Performance Modelling of Computer Systems

Mirco Tribastone

Institut für Informatik
Ludwig-Maximilians-Universität München

Stochastic Processes

Discrete-Time Markov Chains

Outline

From random variables to stochastic processes:

- Classification
- Discrete-time Markov chains
- Sojourn time
- Path probabilities
- Chapman-Kolmogorov equations

We will use the following notions:

- Discrete random variables
- Conditional probability
- Joint probability
- Basics of linear algebra (matrix multiplications)

Stochastic Processes

A stochastic process is a family of random variables $\{X(t), t \in T\}$, where the parameter t usually represents time. The range of $X(t)$ is usually called the state space of the stochastic process and each value in the range is called a state.

Therefore, a stochastic process is often intended to characterise the behaviour of a system as a function of time.

Depending on the nature of $X(t)$ and T, the stochastic process $\{X(t), t \in T\}$ can be classified as:

- Continuous-time, if T is an interval of the reals (typically, $T = \{t \in \mathbb{R} : t \geq 0\}$)
- Discrete-time, if T is discrete (typically, $T = \mathbb{N}$)
- Continuous-state, X is a continuous random variable
- Discrete-state, X is a discrete random variable

Markov Chains: Definitions and Notation

A Markov chain is a special class of stochastic processes in which the joint probability distributions between the random variables of the process enjoy the memoryless property.

For a discrete-time Markov chain (DTMC) the memoryless property says that

$$\mathbb{P}\big(X(t_{n+1}) = x_{n+1} \mid X(t_n) = x_n, X(t_{n-1}) = x_{n-1}, \ldots, X(t_0) = x_0\big) \\ = \mathbb{P}\big(X(t_{n+1}) = x_{n+1} \mid X(t_n) = x_n\big),$$

for all natural numbers n and for all states x_n.

We will often write X_n for $X(t_n)$ and designate the states with a single letter such as i, j, k. For instance, we write $\mathbb{P}(X_{n+1} = i)$ instead of $\mathbb{P}\big(X(t_{n+1}) = x_i\big)$.

Markov Chains: Definitions and Notation

Memoryless Property

$$\mathbb{P}\big(X(t_{n+1}) = x_{n+1} \mid X(t_n) = x_n, X(t_{n-1}) = x_{n-1}, \ldots, X(t_0) = x_0\big) \\ = \mathbb{P}\big(X(t_{n+1}) = x_{n+1} \mid X(t_n) = x_n\big)$$

The conditional probabilities

$$p_{ij}(n) = \mathbb{P}(X_{n+1} = j \mid X_n = i)$$

are called the transition probabilities of the Markov chain. The probability matrix $P(n)$ is formed from the transition probabilities as follows:

$$P(n) = \begin{bmatrix} p_{00}(n) & p_{01}(n) & \cdots & p_{0j}(n) & \cdots \\ p_{10}(n) & p_{11}(n) & \cdots & p_{1j}(n) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ p_{i0}(n) & p_{i1}(n) & \cdots & p_{ij}(n) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

Markov Chain: Properties

$$P(n) = \begin{bmatrix} p_{00}(n) & p_{01}(n) & \cdots & p_{0j}(n) & \cdots \\ p_{10}(n) & p_{11}(n) & \cdots & p_{1j}(n) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ p_{i0}(n) & p_{i1}(n) & \cdots & p_{ij}(n) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

- Since each element of the matrix is a probability

$$0 \leq p_{ij}(n) \leq 1, \qquad \text{for all } i \text{ and } j.$$

- Also, for all i,

$$\sum_j p_{ij}(n) = 1.$$

Markov Chain: Properties

In general, the transition probabilities depend on the time step n. If

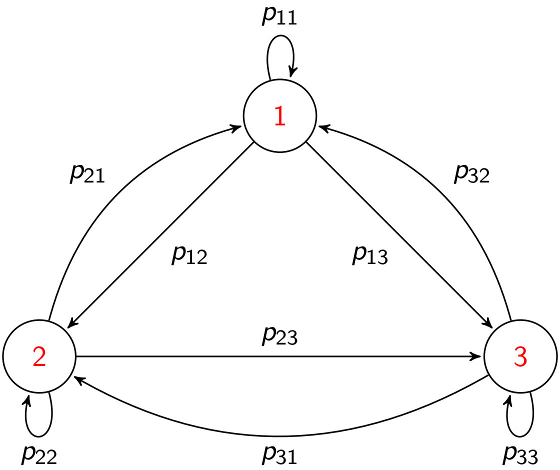
$$\mathbb{P}(X_{n+1} = j \mid X_n = i) = \mathbb{P}(X_{n+1+m} = j \mid X_{n+m} = i)$$

for all n and $m \geq 0$ then the Markov chain is said to be time-homogeneous.

