# 2 Stochastic Processes

## 2.1 Introduction

Computer systems are

1. dynamic - they can pass through a succession of states as time progresses
2. influenced by events which we consider here as random phenomena

c.f. bank, supermarket, stock exchange

Definition: A stochastic process $\boldsymbol{S}$ is a family of random variables $\{X_t \in \Omega | t \in T\}$, each defined on some (the same for each) *sample space* $\Omega$ for a *parameter space* $T$.

- $T, \Omega$ may be either discrete or continuous

- $T$ is normally regarded as time
  - real time: continuous
  - every month or after job completion: discrete
- $\Omega$ is the set of values each $X_t$ may take
  - bank balance: discrete
  - number of active tasks: discrete
  - time delay in communication network: continuous

## 2.2 Example: The Poisson process

The Poisson process is a renewal process with renewal period (interarrival time) distribution and probability density function (pdf)

$$\begin{aligned} F(x) &= P(X \leq x) = 1 - e^{\lambda x} \\ \text{pdf: } f(x) &= F'(x) = \lambda e^{-\lambda x} \end{aligned}$$

$\lambda$ is the parameter or rate of the Poisson process.

## 2.3 Memoryless Property of the (negative) exponential distribution

If $S$ is an exponential random variable

$$P(S \leq t + s | S > t) = P(S \leq s) \quad \forall t, s \geq 0$$

(i.e. it doesn't matter what happened before time $t$)

Proof

$$
\begin{aligned}
P(S \leq t+s | S > t) &= \frac{P(t < S \leq t+s)}{P(S > t)} \\
&= \frac{P(S \leq t+s) - P(S \leq t)}{1 - P(S \leq t)} \\
&= \frac{1 - e^{-\lambda(t+s)} - (1 - e^{-\lambda t})}{e^{-\lambda t}} \\
& \quad (\lambda \text{ is the rate of the exp. distribution}) \\
&= \frac{e^{-\lambda t} - e^{-\lambda(t+s)}}{e^{-\lambda t}} = 1 - e^{-\lambda s} \\
&= P(S \leq s)
\end{aligned}
$$

## 2.4 More properties of the Poisson process

$$\begin{aligned} P(\text{arrival in } (t, t+h)) &= P(R \leq h) = P(S \leq h) \quad \forall t \\ &= 1 - e^{-\lambda h} \\ &= \lambda h + o(h) \end{aligned}$$

Therefore

1. Probability of an arrival in $(t, t+h)$ is $\lambda h + o(h)$ regardless of process history before $t$

2. Probability of more than one arrival in $(t, t+h)$ is $o(h)$ regardless of process history before $t$

In fact we can take this result as an alternative definition of the Poisson process. From it we can derive the distribution function of the interarrival times (negative exponential) and the Poisson distribution for $N_t$ (the *arrival process*)

$$P(N_t = n) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}.$$

## 2.5 Derivation of the Poisson process

The proof of the claim is by induction on the integers. Let $P_n(t) = (P(N_t = n)$.

Base case $(n = 0)$

$$\begin{aligned}
P_0(t+h) &= P(N_t = 0 \text{ \& no arrivals in } (t, t+h)) \\
&= P_0(t)(1 - \lambda h + o(h)) \quad \text{(by independence)} \\
\implies \frac{P_0(t+h) - P_0(t)}{h} &= -\lambda P_0(t) + o(1) \\
P_0'(t) &= -\lambda P_0(t) \quad \text{as } h \to 0 \\
\implies P_0(t) &= Ce^{-\lambda t} \\
P_0(0) &= P(N_0 = 0) = 1 \text{ therefore } C = 1
\end{aligned}$$

Inductive Step ($n > 0$)

$$
\begin{aligned}
P_n(t+h) &= P_n(t)P_0(h) + P_{n-1}(h)P_1(h) + o(h) \\
&= (1-\lambda h)P_n(t) + \lambda h P_{n-1}(t) + o(h) \\
\Longrightarrow \quad P_n'(t) &= -\lambda P_n(t) + \lambda P_{n-1}(t) \\
\Longrightarrow \quad \frac{d}{dt}(e^{\lambda t}P_n(t)) &= \lambda e^{\lambda t} P_{n-1}(t)
\end{aligned}
$$

Inductive hypothesis:

$$P_{n-1}(t) = e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!}$$

This is OK for $n = 1$. Then

$$\begin{aligned} \frac{d}{dt}(e^{\lambda t} P_n(t)) &= \frac{(\lambda)^n}{(n-1)!} t^{n-1} \\ \Longrightarrow \; e^{\lambda t} P_n(t) &= \frac{(\lambda t)^n}{(n)!} + D \end{aligned}$$

For $t = 0$, $P_n(t) = 0$ for $n \geq 2$ and so $D = 0$. Thus

$$P_n(t) = e^{-\lambda t} \frac{\lambda t}{n!}$$

## 2.6 Derivation of the interarrival time distribution

1. Using the derived Poisson distribution

$$\begin{aligned} P(T \leq t) &= 1 - P(0 \text{ arrivals in } (0,t)) \\ &= 1 - e^{-\lambda t} \end{aligned}$$

(distribution for negative exponential random variable)

2. Directly from assumptions (exercise). Use the memoryless property to simplify the equation

$$P(T \leq t+h) = P((T \leq t) \vee [(T > t) \wedge (1 \text{ arrival in } (t, t+h))])$$

hence derive and solve a differential equation for $P(T \leq t) = F(t)$.

## 2.7 Superposition Property

If $A_1, \ldots, A_n$ are independent Poisson Processes with rates $\lambda_1, \ldots, \lambda_n$ respectively, let there be $K_i$ arrivals in an interval of length $t$ from process $A_i$ $(1 \leq i \leq n)$. Then $K = K_1 + \cdots + K_n$ has Poisson distribution with parameter $\lambda t$ where $\lambda = \lambda_1 + \cdots + \lambda_n$. I.e. the superposition of PPs with rates $\lambda_i$ is a PP with rate $\sum_i \lambda_i$.

