Quality of Service, Performance and Introduction to Models

Lecture 2

Communication Theory III

Eng. (Mrs.) PN Karunanayake

International organizations

In the sector of telecommunications, several international organisations aim at publishing commonly agreed level of quality that a network should provide.

-ITU (International Telecommunication Union)

covers all domains of communications (networks, services, image, etc.). It publishes its work as recommendations.

- ETSI (European Telecommunications Standards Institute)

-IETF (Internet Engineering Task Force)

- *Organisation for the IP world, accepting any contribution aimed at helping the development of the internet.*

Quality of Service (QoS)

- The ITU Recommendation E.800 gives a definition of quality of service as
- "the collective effect of service performance which determines the degree of satisfaction of a user of the service".

Network performance

- Network performance is measured in terms of parameters, significant for network and service providers.
- Basis for **system design, for system dimensioning and network provisioning, maintenance**, etc., in order to satisfy both the customer and the provider.

Traffic load conditions

1. Normal load (load A) - Normal day conditions which correspond to the usual daily traffic conditions of the network.

2. High load (load B) -   Less frequent higher busy operating network conditions.

3. Over load - Exceptional conditions which correspond to load conditions far beyond the normal provisioning, and associated with totally unpredictable events.

Normal load, load A

- Corresponds to the busy operating conditions (most frequent) of the network for which normal user service expectation should be met.
- This will be the reference normal load for dimensioning.

