j | X_n = k) P(X_n = k | X_0 = i)$
- $p_{ij}^{(n+l)} = \sum_k p_{ik}^{(n)} p_{kj}^{(l)} = [P^{(n)}P^{(l)}]_{ij}$
- At which step did we use time homogeneity and the Markov property?

n step transition probabilities

- $P = [[p_{ij}]]$ denotes the one step transition probability matrix.
- Let $P^{(n)}$ denote the n-step transition probability matrix.
- From the CK equation we know that $P^{(n+l)} = P^{(n)}P^{(l)}$.
- It is easy to see that $P^{(n)} = P^{(n-1)}P$.
- For an M state DTMC, $p_{ij}^{(2)} = \sum_{k=1}^{M} p_{ik}p_{kj}$.
- This implies that that the n-step transition probability matrix can be obtained as $P^{(n)} = P^n$
- Given X_0 and P, you can generate n-step probabilities or $P_{X_0}(X_n)$

Chapman Kolmogorov Equations for CTMC

- Let $P(t)$ denote the t-time transition probability matrix.
- CK equation for a CTMC is $P(t+l) = P(t)P(l)$.
- $p_{ij}(t+l) = P(X(t+l) = j | X(0) = i)$
- $= \sum_k P(X(t+l) = j, X(t) = k | X(0) = i)$
- $= \sum_k P(X(t+l) = j | X(t) = k, X(0) = i) P(X(t) = k | X(0) = i)$
- $= \sum_k P(X(t+l) = j | X(t) = k) P(X(n) = k | X(0) = i)$
- $p_{ij}(t+l) = \sum_k p_{ik}(t) p_{kj}(l) = [P(t)P(l)]_{ij}$

What generates a CTMC ?

- $P(t+l) = P(t)P(l)$.
- In DTMC, we could use P to generate the chain on Matlab.
- What about CTMC ? Can we use $P(t)$?
- What is $\lim_{h \to 0} P(h)$?
- What is $\frac{dP(h)}{dh}$ evaluated at $h = 0$?

What generates a CTMC ?

- Lets look at $\frac{dP(h)}{dh}|_{h=0} = \lim_{h \to 0} \frac{P(h)-P(0)}{h} = \lim_{h \to 0} \frac{P(h)-I}{h}$.
- Define $Q := \lim_{h \to 0} \frac{P(h)-I}{h}$
- **Does it always exist ?** Yes! (Proposition 2.2 and 2.4 (Anderson))
- Q has terms of the form q_{ii} and q_{ij} for $i, j \in \{1, 2, \ldots, M\}$.
- $q_{ii} = \frac{dp_{ii}(h)}{dh}|_{h=0}$. Similarly $q_{ij} = \frac{dp_{ij}(h)}{dh}|_{h=0}$

What generates a CTMC ?

Theorem

Let $P(t)$ be a transition function. Then the generator matrix $Q = \lim_{h \to 0} \frac{P(h)-I}{h}$ exists.

Theorem

$P(Y_t > u | X(t) = i) = e^{-a_i u}$ *where* $a_i > 0$.

Theorem

(Proposition 2.8 Anderson)
$P(Y_t > u | X(t) = i) := e^{q_{ii} u}$, *i.e.*, $q_{ii} = -a_i$.
$P\left(X(t + Y_t) = j | X(t) = i\right) = \frac{q_{ij}}{|q_{ii}|}$.

Q generates the CTMC

- Cannot generate CTMC directly from $P(t)$.
- From $P(t)$, obtain Q using $Q = \frac{dP(h)}{dh}|_{h=0}$
- Consider Y_t when $X(t) = i$.
- Now use the following theorem for generating the CTMC on a computer

Theorem

(Proposition 2.8 Anderson: we won't see proof)
$P(Y_t > u | X(t) = i) := e^{q_{ii} u}$, i.e., $q_{ii} = -a_i$.
$P\left(X(t + Y_t) = j | X(t) = i\right) = \frac{q_{ij}}{|q_{ii}|}$.

Properties of a conservative Q

Theorem

$P(Y_i > u | X(t) = i) := e^{q_{ii} u}$, *i.e.*, $q_{ii} = -a_i$.
$P\left(X(t + Y_t) = j | X(t) = i\right) = \frac{q_{ij}}{|q_{ii}|}$ *where* $q_{ij} \geq 0$.

- Suppose Q is conservative.
- Recall that q_{ii} is negative. A conservative Q implies $q_{ii} = -\sum_{j \neq i} q_{ij}$.
- $|q_{ii}|$ is the exponential rate at which you leave state i.
- q_{ij} is the exponential rate at which you leave state i to go to state j.
- minimum of exponentials is exponential with aggregated rate.
- This justifies the rate of leaving state i to be $\sum_{j \neq i} q_{ij}$.

Equivalent definition of a CTMC using Q

- Suppose Q is conservative.
- Then in the CTMC, you stay in state i for a random duration that has exponential($|q_{ii}|$) distribution.
- From i, you will move to state j with probability $\frac{q_{ij}}{|q_{ii}|}$.
- Equivalently, in state i, you have $M-1$ exponential(q_{ij}) clocks for $j = 1, 2, \dots, i-1, i+1, \dots M$.
- You move to that state whose clock rings first!

RECAP

CK Equations: $P(t+l) = P(t)P(l)$

Theorem

Let $P(t)$ be a transition function. Then the generator matrix $Q = \lim_{h \to 0} \frac{P(h)-I}{h}$ exists.

Theorem

$P(Y_t > u | X(t) = i) = e^{-a_i u}$ *where* $a_i > 0$.

Theorem

For a CTMC with Q matrix, we have
$P(Y_t > u | X(t) = i) := e^{q_{ii} u}$, *i.e.,* $q_{ii} = -a_i$.
$P(X(t+Y_t) = j | X(t) = i) = \frac{q_{ij}}{|q_{ii}|}$.

Kolmogorov forward/backward equations CTMC

- $\frac{dP(t)}{dt} = \lim_{s \to \infty} \frac{P(t+s) - P(t)}{s}$

- $\frac{dP(t)}{dt} = P(t) \lim_{s \to \infty} \frac{P(s) - I}{s}$

- $\frac{dP(t)}{dt} = P(t)Q.$

- $P(t) = e^{tQ}$ satisfies the above. (Calculus of Matrix exponentials)

- $P(t) = e^{tQ} := I + tQ + \ldots + \frac{(tQ)^n}{n!} \ldots$

Example: Poisson process $N(t)$ as a CTMC

- States $\mathcal{S} = Z_{\geq 0}$.
- Why is it a Markov process / Markov property satisfied?
- $P(N(t) = k | N(t_1) = k_1, \ldots, N(t_m) = k_m) = P(N(t) = k | N(t_m) = k_m)$?
- $P(N(t) = k | N(t_1) = k_1, \ldots, N(t_m) = k_m) = P(N(t - t_k) = k - k_m)$. Therefore the above is true.
- $p_{ij}(t) = P(N(t) = j | N(0) = i)$. $max(j - i, 0)$ arrivals in time t.
- We know that this has Possion distribution.
- How does $P(t)$ look for a Poisson process ?