Proof

The distribution of K is the convolution of the distributions of the $K_i$ which are Poisson. E.g. if $n = 2$

$$
\begin{aligned}
P(K = k) &= \sum_{i=0}^{k} \frac{(\lambda_1 t)^i}{i!} e^{-\lambda_1 t} \frac{(\lambda_2 t)^{k-i}}{(k-i)!} e^{-\lambda_2 t} \\
&= \frac{e^{-(\lambda_1+\lambda_2)t}}{k!} \sum_{i=0}^{k} \binom{k}{i} (\lambda_1 t)^i (\lambda_2 t)^{k-i} \\
&= e^{-(\lambda_1+\lambda_2)t} \frac{[(\lambda_1 + \lambda_2)t]^k}{k!}
\end{aligned}
$$

as required. The proof for arbitrary $n \geq 2$ is an easy induction on $n$.

## 2.8 Decomposition Property

If a Poisson Process is decomposed into processes $B_1, \ldots, B_n$ by assigning each arrival of $A$ to $B_i$ with independent probability $q_i$ ($\sum q_i = 1$), then $B_1, \ldots, B_n$ are independent Poisson Processes with rates $q_1\lambda, \ldots, q_n\lambda$.

Example: Two parallel processors, $B_1$ and $B_2$. Incoming jobs (Poisson arrivals $A$) are directed to $B_1$ with probability $q_1$ and to $B_2$ with probability $q_2$. Then each processor "sees" a Poisson arrival process for its incoming job stream.

Proof

For $n = 2$, let $K_i$ = number of arrivals to $B_i$ in time $t$ $(i = 1, 2)$, $K = K_1 + K_2$

$$\underbrace{P(K_1 = k_1, K_2 = k_2)}_{\text{joint probability}} = \underbrace{P(K_1 = k_1, K_2 = k_2 | K = k_1 + k_2)}_{\text{binomial}} \; \underbrace{P(K = k_1 + k_2)}_{\text{Poisson}}$$

$$= \frac{(k_1+k_2)!}{k_1!k_2!} q_1^{k_1} q_2^{k_2} \frac{(\lambda t)^{k_1+k_2}}{(k_1+k_2)!} e^{-\lambda t}$$

$$= \frac{(q_1\lambda t)^{k_1}}{k_1!} e^{-q_1\lambda t} \frac{(q_2\lambda t)^{k_2}}{k_2!} e^{-q_2\lambda t}$$

I.e. a product of two independent Poisson distributions

($n \geq 2$: easy induction)

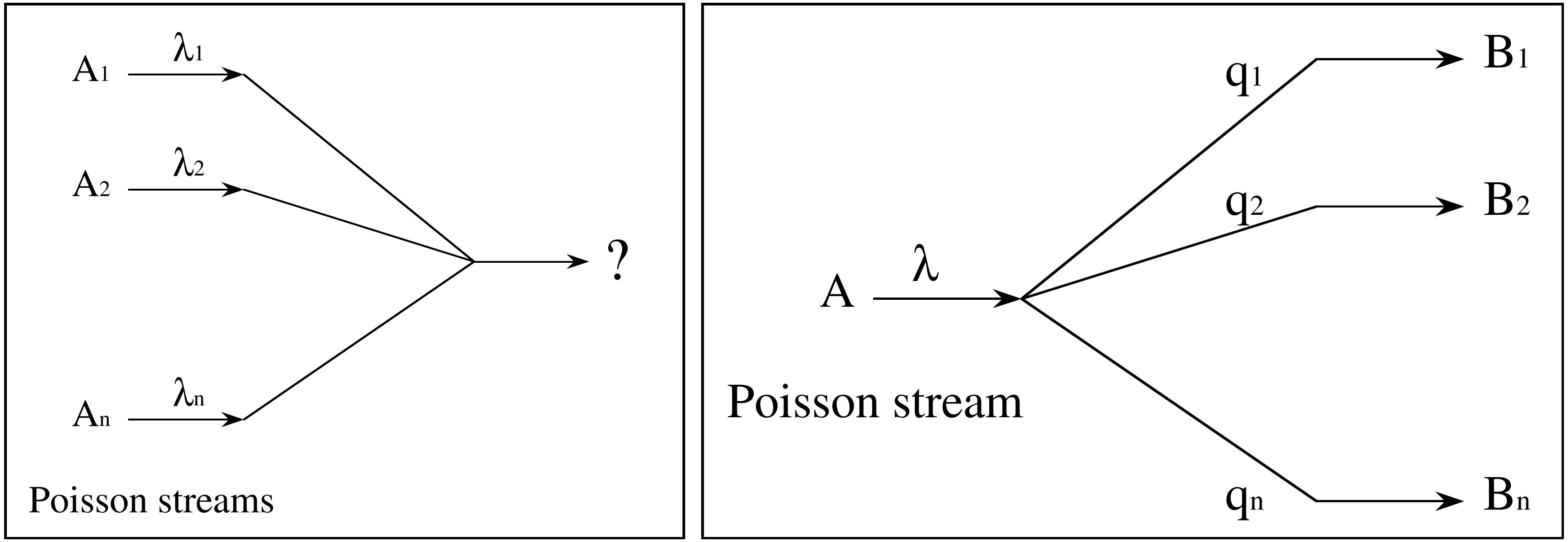


Figure 1: Superposition and decomposition of a Poisson process

## 2.9 Markov Chains and Markov Processes

- Special class of stochastic processes that satisfy the Markov property (MP): Given the state of the process at time $t$, its state at time $t + s$ has probability distribution which is a function of $s$ only.
- i.e. the future behaviour after $t$ is independent of the behaviour before $t$
- Often intuitively reasonable, yet sufficiently "special" to facilitate effective mathematical analysis
- We consider processes with discrete state (sample) space and:
  1. discrete parameter space (times $\{t_0, t_1, \dots\}$), a Markov chain or *Discrete Time Markov Chain*
  2. continuous parameter space (times $t \geq 0, t \in R$), a Markov Process or *Continuous Time Markov Chain.* E.g. number of arrivals in $(0, t)$ from a Poisson arrival process defines a Markov Process because of the memoryless property.

