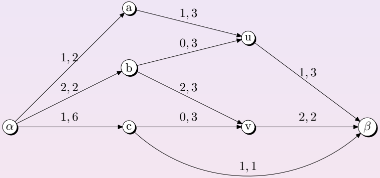
Modelling data networks – stochastic processes and Markov chains



Richard G. Clegg (richard@richardclegg.org)— November 2016
Available online at http://www.richardclegg.org/lectures accompanying printed notes provide full bibliography.
(Prepared using LaTeX and beamer.)

Introduction to stochastic processes and Markov chains

Stochastic processes

A stochastic process describes how a system behaves over time – an arrival process describes how things arrive to a system.

Markov chains

Markov chains describe the evolution of a system in time – in particular they are useful for queuing theory. (A Markov chain is a stochastic process).

Stochastic processes

Stochastic process

Let $X(t)$ be some value (or vector of values) which varies in time t. Think of the stochastic process as the rules for how $X(t)$ changes with t. Note: t may be discrete ($t = 0, 1, 2, \ldots$) or continuous.

Poisson process

A process where the change in $X(t)$ from time t_1 to t_2 is a Poisson distribution, that is $X(t_2) - X(t_1)$ follows a Poisson distribution.

A simple stochastic process – the drunkard's walk

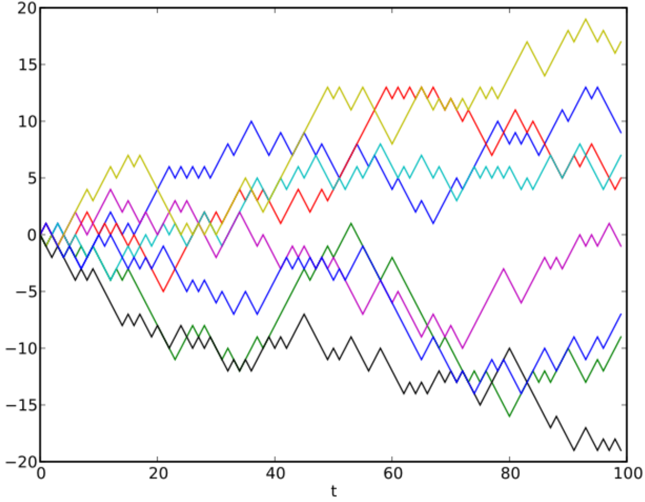
Random walk – or drunkard's walk

A student walks home from a party. They start at a distance $X(0)$ from some point. At every step they (randomly) gets either one unit closer (probability p) or one unit further away.

$$X(t+1) = \begin{cases} X(t)+1 & \text{probability } p \\ X(t)-1 & \text{probability } 1-p. \end{cases}$$

Can answer questions like "where will he be at time t"?

Drunkard's walk – $p = 0.5, X(0) = 0$



Drunkard's walk – $p = 0.5, X(0) = 0$

- What is the average (or expected value) of $X(t)$, that is, $\mathrm{E}[X(t)]$?
- Note – if you've never come across "expected value" or "expectation" before you can think of it as average (see notes).
- $\mathrm{E}[X(t)] = 0.5(X(t-1)+1) + 0.5(X(t-1)-1) = 0.5(X(t-1)+X(t-1)) + 0.5(1-1) = X(t-1)$.

Drunkard's walk – $p = 0.5, X(0) = 0$

- What is the average (or expected value) of $X(t)$, that is, $\mathrm{E}[X(t)]$?
- Note – if you've never come across "expected value" or "expectation" before you can think of it as average (see notes).
- $\mathrm{E}[X(t)] = 0.5(X(t-1)+1) + 0.5(X(t-1)-1) = 0.5(X(t-1)+X(t-1)) + 0.5(1-1) = X(t-1)$.
- Therefore $\mathrm{E}[X(t)] = X(0) = 0$ – the poor drunk makes no progress towards his house (on average).

Drunkard's walk – $p = 0.5, X(0) = 0$

- What is the average (or expected value) of $X(t)$, that is, $\mathrm{E}[X(t)]$?
- Note – if you've never come across "expected value" or "expectation" before you can think of it as average (see notes).
- $\mathrm{E}[X(t)] = 0.5(X(t-1)+1) + 0.5(X(t-1)-1) = 0.5(X(t-1)+X(t-1)) + 0.5(1-1) = X(t-1)$.
- Therefore $\mathrm{E}[X(t)] = X(0) = 0$ – the poor drunk makes no progress towards his house (on average).
- $\mathrm{E}\left[X(t)^2\right] = 0.5(X(t-1)+1)^2 + 0.5(X(t-1)-1)^2 = X(t-1)^2 + 1$.
- Therefore $\mathrm{E}\left[X(t)^2\right] = t$ – on average the drunk does get further from the starting pub.

