Queuing Systems (Q.S.)

Q.S. Components:

Assume that a system gives (delivers) a certain service, then this system has 3 elements as follows:

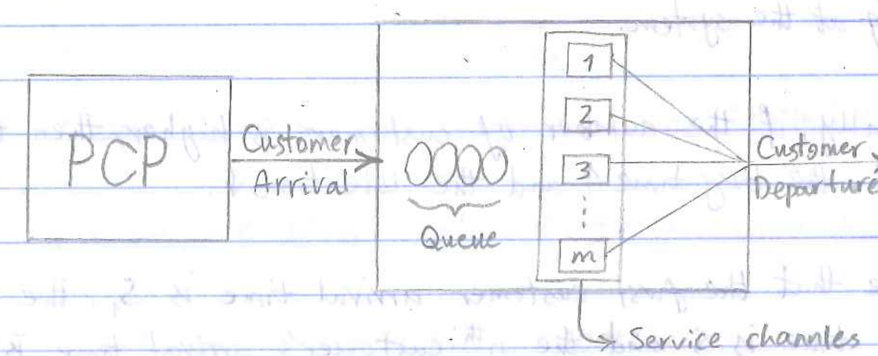
1) Customer: who get the service or request to be served.

2) Server: the machine/person that delivers the service.

* If the number of customers > the number of servers → a queue is formed.

* The number of servers is often called the number of service channles (almost always

3) Potential Customer Population: are the customers that may decide to come to the system.



Important measurements in a queue system:

In order to measure the performance of a system of queue

1) Queue Length: the number of customers waiting in the queue to be served.

* The other measurement in this field is the number of customers in the system including customers in the queue plus customers are served.

2) Waiting time of each customers in the queue or the system.

3) Idle time: the percentage of the time that the system is empty of customers

4) Busy time: the percentage of the time that the system is not empty of customers.

* Note: these measurements tend to have stochastic nature. It means they are random variables (r.vs), so we focuse on their expected value.

System inputs: a system's performance depends on its inputs which are as follows:

A) Customer arrival pattern: the system's performance, i.e. queue length, waiting time, idle time and busy time, depends on the number of customers arriving at the system.

Naturally, if the number of customers is higher, then the customer's waiting time ↑, the busy time ↑ and the idle time ↓.

Assume that the first customer arrival time is S_1, the second customer's arrival time is S_2 and the n^{th} customer's arrival time is S_n. Then the inter-arrival time between the successive arrivals are:

$$\begin{cases} t_1 = S_1 \\ t_2 = S_2 - S_1 \\ \vdots \\ t_n = S_n - S_{n-1} \end{cases}$$

In order to determine the customer arrival pattern, we need to c
$t_1, t_2, \dots, t_n$.

Since the arrival times are r.vs, then the inter-arrival times are
To study these r.vs, we need to study the distribution function of

If we show the CDF of the r.v (t) with $A(x)$, then we have:

$$A(X) = \mathbb{P}(t \leqslant x)$$

Note: a useful quantity for analysing the customers arrival beha
"Customers Arrival Rate", which is the number of customers o
the system per time unit. It is clear that the customers arri
is equal to the receprocal of the average inter-arrival time.

$$\lambda = \frac{1}{E(t)}$$

Arrival of customers can be "individual" or "bulk". For bulk ar
arrival by a bus) we often deal with the following parameters:

$\begin{cases} \text{1) Inter-arrival time between two groups.} \\ \text{2) The number of people in each group.} \end{cases}$

1) Homogeneity: if the customer arrival rate is the same at di
it is called "Homogeneous". (λ)

1) Queue Length: the number of customers waiting in the queue to be served.

* The other measurement in this field is the number of customers in the system including customers in the queue plus customers are served.

2) Waiting time of each customers in the queue or the system.

3) Idle time: the percentage of the time that the system is empty of customers.

4) Busy time: the percentage of the time that the system is not empty of customers.

* Note: these measurements tend to have stochastic nature. It means they are random variables (r.vs), so we focuse on their expected value.

System inputs: a system's performance depends on its inputs which are as follows:

A) Customer arrival pattern: the system's performance, i.e. queue length, waiting time, idle time and busy time, depends on the number of customers arriving at the system.

Naturally, if the number of customers is higher, then the customer's waiting time ↑, the busy time ↑ and the idle time ↓.

Assume that the first customer arrival time is S_1, the second customer's arrival time is S_2 and the n^{th} customer's arrival time is s_n. Then the inter-arrival time between the successive arrivals are:

$$\begin{cases} t_1 = S_1 \\ t_2 = S_2 - S_1 \\ \vdots \\ t_n = S_n - S_{n-1} \end{cases}$$

In order to determine the customer arrival pattern, we need to characterize $t_1, t_2, \dots, t_n$.

Since the arrival times are r.vs, then the inter-arrival times are also r.vs. To study these r.vs, we need to study the distribution function of these r.vs.

If we show the CDF of the r.v (t) with $A(x)$, then we have:

$$A(x) = \mathbb{P}(t \leqslant x)$$

Note: a useful quantity for analysing the customers arrival behavior is "Customers Arrival Rate", which is the number of customers arriving at the system per time unit. It is clear that the customers arrival rate (λ) is equal to the reciprocal of the average inter-arrival time.

$$\lambda = \frac{1}{E(t)}$$

Arrival of customers can be "individual" or "bulk". For bulk arrivals (e.g. arrival by a bus) we often deal with the following parameters:

$$\begin{cases} \text{1) Inter-arrival time between two groups.} \\ \text{2) The number of people in each group.} \end{cases}$$

1) Homogeneity: if the customer arrival rate is the same at different time it is called "Homogeneous". (λ)

2) Heterogeneity: If the customer arrival rate is not the same at different times, it is called "Heterogeneous". $\lambda(t)$

3) Doubly stochastic arrival rate: not only depends on time, but also is a r.v. (Λ_t) e.g. $\Lambda_t \sim \text{Poisson}(5t)$

Note: in some systems the arrival rate depends on the length of queue. For instance, if the queue is long, then the customers may decide not to join the queue or not come to the system.

B) Service pattern: the time that it takes for a customer to be served.

The lower the service time → the lower the queu's length → the lower the waiting time.

Similar to the arrival pattern, to study the service pattern we need to know the distribution of service time received by a customer.

Assume that the service time delivered to a customer is X. If we show the CDF of this r.v. with B(X), then we have:

$$B(X) = \mathbb{P}(X \leqslant x)$$

Service rate (μ): is the average number of customers served per time unit by one server. We have the following relationship between the service rat(μ) and the expected value of the service time.

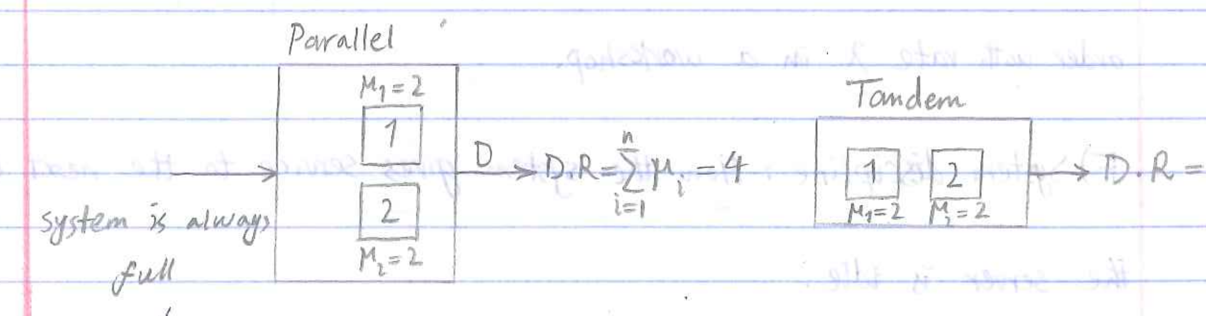
$$\mu = \frac{1}{E(X)}$$

Note: if μ is fixed over time, then it is called homogeneous. If μ on time, then it is called heterogeneous. $(\mu(t))$

Departure rate: is the average number of customers leaving the

The departure rate depends on the arrival rate, the number of cus the system and the service rate.

Note: if the system has only one server and the server is busy all then the departure rate is equal to the service rate.



Note: the service time can be dependent/independent to the queue's

C) The number of servers: these servers often perform in parallel. is called tandem.

2) Heterogeneity: If the customer arrival rate is not the same at different times, it is called "Heterogeneous". $\lambda(t)$

3) Doubly stochastic arrival rate: not only depends on time, but also is a r.v. (Λ_t). e.g. $\Lambda_t \sim \text{Poisson}(5t)$

Note: in some systems the arrival rate depends on the length of queue. For instance, if the queue is long, then the customers may decide not to join the queue or not come to the system.

B) Service pattern: the time that it takes for a customer to be served.

Therlower the service time → the lower the queu's length → the lower the waiting time.

Similar to the arrival pattern, to study the service pattern we need to know the distribution of service time received by a customer.

Assume that the service time delivered to a customer is X. If we show the CDF of this r.v. with B(X), then we have:

$$B(X) = \mathbb{P}(X \leqslant x)$$

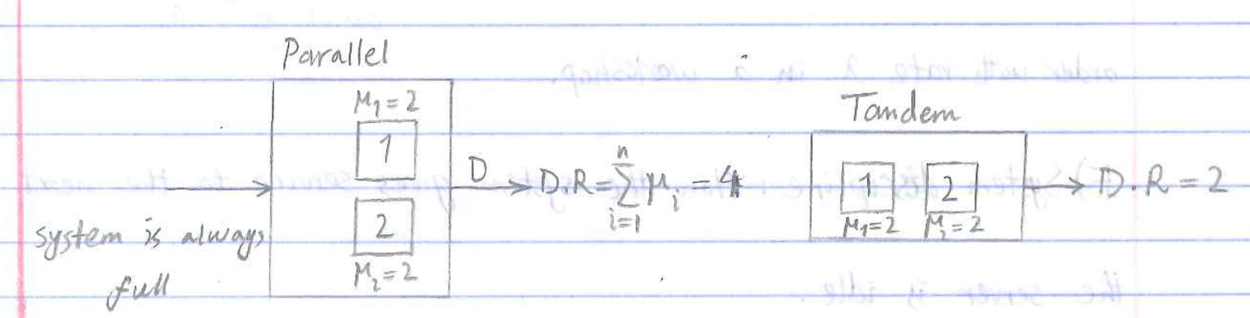
Service rate (μ): is the average number of customers served per time unit by one server. We have the following relationship between the service rat(μ) and the expected value of the service time.

$$\mu = \frac{1}{E(X)}$$

Note: if μ is fixed over time, then it is called homogeneous. If μ depends on time, then it is called heterogeneous. $(\mu(t))$

Departure rate: is the average number of customers leaving the system. The departure rate depends on the arrival rate, the number of customers in the system and the service rate.

Note: if the system has only one server and the server is busy all the time, then the departure rate is equal to the service rate.



Note: the service time can be dependent/independent to the queue's length.

C) The number of servers: these servers often perform in parallel. Otherwise is called tandem.

D) Queue's capacity: the maximum number of customers can wait in the queue. It can be either finite or infinite. For example there is a space constraint in banks or restaurant.

E) Potential Customer Population:

The length of the queue as well as the idle or busy time can depend on the potential customer population. For example, PCP of a call center is the population of the city. If the PCP is very large, we assume that it is infinite. However, it can be indeed finite, e.g. the number of machines that can become out of order with rate λ in a workshop.

F) System discipline: How the system gives service to the next customer when the server is idle.

1) FIFO : First-In-First-Out
2) LIFO : Last-In-First-Out e.g. mails on the desk that are waiting to be typed.
3) SIRO : Service-In-Random-Order e.g. selecting a spare part in a storage or warehouse.

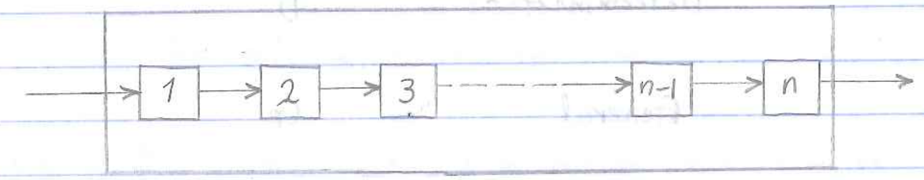
Priority of service: prioritize some groups of customers over others in delivering the service.

1) Pre-emptive: if the customer of higher priority arrives, the server interrupts the service to any customer that has a lower priority, e.g. emergency unit in hospital.

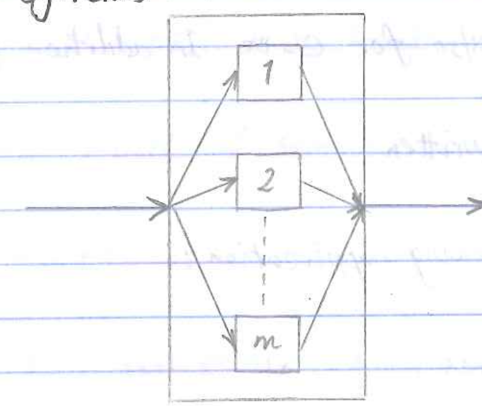
2) Non-emptive: the server continues to give the service to the lowe[r] Customer when a higher priority customer arrives. The higher priority is served after the lower priority customer.

G) Servic stages:

1) Series systems (tandem queue): e.g. production systems



2) Parallel systems:



Denoting Queuing Systems:

We use the Kendal notation: A/B/M/K/C/Z, where each of charaters refer to the one of the main element of the system in[…]

A / A(x) ≡ Inter-arrival time distribution of customers.
B / B(x) ≡ Servic time distribution
M ≡ is the number of servers
K ≡ is system capacity

D) Queue's capacity: the maximum number of customers can wait in the queue. It can be either finite or infinite. For example there is a space constraint in banks or restaurant.

E) Potential Customer Population:

The length of the queue as well as the idle or busy time can depend on the potential customer population. For example, PCP of a call center is the population of the city. If the PCP is very large, we assume that it is infinite. However, it can be indeed finite, e.g. the number of machines that can become out of order with rate λ in a workshop.

F) System discipline: How the system gives service to the next customer when the server is idle.

1) FIFO : First-In-First-Out
2) LIFO : Last-In-First-Out e.g. mails on the desk that are waiting to be typed.
3) SIRO : Service-In-Random-Order e.g. selecting a spare part in a storage or warehouse.

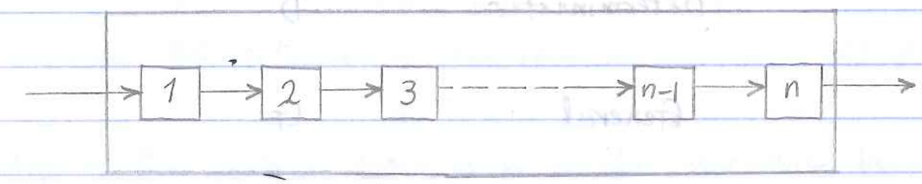
Priority of service: prioritize some groups of customers over others in delivering the service.

1) Pre-emptive: if the customer of higher priority arrives, the server interrupts the service to any customer that has a lower priority, e.g. emergency unit in hospital.

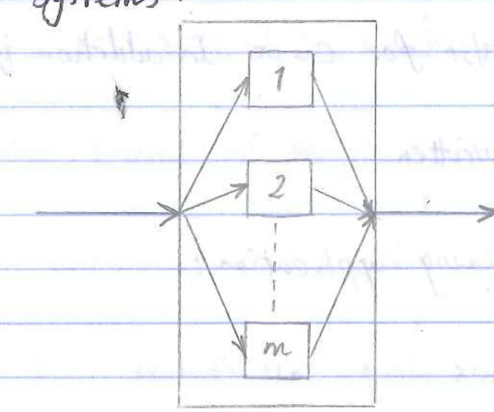
2) Non-emptive: the server continues to give the service to the lower priority customer when a higher priority customer arrives. The higher priority customer is served after the lower priority customer.

G) Servic stages:

1) Series systems (tandem queue): e.g. production systems



2) Parallel systems:



Denoting Queuing Systems:

We use the Kendal notation: A/B/M/K/C/Z, where each of the charaters refer to the one of the main element of the system input.

A / A(x) ≡ Inter-arrival time distribution of customers.
B / B(x) ≡ Servic time distribution
M ≡ is the number of servers
K ≡ is system capacity

$C \equiv$ the size of potential customers population

$Z \equiv$ system discipline

Note: In this convention we use the following notation instead of A and B.

Distribution	Code
Exponential	M
Erlang-r	Er
Deterministic	D
General	G

Note: If the system's capacity is infinite, then K is infinite ($K=\infty$), we do not write it, also for $c=\infty$. In addition if the system's discipline is FIFO, then it is not written.

Some examples of queuing application:

1) Communication systems, e.g. call centers
- Customer: Callers
- Server: Agents

2) Transportation
- Customer: people
- Servers: buses

3) Ports / Airports
- Customers: airplanes
- Servers: airport bands

4) Hospital
- Customers: patients
- Servers: surgery rooms

5) Posting systems
- Customers: mails
- Servers: mailman

6) Maintenance
- Customers: machines
- Servers: repairmans

Overview probability theory:

1) Sample Space: Consider an experiment or trial that we can pr
because the outcome is uncertain. The SS is a set of all possible c

e.g. if the service time is a random variable between 5 to 10 m
SS would be:

$$S = \{ X \mid 5 \leq X \leq 10 \}$$

2) Event: any subset of sample space. e.g. $A = \{5\}$, $B = [5,6]$ or

3) σ-algebra ($P(S)$ or $F(S)$): the set of all subsets of S.

4) Probability: $p : P(S) \sim [0,1]$

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

(2) $P(S) = 1$

(3) $P(E_1 \cup E_2) = P(E_1) + P(E_2)$, $E_1 \cap E_2 = \phi$

Note: modern definition of probability is $(p, S, P(S))$. It is calle
space.

Some conclusions:

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

$C \equiv$ the size of potential customers population
$Z \equiv$ system discipline

Note: In this convention we use the following notation instead of A and B.

Distribution	Code
Exponential	M
Erlang-r	Er
Deterministic	D
General	G

Note: If the system's capacity is infinite, then K is infinite ($K=\infty$), we do not write it, also for $C=\infty$. In addition if the system's discipline is FIFO, then it is not written.

Some examples of queuing application:

1) Communication systems, e.g. call centers
- Customer: Callers
- Server: Agents

2) Transportation
- Customer: people
- Servers: buses

3) Ports / Airports
- Customers: airplanes
- Servers: airport bands

4) Hospital
- Customers: patients
- Servers: surgery rooms

5) Posting systems
- Customers: mails
- Servers: mailman

6) Maintenance
- Customers: machines
- Servers: repairmans

Overview probability theory:

1) Sample Space: Consider an experiment or trial that we can predict, because the outcome is uncertain. The SS is a set of all possible outcomes.

e.g. if the service time is a random variable between 5 to 10 min. Then SS would be:

$$S = \{x \mid 5 \leqslant x \leqslant 10\}$$

2) Event: any subset of sample space. e.g. $A=\{5\}$, $B=[5,6]$ or $[9,10]$.

3) σ-algebra ($\mathcal{P}(s)$ or $F(s)$): the set of all subsets of S.

4) Probability: $P : \mathcal{P}(s) \sim [0,1]$

(1) $0 \leqslant P(E) \leqslant 1$
(2) $P(S) = 1$
(3) $P(E_1 \cup E_2) = P(E_1) + P(E_2)$, $E_1 \cap E_2 = \phi$

Note: modern definition of probability is $(P, S, \mathcal{P}(s))$. It is called probablity space.

Some conclusions:

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

(2) If $E_1 \subset E_2 \Rightarrow P(E_1) \leqslant P(E_2)$

(3) $P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$

5) Random variable: it is a function from sample space to $\mathbb{R}/\mathbb{N}/\mathbb{Z}$.

e.g. $S = \{H, T\} \rightarrow X: S \rightarrow \{0, 1\}$

6) CDF : $F_X(a) = \mathbb{P}(X \leqslant a)$

properties : (1) $a \leqslant b \Rightarrow F(a) \leqslant F(b)$, it means that F is increasing.

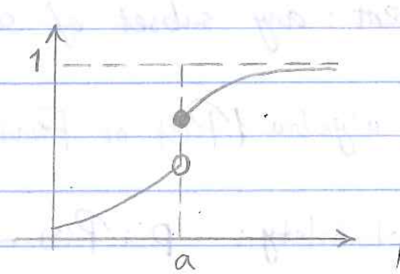
(2) $\lim_{x \to -\infty} F_X(x) = 0$, $\lim_{x \to \infty} F_X(x) = 1$

(3) $\mathbb{P}(a \leqslant x \leqslant b) = F(b) - F(a)$

(4) If F is not continuous at any point, i.e. $X = x$, then:

$$F(x) = \lim_{y \to 0^+} F(x+y)$$

$$\lim_{y \to 0^+} F(x-y) < F(x)$$



F is said to be right continuous with left limit. (RCLL or Cadlag).

Discrete or Continuous : If a variable takes a value from a countable set, then it is called discrete random variable.

e.g. $\mathbb{Z} = \{0, \pm 1, \pm 2, \dots\}$ and $\mathbb{Q} = \left\{ \frac{a}{b} \mid a, b \in \mathbb{Z}, b \neq 0 \right\}$ are countable sets because we can count them and correspond them to natural numbers.

Probability functions

(1) $\mathbb{P}(a) = \mathbb{P}(X = a)$

(2) $F(x) = \mathbb{P}(X \leqslant x) = \sum_{a=-\infty}^{x} \mathbb{P}(a)$

(3) $\sum_{-\infty}^{+\infty} \mathbb{P}(a) = 1$

Bernuli Distribution :

$$\mathbb{P}(a) = \begin{cases} p & ; a = 1 \quad \text{success} \\ 1-p & ; a = 0 \quad \text{failure} \end{cases}$$

Binomial Distribution : n Bernuli trials (each trial can resu success or failure)

Assume that x is the number of success, then we have:

$$P(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad ; \quad \binom{n}{x} = \frac{n!}{x!\,(n-x)!}$$

Geometric Distribution : Assume a number of Bernuli ex that are conducted independently. The number of trials x unt success is :

$$P(x) = p(1-p)^{x-1} \quad ; \quad x = 1, 2, \dots$$

Poisson Distribution :

$$P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}$$

(2) If $E_1 \subset E_2 \Rightarrow P(E_1) \leqslant P(E_2)$

(3) $P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$

5) Random variable: it is a function from sample space to $\mathbb{R}/\mathbb{N}/\mathbb{Z}$.

e.g. $S=\{H,T\} \rightarrow X: S \rightarrow \{0,1\}$

6) CDF: $F_X(a) = P(X \leqslant a)$

properties: (1) $a \leqslant b \Rightarrow F(a) \leqslant F(b)$, it means that F is increasing.

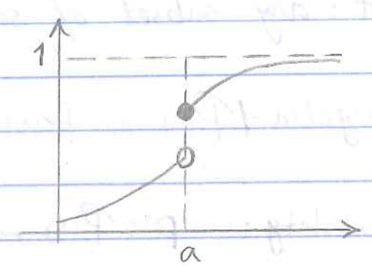
(2) $\lim_{x \to -\infty} F_X(x) = 0$, $\lim_{x \to \infty} F_X(x) = 1$

(3) $P(a \leqslant X \leqslant b) = F(b) - F(a)$

(4) If F is not continuous at any point, i.e. $X=x$, then:

$F(x) = \lim_{y \to 0^+} F(x+y)$

$\lim_{y \to 0^+} F(x-y) < F(x)$



F is said to be right continuous with left limit. (RCLL or Cadlag).

Discrete or Continuous: If a variable takes a value from a countable set, then it is called discrete random variable.

e.g. $\mathbb{Z}=\{0, \pm1, \pm2, \dots\}$ and $\mathbb{Q}=\left\{\frac{a}{b} \mid a,b \in \mathbb{Z}, b \neq 0\right\}$ are countable sets because we can count them and correspond them to natural numbers.

Probability functions:

(1) $\mathbb{P}(a) = \mathbb{P}(X=a)$

(2) $F(x) = \mathbb{P}(X \leqslant x) = \sum_{a=-\infty}^{x} \mathbb{P}(a)$

(3) $\sum_{-\infty}^{+\infty} \mathbb{P}(a) = 1$

Bernuli Distribution:

$$\mathbb{P}(a) = \begin{cases} p & ; a=1 \quad \text{success} \\ 1-p & ; a=0 \quad \text{failure} \end{cases}$$

Binomial Distribution: n Bernuli trials (each trial can result in a success or failure)

Assume that x is the number of success, then we have:

$$P(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad ; \quad \binom{n}{x} = \frac{n!}{x!(n-x)!}$$

Geometric Distribution: Assume a number of Bernuli experiment that are conducted independently. The number of trials x until the first success is:

$$P(x) = p(1-p)^{x-1} \quad ; \quad x=1,2,\dots$$

Poisson Distribution:

$$P(X=x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

Continuous random variable: if a r.v takes a value from non-countable sets, then it is called continuous r.v. e.g. servic time in a bank.

Note: every r.v is characterized by its CDF, $F(x)$, PDF, $f(x)$. the PDF is defined as:

$$f(x)=\frac{d}{dx}F(x)$$

Similarly:

(1) $F(x)=\mathbb{P}(X\leqslant x)=\int_{-\infty}^{x} f(x)\,dx \equiv \int_{-\infty}^{x} dF(x)$

(2) $\mathbb{P}(a'\leqslant X\leqslant b)=F(b)-F(a)=\int_a^b f(x)\,dx$

(3) $\mathbb{P}(X=a)=\int_a^a f(x)\,dx=0 \longrightarrow \mathbb{P}(X\leqslant a)=\mathbb{P}(X<a)$

Expected value:

$$\begin{cases} E(X)=\sum_{i=1}^{n} X_i\,P(X_i) & ;\text{ if } X \text{ is discrete} \\ E(X)=\int_{-\infty}^{+\infty} x\,f(x)\,dx & ;\text{ if } X \text{ is continuous} \end{cases}$$

Theorem: (1) $E(aX+b)=a\,E(X)+b$

(2) $E(aX+bY)=a\,E(X)+b\,E(Y)$

(3)* $E(g(x))=\sum_{x=1}^{\infty} g(x)\,P(g(x))$, $E(g(x))=\int_{-\infty}^{+\infty} g(x)\,f(g(x))\,dx$

* Unaware Statistician:

$$\begin{cases} E(g(x))=\sum_{x=1}^{\infty} g(x)\,P(g(x))=\sum_{x=1}^{\infty} g(x)\,P(x) \\ E(g(x))=\int_{-\infty}^{+\infty} g(x)\,f(g(x))\,dx=\int_{-\infty}^{+\infty} g(x)\,f(x)\,dx \end{cases}$$

Variance:

$$Var(X)=E\left[(X-E(X))^2\right]=E(X^2)-E^2(X)$$

Joint Distribution Function:

$$F(x,y)=\mathbb{P}(X\leqslant x,\ Y\leqslant y)$$

Marginal distribution
$$\begin{cases} F_X(x)=\mathbb{P}(X\leqslant x,\ Y\leqslant \infty) \\ F_X(x)=\lim_{Y\to\infty} F(X,Y)=F(X,\infty) \end{cases}$$

Note: if $(x,y)\in C^2$, where C^2 is a continuous set, then we ha

(1) $\mathbb{P}\left\{(x,y)\in C\right\}=\iint_{(x,y)\in C} f(x,y)\,dx\,dy$

(2) $F(a,b)=\mathbb{P}(X\leqslant a,\ Y\leqslant b) \Rightarrow f(a,b)=\frac{\partial^2}{\partial x\,\partial y}F(a,b$

(3) $f_X(x)=\int_{-\infty}^{+\infty} f(x,y)\,dy$

Example. Joint distribution function of two r.v X and Y is as fo

$$f(x,y)=\begin{cases} 4y(x-y)e^{-(x+y)} & ;\ 0<x<\infty,\ 0\leqslant y\leqslant x \\ 0 & ;\ o.w \end{cases}$$

what is $F_Y(y)=?$

$$F_Y(y)=\int_{x} f(x,y)\,dx=\int_0^{+\infty} f(x,y)\,dx=\int_y^{+\infty} 4y(x-y)e^{-(x+y)}\,dx$$

$$=4y\left[\int_y^{+\infty} x e^{-(x+y)}\,dx-\int_y^{+\infty} y e^{-(x+y)}\,dx\right]=4y\left[-xe^{-(x+y)}-e^{-(x+y)}\right.+$$

Continuous random variable : if a r.v takes a value from non-countable sets, then it is called continuous r.v.. e.g. service time in a bank.

Note: every r.v is characterized by its CDF, $F(x)$, PDF, $f(x)$. the PDF is defined as :

$$f(x) = \frac{d}{dx} F(x)$$

Similarly : (1) $F(x) = \mathbb{P}(X \leqslant x) = \int_{-\infty}^{x} f(x)dx \equiv \int_{-\infty}^{x} dF(x)$

(2) $\mathbb{P}(a < X \leqslant b) = F(b) - F(a) = \int_a^b f(x)\,dx$

(3) $\mathbb{P}(X = a) = \int_a^a f(x)dx = 0 \longrightarrow \mathbb{P}(X \leqslant a) = \mathbb{P}(X < a)$

Expected value :

$$\begin{cases} E(X) = \sum_{i=1}^{n} x_i P(x_i) & ; \text{if } X \text{ is discrete} \\ E(X) = \int_{-\infty}^{+\infty} x f(x)\,dx & ; \text{if } X \text{ is continuous} \end{cases}$$

Theorem : (1) $E(aX+b) = aE(X) + b$

(2) $E(aX+bY) = aE(X) + bE(Y)$

(3)* $E(g(x)) = \sum_{x=1}^{\infty} g(x) P(g(x))$, $E(g(x)) = \int_{-\infty}^{+\infty} g(x) f(g(x))\,dx$

* Unaware statistician :

$$\begin{cases} E(g(x)) = \sum_{x=1}^{\infty} g(x) P(g(x)) = \sum_{x=1}^{\infty} g(x) P(x) \\ E(g(x)) = \int^{+\infty} g(x) f(g(x))dx = \int^{+\infty} g(x) f(x) dx \end{cases}$$

Variance :

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

Joint Distribution Function :

$$F(x,y) = \mathbb{P}(X \leqslant x, Y \leqslant y)$$

Marginal distribution
$$\begin{cases} F_X(x) = \mathbb{P}(X \leqslant x, Y \leqslant \infty) \\ F_X(x) = \lim_{Y \to \infty} F(X,Y) = F(X,\infty) \end{cases}$$

Note : if $(x,y) \in G^2$, where G^2 is a continuous set, then we have :

(1) $\mathbb{P}\{(x,y) \in C\} = \iint_{(x,y)\in C} f(x,y)\,dx\,dy$

(2) $F(a,b) = \mathbb{P}(X \leqslant a, Y \leqslant b) \Rightarrow f(a,b) = \frac{\partial^2}{\partial x \partial y} F(a,b)$

(3) $f_X(x) = \int_{-\infty}^{+\infty} f(x,y)\,dy$

Example. Joint distribution function of two r.v X and Y is as follows :

$$f(x,y) = \begin{cases} 4y(x-y)e^{-(x+y)} & ; 0 < x < \infty \ , \ 0 \leqslant y \leqslant x \\ 0 & ; \text{o.w} \end{cases}$$

what is $F_Y(y) = ?$

$$F_Y(y) = \int_x f(x,y)\,dx = \int_0^{+\infty} f(x,y)\,dx = \int_y^{+\infty} 4y(x-y)e^{-(x+y)}\,dx$$

$$= 4y\left[\int_y^{+\infty} x e^{-(x+y)}dx - \int_y^{+\infty} y e^{-(x+y)}dx\right] = 4y\left[-xe^{-(x+y)} - e^{-(x+y)} + ye^{-(x+y)}\right]_y^{\infty}$$

$= 4y e^{-2y}$

Note : The following technique is called repeated integral by parts.

$$\int f(x)\cdot g(x) = f'(x)\int g(x) - f''(x)\iint g(x) + \cdots + 0$$

Independent variables and events :

The two event A and B are independent if and only if,

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\cdot\mathbb{P}(B)$$

The two random variables X and Y are independent if and only for all sets A and B,

$$\mathbb{P}\{X\in A, Y\in B\} = \mathbb{P}\{X\in A\}\cdot\mathbb{P}\{Y\in B\}$$

Factorization Theorem : from the above relation we have,

$$F(x,y) = F_X(x)\cdot F_Y(y)$$

$$\hookrightarrow \mathbb{P}(X\leq x, Y\leq y) = \mathbb{P}(X\leq x, Y\leq y) = F_X(x)\cdot F_Y(y)$$

$$\mathbb{P}(x,y) = \mathbb{P}_X(x)\cdot\mathbb{P}_Y(y)$$

$$f(x,y) = f_X(x)\cdot f_Y(y)$$

$$^*E(XY) = E(X)\cdot E(Y) \qquad (*)$$

* If X and Y are independent, then (*) holds but if (*) holds, it does not mean independency. The exception ; X, Y ~ Normal, then (*) implies independency.

