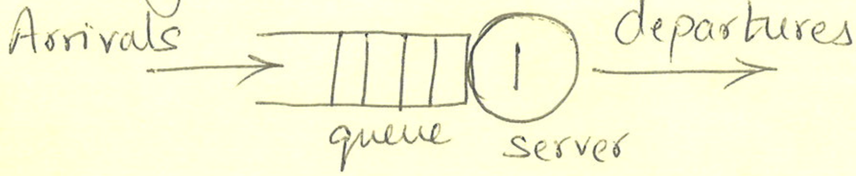
Basic Queuing Theory and Perf. Analysis (1)

A simple model:

1. Single processor, FIFO queue.



Our Assumptions & terms for a steady state FIFO system above

(i) No upper bound on queue length

(ii) Inter-Arrival Times (IATs): time intervals betwn successive arrivals are iid random variables

(iii) Service times: Each job has a service time. Service times of jobs are iid

(iv) avg. IAT > avg. service time (for queue stability)

<u>Note</u>: queue is unstable if avg. queue-size increases with time, without bounds.

Q: Why do we need a mathematical model of a physical sys?

<u>Ans</u>: Because we want to find the foll. performance figures:

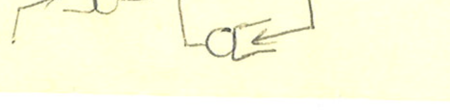
(i) avg. # of customers (pkts) in system.

(ii) avg. time spent by pkt waiting + receiving service (i.e. avg. response time = avg. waiting time + avg. service time).

(iii) System utilization, U

(iv) P(pkt drop) - if dealing with bounded buffers.

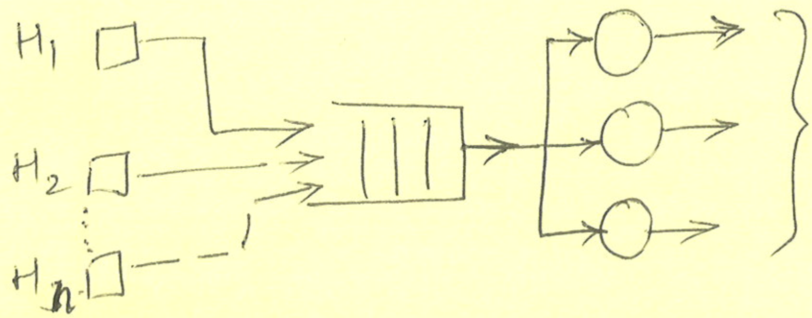
<u>Note</u>: By using interconnections of queues, we can model larger and more complex systems:



etc.