Drunkard's walk – $p = 0.5, X(0) = 0$

- What is the average (or expected value) of $X(t)$, that is, $\mathrm{E}[X(t)]$?
- Note – if you've never come across "expected value" or "expectation" before you can think of it as average (see notes).
- $\mathrm{E}[X(t)] = 0.5(X(t-1)+1) + 0.5(X(t-1)-1) = 0.5(X(t-1)+X(t-1)) + 0.5(1-1) = X(t-1)$.
- Therefore $\mathrm{E}[X(t)] = X(0) = 0$ – the poor drunk makes no progress towards his house (on average).
- $\mathrm{E}\left[X(t)^2\right] = 0.5(X(t-1)+1)^2 + 0.5(X(t-1)-1)^2 = X(t-1)^2 + 1$.
- Therefore $\mathrm{E}\left[X(t)^2\right] = t$ – on average the drunk does get further from the starting pub.
- This silly example has many uses in physics and chemistry (Brownian motion) – not to mention gambling (coin tosses).

The Poisson process

The Poisson process

Let $X(t)$ with $(t \geq 0)$ and $X(0) = 0$ be a Poisson process with rate λ. Let t_2, t_1 be two times such that $t_2 > t_1$. Let $\tau = t_2 - t_1$.

$$\mathbb{P}\left[X(t_2) - X(t_1) = n\right] = \exp[-(\lambda\tau)] \left[\frac{(\lambda\tau)^n}{n!}\right],$$

for $n = 0, 1, 2, \ldots$.

In other words, the number of arrivals in some time period τ follows a Poisson distribution with rate $\lambda\tau$.

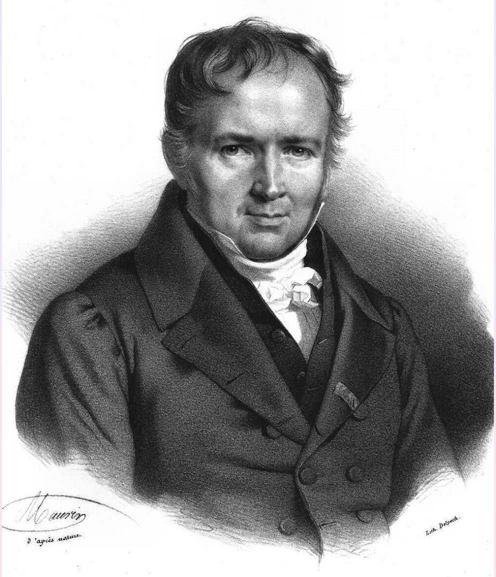
The special nature of the Poisson process

- The Poisson process is in many ways the simplest stochastic process of all.
- This is why the Poisson process is so commonly used.
- Imagine your system has the following properties:
  - The number of arrivals does not depend on the number of arrivals so far.
  - No two arrivals occur at exactly the same instant in time.
  - The number of arrivals in time period τ depends only on the length of τ.
- The Poisson process is the only process satisfying these conditions (see notes for proof).

Some remarkable things about Poisson processes

- The mean number of arrivals in a period τ is $\lambda\tau$ (see notes).
- If two Poisson processes arrive together with rates λ_1 and λ_2 the arrival process is a Poisson process with rate $\lambda_1 + \lambda_2$.
- In fact this is a general result for n Poisson processes.
- If you randomly "sample" a Poisson process – e.g. pick arrivals with probability p, the sampled process is Poisson, rate $p\lambda$.
- This makes Poisson processes easy to deal with.
- Many things in computer networks really are Poisson processes (e.g. people logging onto a computer or requesting web pages).
- The Poisson process is also "memoryless" as the next section explains.

Siméon Denis Poisson(1781-1840)



The interarrival time – the exponential distribution

The exponential distribution

An exponential distribution for a variable T takes this form:

$$\mathbb{P}[T \leq t] = \begin{cases} 1 - \exp[-(\lambda t)], & t \geq 0, \\ 0 & t < 0. \end{cases}$$

- The time between packets is called the interarrival time – the time between arrivals.
- For a Poisson process this follows the exponential distribution (above).
- This is easily shown – the probability of an arrival occurring before time t is one minus the probability of no arrivals occurring up until time t.
- The probability of no arrivals occurring during a time period t is $(\lambda t)^0 \exp[-(\lambda t)]/0! = \exp[-(\lambda t)]$.
- The mean interarrival time is $1/\lambda$.

The memoryless nature of the Poisson process

- There is something strange to be noticed here – the distribution of our interarrival time T was given by $\mathbb{P}[T \leq t] = 1 - \exp[-(\lambda t)$ for $t \geq 0$.
- However, if looked at the Poisson process at any instant and asked "how long must we wait for the next arrival?" the answer is just the same $1/\lambda$.
- Exactly the same argument can be made for any arrival time. The probability of no arrivals in the next t seconds does not change because an arrival has just happened.
- The expected waiting time for the next arrival does not change if you have been waiting for just one second, or for an hour or for many years – the average time to the next arrival is still the same $1/\lambda$.

The Poisson bus dilemma

$\cdots X(t) \cdots$

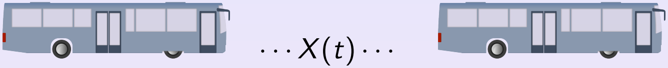
- Consider you arrive at the bus stop at a random time.
- Buses arrive as a Poisson process with a given rate λ.
- Buses are (on average) 30 minutes apart $1/\lambda = 30$ minutes.
- How long do you wait for the bus on average?