Example. in a party n pop throw their hats to the middle of t

and each person takes one of the hats with closed eyes. What is

number of people who takes their own hats? –

$$X_i = \begin{cases} 1 & ; \text{ if person } i \text{ takes his own hat.} \\ 0 & ; \text{ o.w} \end{cases}$$

$$X = X_1 + X_2 + \cdots + X_n \longrightarrow E(X) = E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i) = n\left(\frac{1}{n}\right) = 1$$

Hence, the average number of people who takes their own hats does not de

number of people in the room.

Covariance

$$\mathrm{Cov}(X,Y) = E\Big[(X - E(X))\cdot(Y - E(Y))\Big] = E(XY) - E(X)\cdot E(Y)$$

If X and Y are independent, then $\mathrm{Cov}(X,Y) = 0$, and it is called u

Interpretation of covariance

Assume that $X = \begin{cases} 1 & \text{, if event 1 happens} \\ 0 & \text{, o.w} \end{cases}$ and $Y = \begin{cases} 1 & \text{, if event} \\ 0 & \text{, o.w} \end{cases}$

So, $E(X) = \mathbb{P}(E_1)$ and $E(Y) = \mathbb{P}(E_2)$.

$$XY = \begin{cases} 1 & \text{, if } E_1 \text{ and } E_2 \text{ happen together.} \\ 0 & \text{, o.w} \end{cases} \qquad \Longrightarrow E(XY) = \mathbb{P}(E_1 \cap$$

we a

$$\mathrm{Cov}(X,Y) = E(XY) - E(X)\cdot E(Y) = \mathbb{P}(E_1 \cap E_2) - \mathbb{P}(E_1)\cdot\mathbb{P}(E_2) \geqslant 0$$

$$\Longrightarrow \mathbb{P}(E_1) \leqslant \frac{\mathbb{P}(E_1 \cap E_2)}{\mathbb{P}(E_2)} = \mathbb{P}(E_1 \mid E_2) \quad \text{e.g. } \mathbb{P}(X=1) \leqslant \mathbb{P}(X=1 \mid Y=1)$$

$= 4y e^{-2y}$

Note : The following technique is called repeated integral by parts.

$$\int f(x)\cdot g(x) = f'(x)\int g(x) - f''(x)\iint g(x) + \cdots + 0$$

Independent variables and events :

The two event A and B are independent if and only if,

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\cdot \mathbb{P}(B)$$

The two random variables X and Y are independent if and only for all sets A and B,

$$\mathbb{P}\{X\in A, Y\in B\} = \mathbb{P}\{X\in A\}\cdot \mathbb{P}\{Y\in B\}$$

Factorization Theorem : from the above relation we have,

$$F(x,y) = F_X(x)\cdot F_Y(y)$$

$$\rightarrow \mathbb{P}(X\leq x, Y\leq y) = \mathbb{P}(X\leq x, Y\leq y) = F_X(x)\cdot F_Y(y)$$

$$\mathbb{P}(x,y) = \mathbb{P}_X(x)\cdot \mathbb{P}_Y(y)$$

$$\rightarrow f(x,y) = f_X(x)\cdot f_Y(y)$$

$$^{*}E(XY) = E(X)\cdot E(Y) \qquad (*)$$

* If x and y are independent, then (*) holds but if (*) holds, it does not mean independency. The exception ; $X, Y \sim$ Normal, then (*) implies independency.

Example. in a party n pop throw their hats to the middle of the room and each person takes one of the hats with closed eyes. What is the averge number of people who takes their own hats?

$$X_i = \begin{cases} 1 & \text{; if person } i \text{ takes his own hat.} \\ 0 & \text{, o.w} \end{cases}$$

$$X = X_1 + X_2 + \cdots + X_n \longrightarrow E(X) = E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i) = n\left(\frac{1}{n}\right) = 1$$

Hence, the average number of people who takes their own hats does not depend on the number of people in the room.

Covariance

$$Cov(X,Y) = E\left[(X-E(X))\cdot(Y-E(Y))\right] = E(XY) - E(X)\cdot E(Y)$$

If x and y are independent, then $Cov(X,Y) = 0$, and it is called uncorrelated

Interpretation of covariance

Assume that $X = \begin{cases} 1 & \text{, if event 1 happens} \\ 0 & \text{, o.w} \end{cases}$ and $Y = \begin{cases} 1 & \text{, if event 2 happens} \\ 0 & \text{, o.w} \end{cases}$

So, $E(X) = \mathbb{P}(E_1)$ and $E(Y) = \mathbb{P}(E_2)$.

$$XY = \begin{cases} 1 & \text{, if } E_1 \text{ and } E_2 \text{ happen together.} \\ 0 & \text{, o.w} \end{cases} \implies E(XY) = \mathbb{P}(E_1\cap E_2)$$

we assume that

$$Cov(X,Y) = E(XY) - E(X)\cdot E(Y) = \mathbb{P}(E_1\cap E_2) - \mathbb{P}(E_1)\cdot\mathbb{P}(E_2) \geq 0$$

$$\implies \mathbb{P}(E_1) \leq \frac{\mathbb{P}(E_1\cap E_2)}{\mathbb{P}(E_2)} = \mathbb{P}(E_1|E_2) \qquad \text{e.g. } \mathbb{P}(X=1) \leq \mathbb{P}(X=1|Y=1)$$

Var (X) :

$$Var(X) = Cov(X, X) = E(X^2) - E^2(X)$$

$$Cov(aX+bY, cX+DY) = ac\, Cov(X,X) + ad\, Cov(X,Y) + bc\, Cov(X,Y) + bd\, Cov(Y,Y) = ac\, Var(X) + bd\, Var(Y) + (ad+bc)\, Cov(X,Y)$$

$$Var(X+Y) = Cov(X+Y, X+Y) = Var(X) + Var(Y) + 2\, Cov(X,Y)$$

Conditional Probability :

$$\mathbb{P}(E|F) = \frac{\mathbb{P}(E \cap F)}{\mathbb{P}(F)}$$

Example. two customers have entered the system. If $p = 0.6$ is the probability that a customer is male, what is the probability that both customers are male, given the following information.

(1) No additional information. $\mathbb{P} = 0.6 \times 0.6 = 0.36$

(2) If the first customer is male. ~~at least on of the~~

E : The event that both customers are male.

F : The event that atleast one of customers is male.

$$P(E|F) = \frac{P(E \cap F)}{P(F)} = \frac{P(E)}{P(F)} = \frac{0.6 \times 0.6}{1 - P(F^c)} = \frac{0.36}{0.84^*} = 0.43$$

$* \; 0.6 \times 0.4 + 0.4 \times 0.6 + 0.6 \times 0.6 = 0.84$

Note : if events are defined based on r.vs, then we have the following relationship. $P(X=x|Y=y) = \dfrac{P(X=x, Y=y)}{P(Y=y)}$

* If X and Y are independent then $P(X=x, Y=y) = P(X=x) \cdot P(Y=y)$, if not we must have joint distribution of X and Y.

Example. The number of customers entering the system in the first hour X and the number of customers entering in the second hour is a r.v Y. that X & Y are following Poisson distribution with rate λ_1 & λ_2, res Find the distribution function for X if we know that at the end of hour the total number of customers is n, $X+Y=n$.

$$\mathbb{P}(X=k | X+Y=n) = \frac{\mathbb{P}(X=k, Y=n-k)}{\mathbb{P}(X+Y=n)} = \frac{\mathbb{P}(X=k) \cdot \mathbb{P}(Y=n-k)}{\mathbb{P}(X+Y=n)}$$

$$= \frac{\left(\frac{e^{-\lambda_1} \lambda^k}{k!}\right) \cdot \left(\frac{e^{-\lambda_2} \lambda_2^{(n-k)}}{(n-k)!}\right)}{\frac{e^{-(\lambda_1+\lambda_2)} (\lambda_1+\lambda_2)^n}{n!}} = \frac{n!}{k!\,(n-k)!} \underbrace{\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^k}_{\mathbb{P}} \underbrace{\left(\frac{\lambda_2}{\lambda_1+ \ldots}\right.}_{(1- \ldots}$$

$$\rightarrow (X | X+Y=n) \sim \text{Binomial}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}, n\right)$$

Conditional Expected Value :

The expected value of X given that another r.v Y which takes a cert y, is represented by $E(X|Y=y)$ and defined as :

$$\begin{cases} E(X|Y=y) = \sum_x x\, \mathbb{P}(X=x|Y=y) & ; x \text{ is} \\ E(X|Y=y) = \int_{-\infty}^{+\infty} x \cdot f(X|Y=y) = \dfrac{\int_{-\infty}^{+\infty} x f(x,y)\,dx}{\int_{-\infty}^{+\infty} f(x,y)\,dx} & ; x \text{ is} \end{cases}$$

Application of conditional expected value:

If obtaining E(X) directly is difficult, then we may be able to obtain another r.v. Y and the following theorem.

Var (X) :

$$Var(X) = Cov(X,X) = E(X^2) - E^2(X)$$

$$Cov(aX+bY,\ cX+DY) = ac\ Cov(X,X) + ad\ Cov(X,Y) + bc\ Cov(X,Y) + bd\ Cov(Y,Y) = ac\ Var(X) + bd\ Var(Y) + (ad+bc)\ Cov(X,Y)$$

$$Var(X+Y) = Cov(X+Y, X+Y) = Var(X) + Var(Y) + 2\ Cov(X,Y)$$

Conditional Probability :

$$\mathbb{P}(E|F) = \frac{\mathbb{P}(E \cap F)}{\mathbb{P}(F)}$$

Example. two customers have entered the system. If $P=0.6$ is the probability that a customer is male, what is the probability that both customers are male, given the following information.

(1) No additional information. $P = 0.6 \times 0.6 = 0.36$

(2) If the first customer is male. at least on of the

E : The event that both customers are male.
F : The event that atleast one of customers is male.

$$P(E|F) = \frac{P(E \cap F)}{P(F)} = \frac{P(E)}{P(F)} = \frac{0.6 \times 0.6}{1 - P(F^c)} = \frac{0.36}{0.84^*} = 0.43$$

$* \ 0.6 \times 0.4 + 0.4 \times 0.6 + 0.6 \times 0.6 = 0.84$

Note : if events are defined based on r.vs, then we have the following relationship.

$$P(X=x \mid Y=y) = \frac{P(X=x, Y=y)}{P(Y=y)}$$

* If X and Y are independent then $P(X=x, Y=y) = P(X=x) \cdot P(Y=y)$, if not we must have joint distribution of X and Y.

Example. The number of customers entering the system in the first hour is a r.v X and the number of customers entering in the second hour is a r.v Y. Assume that X & Y are following Poisson distribution with rate λ_1 & λ_2, respectively. Find the distribution function for X if we know that at the end of second hour the total number of customers is n, $X+Y=n$.

$$\mathbb{P}(X=k \mid X+Y=n) = \frac{\mathbb{P}(X=k, Y=n-k)}{\mathbb{P}(X+Y=n)} = \frac{\mathbb{P}(X=k)\cdot\mathbb{P}(Y=n-k)}{\mathbb{P}(X+Y=n)}$$

$$= \frac{\left(\frac{e^{-\lambda_1}\lambda^k}{k!}\right)\cdot\left(\frac{e^{-\lambda_2}\lambda_2^{(n-k)}}{(n-k)!}\right)}{\frac{e^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^n}{n!}} = \frac{n!}{k!\,(n-k)!}\underbrace{\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)}_{\mathbb{P}}^{k}\underbrace{\left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)}_{(1-\mathbb{P})}^{n-k}$$

$$\rightarrow (X \mid X+Y=n) \sim \text{Binomial}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}, n\right)$$

Conditional Expected Value :

The expected value of x given that another r.v Y which takes a certain number y, is represented by $E(X|Y=y)$ and defined as :

$$\begin{cases} E(X|Y=y) = \sum_x x\ \mathbb{P}(X=x|Y=y) & ;\ x \text{ is discrete} \\ E(X|Y=y) = \int_{-\infty}^{+\infty} x \cdot f(x|y=y) = \dfrac{\int_{-\infty}^{+\infty} x f(x,y)\,dx}{\int_{-\infty}^{+\infty} f(x,y)\,dx} & ;\ x \text{ is continuous} \end{cases}$$

Application of conditional expected value:

If obtaining E(X) directly is difficult, then we may be able to obtain it using another r.v. Y and the following theorem.

Theorem: (Tower Property or Law of repeated expectation)

$$E(X) = E_Y\left[E_X(X|Y)\right] \rightarrow \begin{cases} E(X) = \sum_y E(X|Y=y)\,\mathbb{P}(Y=y) & ; Y \text{ is continuous} \\ E(X) = \int_{-\infty}^{+\infty} E(X|Y=y)\, f_Y(y)\, dy & ; Y \text{ is continuous} \end{cases}$$

Proof. X and Y can be either Discrete or Continuous, Then, 4 possible situation can happen. We prove the theorem X and Y are both discrete.

$$\sum_y E(X|Y=y)\,\mathbb{P}(Y=y) = \sum_y \sum_x x \cdot \mathbb{P}(X=x|Y=y) \cdot \mathbb{P}(Y=y)$$

$$= \sum_y \sum_x x \, \frac{\mathbb{P}(X=x, Y=y)}{\cancel{\mathbb{P}(Y=y)}} \cancel{\mathbb{P}(Y=y)}$$

$$= \sum_y \sum_x x \, \mathbb{P}(X=x, Y=y)$$

$$= \sum_x x \sum_y \mathbb{P}(X=x, Y=y) = \sum_x x \, \mathbb{P}(X=x)$$

$$= E(X)$$

Example. Assume that the chance of a coin coming Head is P. We toss the coin until it comes Head for the first time. What is the expected value of the number of times we toss the coin? Geometric Distribution

N = total number of trials until the first head comes.

E(N) = ?

$$Y = \begin{cases} 1 & ; \text{if the coin comes head in the first toss} \\ 0 & , \text{o.w} \end{cases}$$

$$E(N) = E\left[E(N|Y)\right] = \sum_y E(N|Y=y)\,\mathbb{P}(Y=y)$$

$$= E(N|Y=1)\,\mathbb{P}(Y=1) + E(N|Y=0)\,\mathbb{P}(Y=0)$$

$$= 1 \times P + (1 + E(N)) \times (1-P) \Rightarrow E(N) = \frac{1}{P}$$

Note: Actually Tower Property is applied when we tend to calculate e

value of a random variable. In this example being H or T is a ra

variable and also the number of times we toss to achieve the first s

a random variable. (expectation of the sum of a random number of rando

Application of conditional probability in calculating the probability of a

Theorem:

$$\begin{cases} \mathbb{P}(A) = \sum_y \mathbb{P}(A|Y=y) \cdot \mathbb{P}(Y=y) & ; Y \text{ is discret} \\ \mathbb{P}(A) = \int_y \mathbb{P}(A|Y=y)\, f(Y=y) & ; Y \text{ is continu} \end{cases}$$

Proof. define the r.v X as: $X = \begin{cases} 1 & ; \text{if the event A happens} \\ 0 & ; \text{o.w} \end{cases}$

So $E(X) = \mathbb{P}(A)$ and $E(X|Y=y) = \sum_x x\, \mathbb{P}(X=x|Y=y) = \mathbb{P}(X=\overset{A}{1}|Y=y$

also, $\mathbb{P}(A) = E(X) = \sum_y E(X|Y=y) \cdot \mathbb{P}(Y=y)$

$\rightarrow \mathbb{P}(A) = \sum_y \mathbb{P}(A|Y=y)\,\mathbb{P}(Y=y)$

Theorem: (Tower Property or Law of repeated expectation)

$$E(X) = E_y\left[E_x(X|Y)\right] \rightarrow \begin{cases} E(X) = \sum_y E(X|Y=y)\,\mathbb{P}(Y=y) & ; Y \text{ is continuous} \\ E(X) = \int_{-\infty}^{+\infty} E(X|Y=y)\, f_Y(y)\, dy & ; Y \text{ is continuous} \end{cases}$$

Proof. X and Y can be either Discrete or Continuous, Then, 4 possible situation can happen. We prove the theorem x and y are both discrete.

$$\sum_y E(X|Y=y)\,\mathbb{P}(Y=y) = \sum_y \sum_x x \cdot \mathbb{P}(X=x|Y=y)\cdot \mathbb{P}(Y=y)$$

$$= \sum_y \sum_x x \frac{\mathbb{P}(X=x, Y=y)}{\cancel{\mathbb{P}(Y=y)}} \cancel{\mathbb{P}(Y=y)}$$

$$= \sum_y \sum_x x\, \mathbb{P}(X=x, Y=y)$$

$$= \sum_x x \sum_y \mathbb{P}(X=x, Y=y) = \sum_x x\, \mathbb{P}(X=x)$$

$$= E(X)$$

Example. Assume that the chance of a coin coming Head is P. We toss the coin until it comes Head for the first time. What is the expected value of the number of times we toss the coin? Geometric Distribution

N = total number of trials until the first head comes

$E(N) = ?$

$$Y = \begin{cases} 1 & ; \text{if the coin comes head in the first toss} \\ 0 & , \text{o.w} \end{cases}$$

$$E(N) = E\left[E(N|Y)\right] = \sum_y E(N|Y=y)\,\mathbb{P}(Y=y)$$

$$= E(N|Y=1)\,\mathbb{P}(Y=1) + E(N|Y=0)\,\mathbb{P}(Y=0)$$

$$= 1 \times P + (1 + E(N)) \times (1-P) \Rightarrow E(N) = \frac{1}{P}$$

Note: Actually Tower Property is applied when we tend to calculate expected value of a random variable. In this example being H or T is a random variable and also the number of times we toss to achieve the first success is a random variable. (expectation of the sum of a random number of random variable)

Application of conditional probability in calculating the probability of an event.

Theorem:
$$\begin{cases} \mathbb{P}(A) = \sum_y \mathbb{P}(A|Y=y)\cdot \mathbb{P}(Y=y) & ; Y \text{ is discrete} \\ \mathbb{P}(A) = \int_y \mathbb{P}(A|Y=y)\, f(Y=y) & ; Y \text{ is continuous} \end{cases}$$

Proof. define the r.v X as: $X = \begin{cases} 1 & ; \text{if the event A happens} \\ 0 & ; \text{o.w} \end{cases}$

So $E(X) = \mathbb{P}(A)$ and $E(X|Y=y) = \sum_x x\, \mathbb{P}(X=x|Y=y) = \mathbb{P}(X=\overset{A}{1}|Y=y)$

also, $\mathbb{P}(A) = E(X) = \sum_y E(X|Y=y)\cdot \mathbb{P}(Y=y)$

$\rightarrow \mathbb{P}(A) = \sum \mathbb{P}(A|Y=y)\,\mathbb{P}(Y=y)$

Example. The number of customers arriving at a store per day is a r.v with Poisson distribution with average 10 persons per day. Each person with prob. 0.6 and independent of other customers leaves the system before receiving the servic. What is the prob. that in a certain day 15 customers stay in the system to be served.

we define X as the number of customers who stay in the system to be served.

So we tend to obtain $\mathbb{P}(X=15)$.

$$\mathbb{P}(X=15)=\sum_{n=15}^{\infty}\mathbb{P}(X=15|N=n)\cdot\underbrace{\mathbb{P}(\overset{\text{the \# of customers arriving in the system}}{N}=n)}_{\text{Poisson distribution with } \lambda=10}$$

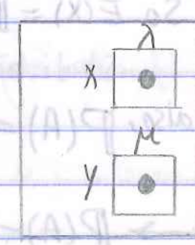
$$*\ \mathbb{P}(X=K|N=n)=\begin{cases}\binom{n}{k}p^{k}(1-p)^{n-k} & ;\ \text{if } n\geqslant K,\ p=0.4\\ 0 & ;\ n<K\end{cases}$$

$$\rightarrow \mathbb{P}(X=K)=\sum_{n=K}^{\infty}\binom{n}{k}p^{k}(1-p)^{n-k}\cdot\frac{e^{-\lambda}\lambda^{n}}{n!}=\sum\frac{n!}{(n-K)!K!}p^{k}(1-p)^{n-k}\cdot\frac{e^{-\lambda}\lambda^{n}}{n!}$$

$$=\frac{e^{-\lambda}(\lambda p)^{k}}{K!}\sum_{n=K}^{\infty}\frac{[\lambda(1-p)]^{n-k}}{(n-K)!}=\frac{e^{-\lambda}(\lambda p)^{k}}{K!}\cdot e^{\lambda(1-p)}=\frac{(\lambda p)^{k}e^{-\lambda p}}{K!}\sim \text{Poisson}(\lambda p)$$

Example. If X and Y are r.vs with Exponential Distribution with parameters λ and μ respectively, what is $\mathbb{P}(X\leqslant Y)=?$

for example a system with 2 server that both of them are serving costumer A and B and the service times are X and Y. A customer enters to the system and tends to know prob. of departure of A before B.



$$\mathbb{P}(Y>X)=\int_0^{\infty}\mathbb{P}(Y>X|X=x)f(x)\,dx=\int_0^{\infty}e^{-\mu x}\lambda e^{-\lambda x}\,dx=\lambda\int_0^{\infty}e^{-x(\lambda+\mu)}$$

$$=\frac{-\lambda}{\lambda+\mu}e^{-(\mu+\lambda)}\Big|_0^{\infty}=\frac{\lambda}{\lambda+\mu}$$

$$*\ \mathbb{P}(Y>K)=e^{-\mu K}$$

Example. Assume that the average number of accidents in a road day and the average number of people injured in each accident is 2 addition, assume that the number of accident is independent of the nu injuries. What is the average number of injuries?

$N=$ the number of accidents.

$X_i=$ the number of injuries in the accident i, $i=1,2,\dots,N$

Then the number of people injured is $\sum_{i=1}^{N}X_i$, So

$$E\Big[\sum_{i=1}^{\overset{\text{r.v}}{N}}\underset{\text{r.v}}{X_i}\Big]=E\Big[\underbrace{E\big[\sum_{i=1}^{N}X_i|N\big]}_{\text{we assume that we know } N}\Big]=E\big[N\,E[X_i]\big]\underset{\text{independency of } X_i \text{ an}}{=}E[N]\cdot E[X]=4\times 2$$

Note: The random variable $\sum_{i=1}^{N}X_i$, equal to the sum of a rando N of independent and identically distributed random variables that are independent of N, is called a "compound random variable". The expecte a compound random variable is $E[X]\cdot E[N]$.

Bayes Rule:

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

Example. The number of customers arriving at a store per day is a r.v with Poisson distribution with average 10 persons per day. Each person with Prob. 0.6 and independent of other customers leaves the system before receiving the servic. What is the prob. that in a certain day 15 customers stay in the system to be served.

we define X as the number of customers who stay in the system to be served.

So we tend to obtain $\mathbb{P}(X=15)$.

$$\mathbb{P}(X=15)=\sum_{n=15}^{\infty}\mathbb{P}(X=15\mid N=n)\cdot\underbrace{\mathbb{P}(N=n)}_{\text{Poisson distribution with } \lambda=10}$$

the # of customers arriving in the system

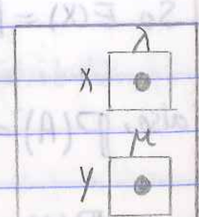
$$*\ \mathbb{P}(X=k\mid N=n)=\begin{cases}\binom{n}{k}p^k\cdot(1-p)^{n-k} & ;\ \text{if } n\geqslant k\ ,\ p=0.4\\ 0 & ;\ n<k\end{cases}$$

$$\to P(X=k)=\sum_{n=k}^{n}\binom{n}{k}p^k(1-p)^{n-k}\cdot\frac{e^{-\lambda}\lambda^n}{n!}=\sum\frac{n!}{(n-k)!\,k!}p^k(1-p)^{n-k}\cdot\frac{e^{-\lambda}\lambda^n}{n!}$$

$$=\frac{e^{-\lambda}(\lambda p)^k}{k!}\sum_{n=k}^{\infty}\frac{[\lambda(1-p)]^{n-k}}{(n-k)!}=\frac{e^{-\lambda}(\lambda p)^k}{k!}\cdot e^{\lambda(1-p)}=\frac{(\lambda p)^k e^{-\lambda p}}{k!}\sim\text{Poisson}(\lambda p)$$

Example. If X and Y are r.vs with Exponential Distribution with parameters λ and μ respectively, what is $\mathbb{P}(X\leqslant Y)=?$

for example a system with 2 server that both of them are serving costumer A and B and the service times are X and Y. A customer enters to the system and tends to know prob. of departure of A before B.



$$\mathbb{P}(Y>X)=\int_0^{\infty}\mathbb{P}(Y>X\mid X=x)f(x)\,dx=\int_0^{\infty}e^{-\mu x}\lambda e^{-\lambda x}\,dx=\lambda\int_0^{\infty}e^{-x(\lambda+\mu)}\,dx$$

$$=\frac{-\lambda}{\lambda+\mu}e^{-(\mu+\lambda)}\Big|_0^{\infty}=\frac{\lambda}{\lambda+\mu}$$

$$*\ \mathbb{P}(Y>k)=e^{-\mu k}$$

Example. Assume that the average number of accidents in a road is 4 per day and the average number of people injured in each accident is 2 persons. In addition, assume that the number of accident is independent of the number of injuries. What is the average number of injuries?

N = the number of accidents.

X_i = the number of injuries in the accident i, $i=1,2,\dots,N$

Then the number of people injured is $\sum_{i=1}^{N}X_i$, So

$$E\Big[\sum_{i=1}^{N}X_i\Big]=E\Big[\underbrace{E\big[\sum_{i=1}^{N}X_i\mid N\big]}_{\text{we assume that we know } N}\Big]=E\big[N\,E[X_i]\big]=E[N]\cdot E[X]=4\times2=8$$

(N → r.v, X_i → r.v; independency of X_i and N)

Note: The random variable $\sum_{i=1}^{N}X_i$, equal to the sum of a random number N of independent and identically distributed random variables that are also independent of N, is called a "compound random variable". The expected value of a compound random variable is $E[X]\cdot E[N]$.

Bayes Rule:

$$\mathbb{P}(A\mid B)=\frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)}=\frac{\mathbb{P}(B\mid A)\mathbb{P}(A)}{\mathbb{P}(B\mid A)\mathbb{P}(A)+\mathbb{P}(B\mid A^c)\mathbb{P}(A^c)}$$

Example. The products of a plant are produced in two workshops. In the first workshop 90% and in the second workshop 40% of products are compatible with the required standards. If a product is compatible with standard, then what is the probability that it is produced in the first workshop? Assume that the number of products in the first workshop is 3 times the second workshop.

B = is the event that the product is compatible with standards.
A = is the event that the product is constructed in the first workshop.

$$\mathbb{P}(A|B) = \frac{P(B|A).P(A)}{P(B|A).P(A)+P(B|A^c).P(A^c)} = \frac{0.9 \times 0.75}{0.9 \times 0.75 + 0.4 \times 0.25}$$

Moment Generating Function:

$$M_X(t) = E[e^{tx}] = \begin{cases} \sum_x e^{tx} p(x) & ; X \text{ is discrete} \\ \int_{-\infty}^{+\infty} e^{tx} f(x)dx & ; X \text{ is continuous} \end{cases}$$

Example. If X is a random variable with Poisson distribution with rate λ, what is MGF of X?

$$M_X(t) = E[e^{tx}] = \sum_{x=0}^{\infty} e^{tx} \frac{e^{-\lambda}\lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda e^t)^x}{x!} = e^{-\lambda} . e^{\lambda e^t} = e^{\lambda(e^t-1)}$$

Example. $X \sim Exp(\lambda) \to MGF(X) = ?$

$$M_X(t) = E(e^{tx}) = \int_0^{+\infty} e^{tx} . \lambda e^{-\lambda x} dx = \lambda \int_0^{+\infty} e^{x(t-\lambda)} dx = \frac{-\lambda}{\lambda - t} e^{-x(\lambda - t)} = \frac{\lambda}{\lambda - t}$$

Theorem. The moment generating function of the sum of two [...] and Y is the product of MGF of each of them.

$$M_{X+Y}(t) = E(e^{t(X+Y)}) = E_X(t) . M_Y(t)$$

Proof. $M_{X+Y}(t) = E(e^{t(x+y)}) = E(e^{tx} . e^{ty}) = E(e^{tx}) . E(e^{ty})$

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

Example. Suppose that X and Y are Poisson r.v. with parameters [...] respectively. What is the distribution function of $Z = X+Y$?

$$\begin{cases} M_X(t) = e^{\lambda_1(t-1)} \\ M_Y(t) = e^{\lambda_2(t-1)} \end{cases} \to M_Z(t) = M_{X+Y}(t) = M_X(t) . M_Y(t) = e^{\lambda_1(t-1)} . e^{\lambda_2(t-1)}$$

$$= e^{(\lambda_1+\lambda_2)(t-1)} = e^{\lambda(t-1)}$$

Note: There is a one-to-one correspondence between MGF and Pd[...]

$\Rightarrow Z = X+Y \sim \text{Poisson}(\lambda), \lambda = \lambda_1 + \lambda_2$

Theorem. The n^{th} moment of the r.v. X, $E[X^n]$, is equal to the [...] derivation of its MGF at $t=0$. $\frac{d M_X^n(t)}{dt^n}\Big|_{t=0} = E[X^n]$

Proof. We use induction to prove the theorem.

1) $\frac{d M_X(t)}{dt} = \frac{dE(e^{tx})}{dt} = \frac{E(d(e^{tx}))}{dt} = E\left(\frac{d(e^{tx})}{dt}\right) = E(xe^{tx})\Big|_{t=0} = E$ [...]

2) $\frac{d^2 M_X(t)}{dt^2} = \frac{d}{dt}[E(xe^{tx})] = E\left(\frac{d(xe^{tx})}{dt}\right) = E(x^2 e^{tx})\Big|_{t=0} = E(x$ [...]

Example. The products of a plant are produced in two workshops. In the first workshop 90% and in the second workshop 40% of products are compatible with the required standards. If a product is compatible with standard, then what is the probability that it is produced in the first workshop? Assume that the number of products in the first workshop is 3 times the second workshop.

B = is the event that the product is compatible with standards.
A = is the event that the product is constructed in the first workshop.

$$\mathbb{P}(A|B) = \frac{P(B|A).P(A)}{P(B|A).P(A)+P(B|A^c).P(A^c)} = \frac{0.9\times0.75}{0.9\times0.75+0.4\times0.25}$$

Moment Generating Function:

$$M_X(t) = E[e^{tx}] = \begin{cases} \sum_x e^{tx}p(x) & ; x \text{ is discrete} \\ \int_{-\infty}^{+\infty} e^{tx}f(x)dx & ; x \text{ is continuous} \end{cases}$$

Example. If X is a random variable with Poisson distribution with rate λ, what is MGF of X?

$$M_X(t) = E[e^{tx}] = \sum_{x=0}^{\infty} e^{tx}\frac{e^{-\lambda}\lambda^x}{x!} = e^{-\lambda}\sum_{x=0}^{\infty}\frac{(\lambda e^t)^x}{x!} = e^{-\lambda}.e^{\lambda e^t} = e^{\lambda(e^t-1)}$$

Example. $X \sim Exp(\lambda) \rightarrow MGF(X) = ?$

$$M_X(t) = E(e^{tx}) = \int_0^{+\infty} e^{tx}.\lambda e^{-\lambda x}dx = \lambda\int_0^{+\infty} e^{x(t-\lambda)}dx = \frac{-\lambda}{\lambda-t}e^{-x(\lambda-t)} = \frac{\lambda}{\lambda-t}$$

Theorem. The moment generating function of the sum of two r.vs X and Y is the product of MGF of each of them.

$$M_{X+Y}(t) = E(e^{t(X+Y)}) = M_X(t).M_Y(t)$$

Proof. $M_{X+Y}(t) = E(e^{t(x+y)}) = E(e^{tx}.e^{ty}) = E(e^{tx}).E(e^{ty})$

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

Example. Suppose that X and Y are Poisson r.v. with parameters λ_1 and λ_2 respectively. What is the distribution function of $Z = X+Y$?

$$\begin{cases} M_X(t) = e^{\lambda_1(t-1)} \\ M_Y(t) = e^{\lambda_2(t-1)} \end{cases} \rightarrow M_Z(t) = M_{X+Y}(t) = M_X(t).M_Y(t) = e^{\lambda_1(t-1)}.e^{\lambda_2(t-1)} = e^{(\lambda_1+\lambda_2)(t-1)} = e^{\lambda(t-1)}$$

Note: There is a one-to-one correspondence between MGF and Pdf.

$\Rightarrow Z = X+Y \sim \text{Poisson}(\lambda),\ \lambda = \lambda_1+\lambda_2$

Theorem. The n^{th} moment of the r.v. X, $E[X^n]$, is equal to the n^{th} derivation of its MGF at $t=0$. $\left.\frac{dM_X^n(t)}{dt^n}\right|_{t=0} = E[X^n]$

Proof. We use induction to prove the theorem.