Example: Poisson process $N(t)$ as a CTMC

- How does $P(t) = [[P(N(t) = j | N(0) = i)]]$ look ?
- Entries below the diagonal are zero.
- Diagonal entries have the value $e^{-\lambda t}$
- ijth entry above the diagonal has the value $e^{-\lambda t} \frac{(\lambda t)^{j-i}}{(j-i)!}$

Example: Poisson process $N(t)$ as a CTMC

- How does $Q = \frac{dP(h)}{dh}|_{h=0}$ look ?
- ijth entry above the diagonal $p_{ij}(t) = e^{-\lambda t}\frac{(\lambda t)^{j-i}}{(j-i)!}$
- what is $\frac{d}{dt}(e^{-\lambda t}\frac{(\lambda t)^{j-i}}{(j-i)!})|_{t=0}$?
- If $j - i = 1$, then $\frac{d}{dt}(e^{-\lambda t}\frac{(\lambda t)^{j-i}}{(j-i)!}) = \lambda$.
- If $j - i > 1$, then $\frac{d}{dt}(e^{-\lambda t}\frac{(\lambda t)^{j-i}}{(j-i)!}) = 0$.
- How does Q for Poisson process look like ?
- $P(t) = e^{tQ} = I + tQ + \ldots + \frac{(tQ)^n}{n!} + \ldots$

Example 3: Binomial process as a DTMC

- DO IT YOURSELF!

Embedded DTMC in a CTMC

- Consider a CTMC over state space $\mathcal{S}$.
- Let $Y_n, n \geq 0$ denote the sequence of times spent in successive states of the CTMC
- Define T_n to be the jump times of the CTMC, i.e., the times of successive state transitions.
- Then $T_n = \sum_{k=1}^{n} Y_k$.
- Define $X_n = X(T_n)$ for $n \geq 0$. X_n is the DTMC embedded in the CTMC.
- The corresponding TPM has $p_{ij} = \frac{q_{ij}}{|q_{ii}|}$.
- $\{X_n,\}$ is such that there are no one step transitions from a state to itself, i.e., $p_{ii} = 0$.

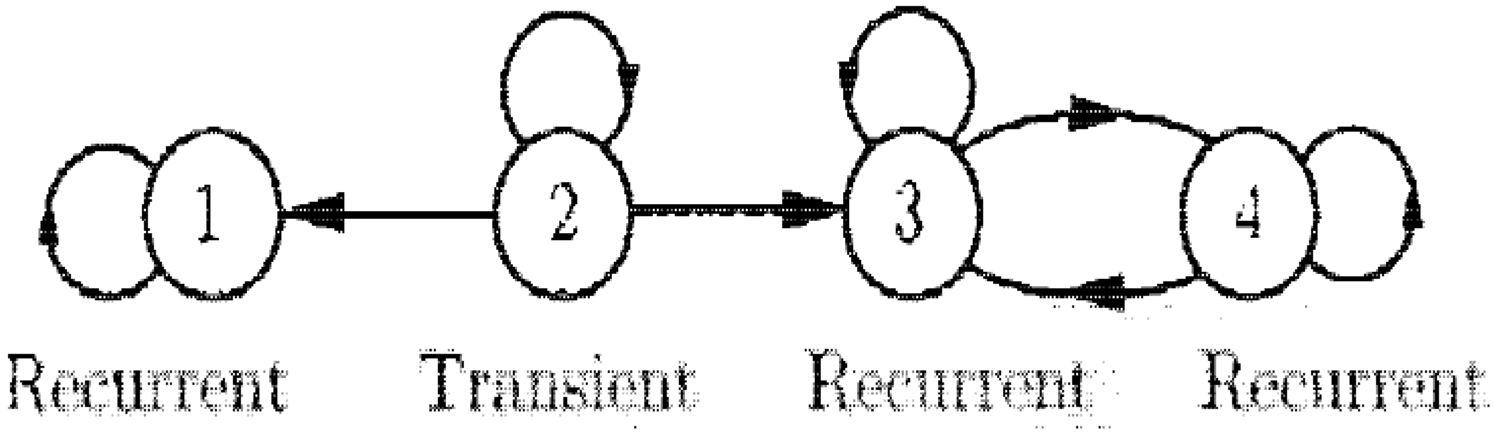
Classification of states

- Consider a Markov process with state space $\mathcal{S}$
- We say that j is accessible from i if $p_{ij}^{n} > 0$ for some n.
- This is denoted by $i \rightarrow j$.
- if $i \rightarrow j$ and $j \rightarrow i$ then we say that i and j communicate.
  This is denoted by $i \leftrightarrow j$.

> A chain is said to be irreducible if $i \leftrightarrow j$ for all $i, j \in \mathcal{S}$.

Recurrent and Transient states

- We say that a state i is recurrent if $F_{ii} = P($ ever returning to i having started in i $) = 1$.
- F_{ii} is not easy to calculate. (We will se ethis after Quiz)
- If a state is not recurrent, it is transient.
- For a transient state i, $F_{ii} < 1$.
- If $i \leftrightarrow j$ and i is recurrent, then j is recurrent.



Limiting probabilities

- $P = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0.6 & 0.4 \\ 0.6 & 0.4 & 0 \end{bmatrix}$ $P^5 = \begin{bmatrix} .06 & .3 & .64 \\ .18 & .38 & .44 \\ .38 & .44 & .18 \end{bmatrix}$ $P^{30} = \begin{bmatrix} .23 & .385 & .385 \\ .23 & .385 & .385 \\ .23 & .385 & .385 \end{bmatrix}$

- $P = \begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix}$ $\lim_{n\to\infty} P^n = \begin{bmatrix} \frac{b}{a+b} & \frac{a}{a+b} \\ \frac{b}{a+b} & \frac{a}{a+b} \end{bmatrix}$

- What is the interpretation of $\lim_{n\to\infty} p_{ij}^{(n)} = [\lim_{n\to\infty} P^n]_{ij}$?

- $\pi_j = \lim_{n\to\infty} p_{ij}^{(n)}$ denotes the probability of being in state j at time n when starting in state i.

- For an M state DTMC, $\bar{\pi} = (\pi_1, \ldots, \pi_M)$ denotes the limiting distribution.