The Poisson bus dilemma

$\cdots X(t) \cdots$

- Consider you arrive at the bus stop at a random time.
- Buses arrive as a Poisson process with a given rate λ.
- Buses are (on average) 30 minutes apart $1/\lambda = 30$ minutes.
- How long do you wait for the bus on average?
- Bus passenger 1: Obviously 15 minutes – the buses are 30 minutes apart, on average I arrive half way through that period.

The Poisson bus dilemma



- Consider you arrive at the bus stop at a random time.
- Buses arrive as a Poisson process with a given rate λ.
- Buses are (on average) 30 minutes apart $1/\lambda = 30$ minutes.
- How long do you wait for the bus on average?
- Bus passenger 1: Obviously 15 minutes – the buses are 30 minutes apart, on average I arrive half way through that period.
- Bus passenger 2: Obviously 30 minutes – the buses are a Poisson process and memoryless. The average waiting time is 30 minutes no matter when the last bus was or when I arrive.

The Poisson bus dilemma

 $\cdots X(t) \cdots$ 

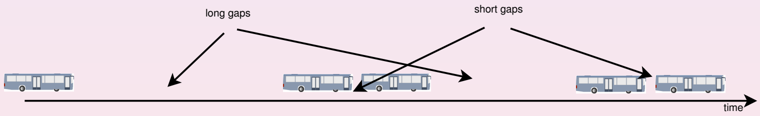
- Consider you arrive at the bus stop at a random time.
- Buses arrive as a Poisson process with a given rate λ.
- Buses are (on average) 30 minutes apart $1/\lambda = 30$ minutes.
- How long do you wait for the bus on average?
- Bus passenger 1: Obviously 15 minutes – the buses are 30 minutes apart, on average I arrive half way through that period.
- Bus passenger 2: Obviously 30 minutes – the buses are a Poisson process and memoryless. The average waiting time is 30 minutes no matter when the last bus was or when I arrive.
- So, who is correct?

The Poisson bus dilemma – solution

- So, is the answer 15 minutes, 30 minutes or something else.

The Poisson bus dilemma – solution

- So, is the answer 15 minutes, 30 minutes or something else.
- 30 minutes is the correct answer (as the Poisson process result show us).
- To see why the 15 minutes answer is wrong consider the diagram.
- The average gap between buses is 30 minutes.
- The average passenger does wait for half of the interarrival gap he or she arrives during.
- However, the average passenger is likely to arrive in a larger than average gap (see diagram).
- We do not need to prove that the answer is 30 minutes – the proof is already there for the Poisson process.



Introducing Markov chains

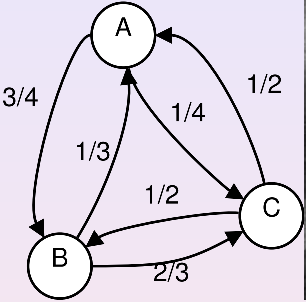
Markov Chains

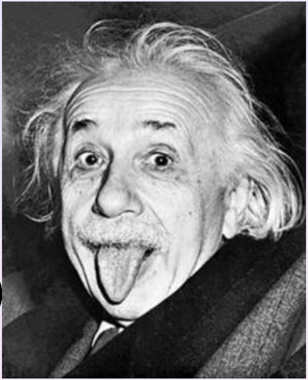
Markov chains are an elegant and useful mathematical tool used in many applied areas of mathematics and engineer but particularly in queuing theory.

- Useful when a system can be in a countable number of "states" (e.g. number of people in a queue, number of packets in a buffer and so on).
- Useful when transitions between "states" can be considered as a probabilistic process.
- Helps us analyse queues.

Andrey Andreyevich Markov (1856 -1922)

Introducing Markov chains – the puzzled professor

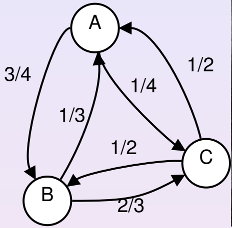


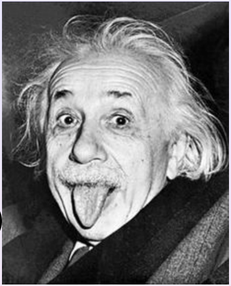


- Professor has lost his class. He tries lecture theatres A, B and C.
- Every day he moves to a different lecture theatre to look.
- He moves with probabilities as shown on the diagram.

The puzzled professor (2)

- Want to answer questions such as:
- What is probability he is in room A on day n?
- Where is he most likely to "end up"?

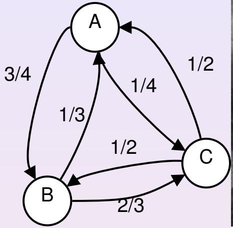


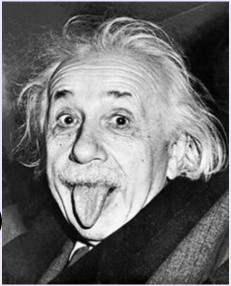


- First step – make system formal. Numbered states for rooms 0, 1 2 for A, B, C.
- Let p_{ij} be the probability of moving from room i to j on a day ($p_{ii} = 0$).
- Let $\lambda_{i,j}$ be the probability he is in room j on day i.
- Let $\boldsymbol{\lambda}_i = (\lambda_{i,0}, \lambda_{i,1}, \lambda_{i,2})$ be the vector of probabilities for day i.
- For example $\boldsymbol{\lambda}_0 = (1, 0, 0)$ means definitely in room A (room 0) on day 0.