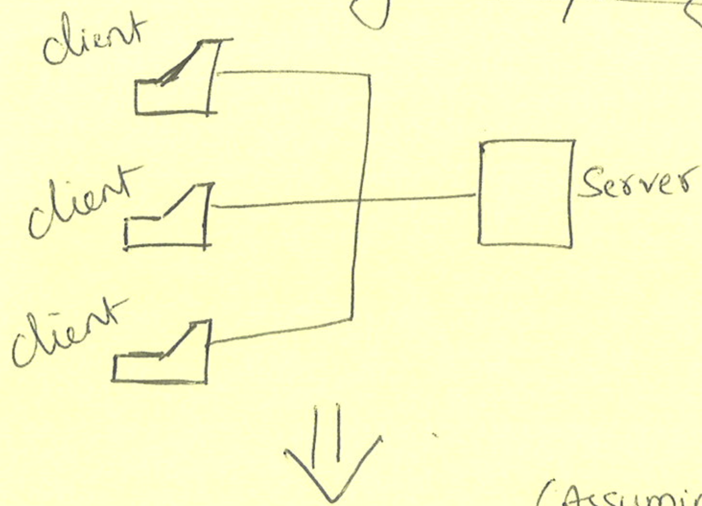
Some more complex models

2. Computationally intensive multiple server system:

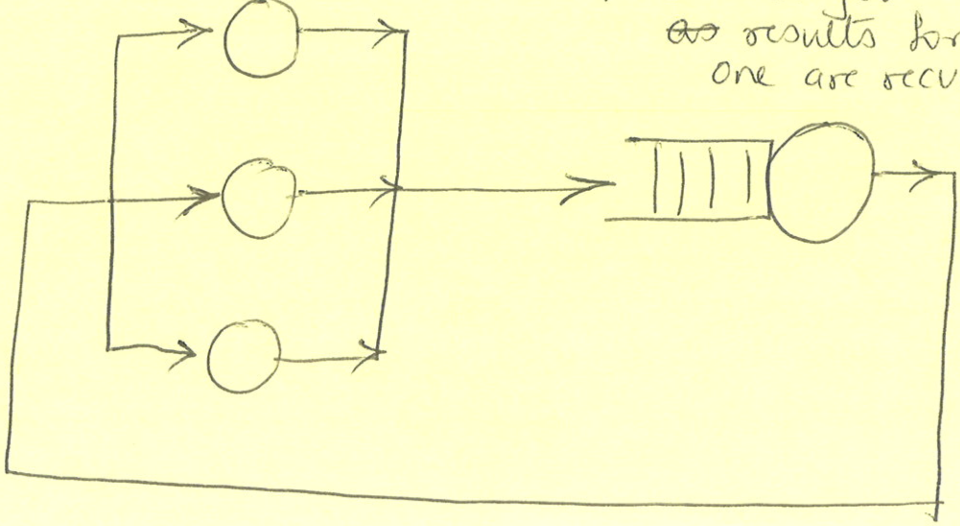


return results to work stations from which they came.

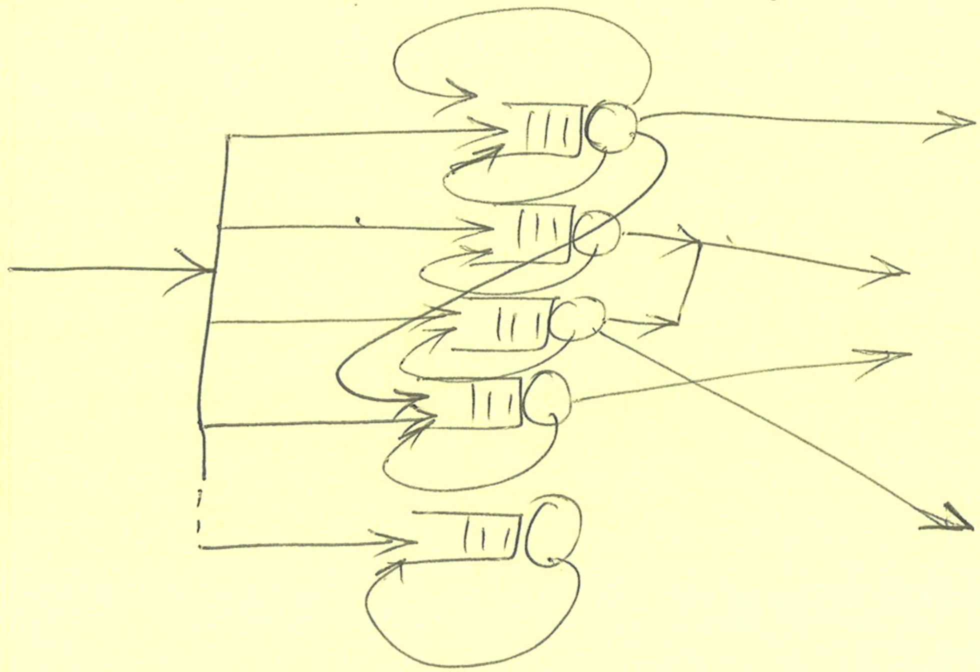
3. Client-server system w/ single server.



(Assuming each client generates only 1 job at a time AND a new job is created ~~when~~ ~~as soon~~ ~~as~~ after results for a previous one are recvd.)



4. A complex telecom/nwk system:



Defn's

1. A random variable is a function x that assigns to each possible outcome in an experiment, a real number.
unique

2. If x can can assume any arbitrary value in some given interval I (the interval need not be bounded), then it is called a continuous r.v.
⇒ e.g: the time taken by a runner to complete a 1 km ~~dis~~ race is a continuous r.v.

3. If x can assume only discrete values, then x is a discrete r.v.
e.g. Rolling a die gives an outcome represented by a discrete r.v.

4. The random number is an outcome of an experiment represented by a random variable.
In other words, when random ~~numbers~~ variables are instantiated, you get a random number.