## 2.10 Markov Chains

- Let $X = \{X_n | n = 0, 1, \dots\}$ be an integer valued Markov chain (MC), $X_i \geq 0, X_i \in Z, i \geq 0$. MP states that:

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

$$\text{for } j, n = 0, 1, \dots$$

- Evolution of a MC is completely described by its 1-step transition probabilities

$$q_{ij}(n) = P(X_{n+1} = j | X_n = i) \text{ for } i, j, n, \geq 0$$

Assumption: $q_{ij}(n) = q_{ij}$ is independent of time $n$ (*time homogeneous*)

$$\sum_{j \in \Omega} q_{ij} = 1 \quad \forall i \in \Omega$$

- MC defines a transition probability matrix

$$Q = \begin{bmatrix} q_{00} & q_{01} & \cdots \\ q_{10} & q_{11} & \cdots \\ \vdots & & \\ q_{i0} & \cdots & q_{ii} \cdots \\ \vdots & & \end{bmatrix} \text{ in which all rows sum to 1}$$

  - dimension = # of states in $\Omega$ if finite, otherwise countably infinite
  - conversely, any real matrix $Q$ s.t. $q_{ij} \geq 0$, $\sum_j q_{ij} = 1$ (called a *stochastic matrix*) defines a MC

### 2.10.1 Examples

1. Telephone line:
   Line is either idle (state 0) or busy (state 1).

$$\begin{aligned} \text{Idle at time } n &\implies \text{idle at time } (n+1) \text{ w.p. } 0.9 \\ \text{Idle at time } n &\implies \text{busy at time } (n+1) \text{ w.p. } 0.1 \\ \text{Busy at time } n &\implies \text{idle at time } (n+1) \text{ w.p. } 0.3 \\ \text{Busy at time } n &\implies \text{busy at time } (n+1) \text{ w.p. } 0.7 \end{aligned} \qquad Q = \begin{bmatrix} 0.9 & 0.1 \\ 0.3 & 0.7 \end{bmatrix}$$

   May be represented by a state diagram 2:

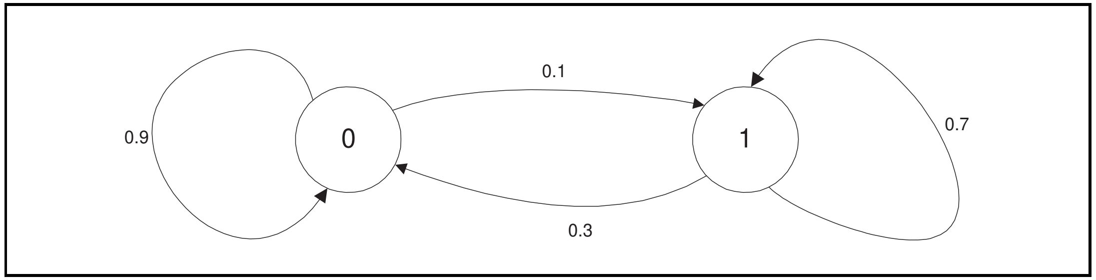


Figure 2: Telephone line example

2. <u>Gambler</u>

   Bets £1 at a time on the toss of fair die. Loses if number is less than 5, wins £2 if 5 or 6. Stops when broke or holds £100. If $X_0$ is initial capital ($0 \leq X_0 \leq 100$) and $X_n$ is the capital held after $n$ tosses $\{X_n | n \leq 0, 1, 2, \ldots\}$ is a MC with $q_{ii} = 1$ if $i = 0$ or $i \geq 100$, $q_{i,i-1} = \frac{2}{3}$, $q_{i,i+2} = \frac{1}{3}$ for $i = 1, 2, \ldots, 99$ $q_{ij} = 0$ otherwise.

3. <u>I/O buffer, capacity M records</u>

   New record added in any unit of time w.p. $\alpha$ (if not full). Buffer emptied in any unit of time w.p $\beta$. If both occur in same interval, insertion done first. Let $X_n$ be the number of records in buffer at (discrete) time $n$. Then, assuming that insertions and emptying are independent of each other and of their own past histories, $\{X_n | n = 0, 1, \ldots\}$ is a MC with state space $\{0, 1, \ldots, M\}$ and state diagram:

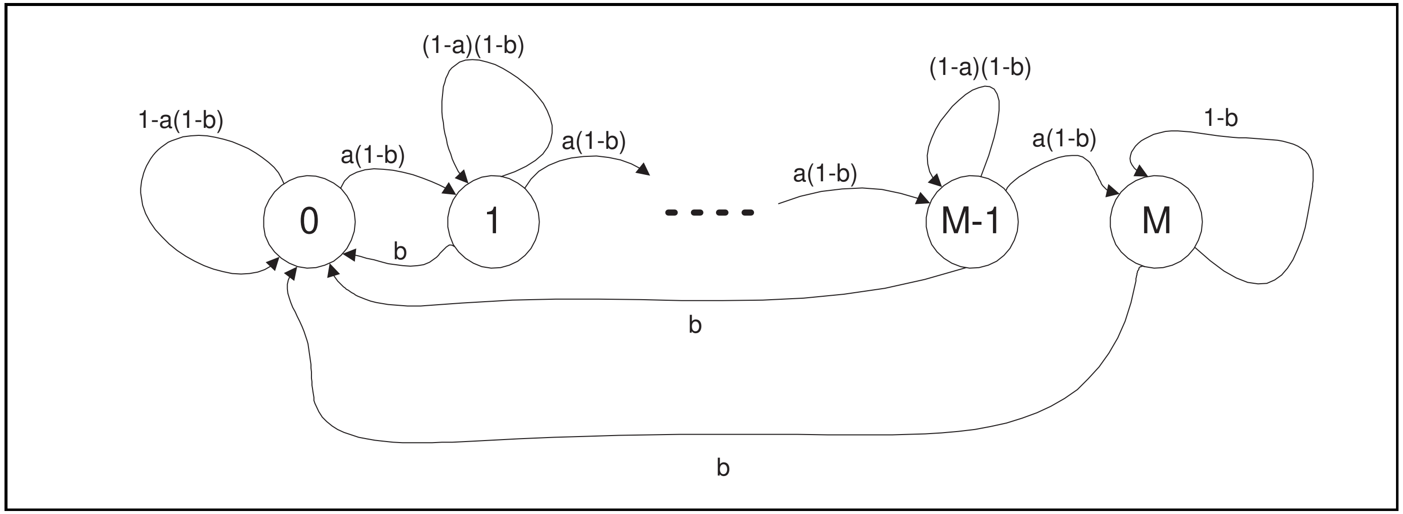


Figure 3: I/O buffer example

The transition matrix follows immediately, e.g.

$$q_{12} = \alpha(1-\beta) = q_{n,n+1}$$

$(0 \leq n \leq M-1)$

$$q_{MM} = 1-\beta$$

etc.