1) $\frac{dM_X(t)}{dt} = \frac{dE(e^{tx})}{dt} = \frac{E(d(e^{tx}))}{dt} = E\left(\frac{d(e^{tx})}{dt}\right) = E(xe^{tx})\Big|_{t=0} = E(x)$

2) $\frac{d^2M_X(t)}{dt^2} = \frac{d}{dt}[E(xe^{tx})] = E\left(\frac{d(xe^{tx})}{dt}\right) = E(x^2e^{tx})\Big|_{t=0} = E(x^2)$

Example. Consider r.v. X to be Poisson distribution with parameter λ and $MGF_X(t)=e^{\lambda(t-1)}$. What is $E(t)$? and $Var(x)$?

$$E(x)=\frac{dM_X(t)}{dt}=\lambda e^t e^{\lambda(t-1)}\Big|_{t=0}=\lambda$$

$$E(x^2)=\frac{d^2M_X(t)}{dt^2}=e^{\lambda(e^t-1)}\left[\lambda e^t+(\lambda e^t)^2\right]\Big|_{t=0}=\lambda^2+\lambda$$

$$Var(x)=E^2(X)-E(x^2)=(\lambda^2+\lambda)-\lambda^2=\lambda$$

Example. $X\sim Exp(\lambda)$, $M_X(t)=\frac{\lambda}{\lambda-t}$. $E(x)=?$, $Var(x)=?$

$$E(x)=\left(\frac{\lambda}{\lambda-t}\right)'_t\Big|_{t=0}=\frac{\lambda}{(\lambda-t)^2}\Big|_{t=0}=\frac{1}{\lambda}$$

$$E(x^2)=\frac{d}{dt}\left(\frac{\lambda}{(\lambda-t)^2}\right)=\frac{2\lambda}{(\lambda-t)^3}\Big|_{t=0}=\frac{2}{\lambda^2}$$

$$Var(x)=\frac{2}{\lambda^2}-\left(\frac{1}{\lambda}\right)^2=\frac{1}{\lambda^2}$$

Convergent Series

A series is a list of numbers $\{a_1,a_2,\dots\}$ where a_n is called the general term. A series is called convergent if the sum of its elements (terms) is finite, that is:

$$\sum_{n=0}^{\infty}a_n<\infty$$

a_n can be either negative or positive. However in queuing we will only deal with positive numbers.

Main properties

1) The general term of the series, a_n, goes to 0, when $n\to\infty$, i.e. $\lim_{n\to\infty}a_n=0$.

Note: The series can be divergent but $\lim_{n\to\infty}a_n=0$.

2) Exponential series.

If the general term is in the form of $a_n=\frac{x^n}{n!}$, where $x<1$, then

$$\sum_{n=0}^{\infty}a_n=e^x$$

3) If the general term of the series is in the form of $a_n=nx^{n-1}$, x

then $\sum a_n$ is obtained as follows:

$$\sum_{n=0}^{\infty}a_n=\sum_{n=0}^{\infty}nx^{n-1}=\sum_{n=0}^{\infty}\frac{d(x^n)}{dx}=\frac{d}{dx}\left(\sum_{n=0}^{\infty}x^n\right)=\frac{d}{dx}\left(\frac{1}{1-x}\right)=\frac{1}{(1-x)}$$

If we want to obtain $\sum_{n=0}^{N}a_n$, then

$$\sum_{n=0}^{N}a_n=\frac{d}{dx}\left(\frac{1-x^{N+1}}{1-x}\right)=\frac{1-x^N}{(1-x)^2}-\frac{Nx^N}{1-x}$$

4) If $\lim_{n\to\infty}\frac{a_{n+1}}{a_n}=0$, then the series is convergent.

5) If $\lim_{n\to\infty}(a_n)^{1/n}<1$, then the series is convergent.

Z-transform (Generating Function)

Z-transform is similar to MGF, but it is used for discrete r.v.s.

Consider the r.v. X that $x\in Z^+$ and $P_i=P(X=i)$, then

$$P(Z)=\sum_{i=0}^{\infty}P_iZ^i$$

is called Z-transform of X, given that $P(Z)$ is convergent. For instan

$|Z|<1$, $P(Z)$ is always convergent.

Example. Consider r.v. X to be Poisson distribution with parameter λ and $MGF_X(t) = e^{\lambda(t-1)}$. What is $E(t)$? and $Var(X)$?

$$E(X) = \frac{dM_X(t)}{dt} = \lambda e^t e^{\lambda(t-1)} \Big|_{t=0} = \lambda$$

$$E(X^2) = \frac{d^2 M_X(t)}{dt^2} = e^{\lambda(e^t-1)} [\lambda e^t + (\lambda e^t)^2] \Big|_{t=0} = \lambda^2 + \lambda$$

$$Var(X) = E^2(X) - E(X^2) = (\lambda^2 + \lambda) - \lambda^2 = \lambda$$

Example. $X \sim Exp(\lambda)$, $M_X(t) = \frac{\lambda}{\lambda - t}$. $E(X) = ?$, $Var(X) = ?$

$$E(X) = \left(\frac{\lambda}{\lambda - t}\right)'_t \Big|_{t=0} = \frac{\lambda}{(\lambda-t)^2}\Big|_{t=0} = \frac{1}{\lambda}$$

$$E(X^2) = \frac{d}{dt}\left(\frac{\lambda}{(\lambda-t)^2}\right) = \frac{2\lambda}{(\lambda-t)^3}\Big|_{t=0} = \frac{2}{\lambda^2}$$

$$Var(X) = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}$$

Convergent Series

A series is a list of numbers $\{a_1, a_2, \dots\}$ where a_n is called the general term. A series is called convergent if the sum of its elements (terms) is finite, that is:

$$\sum_{n=0}^{\infty} a_n < \infty$$

a_n can be either negative or positive. However in queuing we will only deal with positive numbers.

Main properties

1) The general term of the series, a_n, goes to 0, when $n \to \infty$, i.e. $\lim a_n = 0$.

Note: The series can be divergent but $\lim_{n\to\infty} a_n = 0$.

2) Exponential series

If the general term is in the form of $a_n = \frac{x^n}{n!}$, where $x < 1$, then

$$\sum_{n=0}^{\infty} a_n = e^x$$

3) If the general term of the series is in the form of $a_n = n x^{n-1}$, $x < 1$, then $\sum a_n$ is obtained as follows:

$$\sum_{n=0}^{\infty} a_n = \sum_{n=0}^{\infty} n x^{n-1} = \sum_{n=0}^{\infty} \frac{d(x^n)}{dx} = \frac{d}{dx}\left(\sum_{n=0}^{\infty} x^n\right) = \frac{d}{dx}\left(\frac{1}{1-x}\right) = \frac{1}{(1-x)^2}$$

If we want to obtain $\sum_{n=0}^{N} a_n$, then

$$\sum_{n=0}^{N} a_n = \frac{d}{dx}\left(\frac{1-x^{N+1}}{1-x}\right) = \frac{1-x^N}{(1-x)^2} - \frac{Nx^N}{1-x}$$

4) If $\lim_{n\to\infty} \frac{a_{n+1}}{a_n} = 0$, then the series is convergent.

5) If $\lim_{n\to\infty} (a_n)^{1/n} < 1$, then the series is convergent.

Z-transform (Generating Function)

Z-transform is similar to MGF, but it is used for discrete r.v.s.

Consider the r.v. X that $X \in \mathbb{Z}^+$ and $P_i = P(X=i)$, then

$$P(z) = \sum_{i=0}^{\infty} P_i z^i$$

is called Z-transform of X, given that $P(z)$ is convergent. For instance, if $|z| < 1$, $P(z)$ is always convergent.

Soft proof. $p(z)=\sum_{i=0}^{\infty} p_i z^i < \sum_{i=0}^{\infty} z^i = \frac{1}{1-z}$

Example. If X has the geometric distribution, then what is $p(z)$?

$p_i = p(1-p)^i$

$p(z)=\sum_{i=0}^{\infty} p_i z^i = \sum p(1-p)^i z^i = p\sum [(1-p)z]^i = \frac{p}{1-[(1-p)z]} = \frac{p}{1-qz}$

Example. If X has the Binomial distribution, then what is $p(z)$?

$p_i = \binom{n}{i} p^i (1-p)^{n-i}$

$p(z)=\sum \binom{n}{i} p^i (1-p)^{n-i} z^i = \sum \binom{n}{i} (pz)^i (1-p)^{n-i} = (q+pz)^n$

$* \ (a+b)^n = \sum \binom{n}{i} a^i b^{n-i}$

Note: There is a unique correspondence between p_i and $p(z)$.

Theorem. If the z-transform of the r.v. X is $p(z)$, then the probability distribution function is obtained as follows.

$$\begin{cases} p_0 = p(z)\big|_{z=0} \\ p_n = \frac{1}{n!}\frac{d^n p(z)}{dz^n}\Big|_{z=0} \end{cases}$$

Proof. $p(z) = p_0 + p_1 z^1 + p_2 z^2 + \cdots$

$p(z)\big|_{z=0} = p_0$

$p'(z) = p_1 + 2p_2 z + 3p_3 z^2 + \cdots \Rightarrow p'(z)\big|_{z=0} = p_1$

$p''(z) = 2p_2 + 3!\,p_3 z + \cdots \Rightarrow p''(z)\big|_{z=0} = 2p_2$

$\Rightarrow \frac{dp^n(z)}{dz}\Big|_{z=0} = n!\,p_n \Rightarrow p_n = \frac{1}{n!}\frac{dp^n(z)}{dz}\Big|_{z=0}$

Example. Suppose that $p(z) = e^{-\lambda(1-z)}$, what is p_i?

$p_0 = p(z)\big|_{z=0} = e^{-\lambda}$

$p'(z) = \lambda e^{-\lambda(1-z)}\big|_{z=0} = \lambda e^{-\lambda} = p_1$

$p''(z) = \lambda^2 e^{-\lambda(1-z)}\big|_{z=0} = \lambda^2 e^{-\lambda} = 2p_2$

$\frac{d^n p(z)}{dz} = \lambda^n e^{-\lambda(1-z)}\big|_{z=0} = \lambda^n e^{-\lambda} = n!\,p_n \rightarrow p_n = \frac{\lambda^n e^{-\lambda}}{n!} \rightarrow X \sim \text{Poisson}(\lambda)$

Second way (direct approach)

$p(z) = e^{-\lambda(1-z)} = e^{-\lambda}\cdot e^{\lambda z} = e^{-\lambda}\sum \frac{(\lambda z)^n}{n!} = \sum \underbrace{\frac{e^{-\lambda}\lambda^n}{n!}}_{\text{Poisson}(\lambda)} z^n = \sum p_i z^i$

Properties of z-transform

1) $p(1) = 1$

2) $p'(1) = E(X)$

3) $p''(1) + p'(1) - [p'(1)]^2 = Var(X)$

Theorem. If X_1 and X_2 are two independent r.vs and $p(z_1)$ and p

their z-transform, respectively. Then, $p(z)$, the z-transform of X=

$$p(z) = p(z_1)\cdot p(z_2)$$

Proof. Assume that $p_i = p(X_1 = i)$ and $q_i = p(X_2 = i)$, then

$p(X=n) \overset{*}{=} p(X_1 + X_2 = n) = \sum p(X_1 = i)\cdot p(X_2 = n-i) = \sum p_i \cdot q_{n-i}$

↳ Convolution property

Soft proof. $p(z)=\sum_{i=0}^{\infty} P_i z^i < \sum_{i=0}^{\infty} z^i = \frac{1}{1-z}$

Example. If X has the geometric distribution, then what is $p(z)$?

$P_i = p(1-p)^i$

$p(z)=\sum_{i=0}^{\infty} P_i z^i = \sum p(1-p)^i z^i = p\sum[(1-p)z]^i = \frac{p}{1-[(1-p)z]} = \frac{p}{1-qz}$

Example. If X has the Binomial distribution, then what is $p(z)$?

$P_i = \binom{n}{i} p^i (1-p)^{n-i}$

$p(z)=\sum \binom{n}{i} p^i (1-p)^{n-i} z^i = \sum \binom{n}{i} (pz)^i (1-p)^{n-i} = (q+pz)^n$

$* \ (a+b)^n = \sum \binom{n}{i} a^i b^{n-i}$

Note: There is a unique correspondence between P_i and $p(z)$.

Theorem. If the z-transform of the r.v. X is $p(z)$, then the probability distribution function is obtained as follows.

$$\begin{cases} P_0 = p(z)\big|_{z=0} \\ P_n = \frac{1}{n!}\frac{d^n p(z)}{dz^n}\Big|_{z=0} \end{cases}$$

Proof. $p(z) = P_0 + P_1 z^1 + P_2 z^2 + \cdots$

$p(z)\big|_{z=0} = P_0$

$p'(z) = P_1 + 2P_2 z + 3P_3 z^2 + \cdots \Rightarrow p'(z)\big|_{z=0} = P_1$

$p''(z) = 2P_2 + 3!P_3 z + \cdots \Rightarrow p''(z)\big|_{z=0} = 2P_2$

$\Rightarrow \frac{dp^n(z)}{dz}\Big|_{z=0} = n!P_n \Rightarrow P_n = \frac{1}{n!}\frac{dp^n(z)}{dz}\Big|_{z=0}$

Example. Suppose that $p(z)=e^{-\lambda(1-z)}$, what is P_i?

$P_0 = p(z)\big|_{z=0} = e^{-\lambda}$

$p'(z) = \lambda e^{-\lambda(1-z)}\big|_{z=0} = \lambda e^{-\lambda} = P_1$

$p''(z) = \lambda^2 e^{-\lambda(1-z)}\big|_{z=0} = \lambda^2 e^{-\lambda} = 2P_2$

$\frac{d^n p(z)}{dz} = \lambda^n e^{-\lambda(1-z)}\big|_{z=0} = \lambda^n e^{-\lambda} = n!P_n \rightarrow P_n = \frac{\lambda^n e^{-\lambda}}{n!} \rightarrow X \sim \text{Poisson}(\lambda)$

Second way (direct approach)

$p(z) = e^{-\lambda(1-z)} = e^{-\lambda}\cdot e^{\lambda z} = e^{-\lambda}\sum \frac{(\lambda z)^n}{n!} = \sum \underbrace{\frac{e^{-\lambda}\lambda^n}{n!}}_{\text{Poisson}(\lambda)} z^n = \sum P_i z^i$

Properties of z-transform

1) $p(1)=1$

2) $p'(1) = E(X)$

3) $p''(1) + p'(1) - [p'(1)]^2 = Var(X)$

Theorem. If X_1 and X_2 are two independent r.vs and $p(z_1)$ and $p(z_2)$ are their z-transform, respectively. Then, $p(z)$, the z-transform of $X = X_1 + X_2$ is:

$$p(z) = p(z_1)\cdot p(z_2)$$

Proof. Assume that $P_i = P(X_1 = i)$ and $q_i = P(X_2 = i)$, then

$P(X=n) \overset{*}{=} P(X_1+X_2=n) = \underbrace{\sum P(X_1=i)\cdot P(X_2=n-i)}_{\text{Convolution property}} = \sum P_i \cdot q_{n-i}$

* By conditional prob.

$$P(X_1+X_2=n \mid X_1=i) = \sum P(X_1+X_2=n \mid X_1=i)\cdot P(X_1=i)$$

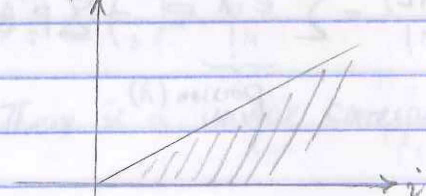
$$= \sum P(X_2=n-i)\cdot P(X_1=i)$$

Hence,

$$P(z) = \sum_{n=1}^{\infty} P(X=n)\cdot z^n = \sum\sum z^n p_i q_{n-i} = \sum_{i=0}^{\infty}\sum_{n=i}^{\infty} z^n p_i q_{n-i}$$

$$= \sum_{i=0}^{\infty} p_i z^i \sum_{n=i}^{\infty} q_{n-i} z^{n-i} = P_1(z)\cdot P_2(z)$$

Note:



Application of Z-transform in solving difference equations (DE):

In DEs, there are functions with integer variables.

$$(\lambda+\mu)P_n = \mu_{n+1}\cdot P_{n+1} + \lambda P_{n-1}$$

$$f(n+2) - f(n+1) - f(n) = 0$$

The aim to solve D.Es is to find $P_1, P_2, \dots$ that are true with the equation.

Example. Assume that P_n is a probability function with following property

$$\begin{cases} P_{n+1} - (1+a)P_n + aP_{n-1} = 0, \quad n=1,2,\dots \quad (*) \\ P_1 - aP_0 = 0 \end{cases}$$

Define the z-transform of P_n as $P(z) = \sum_{n=0}^{\infty} P_n z^n$, then we multip

equation by z^n and then sum all equations together:

$$\sum_{n=1}^{\infty} P_{n+1} z^n - \sum_{n=1}^{\infty}(1-a)P_n z^n + \sum_{n=1}^{\infty} aP_{n-1} z^n = 0$$

$$\sum_{n=1}^{\infty} P_{n+1} z^n = \frac{1}{z}\sum P_{n+1} z^{n+1} = \frac{1}{z}\left[P_2 z^2 + P_3 z^3 + \cdots\right] + (P_1 z + P$$

$$(P_1 z + P_0) = \frac{1}{z}\left[P(z) - P_0 - P_1\right)$$

$$\sum_{n=1}^{\infty} P_n z_n = P(z) - P_0$$

$$\sum_{n=1}^{\infty} P_{n-1} z^n = z\sum_{n=1}^{\infty} P_{n-1} z^{n-1} = zP(z)$$

Replace these 3 terms to the (*):

$$\frac{1}{z}\left[P(z) - P_0 - P_1 z\right] - (1-a)\left(P(z) - P_0\right) + azP(z) = 0$$

$$\Rightarrow P(z) = \frac{P_0}{1-az} = P_0\sum_{n=0}^{\infty}(az)^n = \sum_{n=0}^{\infty} P_0 a^n z^n \Rightarrow P_n = P_0 a^n$$

$$\sum_{n=0}^{\infty} P_n = 1 \Rightarrow \sum P_0 a^n = 1 \Rightarrow P_0 = \frac{1}{\sum a^n} \Rightarrow P_0 = \frac{1}{\frac{1}{1-a}} =$$

$$\Rightarrow P_n = (1-a)a^n, \quad n = 0, 1, 2, 3, \dots$$

* By conditional prob.

$$P(X_1+X_2=n \mid X_1=i) = \sum P(X_1+X_2=n \mid X_1=i)\cdot P(X_1=i)$$

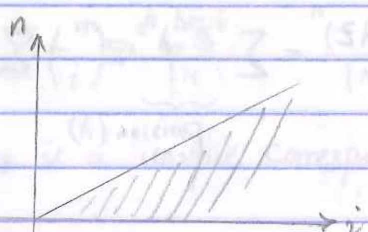
$$= \sum P(X_2=n-i)\cdot P(X_1=i)$$

Hence,

$$P(z)=\sum_{n=1}^{\infty} P(X=n)\cdot z^n = \sum\sum z^n P_i q_{n-i} = \sum_{i=0}^{\infty}\sum_{n=i}^{\infty} z^n P_i q_{n-i}$$

$$=\sum_{i=0}^{\infty} P_i z^i \sum_{n=i}^{\infty} q_{n-i} z^{n-i} = P_1(z)\cdot P_2(z)$$

Note:



Application of z-transform in solving difference equations (DE):

In DEs, there are functions with integer variables.

$(\lambda+\mu)P_n = \mu_{n+1}\cdot P_{n+1} + \lambda P_{n-1}$

$f(n+2)-f(n+1)-f(n)=0$

The aim to solve D.Es is to find $P_1, P_2, \dots$ that are true with the equation.

Example. Assume that P_n is a probability function with following property

$$\begin{cases} P_{n+1}-(1+a)P_n + aP_{n-1}=0, \quad n=1,2,\dots \quad (*) \\ P_1 - aP_0 = 0 \end{cases}$$

Define the z-transform of P_n as $P(z)=\sum_{n=0}^{\infty} P_n z^n$, then we multiply each equation by z^n and then sum all equations together:

$$\sum_{n=1}^{\infty} P_{n+1} z^n - \sum_{n=1}^{\infty}(1-a)P_n z^n + \sum_{n=1}^{\infty} aP_{n-1} z^n = 0$$

$$\sum_{n=1}^{\infty} P_{n+1} z^n = \frac{1}{z}\sum P_{n+1} z^{n+1} = \frac{1}{z}\left[P_2 z^2 + P_3 z^3 + \cdots\right] + (P_1 z + P_0) - (P_1 z + P_0) = \frac{1}{z}\left[P(z) - P_0 - P_1\right)$$

$$\sum_{n=1}^{\infty} P_n z_n = P(z) - P_0$$

$$\sum_{n=1}^{\infty} P_{n-1} z^n = z\sum_{n=1}^{\infty} P_{n-1} z^{n-1} = zP(z)$$

Replace these 3 terms to the (*):

$$\frac{1}{z}\left[P(z)-P_0-P_1 z\right] - (1-a)(P(z)-P_0) + azP(z) = 0$$

$$\Rightarrow P(z) = \frac{P_0}{1-az} = P_0\sum_{n=0}^{\infty}(az)^n = \sum_{n=0}^{\infty} P_0 a^n z^n \Rightarrow P_n = P_0 a^n$$

$$\sum_{n=0}^{\infty} P_n = 1 \Rightarrow \sum P_0 a^n = 1 \Rightarrow P_0 = \frac{1}{\sum a^n} \Rightarrow P_0 = \frac{1}{\frac{1}{1-a}} = 1-a$$

$$\Rightarrow P_n = (1-a)a^n, \quad n=0,1,2,3,\dots$$

Exponential and Poisson Distribution

Exponential Distribution

The r.v. X follows an exponential distribution, if for all $x \geq 0$, the pdf of X is:

$$f(x) = \lambda e^{-\lambda x}$$

Then we have these properties:

1) $F_X(x) = \mathbb{P}(X \leq x) = \int_0^x f(y)\,dy = 1 - e^{-\lambda x}$

$\mathbb{P}(X > x) = e^{-\lambda x}$

2) $E(X) = \int_0^\infty x \lambda e^{-\lambda x}\,dx = \lambda\left(\frac{-x}{\lambda} e^{-\lambda x} - \frac{1}{\lambda^2} e^{-\lambda x}\right)\Big|_0^\infty = \lambda\left(\frac{1}{\lambda^2}\right) = \frac{1}{\lambda}$

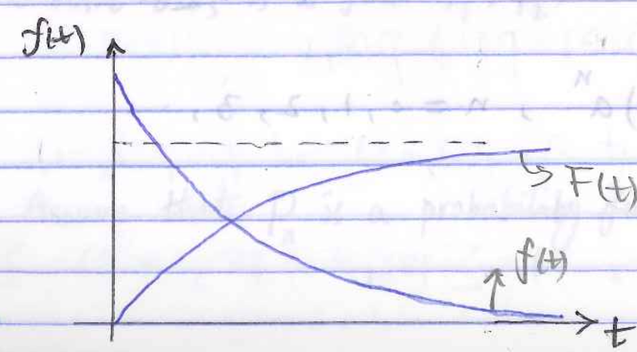
$$\begin{array}{lll} * & x & + e^{-\lambda x} \\ & 1 & -\frac{1}{\lambda} e^{-\lambda x} \\ & 0 & +\frac{1}{\lambda^2} e^{-\lambda x} \end{array}$$

3) $MGF = E(e^{tX}) = \int_0^\infty e^{tx} \lambda e^{-\lambda x}\,dx = \frac{\lambda}{\lambda - t}, \quad \lambda > t$

4) $Var(x) = E(X^2) - E^2(X) = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$

$* \; E(X^2) = \frac{d^2 M_X(t)}{dt^2} = \frac{2\lambda}{(\lambda - t)^3}\Big|_{t=0} = \frac{2}{\lambda^2}$

5)



Properties:

1) Memoryless property:

This is the most important property that says the past does any role in the future.

$$\mathbb{P}(X > s + t \mid X > t) = \mathbb{P}(X > s)$$

Proof. $P(X > t+s \mid X > t) = \frac{P(X > t+s,\, X > t)}{P(X > t)} = \frac{P(X > t+s)}{P(X > t)}$

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

Example. Assume that call durations are r.v. with Exp(5 min

An individual goes to a phone booth and sees that someone is t

but she does not know when the talk has started. What is th

that she waits more than 1 hour?

$$\mathbb{P}(X > 60) = e^{-60(1/5)} = e^{-12}$$

Theorem. The only memoryless continuous r.v. is Exponenti

Proof. $P(X > t+s \mid X > s) = \frac{P(X > t+s)}{P(X > s)}$

by memoryless property: $P(X > t) = \frac{P(X > t+s)}{P(X > s)}$

$\Rightarrow P(X > t+s) = P(X > t) \cdot P(X > s)$

Exponential and Poisson Distribution

Exponential Distribution

The r.v. X follows an exponential distribution, if for all $x \geqslant 0$, the pdf of X is:

$$f(x) = \lambda e^{-\lambda x}$$

Then we have these properties:

1) $F_X(x) = \mathbb{P}(X \leqslant x) = \int_0^x f(y)\,dy = 1 - e^{-\lambda x}$

$\mathbb{P}(X > x) = e^{-\lambda x}$

2) $E(X) = \int_0^\infty x \lambda e^{-\lambda x}\,dx = \lambda\left(\frac{-x}{\lambda} e^{-\lambda x} - \frac{1}{\lambda^2} e^{-\lambda x}\right)\Big|_0^\infty = \lambda\left(\frac{1}{\lambda^2}\right) = \frac{1}{\lambda}$

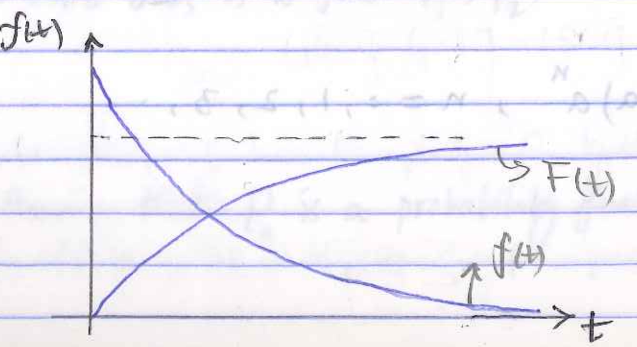
$$* \quad \begin{array}{ccl} x & + & e^{-\lambda x} \\ 1 & - & -\frac{1}{\lambda} e^{-\lambda x} \\ 0 & + & \frac{1}{\lambda^2} e^{-\lambda x} \end{array}$$

3) $MGF = E(e^{tx}) = \int_0^\infty e^{tx} \lambda e^{-\lambda x}\,dx = \frac{\lambda}{\lambda - t}, \quad \lambda > t$

4) $Var(X) = E(X^2) - E^2(X) = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$

$* \; E(X^2) = \frac{d^2 M_X(t)}{dt^2} = \frac{2\lambda}{(\lambda - t)^3}\Big|_{t=0} = \frac{2}{\lambda^2}$

5)



Properties:

1) Memoryless property:

This is the most important property that says the past does not play any role in the future.

$$\mathbb{P}(X > s+t \mid X > t) = \mathbb{P}(X > s)$$

Proof. $P(X > t+s \mid X > t) = \frac{P(X > t+s,\; X > t)}{P(X > t)} = \frac{P(X > t+s)}{P(X > t)}$

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

Example. Assume that call durations are r.v. with $Exp(5^{mins})$ distribut. An individual goes to a phone booth and sees that someone is talking but she does not know when the talk has started. What is the probability that she waits more than 1 hour?

$$P(X > 60) = e^{-60(1/5)} = e^{-12}$$

Theorem. The only memoryless continuous r.v. is Exponential.

Proof. $\underbrace{P(X > t+s \mid X > s)} = \frac{P(X > t+s)}{P(X > s)}$

by memoryless property: $P(X > t) = \frac{P(X > t+s)}{P(X > s)}$

$\Rightarrow P(X > t+s) = P(X > t) \cdot P(X > s)$

$$\begin{cases} \frac{d}{ds} P(X>t+s) = P(X>t) \cdot \frac{d}{ds} P(X>s) \\ \frac{d}{ds} P(X>s) = \frac{d}{ds}(1-F(s)) = -f(s)\,ds \end{cases}$$

$$\Rightarrow dP(X>t+s) = P(X>t)\cdot(-f(s)) \Rightarrow \frac{dP(X>t+s)}{dP(X>t)} = -f(s)\,ds$$

If $s=0 \Rightarrow \frac{dP(X>t)}{d(X>0)} = -f(0)\,ds$

by integral from both sides:

$$Ln\,(P(X>t))' = -f(0)t \Rightarrow P(X>t) = e^{-f(0)t}$$

Hence, X would be Exponential distribution with rate $f(0)$.

2) The minimum of a number of r.v.s. is an exponential r.v..

Theorem. If $X_1, X_2, \dots, X_n$ are iid r.v.s from exponential distribution with parameters $\lambda_1, \lambda_2, \dots, \lambda_n$ and $X = \min\{X_1, X_2, \dots, X_n\}$ then $X \sim Exp(\sum_{i=1}^{n} \lambda_i)$.

proof. $P\{\min\{X_1, X_2, \dots, X_n\} > x\} = P(X_1>x, X_2>x, \dots, X_n>x)$

by factorization theorem $= P(X_1>x)\cdot P(X_2>x) \cdots P(X_n>x)$

$$= e^{-\lambda_1 x}\cdot e^{-\lambda_2 x} \cdots e^{-\lambda_n x} = e^{-\sum \lambda_i x}$$

$$\Rightarrow X \sim Exp(\textstyle\sum \lambda_i)$$

Example. In a queuing system there are two servers. The service time of the two servers are r.v.s from exponential distribution with average of 10 and 15 mins. A customer enters the system and faces that both servers are busy and no one in the queue. If at the arrival time 12 mins have passed from the service time of the person and 8 mins from the second person. Then,

(a) What is the probability that the arrived customer waits no l

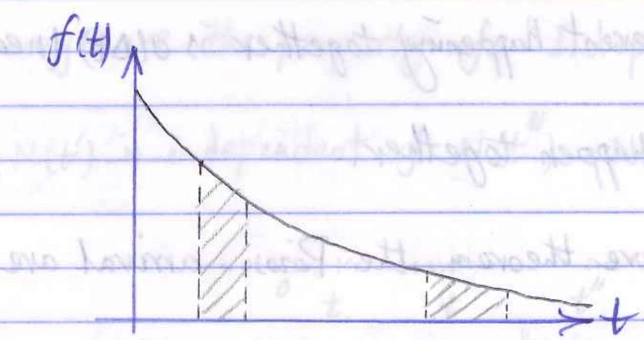
3 mins?

(b) What is the average waiting time?

(a) $P(X>3) = e^{-(\frac{1}{10}+\frac{1}{15})3} = e^{-1/2}$

(b) $E(X) = \frac{1}{\frac{1}{6}} = 6$ mins.

3) The Pdf of exponential distribution is decreasing.



4) The probability of an event in the small time interval Δt is $\lambda \Delta t$.

This means that if an event has not happend until time x, then probability that happens between the interval $(x, x+\Delta t]$ has propor- with λ and Δt.

$$\begin{cases} \frac{d}{ds} P(X>t+s) = P(X>t) \cdot \frac{d}{ds} P(X>s) \\ \frac{d}{ds} P(X>s) = \frac{d}{ds}(1-F(s)) = -f(s)\,ds \end{cases}$$

$$\Rightarrow dP(X>t+s) = P(X>t) \cdot (-f(s)) \Rightarrow \frac{dP(X>t+s)}{dP(X>t)} = -f(s)\,ds$$

If $s=0 \Rightarrow \frac{dP(X>t)}{d(X>0)} = -f(0)\,ds$

by integral from both sides:

$$Ln\,(P(X>t))' = -f(0)t \Rightarrow P(X>t) = e^{-f(0)t}$$

Hence, X would be Exponential distribution with rate $f(0)$.

2) The minimum of a number of r.v.s. is an exponential r.v..

Theorem. If $X_1, X_2, \dots, X_n$ are iid r.v.s from exponential distribution with parameters $\lambda_1, \lambda_2, \dots, \lambda_n$ and $X = \min\{X_1, X_2, \dots, X_n\}$ then $X \sim Exp(\sum_{i=1}^{n} \lambda_i)$.