5. A random variable that is a function of time is given as $x(t)$.
The probabilities of the values that $x(t)$ can take are represented by its probability density function (pdf): $f_X(t)$
e.g. the p.d.f of an exponential r.v. is:

$$f_X(t) = \begin{cases} \lambda e^{-\lambda t} & t \geq 0 \\ 0 & t < 0 \end{cases}$$

The Poisson Process

(traditionally used to model the arrivals of calls to a telephone exchange).

↳ Use of each telephone = Poisson process
↳ Aggregate stream of calls = Aggregate Poisson process
↳ Also used to model TCP connection arrival rates (& pkt arrival rates)

☆ The Poisson process is a purely random arrival process.

Suppose the probability of a call arrival $\propto$ time interval. Consider a very small time interval Δt. Then:

$$P(1 \text{ arrival in the interval } [t, t+\Delta t]) \propto \Delta t$$

$$\therefore P(1 \text{ arrival in } [t, t+\Delta t]) = \lambda \Delta t \qquad -(1)$$

↳ constant.

$$\& \; P(\text{more than } 1 \text{ arrival in } [t, t+\Delta t]) = 0 \qquad -(2)$$

.... since Δt is very small

$$\therefore P(\underline{\text{no}} \text{ arrivals in } [t, t+\Delta t]) = 1 - \lambda \Delta t \qquad -(3).$$

The above is true as long as $\Delta t \to 0$.

~~Now~~

Modeling Poisson (pkt) (call) arrivals.

Assume $t=0$ at the start. Then

let $P_n(t) = P(n \text{ arrivals at time } t)$

let $p_{ij}(\Delta t) = P(\text{going from } i \text{ arrivals to } j \text{ arrivals in time interval } \Delta t)$.

$\therefore P_n(t+\Delta t) = P(n \text{ arr. at time } t)$ AND $p(\text{going from } n \text{ arrivals to } n \text{ arr. in interval } \Delta t)$

OR

$P(n-1 \text{ arr. at time } t)$ AND $p(\text{going from } n-1 \text{ arr. to } n \text{ arr. in interval } \Delta t)$

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

$$\therefore P_n(t+\Delta t) = P_n(t) \cdot \underbrace{(1-\lambda\Delta t)}_{\text{from (3)}} + P_{n-1}(t) \cdot \underbrace{\lambda\Delta t}_{\text{from (1)}} \qquad \ldots\ldots (4)$$

$\therefore$ system looks like this:

$$0 \xrightarrow{\lambda} 1 \xrightarrow{\lambda} 2 \xrightarrow{\lambda} 3 \xrightarrow{\lambda} 4 \xrightarrow{\lambda} \ldots\ldots$$

Each circle above is a "state" in the system.

$\therefore$ the above diagram is called a state transition diagram.

In Eq.(4), substitute $n=0$.

$$\therefore \quad P_0(t+\Delta t) = P_0(t)\cdot(1-\lambda\Delta t) + P_{-1}(t)\cdot\lambda\Delta t \rightarrow \emptyset$$

$$\therefore \quad P_0(t+\Delta t) = P_0(t)\cdot(1-\lambda\Delta t) \quad \cdots\cdots (5)$$

Equations (4) and (5) can be rearranged as:

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

$$\therefore \quad \frac{P_n(t+\Delta t) - P_n(t)}{\Delta t} = -\lambda P_n(t) + \lambda P_{n-1}(t) \quad \cdots\cdots (6)$$

and

$$P_0(t+\Delta t) - P_0(t) = -\lambda\Delta t\cdot P_0(t)$$

$$\therefore \quad \frac{P_0(t+\Delta t) - P_0(t)}{\Delta t} = -\lambda\cdot P_0(t) \quad \cdots\cdots (7)$$

~~The~~ Equations (6) and (7) are difference eqns. If $\Delta t \to 0$, they can be converted to differential equations. So we now, have:

$$\frac{d}{dt}P_n(t) = -\lambda P_n(t) + \lambda P_{n-1}(t) \quad \cdots\cdots (8b)$$

and

$$\frac{d}{dt}P_0(t) = -\lambda P_0(t) \quad \rightarrow (8a)$$

Rearrange (8a) to get:

$$\boxed{dP_0(t)} = -\lambda P_0(t) \cdot \boxed{dt}$$

$$\therefore \quad \frac{dP_0(t)}{P_0(t)} = -\lambda \, dt$$

$\int$ on both sides:

$$\therefore \quad \int \frac{dP_0(t)}{P_0(t)} = \int -\lambda \, dt$$

$$\therefore \quad \int \frac{1}{P_0(t)} \, dP_0(t) = -\lambda \int dt$$

$$\therefore \quad \log[P_0(t)] = -\lambda t$$

$$\therefore \quad e^{\log[P_0(t)]} = e^{-\lambda t}$$

$$\therefore \quad \boxed{P_0(t) = e^{-\lambda t}} \qquad \text{— (9a).}$$

~~Equation (8b) is a linear diffe eqn. of the 1st order of the form: $dy/dt + Py = Q$~~