The puzzled prof (3)

- Define the probability transition matrix P.
- Write down the equation for day n in terms of day $n-1$.
- We have: $\lambda_{n,j} = \sum_i \lambda_{n-1,i} p_{ij}$.





Transition matrix

$$\mathbf{P} = \begin{bmatrix} p_{00} & p_{01} & p_{02} \\ p_{10} & p_{11} & p_{12} \\ p_{20} & p_{21} & p_{22} \end{bmatrix}.$$

Matrix equation is $\boldsymbol{\lambda}_i = \boldsymbol{\lambda}_{i-1}\mathbf{P}$.

Equilibrium probabilities

- The matrix equation lets us calculate probabilities on a given day but where does prof "end up".
- Define "equilibrium probabilities" for states $\pi_i = \lim_{n \to \infty} \lambda_{n,i}$.
- Think of this as probability prof is in room i as time goes on.
- Define equilibrium vector $\boldsymbol{\pi} = (\pi_0, \pi_1, \pi_2)$.
- Can be shown that for a finite connected aperiodic chain this vector exists is unique and does not depend on start.
- From $\boldsymbol{\lambda}_i = \boldsymbol{\lambda}_{i-1}\mathbf{P}$ then $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}$.
- This vector and the requirement that probabilities sum to one uniquely defines π_i for all i.

Equilibrium probabilities – balance equations

- The matrix equation for $\boldsymbol{\pi}$ can also be thought of as "balance equations".
- That is in equilibrium, at every state the flow in a state is the sum of the flow going into it.
- $\pi_j = \sum_i p_{ij}\pi_i$ for all j (in matrix terms $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}$).
- This and $\sum_i \pi_i = 1$ are enough to solve the equations for π_i.

Equilibrium probabilities – balance equations

- The matrix equation for $\boldsymbol{\pi}$ can also be thought of as "balance equations".
- That is in equilibrium, at every state the flow in a state is the sum of the flow going into it.
- $\pi_j = \sum_i p_{ij}\pi_i$ for all j (in matrix terms $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}$).
- This and $\sum_i \pi_i = 1$ are enough to solve the equations for π_i.

$$\begin{aligned}
\pi_0 + \pi_1 + \pi_2 &= 1 && \text{probabilities sum to one} \\
\pi_1 p_{10} + \pi_2 p_{20} &= \pi_0 && \text{balance for room 0} \\
\pi_0 p_{01} + \pi_2 p_{21} &= \pi_1 && \text{balance for room 1} \\
\pi_0 p_{02} + \pi_1 p_{12} &= \pi_2 && \text{balance for room 2}
\end{aligned}$$

Solves as $\pi_0 = 16/55$, $\pi_1 = 21/55$ and $\pi_2 = 18/55$ for prof.

Markov chain summary

- A Markov chain is defined by a set of states and the probability of moving between them.
- This type of Markov chain is a discrete time homogeneous Markov chain.
- Continuous time Markov chains allow transitions at any time not just once per "day".
- Heterogenous Markov chains allow the transition probabilities to vary as time changes.
- Like the Poisson process, the Markov chain is "memoryless".
- Markov chains can be used in many types of problem solving, particularly queues.

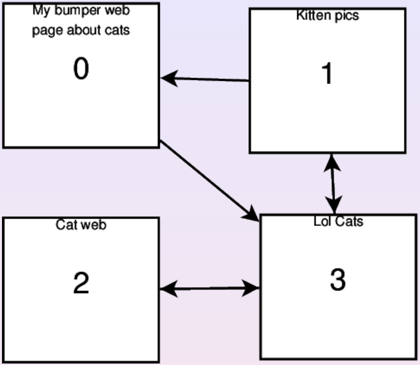
Markov recap

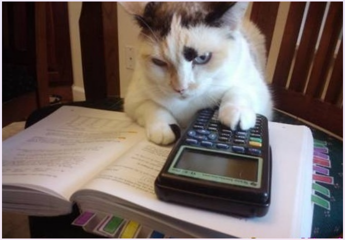
- Before going on to do some examples, a recap.
- p_{ij} is the transition probability – the probability of moving from state i to state j the next iteration of the chain.
- The transition matrix P is the matrix of the p_{ij}.
- π_i is the equilibrium probability – the probability that after a "long time" the chain will be in state i.
- The sum of π_i must be one (the chain must be in some state).
- Each state has a balance equation $\pi_i = \sum_j \pi_j p_{ji}$.
- The balance equations together with the sum of π_i will solve the chain (one redundant equation – why?).