Proof. $P\{\min\{X_1, X_2, \dots, X_n\} > x\} = P(X_1 > x, X_2 > x, \dots, X_n > x)$

by factorization theorem $= P(X_1 > x) \cdot P(X_2 > x) \cdots P(X_n > x)$

$$= e^{-\lambda_1 x} \cdot e^{-\lambda_2 x} \cdots e^{-\lambda_n x} = e^{-\sum \lambda_i x}$$

$$\Rightarrow X \sim Exp(\sum \lambda_i)$$

Example. In a queuing system there are two servers. The service time of the two servers are r.v.s from exponential distribution with average of 10 and 15 mins. A customer enters the system and faces that both servers are busy and no one in the queue. If at the customer arrival time 12 mins have passed from the service time of the first person and 8 mins from the second person. Then,

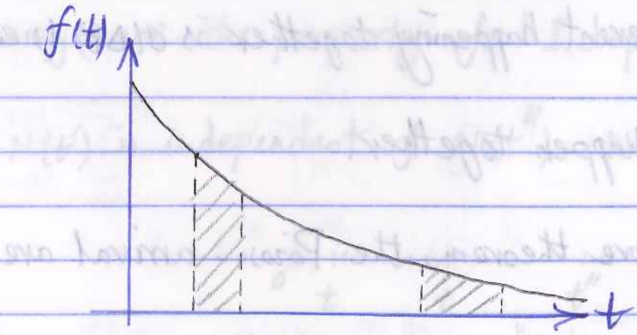
(a) What is the probability that the arrived customer waits no less than 3 mins?

(b) What is the average waiting time?

(a) $P(X>3) = e^{-(\frac{1}{10}+\frac{1}{15})3} = e^{-\frac{1}{2}}$

(b) $E(X) = \frac{1}{\frac{1}{6}} = 6$ mins.

3) The pdf of exponential distribution is decreasing.



4) The probability of an event in the small time interval Δt is almost $\lambda \Delta t$.

This means that if an event has not happend until time x, then the probability that happens between the interval $(x, x+\Delta t]$ has proportional with λ and Δt.

This is:

$$\lim_{\Delta t\to 0}\frac{\mathbb{P}(X\leqslant x+\Delta t\mid X>x)}{\Delta t}=\lambda$$

Proof. $P(X\leqslant x+\Delta t\mid X>x)=P(X\leqslant \Delta t)=1-e^{-\lambda\Delta t}$

$$=1-\left(1-\lambda\Delta t+\frac{(\lambda\Delta t)^2}{2!}-\cdots\right)=\lambda\Delta t+\underbrace{o(\Delta t)}_{\text{order of }\Delta t}$$

where $o(\Delta t)$ is the set of functions for which $\lim_{x\to 0}\frac{o(x)}{x}=0$

for example $x\notin o(x)$ because $\lim_{x\to 0}\frac{x}{x}=1$ but $x^2\in o(x)$ because $\lim_{x\to 0}\frac{x^2}{x}=0$

$$\Rightarrow \lim_{\Delta t\to 0}\frac{P(X\leqslant x+\Delta t\mid X>x)}{\Delta t}=\lim_{\Delta t\to 0}\frac{\lambda\Delta t+\overbrace{o(\Delta t)}^{\text{vanished}}}{\Delta t}$$

Hence, proved.

Note: Based on this property we may show that in a Poisson process, the probability of two event happening together is $o(\Delta t)$ (nearly zero). Hence the two events can not happen together.

Considering the above theorem, the Poisson arrival are sometimes called completely random arrivals, because the arrival chance of a customer in the short time interval Δt is independent of the past and only depends on λ and Δt.

Counting Process.

Counting Process (CP) refers to the number of events up to a certain time. Generally, if $\{N(t),\ t\geqslant 0\}$ is a CP, $N(t)$ is the number of events from time 0 to t.

Example. The number of children born in a city in a certain ti

The number of customers that join a queue in a day.

Based on this definition, $N(t)$ can only take non-negative values

addition $N(t)$ is an increasing function in t. That is

$$t<t' \rightarrow N(t)\leqslant N(t')$$

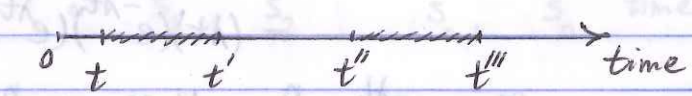
The number of events that happen between the time t and t' is

$$N(t')-N(t)$$

In addition $N(0)=0$.

A CP may have the property of "Independent Increament", t

$N(t)-N(t')$ is independent of $N(t'')-N(t''')$ when $t'<t''$.



A CP may have the property of "stationary increament" if N(t

only depends on $\Delta t=t'-t$.

Poisson Process:

Poisson Process (PP) is a specific case of counting process. The

$\{N(t),\ t\geqslant 0\}$ is said to be a Poisson process with parameter λ, This is:

$$\lim_{\Delta t\to 0} \frac{P(X\leqslant x+\Delta t \mid X>x)}{\Delta t} = \lambda$$

Proof. $P(X\leqslant x+\Delta t \mid X>x) = P(X\leqslant \Delta t) = 1-e^{-\lambda\Delta t}$

$$= 1-\left(1-\lambda\Delta t+\frac{(\lambda\Delta t)^2}{2!}-\dots\right) = \lambda\Delta t + o(\Delta t)$$

↳ order of Δt

where $o(\Delta t)$ is the set of functions for which $\lim_{x\to 0}\frac{o(x)}{x}=0$

for example $x\notin o(x)$ because $\lim_{x\to 0}\frac{x}{x}=1$ but $x^2\in o(x)$ because $\lim_{x\to 0}\frac{x^2}{x}=0$

$$\Rightarrow \lim_{\Delta t\to 0}\frac{P(X\leqslant x+\Delta t\mid X>x)}{\Delta t} = \lim_{\Delta t\to 0}\frac{\lambda\Delta t+o(\Delta t)}{\Delta t}$$

vanished

Hence, proved.

Note: Based on this property we may show that in a Poisson process, the probability of two event happening together is $o(\Delta t)$ (nearly zero). Hence the two events can not happen together.

Considering the above theorem, the Poisson arrival are sometimes called completely random arrivals, because the arrival chance of a customer in the short time interval Δt is independent of the past and only depends on λ and Δt.

Counting Process.

Counting Process (CP) refers to the number of events up to a certain time. Generally, if $\{N(t), t\geqslant 0\}$ is a CP, $N(t)$ is the number of events from time 0 to t.

Example. The number of children born in a city in a certain time period

The number of customers that join a queue in a day.

Based on this definition, $N(t)$ can only take non-negative values. In addition $N(t)$ is an increasing function in t. That is

$$t<t' \rightarrow N(t)\leqslant N(t')$$

The number of events that happen between the time t and t' is

$$N(t')-N(t)$$

In addition $N(0)=0$.

A CP may have the property of "Independent Increament", that is $N(t)-N(t')$ is independent of $N(t'')-N(t''')$ when $t'<t''$.

A CP may have the property of "stationary increament" if $N(t')-N(t)$ only depends on $\Delta t = t'-t$.

Poisson Process:

Poisson Process (PP) is a specific case of counting process. The CP $\{N(t), t\geqslant 0\}$ is said to be a Poisson process with parameter λ, if the number of events in the short time interval $[s, s+t)$ follows the Poisson distribution as below:

$$P[N(t+s)-N(s)=n] = \frac{e^{-\lambda t}(\lambda t)^n}{n!}$$

Furthermore, we assume that the independent increament also exists.

Note: Poisson process is different from Poisson distribution. In addition for any constant time value, λt is also constant. Hence, PP become a Poisson dist. (λt written over λ)

Note: Based on Poisson process definition, the stationary increament property would also exist in the Poisson process.

Expected value and variance.

If $N(t)$ is a Poisson process, then

$$E[N(t)] = \sum_{n=0}^{\infty} n P(N(t)=n) = \sum n \frac{e^{-\lambda t}(\lambda t)^n}{n!} = e^{-\lambda t}\sum_{n=0}^{\infty}\frac{(\lambda t)^n}{(n-1)!} = \lambda t(e^{-\lambda t})\sum_{n=1}^{\infty}\frac{(\lambda t)^{n-1}}{(n-1)!}$$

$$= (\lambda t)(e^{-\lambda t})(e^{\lambda t}) = \lambda t$$

$$E[N^2(t)] = \sum_{n=0}^{\infty} n^2 \frac{e^{-\lambda t}(\lambda t)^n}{n!} = e^{-\lambda t}\sum_{n=0}^{\infty}\frac{n(\lambda t)^n}{(n-1)!} = e^{-\lambda t}(\lambda t)\sum_{n=1}^{\infty}\frac{n(\lambda t)^{n-1}}{(n-1)!}$$

$$= (\lambda t)e^{-\lambda t}\sum\left[\frac{(n-1)(\lambda t)^{n-1}}{(n-1)!} + \frac{(\lambda t)^{n-1}}{(n-1)!}\right]$$

$$\sum_{n=2}^{\infty}\frac{(\lambda t)^{n-1}}{(n-2)!}$$

$$= (\lambda t)\left[(\lambda t)e^{\lambda t} + e^{\lambda t}\right] = \lambda t + (\lambda t)^2$$

$$Var(N(t)) = E[N^2(t)] - [E(N(t))]^2 = \lambda t + (\lambda t)^2 - (\lambda t)^2 = \lambda t.$$

The other approach is applying MGF

$$M_{N(t)}(s) = e^{\lambda t(e^s-1)}$$

$$E(N(t)) = \frac{d M_{N(t)}(s)}{ds}\bigg|_{s=0} = \lambda t e^s e^{\lambda t(e^s-1)}\bigg|_{s=0} = \lambda t$$

$$E(N^2(t)) = \frac{d^2 M_{N(t)}(s)}{ds^2}\bigg|_{s=0} = \left(e^s \lambda t\, e^{\lambda t(e^s-1)} + \lambda t e^s \lambda t e^s e^{\lambda s(e^s-1)}\right)\bigg|_{s=} =$$

The relationship between Poisson process and Exponential distributi

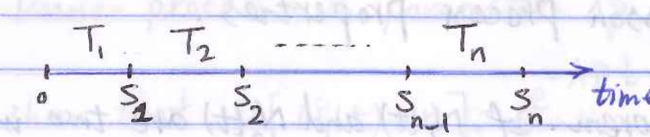
Suppose that the number of events, $N(t)$, follows a Poisson process. I

first event take place at time S_1 and the second happens at tim

and the n^{th} event happens at time S_n and we lable the times b

each two events as $T_1, T_2, \ldots, T_n$, then $T_1, T_2, \ldots, T_n$ are indepen

following exponential distribution with the same parameter λ a



Proof. We first show that T_1 has exponential distribution wit

λ.

$$P(T_1 > x) = P(\text{No enters or events take place until time } x)$$

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

Next we show that T_2 is exponential.

number of events in the short time interval $[s, s+t)$ follows the Poisson distribution as below:

$$P[N(t+s)-N(s)=n]=\frac{e^{-\lambda t}(\lambda t)^n}{n!}$$

Furthermore, we assume that the independent increament also exists.

Note: Poisson process is different from Poisson distribution. In addition for any constant time value, $\overset{\lambda t}{\lambda t}$ is also constant. Hence, PP become a Poisson dist. $\overset{\lambda}{}$

Note: Based on Poisson process definition, the stationary increament property would also exist in the Poisson process.

Expected value and variance.

If $N(t)$ is a Poisson process, then

$$E[N(t)]=\sum_{n=0}^{\infty} n P(N(t)=n)=\sum n\frac{e^{-\lambda t}(\lambda t)^n}{n!}=e^{-\lambda t}\sum_{n=0}^{\infty}\frac{(\lambda t)^n}{(n-1)!}=\lambda t(e^{-\lambda t})\sum_{n=1}^{\infty}\frac{(\lambda t)^{n-1}}{(n-1)!}$$

$$=(\lambda t)(e^{-\lambda t})(e^{\lambda t})=\lambda t$$

$$E[N^2(t)]=\sum_{n=0}^{\infty} n^2\frac{e^{-\lambda t}(\lambda t)^n}{n!}=e^{-\lambda t}\sum_{n=0}^{\infty}\frac{n(\lambda t)^n}{(n-1)!}=e^{-\lambda t}(\lambda t)\sum_{n=1}^{\infty}\frac{n(\lambda t)^{n-1}}{(n-1)!}$$

$$=(\lambda t)e^{-\lambda t}\sum\left[\frac{(n-1)(\lambda t)^{n-1}}{(n-1)!}+\frac{(\lambda t)^{n-1}}{(n-1)!}\right]$$

$$\sum_{n=2}^{\infty}\frac{(\lambda t)^{n-1}}{(n-2)!}$$

$$=(\lambda t)\left[(\lambda t)e^{\lambda t}+e^{\lambda t}\right]=\lambda t+(\lambda t)^2$$

$$Var(N(t))=E[N^2(t)]-[E(N(t))]^2=\lambda t+(\lambda t)^2-(\lambda t)^2=\lambda t.$$

The other approach is applying MGF

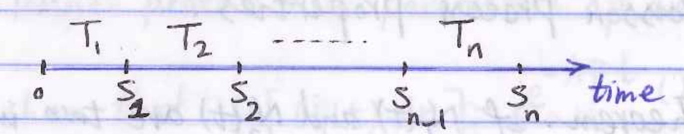
$$M_{N(t)}(s)=e^{\lambda t(e^s-1)}$$

$$E(N(t))=\frac{dM_{N(t)}(s)}{ds}\Big|_{s=0}=\lambda t e^s e^{\lambda t(e^s-1)}\Big|_{s=0}=\lambda t$$

$$E(N^2(t))=\frac{d^2M_{N(t)}(s)}{ds^2}\Big|_{s=0}=\left(e^s\lambda t\, e^{\lambda t(e^s-1)}+\lambda t e^s \lambda t e^s e^{\lambda s(e^s-1)}\right)\Big|_{s=0}=\lambda t+(\lambda t)^2$$

The relationship between Poisson process and Exponential distribution.

Suppose that the number of events, $N(t)$, follows a Poisson process. If the first event take place at time S_1 and the second happens at time S_2, ..., and the n^{th} event happens at time S_n and we lable the times between each two events as $T_1, T_2, \ldots, T_n$, then $T_1, T_2, \ldots, T_n$ are independent r.v.s following exponential distribution with the same parameter λ as in PP.



Proof. We first show that T_1 has exponential distribution with parameter λ.

$$P(T_1>x)=P(\text{No enters or events take place until time } x)$$

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

Next we show that T_2 is exponential.

$P(T_2 > x \mid T_1 = s) = P(\text{No events in the interval } [s, s+x] \mid T_1 = s)$

$= P(\text{No events in the interval } [s, s+x)) = e^{-\lambda x}$

Similarly, one can show that $T_3, T_4, \dots, T_n$ have exponential distribution.

Example. The customers of a bank arrive based on a Poisson process with rate 10 people per hour. After the bank opens the door in the morning, what is the probability that no one arrives in the first 15 mins. What is the probability that the time between the 7^{th} and 8^{th} customer arrival is greater than 1 hour.

$P(T_1 > \frac{1}{4}^{\text{hour}}) = e^{-10(1/4)} = e^{-2.5}$

$P(T_8 > 1^{\text{hour}}) = e^{-10(1)} = e^{-10}$

Poisson process properties.

Theorem. If $N_1(t)$ and $N_2(t)$ are two independent Poisson process with parameters λ_1 and λ_2, then the process $N(t) = N_1(t) + N_2(t)$ is also a Poisson process with $\lambda = \lambda_1 + \lambda_2$.

Proof. we know that $MGF = M_{N(t)}(s) = M_{N_1(t)+N_2(t)}(s)$

$= M_{N_1(t)}(s) \cdot M_{N_2(t)}(s)$

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

From this proposition we can conclude if a system has n types customers and each types arrive independently of other types, then number of customers arriving at the system is a Poisson process summation of all parameters.

Theorem. (Spliting property of Poisson process)

Consider a system that the customers arrival at it follows a P[oisson] process, $N(t)$, with parameter λ. There are two types of custo[mers] 1, 2. Each customer is of type 1 with probability p and is of ty[pe 2 with] probability $(1-p)$. If $N_1(t)$ and $N_2(t)$ are the number of custome[rs of type 1] and 2 that arrive at the system up to time t, then $N_1(t)$ and [$N_2(t)$] are two independent Poisson process with parameters λp and $\lambda(1$[$-p)$].

Proof. we want to show that $P[N_1(t) = n] = \frac{e^{-\lambda p t}(\lambda p t)^n}{n!}$

$P(N_1(t) = n) = \sum_{m=n}^{\infty} P(N_1(t) = n \mid N(t) = m) \cdot P(N(t) = m)$

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

$= \frac{e^{-\lambda t}(\lambda t)^n p^n}{n!} \underbrace{\sum \frac{(1-p)^{m-n}(\lambda t)^{m-n}}{(m-n)!}}_{e^{\lambda t(1-p)}} = \frac{e^{-\lambda t p}(\lambda t p)^n}{n!}$

$$P(T_2 > x \mid T_1 = s) = P(\text{No events in the interval } [s, s+x] \mid T_1 = s)$$
$$= P(\text{No events in the interval } [s, s+x]) = e^{-\lambda x}$$

Similarly, one can show that $T_3, T_4, \ldots, T_n$ have exponential distribution.

Example. The customers of a bank arrive based on a Poisson process with rate 10 people per hour. After the bank opens the door in the morning, what is the probability that no one arrives in the first 15 mins. What is the probability that the time between the 7^{th} and 8^{th} customer arrival is greater than 1 hour.

$$P(T_1 > \tfrac{1}{4} \text{ hour}) = e^{-10(1/4)} = e^{-2.5}$$

$$P(T_8 > 1 \text{ hour}) = e^{-10(1)} = e^{-10}$$

Poisson process properties.

Theorem. If $N_1(t)$ and $N_2(t)$ are two independent Poisson process with parameters λ_1 and λ_2, then the process $N(t) = N_1(t) + N_2(t)$ is also a Poisson process with $\lambda = \lambda_1 + \lambda_2$.

Proof. We know that $MGF = M_{N(t)}(s) = M_{N_1(t)+N_2(t)}(s)$

$$= M_{N_1(t)}(s) \cdot M_{N_2(t)}(s)$$

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

From this proposition we can conclude if a system has n types of customers and each types arrive independently of other types, then the number of customers arriving at the system is a Poisson process with the summation of all parameters.

Theorem. (Spliting property of Poisson process)

Consider a system that the customers arrival at it follows a Poisson process, $N(t)$, with parameter λ. There are two types of customers 1 and 2. Each customer is of type 1 with probability p and is of type 2 with probability $(1-p)$. If $N_1(t)$ and $N_2(t)$ are the number of customers type 1 and 2 that arrive at the system up to time t, then $N_1(t)$ and $N_2(t)$ are two independent Poisson process with parameters λp and $\lambda(1-p)$.

Proof. We want to show that $P[N_1(t) = n] = \frac{e^{-\lambda p t}(\lambda p t)^n}{n!}$

$$P(N_1(t) = n) = \sum_{m=n}^{\infty} P(N_1(t) = n \mid N(t) = m) \cdot P(N(t) = m)$$

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

$$= \frac{e^{-\lambda t}(\lambda t)^n p^n}{n!} \underbrace{\sum \frac{(1-p)^{m-n}(\lambda t)^{m-n}}{(m-n)!}}_{e^{\lambda t(1-p)}} = \frac{e^{-\lambda t p}(\lambda t p)^n}{n!}$$

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

(a) $N(0) = 0$

(b) $N(t)$ has independent increament property.

(c) The probability of an event in a short time interval Δt is

$$p[N(\Delta t) = 1] = \lambda \Delta t + o(\Delta t)$$

(d) The probability of two events or more in the short time interval Δt is $P[N(\Delta t) \geq 2] = o(\Delta t)$ and $o(\Delta t)$ is a infinitly small function of $(\Delta t)^2$ when $\Delta t \to 0$. That is

$$\lim_{\Delta t \to 0} \frac{o(\Delta t)}{\Delta t} = 0$$

Proof. For convenience let us $P[N(t) = n]$ with $P_n(t)$. We need to show that $P_n(t) = \frac{e^{-\lambda t}(\lambda t)^n}{n!}$

Using conditional probability we have:

$$P_n(t+\Delta t) = p(N(t+\Delta t) = n) = \sum_{m=0}^{\infty} p(N(t+\Delta t) = n \mid N(t) = m) \cdot p(N(t) = m)$$

$$= \sum_{m=0}^{n-2} \underbrace{p(N(t+\Delta t) = n \mid N(t) = m) \cdot P_m(t)}_{o(\Delta t)}$$

$$+ \underbrace{p(N(t+\Delta t) = n \mid N(t) = n-1)}_{\substack{\mathrm{I} \\ \lambda t + o(\Delta t)}} \cdot P_{n-1}(t) + \underbrace{p(N(t+\Delta t) = n \mid N(t) = n)}_{\substack{1 - P_1(\Delta t) - P_{\geq 2}(\Delta t) \\ \mathrm{II}}} P_n(t)$$

$$+ \underbrace{\sum_{m=n+1}^{\infty} p(N(t+\Delta t) = n \mid N(t) = m) \cdot P_m(t)}_{\text{Zero}}$$

(I) $p(N(t+\Delta t) = n \mid N(t) = n-1) = p(\text{exactly one event in } \Delta t) = \lambda \Delta t$

(II) $p(N(t+\Delta t) = n \mid N(t) = n) = p(\text{No event in } \Delta t) = 1 - P_1(\Delta t) - P_{n \geq 2}(\Delta t)$

$$= 1 - (\lambda \Delta t + o(\Delta t)) - o(\Delta t) = 1 - \lambda \Delta t + o(\Delta t)$$

Note: $o_1(\Delta t) + o_2(\Delta t) = o_3(\Delta t)$ and $o_1(\Delta) \cdot o_2(\Delta t) = o_3(\Delta t)$

$$P_n(t+\Delta t) = \lambda \Delta t . P_{n-1}(t) + (1 - \lambda \Delta t) P_n(t) + o(\Delta t)$$

$$\Rightarrow P_n(t+\Delta t) - P_n(t) = \lambda \Delta t\, P_{n-1}(t) - \lambda \Delta t\, P_n(t) + o(\Delta t)$$

$$\Rightarrow \lim_{\Delta t \to 0} \frac{P_n(t+\Delta t) - P_n(t)}{\Delta t} = \lambda P_{n-1}(t) - \lambda P_n(t) + \underbrace{\frac{o(\Delta t)}{\Delta t}}_{0}$$

$$\Rightarrow P'_n(t) = \lambda P_{n-1}(t) \mp \lambda P_n(t) \quad , n = 0, 1, 2, \ldots$$

Lets define the z-transform of $P_n(t)$ as $p(z) = \sum_{n=0}^{\infty} z^n P_n(t)$

If $n = 0 \Rightarrow P'_0(t) = -\lambda P_0(t) \Rightarrow P_0(t) = e^{-\lambda t} + K$

Using the initial condition $\begin{cases} P_0(0) = 1 \\ P_n(0) = 0 \end{cases} \Rightarrow K = 1 \Rightarrow P_0(t) = e^{-\lambda t}$

$$P_t(z) = \sum_{n=0}^{\infty} z^n P_n(t)$$

$$\begin{cases} P'_n(t) = -\lambda (P_n(t) - P_{n-1}(t)) \quad , n \geq 1 \quad , P_n(0) = 0 \quad , P_0(0) = 1 \\ P'_0(t) = -\lambda P_0(t) \end{cases}$$

$$\sum_{n=1}^{\infty} P'_n(t) z^n = -\lambda \left(\sum_{n=1}^{\infty} P_n(t) z^n \overset{+P_0 z^0 - P_0 z^0}{} - \sum_{n=1}^{\infty} P_{n-1}(t) z^n \right)$$

$$\Rightarrow \sum_{n=0}^{\infty} P'_n(t) z^n - P'_0(t) = -\lambda (P_t(z) - P_0(t) - z P_t(z))$$

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

(a) $N(0) = 0$

(b) $N(t)$ has independent increament property.

(c) The probability of an event in a short time interval Δt is

$$p[N(\Delta t) = 1] = \lambda \Delta t + o(\Delta t)$$

(d) The probability of two events or more in the short time interval Δt is $P[N(\Delta t) \geq 2] = o(\Delta t)$ and $o(\Delta t)$ is a infinitly small function of $(\Delta t)^2$ when $\Delta t \to 0$. That is

$$\lim_{\Delta t \to 0} \frac{o(\Delta t)}{\Delta t} = 0$$

Proof. For convenience let us $p[N(t) = n]$ with $P_n(t)$. We need to show that $P_n(t) = \frac{e^{-\lambda t}(\lambda t)^n}{n!}$

Using conditional probability we have:

$$P_n(t+\Delta t) = p(N(t+\Delta t) = n) = \sum_{m=0}^{\infty} p(N(t+\Delta t) = n \mid N(t) = m) \cdot p(N(t) = m)$$

$$= \underbrace{\sum_{m=0}^{n-2} p(N(t+\Delta t) = n \mid N(t) = m) \cdot P_m(t)}_{o(\Delta t)}$$

$$+ \underbrace{p(N(t+\Delta t) = n \mid N(t) = n-1)}_{\mathrm{I}:\ \lambda t + o(\Delta t)} \cdot P_{n-1}(t) + \underbrace{p(N(t+\Delta t) = n \mid N(t) = n)}_{\mathrm{II}:\ 1 - P_1(\Delta t) - P_{\geq 2}(\Delta t)} P_n(t)$$

$$+ \underbrace{\sum_{m=n+1}^{\infty} p(N(t+\Delta t) = n \mid N(t) = m) \cdot P_m(t)}_{\text{zero}}$$

(I) $p(N(t+\Delta t) = n \mid N(t) = n-1) = p(\text{exactly one event in } \Delta t) = \lambda \Delta t + o(\Delta t)$.

(II) $p(N(t+\Delta t) = n \mid N(t) = n) = p(\text{No event in } \Delta t) = 1 - P_1(\Delta t) - P_{n \geq 2}(\Delta t)$

$$= 1 - (\lambda \Delta t + o(\Delta t)) - o(\Delta t) = 1 - \lambda \Delta t + o(\Delta t)$$

Note: $o_1(\Delta t) + o_2(\Delta t) = o_3(\Delta t)$ and $o_1(\Delta) \cdot o_2(\Delta t) = o_3(\Delta t)$

$$P_n(t+\Delta t) = \lambda \Delta t \, P_{n-1}(t) + (1 - \lambda \Delta t) P_n(t) + o(\Delta t)$$

$$\Rightarrow P_n(t+\Delta t) - P_n(t) = \lambda \Delta t \, P_{n-1}(t) - \lambda \Delta t \, P_n(t) + o(\Delta t)$$

$$\Rightarrow \lim_{\Delta t \to 0} \frac{P_n(t+\Delta t) - P_n(t)}{\Delta t} = \lambda P_{n-1}(t) - \lambda P_n(t) + \underbrace{\frac{o(\Delta t)}{\Delta t}}_{0}$$

$$\Rightarrow P_n'(t) = \lambda P_{n-1}(t) - \lambda P_n(t) \quad , n = 0, 1, 2, \ldots$$

Lets define the z-transform of $P_n(t)$ as $p(z) = \sum_{n=0}^{\infty} z^n P_n(t)$

If $n = 0 \Rightarrow P_0'(t) = -\lambda P_0(t) \Rightarrow P_0(t) = e^{-\lambda t} + K$

Using the initial condition $\begin{cases} P_0(0) = 1 \Rightarrow K = 1 \Rightarrow P_0(t) = e^{-\lambda t} \\ P_n(0) = 0 \end{cases}$

$$P_t(z) = \sum_{n=0}^{\infty} z^n P_n(t)$$

$$\begin{cases} P_n'(t) = -\lambda(P_n(t) - P_{n-1}(t)) \quad , n \geq 1 \quad , P_n(0) = 0 \quad , P_0(0) = 1 \\ P_0'(t) = -\lambda P_0(t) \end{cases}$$

$$\sum_{n=1}^{\infty} P_n'(t) z^n = -\lambda \Big(\sum_{n=1}^{\infty + P_0 z^0 - P_0 z^0} P_n(t) z^n - \sum_{n=1}^{\infty} P_{n-1}(t) z^n \Big)$$

$$\Rightarrow \sum_{n=0}^{\infty} P_n'(t) z^n - P_0'(t) = -\lambda (P_t(z) - P_0(t) - z P_t(z))$$

Note: The Z-transform of $P_n'(t)$ is $\frac{\partial P_t(Z)}{\partial(t)}$

$$\Rightarrow \frac{\partial P_t(Z)}{\partial t} = P_0'(t) = -\lambda(P_t(Z) - P_0(t) - Z P(Z))$$

$$\Rightarrow \frac{\partial P_t(Z)}{\partial t} + \lambda(1-Z)P_t(Z) = \underbrace{\lambda P_0(t) + \overset{-\lambda P_0(t)}{P_0'(t)}}_{0}$$

In addition we know that $P_0'(t) = \lambda P_0(t)$. Hence $\frac{dP_t(Z)}{dt} + \lambda(1-Z)P_t(Z) = 0$

$$\Rightarrow \frac{dP_t(Z)}{P_t(Z)} + \lambda(1-Z)dt = 0$$

by integral $\rightarrow \ln P_t(Z) + \lambda(1-Z)t = \ln C \Rightarrow P_t(Z) = e^{\ln C - \lambda(1-Z)t}$

$$\Rightarrow P_t(Z) = e^{\ln c} \cdot e^{-\lambda(1-Z)t} \Rightarrow P_t(Z) = Ce^{-\lambda(1-Z)t}$$

$$\Rightarrow P_t(Z) = Ce^{-\lambda t} e^{\lambda Z t} = e^{-\lambda t} \sum C \frac{(\lambda Z t)^n}{n!} = \sum_{n=0}^{\infty} \frac{Ce^{-\lambda t}(\lambda t)^n}{n!} Z^n$$

$$\Rightarrow \begin{cases} P_n(t) = \frac{Ce^{-\lambda t}(\lambda t)^n}{n!} \\ P_0(0) = 1 \end{cases} \Rightarrow C = 1 \Rightarrow P_n(t) = \frac{e^{-\lambda t}(\lambda t)^n}{n!}$$

Theorem. Assume that customers arrivals are Poisson Process with parameter λ. If exactly one customer has entered within the time interval $(0,t)$, then the arrival distribution of this customer is uniform.

Proof. If t is the arrival time of the customer, then

$$P(t \leqslant x \mid N(t)=1) = \frac{P(T \leqslant x, N(t)=1)^*}{P(N(t)=1)} = \frac{(\lambda x)e^{-\lambda x} e^{-\lambda(t-x)}}{(\lambda t)e^{-\lambda t}} = \frac{x}{t}$$

$* = P($no customer arrive in (x,t) and only one customer in $(0,x))$

Note: The above theorem can be generalized. If we know that up to a certain time n customers have entered the system exactly, then the arrival t... each customer is an uniform distribution.

Example. The number of customers ariving at a sytem is a Poisson proces... parameter $\lambda = 0.5$ customer per hour. If in the first 3 hours, 2 custome... arrived. What is the probability that both have entered the system in the...

$$P = \frac{x}{t} \cdot \frac{y}{t} = \frac{1}{3} \cdot \frac{1}{3} = \frac{1}{9}$$

Erlang Distribution Function

The random variable X has an Erlang density function, if the function follows:

$$f(x) = \lambda e^{-\lambda x} \frac{(\lambda x)^{n-1}}{(n-1)!}, \quad x \geqslant 0$$

where λ and n are parameters of Erlang function and both of them ...

In addition, n is an integer value.

Note: Erlang function is a specific form of Gamma function when n...

In Gamma distribution n can be non-integer and $\Gamma(n)$ is used as follo...

$$\Gamma(n) = \int e^{-u} u^{n-1} du$$

The Erlang r.v. has the mean $E(x) = \frac{n}{\lambda}$, and variance $Var(x) = \frac{n}{\lambda^2}$, ...

$$M_\lambda(t) = \left(\frac{\lambda}{\lambda - t}\right)^n.$$

Note: The Z-transform of $P_n'(t)$ is $\frac{\partial P_t(Z)}{\partial(t)}$

$$\Rightarrow \frac{\partial P_t(Z)}{\partial t} \doteq P_0'(t) = -\lambda(P_t(Z) - P_0(t) - Z P(Z))$$

$$\Rightarrow \frac{\partial P_t(Z)}{\partial t} + \lambda(1-Z)P_t(Z) = \underbrace{\lambda P_0(t) + \overset{-\lambda P_0(t)}{P_0'(t)}}_{0}$$

In addition we know that $P_0'(t) = \lambda P_0(t)$. Hence $\frac{dP_t(Z)}{dt} + \lambda(1-Z)P_t(Z) = 0$

$$\Rightarrow \frac{dP_t(Z)}{P_t(Z)} + \lambda(1-Z)dt = 0$$

by integral → $\ln P_t(Z) + \lambda(1-Z)t = \ln C \Rightarrow P_t(Z) = e^{\ln C - \lambda(1-Z)t}$

$$\Rightarrow P_t(Z) = e^{\ln C}\cdot e^{-\lambda(1-Z)t} \Rightarrow P_t(Z) = Ce^{-\lambda(1-Z)t}$$

$$\Rightarrow P_t(Z) = Ce^{-\lambda t}e^{\lambda Z t} = e^{-\lambda t}\sum C\cdot\frac{(\lambda Z t)^n}{n!} = \sum_{n=0}^{\infty}\frac{Ce^{-\lambda t}(\lambda t)^n}{n!}Z^n$$

$$\Rightarrow \begin{cases} P_n(t) = \frac{Ce^{-\lambda t}(\lambda t)^n}{n!} \\ P_0(0) = 1 \end{cases} \Rightarrow C = 1 \Rightarrow P_n(t) = \frac{e^{-\lambda t}(\lambda t)^n}{n!}$$

Theorem. Assume that customers arrivals are Poisson Process with parameter λ. If exactly one customer has entered within the time interval $(0,t)$, then the arrival distribution of this customer is uniform.