In a time-homogenous Markov chain, the transition probabilities and the probability matrix are denoted by p_{ij} and P, respectively (i.e., the parameter n is dropped because the quantities do not depend on n).

We will mostly be working with time-homogeneous Markov chains from now on.

Graphical Representation of a DTMC



Sample-Path Probabilities

Given that the chain is in state i at time t_n, we wish to compute the probability that it is in state j at time t_{n+1} and in state k at time t_{n+2}:

$$\begin{aligned}\mathbb{P}(X_{n+2} = k, X_{n+1} = j \mid X_n = i) &= \mathbb{P}(X_{n+2} = k \mid X_{n+1} = j, X_n = i)\mathbb{P}(X_{n+1} = j \mid X_n = i) \\ &= \mathbb{P}(X_{n+2} = k \mid X_{n+1} = j)\,\mathbb{P}(X_{n+1} = j \mid X_n = i) \\ &= p_{jk}(n+1)p_{ij}(n).\end{aligned}$$

(see also tutorial sheet. . .)

In general, given a sample path $i, j, k, \dots, w, z$ at time steps $t_n, t_{n+1}, t_{n+2}, \dots, t_{n+m-1}, t_{n+m}$,

$$\begin{aligned}&\mathbb{P}(X_{n+m} = z, X_{n+m-1} = y, \dots, X_{n+2} = k, X_{n+1} = j \mid X_n = i) \\ &\quad = \mathbb{P}(X_{n+m} = z \mid X_{n+m-1} = y)\mathbb{P}(X_{n+m-1} = y \mid X_{n+m-2} = x)\cdots \\ &\qquad \cdots\mathbb{P}(X_{n+2} = k \mid X_{n+1} = j)\mathbb{P}(X_{n+1} = j \mid X_n = i) \\ &\quad = p_{yz}(n+m-1)p_{xy}(n+m-2)\cdots p_{jk}(n+1)p_{ij}(n) \\ &\quad = p_{ij}(n)p_{jk}(n+1)\cdots p_{xy}(n+m-2)p_{yz}(n+m-1).\end{aligned}$$

Sojourn Time in a DTMC

The probability of the sample path $\underbrace{i \to i \to \cdots \to i}_{n} \to j$ is

$$\underbrace{p_{ii}\, p_{ii}\, \cdots p_{ii}}_{n-1}\, p_{ij},$$

and for some other state $k \neq j$ the probability of the sample path $\underbrace{i \to i \to \cdots \to i}_{n} \to k$ is

$$\underbrace{p_{ii}\, p_{ii}\, \cdots p_{ii}}_{n-1}\, p_{ik}.$$

So, the probability that the $(n+1)$-th state is not i is

$$\underbrace{p_{ii}\, p_{ii}\, \cdots p_{ii}}_{n-1} \left(1 - \sum_{j \neq i} p_{ij}\right) = p_{ii}^{n-1}(1 - p_{ii}).$$

This is the probability that the chain exits state i after n steps (sojourn time) and is denoted by the random variable R_i.

Sojourn Time in a DTMC

Thus, in a time-homogeneous DTMC, the sojourn time R_i of the state i has the following distribution:

$$\mathbb{P}(R_i = k) = \begin{cases} p_{ii}^{k-1}(1 - p_{ii}) & \text{if } k = 1, 2, \ldots \\ 0 & \text{otherwise.} \end{cases}$$

This is the geometric distribution. It can be shown that the geometric distribution is the only discrete random variable which enjoys the memoryless property, i.e.,

$$\mathbb{P}(R_i > m + n \mid R_i > m) = \mathbb{P}(R_i > n), \qquad n > 0.$$

We have also that

$$\mathbb{E}[R_i] = \frac{1}{1 - p_{ii}} \qquad \text{and} \qquad \text{Var}[R_i] = \frac{p_{ii}}{(1 - p_{ii})^2}.$$