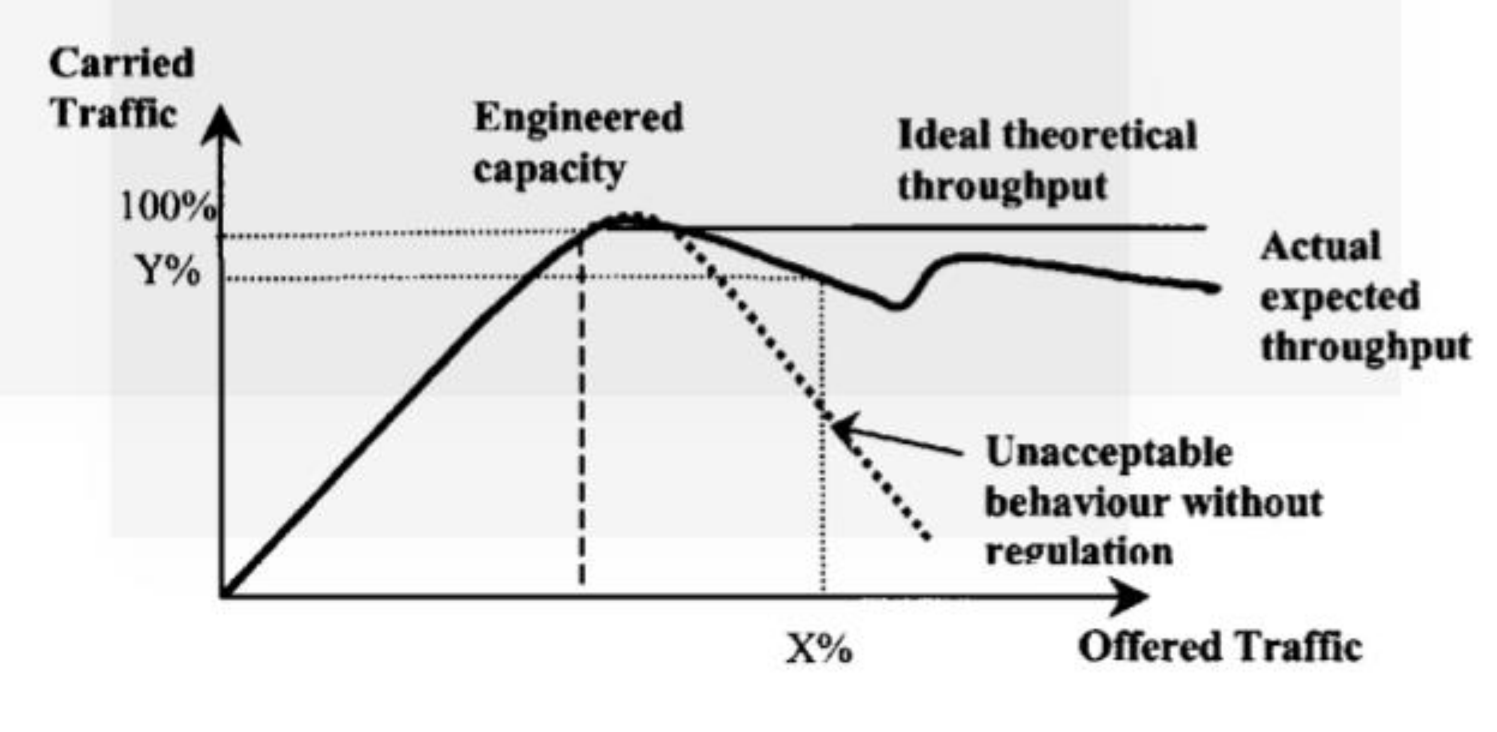
High load, load B

- Not very frequently encountered operating conditions.
- User service expectation would not necessarily be met.
- The level of performance is high enough to avoid significant user dissatisfaction.

Overload

- Adresses situations that can be said to be uncommon or exceptional.
- Offered traffic is greatly larger than the available network resources.
- To cater this kind of a situation, large investment is required due to the unavoidable overdimensioning

Traffic carried in overload situation



Synthesis tables - Trafficability

Parameter	Normal load		High load	
	Mean	95%	Mean	95%
Set-up delay for a national call (E.721)	5000ms	8000 ms	7500 ms	12 000 ms
Set up delay at the access for an outgoing call Q.543)	600 ms	800 ms	800 ms	1200 ms
Incoming call indication sending delay (Q.543)	650 ms	900 ms	1000 ms	1600 ms
Clearing delay for a national call (I.352)	1250ms	1750 ms	Not specified	Not specified
Release delay at the access Q.543)	250 ms	300 ms	400 ms	700 ms
Blocking probability: – national call – international call (E.721)	3% 5%	Not applicable	4,5% 7.5%	Not applicable
Blocking probability at the access: – originating call – terminating call (Q.543)	0.5% 0.5%	Not applicable	3% 3%	Not applicable
Transfer delay at a node for an new call demand signalling message (Q.766)	180 ms	360 ms	450 ms	900 ms
Transfer delay at a node for a voice sample (Q.551)	0.9 ms	1.5 ms	Not specified	Not specified
End-to-end transfer delay for an IP packet: – real time – file (Y.1541)	100 ms 1000ms	Not specified	Not specified	Not specified
Transfer delay variation for an IP packet (10^{-3} quantile): – real time (Y.1541)	50 ms	Not specified	Not specified	Not specified
End-to-end IP packet loss probability (Y.1541)	10^{-3}	Not applicable	Not specified	Not applicable
Handover failure probability (E.771)	0.5%	Not specified	Not specified	Not specified

Synthesis tables - Dependability parameters

Mean network unavailability (E.846)	$6\ 10^{-2}$
Unavailability at the access nod for: – a user – a large group of users (Q541 and network providers)	 30 mn/year 3 mn/year
Premature release probability for an international communication (E850): – typical configuration – worst case	 $4\ 10^{-4}$ $1.6\ 10^{-3}$
Premature release probability at a node (Q543)	$2\ 10^{-5}$
Maintenance load at an access node (network providers)	< 15 failure per year and per 10 000 users
Logistic delays (network providers): – immediate – postponed – without urgency	 3.5 hours 12 hours 72 hours