Proof. If t is the arrival time of the customer, then

$$P(t \leq x \mid N(t)=1) = \frac{P(T \leq x, N(t)=1)^*}{P(N(t)=1)} = \frac{(\lambda x)e^{-\lambda x}e^{-\lambda(t-x)}}{(\lambda t)e^{-\lambda t}} = \frac{x}{t}$$

$*$ = P(no customer arrive in (x,t) and only one customer in $(0,x)$)

Note: The above theorem can be generalized. If we know that up to a certain time n customers have entered the system exactly, then the arrival time of each customer is an uniform distribution.

Example. The number of customers arriving at a system is a Poisson process with parameter $\lambda = 0.5$ customer per hour. If in the first 3 hours, 2 customers have arrived. What is the probability that both have entered the system in the first hour?

$$P = \frac{x}{t}\cdot\frac{y}{t} = \frac{1}{3}\cdot\frac{1}{3} = \frac{1}{9}$$

Erlang Distribution Function

The random variable X has an Erlang density function, if the function is like follows:

$$f(x) = \lambda e^{-\lambda x}\frac{(\lambda x)^{n-1}}{(n-1)!}, \quad x \geq 0$$

where λ and n are parameters of Erlang function and both of them are positive. In addition, n is an integer value.

Note: Erlang function is a specific form of Gamma function when n is integer. In Gamma distribution n can be non-integer and, $\Gamma(n)$ is used as follows:

$$\Gamma(n) = \int e^{-u}u^{n-1}du$$

The Erlang r.v. has the mean $E(x) = \frac{n}{\lambda}$, and variance $Var(x) = \frac{n}{\lambda^2}$, and also $M_X(t) = \left(\frac{\lambda}{\lambda - t}\right)^n$.

Theorem. The sum of n independent exponential r.v. with parameter λ is a r.v. with Erlang density function with parameters (n, λ).

Proof. Assume that $X = X_1 + X_2 + \cdots + X_n$, $X_i \sim Exp(\lambda)$

$$M_X(t) = M_{\sum X_i}(t) = \prod_{i=1}^{n} M_{X_i}(t) = \prod_{i=1}^{n} \left(\frac{\lambda}{\lambda - t}\right)^n$$

Theorem. If the arrival of customers are based on a Poisson process with parameter λ, the the arrival time of the n^{th} customer is a r.v. with Erlang (λ, n).

Proof. We know that the arrival time of the n^{th} customer S_n is

$$S_n = T_1 + T_2 + \cdots + T_n$$

where T_i is the inter arrival time between customers $i-1$ and i.

Based on a theorem we know that $T_i \sim Exp(\lambda)$ and also based on the previous theorem $\sum T_i \sim Erlang(n, \lambda)$

Usually, when dealing with problems that involves Erlang distribution, we need to deal with quite complicated integrals. The following theorem can reduce the complication.

Theorem. The CDF of the Erlang distribution with parameters (n, λ) can be obtained as follows.

$$F(x) = (1 - e^{-\lambda x}) \sum_{k=0}^{n-1} \frac{(\lambda x)^k}{k!}$$

Proof.

$$F_X(x) = P(X \leqslant x) = \int_0^x f(t)\,dt = \int_0^x \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!}\,dt = 1 - \int_x^\infty \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!}$$

using the change of variable $y = t - x$, gives

$$F(x) = 1 - \int_0^\infty \frac{\lambda^n e^{-\lambda(x+y)} (x+y)^{n-1}}{(n-1)!}\,dy$$

Knowing that $(a+b)^{n-1} = \sum_{k=0}^{n-1} \binom{n-1}{k} x^k y^{n-k-1}$, so we have

$$F_X(x) = 1 - e^{-\lambda x} \int_0^\infty \frac{\lambda e^{-\lambda y}}{(n-1)!} \sum_{k=0}^{n-1} \binom{n-1}{k} x^k y^{n-k-1}\,dy$$

$$= 1 - e^{-\lambda x} \int_0^\infty \lambda^n e^{-\lambda y} \sum_{k=0}^{n-1} \frac{x^k y^{n-k-1}}{k!\,(n-k-1)!}\,dy$$

$$= 1 - e^{-\lambda x} \sum_{k=0}^{n-1} \frac{(\lambda x)^k}{k!\,(n-k-1)!} \underbrace{\int_0^\infty \lambda e^{-\lambda y} (\lambda y)^{n-k-1}\,dy}_{I}$$

$u = \lambda y \rightarrow du = \lambda\,dy$, then $I = \int e^{-u} u^{(n-k-1)}\,dy = \Gamma(n-k) = (n-k-1)!$

$$\Rightarrow F_X(x) = 1 - e^{-\lambda x} \sum_{k=0}^{n-1} \frac{(\lambda x)^k}{k!}$$

Example. Consider a queuing system with a single server. The servi[ce] is exponential with the average of 20 mins. Assume that a customer [enters] the system and faces that one other person waiting in the queue and [one] customer is being served. What is the probability that this customer wa[its more] than 1 hour?

Theorem. The sum of n independent exponential r.v. with parameter λ is a r.v. with Erlang density function with parameters (n, λ).

Proof. Assume that $X = X_1 + X_2 + \cdots + X_n$, $X_i \sim Exp(\lambda)$

$$M_X(t) = M_{\sum X_i}(t) = \prod_{i=1}^{n} M_{X_i}(t) = \prod_{i=1}^{n} \left(\frac{\lambda}{\lambda - t}\right)^n$$

Theorem. If the arrival of customers are based on a Poisson process with parameter λ, the the arrival time of the n^{th} customer is a r.v. with Erlang (λ, n).

Proof. We know that the arrival time of the n^{th} customer S_n is

$$S_n = T_1 + T_2 + \cdots + T_n$$

where T_i is the inter arrival time between customers $i-1$ and i.

Based on a theorem we know that $T_i \sim Exp(\lambda)$ and also based on the previous theorem $\sum T_i \sim$ Erlang (n, λ)

Usually, when dealing with problems that involves Erlang distribution, we need to deal with quite complicated integrals. The following theorem can reduce the complication.

Theorem. The CDF of the Erlang distribution with parameters (n, λ) can be obtained as follows.

$$F(x) = (1 - e^{-\lambda x}) \sum_{k=0}^{n-1} \frac{(\lambda x)^k}{k!}$$

Proof.

$$F_X(x) = P(X \leqslant x) = \int_0^x f(t)\,dt = \int_0^x \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!}\,dt = 1 - \int_x^\infty \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!}$$

using the change of variable $y = t - x$, gives

$$F(x) = 1 - \int_0^\infty \frac{\lambda^n e^{-\lambda(x+y)} (x+y)^{n-1}}{(n-1)!}\,dy$$

Knowing that $(a+b)^{n-1} = \sum_{k=0}^{n-1} \binom{n-1}{k} x^k y^{n-k-1}$, so we have

$$F_X(x) = 1 - e^{-\lambda x} \int_0^\infty \frac{\lambda e^{-\lambda y}}{(n-1)!} \sum_{k=0}^{n-1} \binom{n-1}{k} x^k y^{n-k-1}\,dy$$

$$= 1 - e^{-\lambda x} \int_0^\infty \lambda^n e^{-\lambda y} \sum_{k=0}^{n-1} \frac{x^k y^{n-k-1}}{k!(n-k-1)!}\,dy$$

$$= 1 - e^{-\lambda x} \sum_{k=0}^{n-1} \frac{(\lambda x)^k}{k!(n-k-1)!} \underbrace{\int_0^\infty \lambda e^{-\lambda y} (\lambda y)^{n-k-1}\,dy}_{I}$$

$u = \lambda y \rightarrow du = \lambda\,dy$, then $I = \int e^{-u} u^{(n-k-1)}\,dy = \Gamma(n-k) = (n-k-1)!$

$$\Rightarrow F_X(x) = 1 - e^{-\lambda x} \sum_{k=0}^{n-1} \frac{(\lambda x)^k}{k!}$$

Example. Consider a queuing system with a single server. The service time is exponential with the average of 20 mins. Assume that a customer arrives at the system and faces that one other person waiting in the queue and another customer is being served. What is the probability that this customer waits more than 1 hour?

$$X = X_1 + X_2 \sim \text{Erlang}(2, \overset{\lambda=3}{\lambda})$$

$$\mathbb{P}(X \geq 1) = e^{-\mu}\sum_{k=0}^{1}\frac{\mu^k}{k!} = e^{-3}\left[\frac{\mu^0}{0!} + \frac{\mu}{1}\right] = 4e^{-3}$$

From the preceding theorem,

$$F(x) = 1 - \sum_{k=0}^{n-1}\frac{e^{-\lambda x}(\lambda x)^k}{k!} \equiv \mathbb{P}(S_n \leq x) \Longleftrightarrow \mathbb{P}(N(x) \geq n)$$

That is the relastionship between Erlang and Poisson distribution.

Markov Chains

Markov Process;

Stochastic Process: is a set of $\{X(t),\ t \geq 0\}$ where $X(t)$ is the state of process at time t, for every $t \in T$, $X(t)$ is a random variable and T is called index set.

If we devide the time into three phases past, present and future, Hence in Markov Process the future is independent of past and it only depends on the present state of system.

Example. Poisson process is a Markov process because the number of events in the future is independent of the past.

The Markov property can be shown more formally as follows. Consider a set of r.vs. $\{X(t),\ t \geq 0\}$, If $X(t)$ operates according to a Markov process, then for all $x_1, x_2, \dots, x_n$:

$$\mathbb{P}(X(t_{n+1}) \leqslant x \mid X(t_n) = x_n, \dots, X(t_2) = x_2, X(t_1) = x_1) = \mathbb{P}(X(t_{n+1}) \leqslant x \mid X(t_n) = x_n)$$

The Markov processes are grouped based on two factors:

A) The parameter t which can be either continuous or discrete. The discreteness of t means that we study the states of the system at certain times. If t is discrete, we usually show $X(t)$ as $X_1, X_2, \dots, X_n$.

B) The set of values for $X(t)$ which can be either continuous or discrete. $X(t)$ is called the state of system.

Note: Markov chains are special cases of Markov process in which both t and $X(t)$ are discrete.

Markov Chains:

The r.v.s. $X_1, X_2, \dots, X_n$ are a Markov chain if for all n and all states i and j :

$$\mathbb{P}(X_{n+1} = j \mid X_1 = i_1, X_2 = i_2, \dots, X_n = i_n) = \mathbb{P}(X_{n+1} = j \mid X_n = i_n)$$

where n is called "stage" or "time" of the process.

Example. Consider a point on the x-axis where the point can move forward one unit on the axis with the probability p and move backward with probability $(1-p)$. Then $X(t)$ (the position of this point on the ...) forms a Markov chain, because the next state depends only on the ... state and is independent of the past. The probability of the states' tr... from i to j is:

$$\begin{cases} \mathbb{P}(X_{n+1} = i+1 \mid X_n = i) = p \\ \mathbb{P}(X_{n+1} = i-1 \mid X_n = i) = 1-p \\ \mathbb{P}(X_{n+1} = j \mid X_n = i) = 0 \quad \text{for } j \neq i+1 \text{ or } j = i-1 \end{cases}$$

Homogeneous Markov chains: are special cases of MC where ... transition probability from one state to another state is independent ... stage or time n. That is:

$$\mathbb{P}(X_{n+1} = j \mid X_n = i) = \mathbb{P}(X_1 = j \mid X_0 = i)$$

Hence, in a homogeneous MC, P_{ij} is the probability that the stat... system makes a transition from state i to j.

Transition Matrix: is a matrix where its elements in row i and ... is P_{ij}. If we assume that the number of states is M, then

$$P = \begin{pmatrix} P_{11} & P_{12} & \cdots & P_{1M} \\ P_{21} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ P_{M1} & \cdots & \cdots & P_{MM} \end{pmatrix}, \quad P_{ij} \geqslant 0, \quad \sum_j P_{ij} = 1 \; ; \; i ...$$

all $x_1, x_2, \dots, x_n$:

$$\mathbb{P}(X(t_{n+1}) \leqslant x \mid X(t_n)=x_n, \dots, X(t_2)=x_2, X(t_1)=x_1) = \mathbb{P}(X(t_{n+1}) \leqslant x \mid X(t_n)=x_n)$$

The Markov processes are grouped based on two factors:

A) The parameter t which can be either continuous or discrete. The discreteness of t means that we study the states of the system at certain times. If t is discrete, we usually show $X(t)$ as $X_1, X_2, \dots, X_n$.

B) The set of values for $X(t)$ which can be either continuous or discrete. $X(t)$ is called the state of system.

Note: Markov chains are special cases of Markov process in which both t and $X(t)$ are discrete.

Markov Chains:

The r.v.s. $X_1, X_2, \dots, X_n$ are a Markov chain if for all n and all states i and j :

$$\mathbb{P}(X_{n+1}=j \mid X_1=i_1, X_2=i_2, \dots, X_n=i_n) = \mathbb{P}(X_{n+1}=j \mid X_n=i_n)$$

where n is called "stage" or "time" of the process.

Example. Consider a point on the x-axis where the point can move forward one unit on the axis with the probability p and move backward with probability $(1-p)$. Then $X(t)$ (the position of this point on the x-axis) forms a Markov chain, because the next state depends only on the current state and is independent of the past. The probability of the states' transition from i to j is:

$$\begin{cases} \mathbb{P}(X_{n+1}=i+1 \mid X_n=i) = p \\ \mathbb{P}(X_{n+1}=i-1 \mid X_n=i) = 1-p \\ \mathbb{P}(X_{n+1}=j \mid X_n=i) = 0 \quad \text{for } j \neq i+1 \text{ or } j=i-1 \end{cases}$$

Homogeneous Markov chains: are special cases of MC where the transition probability from one state to another state is independent of the stage or time n. That is:

$$\mathbb{P}(X_{n+1}=j \mid X_n=i) = \mathbb{P}(X_1=j \mid X_0=i)$$

Hence, in a homogeneous MC, P_{ij} is the probability that the state of the system makes a transition from state i to j.

Transition Matrix: is a matrix where its elements in row i and column j is P_{ij}. If we assume that the number of states is M, then

$$P = \begin{pmatrix} P_{11} & P_{12} & \cdots & P_{1M} \\ P_{21} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ P_{M1} & \cdots & \cdots & P_{MM} \end{pmatrix}, \quad P_{ij} \geqslant 0, \ \sum_j P_{ij} = 1 \ ; \ i=1,\dots,M$$

Example. For the previous example TM is:

$$P=\begin{array}{c} \begin{array}{ccc} i-1 & i & i+1 \end{array} \\ \begin{bmatrix} \ddots & & & & \\ \cdots q & 0 & p \cdots & & \\ & \cdots q & 0 & p \cdots & \\ & & & & \ddots \end{bmatrix} \end{array}$$

the dimantion is infinite.

Note: Generally any matrix where all its elements are positive and the sum of elements in every row is equal to 1 is called "Markov matrix".

Example. Suppose a person's income per day is y where y is a non-negative r.v. and the probility that $P(y=i)=a_i$. It is clear that $\sum_{i=0}^{\infty} a_i = 1$, $a_i \geq 0$. If X_n is the sum of the individual's income in the first n days, then X_n is a Markov chain because by knowing his income in the first n days we can obtain the person's income for the first $(n+1)$ days as follows. The TM for this MC is:

$$P=\begin{bmatrix} a_0 & a_1 & a_2 & \cdots \\ 0 & a_0 & a_1 & a_2 & \cdots \\ 0 & 0 & a_0 & \cdots \\ \vdots & \vdots & & \ddots \end{bmatrix}$$

m-step Transition Matrix:

Using TM we can identify the relation between two successive st... another words, by knowing the state of the system at any stage we c... the probability distribution of the system states for the next stage. ... that we want to see what is the relationship between the system ... this stage and the state of the system at m stages or m time u... That is, If the state of the system at the stage n is $X_n=i$, wha... probability of $X_{n+m}=j$

Definitions: A) $P_{ij}^{(m)}$ is the probability of the systems state transiti... i to j in m-steps that is:

$$P_{ij}^{(m)}=(X_m=j \mid X_0=i)$$

B) $P^{(m)}$ is the m-step TM:

$$P^{(m)}=\begin{bmatrix} P_{11}^{(m)} & \cdots & P_{1M}^{(m)} \\ \vdots & \ddots & \vdots \\ P_{M1}^{(m)} & \cdots & P_{MM}^{(m)} \end{bmatrix}$$

Theorem. For all i and j, and all K and K' we have the following

a) $P_{ij}^{(K+K')}=\sum_{s=0}^{\infty} P_{is}^{(K)} P_{sj}^{(K')}$ (Chapman - Kolmogrov)

Example. For the previous example TM is:

$$P = \begin{array}{c} \begin{array}{ccc} i-1 & i & i+1 \end{array} \\ \begin{bmatrix} \cdots & q & 0 & p & \cdots & \\ & \cdots & q & 0 & p & \cdots \end{bmatrix} \end{array}$$

the dimantion is infinite.

Note: Generally any matrix where all its elements are positive and the sum of elements in every row is equal to 1 is called "Markov matrix".

Example. Suppose a person's income per day is y where y is a non-negative r.v. and the probility that $P(y=i)=a_i$. It is clear that $\sum_{i=0}^{\infty} a_i = 1$, $a_i \geq 0$. If X_n is the sum of the individual's income in the first n days, then X_n is a Markov chain because by knowing his income in the first n days we can obtain the person's income for the first $(n+1)$ days as follows. The TM for this MC is:

$$P = \begin{bmatrix} a_0 & a_1 & a_2 & \cdots \\ 0 & a_0 & a_1 & a_2 \\ 0 & 0 & a_0 & \ddots \\ \vdots & \vdots & & \end{bmatrix}$$

m-step Transition Matrix:

Using TM we can identify the relation between two successive stages. In another words, by knowing the state of the system at any stage we can identify the probability distribution of the system states for the next stage. Assume that we want to see what is the relationship between the system state at this stage and the state of the system at m stages or m time unit later. That is, If the state of the system at the stage n is $X_n = i$, what is the probability of $X_{n+m} = j$

Definitions: A) $P_{ij}^{(m)}$ is the probability of the systems state transition from i to j in m-steps that is:

$$P_{ij}^{(m)} = (X_m = j \mid X_0 = i)$$

B) $P^{(m)}$ is the m-step TM:

$$P^{(m)} = \begin{bmatrix} P_{11}^{(m)} & \cdots & P_{1M}^{(m)} \\ \vdots & \ddots & \vdots \\ P_{M1}^{(m)} & \cdots & P_{MM}^{(m)} \end{bmatrix}$$

Theorem. For all i and j, and all K and K' we have the following relationships:

a) $$P_{ij}^{(K+K')} = \sum_{s=0}^{\infty} P_{is}^{(K)} P_{sj}^{(K')} \quad \text{(Chapman – Kolmogrov)}$$

b) $\mathbb{P}^{(m)} = \underbrace{\mathbb{P}\cdot\mathbb{P}\cdot\dots\cdot\mathbb{P}}_{m} = \mathbb{P}^{m}$

Proof. Using the conditional probability we have:

$$P_{ij}^{(k+k')} = \mathbb{P}(X_{k+k'}=j \mid X_0=i) = \sum_{s=0}^{\infty} \mathbb{P}(X_{k+k'}=j \mid X_k=s, \cancel{X_0=i})\cdot \mathbb{P}(X_k=s \mid X_0=i)$$

Markov property

$$= \sum P_{sj}^{k'} \cdot P_{is}^{k}$$

We can also use $\mathbb{P}^{(k+k')} = \mathbb{P}^{(k)} \cdot \mathbb{P}^{(k')}$ $(*)$

Since the relationship $(*)$ is true with every value of k and k', hence:

$$\begin{cases} P^{(m)} = P\cdot P^{(m-1)} \\ P^{(m-1)} = P\cdot P^{(m-2)} \\ \vdots \\ P^{(2)} = P\cdot P \end{cases} \Rightarrow P^{(m)} = P\cdot P\cdot\dots\cdot P = P^{m}$$

Example. Consider a MC with the following TM:

$$P = \begin{bmatrix} 0.75 & 0.25 & 0 \\ 0.25 & 0.5 & 0.25 \\ 0 & 0.75 & 0.25 \end{bmatrix}$$

obtain $P_{13}^{(2)}$ and $P_{31}^{(2)}$.

$$P^{(2)} = P\cdot P = \begin{bmatrix} 0.625 & 0.3125 & 0.0625 \\ 0.3125 & 0.5 & 0.1875 \\ 0.1875 & 0.5625 & 0.25 \end{bmatrix} \Rightarrow P_{13}^{(2)} = 0.0625,\ P_{31}^{(2)} = 0.1875$$

The second way:

$$P_{13}^{(2)} = P_{11}P_{13} + P_{12}P_{23} + P_{13}\cdot P_{33} = [P_{11}\ P_{12}\ P_{13}]\begin{bmatrix} P_{13} \\ P_{23} \\ P_{33} \end{bmatrix}$$

Obtaining the distribution of the system state for each state:

We may need to know the probability distribution of the system state

is $P(X_n=j)$ for all value of j and n. This probability is denoted by π

probability distribution of the system state at stage 0 is $P(X_0=i)$ th

conditional probability we have:

$$\pi_j^{(n)} = P(X_n=j) = \sum P(X_n=j \mid X_0=i)\, P(X_0=i) = \sum P_{ij}^{(n)} \pi_i^{(0)}$$

Or $\pi_j^{(n)} = \pi^{(0)} \times \mathbb{P}^{(n)}$ (Matrix form)

$$[\pi_1^{(n)}, \pi_2^{(n)}, \dots, \pi_M^{(n)}] = [\pi_1^{(0)}, \pi_2^{(0)}, \dots, \pi_M^{(0)}]\,[P_{ij}^{(n)}]_{M\times M}$$

Where $\pi_i^{(0)}$ and $\pi_j^{(n)}$ are the vectors that their elements are π_i

Obtaining the probability of a path:

$$P(X_1=j, X_2=k \mid X_0=i) = P(X_2=k \mid X_1=j, X_0=i)\cdot P(X_1=j \mid X_0=$$

$$= P_{ij}\cdot P_{jk}$$

Example. In the previous example find the following probabilities

$A_1 = P(X_3=2 \mid X_0=1) = P_{12}^{(3)} = \mathbb{P}\cdot\mathbb{P}^{(2)} = 0.3594$ $\qquad \mathbb{P}^{(3)} = \begin{bmatrix} .5469 \\ .3594 \\ .2812 \end{bmatrix}$

$A_2 = P(X_3=2) = \sum_{i=1}^{3} P(X_3=2 \mid X_0=i)\, P(X_0=i) = \frac{1}{3}(0.3594 + 0.4687 + 0.575$

$A_3 = P(X_3=2, X_2=3, X_1=2 \mid X_0=1) = P_{12}\cdot P_{23}\cdot P_{32} = (0.25)(0.25)(0.75) = 3$

$A_4 = P(X_3=2, X_2=3, X_1=2) = \sum_{i=1}^{3} P(X_3=2, X_2=3, X_1=2 \mid X_0=i)\, P(X_0=i)$

b) $\mathbb{P}^{(m)} = \underbrace{\mathbb{P}\cdot\mathbb{P}\cdots\mathbb{P}}_{m} = \mathbb{P}^m$

Proof. Using the conditional probability we have:

$$P_{ij}^{(k+k')} = \mathbb{P}(X_{k+k'}=j \mid X_0=i) = \sum_{s=0}^{\infty} \mathbb{P}(X_{k+k'}=j \mid X_k=s, X_0=i)\cdot \mathbb{P}(X_k=s \mid X_0=i)$$

Markov property

$$= \sum P_{sj}^{k'} \cdot P_{is}^{k}$$

We can also use $\mathbb{P}^{(k+k')} = \mathbb{P}^{(k)} \cdot \mathbb{P}^{(k')} \quad (*)$

Since the relationship $(*)$ is true with every value of k and k', hence:

$$\left\{\begin{array}{l} P^{(m)} = P \cdot P^{(m-1)} \\ P^{(m-1)} = P \cdot P^{(m-2)} \\ \vdots \\ P^{(2)} = P \cdot P \end{array}\right. \Rightarrow P^{(m)} = P \cdot P \cdots P = P^m$$

Example. Consider a MC with the following TM:

$$P = \begin{bmatrix} 0.75 & 0.25 & 0 \\ 0.25 & 0.5 & 0.25 \\ 0 & 0.75 & 0.25 \end{bmatrix}$$

obtain $P_{13}^{(2)}$ and $P_{31}^{(2)}$.

$$P^{(2)} = P \cdot P = \begin{bmatrix} 0.625 & 0.3125 & 0.0625 \\ 0.3125 & 0.5 & 0.1875 \\ 0.1875 & 0.5625 & 0.25 \end{bmatrix} \Rightarrow P_{13}^{(2)} = 0.0625,\ P_{31}^{(2)} = 0.1875$$

The second way:

$$P_{13}^{(2)} = P_{11}P_{13} + P_{12}P_{23} + P_{13}\cdot P_{33} = [P_{11}\ P_{12}\ P_{13}]\begin{bmatrix} P_{13} \\ P_{23} \\ P_{33} \end{bmatrix}$$

Obtaining the distribution of the system state for each state:

We may need to know the probability distribution of the system state (n). That is $P(X_n=j)$ for all value of j and n. This probability is denoted by $\pi_j^{(n)}$. If the probability distribution of the system state at stage 0 is $P(X_0=i)$ then using conditional probability we have:

$$\pi_j^{(n)} = P(X_n=j) = \sum P(X_n=j \mid X_0=i)\, P(X_0=i) = \sum P_{ij}^{(n)} \pi_i^{(0)}$$

Or $\pi_j^{(n)} = \pi^{(0)} \times \mathbb{P}^{(n)}$ (Matrix form)

$$[\pi_1^{(n)}, \pi_2^{(n)}, \ldots, \pi_M^{(n)}] = [\pi_1^{(0)}, \pi_2^{(0)}, \ldots, \pi_M^{(0)}]\,[P_{ij}^{(n)}]_{M\times M}$$

Where $\pi_i^{(0)}$ and $\pi_j^{(n)}$ are the vectors that their elements are $\pi_i^{(0)}$ and $\pi_j^{(n)}$.

Obtaining the probability of a path:

$$P(X_1=j, X_2=k \mid X_0=i) = P(X_2=k \mid X_1=j, X_0=i)\cdot P(X_1=j \mid X_0=i)$$

$$= P_{ij} \cdot P_{jk}$$

Example. In the previous example find the following probabilities:

$$A_1 = P(X_3=2 \mid X_0=1) = P_{12}^{(3)} = \mathbb{P}\cdot\mathbb{P}^{(2)} = 0.3594 \qquad \mathbb{P}^{(3)} = \begin{bmatrix} .5469 & .3594 & .0937 \\ .3594 & .4687 & .1719 \\ .2812 & .5156 & .2031 \end{bmatrix}$$

$$A_2 = P(X_3=2) = \sum_{i=1}^{3} P(X_3=2 \mid X_0=i)\, P(X_0=i) = \tfrac{1}{3}(0.3594 + 0.4687 + 0.5156)$$

$$A_3 = P(X_3=2, X_2=3, X_1=2 \mid X_0=1) = P_{12}\cdot P_{23}\cdot P_{32} = (0.25)(0.25)(0.75) = 3/64$$

$$A_4 = P(X_3=2, X_2=3, X_1=2) = \sum_{i=1}^{3} P(X_3=2, X_2=3, X_1=2 \mid X_0=i)\, P(X_0=i)$$

$= \frac{1}{3}\left(\sum P(X_3=2, X_2=3, X_1=2 \mid X_0=i)\right) = \frac{1}{3}\sum_{i=1}^{3} P_{i2}\, P_{23}\, P_{32} = 6/64$

Classification of states:

The aim of classifying the system states is to study the limiting probabilities.

Definitions:

1) The state j is accessible from state i, if there is a positive probability that the system can reach j from i in a number of steps or time units.

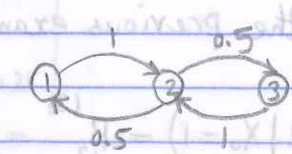
That is, j is accessible from i if for some n, $P_{ij}^{(n)} > 0$. $(i \to j)$

2) If i is accessible from j and j is also accessible from i, then we say that the two states i and j communicate, and denote with $(i \leftrightarrow j)$ double arrow

It is clear that $P_{ii}^{(0)} = 1$, which implies that each state communicate with itself $(i \leftrightarrow i)$

Example. consider a MC with the TM.

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



$\to 1 \leftrightarrow 2,\ 2 \leftrightarrow 3,\ 1 \leftrightarrow 3$

3) A) The set of all states that communicate to each other is called a "class".

B) If the system contains only one class, it is called "irreducible".

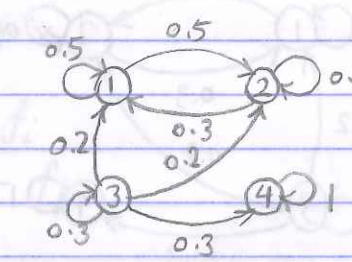
C) The set of the states are called "closed", if there is no way of reaching from any state in the set to any state out of the set. That is, $P_{ij}^{(n}$ all n, $i \in$ set, $j \notin$ set.

D) If a single state forms a closed set, then it is called an "absorb like black

That is $P_{ii} = 1$

Example.

$$P = \begin{bmatrix} 0.5 & 0.5 & 0 & 0 \\ 0.3 & 0.7 & 0 & 0 \\ 0.2 & 0.2 & 0.3 & 0.3 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$



$\{1,2\}, \{4\}$ are closed classes.

$\{3\}$ is not closed.

There are 3 classes $\{1,2\}, \{4\}, \{3\}$ where $\{4\}$ is an absorbing state

Note: An irreducible MC is a closed set of states where non of its of states can form a closed set.

Theorem. If $i \leftrightarrow j$ and $j \leftrightarrow k$, then $i \leftrightarrow k$.

Proof. Since $i \leftrightarrow j$, then there is some n and n' such that $P_{ij}^{(n)} > 0$,

Similarly $\exists\ m, m' > 0$ so that $P_{jk}^{(m)} > 0$ and $P_{kj}^{(m')} > 0$, hence,

$$P_{ik}^{(n+m)} = \sum P_{is}^{(n)} P_{sk}^{(m)} > P_{ij}^{(n)} P_{jk}^{(m)} > 0 \Rightarrow i \to k$$

Similarly, it can be shown that $k \to i$.

$= \frac{1}{3}\left(\sum P(X_3=2, X_2=3, X_1=2 \mid X_0=i)\right) = \frac{1}{3}\sum_{i=1}^{3} P_{i2}\, P_{23}\, P_{32} = 6/64$

Classification of states:

The aim of classifying the system states is to study the limiting probabilities.

Definitions:

1) The state j is accessible from state i, if there is a positive probability that the system can reach j from i in a number of steps or time units.

That is, j is accessible from i if for some n, $P_{ij}^{(n)} > 0$. $(i \to j)$

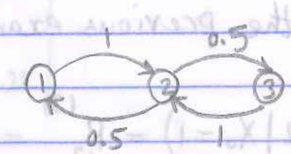
2) If i is accessible from j and j is also accessible from i, then we say that the two states i and j communicate, and denote with $(i \leftrightarrow j)$

double arrow

It is clear that $P_{ii}^{(0)} = 1$, which implies that each state communicate with itself. $(i \leftrightarrow i)$

Example. consider a MC with the TM.

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



$\to 1 \leftrightarrow 2,\ 2 \leftrightarrow 3,\ 1 \leftrightarrow 3$

3) A) The set of all states that communicate to each other is called a "class".

B) If the system contains only one class, it is called "irreducible".

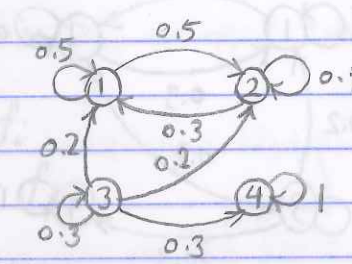
C) The set of the states are called "closed", if there is no way of reaching from any state in the set to any state out of the set. That is, $P_{ij}^{(n)} = 0$ for all n, $i \in$ set, $j \notin$ set.

D) If a single state forms a closed set, then it is called an "absorbing state".
like black

That is $P_{ii} = 1$

Example.

$$P = \begin{bmatrix} 0.5 & 0.5 & 0 & 0 \\ 0.3 & 0.7 & 0 & 0 \\ 0.2 & 0.2 & 0.3 & 0.3 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$



$\{1,2\}, \{4\}$ are closed classes.

$\{3\}$ is not closed.