Chapman-Kolmogorov Equations

The path probability gives the probability that the chain follows a given sample path given some initial state. We now wish to compute the probability that, after m steps from some initial state i, the chain is in state l, i.e.,

$$p_{il}^{(m)} := \mathbb{P}(X_{n+m} = l \mid X_n = i).$$

For example, in a time-homogeneous DTCM with states $\{i, j, k\}$ we may wish to compute $p_{ik}^{(2)} = \mathbb{P}(X_{n+2} = k \ \mid X_n = i)$.
We compute the probabilities of all possible two-step paths from i to k:

$$\mathbb{P}(X_{n+2} = k, X_{n+1} = i \mid X_n = i) = p_{ii}\, p_{ik},$$
$$\mathbb{P}(X_{n+2} = k, X_{n+1} = j \mid X_n = i) = p_{ij}\, p_{jk},$$
$$\mathbb{P}(X_{n+2} = k, X_{n+1} = k \mid X_n = i) = p_{ik}\, p_{kk},$$

Since all these events are disjoint, by the law of total probability

$$p_{ik}^{(2)} = \mathbb{P}(X_{n+2} = k \mid X_n = i) = p_{ii} p_{ik} + p_{ij} p_{jk} + p_{ik} p_{kk} = \sum_{z \in \{i,j,k\}} p_{iz} p_{zk}.$$

A Matrix Interpretation

$$p_{ik}^{(2)} = \mathbb{P}(X_{n+2} = k \mid X_n = i) = p_{ii}p_{ik} + p_{ij}p_{jk} + p_{ik}p_{kk} = \sum_{z \in \{i,j,k\}} p_{iz}p_{zk}.$$

A matrix interpretation:

$$P \times P = \begin{bmatrix} p_{ii} & p_{ij} & p_{ik} \\ p_{ji} & p_{jj} & p_{jk} \\ p_{ki} & p_{kj} & p_{kk} \end{bmatrix} \times \begin{bmatrix} p_{ii} & p_{ij} & p_{ik} \\ p_{ji} & p_{jj} & p_{jk} \\ p_{ki} & p_{kj} & p_{kk} \end{bmatrix}$$

$$(P \times P)_{i,k} = p_{ii}p_{ik} + p_{ij}p_{jk} + p_{ik}p_{kk} = \mathbb{P}(X_{n+2} = k \mid X_n = i)$$

In general $p_{ij}^{(2)} = (P^2)_{ij}$, for a time-homogeneous DTMC.

Chapman-Kolmogorov Equations

Let us compute

$$p_{il}^{(m)} = \mathbb{P}(X_{n+m} = l \mid X_n = i)$$

for $m = 3$.

The probability of a sample path from i to l has the following general form

$$\mathbb{P}(X_{n+3} = l, X_{n+2} = k, X_{n+1} = j \mid X_n = i) = p_{ij}\, p_{jk}\, p_{kl}.$$

Applying the law of total probability:

$$\begin{aligned} p_{il}^{(3)} = \mathbb{P}(X_{n+3} = l \mid X_n = i) &= \sum_j \sum_k p_{ij} p_{jk} p_{kl} \\ &= \sum_j p_{ij} \sum_k p_{jk} p_{kl} = \sum_j p_{ij} p_{jl}^{(2)} = (P^3)_{il}. \end{aligned}$$

Chapman-Kolmogorov Equations

The result

$$p_{il}^{(3)} = \mathbb{P}(X_{n+3} = l \mid X_n = i) = \sum_j p_{ij} p_{jl}^{(2)}$$

may be generalised as follows:

$$p_{ij}^{(m)} := \mathbb{P}(X_{n+m} = j \mid X_n = i) = \sum_k p_{ik}^{(l)} p_{kj}^{(m-l)}, \qquad \text{for } 0 < l < m.$$

For a homogeneous DTMC:

$$\begin{aligned}
p_{ij}^{(m)} &:= \mathbb{P}(X_m = j \mid X_0 = i) \\
&= \sum_k \mathbb{P}(X_m = j, X_l = k \mid X_0 = i) \qquad \text{for } 0 < l < m \\
&= \sum_k \mathbb{P}(X_m = j \mid X_l = k, X_0 = i)\mathbb{P}(X_l = k \mid X_0 = i) \\
&= \sum_k \mathbb{P}(X_m = j \mid X_l = k)\mathbb{P}(X_l = k \mid X_0 = i) = \sum_k p_{ik}^{(l)} p_{kj}^{(m-l)}.
\end{aligned}$$