### 2.10.2 Multi-step transition probabilities

Let

$$
\begin{aligned}
q_{ij}^{(2)} &= P(X_{n+2} = j | X_n = i) \\
&= \sum_{k \in \Omega} P(X_{n+1} = k, X_{n+2} = j | X_n = i) \text{ from law of tot. prob.} \\
&= \sum_{k \in \Omega} P(X_{n+2} = j | X_n = i, X_{n+1} = k) P(X_{n+1} = k | X_n = i) \\
&= \sum_{k \in \Omega} P(X_{n+2} = j | X_{n+1} = k) P(X_{n+1} = k | X_n = i) \text{ by MP} \\
&= \sum_{k \in \Omega} q_{ik} q_{kj} \text{ by TH} \\
&= (Q^2)_{ij}
\end{aligned}
$$

Similarly, $m$-step transition probabilities :

$$\begin{aligned} q_{ij}^{(m)} &= P(X_{n+m} = j | X_n = i) \quad (m \geq 1) \\ &= (Q^m)_{ij} \end{aligned}$$

by induction on $m$. Therefore we can compute probabilistic behaviour of a MC over any finite period of time, in principle. E.g. average no. of records in buffer at time 50

$$E[X_{50} | X_0 = 0] = \sum_{j=1}^{\min(M,50)} j q_{0j}^{(50)}$$

Computationally intractable. Need to determine long-term behaviour: hope that asymptotically MC approaches a steady-state (probabilistically), i.e. that $\exists\{p_j | j = 0, 1, 2, \ldots\}$ s.t.

$$p_j = \lim_{n\to\infty} P(X_n = j | X_0 = i) = \lim_{n\to\infty} q_{ij}^{(n)}$$

independent of $i$

### 2.10.3 Definitions and properties of MCs

Let $v_{ij}$ = prob. that state $j$ will eventually be entered, given that the MC has been in state $i$, i.e.

$$v_{ij} = P(X_n = j \text{ for some } n | X_0 = i)$$

- If $v_{ij} \neq 0$, then $j$ is *reachable* from $i$
- If $v_{ij} = 0 \quad \forall j \neq i, v_{ii} = 1$, then $i$ is an *absorbing* state

Example (1) & (3): all states reachable from each other (e.g. look at transition diagrams)
Example (2): all states in $\{1, 2, \ldots, 99\}$ reachable from each other, 0 & 100 absorbing

A subset of states, $C$ is called *closed* if $j \notin C$, implies $j$ cannot be reached from any $i \in C$. E.g. set of all states, $\Omega$, is closed. If $\nexists$ proper subset $C \subset \Omega$ which is closed, the MC is called *irreducible.*
A Markov Chain is irreducible if and only if every state is reachable from every other state. E.g. examples (1) & (3) and $\{0\}$, $\{100\}$ are closed in example (2).

### 2.10.4 Recurrent states

If $v_{ii} = 1$, state $i$ is *recurrent* (once entered, guaranteed eventually to return). Thus we have:
$i$ recurrent implies $i$ is either not visited by the MC or is visited an infinite number of times – no visit can be the last (with non-zero probability).
If $i$ is not recurrent it is called *transient*. This implies the number of visits to a transient state is finite w.p. 1 (has geometric distribution).
E.g. if $C$ is a closed set of states, $i \notin C$, $j \in C$ and $v_{ij} \neq 0$, then $i$ is transient since MC will eventually enter $C$ from $i$, never to return.
E.g. in example (2), states 1 to 99 are transient, states 0 and 100 are recurrent (as is any absorbing state)

Proposition: If $i$ is recurrent and $j$ is reachable from $i$, then $j$ is also recurrent.

Proof: $v_{jj} \neq 1$ implies $v_{ij} \neq 1$ since if the MC visits $i$ after $j$ it will visit $i$ repeatedly and with probability 1 will eventually visit $j$ since $v_{ij} \neq 0$. This implies $v_{ii} \neq 0$ since with non-zero probability the MC can visit $j$ after $i$ but not return to $i$ (w.p. $1 - v_{ji} > 0$) **qed.**

Thus, in an irreducible MC, either all states are transient (a transient MC) or recurrent (a recurrent MC).

### 2.10.5 Further properties of MCs

Let $X$ be an irreducible, recurrent MC and let $n_1^j, n_2^j, \ldots$ be the times of successive visits to state $j$, $m_j = E[n_{k+1}^j - n_k^j]$ $(k = 1, 2, \ldots)$, the mean interval between visits (independent of $k$ by the MP).

Intuition:

$$p_j = \frac{1}{m_j}$$

Because the proportion of time spent, on average, in state $j$ is $1/m_j$. This is not necessarily true, because the chain may exhibit some *periodic* behaviour:

The state $j$ is *periodic* with period $m > 1$ if $q_{ii}^{(k)} = 0$ for $k \neq rm$ for any $r \geq 1$ and $P(X_{n+rm} = j \text{ for some } r \geq 1 | X_n = j) = 1$. Otherwise it is aperiodic, or has period 1. (Note that a periodic state is recurrent)

Proposition: In an irreducible MC, either all states are aperiodic or periodic with the same period. The MC is then called aperiodic or periodic respectively.

Proposition: If $\{X_n | n = 0, 1, \dots\}$ is an irreducible, aperiodic MC, then the limiting probabilities $\{p_j | j = 0, 1, \dots\}$ exist and $p_j = 1/m_j$.

### 2.10.6 Null & positive recurrence

If $m_j = \infty$ for state $j$ then $p_j = 0$ and state $j$ is *recurrent null*. Conversely, if $m_j < \infty$, then $p_j > 0$ and state $j$ is *recurrent non-null* or *positive recurrent*.

Proposition: In an irreducible MC, either all states are positive recurrent or none are. In the former case, the MC is called positive recurrent (or recurrent non-null).

A state which is aperiodic and positive recurrent is often called *ergodic* and an irreducible MC whose states are all ergodic is also called ergodic. When the limiting probabilities $\{p_j | j = 0, 1, \dots\}$ do exist they form the

$$\left.\begin{array}{l}\text{steady state}\\[1ex] \text{equilibrium}\\[1ex] \text{long term}\end{array}\right\}\ \text{probability distribution (SSPD)}$$

of the MC. They are determined by the following theorem.

### 2.10.7 Steady state theorem for Markov chains

An irreducible, aperiodic Markov Chain, $X$, with state space $S$ and one-step transition probability matrix $Q = (q_{ij}|i, j \in S)$, is positive recurrent if and only if the system of equations

$$p_j = \sum_{i \in S} p_i q_{ij}$$

with $\sum_{i \in S} p_i = 1$ (normalisation) has a solution. If it exists, the solution is unique and is the SSPD of $X$.