The google page rank example

- Did you know google owes part of its success to Markov chains?
- "Pagerank" (named after Larry Page) was how google originally ranked search queries.
- Pagerank tries to work out which web page matching a search term is the most important.
- Pages with many links to them are very "important" but it is also important that the "importance" of the linking page counts.
- Here we consider a very simplified version.
- (Note that Larry Page is now a multi-billionaire thanks to Markov chains).

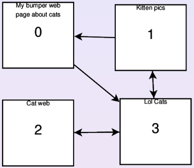
kittenweb – pagerank example

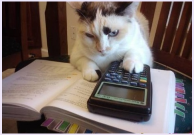




- Imagine these four web pages are every web page about kittens and cats on the web.
- An arrow indicates a link from one page to another – e.g. "Lol cats" and "Cat web" link to each other.

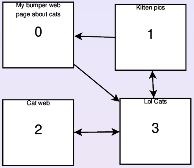
Kittenweb – pagerank example

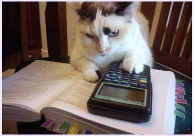




- Now think of a user randomly clicking on “cats/kittens” links.
- What page will the user visit most often – this is a Markov chain.
- “Lolcats” links to two other pages so 1/2 probability of visiting “Cat web” next.
- “Cat web” only links to “Lol cats” so probability 1 of visiting that next.

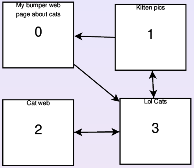
Kittenweb – pagerank example

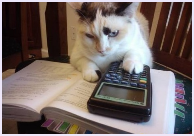




$$\mathbf{P} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1/2 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 1 \\ 0 & 1/2 & 1/2 & 0 \end{bmatrix}.$$

Kittenweb – pagerank example





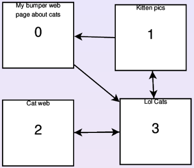
$$\mathbf{P} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1/2 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 1 \\ 0 & 1/2 & 1/2 & 0 \end{bmatrix}.$$

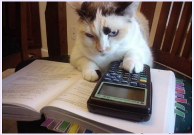
$$\pi_0 = \pi_1/2$$
$$\pi_1 = \pi_3/2$$
$$\pi_2 = \pi_3/2$$

miss equation for π_3

$$\pi_0 + \pi_1 + \pi_2 + \pi_3 = 1$$

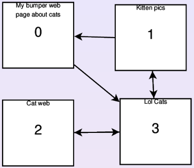
Kittenweb – pagerank example

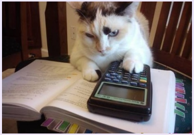




$$\pi_0 = \pi_1/2$$
$$\pi_1 = \pi_3/2$$
$$\pi_2 = \pi_3/2 \qquad \text{miss equation for } \pi_3$$
$$\pi_0 + \pi_1 + \pi_2 + \pi_3 = 1$$

Kittenweb – pagerank example





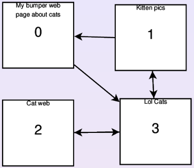
$$\pi_0 = \pi_1/2$$
$$\pi_1 = \pi_3/2$$
$$\pi_2 = \pi_3/2$$

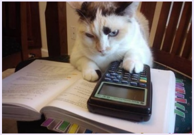
miss equation for π_3

$$\pi_0 + \pi_1 + \pi_2 + \pi_3 = 1$$

$\pi_1 = \pi_2$ from lines 2 and 3.

Kittenweb – pagerank example

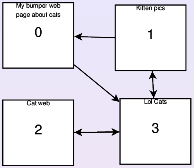


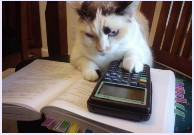


$$\pi_0 = \pi_1/2$$
$$\pi_1 = \pi_3/2$$
$$\pi_2 = \pi_3/2 \qquad \text{miss equation for } \pi_3$$
$$\pi_0 + \pi_1 + \pi_2 + \pi_3 = 1$$

$\pi_1 = \pi_2$ from lines 2 and 3.
$\pi_1 = 2\pi_0 = \pi_3/2$ from line 1 and 3.

Kittenweb – pagerank example

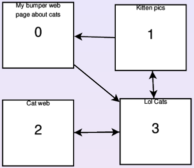


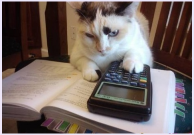


$$\pi_0 = \pi_1/2$$
$$\pi_1 = \pi_3/2$$
$$\pi_2 = \pi_3/2 \qquad \text{miss equation for } \pi_3$$
$$\pi_0 + \pi_1 + \pi_2 + \pi_3 = 1$$

$\pi_1 = \pi_2$ from lines 2 and 3.
$\pi_1 = 2\pi_0 = \pi_3/2$ from line 1 and 3.
$\pi_1/2 + \pi_1 + \pi_1 + 2\pi_1 = 1$ from line 4 and above lines.