Chapman-Kolmogorov Equations

$$p_{ij}^{(m)} := \mathbb{P}(X_{n+m} = j \mid X_n = i) = \sum_k p_{ik}^{(l)} p_{kj}^{(m-l)}, \qquad \text{for } 0 < l < m.$$

In matrix notation,

$$P^{(m)} = P^{(l)} P^{(m-l)}.$$

For $l = 1$,

$$P^{(m)} = P\, P^{(m-1)} = P^{(m-1)}\, P = P^m.$$

Probability Distribution of a DTMC

Let us define

$$\pi_i^{(n)} := \mathbb{P}(X_n = i).$$

The row vector

$$\pi^{(n)} = [\pi_0^{(n)}, \pi_1^{(n)}, \ldots, \pi_i^{(n)}, \ldots, \pi_j^{(n)}, \ldots]$$

is called the probability distribution of the DTMC at time n.

Given $\pi^{(0)}$ of a DTMC with state space $\mathcal{I}$, $\pi^{(n)}$ may be computed from the one-step transition probabilities of the DTMC. For all $i \in \mathcal{I}$,

$$\begin{aligned}\pi_i^{(1)} = \mathbb{P}(X_1 = i) &= \mathbb{P}(X_1 = i \mid X_0 = 0)\mathbb{P}(X_0 = 0) \\ &\quad + \mathbb{P}(X_1 = i \mid X_0 = 1)\mathbb{P}(X_0 = 1) + \ldots \\ &= p_{0i}(0)\pi_0^{(0)} + p_{1i}(0)\pi_1^{(0)} + \ldots = \sum_{s \in \mathcal{I}} p_{si}(0)\pi_s^{(0)}.\end{aligned}$$

Probability Distribution of a DTMC

In matrix notation,

$$\pi^{(1)} = \pi^{(0)} P(0).$$

Using the same arguments,

$$\pi^{(2)} = \pi^{(1)} P(1) = \pi^{(0)} P(0) P(1).$$

If the DTMC is time-homogeneous, this simplifies to

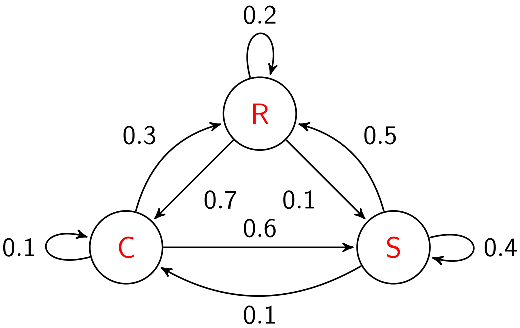
$$\pi^{(2)} = \pi^{(1)} P = \pi^{(0)} P^2.$$

These results generalise as follows:

$$\pi^{(n)} = \pi^{(n-1)} P(n-1) = \pi^{(0)} P(0) P(1) \cdots P(n-1)$$
$$\pi^{(n)} = \pi^{(n-1)} P = \pi^{(0)} P^n.$$

$\pi^{(n)}$ is the transient probability distribution of the Markov chain at step n.

Example: Munich Weather Model



$$P = \begin{bmatrix} 0.1 & 0.3 & 0.6 \\ 0.7 & 0.2 & 0.1 \\ 0.1 & 0.5 & 0.4 \end{bmatrix} \begin{matrix} \color{red}{C} \\ \color{red}{R} \\ \color{red}{S} \end{matrix}$$

Suppose that $\pi^{(0)} = \begin{bmatrix} 0.5000 & 0.5000 & 0.0000 \end{bmatrix}$.

$$\pi^{(1)} = \begin{bmatrix} 0.5000 & 0.5000 & 0.0000 \end{bmatrix} \begin{bmatrix} 0.1 & 0.3 & 0.6 \\ 0.7 & 0.2 & 0.1 \\ 0.1 & 0.5 & 0.4 \end{bmatrix} = \begin{bmatrix} 0.4000 & 0.2500 & 0.3500 \end{bmatrix}.$$

$$\pi^{(2)} = \begin{bmatrix} 0.4000 & 0.2500 & 0.3500 \end{bmatrix} \begin{bmatrix} 0.1 & 0.3 & 0.6 \\ 0.7 & 0.2 & 0.1 \end{bmatrix} = \begin{bmatrix} 0.2500 & 0.3450 & 0.4050 \end{bmatrix}$$