~~where $y = P_n(t)$, $P = \lambda$ and $Q =$~~

Now, from 8(b) and substituting $n=1$;

$$\frac{d}{dt} P_1(t) = -\lambda P_1(t) + \lambda P_0(t)$$

i.e. $$\frac{d}{dt} P_1(t) = -\lambda P_1(t) + \lambda \cdot \underbrace{e^{-\lambda t}}_{\text{from (9a)}}$$

This is a linear diffe eqn. of the 1st order of the form $\boxed{\frac{dy}{dt} + Py = Q}$

where $y = P_1(t)$, $P = \lambda$ and $Q = \lambda e^{-\lambda t}$.

The soln is given by :

$$\boxed{y \cdot IF = \int Q \cdot IF \cdot dt + C}$$

where $\boxed{IF}$ = Integrating factor $= \boxed{e^{\int P dt}}$

i.e. $IF = e^{\int \lambda dt} = e^{\lambda t}$

$$\therefore y \cdot e^{\lambda t} = \int \lambda \cancel{e^{-\lambda t}} \cdot \cancel{e^{\lambda t}} \, dt + C$$

$$\therefore P_1(t) \cdot e^{\lambda t} = \int \lambda \, dt$$

$$= \lambda t$$

$$\therefore \underline{P_1(t) = \lambda t \cdot e^{-\lambda t}} \quad \text{—} \cdots \text{(9b)}$$

Similarly we get

$$P_2(t) = \frac{\lambda^2 t^2}{2} e^{-\lambda t}$$

using a method similar to that in (9a)

— .... (9c)

$\therefore$ from equations (9a), (9b), (9c) ....
we can use mathematical induction to get

$$\boxed{P_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}} \quad \text{——} \dots (10).$$

This is the Poisson distribution or Poisson PMF

It gives us the probability of n arrivals in an interval of t seconds, when the (arrival) rate of the Poisson process is λ.

Example: A telephone exchange receives 120 calls a minute, on average according to a Poisson process. What is the probability that no calls are received in an interval of 5 seconds?

Sol^n: We need to find $P_0(5)$.
Using the data given and equation (10), we know:

$$P_0(5) = \frac{(2\times5)^0}{0!} \cdot e^{-(2\times5)}$$
$$= 1 \cdot e^{-10}$$
$$= e^{-10}$$
$$= \dots\dots$$

note; here:

$\lambda = 120$ calls/min.
$= 2$ calls/sec
$t = 5$ seconds.
$n = 0$ calls.

Some ~~Other~~ properties of the Poisson ~~~~ distribution ~~~~.

(1). Mean, $E[N] = \sum_{n=0}^{\infty} n \cdot P_n(t) = \sum_{n=0}^{\infty} n \cdot \frac{(\lambda t)^n}{n!} \cdot e^{-\lambda t}$

$\therefore \boxed{E[N] = \lambda t}$ (on solving)

$E[N]$ is also represented as $\bar{n}$.

$\therefore \boxed{\bar{n} = \lambda t}$

(2). Variance, ~~$\sigma(N)$~~ $\sigma_N^2 = E[N^2] - E^2[N]$

$= E[N^2] - \bar{n}^2$

$= \lambda t$.

$\therefore \boxed{\sigma_N^2 = \lambda t}$

(The complete derivation can be found in any book on statistics).

(3.) If the arrival process is Poisson, then the IATs are exponential r.v.s.

Note that ~~the~~ exponential ~~distribution~~ ~~for~~ r.Vs have the "memoryless" property.

i.e. we cannot predict future event(s) based on past observations.

e.g. train station $\longleftrightarrow$ 10 min avg wait betwn trains

Note: Memoryless property $\Rightarrow$ "Markov" property

(4) Merging 2 independent Poisson streams with rates $\lambda_1 + \lambda_2$ results in a Poisson stream with rate $\lambda_1 + \lambda_2$

Service Time Distribution ($\rightarrow$ Exponential distribution).