Kittenweb – pagerank example

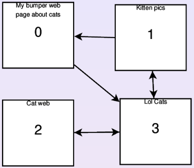


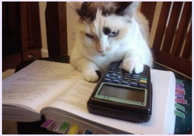


$$\pi_0 = \pi_1/2$$
$$\pi_1 = \pi_3/2$$
$$\pi_2 = \pi_3/2 \qquad \text{miss equation for } \pi_3$$
$$\pi_0 + \pi_1 + \pi_2 + \pi_3 = 1$$

$\pi_1 = \pi_2$ from lines 2 and 3.
$\pi_1 = 2\pi_0 = \pi_3/2$ from line 1 and 3.
$\pi_1/2 + \pi_1 + \pi_1 + 2\pi_1 = 1$ from line 4 and above lines.
$\pi_1 = 2/9 \quad \pi_0 = 1/9 \quad \pi_2 = 2/9 \quad \pi_3 = 4/9$

Kittenweb – pagerank example





$\pi_1 = 2/9 \quad \pi_0 = 1/9 \quad \pi_2 = 2/9 \quad \pi_3 = 4/9$

- So this page shows “Lol Cats” is the most important page, followed by “Cat web” and “Kitten pics” equally important.
- Note that pages 0,1 and 2 all have only one incoming link but are not equally important.
- Nowadays google has made many optimisations to their algorithm (and this is a simplified version anyway).
- Nonetheless this “random walk on a graph” principle remains important in many network models.

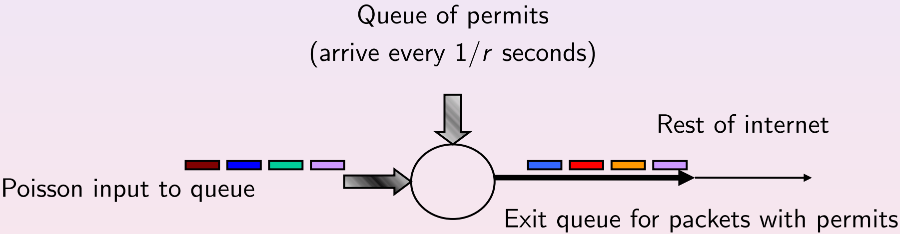
Queuing analysis of the leaky bucket model

- A "leaky bucket" is a mechanism for managing buffers and to smooth downstream flow.
- What is described here is sometimes known as a "token bucket".
- A queue holds a stock of "permit" generated at a rate r (one permit every $1/r$ seconds) up to a maximum of W.
- A packet cannot leave the queue without a permit – each packet takes one permit.
- The idea is that a short burst of traffic can be accommodated but a longer burst is smoothed to ensure that downstream can cope.
- Assume that packets arrive as a Poisson process at rate λ.
- A Markov model will be used [Bertsekas and Gallager page 515].

Modelling the leaky bucket

Use a discrete time Markov chain where we stay in each state for time $1/r$ seconds (the time taken to generate one permit). Let a_k be the probability that k packets arrive in one time period. Since arrivals are Poisson,

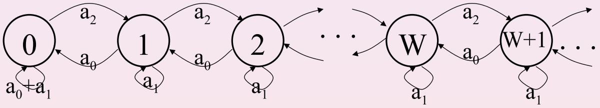
$$a_k = \frac{e^{-\lambda/r}(\lambda/r)^k}{k!}.$$



A Markov chain model of the situation

- In one time period (length $1/r$ secs) one token is generated (unless W exist) and some may be used sending packets.
- States $i \in \{0, 1, \ldots, W\}$ represent no packets waiting and $W - i$ permits available. States $i \in \{W+1, W+2, \ldots\}$ represent 0 tokens and $i - W$ packets waiting.
- If k packets arrive we move from state i to state $i + k - 1$ (except from state 0).
- Transition probabilities from i to j, $p_{i,j}$ given by

$$p_{i,j} = \begin{cases} a_0 + a_1 & i = j = 0 \\ a_{j-i+1} & j \geq i - 1 \\ 0 & \text{otherwise} \end{cases}$$



Continuous time Markov chains

- The Markov chains we looked at are "discrete" time – assume that one transition occurs every time unit.
- What if we want to drop this?
- We need to study "continuous time" Markov chains.
- As it turns out this is quite easy if the maths is treated carefully answers are nearly the same.

Continuous time Markov chains

- Consider a chain with states numbered from 0.
- Time step is some small δt and
- Transition probabilities from i to j given by $p_{ij}\delta t$.

$$\mathbf{P}(\delta t) = \begin{bmatrix} 1 - p_{00}\delta t & p_{01}\delta t & p_{02}\delta t & \dots \\ p_{10}\delta t & 1 - p_{11}\delta t & p_{12}\delta t & \dots \\ p_{20}\delta t & p_{21}\delta t & 1 - p_{22}\delta t & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

Note slightly "strange" definition of p_00 (why?)

Continuous time Markov chains

Define the following (assuming that the states of the chain are numbered $(0, 1, 2, \dots)$.

- $X(t)$ is the state of the chain at some time $t \geq 0$.
- $\mathbf{f}(t) = (f_0(t), f_1(t), \dots)$ is the vector of probabilities at time t, formally $f_i(t) = \mathbb{P}[X(t) = i]$.
- $q_{ij}(t_1, t_2)$ where $t_1 < t_2$ is $\mathbb{P}[X(t_2) = j | X(t_1) = i]$.