Limiting Distribution

We observe that there exists

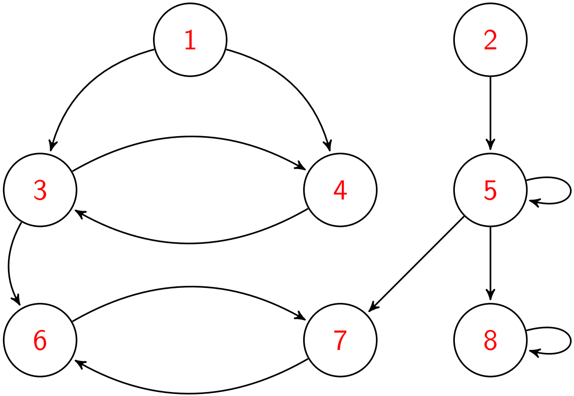
$$\lim_{n \to \infty} \pi^{(n)}$$

which is called a limiting distribution of the DTMC.

We need to check some properties on the states of the Markov chain to be able answer the following questions:

- Does it always exist?
- Is it unique?
- Does it depend on the initial distribution?

Classification of States



State Classification

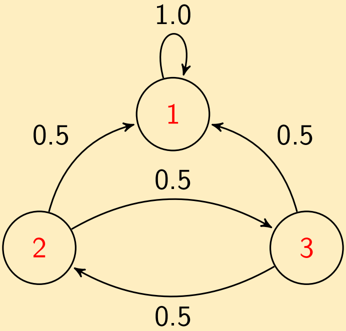
A state is said to be

- Recurrent if the Markov chain may return to the state infinitely often
  - A recurrent state is said to be periodic of period k if the chain returns to that state every k time steps.
  - The mean recurrence time is the average number of steps to return to a recurrent state.
    - If it is finite, the state is said to be positive recurrent
    - Else, the state is null recurrent (this may happen only if the state space is infinite)
  - A recurrent state i is absorbing if $p_{ii} = 1$
- Transient if there is non-zero probability that the chain will never return to that state.

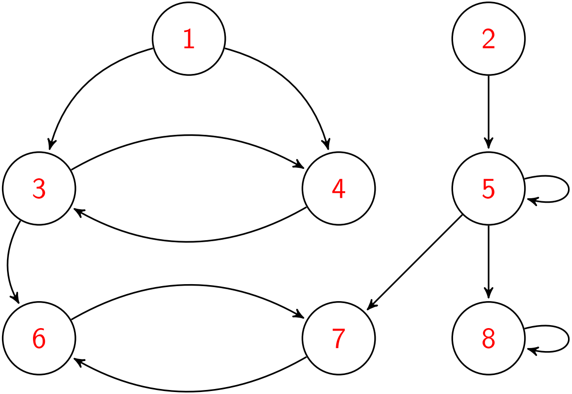
Ergodicity

The period of a periodic state i is defined as the greatest common divisor of the set of integers n such that $p_{ii}^{(n)} > 0$. If this is equal to one, then the state is said to be aperiodic. A positive recurrent and aperiodic state is said to be ergodic. A Markov chain is said to be ergodic if every state is ergodic.

Example of a non-ergodic DTMC



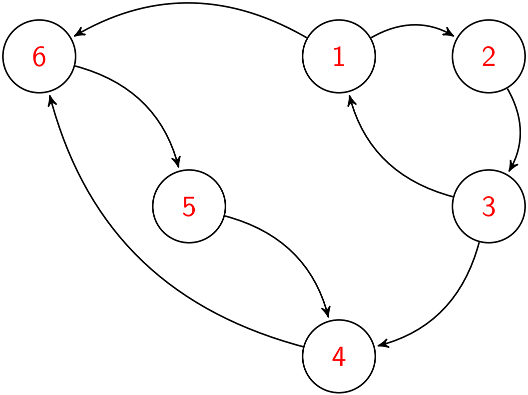
State Classification



Classification of Groups of States

- A nonempty subset S of the state space is said to be closed if none of the states in S have paths that lead outside S in any number of steps.
- For instance, an absorbing state is a closed subset of the state space.
- A subset that is not closed is said to be open.
- If the state space is closed and no proper subset is closed then the Markov chain is said to be irreducible. Else, the Markov chain is reducible.
- An alternative definition of irreducibility is that for every pair of states (i, j) there exists at least a path that leads from i to j, i.e., $p_{ij}^{(n)} > 0$ for some n.

Classification of Groups of States



Properties of Markov chains

- If a Markov chain is finite and irreducible then it is positive recurrent.
- The states of an aperiodic, finite, irreducible Markov chain are ergodic.