Limiting probabilities

- Do the limiting probabilities always exist ?
- $P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ In this case the limiting probabilities do not exist.
- Now suppose $\bar{\mu} = [.5, .5]$. Then $P(X_1 = 1) = 0.5$. But this is true for every X_n, i.e., $P(X_n = 1) = 0.5$. (already in steady state)
- What is happening ?
- Now suppose $\bar{\mu} = [.1, .9]$. Then $P(X_1 = 1) = 0.9$. But this is **not** true for every X_n, i.e., $P(X_2 = 1) = 0.9, P(X_3 = 1) = 0.1$. (Never in steady-state)

Stationary distribution

- $P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ is a periodic chain for which the the limiting probabilities do not exist.

- What however always exists is known as stationary distribution (not necessarily unique)

- A **stationary distribution** is a probability (row) vector on $\mathcal{S}$ that satisfies $\pi = \pi P$ in case of DTMC.

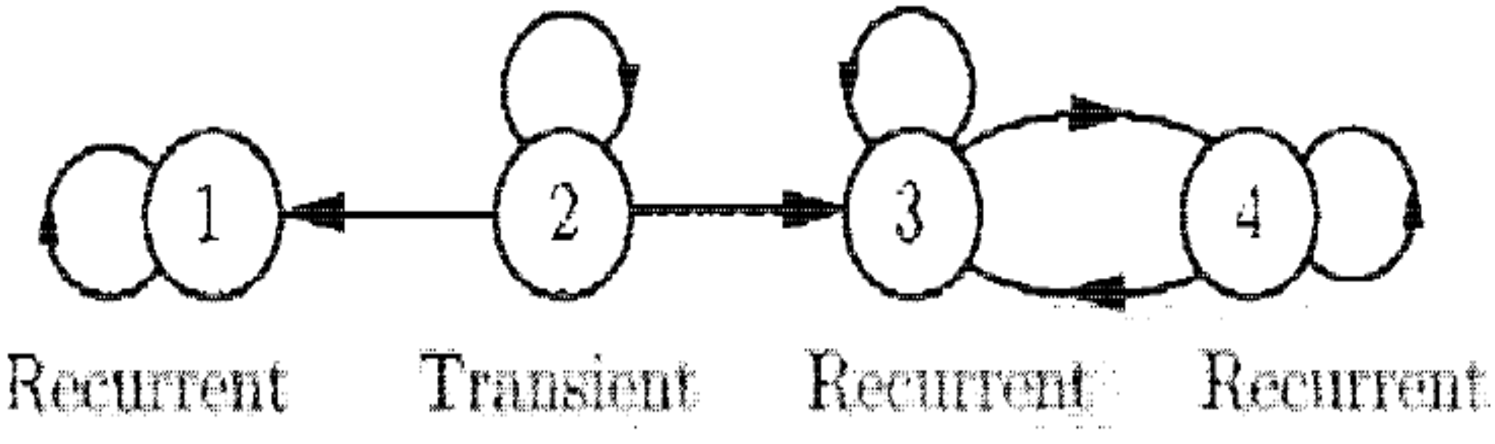
- For a CTMC, we know that $\frac{dP(t)}{dt} = P(t)Q$. When $\lim_{t\to\infty} P(t) = \Pi$, this means that at stationarity $\frac{dP(t)}{dt} = 0$. Therefore we have $\pi Q = 0$ in case of CTMC.

- If the limiting distribution exists, it is equal to its stationary distirbution.

More on Transience and Recurrence

Recall: Recurrent and Transient states

- We say that a state i is recurrent if $F_{ii} = P($ ever returning to i having started in i $) = 1$.
- F_{ii} is not easy to calculate. (We will see this today)
- If a state is not recurrent, it is transient.
- For a transient state i, $F_{ii} < 1$.
- If $i \leftrightarrow j$ and i is recurrent, then j is recurrent.



First passage probabilities

- Consider a DTMC with state space denoted by $\mathcal{S}$.
- $f_{ij}^n := P(X_n = j, X_k \neq j \text{ for } 1 \leq k \leq n-1 | X_0 = i)$. ($f_{ij}^0 = 0$).
- $\{f_{ij}^n, n \geq 0, i, j \in \mathcal{S}\}$ is called the first passage probabilities
- For a fixed n, f^n denotes the matrix $[[f_{ij}^n]]$.
- For a fixed ij pair, $\{f_{ij}^n, n \geq 0\}$ represents a probability mass function on $\mathbb{Z}_+$.
- This mass function can be degenerate which means it can have a point mass at infinity.
- Define: $F_{ij} = \sum_{n=1}^{\infty} f_{ij}^n$.

First passage probabilities

- Define: $F_{ij} = \sum_{n=1}^{\infty} f_{ij}^{n}$.
- F_{ij} has the interpretation of the probability of ever visiting state j when starting from state i.
- If $F_{ij} = p < 1$, then there is a finite probability $1 - p$ with which you may not ever reach j when starting from state i.
- In this case $\{f_{ij}^{n}, n \geq 0\}$ is not a proper mass function since it does not sum to 1.
- If $F_{ij} = 1$, then the from i you can certainly reach j.
- In this case $\{f_{ij}^{n}, n \geq 0\}$ is a proper mass function.
- Let T_{ij} denote the first passage time from i to j.
- T_{ij} has the probability mass function $\{f_{ij}^{n}, n \geq 0\}$. In other words, $P(T_{ij} = k) = f_{ij}^{k}$.

First recurrence probabilities

- Define: $F_{ii} = \sum_{n=1}^{\infty} f_{ii}^{n}$.
- f_{ii}^{n} : probability of starting in i and returning to state i for the first time at time n
- F_{ii} has the interpretation of the probability of ever returninig to state i.
- If $F_{ii} = p < 1$, then there is a finite probability $1 - p$ with which you may not return to state i .
- If $F_{ii} = 1$, then the from i you can certainly return to i.
- For any $i \in \mathcal{M}$, the first return time T_{ii} has the probability mass function $\{f_{ii}^{n}, n \geq 0\}$.

Mean passage and recurrence times

- Let μ_{ij} be the mean (first passage) time from state i to j.
- $\mu_{ij} = \sum_{n=1}^{\infty} n f_{ij}^{n}$
- Let μ_{ii} be the mean recurrence time at state i.
- $\mu_{ii} = \sum_{n=1}^{\infty} n f_{ii}^{n}$
- All the above definitions have an equivalent counterpart in a CTMC.
- For eg: f_{ij}^{t} has a natural interpretation. We wont go further into this.