Note: If $S$ is finite, $X$ is always positive recurrent, which implies the equations have a unique solution. The solution is then found by discarding a balance equation & replacing it with the normalising equation (the balance equations are dependent since the rows of homogeneous equations $\underline{p} - \underline{p}.\underline{Q} = \underline{0}$ all sum to zero).

## 2.11 Markov Processes

- Continuous time parameter space, discrete state space

$$X = \{X_t | t \geq 0\} \quad X_t \in \Omega$$

  where $\Omega$ is a countable set.

- Markov property

$$P(X_{t+s} = j | X_u, u \leq t) = P(X_{t+s} = j | X_t)$$

- Markov process is *time homogeneous* if the r.h.s. of this equation does not depend on $t$
  - $q_{ij}(s) = P(X_{t+s} = j | X_t = i) = P(X_s = j | X_0 = i)$ $(i, j = 0, 1, \dots)$
  - *Transition probability functions* of the MP

- Markov Property and time homogeneity imply:

If at time $t$ the process is in state $j$, the time remaining in state $j$ is independent of the time already spent in state $j$: *memoryless property*

Proof:

$$\begin{aligned}
P(S > t + s | S > t) &= P(X_{t+u} = j, 0 \leq u \leq s | X_u = j, 0 \leq u \leq t) \\
&\quad \text{where S} = \text{time spent in state } j \\
&\quad \text{state } j \text{ entered at time } 0 \\
&= P(X_{t+u} = j, 0 \leq u \leq s | X_t = j) \text{ by MP} \\
&= P(X_u = j, 0 \leq u \leq s | X_0 = j) \text{ by T.H.} \\
&= P(S > s) \\
\Longrightarrow P(S \leq t + s | S > t) &= P(S \leq s)
\end{aligned}$$

- Time spent in state $i$ is exponentially distributed, parameter $\mu_i$.

### 2.11.1 Time homogeneous Markov Processes

The MP implies next state, $j$, after current state $i$ depends only on $i, j$ – state transition probability $q_{ij}$. This implies the Markov Process is uniquely determined by the products:

$$a_{ij} = \mu_i q_{ij}$$

where $\mu_i$ is the rate out of state $i$ and $q_{ij}$ is the probability of selecting state $j$ next. The $a_{ij}$ are the *generators* of the Markov Process.

- intuitively reasonable

### 2.11.2 Instantaneous transition rates

Consider now the (small) interval $(t, t+h)$

$$\begin{aligned} P(X_{t+h} = j | X_t = i) &= \mu_i h q_{ij} + o(h) \\ &= a_{ij} h + o(h) \end{aligned}$$

- $a_{ij}$ is the *instantaneous transition rate* $i \to j$ i.e. the average number of transitions $i \to j$ per unit time spent in state $i$

### 2.11.3 Examples

1. <u>Poisson process rate $\lambda$</u>

$$a_{ij} = \begin{cases} \lambda \text{ if } j = i+1 \\ 0 \text{ if } j \neq i, i+1 \\ \text{not defined if } j = i \end{cases}$$

because

$$\begin{array}{rcl} P(\text{arrival in } (t, t+h)) & = & \lambda h + o(h) \\ i \to i+1 & & a_{i,i+1} h \end{array}$$

2. <u>The buffer problem</u>
   - Records arrive as a P.P. rate $\lambda$
   - Buffer capacity is $M$
   - Buffer cleared at times spaced by intervals which are

exponentially distributed, parameter $\mu$. Clearance times i.i.d. & independent of arrivals (i.e. clearances are an independent PP, rate $\mu$)

$$a_{ij} = \begin{cases} \lambda \text{ if } j = i+1 \quad (0 \le i \le M-1) \\ \mu \text{ if } j = 0 \quad (i \neq 0) \\ 0 \text{ otherwise } (j \neq i) \end{cases}$$

[Probability of $> 1$ arrivals or clearances, or $\geq 1$ arrivals <u>and</u> clearances in $(t, t+h)$ is $o(h)$]

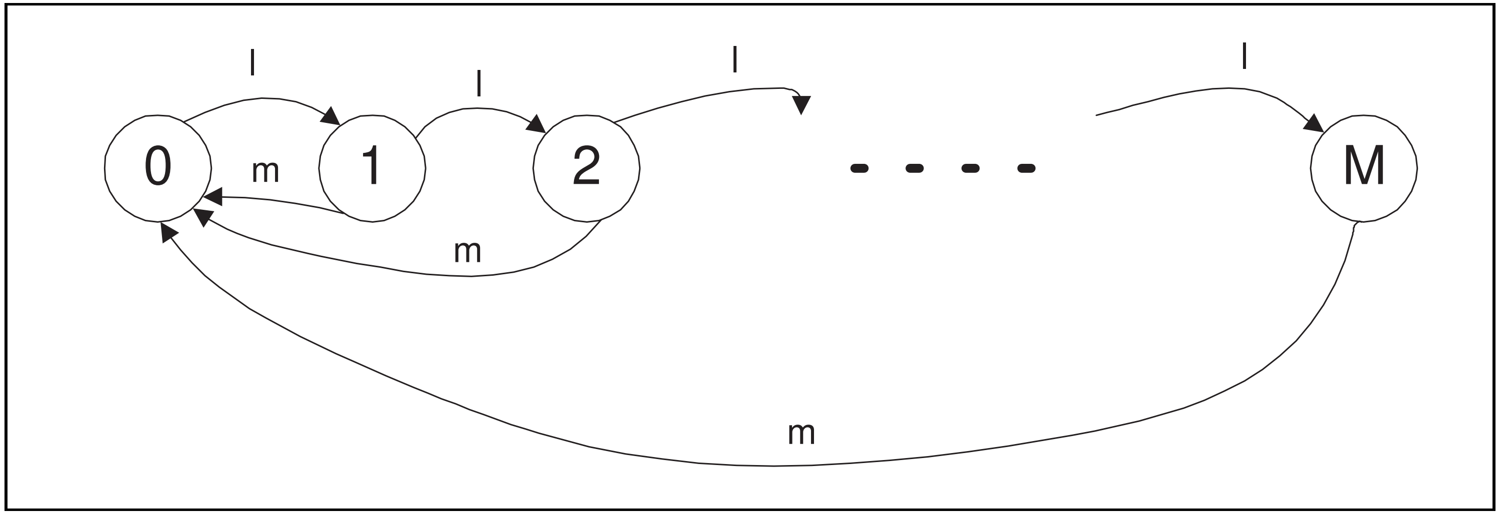


Figure 4: Buffer problem

3. Telephone exchange
   Suppose there 4 subscribers and any two calls between different callers can be supported simultaneously. Calls are made by non-connected subscribers according to independent P.Ps, rate $\lambda$ (callee chosen at random). Length of a call is (independent) exponentially distributed, parameter $\mu$. Calls are lost if called