Limiting Distribution

We saw earlier on that the weather model admitted a limiting distribution. The general definition follows.

Limiting Distribution

Given a transition probability matrix P of a time-homogeneous DTMC and $\pi^{(0)}$ an initial probability distribution, if the limit

$$\lim_{n\to\infty} P^{(n)} = \lim_{n\to\infty} P^n$$

exists then the probability distribution

$$\pi_l := \lim_{n\to\infty} \pi^{(n)} = \pi^{(0)} \lim_{n\to\infty} P^{(n)} = \pi^{(0)} \lim_{n\to\infty} P^n$$

exists and is called a limiting distribution of the DTMC.

Results

Limiting Distribution of an Ergodic Chain

If the states of the Markov chain are ergodic, then the limiting distribution exists and is unique.

The same results holds if the Markov chain is finite, irreducible and aperiodic because the first two properties imply that the chain is positive recurrent (see earlier).

If these conditions do not hold, the chain may not have a limiting distribution . . .

Example

Consider the following transition probability matrix:

$$P = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

It may be seen that

$$P^2 = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, P^3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, P^4 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \ldots$$

that is,

$$P, P^2, P^3, P^4 = P, P^5 = P^2, P^6 = 3, P^7 = P, \ldots$$

which clearly does not admit any limiting distribution.
What goes wrong with P?

Steady-State Distribution

When $P^{(n)}$ converges, different initial conditions may lead to different limiting distributions.

Steady-state Distribution

A limiting distribution π is a steady-state distribution if it converges, independently of the initial distribution $\pi^{(0)}$, to a vector whose components are strictly positive and sum to 1. If a steady-state distribution exists, it is unique.

The steady-state distribution gives the distribution of the probability mass in the Markov chain when the process is sufficiently away from the initial condition such that it is no influence on the behaviour of the system. For this reason it is also called the *long-run* probability distribution.

Stationary Distribution

Let P be the transition probability matrix of a DTMC, and let the vector $z = (\ldots, z_j, \ldots)$ be a probability distribution, i.e.,

$$0 \leq z_j < 1, \quad \text{for all} \quad j \quad \text{and} \quad \sum_j z_j = 1.$$

The vector z is said to be a stationary distribution if

$$z = zP.$$

Therefore, it follows that

$$z = zP = zP^2 = \ldots = zP^n = \ldots$$

Stationary and Steady-State Distributions

- The steady-state distribution gives the distribution after the influence of the initial condition has passed.
- The stationary distribution instead is a distribution that, if reached, never changes after any length of time.
- When a steady-state distribution exists, then that distribution is also the unique stationary distribution.
- However, the existence of a stationary distribution does not imply that a steady-state distribution exists . . .

Stationary and Steady-State Distributions

The DTMC with transition probability matrix

$$P = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

has a stationary distribution $\begin{bmatrix} 1/3 & 1/3 & 1/3 \end{bmatrix}$ but $P^{(n)}$ does not converge.

Stationary and Steady-State Distribution

A DTMC may admit more than one stationary distributions. For instance, given

$$P = \begin{bmatrix} 0.4 & 0.6 & 0.0 & 0.0 \\ 0.6 & 0.4 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.5 & 0.5 \\ 0.0 & 0.0 & 0.5 & 0.5 \end{bmatrix}$$

all vectors $z = \left[\alpha/2, \alpha/2, (1-\alpha)/2, (1-\alpha)/2\right], 0 \leq \alpha \leq 1$ are stationary distributions.

Sufficient Condition for Uniqueness

However, if the Markov chain is finite and irreducible, then a unique stationary distribution exists. This unique distribution can be found by solving the linear system of equations

$$z(P - I) = 0, \qquad \text{and} \qquad \sum_i z_i = 1.$$

Stationary and Steady-State Distribution

- In many cases of interest we will be dealing with finite, irreducible, and aperiodic Markov chains.
- For such chains there exist a unique stationary distribution, which is also the unique steady-state distribution of the chain.

If a stationary distribution exists, it means that in the system of linear equations

$$z(P - I) = 0$$

the coefficient matrix $(P - I)$ must be singular, otherwise it would have the only solution $z = 0$. Since it is singular, it gives rise to linearly dependent equations ...