Since the context is still homogeneous chains then these probabilities are just a function of $\tau = t_2 - t_1$. Hence, define for $i \neq j$

$$q_{ij}(\tau) = q_{ij}(t_2 - t_1) = q_{ij}(t_1, t_2) = \mathbb{P}[X(\tau) = j | X(0) = i] .$$

Continuous time Markov chains

Define the following (assuming that the states of the chain are numbered $(0, 1, 2, \dots)$.

- $X(t)$ is the state of the chain at some time $t \geq 0$.
- $\mathbf{f}(t) = (f_0(t), f_1(t), \dots)$ is the vector of probabilities at time t, formally $f_i(t) = \mathbb{P}[X(t) = i]$.
- $q_{ij}(t_1, t_2)$ where $t_1 < t_2$ is $\mathbb{P}[X(t_2) = j | X(t_1) = i]$.

Since the context is still homogeneous chains then these probabilities are just a function of $\tau = t_2 - t_1$. Hence, define for $i \neq j$

$$q_{ij}(\tau) = q_{ij}(t_2 - t_1) = q_{ij}(t_1, t_2) = \mathbb{P}[X(\tau) = j | X(0) = i].$$

Define the limit

$$q_{ij} = \lim_{\tau \to 0} \frac{q_{ij}(\tau)}{\tau}.$$

Continuous time Markov chains

We can show in the limit if transitions are Poisson (see notes)

$$\frac{df_i(t)}{dt} = -\sum_{j \neq i} f_i(t) q_{ij} + \sum_{j \neq i} f_j(t) q_{ji}.$$

Continuous time Markov chains

We can show in the limit if transitions are Poisson (see notes)

$$\frac{df_i(t)}{dt} = -\sum_{j \neq i} f_i(t) q_{ij} + \sum_{j \neq i} f_j(t) q_{ji}.$$

It is handy to define

$$q_{ii} = -\sum_{i \neq j} q_{ij}.$$

Continuous time Markov chains

We can show in the limit if transitions are Poisson (see notes)

$$\frac{df_i(t)}{dt} = -\sum_{j \neq i} f_i(t) q_{ij} + \sum_{j \neq i} f_j(t) q_{ji}.$$

It is handy to define

$$q_{ii} = -\sum_{i \neq j} q_{ij}.$$

Now, define the matrix

$$\mathbf{Q} = \begin{bmatrix} q_{00} & q_{01} & q_{02} & \cdots \\ q_{10} & q_{11} & q_{12} & \cdots \\ q_{20} & q_{21} & q_{22} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

Continuous time Markov chains

We can show in the limit if transitions are Poisson (see notes)

$$\frac{df_i(t)}{dt} = -\sum_{j\neq i} f_i(t) q_{ij} + \sum_{j \neq i} f_j(t) q_{ji}.$$

It is handy to define

$$q_{ii} = -\sum_{i \neq j} q_{ij}.$$

Now, define the matrix

$$\mathbf{Q} = \begin{bmatrix} q_{00} & q_{01} & q_{02} & \cdots \\ q_{10} & q_{11} & q_{12} & \cdots \\ q_{20} & q_{21} & q_{22} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

It can now be seen that

$$\frac{d\mathbf{f}(t)}{dt} = \mathbf{f}(t)\mathbf{Q}.$$

Completing the continuous Markov Chain (1)

Notice that

$$\mathbf{Q} = \mathbf{P}(1) - \mathbf{I},$$

where $\mathbf{I}$ is the identity matrix.
Assume the chain is finite and there are no disconnected states.
Now the equilibrium probabilities can be calculated. In this case

$$\boldsymbol{\pi} = \lim_{t \to \infty} \mathbf{f}(t).$$

Completing the continuous Markov Chain (1)

Notice that

$$\mathbf{Q} = \mathbf{P}(1) - \mathbf{I},$$

where $\mathbf{I}$ is the identity matrix.
Assume the chain is finite and there are no disconnected states.
Now the equilibrium probabilities can be calculated. In this case

$$\boldsymbol{\pi} = \lim_{t \to \infty} \mathbf{f}(t).$$

Therefore

$$\boldsymbol{\pi}\mathbf{Q} = \frac{d \lim_{t \to \infty} \mathbf{f}(t)}{dt} =$$

Completing the continuous Markov Chain (1)

Notice that

$$\mathbf{Q} = \mathbf{P}(1) - \mathbf{I},$$

where $\mathbf{I}$ is the identity matrix.
Assume the chain is finite and there are no disconnected states.
Now the equilibrium probabilities can be calculated. In this case

$$\boldsymbol{\pi} = \lim_{t\to\infty} \mathbf{f}(t).$$

Therefore

$$\boldsymbol{\pi}\mathbf{Q} = \frac{d \lim_{t\to\infty} \mathbf{f}(t)}{dt} = \frac{d}{dt} \lim_{t\to\infty} \mathbf{f}(t) =$$

Completing the continuous Markov Chain (1)