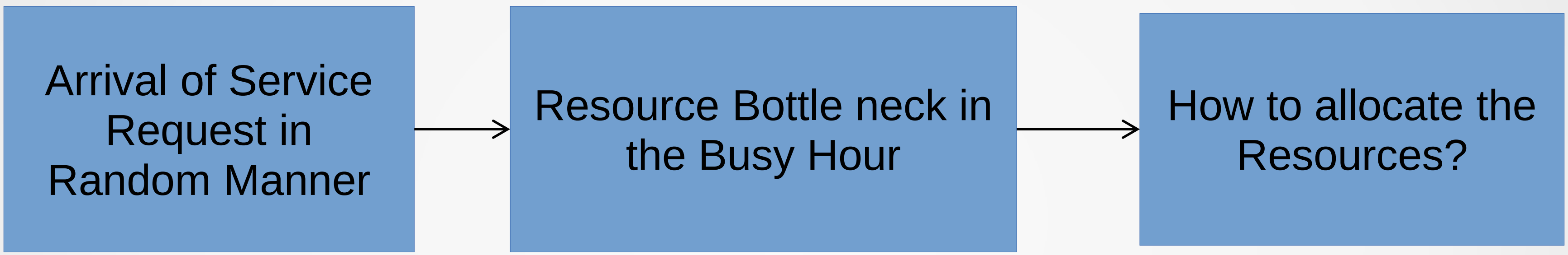
For the access phase, the dependability performance parameters concern the probability that the user is able to access the network (availability of the equipments)

Why an accurate estimation of network performance is vital?

- Service is essential for maintaining customer satisfaction.
- Cost is always a factor in maintaining profitability.

- To Develop effective congestion control techniques.
- To model the performance to decide the type of congestion control policies to be implemented.
- To avoid under-estimation or over-estimation of the performance of the network.

Network Models



1. Loss System
2. Queuing System

Both involves an additional Delay

Point Process

- A very special class of stochastic processes called *point* processes that possess two properties,

- *Orderliness*
- *Memorylessness*

point process is a sequence of events which is called *arrivals* occurring at random in points of time t_i, $i = 1; 2; : : :$, $t_{i+1} > t_i$

- To the context of queueing theory, where a point process typically corresponds to points of arrivals, i.e., t_i is the time of the i^{th} arrival that joints a queue.
- A point process can be defined by its counting process $\{N(t); t \geq 0\}$, where **N(t) is the number of arrivals occurred within [0; t).**

Point Process

- Properties of a counting process $\{N(t)\}$:

1. $N(t) \geq 0$,
2. $N(t)$ is integer,
3. if $s > t$, then $N(s) \geq N(t)$ and $N(s) - N(t)$ is the number of occurrences within $(t; s]$.

Note that $N(t)$ is not an independent process because, for example, if $t2 > t1$, then $N(t2)$ is dependent on the number of arrivals in $[0; t1)$, namely, $N(t1)$.

Orderliness
The probability that two or more arrivals happen at once is negligible.

Memoryless
At any point in time, the future evolution of the process is statistically independent of its past.

Bernoulli Process

- The Bernoulli process is a discrete-time stochastic process made up of a sequence of IID Bernoulli distributed random variables

$\{X_i; i = 0; 1; 2; 3; \ldots\}$ where for all i, $P(X_i = 1) = p$ and $P(X \; = 0) = 1\text{-}p$.

- The Bernoulli process is both orderly and memoryless.
- The counting process for the Bernoulli process is another discrete-time stochastic process,
  $\{N(n); n \geq 0\}$

which is a sequence of Binomial random variables $N(n)$ representing the total number of arrivals occurring within the first n time slots.

$$P[N(n) = i] = \binom{n}{i} p^i (1-p)^{n-i} \qquad i = 0,\ 1,\ 2,\ \ldots,\ n.$$

Bernoulli Process

- The random variable Δ represent any inter-arrival time

$$P(\Delta = i) = p(1-p)^{i-1} \qquad i = 1,\ 2,\ \ldots\ .$$

- The time it takes *n* until the *i*th arrival

$$P[\text{the } i\text{th arrival occurs in time slot } n] = \binom{n-1}{i-1} p^i (1-p)^{n-i} \qquad i = i,\ i+1,\ i+2,\ \ldots\ .$$