Transient and recurrent states

- Suppose for a state i we have $F_{ii} = 1$. Then we say that state i is recurrent.
- Once in i, you are certain to come back to i.
- If $\mu_{ii} = \infty$, it is called null recurrent. The chain is bound to return to state i, but possibly after an infinite time.
- $\mu_{ii} < \infty$, it is called positive recurrent.
- If all states of the Markov chain are (null /positive) recurrent, it is called as a (null /positive) recurrent Markov chain.
- Null recurrence is possible in infinite state space models.
- State i is transient if $F_{ii} < 1$.
- You may not return back to i with a finite probability.

Transient and recurrent states

- Consider a DTMC and consider $X_0 = i$.
- Let us count the number of times the chain is in state i.
- Let I_n denote an indicator variable which is 1 if $X_n = i$ and 0 if $X_n \neq i$.
- $\sum_{n=1}^{\infty} I_n$ counts the number of times state i was visited.
- $E[I_n] = P(X_n = i | X_0 = i) = p_{ii}^{(n)}$.
- The mean total number of visits to state i is given by $\sum_{n=1}^{\infty} p_{ii}^{n}$
- Convergence or divergence of this sum also defines transient or recurrent states.

Recurrent criteria

- The mean total number of visits to state i is given by $\sum_{n=1}^{\infty} p_{ii}^{n}$
- Suppose the chain visits state i only exactly n times.
- The P(exactly n visits to i) $= F_{ii}^{n}(1 - F_{ii})$.
- For a recurrent state, $F_{ii} = 1$. Hence P(exactly n visits to i) $=$ 0.
- P(exactly infinite visits to i) $= 1$.
- Mean total number of visits is also infinite and hence $\sum_{n=1}^{\infty} p_{ii}^{n}$ diverges.

Transient state criteria

- The mean total number of visits to state i is given by $\sum_{n=1}^{\infty} p_{ii}^{n}$
- Suppose the chain visits state i only exactly n times.
- The P(exactly n visits to i) $= F_{ii}^{n}(1 - F_{ii})$.
- For transient state i, $F_{ii} < 1$.
- The P(exactly n visits to i) $= F_{ii}^{n}(1 - F_{ii})$. Compare this with geometric random variable.
- Mean total number of visits to state i is $\frac{F_{ii}}{1-F_{ii}}$ which is finite.
- Hence for transient state $\sum_{n=1}^{\infty} p_{ii}^{n}$ must converge.

Thursday

09/02/2023

Lecture #9: Elementary Queues.

$$\text{Geom}(p) / \text{Geom}(q) / 1 / 1$$

- $\text{Geom}(p)$ → Arrival process → Bernoulli process. $\longleftrightarrow$ Geometric inter-arrival.
- $\text{Geom}(q)$ → Service rate
- 1 → No. of server
- 1 → No. of positions in the system.

$$X(n) = \begin{cases} 1 & \text{if service is busy at } n \\ 0 & \text{otherwise.} \end{cases}$$

"Arrivals when server is busy are dropped"

$P(1 \mid 0) = p$

$P_{01} = p$

$P_{00} = 1-p$

$P_{10} = q$

$$\begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}$$

Memoryless prop. → Markov prop.

Assumption : No ~~two~~ arrival & completion happen at the same time

(No simultaneous Transitions)

Arrivals can happen, but they will be dropped.

$$\pi = \pi P$$

$$[x \; y] \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix} = [x \; y]$$

$$x = \frac{q}{p+q}, \quad y = \frac{p}{p+q}$$

$$E(X_n) = \frac{p}{p+q}$$

(At stationary)

/ when the markov chain has become stationary.

___/___/___

M / M / 1 / 1

$\downarrow$ PP(λ)

$\downarrow$ service ~~ame~~ exp(μ)

$$X(t) = \begin{cases} 1 \\ 0 \end{cases}$$

$$Q = \begin{bmatrix} q_{00} & q_{01} \\ q_{10} & q_{11} \end{bmatrix}$$

$q_{00} = -\lambda \qquad q_{01} = \lambda$

$q_{10} = \mu \qquad q_{11} = -\mu.$

q_{ij} = rate at which the CTMC moves from state i to state j.

$$P(t) = e^{Qt} = \begin{bmatrix} e^{-\lambda t} & e^{\lambda t} \\ e^{\mu t} & e^{-\mu t} \end{bmatrix}$$

__/__/____

$$\pi_0 = \frac{\mu}{\lambda+\mu}, \quad \pi_1 = \frac{\lambda}{\lambda+\mu}.$$

$$E(X) = \frac{\lambda}{\lambda+\mu}.$$

M/M/1/∞

$N(t)$ = No. of jobs of the system at time t.

$$Q = \begin{array}{c|c} & \begin{array}{ccccc} 0 & 1 & 2 & 3 & \cdots \end{array} \\ \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \end{array} & \begin{bmatrix} -\lambda & \lambda & 0 & 0 & \\ \mu & -(\lambda+\mu) & \lambda & 0 & \cdots \\ 0 & \mu & -(\lambda+\mu) & \lambda & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \end{array}$$

or
$$q_{ij} = \begin{cases} \mu, & \text{if } i = j+1,\ i \neq 0 \\ -(\mu+\lambda), & \text{if } i = j,\ i \neq 0 \\ \lambda, & \text{if } i = j-1,\ i \neq 0 \\ -\lambda, & i = 0, j = 0 \\ \lambda, & i = 0, j = 1 \end{cases}$$

___/___/___

$$\pi Q = 0$$

$$\begin{bmatrix}\pi_0 & \pi_1 & \cdots\end{bmatrix} \begin{bmatrix} -\lambda & \lambda & 0 & 0 \\ \mu & -(\mu+\lambda) & \lambda & 0 \\ 0 & \mu & -(\mu+\lambda) & \lambda \cdots \end{bmatrix} = 0.$$

$$-\lambda \pi_0 + \mu \pi_1 = 0$$

Gen. eq.: $\lambda \pi_{n-1} + \mu \pi_{n+1} = (\lambda+\mu)\pi_n$

$$\pi_1 = \frac{\lambda}{\mu} \pi_0$$

$$\lambda \pi_0 - (\lambda+\mu)\pi_1 + \mu \pi_2 = 0$$

$$\pi_2 = \frac{\lambda}{\mu} \pi_1$$

$\therefore$ generalizing, $\pi_i = \frac{\lambda}{\mu} \pi_{i-1}$.