There are 3 classes $\{1,2\}, \{4\}, \{3\}$ where $\{4\}$ is an absorbing state.

Note: An irreducible MC is a closed set of states where non of its subsets of states can form a closed set.

Theorem. If $i \leftrightarrow j$ and $j \leftrightarrow k$, then $i \leftrightarrow k$.

Proof. Since $i \leftrightarrow j$, then there is some n and n' such that $P_{ij}^{(n)} > 0$, $P_{ji}^{(n')} > 0$.

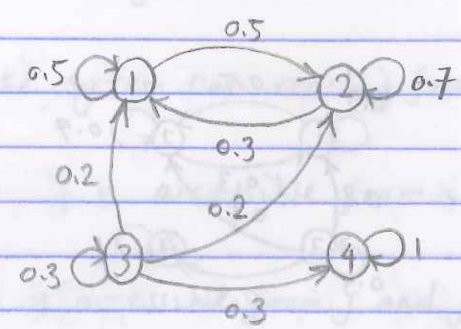
Similarly $\exists\ m, m' > 0$ so that $P_{jk}^{(m)} > 0$ and $P_{kj}^{(m')} > 0$, hence,

$$P_{ik}^{(n+m)} = \sum P_{is}^{(n)} P_{sk}^{(m)} > P_{ij}^{(n)} P_{jk}^{(m)} > 0 \implies i \to k$$

Similarly, it can be shown that $k \to i$.

Definition. Suppose that the state of the system is i, denote the probability that the system returns to i at some time with f_i. If $f_i = 1$, then the state i is called "recurrent", otherwise, if $f_i < 1$, then it is called "transient", if starts from i.

Example.



$f_1 = \sum_{n=1}^{\infty} f_1^{(n)} = f_1^{(1)} + f_1^{(2)} + \cdots$

$= 0.5 + 0.5 \times 0.3 + 0.5 \times 0.7 \times 0.3 + \cdots$

$= 0.5 + 0.5 \times 0.3 + 0.5 \times 0.7 \times 0.3$

$+ 0.5 \times (0.7)^2 (0.3) + \cdots$

$= 0.5 + (0.5)(0.3)[1 + 0.7 + (0.7)^2 + \cdots] = 0.5 + (0.5)(0.3)\frac{1}{1-0.7} = 1$ So state 1 is recurrent.

$f_2 = \sum_{n=1}^{\infty} f_2^{(n)} = 1 \rightarrow$ state 2 is recurrent.

$f_3 = f_3^{(1)} + f_3^{(2)} + \cdots = 0.3 + 0 + 0 \cdots = 0.3 \rightarrow$ state 3 is transient.

The expected number of visits:

Based on the definition of f_i, if the state i is recurrent and the system enters this state, it will surely return to i at some time in the future. Hence, the process starts from i again and again, So there will be some other time that the process will surely visit state i, and so on. Therefore,

the number of visits to state i is infinite. However, if i is transien
Starting from i, there is a probability f_i that the system will retur
at some times and $(1-f_i)$ that the system will never come back any
the probability that the system will return to i in n stages later a
never return anymore is $f_i^n(1-f_i)$, which is a geometric distribut
expected number of visits, $E(N)$ is

$$E(N) = \frac{f_i}{1-f_i}$$

which is finite.

Theorem. In a MC if for a state i, $E(N) = \sum_{n=1}^{\infty} P_{ii}^{(n)} = \infty$, then th
i is recurrent, otherwise $\sum P_{ii}^{(n)} < \infty$, then i is transient.

Proof. $E[R_{ii}^{(n)}] = 1 \times P_{ii}^{(n)} + 0\,(1 - P_{ii}^{(n)}) = P_{ii}^{(n)}$

$$R_{ii}^{(n)} = \begin{cases} 1, & \text{if after } n \text{ time units the system visits } i \\ 0, & \text{o.w.} \end{cases}$$

$$E(N) = E(R_{ii}) = \sum_{n=1}^{\infty} E(R_{ii}^{(n)}) = \sum_{n=1}^{\infty} P_{ii}^{(n)}$$

Example.

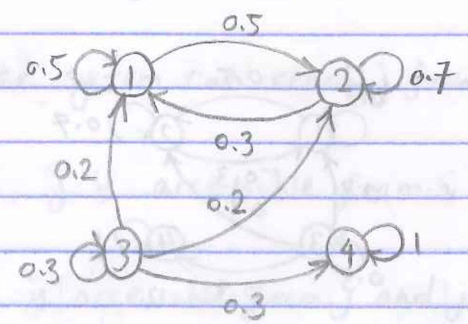
$P_{33}^{(1)} = 0.3$

$P_{33}^{(2)} = P_{31}P_{13} + P_{32}P_{23} + P_{33}P_{33} + P_{34}P_{43} = 0.3^2$

$P_{33}^{(3)} = 0.3^3$

Definition. Suppose that the state of the system is i, denote the probability that the system returns to i at some time with f_i. If $f_i = 1$, then the state i is called "recurrent", otherwise, if $f_i < 1$, then it is called "transient", if starts from i.

Example.



$$f_1 = \sum_{n=1}^{\infty} f_1^{(n)} = f_1^{(1)} + f_1^{(2)} + \cdots$$

$$= 0.5 + 0.5\times 0.3 + 0.5\times 0.7 \times 0.3 + \cdots$$

$$= 0.5 + 0.5\times 0.3 + 0.5\times 0.7\times 0.3$$

$$+ 0.5\times(0.7)^2(0.3) + \cdots$$

$$= 0.5 + (0.5)(0.3)[1 + 0.7 + (0.7)^2 + \cdots] = 0.5 + (0.5)(0.3)\frac{1}{1-0.7} = 1$$ So state 1 is recurrent.

$f_2 = \sum_{n=1}^{\infty} f_2^{(n)} = 1 \rightarrow$ state 2 is recurrent.

$f_3 = f_3^{(1)} + f_3^{(2)} + \cdots = 0.3 + 0 + 0 \ldots = 0.3 \rightarrow$ state 3 is transient.

The expected number of visits:

Based on the definition of f_i, if the state i is recurrent and the system enters this state, it will surely return to i at some time in the future. Hence, the process starts from i again and again, So there will be some other time that the process will surely visit state i, and so on. Therefore, the number of visits to state i is infinite. However, if i is transient, then starting from i, there is a probability f_i that the system will return to i at some times and $(1-f_i)$ that the system will never come back anymore. So, the probability that the system will return to i in n stages later and will never return anymore is $f_i^n(1-f_i)$, which is a geometric distribution, so the expected number of visits, $E(N)$ is

$$E(N) = \frac{f_i}{1-f_i}$$

which is finite.

Theorem. In a MC if for a state i, $E(N) = \sum_{n=1}^{\infty} P_{ii}^{(n)} = \infty$, then the state i is recurrent, otherwise $\sum P_{ii}^{(n)} < \infty$, then i is transient.

Proof. $E[R_{ii}^{(n)}] = 1 \times P_{ii}^{(n)} + 0\,(1 - P_{ii}^{(n)}) = P_{ii}^{(n)}$

$$R_{ii}^{(n)} = \begin{cases} 1, & \text{If after } n \text{ time units the system visits } i \\ 0, & \text{o.w.} \end{cases}$$

$$E(N) = E(R_{ii}) = \sum_{n=1}^{\infty} E(R_{ii}^{(n)}) = \sum_{n=1}^{\infty} P_{ii}^{(n)}$$

Example.

$P_{33}^{(1)} = 0.3$

$P_{33}^{(2)} = P_{31}P_{13} + P_{32}P_{23} + P_{33}\cdot P_{33} + P_{34}P_{43} = 0.3^2$

$P_{33}^{(3)} = 0.3^3$

$$\Rightarrow \sum_{n=1}^{\infty} P_{33}^{(n)} = \frac{0.3}{1-0.3} = 3/7$$

$$f_3 = 0.3 \Rightarrow E(N) = \frac{f_3}{1-f_3} = \frac{0.3}{1-0.3} = \sum_{n=1}^{\infty} P_{33}^{(n)} \quad *$$

Theorem. In a MC with a finite number of states, all st Cannot be transient.

Proof. Assume that the number of states which is $M < \infty$ and all of them are transient. Suppose we are in state 1, then after some finite time, $t_1 > 0$, the system will never return to state 1 again. Similarly, after some time, $t_2 > 0$, the system will not return to state 2 and so on. So, after some time the system will not return to any state which is impossible.

Theorem. If $i \leftrightarrow j$ and i is recurrent, then j is also recurrent.

Proof. As $i \leftrightarrow j$, then $\underset{\text{there exist}}{\exists}$ $m, m' > 0$ such that $P_{ij}^{(m)} > 0$, $P_{ji}^{(m')} > 0$, Hence, for $n \geq m+m'$, using C-K, we have

$$P_{jj}^{(n)} \geq P_{ji}^{(m)} \cdot P_{ii}^{(n-m-m')} \cdot P_{ij}^{(m')} \; ; \; n \geq m+m'$$

$$\sum_{n=m+m'}^{\infty} P_{jj}^{(n)} \geq P_{ji}^{(m)} P_{ij}^{(m')} \underbrace{\sum P_{ii}^{(n-m-m')}}_{\infty \text{ because it is recurrent.}} \quad (*)$$

Since m & m' are fixed values, then $(*)$ is infinite, so

$$\sum P_{jj}^{(n)} \geq \sum_{n=m+m'} P_{jj}^{(n)} = \infty \Rightarrow j \text{ is recurrent}$$

Theorem. If $i \leftrightarrow j$ and i is transient, then j is also transient.

Proof. If the state j is recurrent, then based on the previous t
i should be recurrent.

Theorem. In an irreducible MC with finite number of states, a
states are recurrent.

Proof. Based on the last two theorems all states should be either
or transient. In addition, we know that in a MC with finite numb
all states cannot be transient. Therefore the proof follows.

Definition. The average return time to state i is defined as t
time that the system visits i for the first time after leaving it
defined by $M_i = \sum_{n=1}^{\infty} n f_i^{(n)}$. Based on this definition:

A) State i is called "positive recurrent", if $M_i < \infty$.

B) State i is called "Null recurrent", if $M_i = \infty$.

Theorem. If $i \leftrightarrow j$ and i is positive recurrent, then j is also po
recurrent. In addition, in an irreducible system with finite n
states, all the states are positive recurrent.

Definition. If the system starts from state i and only in the sta

$$\Rightarrow \sum_{n=1}^{\infty} P_{33}^{(n)} = \frac{0.3}{1-0.3} = 3/7$$

$$f_3 = 0.3 \Rightarrow E(N) = \frac{f_3}{1-f_3} = \frac{0.3}{1-0.3} = \sum_{n=1}^{\infty} P_{33}^{(n)} \quad *$$

Theorem. In a MC with a finite number of states, all st... cannot be transient.

Proof. Assume that the number of states which is $M<\infty$ and all of them are transient. Suppose we are in state 1, then after some finite time, $t_1>0$, the system will never return to state 1 again. Similarly, after some time, $t_2>0$, the system will not return to state 2 and so on. So, after some time the system will not return to any state which is impossible.

Theorem. If $i \leftrightarrow j$ and i is recurrent, then j is also recurrent.

Proof. As $i \leftrightarrow j$, then $\underset{\text{there exist}}{\exists}$ $m, m'>0$ such that $P_{ij}^{(m)}>0$, $P_{ji}^{(m')}>0$, Hence, for $n \geq m+m'$, using C-K, we have

$$P_{jj}^{(n)} \geq P_{ji}^{(m)} \cdot P_{ii}^{(n-m-m')} \cdot P_{ij}^{(m')} \ ; \ n \geq m+m'$$

$$\sum_{n=m+m'}^{\infty} P_{jj}^{(n)} \geq P_{ji}^{(m)} P_{ij}^{(m')} \underbrace{\sum P_{ii}^{(n-m-m')}}_{\infty \text{ because it is recurrent.}} \quad (*)$$

Since m & m' are fixed values, then $(*)$ is infinite, so

$$\sum P_{jj}^{(n)} \geq \sum_{n=m+m'} P_{jj}^{(n)} = \infty \Rightarrow j \text{ is recurrent}$$

Theorem. If $i \leftrightarrow j$ and i is transient, then j is also transient.

Proof. If the state j is recurrent, then based on the previous theorem, i should be recurrent.

Theorem. In an irreducible MC with finite number of states, all the states are recurrent.

Proof. Based on the last two theorems all states should be either recurrent or transient. In addition, we know that in a MC with finite number of states, all states cannot be transient. Therefore the proof follows.

Definition. The average return time to state i is defined as the average time that the system visits i for the first time after leaving it. It is defined by $M_i = \sum_{n=1}^{\infty} n f_i^{(n)}$. Based on this definition:

A) State i is called "positive recurrent", if $M_i < \infty$.

B) State i is called "Null recurrent", if $M_i = \infty$.

Theorem. If $i \leftrightarrow j$ and i is positive recurrent, then j is also positive recurrent. In addition, in an irreducible system with finite number of states, all the states are positive recurrent.

Definition. If the system starts from state i and only in the stages $d, 2d,$

3d, ... returns to i, then we say that state i has a period d. Hence, for all integer valuation except for $n=d, 2d, 3d, \dots$ we have $P_{ii}^{(n)}=0$.

As a special case, if $d=1$, then the state is called "aperiodic".

The limiting probabilities in a Markov chain:

One of the properties of a MC is that under certain conditions, the state of the system in long run is independent of the initial state that the system starts from. For example, Consider $P_{ik}^{(m)}$ and $P_{jk}^{(m)}$. These two probabilities are different from each other. Another words, the probability that the systems goes to k from i in m-steps is different from it starts from j. However, under certain conditions these two probabilities are the same. Hence, regardless of the initial state, the probability that the system is in state k is a fixed number. We denote this limiting probability with

$$\pi_k = \lim_{m\to\infty} P_{ik}^{(m)} = \lim_{m\to\infty} P_{jk}^{(m)}$$

Similarly, it is clear that when the system's long run movement is independent of the initial state, the following also holds:

$$\lim_{n\to\infty} P(X_n=k) = \lim_{n\to\infty} \pi_k^{(n)} = \pi_k$$

Proof. we know that $\pi^{(m)} = \pi^{(0)} \cdot \mathbb{P}^m$. That is, $\pi_j^{(m)} = \sum_i \pi_i^{(0)} P_{ij}$

Hence, $\lim_{m\to\infty} \pi_j^{(m)} = \lim_{m\to\infty} \sum \pi_i^{(0)} P_{ij}^{(m)}$ but we can writ the right side as follows

$$\lim_{m\to\infty} \sum \pi_i^{(0)} P_{ij}^{(m)} = \sum \pi_i^{(0)} \lim_{m\to\infty} P_{ij}^{(m)} = \underbrace{\sum \pi_i^{(0)}}_{1} \pi_j = \pi_j$$

Example. consider a MC with the following TM. what is the 2 ... 8 steps transition matrices? $P=\begin{bmatrix} 0.2 & 0.8 \\ 0.6 & 0.4 \end{bmatrix}$

$$P^{(2)}=\begin{bmatrix} 0.52 & 0.48 \\ 0.36 & 0.64 \end{bmatrix}, \quad P^{(4)}=\begin{bmatrix} 0.4432 & 0.5568 \\ 0.4172 & 0.5824 \end{bmatrix}, \quad P^{(8)}=\begin{bmatrix} 0.4289 & 0.5711 \\ 0.4283 & 0.5717 \end{bmatrix}$$

Theorem. Consider an irreducible MC with aperiodic states, the

A) The limiting probabilities always exist.

B) If the states of the system are positive recurrent, for all j,

If the states are either transient or null recurrent, then for a

Theorem. In an irreducible MC with aperiodic positive recurrent the limiting probabilities π_j are obtained as follows.

$$\begin{cases} (\pi \ \pi \ \cdots \ \pi_n) = (\pi_0 \ \pi_1 \cdots \pi_n) \cdot \mathbb{P} \quad \text{or } \pi = \pi P \quad \text{or } \pi_j = \sum_{i=}^{\infty} \\ \sum_{i=1}^{\infty} \pi_i = 1 \end{cases}$$

Proof. we know that $\pi^{(m)} = \pi^{(m-1)} \cdot P$.

$$\pi^{(m)} = \pi^{(0)} P^m = \underbrace{\pi^{(0)} \cdot P^{m-1}}_{\pi^{(m-1)}} \cdot P \Rightarrow \lim_{m\to\infty} \pi^{(m)} = \overbrace{\lim_{m\to\infty} \pi^{(m-1)}}^{\pi} \cdot P \Rightarrow \pi = \pi$$

$3d, \dots$ returns to i, then we say that state i has a period d. Hence, for all integer valuation except for $n=d, 2d, 3d, \dots$ we have $P_{ii}^{(n)}=0$.

As a special case, if $d=1$, then the state is called "aperiodic".

The limiting probabilities in a Markov chain:

One of the properties of a MC is that under certain conditions, the state of the system in long run is independent of the initial state that the system starts from. For example, Consider $P_{ik}^{(m)}$ and $P_{jk}^{(m)}$. These two probabilities are different from each other. Another words, the probability that the systems goes to k from i in m-steps is different from it starts from j. However, under certain conditions these two probabilities are the same. Hence, regardless of the initial state, the probability that the system is in state k is a fixed number. We denote this limiting probability with

$$\pi_k = \lim_{m\to\infty} P_{ik}^{(m)} = \lim_{m\to\infty} P_{jk}^{(m)}$$

Similarly, it is clear that when the system's long run movement is independent of the initial state, the following also holds:

$$\lim_{n\to\infty} P(X_n=k) = \lim_{n\to\infty} \pi_k^{(n)} = \pi_k$$

Proof. we know that $\pi^{(m)} = \pi^{(0)} \cdot \mathbb{P}^m$. That is, $\pi_j^{(m)} = \sum_i \pi_i^{(0)} P_{ij}^{(m)}$.

Hence, $\lim_{m\to\infty} \pi_j^{(m)} = \lim_{m\to\infty} \sum \pi_i^{(0)} P_{ij}^{(m)}$ but we can writ the right hand side as follows

$$\lim_{m\to\infty} \sum \pi_i^{(0)} P_{ij}^{(m)} = \sum \pi_i^{(0)} \lim_{m\to\infty} P_{ij}^{(m)} = \underbrace{\sum \pi_i^{(0)}}_{1} \pi_j = \pi_j$$

Example. consider a MC with the following TM. what is the 2, 4 and 8 steps transition matrices? $P=\begin{bmatrix}0.2 & 0.8\\ 0.6 & 0.4\end{bmatrix}$

$$P^{(2)}=\begin{bmatrix}0.52 & 0.48\\ 0.36 & 0.64\end{bmatrix}, \quad P^{(4)}=\begin{bmatrix}0.4432 & 0.5568\\ 0.4172 & 0.5824\end{bmatrix}, \quad P^{(8)}=\begin{bmatrix}0.4289 & 0.5711\\ 0.4283 & 0.5717\end{bmatrix}$$

Theorem. Consider an irreducible MC with aperiodic states, then

A) The limiting probabilities always exist.

B) If the states of the system are positive recurrent, for all j, $\pi_j > 0$.

If the states are either transient or null recurrent, then for all j, $\pi_j = 0$.

Theorem. In an irreducible MC with aperiodic positive recurrent states, the limiting probabilities π_j are obtained as follows.

$$\begin{cases} (\pi \ \pi \ \dots \ \pi_n) = (\pi_0 \ \pi_1 \dots \pi_n) \cdot \mathbb{P} \quad \text{or} \quad \pi = \pi P \quad \text{or} \quad \pi_j = \sum_{i=0}^{\infty} \pi_i P_{ij} \\ \sum_{i=1}^{\infty} \pi_i = 1 \end{cases}$$

Proof. we know that $\pi^{(m)} = \pi^{(m-1)} \cdot P$.

$$\pi^{(m)} = \pi^{(0)} P^m = \underbrace{\pi^{(0)} P^{m-1}}_{\pi^{(m-1)}} \cdot P \Rightarrow \lim_{m\to\infty} \pi^{(m)} = \overbrace{\lim_{m\to\infty} \pi^{(m-1)}}^{\pi} \cdot P \Rightarrow \pi = \pi \cdot P$$

Example. $P=\begin{bmatrix}0.2 & 0.8\\ 0.6 & 0.4\end{bmatrix}$

We first check the conditions of existence of limiting probabilities.

$\left\{\begin{array}{l}\text{1) the only class is } \{1,2\} \text{ so it is irreducible.}\\ \text{2) } d=1, \text{ aperiodicity exists.}\end{array}\right. \Rightarrow$ Limiting probabilities exist.

$$\begin{cases}\pi=\pi P & (\pi_1\ \pi_2)=(\pi_1\ \pi_2)\begin{pmatrix}0.2 & 0.8\\ 0.6 & 0.4\end{pmatrix}\\ \sum \pi=1\end{cases}$$

$$\begin{cases}\pi_1=0.2\pi_1+0.6\pi_2\\ \pi_2=0.8\pi_1+0.4\pi_2 \quad\Rightarrow\quad \pi_1=0.5714,\ \pi_2=0.4286\\ \pi_1+\pi_2=1\end{cases}$$

Note: Always we have m unknowns and $m+1$ equalities. So, at most there is a (may be not) unique solution.

Note: π_1 & π_2 are both positive, because all the states are positive recurrent.

Limiting probabilities in MC with period d:

Recall that if at the starting time the system's state is j, then it only can return to j at steps $d, 2d, 3d, \ldots$ However, as n goes to infinity, the value of $P_{jj}^{(n)}$ is either zero or π_j (at the periods $d, 2d, 3d, \ldots$). In addition,

$$\lim_{n\to\infty} \mathbb{P}(X_n=j)=\begin{cases}\pi_j & , n=kd\\ 0 & , n\neq kd\end{cases}$$

Theorem. In an irreducible MC with the period d, the limitin[g]

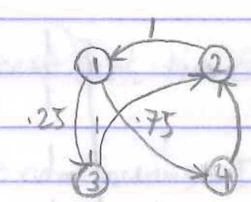
π_i is obtained as follows:

$$\begin{cases}(\pi_0\ \pi_1\ \pi_2 \cdots)=(\pi_0\ \pi_1\ \pi_2\cdots)\mathbb{P}\\ \sum_i \pi_i=d\end{cases}$$

Note: The important point in this theorem is that the sum of the probabilities is d instead of 1. The reason is that in the stage m[…] the system goes to a certain number of states. Hence, $\sum \pi_i=1$ i[…] all states i that the system can indeed go to at the time $n=md$. the overall sum of all probabilities for all states would be $\sum_i P_i=$[…]

Example.

$$P=\begin{bmatrix}0 & 0 & 0.25 & 0.75\\ 1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 1 & 0 & 0\end{bmatrix}$$



the only class is $\{1,2,3,4\}$ and it is irreducible.

$$\left.\begin{array}{l}P_{11}^{(1)}=0\\ P_{11}^{(2)}=0\\ P_{11}^{(3)}>0\end{array}\right\}\Rightarrow d=1 \qquad \left.\begin{array}{l}\pi_1=1\\ \pi_2=1\\ \pi_3=0.25\\ \pi_4=0.75\end{array}\right\}\rightarrow \sum \pi_i=3$$

In this example $d=3$ and the MC is irreducible. Assume that at the beg[…] system is in state 2. Then it goes to 1 with prob. 1, then with prob. $\frac{1}{4}$[…]

Example. $P=\begin{bmatrix}0.2 & 0.8\\ 0.6 & 0.4\end{bmatrix}$

We first check the conditions of existence of limiting probabilities.

$\left\{\begin{array}{l}\text{1) the only class is } \{1,2\} \text{ so it is irreducible.}\\ \text{2) } d=1, \text{ aperiodicity exists.}\end{array}\right. \Rightarrow$ Limiting probabilities exist.

$$\begin{cases}\pi=\pi P & (\pi_1\ \pi_2)=(\pi_1\ \pi_2)\begin{pmatrix}0.2 & 0.8\\ 0.6 & 0.4\end{pmatrix}\\ \sum\pi=1\end{cases}$$

$$\begin{cases}\pi_1=0.2\pi_1+0.6\pi_2\\ \pi_2=0.8\pi_1+0.4\pi_2\\ \pi_1+\pi_2=1\end{cases} \Rightarrow \pi_1=0.5714,\ \pi_2=0.4286$$

Note: Always we have m unknowns and m+1 equalities. So, at most there is a (may be not) unique solution.

Note: π_1 & π_2 are both positive, because all the states are positive recurrent.

Limiting probabilities in MC with period d:

Recall that if at the starting time the system's state is j, then it only can return to j at steps d, 2d, 3d, ... However, as n goes to infinity, the value of $P_{jj}^{(n)}$ is either zero or π_j (at the periods d, 2d, 3d, ...). In addition,

$$\lim_{n\to\infty} \mathbb{P}(X_n=j)=\begin{cases}\pi_j & , n=kd\\ 0 & , n\neq kd\end{cases}$$

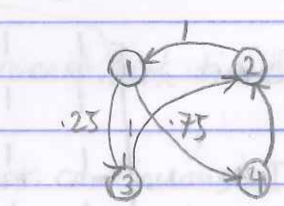
Theorem. In an irreducible MC with the period d, the limiting prob. π_i is obtained as follows:

$$\begin{cases}(\pi_0\ \pi_1\ \pi_2\cdots)=(\pi_0\ \pi_1\ \pi_2\cdots)\mathbb{P}\\ \sum_i \pi_i=d\end{cases}$$

Note: The important point in this theorem is that the sum of the limiting probabilities is d instead of 1. The reason is that in the stage md+k, the system goes to a certain number of states. Hence, $\sum\pi_i=1$ is only for all states i that the system can indeed go to at the time n=md+k. Hence, the overall sum of all Probabilities for all states would be $\sum_i P_i=d$.

Example.

$$P=\begin{bmatrix}0 & 0 & 0.25 & 0.75\\ 1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 1 & 0 & 0\end{bmatrix}$$



the only class is $\{1,2,3,4\}$ and it is irreducible.

$$\left.\begin{array}{l}P_{11}^{(1)}=0\\ P_{11}^{(2)}=0\\ P_{11}^{(3)}>0\end{array}\right\} \Rightarrow d=1 \qquad \left.\begin{array}{l}\pi_1=1\\ \pi_2=1\\ \pi_3=0.25\\ \pi_4=0.75\end{array}\right\} \to \sum\pi_i=3$$

In this example d=3 and the MC is irreducible. Assume that at the begining, the system is in state 2. Then it goes to 1 with Prob. 1, then with prob. 1/4 or 3/4 it goes to either state 3 or state 4. I step 4 (time 4), it goes back to 2 with prob. 1. Hence, $\pi_1 = 1$, because the system at time 2, 5, 8, ... reaches to 1 with prob 1 and at other times 1 cannot be accessed. Similarly, $\pi_2 = 1$, $\pi_3 = 0.25$, $\pi_4 = 0.75$.

The limiting matrices for P can be written as:

$$\lim_{m\to\infty} P^{(3m)} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1/4 & 3/4 \\ 0 & 0 & 1/4 & 3/4 \end{bmatrix}, \quad \lim_{m\to\infty} P^{(3m+1)} = \begin{bmatrix} 0 & 0 & 1/4 & 3/4 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

$$\lim_{m\to\infty} P^{(3m+2)} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1/4 & 3/4 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

If we add them together so:

$$\begin{bmatrix} 1 & 1 & 1/4 & 3/4 \\ 1 & 1 & 1/4 & 3/4 \\ 1 & 1 & 1/4 & 3/4 \\ 1 & 1 & 1/4 & 3/4 \end{bmatrix}^{d}$$

Theorem. In an irreducible MC

$$\pi_j = \frac{d}{M_j} \quad \text{and} \quad M_j = \sum_{n=1}^{\infty} n f_j^{(n)}$$

→ The average return time to j

Example.

$$M_1 = M_2 = \frac{d}{\pi_1} = \frac{d}{\pi_2} = \frac{3}{1} = 3$$

$$M_3 = \frac{d}{\pi_3} = \frac{3}{0.25} = 12, \quad M_4 = \frac{3}{0.75} = 4$$

Proof. By conditioning

$$M_i = E[T_i] = \sum E[T_i \mid \text{very first outcome}] \cdot \mathbb{P}(\text{very first out}$$

$$= \pi_i \cdot d + (1-\pi_i)(d + M_i) \Rightarrow M_i = \frac{d}{\pi_i}$$

The other way

$$M_i = \pi_i d + (1-\pi_i)\pi_i (2d) + (1-\pi_i)^2 \pi_i 3d + \dots = d \underbrace{\sum_{n=1}^{\infty} n(1-\pi_i)}_{\text{Geometric d}}$$

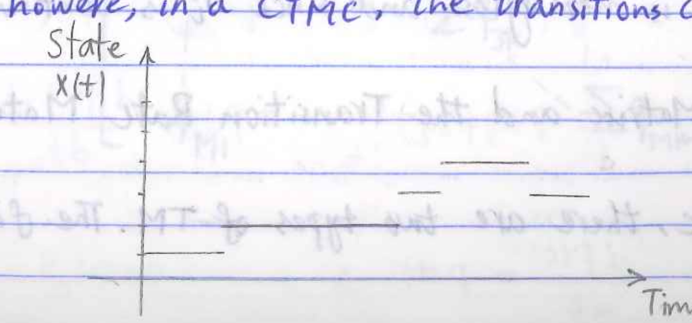
so $E = \frac{1}{\pi_i}$

Limiting probabilities in non-irreducible MC:

If the system is not irreducible, then each class of the MC can role of an irreducible MC, and for any transient states j, π

Continuous time Markov chain:

As mentioned before in a Markov process both time and the sta system X(t), can be either discrete or continuous. In a MC bot X(t) are discrete. In CTMC, t is continuous but X(t) is still discret we assume that the state transition takes place at the end of or stage, howere, in a CTMC, the transitions can take place time.



goes to either state 3 or state 4. I step 4 (time 4), it goes back to 2 with prob. 1. Hence, $\pi_1 = 1$, because the system at time 2, 5, 8, ... reaches to 1 with prob 1 and at other times 1 cannot be accessed. Similarly, $\pi_2 = 1$, $\pi_3 = 0.25$, $\pi_4 = 0.75$.

The limiting matrices for P can be written as:

$$\lim_{m\to\infty} P^{(3m)} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1/4 & 3/4 \\ 0 & 0 & 1/4 & 3/4 \end{bmatrix}, \quad \lim_{m\to\infty} P^{(3m+1)} = \begin{bmatrix} 0 & 0 & 1/4 & 3/4 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

$$\lim_{m\to\infty} P^{(3m+2)} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1/4 & 3/4 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

If we add them together so:

$$\begin{array}{c} d \\ \begin{bmatrix} 1 & 1 & 1/4 & 3/4 \\ 1 & 1 & 1/4 & 3/4 \\ 1 & 1 & 1/4 & 3/4 \\ 1 & 1 & 1/4 & 3/4 \end{bmatrix} \end{array}$$

Theorem. In an irreducible MC

The average return time to j

$$\pi_j = \frac{d}{M_j} \quad \text{and} \quad M_j = \sum_{n=1}^{\infty} n f_j^{(n)}$$

Example.

$$M_1 = M_2 = \frac{d}{\pi_1} = \frac{d}{\pi_2} = \frac{3}{1} = 3$$

$$M_3 = \frac{d}{\pi_3} = \frac{3}{0.25} = 12, \quad M_4 = \frac{3}{0.75} = 4$$

Proof. By conditioning

$$M_i = E[T_i] = \sum E[T_i \mid \text{very first outcome}] \cdot \mathbb{P}(\text{very first outcome})$$

$$= \pi_i d + (1-\pi_i)(d + M_i) \Rightarrow M_i = \frac{d}{\pi_i}$$

The other way

$$M_i = \pi_i d + (1-\pi_i)\pi_i(2d) + (1-\pi_i)^2 \pi_i 3d + \dots = d \underbrace{\sum_{n=1}^{\infty} n(1-\pi_i)^n}_{\text{Geometric distribution}} = \frac{d}{\pi_i}$$

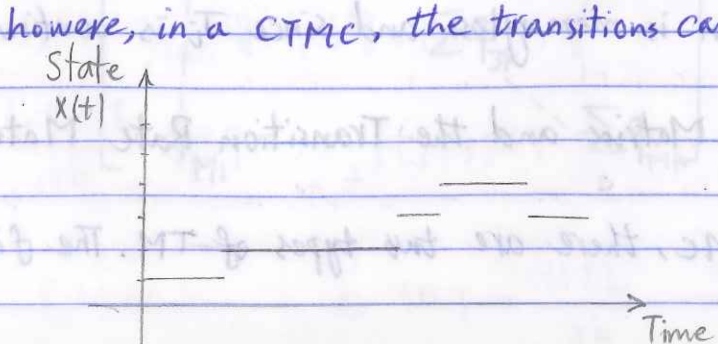
so $E = \frac{1}{\pi_i}$

Limiting probabilities in non-irreducible MC:

If the system is not irreducible, then each class of the MC can play a role of an irreducible MC, and for any transient states j, $\pi_j = 0$.