Solving for the Stationary Distribution

Given the following time-homogeneous DTMC,

$$P = \begin{bmatrix} 0.1 & 0.3 & 0.6 \\ 0.7 & 0.2 & 0.1 \\ 0.1 & 0.5 & 0.4 \end{bmatrix}$$

we have that

$$z(P - I) = \begin{bmatrix} z_1 & z_2 & z_3 \end{bmatrix} \begin{bmatrix} -0.9 & 0.3 & 0.6 \\ 0.7 & -0.8 & 0.1 \\ 0.1 & 0.5 & -0.6 \end{bmatrix}$$

$$= \begin{cases} -0.9z_1 + 0.7z_2 + 0.1z_3 = 0 \\ 0.3z_1 - 0.8z_2 + 0.5z_3 = 0 \\ 0.6z_1 + 0.1z_2 - 0.6z_3 = 0 \end{cases}$$

Solving for the Stationary Distribution

$$\begin{cases} -0.9z_1 + 0.7z_2 + 0.1z_3 = 0 \\ 0.3z_1 - 0.8z_2 + 0.5z_3 = 0 \\ 0.6z_1 + 0.1z_2 - 0.6z_3 = 0 \end{cases}$$

Since only two of such equations are sufficient (e.g., the first two), we replace one (e.g., the third) with the normalising condition $\sum_i z_i = 1$, i.e.:

$$\begin{cases} -0.9z_1 + 0.7z_2 + 0.1z_3 = 0 \\ 0.3z_1 - 0.8z_2 + 0.5z_3 = 0 \\ z_1 + z_2 + z_3 = 1 \end{cases}$$

In matrix notation, this corresponds to having the following system of equations

$$\begin{bmatrix} z_1 & z_2 & z_3 \end{bmatrix} \begin{bmatrix} -0.9 & 0.3 & 1.0 \\ 0.7 & -0.8 & 1.0 \\ 0.1 & 0.5 & 1.0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Continuous-Time Markov Chains

The $\mathcal{O}$ Notation

We define for a $c \in \mathbb{R} \cup \{\pm\infty\}$

$$f(x) \in o(g(x)) \text{ for } x \to c :\Leftrightarrow \lim_{x \to c} \left| \frac{f(x)}{g(x)} \right| = 0$$

$$f(x) \in \mathcal{O}(g(x)) \text{ for } x \to c :\Leftrightarrow \overline{\lim_{x \to c}} \left| \frac{f(x)}{g(x)} \right| < \infty,$$

if g is non-zero for values which are sufficiently close to c.
A common abbreviation is

$$f_1(x) = f_2(x) + o(g(x)) :\Leftrightarrow f_1(x) - f_2(x) \in o(g(x))$$

$$f_1(x) = f_2(x) + \mathcal{O}(g(x)) :\Leftrightarrow f_1(x) - f_2(x) \in \mathcal{O}(g(x)) \; .$$

For instance

- $x \in \mathcal{O}(x^2)$ for $x \to \infty$
- $x^2 \in o(x)$ for $x \to 0$
- $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!} = 1 + x + \frac{x^2}{2} + \mathcal{O}(x^3)$ for $x \to 0$

CTMC: Definitions

Like a DTMC, a CTMC is characterised by discrete random variables. However, unlike a DTMC, time can vary continuously. Without loss of generality we shall use $\{t \in \mathbb{R} : t \geq 0\}$ as the interval of time in which the CTMC is defined.

A stochastic process $\{X(t)\}$ is said to be a CTMC if for all states and for any sequence $t_0 < t_1 < \ldots < t_n < t_{n+1}$,

$$\mathbb{P}\big(X(t_{n+1}) = x_{n+1} \mid X(t_n) = x_n, X(t_{n-1}) = x_{n-1}, \ldots, X(t_0) = x_0\big) \\ = \mathbb{P}\big(X(t_{n+1}) = x_{n+1} \mid X(t_n) = x_n\big).$$

Notice that this definition is similar to that of a DTMC, however here changes of state can happen at any moment in time.