$$\therefore \pi_i = \left(\frac{\lambda}{\mu}\right)^i \pi_0$$

$$\sum_{i=0}^{\infty} \pi_i = 1$$

$$\therefore \pi_0 = 1 - \frac{\lambda}{\mu}$$

$$\rho = \text{load} = \frac{\lambda}{\mu}.$$

$$\pi_i = \rho^i (1-\rho)$$

Homework: $E(N) = ?$

$$= \sum_{0}^{\infty} i \, \pi_i$$

$$= \sum i \, \rho^i (1-\rho)$$

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

$$= \frac{\rho}{1-\rho}.$$

$$= \frac{\frac{\lambda}{\mu}}{1-\frac{\lambda}{\mu}}$$

$$= \frac{\lambda}{\mu-\lambda}.$$

Lecture #10:

Global Blalance equations.

$$\lambda \pi_{n-1} + \mu \pi_{n+1} = (\lambda+\mu)\pi_n$$

for any general equation.

$$\begin{tikzcd}
n-1 \arrow[r, bend right, "\lambda"'] & n \arrow[l, bend right, "\mu"'] \arrow[r, bend left, "\lambda"] & n+1 \arrow[l, bend left, "\mu"]
\end{tikzcd}$$

$$\left\{\begin{matrix}\text{Rate of leaving } n \\ \pi_n(\lambda+\mu)\end{matrix}\right\} = \left\{\begin{matrix}\text{Rate of entry } n \\ \pi_{n-1}\lambda + \pi_{n+1}\mu.\end{matrix}\right\}$$

Global balance equations used for solving solving π stationary distribution.

Little's Law:

$$E(N) = \lambda \cdot E(T)$$

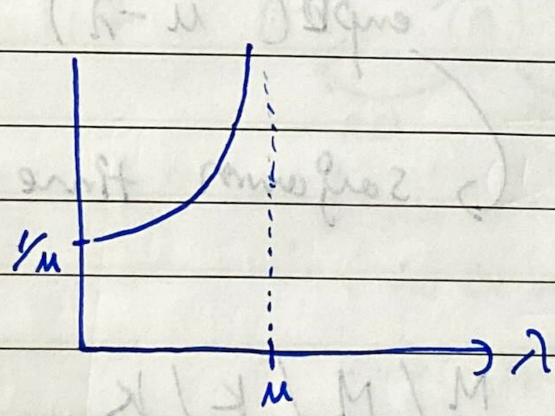
$N \to$ No. of people in the system

$T \to$ Response time

T = service time + waiting time
= time spent by each job in the system.

Renewal process → Generalization of poisson process.

$$E(T) = \frac{E(N)}{\lambda} = \frac{1}{\mu - \lambda}.$$



$$T = W + S$$

$$N = N_q + 1_s$$

$E[N_q] = \lambda E(W)$ → Try to derive (Homework).

$P(T > t)$

$\Rightarrow \sum_n P(T > t / N = n)\, P(N = n)$

$\Rightarrow \sum_n P\left(\sum_{j=0}^{n} B_j > t\right)(1-\rho)\rho^n$ → Gamma / Erlang.

$\Rightarrow$ ↓ ⟶ Homework

$Exp(\mu(1-\rho))$

$exp(\mu - \lambda)$

↳ sojourn time distribution.*

M / M / K / K.

① ② ③ . . . Ⓚ

One possible descriptor

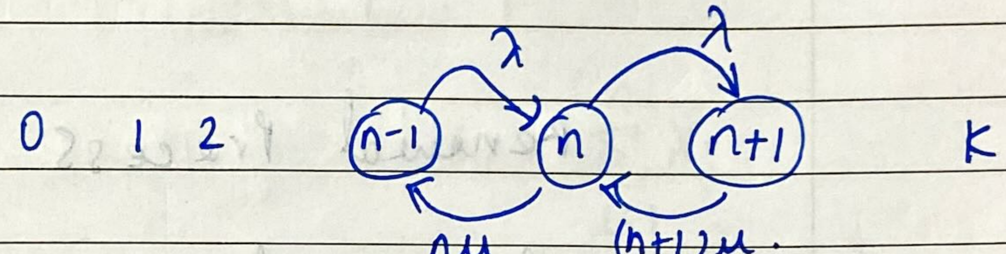
(Hard) $Desc\ 1 = (I_1, I_2, \ldots, I_k)$

$$I_i = \begin{cases} 1 & \text{if busy} \\ 0 & \text{else} \end{cases}$$

This makes sense if μ_i's are different, but if they are same then

Desc. 2 = No. of busy servers

Easier to describe.



This will only go up till K.

min(All the busy server)

= exp(μ + μ + ... + μ)

= exp(nμ).

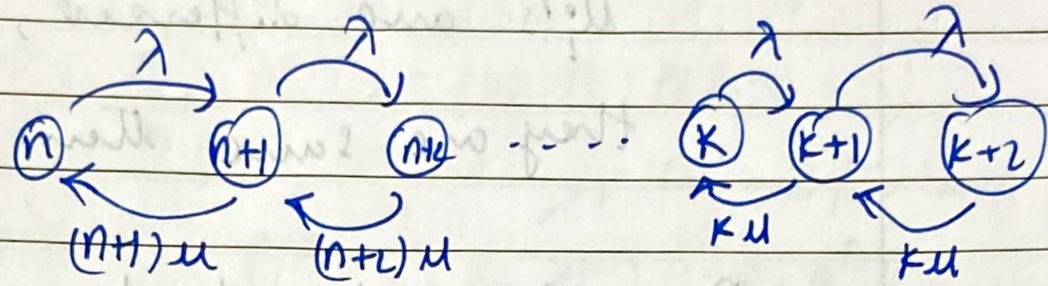
Home work: Calculate E(N) & π_n.

The quantities we have to obtain

E(W), E(T) using Littles Law

π_n E(W) E(N_q)

$M/M/K/\infty$



Check book by Mohr Balter.

Renewal Process

→ A renewal process is a counting process for which the interarrival

→ Renewal process is generalization of poisson process.

Performance modeling of CS

Tejas Bodas

Assistant Professor, IIIT Hyderabad

Renewal Processes

A Renewal process is a counting process for which the inter-arrival times are i.i.d with an arbitrary distribution.

Renewal process is a generalization of Poisson process where the inter-arrival times were i.i.d exponential.

Renewal Processes - Notations

- $\{X_n, n \geq 0\}$ denote the sequence of inter-arrival times of a renewal process.
- X_n is the time between $n-1$th and nth renewal.
- $\{X_n, n \geq 0\}$ are non negative iid random variables with law F.
- Let $S_0 = 0$ and $S_n = \sum_{i=1}^{n} X_i$

> The term renewal process refers to any of the following:
> 1) The sequence $\{X_n, n \geq 0\}$ of inter-arrival times
> 2) The sequence $\{S_n, n \geq 0\}$ of arrival times
> 3) The associated counting process $\{N(t), t \geq 0\}$.