Continuous time Markov chain:

As mentioned before in a Markov process both time and the state of the system $X(t)$, can be either discrete or continuous. In a MC both t and $X(t)$ are discrete. In CTMC, t is continuous but $X(t)$ is still discrete. In DTMC we assume that the state transition takes place at the end of each time or stage, howere, in a CTMC, the transitions can take place at any time.



Example. Each Poisson process is a CTMC. Because of the independent increment, the number of events in the time interval $[s, s+t)$ is independent of the number of customers have arrived up to time s.

$$\begin{cases} P_{ij}(t) = \mathbb{P}(\text{the \# of customers during } t \text{ time units} = j-i) = \dfrac{e^{-\lambda t}(\lambda t)^{j-i}}{(j-i)!} & ; j \geq i \\ P_{ij}(t) = 0 & ; j < i \end{cases}$$

The distribution of the stopping time in a given state:

Once a system reaches to a state, it may stay at that state for some time and then makes transition to other states. If T_i is the amount of time that the system remains in state i, then T_i would have an exponential distribution.

Proof. Assume that at time zero the system has entered the state i and has remained in that state by the time s. Since the process is Markov, the future movement of the process is independent of the past. Hence, $\mathbb{P}(T_i > t+s \mid T_i > s) = \mathbb{P}(T_i > t)$ which in the case that the T_i distribution is memoryless and since T_i is continuous, so $T_i \sim \text{Exp}$.

Transition Matrix and the Transition Rate Matrix:

In a CTMC, there are two types of TM. The first which we show it, is $P(t)$, which is similar to DTMC. Each elements of this matrix is the probability of going from state i to j during t time interval. A specific case of $P(t)$ is $P(0)$:

$$P_{ii}(0) = 1, \quad P_{ij}(0) = 0$$

So, $P(0) = I$ (identical matrix).

The second matrix is called transition rate matrix, which we denote by Q. Each element of Q, q_{ij}, is defined as:

$$\begin{cases} q_{ij} = \lim\limits_{t \to 0} \dfrac{P_{ij}(t)}{t} = \dfrac{d}{dt} P_{ij}(t) & , i \neq j \\ q_{ii} = -\sum\limits_{j \neq i} q_{ij} = \dfrac{d}{dt} P_{ii}(t) \end{cases}$$

Proof.

$$q_{ii} = -\sum_{j \neq i} q_{ij} = -\sum_{j \neq i} \frac{d}{dt} P_{ij}(t) = \frac{d}{dt}(-1 + P_{ii}(t)) = \frac{d}{dt} P_{ii}(t)$$

$$Q = \begin{bmatrix} -\sum\limits_{k=1}^{M} q_{1k} & q_{12} & q_{13} & \cdots & q_{1M} \\ q_{21} & -\sum q_{2k} & q_{23} & \cdots & q_{2M} \\ \vdots & & -\sum q_{3k} & & \vdots \\ \vdots & & & \ddots & \\ q_{M1} & \cdots & \cdots & \cdots & -\sum q_{Mk} \end{bmatrix}$$

Example. Each Poisson process is a CTMC. Because of the independent increment, the number of events in the time interval $[s, s+t)$ is independent of the number of customers have arrived up to time s.

$$\begin{cases} P_{ij}(t) = \mathbb{P}(\text{the \# of customers during } t \text{ time units} = j-i) = \dfrac{e^{-\lambda t} (\lambda t)^{j-i}}{(j-i)!} & ; j \geqslant i \\ P_{ij}(t) = 0 & ; j < i \end{cases}$$

The distribution of the stopping time in a given state:

Once a system reaches to a state, it may stay at that state for some time and then makes transition to other states. If T_i is the amount of time that the system remains in state i, then T_i would have an exponential distribution.

Proof. Assume that at time zero the system has entered the state i and has remained in that state by the time s. Since the process is Markov, the future movement of the process is independent of the past. Hence, $\mathbb{P}(T_i > t+s \mid T_i > s) = \mathbb{P}(T_i > t)$ which in the case that the T_i distribution is memoryless and since T_i is continuous. So $T_i \sim \text{Exp}$.

Transition Matrix and the Transition Rate Matrix:

In a CTMC, there are two types of TM. The first which we show it, is $P(t)$, which is similar to DTMC. Each elements of this matrix, $P_{ij}(t)$ is the probability of going from state i to j during t time interval. The specific case of $P(t)$ is $P(0)$:

$$P_{ii}(0) = 1 \,, \quad P_{ij}(0) = 0$$

So, $P(0) = I$ (identical matrix).

The second matrix is called transition rate matrix, which we denote with Q. Each element of Q, q_{ij}, is defined as:

$$\begin{cases} q_{ij} = \lim\limits_{t \to 0} \dfrac{P_{ij}(t)}{t} = \dfrac{d}{dt} P_{ij}(t) & , i \neq j \\ q_{ii} = -\sum\limits_{j \neq i} q_{ij} = \dfrac{d}{dt} P_{ii}(t) \end{cases}$$

Proof.

$$q_{ii} = -\sum_{j \neq i} q_{ij} = -\sum_{j \neq i} \frac{d}{dt} P_{ij}(t) = \frac{d}{dt}\left(-1 + P_{ii}(t)\right) = \frac{d}{dt} P_{ii}(t)$$

$$Q = \begin{bmatrix} -\sum\limits_{k=1}^{M} q_{1k} & q_{12} & q_{13} & q_{1M} \\ q_{21} & -\sum q_{2k} & q_{23} & q_{2M} \\ \vdots & & -\sum q_{3k} & \vdots \\ q_{M1} & & & -\sum q_{Mk} \end{bmatrix}$$

Example. Find $P_{ij}(t)$ and q_{ij} in the Poisson process.

$$P_{ij}=\begin{cases}\lambda t+o(t) & ,j=i+1\\ o(t) & ,j>i+1\\ 0 & ,j<i\end{cases} \Rightarrow q_{ij}=\begin{cases}\lim_{t\to 0}\frac{o(t)}{t}=0 & ,j\neq i+1\\ \lim_{t\to 0}\frac{\lambda t+o(t)}{t}=\lambda & ,j>i+1\\ -\lambda & ,i=j\end{cases}$$

Obtaining the parameter of distribution of T_i:

$T_i \sim Exp(\cdot)$. The parameter is the rate of staying at state i which is equal to $q_{ii}=-\sum_{j\neq i} q_{ij}$

So,

$$\mathbb{P}(T_i>x)=e^{q_{ii}x}=e^{-\sum_{i\neq j} q_{ij}\,x}$$

C-K equation:

$$P_{ij}(t_1+t_2)=\sum_k P_{ik}(t)\cdot P_{kj}(t_2)$$

Kolmogrov Backward and Forward Equations:

$Q=\frac{d}{dt}P(t)$. We can obtain Q from P by this relation. But, in practice obtaining q is far easier than $P(t)$. To obtain $P(t)$ from Q we use C-K as follows:

$$\mathbb{P}(t_1+t_2)=\mathbb{P}(t_1)\cdot\mathbb{P}(t_2) \quad \text{(Matrix form)}$$

$$\Rightarrow \frac{d}{dt_2}P(t_1+t_2)\Big|_{t_2=0}=P(t_1)\,\frac{dP(t_2)}{dt_2}\Big|_{t_2=0}$$

$$\Rightarrow \frac{dP(t)}{dt}=P(t)\cdot Q$$

This is called Kolmogrov Forward equation.

Similarly, we have:

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

which is called Kolmogrov Backward equation. This is obtained

derivetive from both sides with respect to t_1.

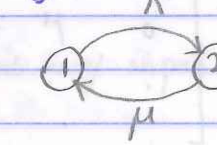
Example. Consider a machine which either work or out of fun

operating time is exponential distribution with parameter λ and

time is exponential with parameter μ. If at the machine

what is the probability that it is working by the time $t=15$? $P($

P_{11}



$$Q=\begin{bmatrix}-\lambda & \lambda\\ \mu & -\mu\end{bmatrix}$$

$$\begin{bmatrix}\frac{dP_{11}(t)}{dt} & \frac{dP_{12}(t)}{dt}\\ \frac{dP_{21}(t)}{dt} & \frac{dP_{22}(t)}{dt}\end{bmatrix}=\begin{bmatrix}-\lambda & \lambda\\ \mu & -\mu\end{bmatrix}\begin{bmatrix}P_{11}(t) & P_{12}(t)\\ P_{21}(t) & P_{22}(t)\end{bmatrix}$$

$*\ \frac{dP_{11}(t)}{dt}=-\lambda[P_{11}(t)-P_{21}(t)] \quad (\times\mu)$

$\frac{dP_{21}(t)}{dt}=\mu[P_{11}(t)-P_{21}(t)] \quad (\times\lambda)$

$+ \Big\}\ \mu\frac{dP_{11}(t)}{dt}+\lambda\frac{dP_{21}(t)}{dt}=0$ int

bo

$$\Rightarrow \begin{cases}\mu P_{11}(t)+\lambda P_{21}(t)=C\\ P_{11}(0)=1,\ P_{21}(0)=0\end{cases} \Rightarrow C=\mu \Rightarrow \lambda P_{21}(t)=\mu(1-P_{11}(t))$$ replace bac

Example. Find $P_{ij}(t)$ and q_{ij} in the Poisson process.

$$P_{ij} = \begin{cases} \lambda t + o(t) & , j = i+1 \\ o(t) & , j > i+1 \\ 0 & , j < i \end{cases} \Rightarrow q_{ij} = \begin{cases} \lim_{t \to 0} \frac{o(t)}{t} = 0 & , j \neq i+1 \\ \lim_{t \to 0} \frac{\lambda t + o(t)}{t} = \lambda & , j > i+1 \\ -\lambda & , i = j \end{cases}$$

Obtaining the parameter of distribution of T_i :

$T_i \sim Exp(\cdot)$. The parameter is the rate of staying at state i which is equal to $q_{ii} = -\sum_{j \neq i} q_{ij}$

So,

$$\mathbb{P}(T_i > x) = e^{q_{ii} x} = e^{-\sum_{i \neq j} q_{ij} x}$$

C-K equation :

$$P_{ij}(t_1 + t_2) = \sum_k P_{ik}(t) \cdot P_{kj}(t_2)$$

Kolmogrov Backward and Forward Equations :

$Q = \frac{d}{dt} P(t)$. We can obtain Q from P by this relation. But, in practice obtaining q is far easier than $P(t)$. To obtain $P(t)$ from Q we use C-K as follows :

$$\mathbb{P}(t_1 + t_2) = \mathbb{P}(t_1) \cdot \mathbb{P}(t_2) \qquad \text{(Matrix form)}$$

$$\Rightarrow \frac{d}{dt_2} P(t_1 + t_2) \Big|_{t_2 = 0} = P(t_1) \frac{d\,P(t_2)}{dt_2} \Big|_{t_2 = 0}$$

$$\Rightarrow \frac{d\,P(t)}{dt} = P(t) \cdot Q$$

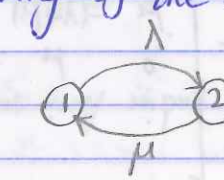
This is called Kolmogrov Forward equation.

Similarly, we have :

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

Which is called Kolmogrov Backward equation. This is obtained by taking derivative from both sides with respect to t_1.

Example. Consider a machine which either work or out of function. The operating time is exponential distribution with parameter λ and the repair time is exponential with parameter μ. If at the machine is working what is the probability that it is working by the time $t = 15$? $P_{11}(15) = ?$ $P_{11}(t) = ?$



$$Q = \begin{bmatrix} -\lambda & \lambda \\ \mu & -\mu \end{bmatrix}$$

$$\begin{bmatrix} \frac{d\,P_{11}(t)}{dt} & \frac{d\,P_{12}(t)}{dt} \\ \frac{d\,P_{21}(t)}{dt} & \frac{d\,P_{22}(t)}{dt} \end{bmatrix} = \begin{bmatrix} -\lambda & \lambda \\ \mu & -\mu \end{bmatrix} \begin{bmatrix} P_{11}(t) & P_{12}(t) \\ P_{21}(t) & P_{22}(t) \end{bmatrix}$$

$$\left.\begin{array}{l} * \ \frac{dP_{11}(t)}{dt} = -\lambda [P_{11}(t) - P_{21}(t)] \quad (\times \mu) \\ \frac{dP_{21}(t)}{dt} = \mu [P_{11}(t) - P_{21}(t)] \quad (\times \lambda) \end{array}\right\} + \quad \mu \frac{dP_{11}(t)}{dt} + \lambda \frac{dP_{21}(t)}{dt} = 0 \quad \text{integral from both sides}$$

$$\Rightarrow \begin{cases} \mu P_{11}(t) + \lambda P_{21}(t) = C \\ P_{11}(0) = 1 , \ P_{21}(0) = 0 \end{cases} \Rightarrow C = \mu \Rightarrow \lambda P_{21}(t) = \mu (1 - P_{11}(t)) \quad \text{replace back in } *$$

$$\frac{dP_{11}(t)}{dt} = -(\lambda+\mu)P_{11}(t)+\mu \Rightarrow e^{(\lambda+\mu)t}\left(\frac{dP_{11}(t)}{dt}+(\lambda+\mu)P_{11}(t)\right)=\mu e^{(\lambda+\mu)t}$$

$$\Rightarrow \frac{d}{dt}\left[e^{(\lambda+\mu)t}P_{11}(t)\right]=\mu e^{(\lambda+\mu)t} \Rightarrow e^{(\lambda+\mu)t}P_{11}(t)=\frac{\mu}{\mu+\lambda}e^{(\lambda+\mu)t}+C$$

$$\Rightarrow P_{11}(0)=1 \Rightarrow C=\frac{\lambda}{\lambda+\mu} \Rightarrow e^{(\lambda+\mu)t}P_{11}(t)=\frac{\mu}{\lambda+\mu}e^{(\lambda+\mu)t}+\frac{\lambda}{\lambda+\mu}$$

$$\Rightarrow P_{11}(t)=\frac{\mu}{\lambda+\mu}+\frac{\lambda}{\lambda+\mu}e^{-(\lambda+\mu)t} \xrightarrow{t=15} P_{11}(15)=\frac{\mu}{\mu+\lambda}+\frac{\lambda}{\lambda+\mu}e^{-(\lambda+\mu)15}$$

Transforming CTMC to DTMC:

$$\begin{cases} P_{ij}=0 & ; i=j \\ P_{ij}=\frac{-1}{q_{ii}}q_{ij} & ; i\neq j \end{cases}$$

Example.

$$Q=\begin{bmatrix} -\lambda & \lambda & 0 & 0 & \cdots \\ \mu & -(\lambda+\mu) & \lambda & 0 & \\ 0 & \mu & -(\lambda+\mu) & \lambda & 0 \cdots \\ \vdots & \vdots & & & \ddots \end{bmatrix}$$

$$P=\begin{bmatrix} 0 & 1 & 0 & 0 & \cdots \\ \frac{\mu}{(\lambda+\mu)} & 0 & \frac{\lambda}{(\lambda+\mu)} & 0 & \cdots \\ 0 & \frac{\mu}{(\lambda+\mu)} & 0 & \frac{\lambda}{(\lambda+\mu)} & \cdots \\ \vdots & \vdots & \vdots & & \end{bmatrix}$$

Limiting probabilities in CTMC:

Theorem. If the system is irreducible, then the $\lim_{t\to\infty} P_{ij}(t)$ converge to a fixed number. When we denote with π_j which is equal to:

$$\pi_j = \lim_{t\to\infty} P_{ij}(t)$$

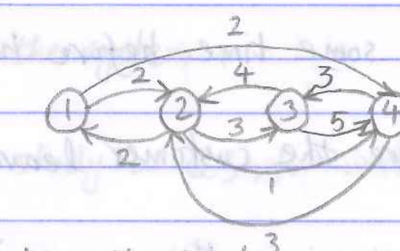
In addition, if all the states are positive recurrent, then:

$$\begin{cases} [\pi_1\ \pi_2 \ldots]\,Q=0 \\ \sum_{i=1}^{\infty}\pi_i=1 \end{cases}$$

*A CTMC is "Ergodic" if it i
and positive recurrent.

Example.

$$Q=\begin{bmatrix} -4 & 2 & 0 & 2 \\ 2 & -6 & 3 & 1 \\ 0 & 4 & -9 & 5 \\ 0 & 3 & 3 & -6 \end{bmatrix}$$



This is irreducible and positive recurrent, so the limiting prob. exist.

Note: In CTMC, we donot study the periodicity condition as DTM

We can use $\pi Q=0$ by multiplying π by Q. The other way is to use equations directly from the graph. In this way input flows to a state is output flows to the state. So,

$$\begin{cases} 2\pi_2 = 4\pi_1 \\ 2\pi_1+4\pi_3+3\pi_4=6\pi_2 \\ 3\pi_3+3\pi_4=9\pi_3 \\ 2\pi_1+\pi_2+5\pi_3=6\pi_4 \\ \pi_1+\pi_2+\pi_3+\pi_4=1 \end{cases}$$

$$\frac{dP_{11}(t)}{dt} = -(\lambda+\mu)P_{11}(t) + \mu \Rightarrow e^{(\lambda+\mu)t}\left(\frac{dP_{11}(t)}{dt} + (\lambda+\mu)P_{11}(t)\right) = \mu e^{(\lambda+\mu)t}$$

$$\Rightarrow \frac{d}{dt}\left[e^{(\lambda+\mu)t}P_{11}(t)\right] = \mu e^{(\lambda+\mu)t} \Rightarrow e^{(\lambda+\mu)t}P_{11}(t) = \frac{\mu}{\mu+\lambda}e^{(\lambda+\mu)t} + C$$

$$\Rightarrow P_{11}(0)=1 \Rightarrow C = \frac{\lambda}{\lambda+\mu} \Rightarrow e^{(\lambda+\mu)t}P_{11}(t) = \frac{\mu}{\lambda+\mu}e^{(\lambda+\mu)t} + \frac{\lambda}{\lambda+\mu}$$

$$\Rightarrow P_{11}(t) = \frac{\mu}{\lambda+\mu} + \frac{\lambda}{\lambda+\mu}e^{-(\lambda+\mu)t} \xrightarrow{t=15} P_{11}(15) = \frac{\mu}{\mu+\lambda} + \frac{\lambda}{\lambda+\mu}e^{-(\lambda+\mu)15}$$

Transforming CTMC to DTMC:

$$\begin{cases} P_{ij} = 0 & ; i=j \\ P_{ij} = \frac{-1}{q_{ii}}q_{ij} & ; i\neq j \end{cases}$$

Example.

$$Q = \begin{bmatrix} -\lambda & \lambda & 0 & 0 & \cdots \\ \mu & -(\lambda+\mu) & \lambda & 0 & \\ 0 & \mu & -(\lambda+\mu) & \lambda & 0 \cdots \\ \vdots & \vdots & & \ddots & \end{bmatrix}$$

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots \\ \frac{\mu}{(\lambda+\mu)} & 0 & \frac{\lambda}{(\lambda+\mu)} & 0 & \cdots \\ 0 & \frac{\mu}{(\lambda+\mu)} & 0 & \frac{\lambda}{(\lambda+\mu)} & \cdots \\ \vdots & \vdots & \vdots & & \end{bmatrix}$$

Limiting probabilities in CTMC:

Theorem. If the system is irreducible, then the $\lim_{t\to\infty} P_{ij}(t)$ converge to a fixed number. When we denote with π_j which is equal to:

$$\pi_j = \lim P_{ij}(t)$$

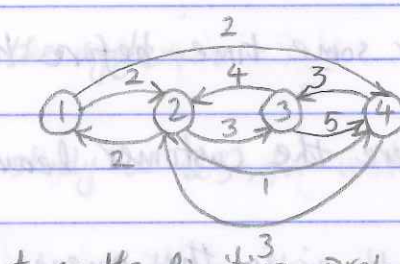
In addition, if all the states are positive recurrent, then:

$$\begin{cases} [\pi_1\ \pi_2 \cdots]Q = 0 \\ \sum_{i=1}^{\infty} \pi_i = 1 \end{cases}$$

* A CTMC is "Ergodic" if it is irreducible and positive recurrent.

Example.

$$Q = \begin{bmatrix} -4 & 2 & 0 & 2 \\ 2 & -6 & 3 & 1 \\ 0 & 4 & -9 & 5 \\ 0 & 3 & 3 & -6 \end{bmatrix}$$



This is irreducible and positive recurrent, so the limiting prob. exist.

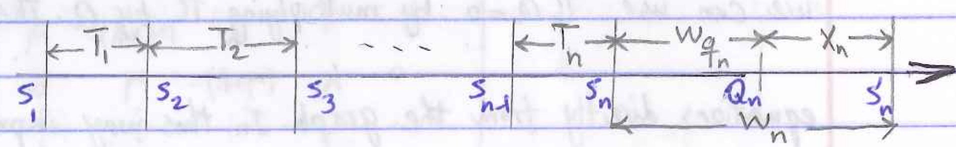
Note: In CTMC, we donot study the periodicity condition as DTMC.

We can use $\pi Q = 0$ by multiplying π by Q. The other way is to use balance equations directly from the graph. In this way input flows to a state is equal to output flows to the state. So,

$$\begin{cases} 2\pi_2 = 4\pi_1 \\ 2\pi_1 + 4\pi_3 + 3\pi_4 = 6\pi_2 \\ 3\pi_2 + 3\pi_4 = 9\pi_3 \\ 2\pi_1 + \pi_2 + 5\pi_3 = 6\pi_4 \\ \pi_1 + \pi_2 + \pi_3 + \pi_4 = 1 \end{cases}$$

The general framework of queuing systems:

Assume that the first customer arrives at S_1, the second at $S_2, \dots, n^{th}$ at S_n and the time interval of the arrival of customer n and $n-1$ is T_n. A customer, for instance n, after arriving at the system may have to wait for some time before his service starts at Q_n and will finish at S'_n where the customer leaves the system. The waiting time of this customer in the queue is W_{q_n}, his sojurn time (the waiting time in the system) is W_n and his service time is X_n.



Then we can write these relationships:

$$\begin{cases} W_{q_n} = Q_n - S_n \\ W_n = S'_n - S_n \\ X_n = S'_n - Q_n \\ W_n = W_{q_n} + X_n \end{cases}$$

Let us denote the number of customers arriving at the system by the time t with $X(t)$ and the number of customers leaving by time t by $X'(t)$, then

$$X(t) = \{ n : S_n \leqslant t,\ S_{n+1} > t \}$$

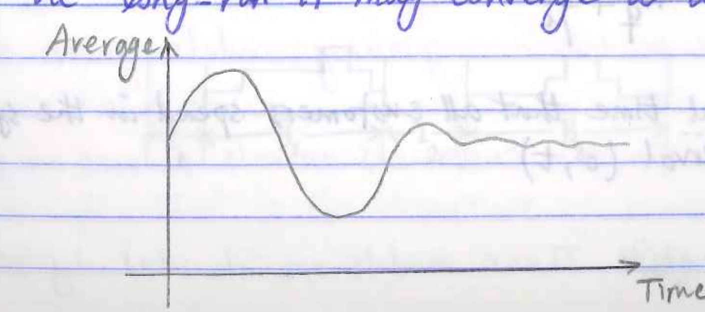
$$X'(t) = \{ n : S'_n \leqslant t,\ S'_{n+1} > t \}$$

$$N(t) = X(t) - X'(t)$$

Where $N(t)$ is the number of customers in the system at time

Transient and stationary state of the system:

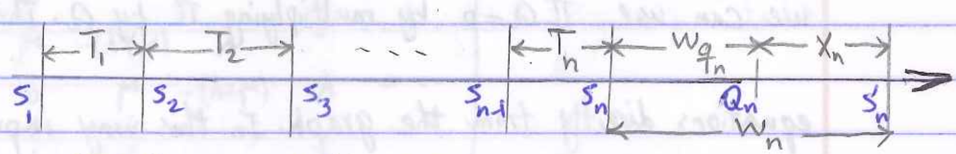
The ma or criteria that we use to study a QS, for instance, th customers in the queue, the sojurn time, the waiting time in the often have stochastic nature and hence we focus on their averag However, these average quantities are functions of time themselv many systems as time surpasses a certain value, these average converge to some fixed value which are independent of time an initial conditions of the system. For instance, if a bank at th of a day may have many customers, then the average number o in the system is high, however as time passes, this average m and in the long-run it may converge to a fixed value.



The general framework of queuing systems:

Assume that the first customer arrives at S_1, the second at $S_2, \dots$, n^{th} at S_n and the time interval of the arrival of customer n and $n-1$ is T_n.

A customer, for instance n, after arriving at the system may have to wait for some time before his service starts at Q_n and will finish at S'_n where the customer leaves the system. The waiting time of this customer in the queue is W_{q_n}, his sojurn time (the waiting time in the system) is W_n and his service time is X_n.



Then we can write these relationships:

$$
\begin{cases}
W_{q_n} = Q_n - S_n \\
W_n = S'_n - S_n \\
X_n = S'_n - Q_n \\
W_n = W_{q_n} + X_n
\end{cases}
$$

Let us denote the number of customers arriving at the system by the time t with $X(t)$ and the number of customers leaving by time t by $X'(t)$, then

$$X(t) = \{n : S_n \leqslant t,\ S_{n+1} > t\}$$

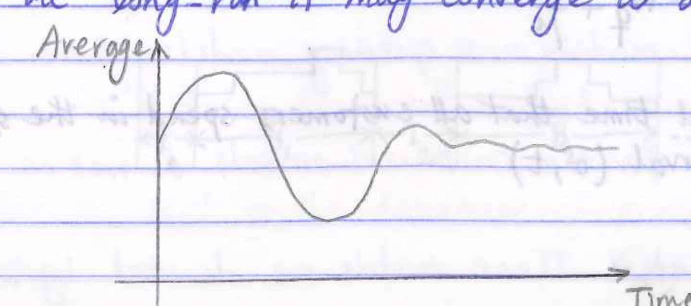
$$X'(t) = \{n : S'_n \leqslant t,\ S'_{n+1} > t\}$$

$$N(t) = X(t) - X'(t)$$

Where $N(t)$ is the number of customers in the system at time t.

Transient and stationary state of the system:

The major criteria that we use to study a QS, for instance, the # of customers in the queue, the sojurn time, the waiting time in the queue, ... often have stochastic nature and hence we focus on their average value. However, these average quantities are functions of time themselves. In many systems as time surpasses a certain value, these average quantities converge to some fixed value which are independent of time and the initial conditions of the system. For instance, if a bank at the begining of a day may have many customers, then the average number of customers in the system is high, however as time passes, this average may reduce and in the long-run it may converge to a fixed value.



we define the transient period as the time interval that the average of the system's performance criteria changes frequently. The stationary period is the time interval after which these quantities become independent of time and the initial conditions of the system.

The relationship between the performance criteria in the system:

The major criteria for studying a QS in the long-run are:

$$L = \lim_{t\to\infty} \{ \underbrace{E[N(t)]}_{L(t)} \}$$

$$W = \lim_{t\to\infty}\{E[W_t(N(t))]\} = \lim_{t\to\infty} E\{\text{a customer waiting time in the system at time } t\}$$

Sojourn time

$$W_q = \lim \{E[W_{q_t}(N(t))]\}$$

$$\pi_n = \lim_{t\to\infty} P_n(t) = \lim_{t\to\infty} \mathbb{P}(n \text{ customers in the system at time } t)$$

Little's Law:

$$\begin{cases} L = \lambda W \\ L_q = \lambda W_q \\ W = W_q + \frac{1}{\mu} \end{cases}$$

Proof. T = total time that all customers spend in the system in the interval $(0,t)$

$$\begin{cases} T = W_t \cdot X(t) \\ T = L_t \cdot t \end{cases} \Rightarrow W_t \cdot X(t) = L_t \cdot t \Rightarrow L_t = \frac{X(t)}{t} \cdot W_t$$

by taking the limit $L = \lambda \cdot W$

The relationship between L and π_n:

$$L = \sum_{n=0}^{\infty} n\,\pi_n$$

Obtaining L by Z-transform:

$$P(z) = \sum_{n=0}^{\infty} \pi_n z^n \Rightarrow L = \frac{dP(z)}{dz}\Big|_{z=1}$$

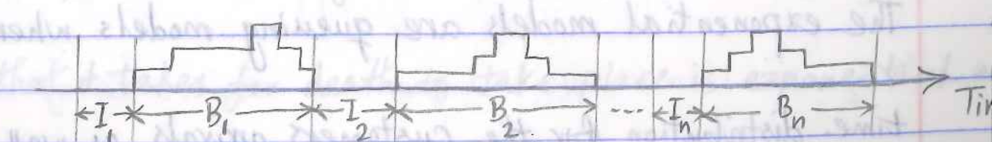
Similarly,

$$L_q = \sum_{n=0}^{\infty} n\,\mathbb{P}\{n \text{ customers in the queue}\} = \sum_{n=m}^{\infty} (n-m)\,\pi_n$$

↳ the # of serve

Idle and Busy period in a queuing system:

Any queuing system can be idle or inactive for some time. The i starts from the time the last customer departs the system and until a new customer arrives at the system. Then the busy perio and continues until the system becomes fully empty again.



we define the transient period as the time interval that the average of the system's performance criteria changes frequently. The stationary period is the time interval after which these quantities become independent of time and the initial conditions of the system.

The relationship between the performance criteria in the system :

The major criteria for studying a QS in the long-run are:

$$L = \lim_{t\to\infty}\{\underbrace{E[N(t)]}_{L(t)}\}$$

$$\underbrace{W}_{\text{Sojourn time}} = \lim_{t\to\infty}\{E[W_t(N(t))]\} = \lim_{t\to\infty} E\{\text{a customer waiting time in the system at time } t\}$$

$$W_q = \lim\{E[W_{q_t}(N(t))]\}$$

$$\pi_n = \lim_{t\to\infty} P_n(t) = \lim_{t\to\infty} \mathbb{P}(n \text{ customers in the system at time } t)$$

Little's Law :

$$\begin{cases} L = \lambda W \\ L_q = \lambda W_q \\ W = W_q + \frac{1}{\mu} \end{cases}$$

Proof. T = total time that all customers spend in the system in the interval $(0,t)$

$$\begin{cases} T = W_t \cdot X(t) \\ T = L_t \cdot t \end{cases} \Rightarrow W_t \cdot X(t) = L_t \cdot t \Rightarrow L_t = \frac{X(t)}{t} \cdot W_t$$

by taking the limit $L = \lambda \cdot W$

The relationship between L and π_n :

$$L = \sum_{n=0}^{\infty} n\,\pi_n$$

Obtaining L by Z-transform :

$$P(z) = \sum_{n=0}^{\infty} \pi_n z^n \Rightarrow L = \frac{dP(z)}{dz}\Big|_{z=1}$$

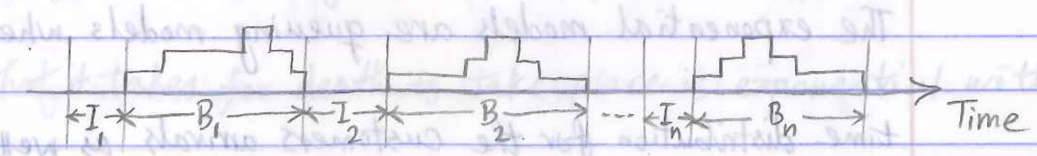
Similarly,

$$L_q = \sum_{n=0}^{\infty} n\,\mathbb{P}\{n \text{ customers in the queue}\} = \sum_{n=m}^{\infty} (n-m)\,\pi_n$$

↳ the # of servers

Idle and Busy period in a queuing system :

Any queuing system can be idle or inactive for some time. The idle period starts from the time the last customer departs the system and continues until a new customer arrives at the system. Then the busy period starts and continues until the system becomes fully empty again.



It is clear that idle and busy period are r.v.s. Assume that P_b is the percentage of time that the system is busy, then

$$P_b = \lim_{n\to\infty} \frac{(B_1+B_2+\cdots+B_n)}{(B_1+B_2+\cdots+B_n)+(I_1+I_2+\cdots+I_n)} = \lim_{n\to\infty} \frac{\frac{\sum B_i}{n}}{\frac{\sum B_i}{n}+\frac{\sum I_i}{n}}$$

However, by strong law of large numbers we know that:

$$\mathbb{P}\left(\lim_{n\to\infty} \frac{\sum B_i}{n} = E(B)\right) = 1$$

Hence,

$$P_b = \frac{E(B)}{E(B)+E(I)}$$

Utilization Factor:

$$\rho = \frac{\text{the average demand rate for recieving the service}}{\text{the average service rate}}$$

$$= \frac{\overset{\to \text{ demand rate (arrival rate)}}{\lambda}}{\underset{\#\text{ of servers}}{m}\cdot\underset{\text{service rate}}{\mu}}$$

We assume that $\rho < 1$, if $\rho > 1$ then the system will explode.

Exponential models of QS:

The exponential models are queuing models where the inter-arrival time distribution for the customers arrivals as well as the service process are both exponential. These models are denoted by M/M/...