Poisson Process

- Most widely used and oldest traffic model in traditional telephony networks.
- A **memoryless & orderly model**.
- In a Poisson process the inter-arrival times are **exponentially distributed with a rate parameter λ**.
- The Poisson distribution is appropriate if the arrivals are from a large number of independent sources, referred to as Poisson sources.
- Conuting Process (the number of arrivals in the interval [0; t]):

$$\{N(t),\ t \geq 0\}$$

Poisson Process

A counting process *{N(t)}* is defined as *a Poisson process* with rate $\lambda > 0$ if it satisfies the following three conditions.

***1. N*(0) = 0.**

2. The numbers of arrivals in two non-overlapping intervals are independent.
That is, for
any $s > t > u > v > 0$, the random variable $N(s) - N(t)$ (representing the number of arrivals between time s and time t), and the random variable $N(u) - N(v)$ are independent. **This means that the Poisson process has what is called *independent increments*.**

3. **The number of arrivals** in an interval of length t, for $t > 0$, has a Poisson distribution with **mean λt,**

Properties of Poisson Process

1. *Time-homogeneity*
   Every point in time has the same chance of having an arrival

2. Inter-arrival times is exponentially distributed with parameter λ

$$E[T] = \frac{1}{\lambda}.$$

Assumptions of Poisson Model

1. The number of sources is infinite
2. The traffic arrival pattern is random.

- The cumilative distribution function given as

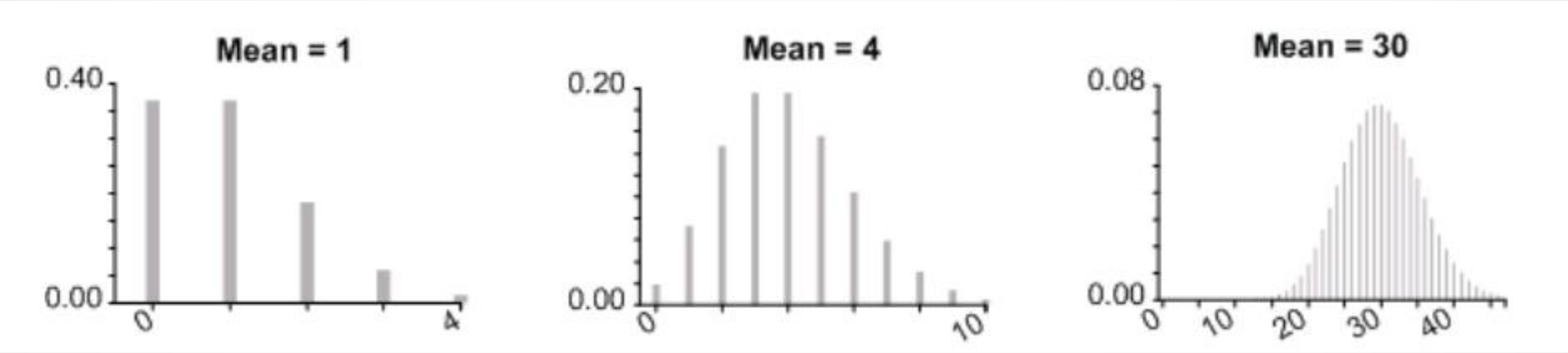
$$F(t) = 1 - e^{-\lambda t}$$

- Probability density function of the model given as:

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

Poisson Distribution

- As the mean increases, the properties of the Poisson distribution approach those of the normal distribution.



Bernoulli processes are the discrete time analog of Poisson processes.

Markov and Embedded Markov Models

- Markov models attempt to model the activities of a traffic source on a network, by a finite number of states.
- The accuracy and the complexity of the model increases linearly with the number of states used in the model.
- The next (future) state depends only on the current state.
- In a simple Markov Traffic model, each of the state transition represents a new arrival process on the network.