Renewal Process Examples

- Bernoulli/Binomial process.
- Poisson process.
- Successive times between your water bottle gets empty.
- Time instants when number of customers in Ikea is exactly 10.
- Time between successive visits to a particular state of a Markov Chain.

Relation between S_n and $N(t)$

- Define $N(t) = sup\{n : S_n \leq t\}$. $N(t)$ signifies the number of renewals until time t.

$$N(t) \geq n \Leftrightarrow S_n \leq t$$

- $P\{N(t) \geq n\} = P\{S_n \leq t\}$
- $P\{N(t) = n\} = P\{N(t) \geq n\} - P\{N(t) \geq n+1\}$.
- $P\{N(t) = n\} = P\{S_n \leq t\} - P\{S_{n+1} \leq t\}$.
- $P\{N(t) = n\} = F_n(t) - F_{n+1}(t)$.
- How do you obtain F_n from F ?

Convolution basics

- Convolution of two functions functions:
  $(f * g)(t) := \int_{-\infty}^{\infty} f(t-u)g(u)du$
- Convolution of two positive functions:
  $(f * g)(t) := \int_{0}^{t} f(t-u)g(u)du.$
- Convolution of a function w.r.t a distribution:
  $(f * G)(t) := \int_{-\infty}^{\infty} f(t-u)dG(u) = \int_{-\infty}^{\infty} f(t-u)g(u)du$

Convolution basics

- Convolution of distirbutions is also a distribution.
- If F and G are two distributions then $(F * G)(t) := \int_{-\infty}^{\infty} F(t-u)dG(u)$. $F^{(*2)}(t) = (F * F)(t)$.
- $P(S_n \leq t) := F_n(t)$. We will now express $F_n(t)$ as a convolution!
- $P(S_1 \leq t) = P(X_1 \leq t) = F(t)$ where $F(t)$ is the cdf of the interarrival time X_1.
- $P(S_2 \leq t) = P(X_1 + X_2 \leq t) = \int_0^t F(t-u)f(u)du$
- $= \int_0^t F(t-u)dF(u) = (F * F)(t)$.

$$F_n(t) = F^{(*n)}(t)$$

Laplace transform basics

- Bilateral Laplace transform of function $f(\cdot)$ is given by $\bar{f}(s) := \int_{-\infty}^{\infty} e^{-st} f(t) dt$.
- Consider a random variable X with distribution F. Then recall that $M_X(s) = E[e^{sX}] = \int e^{st} dF(t)$.
- The laplace transform of a distribution F is defined as $\bar{F}(s) = \int_{-\infty}^{\infty} e^{-st} dF(t) = E[e^{-sX}]$.Here X is a random variable with distribution F.
- Property: Consider $Z(t) = (f * F)(t)$. Then $\bar{Z}(s) = \bar{f}(s)\bar{F}(s)$
- This implies that if $Z(t) = (F * F)(t)$, then $\bar{Z}(s) = \bar{F}^2(s)$.
- Then by the same logic, $LT\{F^{(*n)}(t)\} = \bar{F}^n(s)$.

Renewal equation $m(t)$

Let $m(t)$ denote the mean number of arrivals by time t, i.e., $m(t) := E[N(t)]$. Then $m(t) = \sum_{n=1}^{\infty} F_n(t)$.

What is $m(t)$ for the Poisson process?

Let $\bar{m}(s), \bar{F}(s)$ and $\bar{F}_n(s)$ denote the Laplace transform of $m(t), F(t)$ and $F_n(t)$ respectively. Then $\bar{m}(s) = \frac{\bar{F}(s)}{1-\bar{F}(s)}$.

- $\bar{m}(s) = \bar{F}(s) + \bar{m}(s)\bar{F}(s)$. Inverse Laplace transform gives
- $m(t) = F(t) + (m * F)(t)$.

Renewal equation

- Renewal equation is an integral equation for $m(t)$ that is obtained by conditioning on time for first renewal.
- Suppose $X_1 = x$. Since this is the time interval between 0th and 1st arrival, $S_1 = x$ and the first arrival has happened at x.
- $m(t) = E[N(t)] = E_F[E[N(t)/X_1]]$.
- Therefore $m(t) = \int_0^\infty E[N(t)/X_1 = x]dF(x)$
- What if $t < x$? Then $E[N(t)/X_1 = x] = 0$.
- What happens when $t \geq x$?
- $E[N(t)/X_1 = x] = 1 + m(t-x)$.
- This gives us the renewal equation $m(t) = \int_0^t (1 + m(t-x))dF(x)$.

Stopping times

- Let $X_1, X_2, \ldots$ be a sequence of independent random variables.
- An integer valued positive random variable N is said to be a stopping time for this sequence if the event $\{N = n\}$ is independent of $X_{n+1}, X_{n+2}, \ldots$ for $n = 1, 2, \ldots$.
- N is not independent of the entire sequence $\{X_i\}$.
- Think as if we are seeing $X_n's$ one at a time and stop after a stopping criteria is met.
- So if we stop after seeing $X_1, X_2, \ldots, X_n$, then $N = n$.
- Suppose $P(X_n = 1) = P(X_n = -1) = 0.5$. Then $N = min\{n : X_1 + \ldots + X_n = 1\}$ is a stopping time.
- Stop one roll before you see 6. Is this a stopping time ?No.

Stopping times for Renewal process

- Is $N(t)$ a stopping time for the sequence of interarrivals X_i?
- Suppose $N(t) = n$, i.e., by time t there have been only n arrivals. Then what we know is that $S_n \leq t$ and that $S_{n+1} > t$.
- Therefore $N(t) = n$ depends on X_{n+1}. For it to be a stopping time, it should have been independent of X_{n+1}.
- Therefore $N(t)$ is not a stopping time.
- However $N(t) + 1$ is a stopping time. This is because $N(t) + 1 = n$ implies $N(t) = n - 1$ for which $S_{n-1} \leq t$ and that $S_n > t$.
- $N(t) + 1 = n$ depends on $X_1, \ldots X_n$ and is independent of $X_{n+1}, X_{n+2}, \ldots$

Wald's Equation

Theorem

If $X_1, X_2, \ldots,$ are independent and identically distributed random variables having finite expectations, and if N is a stopping time for $X_1, X_2, \ldots$ such that $E[N] < \infty$, then

$$E[\sum_{i=1}^{N} X_i] = ENEX$$

Proof on board. Also Refer Sheldon Ross, Thm 3.3.2.