The exponential distribution
- $\hookrightarrow$ originally used to model duration of telephone calls.
- $\hookrightarrow$ Also used to model time required to transmit data packets (here service time = transmission time).

Some properties of the exponential distribution:

(1) The pdf is given by $f_X(t) = \begin{cases} \cancel{\lambda}\, \mu e^{-\mu t}, & t \geq 0 \\ 0, & t < 0 \end{cases}$

(2) The CDF is $\underbrace{P(X \leq t)}_{F_X(t)} = \cancel{1-e^{-\lambda t}}\,.\ 1 - e^{-\mu t}$

(3) The mean of the exp. r.v. is

$$E[X] = \frac{1}{\mu}$$

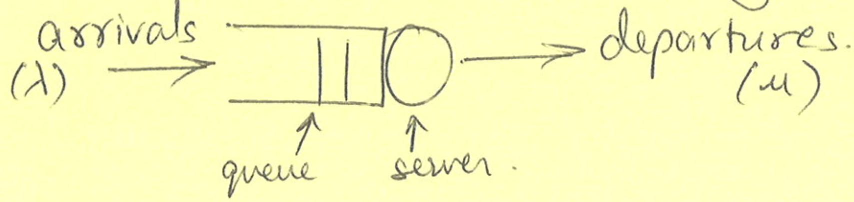
(4) Variance of the exp. r.v. is.

$$\sigma_X^2 = \frac{1}{\mu^2}$$

The M/M/1 Queueing system.

When the arrival process is Poisson, and the departure process (service times) is exponential, then such a system is called a "Markovian system". The server itself is called an "exponential server".

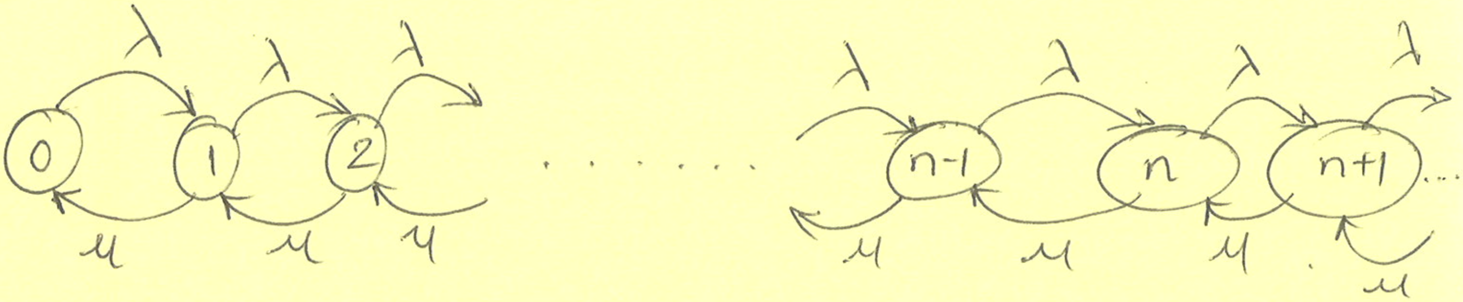
Consider a simple FIFO queuing system:



rate of pkt arrivals is λ pkts/sec.
rate of pkt departures is μ pkts/sec.

Think of the # of pkts in the above system as population size. It increases & decreases due to arrivals/departures. $\Longrightarrow$ called a birth-death process.

A transition diagram for a birth-death process is:



* **Note:** This is a 1-D diagram.

→ this transition is not valid.

Why? [Instantaneous double arrivals/deps not possible]

M/M/1 is short for M/M/1/∞.