subscriber is engaged. State is the number of calls in progress $\in \{0, 1, 2\}$

$$\begin{aligned}
a_{01} &= 4\lambda \quad \text{(all free)} \\
a_{12} &= \frac{2}{3}\lambda \quad \text{(caller has } \frac{1}{3} \text{ probability of successful connection)} \\
a_{10} &= \mu \\
a_{21} &= 2\mu \quad \text{(either call may end)}
\end{aligned}$$

<u>Exercise</u>: The last equation follows because the distribution of the random variable $Z = \min(X, Y)$ where $X, Y$ are independent exponential random variables with parameters $\lambda, \mu$ respectively is exponential with parameter $\lambda + \mu$. Prove this.

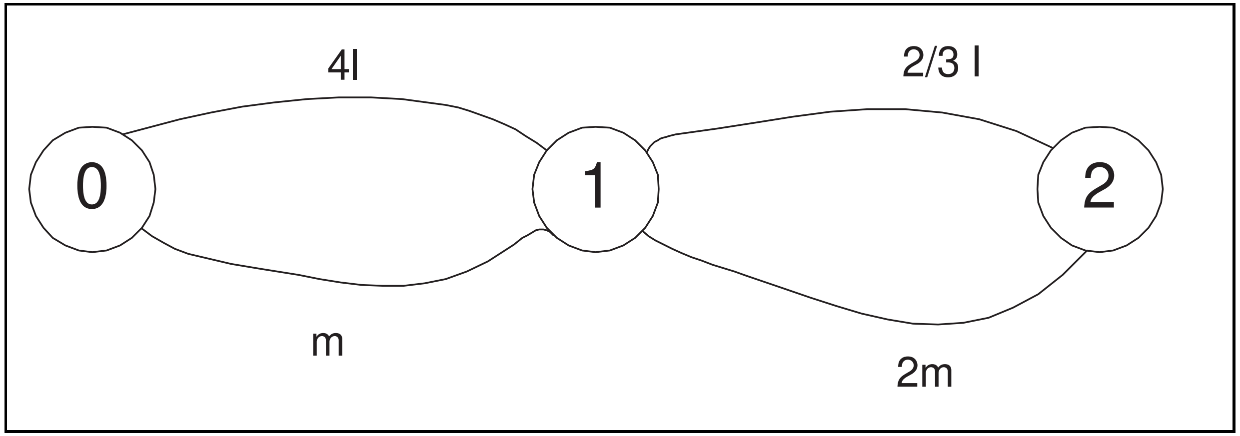


Figure 5: Telephone exchange

### 2.11.4 The Generator Matrix

$$\sum_{j\in\Omega, j\neq i} q_{ij} = 1 \implies \mu_i = \sum_{j\in\Omega, j\neq i} a_{ij}$$

(hence *instantaneous transition rate out of* $i$). Let $a_{ii} = -\mu_i$ (undefined so far)

$$(a_{ij}) = A = \begin{pmatrix} a_{00} & a_{01} & \dots & & \\ a_{10} & a_{11} & \dots & & \\ \dots & & & & \\ a_{i0} & a_{i1} & \dots & a_{ii} & \dots \\ \dots & & & & \end{pmatrix}$$

$A$ is the generator matrix of the Markov Process. Rows of $A$ sum to zero ($\sum_{j\in S} a_{ij} = 0$)

### 2.11.5 Transition probabilities & rates

- determined by $A$

$$
\begin{aligned}
q_{ij} &= \frac{a_{ij}}{\mu_i} = -\frac{a_{ij}}{a_{ii}} \\
\mu_i &= -a_{ii}
\end{aligned}
$$

- hence $A$ is all we need to determine the Markov process
- Markov processes are instances of state transition systems with the following properties
  - The state holding time are *exponentially* distributed
  - The probability of transiting from state $i$ to state $j$ depends only on $i$ and $j$
- In order for the analysis to work we require that the process in **irreducible** – every state must be reachable from every other, e.g.