Notice that

$$\mathbf{Q} = \mathbf{P}(1) - \mathbf{I},$$

where $\mathbf{I}$ is the identity matrix.
Assume the chain is finite and there are no disconnected states.
Now the equilibrium probabilities can be calculated. In this case

$$\boldsymbol{\pi} = \lim_{t \to \infty} \mathbf{f}(t).$$

Therefore

$$\boldsymbol{\pi}\mathbf{Q} = \frac{d \lim_{t\to\infty} \mathbf{f}(t)}{dt} = \frac{d}{dt} \lim_{t\to\infty} \mathbf{f}(t) = \frac{d\boldsymbol{\pi}}{dt} = 0.$$

Completing the continuous Markov Chain (2)

This gives a new version of our balance equations. For all i then

$$\sum_j \pi_j q_{ji} = 0$$

Expanding q_{jj} from its definition and multiplying by -1 gives

$$\sum_{j \neq i} \pi_i q_{ij} - \sum_{j \neq i} \pi_j q_{ji} = 0.$$

This can also be seen as a balance of flows into and out of the state. For all i:

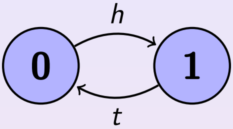
$$\sum_{j \neq i} \pi_i q_{ij} = \sum_{j \neq i} \pi_j q_{ji} \qquad (output) = (input)$$

Also as usual $\sum_i \pi_i = 1$.

The "talking on the phone" example

- If I am talking on the phone, I will hang up as a Poisson process with a rate h (for hang up).
- If I am not talking on the phone, I will decide to start a new call as a Poisson process with rate t (for talk).
- At a given time what is the probability I am talking on the phone.
- Unsurprisingly this can be modelled as a Markov chain.
- This example may seem "trivial" but several such chains could be use to model how occupied the phone network is.

The "talking on the phone" example

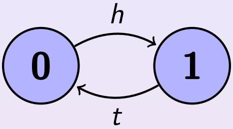


Our chain has two states 0 (talking) and 1 (not talking) and the transition matrix:

$$\mathbf{Q} = \begin{bmatrix} -h & h \\ t & -t \end{bmatrix}.$$

We need our new balance equations:

The "talking on the phone" example



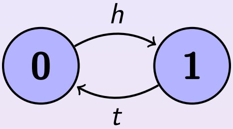
Our chain has two states 0 (talking) and 1 (not talking) and the transition matrix:

$$\mathbf{Q} = \begin{bmatrix} -h & h \\ t & -t \end{bmatrix}.$$

We need our new balance equations:

- State 0 – (output) $h\pi_0 = t\pi_1$ (input)

The "talking on the phone" example



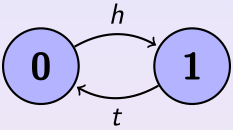
Our chain has two states 0 (talking) and 1 (not talking) and the transition matrix:

$$\mathbf{Q} = \begin{bmatrix} -h & h \\ t & -t \end{bmatrix}.$$

We need our new balance equations:

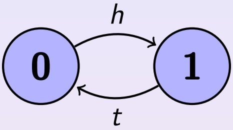
- State 0 – (output) $h\pi_0 = t\pi_1$ (input)
- State 1 – (output) $t\pi_1 = h\pi_0$ (input)

The “talking on the phone” example



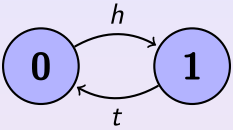
- State 0 – (output) $h\pi_0 = t\pi_1$ (input)
- State 1 – (output) $t\pi_1 = h\pi_0$ (input)

The "talking on the phone" example



- State 0 – (output) $h\pi_0 = t\pi_1$ (input)
- State 1 – (output) $t\pi_1 = h\pi_0$ (input)
- We also need $\pi_0 + \pi_1 = 1$ which gives from state 0 $h\pi_0 = t(1 - \pi_0)$.

The "talking on the phone" example



- State 0 – (output) $h\pi_0 = t\pi_1$ (input)
- State 1 – (output) $t\pi_1 = h\pi_0$ (input)
- We also need $\pi_0 + \pi_1 = 1$ which gives from state 0 $h\pi_0 = t(1 - \pi_0)$.
- Rearrange to $\pi_0 = t/(h + t)$ and $\pi_1 = h/(h + t)$.
- Interpretation – the proportion of time talking (state 0) is proportional to the talking rate t and the proportion hung up is proportional to the hangup rate.

Markov Chain reminder

Balance equations

N equations for N states of chain – but only $N-1$ independent.

- Discrete – For all i then

$$\sum_j \pi_j p_{ji} = \pi_i \qquad \text{(input)} = \text{(original probability)}$$

- Continuous – For all i then

$$\sum_{j \neq i} \pi_j q_{ji} = \sum_{j \neq i} \pi_i q_{ij} \qquad (input) = (output)$$

Probabilities sum to one

For both types $\sum_i \pi_i = 1$.

Lecture summary

Stochastic processes

Stochastic processes are processes describing how a system evolves which are in some way "random" but can lead to interesting behaviour.

Poisson processes

Poisson processes are stochastic processes which can be thought of as describing arrivals. They are memoryless and have many useful mathematical properties.

Markov chains

Markov chains are stochastic processes which describe the evolution of a system between states. They remember only their current state and can be used to model a wide variety of situations.