Corollary

$E[S_{N(t)+1}] = E[X](m(t) + 1)$

Time average versus Ensemble average

$$\bar{X}^{time-avg} = \lim_{t\to\infty} \frac{\int_0^t X(u,\omega)du}{t}$$

$$\bar{X}^{ensemble} = \lim_{t\to\infty} E(X(t))$$

For an ergodic process, $\bar{X}^{time-avg} = \bar{X}^{ensemble}$

- Consider a Markov coin (with unknown transition probabilities) and given a budget of 10,00,000 (10 lakh) tosses, how will you find the stationary probability of head?
- Exhaust all at once (time average)
- Perform 100 runs each of length 10000 and average across the last toss in each run! (ensemble average)

Renewal theorem

Lemma

- *With probability 1,* $\frac{N(t)}{t} \to \frac{1}{E[X_1]}$ *as* $t \to \infty$.
  (Proof hint:- $S_{N(t)} \leq t \leq S_{N(t)+1}$*)*
- $\frac{m(t)}{t} \to \frac{1}{E[X_1]}$ *as* $t \to \infty$.

See Sheldon Ross (Stochastic Processes, 2nd edition) Proposition 3.3.1 and Thm 3.3.4 for proof.
NOTE: $S_{N(t)+1} > t$. Taking Expectations on both sides, invoking Wald's lemma, and rearranging gives us $\liminf_{t\to\infty} \frac{m(t)}{t} \geq \frac{1}{E[X_1]}$

Renewal Reward theorem

- Consider a renewal process with interarrival times $X_i, i = 1, 2, \ldots$. Suppose a random reward Y_i is earned at the time of the ith arrival. While Y_i may depend on X_i, the pairs (X_i, Y_i) are independent and identically distributed.

- Let $Y(t)$ denote the total reward accrued till time t. Then $Y(t) = \sum_{i=1}^{N(t)} Y_i$.

Lemma

- *With probability 1,* $\frac{Y(t)}{t} \to \frac{E[Y]}{E[X]}$ *as* $t \to \infty$.

- $\frac{E[Y(t)]}{t} \to \frac{E[Y]}{E[X]}$ *as* $t \to \infty$.

See Sheldon Ross Theorem 3.6.1 for proof.

__/__/__

For an ergodic process,

$\bar{X}_{\text{time-avg}} = \bar{X}_{\text{ensemble}}$.

Lecture #13: (Lecture Recording)

Renewal theorem:

N(t) is the counting process associated with the renewal process.

$$L(t) = \begin{cases} \infty & \text{where the event happened} \\ 0 & \text{otherwise} \end{cases}$$

$L(t) \sim$ dirac delta function

Dirac delta function $\delta(x)$

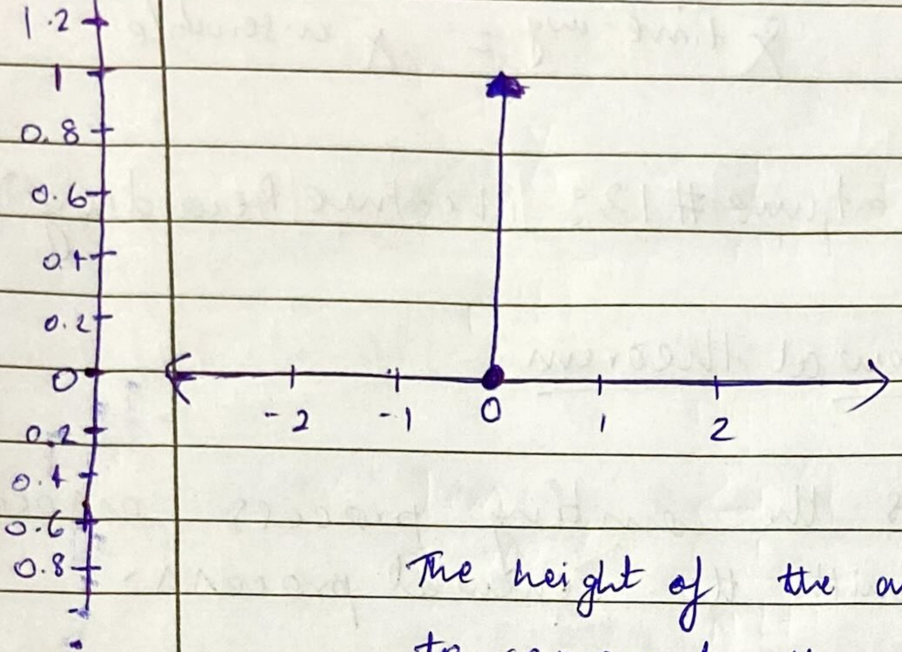
$$\delta(x) \triangleq \begin{cases} +\infty & x = 0 \\ 0 & x \neq 0 \end{cases}$$