- M → Markovian arrivals (Poisson)
- M → Markovian service (exponential)
- 1 → # of servers. (= 1, here)
- ∞ → max. queue capacity

In the figure, and from our ~~assumptions~~ development of the model:

For arrivals.

$$P(1 \text{ arrival in } [t, t+\Delta t]) = \lambda \Delta t \qquad \text{(see eqns (1), (2) and (3))} \quad \text{------ (11)}$$

$$P(\text{no arrivals in } [t, t+\Delta t]) = 1 - \lambda \Delta t \quad \text{------ (12)}$$

For departures:

$$P(1 \text{ service completion in } [t, t+\Delta t]) = \mu \Delta t \quad \text{------ (13)}$$

$$P(\text{no service completions in } [t, t+\Delta t]) = 1 - \mu \Delta t \quad \text{------ (14)}$$

~~and let~~

~~$P_n(t) \equiv P(n \text{ arrivals at time } t)$.~~

~~$P_{ij}(\Delta t) \equiv P(\text{going from } i \text{ arrivals to } j \text{ arrivals in an interval of } \Delta t \text{ seconds})$.~~

and let

* $P_n(t) = P(n \text{ customers in sys. at time } t)$ and

* $P_{ij}(\Delta t) = P(\text{going from } i \text{ customers in sys. to } j \text{ customers in sys. in interval } \Delta t)$.

(new def^ns. for $P_n(t)$ & $P_{ij}(\Delta t)$).

Note: # in system = # in queue + # in server.

Then.

$$P_n(t+\Delta t)$$

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

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

[note no other transitions are allowed by the diagram].

$$1-\mu\Delta t-\lambda\Delta t+\lambda\mu\Delta t^2$$
$$= \underline{\underline{1-\mu\Delta t-\lambda\Delta t}}$$

(if $\Delta t \to 0$ the Δt^2 can be ignored)

$$\lambda\Delta t - \mu\lambda(\Delta t)^2$$
$$= \lambda\Delta t - 0$$
$$= \underline{\underline{\lambda\Delta t}}$$

$$\mu\Delta t - \lambda\mu(\Delta t)^2$$
$$= \underline{\underline{\mu\Delta t}}$$

$\therefore$ $P_n(t+\Delta t)$

$$= P_n(t)\left[1-\mu\Delta t-\lambda\Delta t\right] + P_{n-1}(t)\cdot\lambda\Delta t + P_{n+1}(t)\cdot\mu\Delta t$$

i.e. $$\frac{P_n(t+\Delta t)-P_n(t)}{\Delta t} = -(\lambda+\mu)P_n(t) + \lambda P_{n-1}(t) + \mu P_{n+1}(t)$$

as $\Delta t \to 0$, we can convert the above to a diffe. eqn.

i.e. $$\boxed{\frac{d}{dt}P_n(t) = (-\lambda+\mu)P_n(t) + \lambda\cdot P_{n-1}(t) + \mu\cdot P_{n+1}(t) \quad \text{for } n>0 \qquad (15)}$$

Substitute $n = 0$ in Eq. (15) and we get:

$$\frac{d}{dt} P_0(t) = (-\lambda + \cancel{\mu}) P_0(t) + 0 + \mu \cdot P_1(t).$$

$$\boxed{\therefore \frac{d}{dt} P_0(t) = -\lambda P_0(t) + \mu P_1(t). \quad \text{for } n=0 \quad \text{--- (16)}}$$

~~Note that $P_n(t)$ actually represents the state of system wherein the system has n customers~~ at ~~time t.~~

~~i.e. P_0~~

☆ Note that $P_n(t)$ represents the Probability that the system will have n customers at time t. If the state of the system is defined by the # of customers in that system, then $P_n(t)$ represents the state probability at time t. i.e. $P_n(t)$ represents the probability that the system is in state n at time t.

Stability of system:

If a system is in equilibrium, or, if it can reach equilibrium, then it is stable.

Equilibrium ≡ "steady state".

When a system reaches equilibrium, its state probabilities will no longer be a function of time.

i.e. $P_0(t) = p_0$, say
$P_1(t) = p_1$, say
$\vdots$
$P_n(t) = p_n$, say

under equilibrium.

* ~~i.e.~~ $p_0, p_1, \ldots, p_n$ are also called the steady state probabilities.

$\therefore$ Under steady state, eqn. (15) yields:

$$\frac{d}{dt} p_n = (-\lambda+\mu) p_n + \lambda p_{n-1} + \mu p_{n+1}$$

$\Rightarrow \emptyset$

$$\boxed{\therefore (\lambda+\mu) p_n = \lambda p_{n-1} + \mu p_{n+1}} \quad \text{— (17)}$$

for $n > 0$

And from (16):

$$0 = -\lambda p_0 + \mu p_1$$

$$\boxed{\therefore \lambda p_0 = \mu p_1} \quad \text{— (18).}$$

for $n = 0$.

* Equations (17) and (18) are called global ~~bance~~ balance equations. why?