Birth-Death process:

Consider a population where its size at each moment can incre(birth) or decrease (death). In a queuing systems, the custom the system play the role of population. In a B-D process arrival new customer is called a birth and departure of a customer recieve the service is called a death. There are 2 assumption a B-D process.

Assumption 1) Consider a moment where the system's populat The time that it takes for a new birth to take place (a new cus arrives) is a r.v. with exponential distribution with paramete other words, the prob. of arrival of a new customer at the syst short time Δt is $\lambda_n \Delta t$ and the prob. that more than one custo arriving is $o(\Delta t)$. The arrival rate (or the birth rate) may depe size of the population, but it is independent of time.

Assumption 2) Consider a system with population n (at a certa The time that it takes for death to take place is exponential w parameter μ_n, that is the departure of one customer in a short

It is clear that idle and busy period are r.v$_s$. Assume that P_b is the percentage of time that the system is busy, then

$$P_b = \lim_{n\to\infty} \frac{(B_1+B_2+\cdots+B_n)}{(B_1+B_2+\cdots+B_n)+(I_1+I_2+\cdots+I_n)} = \lim_{n\to\infty} \frac{\frac{\sum B_i}{n}}{\frac{\sum B_i}{n}+\frac{\sum I_i}{n}}$$

However, by strong law of large numbers we know that:

$$\mathbb{P}\left(\lim_{n\to\infty} \frac{\sum B_i}{n} = E(B)\right) = 1$$

Hence,

$$P_b = \frac{E(B)}{E(B)+E(I)}$$

Utilization Factor:

$$\rho = \frac{\text{the average demand rate for recieving the service}}{\text{the average service rate}}$$

$$= \frac{\overset{\text{demand rate (arrival rate)}}{\lambda}}{\underset{\text{\# of servers}}{m}\cdot\underset{\text{service rate}}{\mu}}$$

We assume that $\rho < 1$, if $\rho > 1$ then the system will explode.

Exponential models of QS:

The exponential models are queuing models where the inter-arrival time distribution for the customers arrivals as well as the service process are both exponential. These models are denoted by $M/M/\ldots$

Birth-Death process:

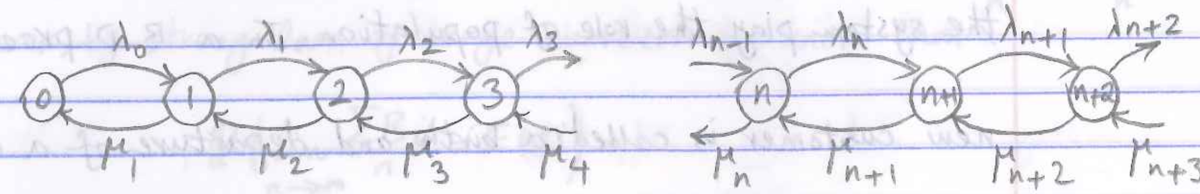
Consider a population where its size at each moment can increase (birth) or decrease (death). In a queuing systems, the customers in the system play the role of population. In a B-D process arrival of a new customer is called a birth and departure of a customer after she recieve the service is called a death. There are 2 assumption made in a B-D process.

Assumption 1) Consider a moment where the system's population is n. The time that it takes for a new birth to take place (a new customer arrives) is a r.v. with exponential distribution with parameter λ_n. In other words, the prob. of arrival of a new customer at the system at a short time Δt is $\lambda_n \Delta t$ and the prob. that more than one customer arriving is $o(\Delta t)$. The arrival rate (or the birth rate) may depend on the size of the population, but it is independent of time.

Assumption 2) Consider a system with population n (at a certain time). The time that it takes for death to take place is exponential with parameter μ_n, that is the departure of one customer in a short time is

$\mu_n \Delta t$, and the departure of more than one customer in Δt is $o(\Delta t)$. μ_n may depends on n but it is independent of t.

Birth – Death process as MC :



$$\begin{cases}\lambda_0 \pi_0 = \mu_1 \pi_1 \\ (\lambda_n + \mu_n)\pi_n = \lambda_{n-1}\pi_{n-1} + \mu_{n+1}\pi_{n+1}\end{cases}$$

$$\pi_1 = \frac{\lambda_0}{\mu_1}\pi_0$$

$$\pi_2 = \left(\frac{\lambda_0}{\mu_1}\right)\left(\frac{\lambda_1}{\mu_2}\right)\pi_0$$

$$\vdots$$

$$\pi_n = \underbrace{\frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n}}_{C_n} \pi_0 \Rightarrow \begin{cases}\pi_n = C_n \pi_0 \; ; \; n = 1, 2, \ldots \\ \sum_{n=0}^{\infty} \pi_n = 1\end{cases}$$

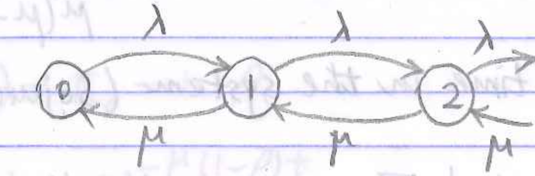
$$\Rightarrow \pi_0 \left(1 + \sum_{n=1}^{\infty} C_n\right) = 1 \Rightarrow \pi_0 = \frac{1}{1 + \sum_{n=1}^{\infty} C_n}, \quad C_n = \frac{\prod_{i=0}^{n-1} \lambda_i}{\prod_{i=1}^{n} \mu_i}$$

The stability condition :

$\sum_{n=1}^{\infty} C_n$ must be convergent. Suppose that $\sum C_n$ is not convergent, then $\pi_0 = 0$. Hence, $\pi_n = 0$ for any real value of n. That means that the system with prob. 1 has infinite number of customers.

Special cases of B-D process :

1. M/M/1 model :



$$\begin{cases} C_n = \left(\frac{\lambda}{\mu}\right)^n = \rho^n \\ \pi_0 = \left[1 + \sum_{n=1}^{\infty} \rho^n\right]^{-1} = \left[1 + \frac{\rho}{1-\rho}\right]^{-1} = \left[\frac{1}{1-\rho}\right]^{-1} = 1 - \rho \\ \pi^n = \rho^n (1-\rho) \end{cases}$$

Example. Consider a library such that :

The arrival rate of the members $= \lambda = 10$ members / hour

The service time of each member $= \frac{1}{\mu} = 5$ mins $\Rightarrow \mu = 12$ /h

A) What is the percentage of idle time?

B) What is the prob. that 3 members wait in the queue?

A) $\pi_0 = 1 - \rho = 1 - \frac{10}{12} = \frac{1}{6}$ Note: Busy time P_b, idle time π_0

B) $\pi_4 = \rho^4 (1-\rho) = (5/6)^4 (1/6) = 0.08$

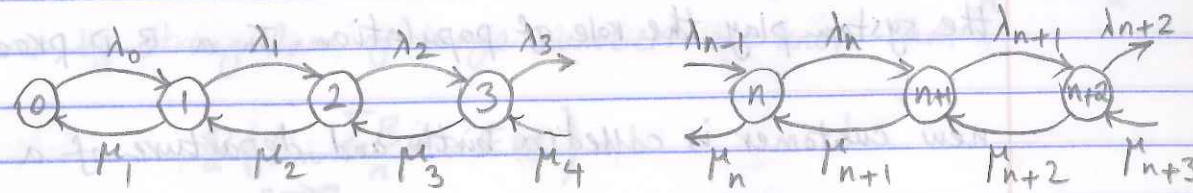
The average number of customers in the system :

$$L = \sum_{n=0}^{\infty} n \pi_n = \sum_{n=0}^{\infty} n \rho^n (1-\rho) = (1-\rho)\rho \underbrace{\sum_{n=0}^{\infty} n \rho^{n-1}}_{\frac{1}{(1-\rho)^2}} = \frac{\rho}{1-\rho}$$

$$L_q = \sum_{n=1}^{\infty} (n-1)\pi_n = \sum_{n=1}^{\infty} n \pi_n - \sum_{n=1}^{\infty} \pi_n = L - (1 - \pi_0)$$

$\mu_n \Delta t$, and the departure of more than one customer in Δt is $o(\Delta t)$. μ_n may depends on n but it is independent of t.

Birth – Death process as MC :



$$\begin{cases} \lambda_0 \pi_0 = \mu_1 \pi_1 \\ (\lambda_n + \mu_n)\pi_n = \lambda_{n-1}\pi_{n-1} + \mu_{n+1}\pi_{n+1} \end{cases}$$

$$\pi_1 = \frac{\lambda_0}{\mu_1}\pi_0$$

$$\pi_2 = \left(\frac{\lambda_0}{\mu_1}\right)\left(\frac{\lambda_1}{\mu_2}\right)\pi_0$$

$$\vdots$$

$$\pi_n = \underbrace{\frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n}}_{C_n}\pi_0 \Longrightarrow \begin{cases} \pi_n = C_n \pi_0 \quad ; n = 1, 2, \dots \\ \sum_{n=0}^{\infty} \pi_n = 1 \end{cases}$$

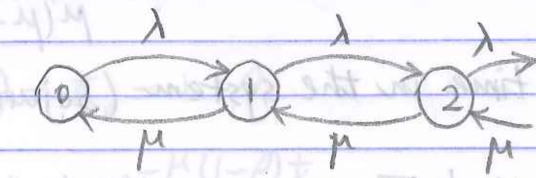
$$\Rightarrow \pi_0 \left(1 + \sum_{n=1}^{\infty} C_n\right) = 1 \Rightarrow \pi_0 = \frac{1}{1 + \sum_{n=1}^{\infty} C_n} \quad , \quad C_n = \frac{\prod_{i=0}^{n-1} \lambda_i}{\prod_{i=1}^{n} \mu_i}$$

The stability condition :

$\sum_{n=1}^{\infty} C_n$ must be convergent. Suppose that $\sum C_n$ is not convergent, then $\pi_0 = 0$. Hence, $\pi_n = 0$ for any real value of n. That means that the system with prob. 1 has infinite number of customers.

Special cases of B-D process :

1. M/M/1 model :



$$\begin{cases} C_n = \left(\frac{\lambda}{\mu}\right)^n = \rho^n \\ \pi_0 = \left[1 + \sum_{n=1}^{\infty} \rho^n\right]^{-1} = \left[1 + \frac{\rho}{1-\rho}\right]^{-1} = \left[\frac{1}{1-\rho}\right]^{-1} = 1 - \rho \\ \pi^n = \rho^n (1-\rho) \end{cases}$$

Example. Consider a library such that :

The arrival rate of the members $= \lambda = 10$ members / hour

The service time of each member $= \frac{1}{\mu} = 5$ mins $\Longrightarrow \mu = 12$ /hr

A) What is the percentage of idle time?

B) What is the prob. that 3 members wait in the queue?

A) $\pi_0 = 1 - \rho = 1 - \frac{10}{12} = \frac{1}{6}$ Note : Busy time P_b, idle time π_0

B) $\pi_4 = \rho^4 (1-\rho) = (5/6)^4 (1/6) = 0.08$

The average number of customers in the system :

$$L = \sum_{n=0}^{\infty} n \pi_n = \sum_{n=0}^{\infty} n \rho^n (1-\rho) = (1-\rho)\rho \underbrace{\sum_{n=0}^{\infty} n \rho^{n-1}}_{\frac{1}{(1-\rho)^2}} = \frac{\rho}{1-\rho}$$

$$L_q = \sum_{n=1}^{\infty} (n-1)\pi_n = \sum_{n=1}^{\infty} n \pi_n - \sum_{n=1}^{\infty} \pi_n = L - (1 - \pi_0)$$

$$\Rightarrow L_q = L - (1-\pi_0) \Rightarrow L_q = L - \rho = \frac{\rho}{1-\rho} - \rho = \frac{\rho^2}{1-\rho}$$
$$= \frac{\lambda^2}{\mu(\mu-\lambda)}$$

Waiting time in the system (Sojurn time):

Suppose that T_s represents the time spent in the system and T_q is the time spent in the queue. Then

$$W = E(T_s) \quad \text{and} \quad W_q = E(T_q)$$

by Little's Law:

$$L = \lambda W \Rightarrow W = \frac{L}{\lambda} = \frac{1}{\lambda}\left(\frac{\lambda}{\mu-\lambda}\right) = \frac{1}{\mu-\lambda}$$

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

or

$$W_q = \frac{L_q}{\lambda} \Rightarrow W_q = \frac{1}{\lambda}\left(\frac{\lambda^2}{\mu(\mu-\lambda)}\right) = \frac{\lambda}{\mu(\mu-\lambda)}$$

Example. For the previous example:

$W = \frac{1}{12-10} = \frac{1}{2}$ hour , $W_q = \frac{\lambda}{\mu(\mu-\lambda)} = \frac{10}{12(2)} = \frac{10}{24} = 25$ mins

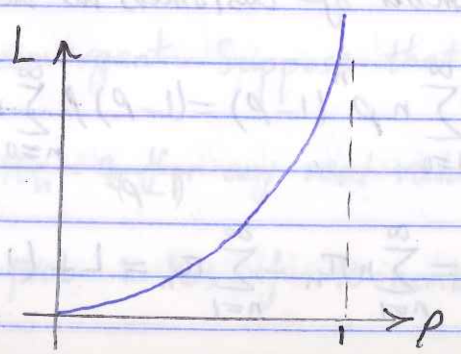
$L = \lambda \cdot W = (10)\left(\frac{1}{2}\right) = 5$, $L_q = \lambda \cdot W_q = 10\left(\frac{10}{24}\right) = 4.17$

or

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

Note:

$L = \frac{\rho}{1-\rho} \Rightarrow$



Waiting time distribution:

Theorem. In an M/M/1 model we have

$$\mathbb{P}(T_q = 0) = \pi_0 = 1-\rho$$
$$\mathbb{P}(T_q > t) = \rho e^{-\mu(1-\rho)t}$$
$$\mathbb{P}(T > t) = e^{-\mu(1-\rho)t}$$

Proof.

$$\mathbb{P}(T_s > t) = \sum_{n=0}^{\infty} \mathbb{P}(T_s > t \mid N=n) \underbrace{\mathbb{P}(N=n)} \quad \pi_n = \rho^n(1-\rho)$$

on the other side, the time that a customer waits in the queue is

$T_q = X_1 + X_2 + \cdots + X_n$ where X_i is the service time to customer i.

Similarly, $T_s = T_q + X_{n+1}$ where X_{n+1} is the service time of the customer himself.

$$\mathbb{P}(T_s > t \mid N=n) = \mathbb{P}\left(\sum_{i=1}^{n+1} X_i > t\right) = \int_t^{\infty} \mu e^{-\mu y} \frac{(\mu y)^n}{n!}\, dy$$

$* \sum_{i=1}^{n+1} X_i \sim \text{Erlang}(\mu, n+1)$

Hence,

$$\mathbb{P}(T_s > t) = \sum_{n=0}^{\infty} \rho^n(1-\rho) \int_t^{\infty} \mu e^{-\mu y} \frac{(\mu y)^n}{n!}\, dy$$
$$= \mu(1-\rho)\int_t^{\infty} e^{-\mu y} \overbrace{\sum \frac{\rho^n(\mu y)^n}{n!}}^{e^{\rho\mu y} = e^{\lambda y}} dy$$
$$= \mu(1-\rho)\int_t^{\infty} e^{(\lambda-\mu)y}\, dy = \underbrace{\mu(1-\rho)\frac{1}{\lambda-\mu}}_{\rho-1} e^{(\lambda-\mu)y}\Big|_t^{\infty}$$

$$= e^{(\lambda-\mu)t} = e^{-\mu(1-\rho)t} = e^{-\lambda t}$$

As a result, the sojurn time distribution is exponential with parameter $(\mu-\lambda)$.

Hence,

$$E(T_s) = \frac{1}{\mu-\lambda} = W$$

Obtaining the desired Q length:

$$\mathbb{P}(\text{queue length} \geq n) = \mathbb{P}(\#\text{ of customers in the system} \geq n+1)$$

$$= \sum_{i=n+1}^{\infty} \pi_i = \sum_{i=n+1}^{\infty} \rho^i (1-\rho) = (1-\rho)\rho^{n+1} \underbrace{\sum_{i=n+1}^{\infty} \rho^{i-(n+1)}}_{\frac{1}{1-\rho}} = \rho^{n+1}$$

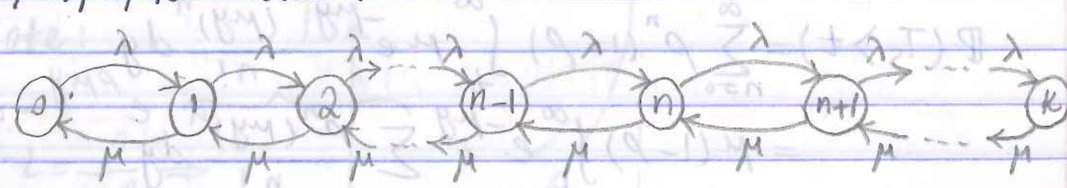
Similarly,

$$L'_q = E[N_q \mid N_q = 0] = \sum_{n=1}^{\infty} (n-1) P'_n = \sum (n-1) \overbrace{\mathbb{P}(n \text{ customers in the system} \mid n \geq 2)}^{A}$$

$$A = \frac{\mathbb{P}(n \text{ customers in the system})}{\mathbb{P}(n \geq 2)} = \frac{\pi_n}{\rho^{1+1}} = \frac{\pi_n}{\rho^2}$$

$$\Rightarrow L'_q = \sum_{n=0}^{\infty} (n-1)\frac{\pi_n}{\rho^2} = \frac{1}{\rho^2}\sum_{n=1}^{\infty}(n-1)\pi_n = \frac{L_q}{\rho^2} = \frac{1}{\rho^2}\left(\frac{\rho^2}{1-\rho}\right) = \frac{1}{1-\rho} = \frac{1}{\pi_0}$$

M/M/1/K model:



$$\lambda_n = \begin{cases} \lambda & ; n < k \\ 0 & ; n \geq k \end{cases}, \quad \mu_n = \mu \; ; n = 1, 2, \ldots, k$$

Customer effective arrival rate:

$$\bar{\lambda} = \lambda \times \mathbb{P}(n < k) + 0 \times \mathbb{P}(n \geq k) \Rightarrow \bar{\lambda} = \lambda\,\mathbb{P}(n < k)$$

Utilization factor:

$$\rho = \frac{\bar{\lambda}}{\mu} = \frac{\lambda \mathbb{P}(n < k)}{\mu}$$

Limiting probability:

$$C_n = \begin{cases} (\lambda/\mu)^n = r^n & ; n \leq k \\ 0 & ; n > k \end{cases}$$

$$\pi_0 = \left(1 + \sum_{n=1}^{k} C_n\right)^{-1} = \frac{1-r}{1-r^{k+1}}, \quad \pi_n = C_n \cdot \pi_0 = \frac{(1-r)r^n}{1-r^{k+1}}$$

For the specific case where $r=1$, we have: $\pi_n = \frac{1}{1+k}$

The average number of customers in the system:

$$L = \sum_{n=0}^{k} n\pi_n = \frac{(1-r)}{1-r^{k+1}} \sum_{n=0}^{k} n r^{n-1}$$

$$* \sum_{n=0}^{k} n r^{n-1} = \sum_{n=0}^{k} \frac{d(r^n)}{dr} = \frac{d}{dr}\left(\sum r^n\right) = \frac{d}{dr}\left(\frac{1-r^{k+1}}{1-r}\right) = \frac{1-(1+k)r^{k+}}{(1-r)^2}$$

$$\Rightarrow L = \frac{r}{1-r} - \frac{(k+1)r^{k+1}}{1-r^{k+1}}$$

The average number of customers in the queue:

$$L_q = L - (1-\pi_0) = L - \frac{r(1-r^k)}{1-r^{k+1}}$$

In order to obtain W and W_q we use Little's Law, we notice th[e]
Little's law we need to use $\bar{\lambda}$ instead of λ, that is $L = \bar{\lambda} W$.

$$= e^{(\lambda-\mu)t} = e^{-\mu(1-\rho)t} = e^{-\lambda t}$$

As a result, the sojurn time distribution is exponential with parameter $(\mu-\lambda)$.

Hence,

$$E(T_s) = \frac{1}{\mu-\lambda} = W$$

Obtaining the desired Q Length:

$$\mathbb{P}(\text{queue length} \geq n) = \mathbb{P}(\# \text{ of customers in the system} \geq n+1)$$

$$= \sum_{i=n+1}^{\infty} \pi_i = \sum_{i=n+1}^{\infty} \rho^i (1-\rho) = (1-\rho)\rho^{n+1} \underbrace{\sum_{i=n+1}^{\infty} \rho^{i-(n+1)}}_{\frac{1}{1-\rho}} = \rho^{n+1}$$

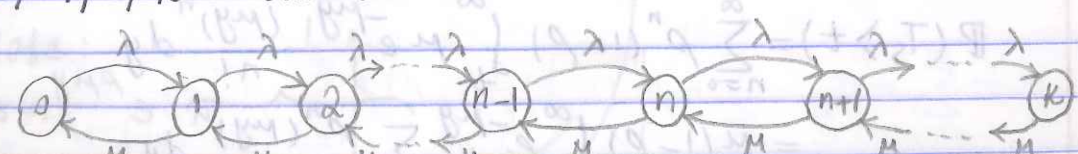
Similarly,

$$L_q' = E[N_q \mid N_q = 0] = \sum_{n=1}^{\infty} (n-1) P_n' = \sum (n-1) \overbrace{\mathbb{P}(n \text{ customers in the system} \mid n \geq 2)}^{A}$$

$$A = \frac{\mathbb{P}(n \text{ customers in the system})}{\mathbb{P}(n \geq 2)} = \frac{\pi_n}{\rho^{1+1}} = \frac{\pi_n}{\rho^2}$$

$$\Rightarrow L_q' = \sum_{n=0}^{\infty} (n-1)\frac{\pi_n}{\rho^2} = \frac{1}{\rho^2}\sum_{n=1}^{\infty}(n-1)\pi_n = \frac{L_q}{\rho^2} = \frac{1}{\rho^2}\left(\frac{\rho^2}{1-\rho}\right) = \frac{1}{1-\rho} = \frac{1}{\pi_0}$$

M/M/1/K model:



$$\lambda_n = \begin{cases} \lambda & ; n < k \\ 0 & ; n > k \end{cases}, \quad \mu_n = \mu \; ; n = 1, 2, \dots, k$$

Customer effective arrival rate:

$$\bar{\lambda} = \lambda \times \mathbb{P}(n<k) + 0 \times \mathbb{P}(n \geq k) \Rightarrow \bar{\lambda} = \lambda \mathbb{P}(n<k)$$

Utilization factor:

$$\rho = \frac{\bar{\lambda}}{\mu} = \frac{\lambda \mathbb{P}(n<k)}{\mu}$$

Limiting probability:

$$C_n = \begin{cases} (\lambda/\mu)^n = r^n & ; n \leq k \\ 0 & ; n > k \end{cases}$$

$$\pi_0 = \left(1 + \sum_{n=1}^{k} C_n\right)^{-1} = \frac{1-r}{1-r^{k+1}}, \quad \pi_n = C_n \pi_0 = \frac{(1-r)r^n}{1-r^{k+1}}$$

For the specific case where $r=1$, we have: $\pi_n = \frac{1}{1+k}$

The average number of customers in the system:

$$L = \sum_{n=0}^{k} n \pi_n = \frac{(1-r)}{1-r^{k+1}} \sum_{n=0}^{k} n r^{n-1}$$

$$* \sum_{n=0}^{k} n r^{n-1} = \sum_{n=0}^{k} \frac{d(r^n)}{dr} = \frac{d}{dr}\left(\sum r^n\right) = \frac{d}{dr}\left(\frac{1-r^{k+1}}{1-r}\right) = \frac{1-(1+k)r^{k+1}}{(1-r)^2}$$

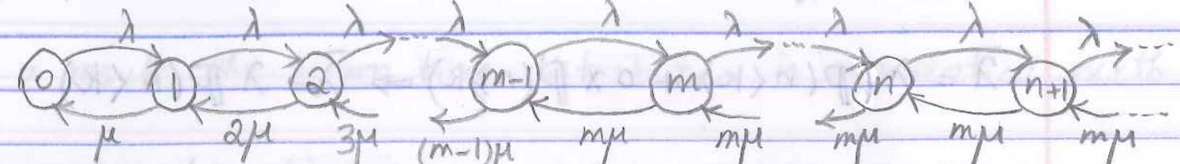
$$\Rightarrow L = \frac{r}{1-r} - \frac{(k+1)r^{k+1}}{1-r^{k+1}}$$

The average number of customers in the queue:

$$L_q = L - (1-\pi_0) = L - \frac{r(1-r^k)}{1-r^{k+1}}$$

In order to obtain W and W_q we use Littles Law, we notice that in Little's law we need to use $\bar{\lambda}$ instead of λ, that is $L = \bar{\lambda} W$.

M/M/m model :



$$\mu_n = \begin{cases} n\mu & n \leq m \\ m\mu & n > m \end{cases}$$

That is because for example when n customers is being served then the time that the system make a transition from n to (n-1) is the minimum of service time for each customer, on the other hand service time has exponential distribution and so the minimum of sum of exponential distribution is an exponential distribution with parameter $\sum \mu_i = n\mu$

$\lambda_n = \lambda \; ; n = 1, 2, \ldots$

$$C_n = \frac{\lambda_0 \lambda_1 \lambda_2 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \mu_3 \cdots \mu_n} = \begin{cases} \dfrac{\lambda^n}{\mu^n \cdot n!} & ; \; n < m \\[2mm] \dfrac{\lambda^n}{\mu^n m! \, m^{n-m}} = \dfrac{r^n}{m! \, m^{n-m}} & ; \; n \geq m \end{cases}$$

$$\pi_0 = \left[1 + \sum_{n=1}^{\infty} C_n\right]^{-1} = \left[1 + \sum_{n=1}^{m-1} \frac{r^n}{n!} + \underbrace{\sum_{n=m}^{\infty} \frac{r^n}{m!} \frac{1}{m^{n-m}}}_{*}\right]^{-1}$$

$$* = \frac{r^m}{m!} \sum_{n=m}^{\infty} \left(\frac{r}{m}\right)^{n-m} = \frac{r^m}{m!} \sum_{m=0}^{\infty} (r/m)^m = \frac{r^m}{m!} \sum_{m=0}^{\infty} (r/m)^m = \frac{r^m}{m!} \cdot \frac{1}{1 - r/m} \quad \rho$$

$$\Rightarrow \pi_0 = \left[\sum_{n=0}^{m-1} \frac{r^n}{n!} + \frac{r^m}{m!}\left(\frac{1}{1-\rho}\right)\right]^{-1} \Rightarrow \pi_n = C_n \pi_0 = \begin{cases} \dfrac{r^n}{n!} \pi_0 & , \; n < m \\[2mm] \dfrac{r^n}{m! \, m^{n-m}} \pi_0 & , \; n \geq m \end{cases}$$

Specific case :

$$M/M/2 \Rightarrow \pi_0 = \frac{1-\rho}{1+\rho}$$

$$M/M/3 \Rightarrow \pi_0 = \frac{1-\rho}{1 + 2\rho + 3/2 \, \rho^2}$$

$$L_q = \sum_{n=m}^{\infty} (n-m) \pi_n = \sum_{n=m}^{\infty} (n-m) \frac{r^n \pi_0}{m! \, m^{n-m}} = \frac{\pi_0 r^m}{m!} \overbrace{\sum_{n=m}^{\infty} (n-m) \rho^{n-m}}^{\frac{\rho}{(1-\rho)^2}}$$

$$= \frac{\pi_0}{m!} \cdot r^m \cdot \frac{\rho}{(1-\rho)^2}$$

Specific case :

$$M/M/2 \Rightarrow L_q = \frac{2\rho^3}{1-\rho^2}$$

In order to obtain W, W_q and L, we use Little's Law :

$$W_q = \frac{L_q}{\lambda}, \quad W = W_q + \frac{1}{\mu}, \quad L = \lambda W = \lambda\left(W_q + \frac{1}{\mu}\right) = L_q + \underbrace{\frac{\lambda}{\mu}}_{r}$$

$$\Rightarrow L = \frac{\pi_0}{m!} r^m \frac{\rho}{(1-\rho)^2} + r$$

Waiting time distribution in M/M/m

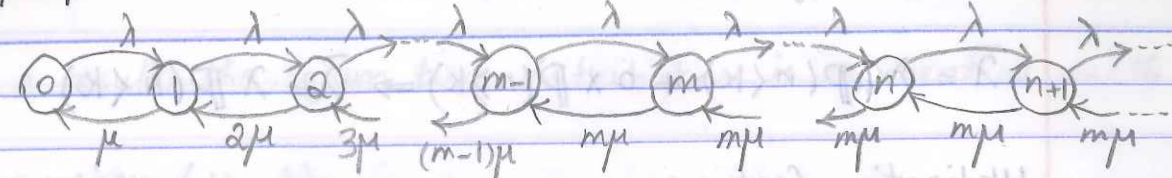
Theorem. $\mathbb{P}(T_q = 0) = 1 - r^m \dfrac{\pi_0}{(1-\rho) m!}$

$$\mathbb{P}(T_q > t) = (1 - \mathbb{P}(T_q = 0)) \cdot e^{-(m\mu - \lambda)t}$$

thi… a rec… distr…

Proof.

M/M/m model:



$$\mu_n = \begin{cases} n\mu & n \le m \\ m\mu & n > m \end{cases}$$

That is because for example when n customers is being served then the time that the system make a transition from n to (n-1) is the minimum of service time for each customer, on the other hand service time has exponential distribution and so the minimum of sum of exponential distribution is an exponential distribution with parameter $\sum \mu_i = n\mu$

$\lambda_n = \lambda \; ; n = 1, 2, \dots$

$$C_n = \frac{\lambda_0 \lambda_1 \lambda_2 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \mu_3 \cdots \mu_n} = \begin{cases} \dfrac{\lambda^n}{\mu^n \cdot n!} \; ; & n < m \\[2ex] \dfrac{\lambda^n}{\mu^n m! \, m^{n-m}} = \dfrac{r^n}{m! \, m^{n-m}} \; ; & n \ge m \end{cases}$$

$$\pi_0 = \Big[1 + \sum_{n=1}^{\infty} C_n\Big]^{-1} = \Big[1 + \sum_{n=1}^{m-1} \frac{r^n}{n!} + \underbrace{\sum_{n=m}^{\infty} \frac{r^n}{m!} \frac{1}{m^{n-m}}}_{*}\Big]^{-1}$$

$$* = \frac{r^m}{m!} \sum_{n=m}^{\infty} \left(\frac{r}{m}\right)^{n-m} = \frac{r^m}{m!} \sum_{m=0}^{\infty} (r/m)^m = \frac{r^m}{m!} \sum_{m=0}^{\infty} (r/m)^m = \frac{r^m}{m!} \cdot \frac{1}{1 - \underbrace{r/m}_{\rho}}$$

$$\Rightarrow \pi_0 = \Big[\sum_{n=0}^{m-1} \frac{r^n}{n!} + \frac{r^m}{m!}\Big(\frac{1}{1-\rho}\Big)\Big]^{-1} \Rightarrow \pi_n = C_n \pi_0 = \begin{cases} \dfrac{r^n}{n!} \pi_0 & , n < m \\[2ex] \dfrac{r^n}{m! \, m^{n-m}} \pi_0 & , n \ge m \end{cases}$$

Specific case:

$M/M/2 \Rightarrow \pi_0 = \dfrac{1-\rho}{1+\rho}$

$M/M/3 \Rightarrow \pi_0 = \dfrac{1-\rho}{1 + 2\rho + 3/2 \, \rho^2}$

$$L_q = \sum_{n=m}^{\infty} (n-m) \pi_n = \sum_{n=m}^{\infty} (n-m) \frac{r^n \pi_0}{m! \, m^{n-m}} = \frac{\pi_0 r^m}{m!} \overbrace{\sum_{n=m}^{\infty} (n-m) \rho^{n-m}}^{\frac{\rho}{(1-\rho)^2}}$$

$$= \frac{\pi_0}{m!} \cdot r^m \cdot \frac{\rho}{(1-\rho)^2}$$

Specific case:

$M/M/2 \Rightarrow L_q = \dfrac{2\rho^3}{1-\rho^2}$

In order to obtain W, W_q and L, we use Little's Law:

$$W_q = \frac{L_q}{\lambda} \; , \; W = W_q + \frac{1}{\mu} \; , \; L = \lambda W = \lambda\Big(W_q + \frac{1}{\mu}\Big) = L_q + \underbrace{\frac{\lambda}{\mu}}_{r}$$

$$\Rightarrow L = \frac{\pi_0}{m!} r^m \frac{\rho}{(1-\rho)^2} + r$$

Waiting time distribution in M/M/m

Theorem.
$$\mathbb{P}(T_q = 0) = 1 - r^m \frac{\pi_0}{(1-\rho) m!}$$
$$\mathbb{P}(T_q > t) = (1 - \mathbb{P}(T_q = 0)) \cdot e^{-(m\mu - \lambda)t}$$
this is not a recognized distribution.

Proof.