and

$$\int_{-\infty}^{\infty} \delta(x)\,dx = 1$$

___/___/___

Schematic representation of dirac-delta

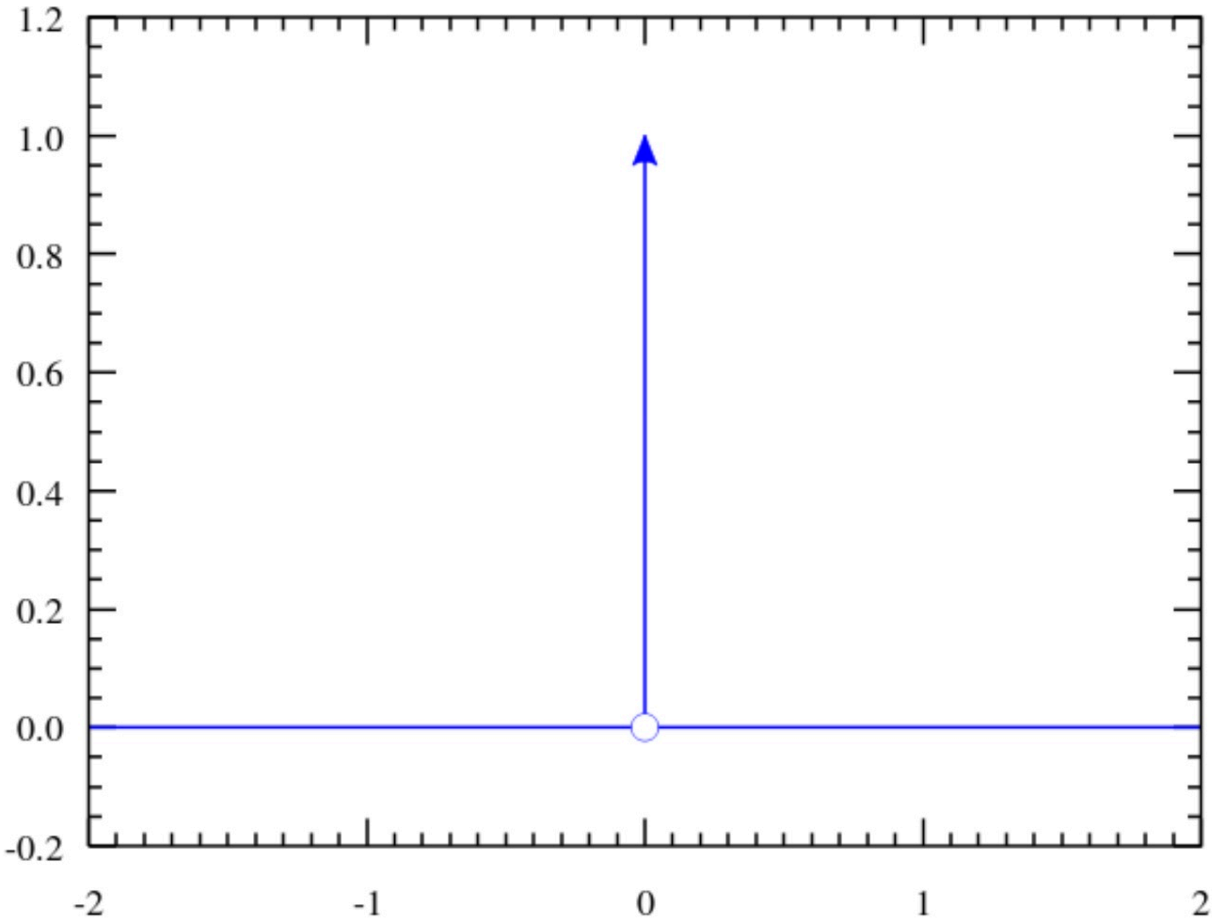


The height of the arrow is used to represent the area under the curve/function.

So $L(t)$ consists of ∞ many dirac delta functions summed & shifted from their origin ($t=0$) to other times where the events have happened.

$$\therefore \quad N(t) = \int_0^t L(t)\,dt.$$

$$\bar{X}^{\text{time-avg}} = \lim_{t\to\infty} \frac{N(t,\omega)}{t} = \lim_{t\to\infty} \frac{\int_0^t L(t,\omega)\,dt}{t}$$



Schematic representation of the Dirac delta by a line surmounted by an arrow. The height of the arrow is usually meant to specify the value of any multiplicative constant, which will give the area under the function. The other convention is to write the area next to the arrowhead.

Lemma :

$$\lim_{t \to \infty} \frac{N(t)}{t} = \frac{1}{E(X_1)} \quad \text{almost surely}$$

Meaning, $\displaystyle \lim_{t \to \infty} \frac{N(t, \omega)}{t} = \frac{1}{E(X_1)}$

$\forall \omega \in \Omega$ ~~where~~ such that $P(\omega) = 1$

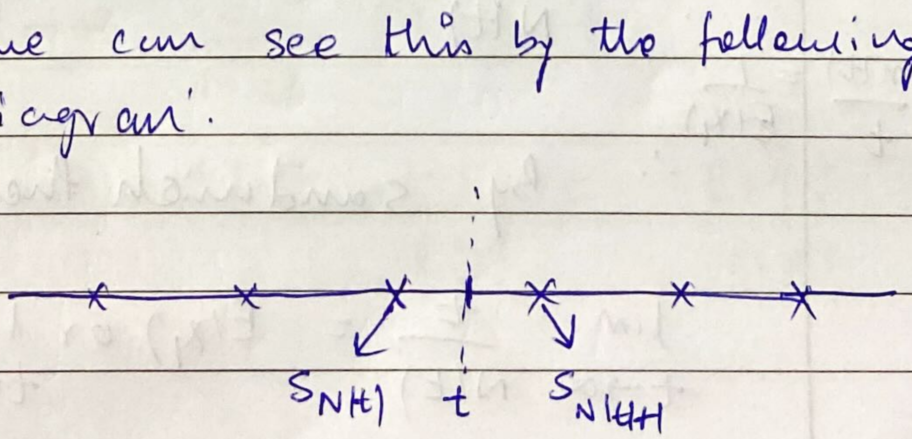
~~i.e where the time average does~~

i.e. the realization on which the time average does not equal to $\frac{1}{E(X_1)}$, the probability of that realization happening is 0.

Proof: Let S_n be the time of the n^{th} renewal.

Then $S_{N(t)} \leq t < S_{N(t)+1}$

we can see this by the following diagram:



__/__/__

$\therefore \quad \sum_{i=1}^{N(t)} s_i \leq t \leq \sum_{i=1}^{N(t)+1} s_i$

$$\Rightarrow \quad \frac{\sum_{i=1}^{N(t)} X_i}{N(t)} \leq \frac{t}{N(t)} \leq \frac{\sum_{i=1}^{N(t)+1} X_i}{N(t)+1} \cdot \frac{N(t)+1}{N(t)}$$

as $\lim_{t \to \infty}, \quad \frac{\sum_{i=1}^{N(t)} X_i}{N(t)} = E(X_i) = E(X_1)$

By strong Law of Large Numbers

and similarly,

$$\lim_{t \to \infty} \frac{\sum_{i=1}^{N(t)+1} X_i}{N(t)+1} = E(X_i) = E(X_1)$$

$$\frac{N(t)+1}{N(t)} \to 1 \text{ as } t \to \infty$$

$\lim_{t \to \infty} \frac{m(t)}{t} = \frac{1}{E(X_1)}$

$\therefore$ By sandwich theorem,

$$\lim_{t \to \infty} \frac{t}{N(t)} = E(X_1) \text{ or } \lim_{t \to \infty} \frac{N(t)}{t} = \frac{1}{E(X_1)}$$

Renewal Reward Theorem

→ Consider a renewal process with interarrival time X_i's and suppose a random reward Y_i is earned at the time of the i^{th} arrival.

Let $Y(t)$ be the total reward earned till time t. Then

$$Y(t) = \sum_{i=1}^{N(t)} Y_i$$

Lemma:

$$\lim_{t \to \infty} \frac{Y(t)}{t} = \frac{E(Y)}{E(X)} \text{ almost surely.}$$

Proof:

$$\lim_{t\to\infty} \frac{Y(t)}{t} = \lim_{t\to\infty} \frac{Y(t)}{N(t)} \cdot \frac{N(t)}{t}$$

$$\Rightarrow \lim_{t\to\infty} \frac{\sum_{i=1}^{N(t)} Y_i}{N(t)} \cdot \lim_{t\to\infty} \frac{N(t)}{t}$$

$$\Rightarrow E(Y) \cdot \frac{1}{E(X)} = \frac{E(Y)}{E(X)}.$$