Discrete-time Markov chains

- Markov chains are certain discrete space stochastic processes.
- Popular for analyzing traffic characterization, and modeling of queueing and telecommunications networks and systems.
- Can be classified into two groups:
    1. Discrete time Markov chains
    2. Continuous-time Markov chains

Discrete-time Markov chains

- A discrete-time Markov chain is a discrete-time stochastic process $\{X_n; n = 0; 1; 2; \ldots\}$ **with the Markov property**;

that at any point in time *n*, **the future evolution of the process is dependent only on the state of the process at time *n*, and is independent of the past evolution of the process**.

- The state of the process can be a scalar or a vector

Discrete-time Markov chains

- The discrete-time Markov chain $\{X_n; n = 0; 1; 2, \ldots\}$ at any point in time may take many possible values.
- The set of these possible values is finite or countable and it is called the state space of the Markov chain, denoted by Θ.

$$P(X_{n+1} = i \mid X_n = j) = P(X_n = i \mid X_{n-1} = j) \quad \text{for all } n.$$

Transition Probability Matrix

A Markov chain is characterized by the *Transition Probability Matrix* **P** which is a matrix of one-step transition probabilities P_{ij} defined by :

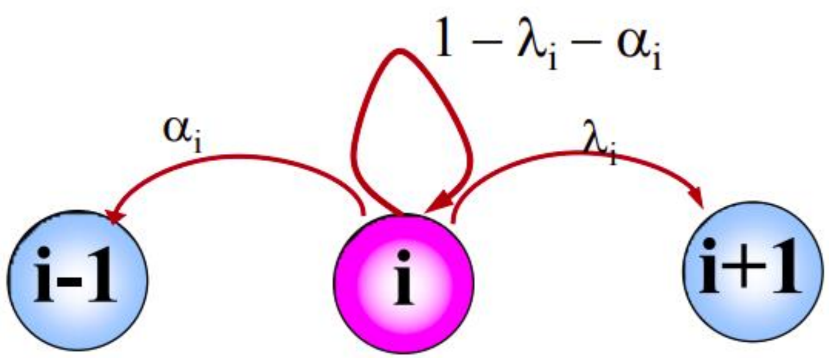
$$P_{ij} = P(X_{n+1} = j \mid X_n = i) \quad \text{for all } n.$$

Birth-Death Processes

Size of population

- System is in state E_k when consists of k members
- Changes in population size occur by at most one
- Size increased by one → "Birth"
- Size decreased by one → "Death"
- Transition probabilities p_{ij} do not change with time

$$p_{ij} = \begin{cases} \alpha_i & j = i - 1 \\ 1 - \lambda_i - \alpha_i & j = i \\ \lambda_i & j = i + 1 \\ 0 & \text{Otherwise} \end{cases}$$



Birth-Death Processes

- α_i = death (less one in population size)
- $\alpha_0 = 0$ (no population → no death)
- λ_i = birth (increase one in population)
- $\lambda_i > 0$ (birth is allowed)
- Pure Birth = no decrement, only increment
- Pure Death = no increment, only decrement

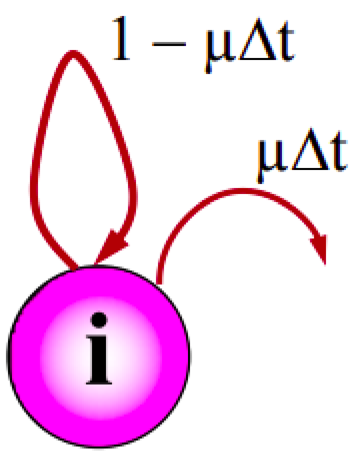
For Queueing Theory Mode

- Population = customers in the queueing system
- Death = a customer departure from the system
- Birth = a customer arrival to the system