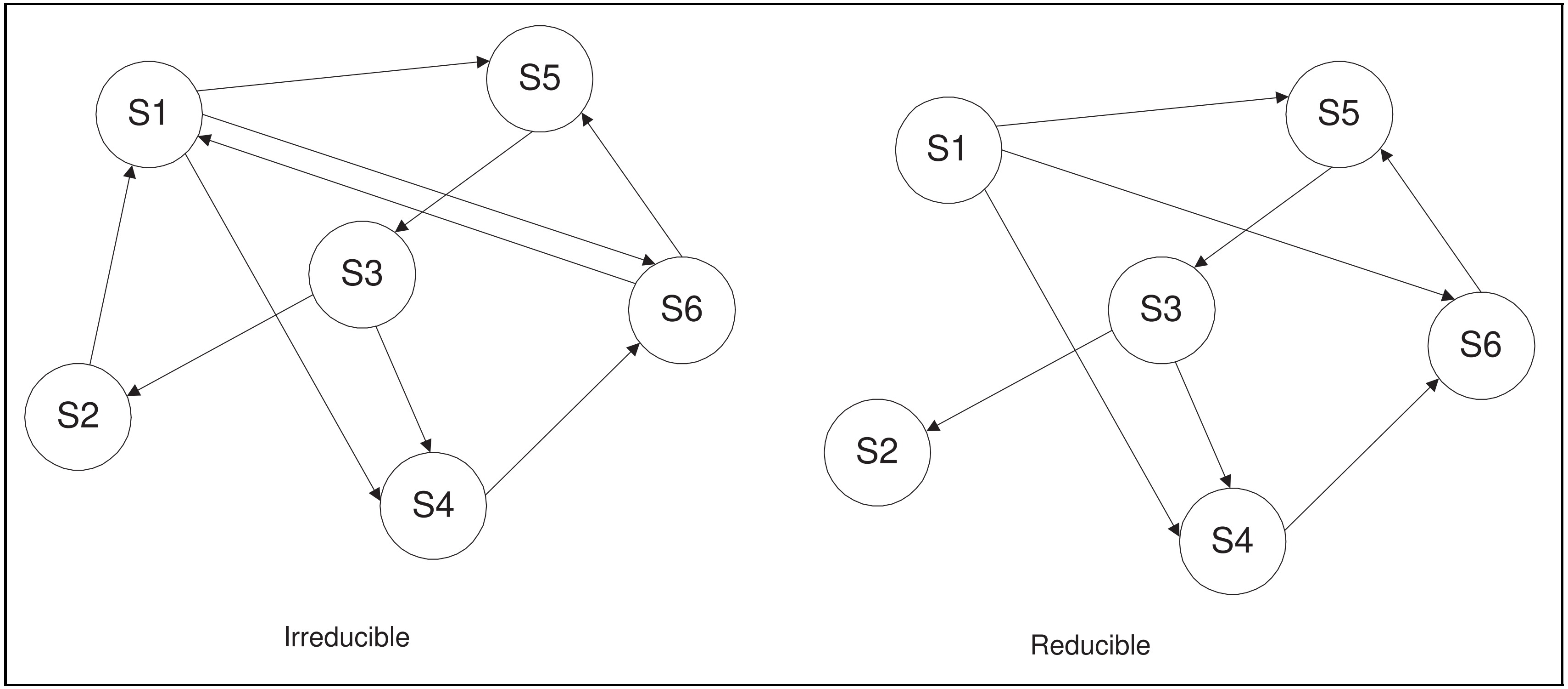


Figure 6: (Ir)reducible Markov chains

- Associated with each arc in a state transition diagram is a *rate*, e.g.:

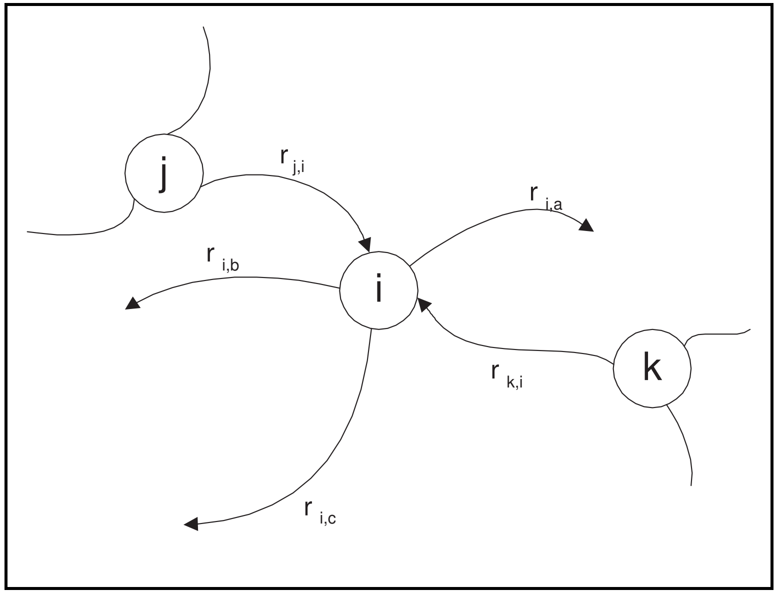


Figure 7: Transition rates in Markov chains

- The state holding time depends on the departure rates $(1/(r_{i,a} + r_{i,b} + r_{i,c})$ is the mean holding time for state $i$ in the example).

- The *probability flux* from a state $i$ to a state $j$ is the average number of transitions from $i \to j$ per unit time at equilibrium:

$$\text{Probability flux}(i \to j) = r_{i,j} p_i$$

- At equilibrium (if it exists) the total flux into each state must balance the total flux out of it
- For example, in the example, for state $i$ alone:

$$(r_{i,a} + r_{i,b} + r_{i,c}) p_i = r_{j,i} p_j + r_{k,i} p_k$$

- Treating all states the same way leads to a set of *linear* equations:

$$\sum_{j \neq i} a_{ij} p_i = \sum_{j \neq i} a_{ji} p_j = -a_{ii} p_i \text{ for all } j \in S$$

- Called the *balance equations* and subject to the *normalising equation*

$$\sum_{i \in S} p_i = 1$$

- The **steady-state theorem** tells us that if these equations have a solution then there is a steady-state and that the solution is unique
- **Important generalisation**: At equilibrium the fluxes also balance into and out of *every closed contour* drawn around *any* collection of states
- This often provides a short-cut to the solution – see the M/M/1 queue later.

## 2.12 Steady state results

Theorem

(a) if a Markov process is transient or recurrent null, $p_j = 0 \quad \forall j \in S$ and we say a SSPD does not exist

(b) If a Markov Process is positive recurrent, the limits $p_j$ exist, $p_j > 0, \sum_{j \in S} p_j = 1$ and we say $\{p_j | j \in S\}$ constitute the SSPD of the Markov Process.

### 2.12.1 Steady State Theorem for Markov Processes

An irreducible Markov Process $X$ with state space $S$ and generator matrix $A = (a_{ij})$ $(i, j \in S)$ is positive recurrent if and only if

$$\begin{aligned} \sum_{i \in S} p_i a_{ij} &= 0 \quad \text{for } j \in S \text{ [ Balance equations]} \\ \sum_{i \in S} p_i &= 1 \quad \text{[Normalising equation]} \end{aligned}$$

have a solution. This solution is unique and is the SSPD.