Transition matrix

$$
P = \begin{bmatrix}
1-\lambda_0 & \lambda_0 & 0 & 0 & 0 & 0 & \cdots \\
\alpha_1 & 1-\lambda_1-\alpha_1 & \lambda_1 & 0 & 0 & 0 & \\
0 & \alpha_2 & 1-\lambda_2-\alpha_2 & \lambda_2 & & & \\
0 & & \cdots & & & & \\
0 & & & & & & \\
 & & & \alpha_i & 1-\lambda_i-\alpha_i & \lambda_i & \\
\cdots & & & & & & \\
\end{bmatrix}
$$

Markov Process Property



For Continuous Time Markov Chain

- P[system in state i for time T | system in current state i]

$$= (1 - \mu\Delta t)^{T/\Delta t}$$

$$= e^{-\mu T} \qquad \text{where } \Delta t \rightarrow 0$$

- Exponentially distributed state times

Continuous Time Birth-Death Markov Chains

- Let λ_i = birth rate in state i

  μ_i = death rate in state i
- Then

  P[state i to state i – 1 in Δt] $= \mu_i \Delta t$

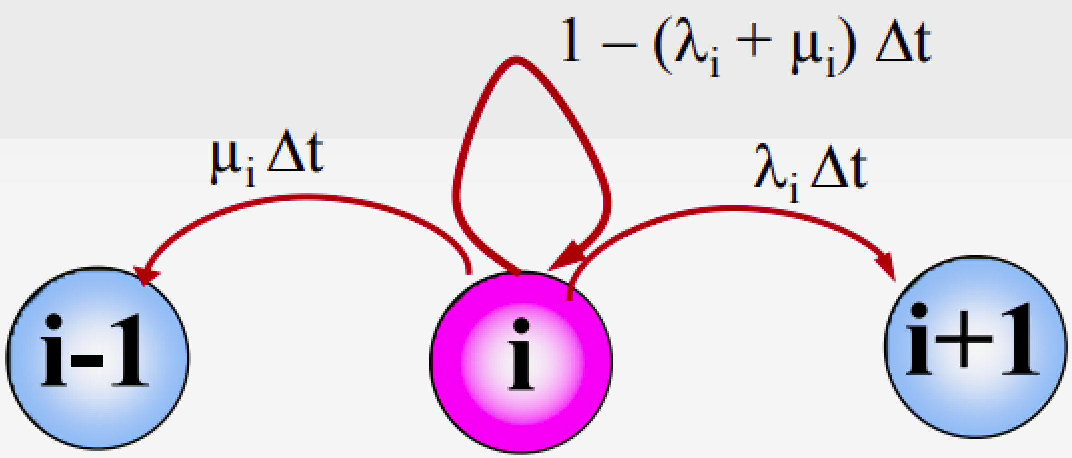
  P[state i to state i + 1 in Δt] $= \lambda_i \Delta t$

  P[state i to state i in Δt] $= 1 - (\lambda_i + \mu_i) \Delta t$

  P[state i to other state in Δt] $= 0$

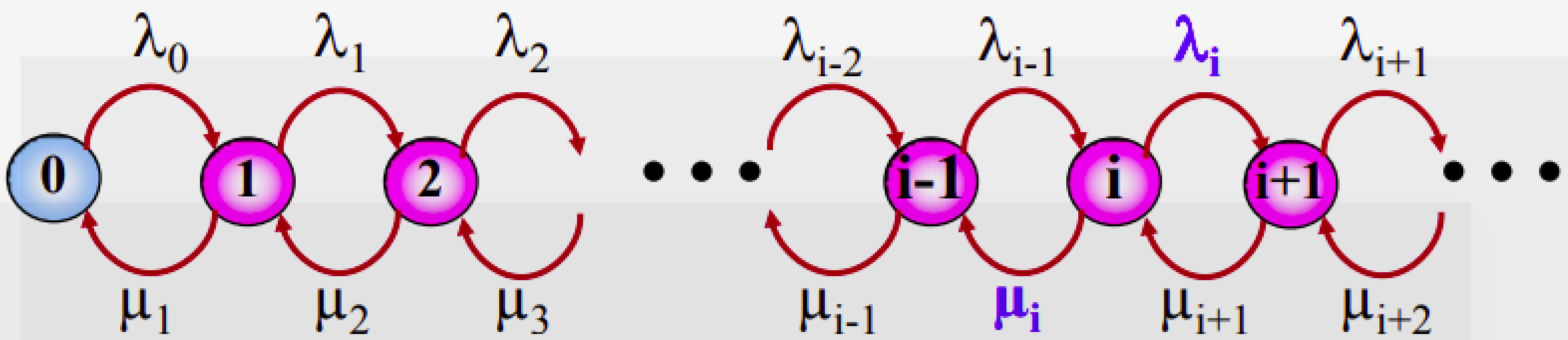
Continuous Time Birth-Death Markov Chains

$$p_{ij} = \begin{cases} \mu_i \,\Delta t & j = i - 1 \\ 1 - (\lambda_i + \mu_i)\,\Delta t & j = i \\ \lambda_i \,\Delta t & j = i + 1 \\ 0 & \text{Otherwise} \end{cases}$$



$\Delta t \rightarrow 0$

State Transition Diagram



- $X(t)$ = # in the system at time t

  = birth – death in $(0,t)$
- $p_i(t) = P[\, X(t) = i \,]$

  = Prob. that system is in state i at time t

- A weather model. Consider a homogeneous, discrete-time Markov chain that describes the daily weather pattern in Belfast, Northern Ireland .
- Simplify the situation by considering only three types of weather pattern: **rainy, cloudy, and sunny.**
- Markov chain: state 1 (R) represents a (mostly) rainy day; state 2 (C), a (mostly) cloudy day; and state 3 (S), a (mostly) sunny day.
- The weather is observed daily. On any given rainy day, the probability that it will rain the next day is estimated at 0.8; the probability that the next day will be cloudy is 0.15, while the probability that tomorrow will be sunny is only 0.05.

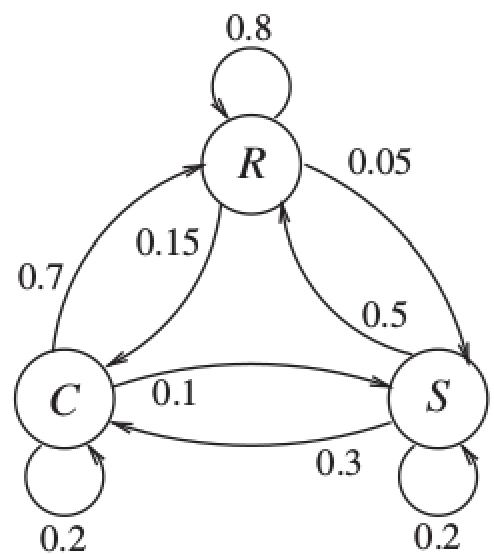


Figure 9.2. Transition diagram for weather at Belfast.

The transition probability matrix P for this Markov chain is

$$P = \begin{pmatrix} 0.8 & 0.15 & 0.05 \\ 0.7 & 0.2 & 0.1 \\ 0.5 & 0.3 & 0.2 \end{pmatrix}.$$

- Elements in the matrix represent conditional probabilities. For example, the element p_{32} tells us that the probability that tomorrow is cloudy, given that today is sunny, is 0.3.
- we may compute quantities such as the probability that tomorrow is cloudy and the day after is rainy, given that it is sunny today. In this case we have, for all n = 0, 1, 2, . . . ,

$$\text{Prob}\{X_{n+2} = R, X_{n+1} = C \mid X_n = S\} = p_{SC}\, p_{CR} = 0.3 \times 0.7 = 0.21.$$

This is the probability of the sample path $S \to C \to R$.