### 2.12.2 Justification of the Balance Equation

The rate going from state $i$ to $j(\neq i)$ is $a_{ij}$, the fraction of time spent in $i$ is $p_i$

$$\begin{aligned} \sum_{i\neq j} a_{ij}p_i &= \sum_{i\neq j} a_{ji}p_j \\ \text{Avg. no. of transitions} & \\ i \to j \text{ in unit time} & \qquad \text{Avg. no. of transitions} \\ & \qquad j \to i \text{ in unit time} \\ \sum_{i\neq j} \text{flux}(i \to j) &= \sum_{i\neq j} \text{flux}(j \to i) \quad \forall j \end{aligned}$$

### 2.12.3 Buffer example (DTMC)

$\underline{1 \leq i \leq M-1}$

$$
\begin{aligned}
(\text{Net "rate" in}) &= (\text{Net "rate" out}) \\
(\alpha(1-\beta)+\beta)p_i &= \alpha(1-\beta)p_{i-1}
\end{aligned}
$$

$$\{= p_i = \alpha(1-\beta)p_{i-1} + p_i(1-\alpha)(1-\beta) \text{ from S.S. Theorem}\}$$

Therefore

$$p_i = kp_{i-1} = k^i p_0$$

where $k = \frac{\alpha(1-\beta)}{\alpha+\beta-\alpha\beta}$ $(1 \leq i \leq M-1)$

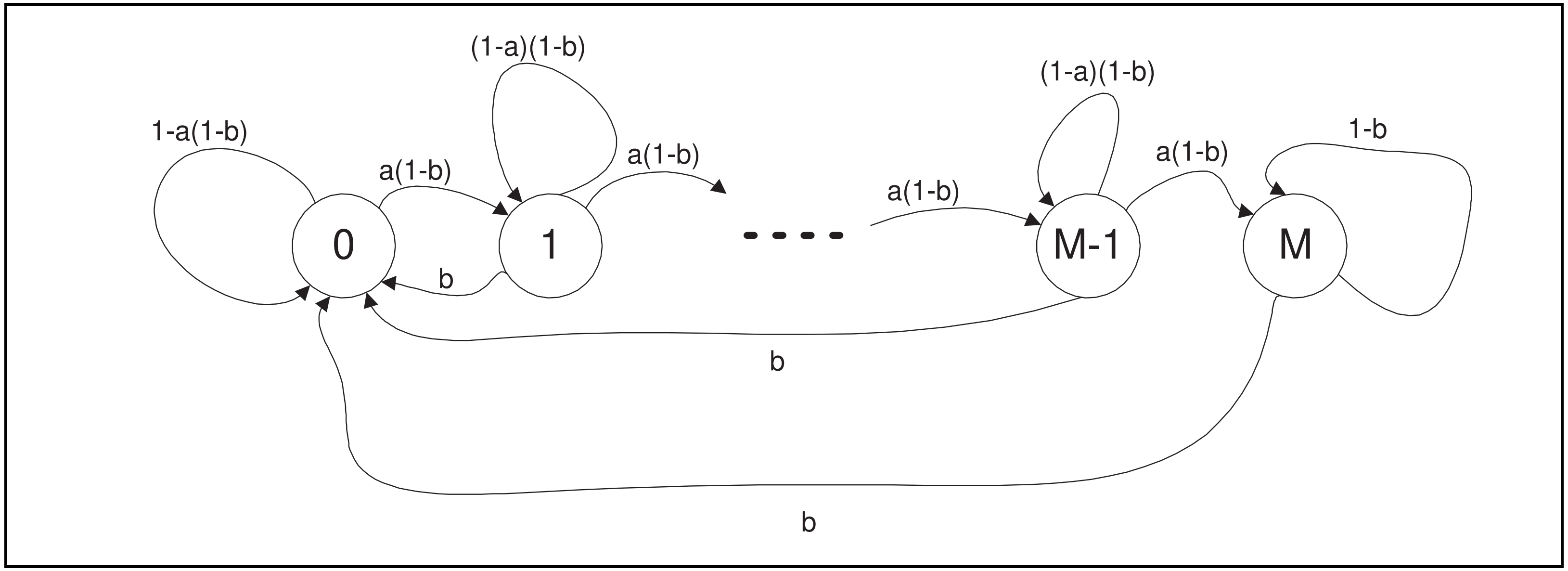


Figure 8: I/O buffer.

$\underline{i = M}$

$$\beta p_M = \alpha(1-\beta)p_{M-1}$$

therefore

$$p_M = \frac{\alpha(1-\beta)}{\beta} k^{M-1} p_0$$

$\underline{i = 0}$

Redundant equation (why?) - use as check

$$\begin{aligned}
\alpha(1-\beta)p_0 &= \sum_{i=1}^{M} \beta p_i = \beta \sum_{i=1}^{M-1} k^i p_0 + k^{M-1}\alpha(1-\beta)p_0 \\
&= p_0\Big(\beta k \frac{1-k^{M-1}}{1-k} + \alpha(1-\beta)k^{M-1}\Big)
\end{aligned}$$

rest is exercise

### 2.12.4 Examples (CTMC)

1. I/O buffer:
   By considering the flux into and out of each state $0, 1, \ldots, M$ we obtain the balance equations:

$$\begin{aligned} \lambda p_0 &= \mu(p_1 + \cdots + p_M) \text{ (State 0)} \\ (\lambda + \mu)p_i &= \lambda p_{i-1} \quad (1 \le i \le M-1) \text{ State } i \\ \mu p_M &= \lambda p_{M-1} \text{ State } M \end{aligned}$$

   Normalising equation:

$$\begin{aligned} p_0 &+ \cdots + p_M = 1 \\ \Longrightarrow p_j &= \left(\frac{\lambda}{\lambda+\mu}\right)^j \frac{\mu}{\lambda+\mu} \quad (0 \le j \le M-1) \\ p_M &= \left(\frac{\lambda}{\lambda+\mu}\right)^M \end{aligned}$$

Thus (for example) mean number of records in the buffer in the steady state $= M\alpha^M + \sum_{j=0}^{M-1} j\alpha^j \frac{\mu}{\lambda+\mu}$ where $\alpha = \frac{\mu}{\lambda+\mu}$

2. <u>Telephone network</u>

$$\begin{aligned} 4\lambda p_0 &= \mu p_1 \text{ (State 0)} \\ \left(\mu + \frac{2}{3}\lambda\right)p_1 &= 4\lambda p_0 + 2\mu p_2 \text{ (State 1)} \\ 2\mu p_2 &= \lambda\frac{2}{3}p_1 \text{ (State 2)} \end{aligned}$$

Thus $p_1 = \frac{4\lambda}{\mu}p_0$, $p_2 = \frac{\lambda}{3\mu}p_1 = \frac{4\lambda^2}{3\mu^2}p_0$ with $p_0 + p_1 + p_2 = 1 \implies p_0 = \left(1 + \frac{4\lambda}{\mu} + \frac{4\lambda^2}{3\mu^2}\right)^{-1}$ Average number of calls in progress in the steady state $= 1.p_1 + 2.p_2 = \frac{\frac{4\lambda}{\mu}\left(1+\frac{2\lambda}{3\mu}\right)}{1+\frac{4\lambda}{\mu}+\frac{4\lambda^2}{3\mu^2}}$