CTMC: Definitions

As an equivalent definition, we say that the stochastic process $\{X(t)\}$ is a CTMC if, for any states i, j, k and for all time points s, t, u such that $t \geq 0$, $s \geq 0$, and $0 \leq u \leq s$, we have that

$$\mathbb{P}\big(X(s+t) = j \mid X(s) = i, X(u) = k\big) = \mathbb{P}\big(X(s+t) = j \mid X(s) = i\big)$$

Similarly to the DTMC case, we write

$$p_{ij}(s,t) := \mathbb{P}\big(X(t) = j \mid X(s) = i\big), \quad \text{for } t \geq s$$

which is the transition probability for a nonhomogeneous CTMC.
In a time-homogeneous CTMC, the transition probabilities do not depend on the specific time instants s and t, but only on the difference $\tau = t - s$,

$$p_{ij}(\tau) := \mathbb{P}\big(X(t+\tau) = j \mid X(t) = i\big), \quad \text{for } \tau \geq 0.$$

We must have that $\sum_j p_{ij}(\tau) = 1$ for any i and all τ.

Transition Rates

Usually the notion of transition rate is preferred to that of transition probability when dealing with CTMCs. Since the stochastic process may change state at any time instant, it makes sense to reason about the speed at which changes may occur.

Mathematically, this is captured by the following definition of transition rate:

$$q_{ij}(t) = \lim_{\Delta t \to 0} \frac{p_{ij}(t, t + \Delta t)}{\Delta t}, \quad \text{for } i \neq j,$$

from which we have

$$p_{ij}(t, t + \Delta t) = q_{ij}(t)\Delta t + o(\Delta t), \text{for } i \neq j.$$

Intuitively, this captures the fact that, as $\Delta t \to 0$, the probability of leaving state i goes to zero.

Transition Rates

Using

$$p_{ij}(t, t + \Delta t) = q_{ij}(t)\Delta t + o(\Delta t), \text{for } i \neq j$$

and the property of conservation of probability

$$1 = p_{ii}(t, t + \Delta t) + \sum_{i \neq j} p_{ij}(t, t + \Delta t),$$

we may write

$$p_{ii}(t, t + \Delta t) = 1 - \sum_{i \neq j} p_{ij}(t, t + \Delta t) = 1 - \sum_{i \neq j} q_{ij}(t)\Delta t + o(\Delta t).$$

Therefore

$$\lim_{\Delta t \to 0} \frac{1 - p_{ii}(t, t + \Delta t)}{\Delta t} = \sum_{i \neq j} q_{ij}(t).$$

We let $q_{ii}(t) = -\sum_{i \neq j} q_{ij}(t)$.

Generator Matrix

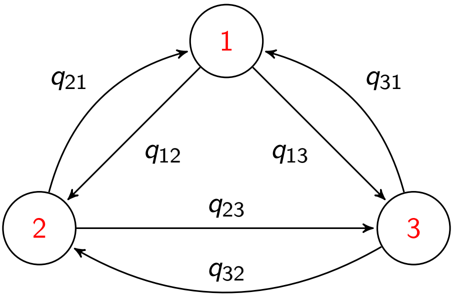
The matrix $Q(t)$ with ij-th element equal to $q_{ij}(t)$ is called the generator matrix of the CTMC.

In general, $Q(t)$ depends on time. However, if the CTMC is time homogeneous, then the following quantities are independent from t

$$\lim_{\Delta t \to 0} \frac{p_{ij}(t, t + \Delta t)}{\Delta t} = \lim_{\Delta t \to 0} \frac{p_{ij}(\Delta t)}{\Delta t} := q_{ij},$$
$$q_{ii} := - \sum_{i \neq j} q_{ij},$$

and the generator matrix is written Q.

Graphical Representation



$$Q = \begin{bmatrix} -(q_{12}+q_{13}) & q_{12} & q_{13} \\ q_{21} & -(q_{21}+q_{23}) & q_{23} \\ q_{31} & q_{32} & -(q_{31}+q_{32}) \end{bmatrix}$$

- No self loops
- The matrix is not stochastic (but the sum across rows equals zero)
- q_{ij} may be greater than zero!

CTMC and the Exponential Distribution

Given a non-absorbing state i, define T_i as the sojourn time in i. It can be shown that, for positive $s > 0$ and $t > 0$,

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

That is, the distribution of the residual time is equal to the distribution itself. But this property is satisfied iff T_i is an exponential distribution, therefore the sojourn time in a non-absorbing state of a CTMC is exponentially distributed.

CTMC and the Exponential Distribution

In a homogeneous CTMC:

- The distribution of the time until a transition from state i to j occurs is exponentially distributed with parameter q_{ij}.
- Since more states can be reached from a state i, a race condition occurs. The transition with the fastest time until occurrence is chosen. This means that the sojourn time is the minimum of a number of exponentially distributed random variables with rates q_{ij}.
- But we have seen that the minimum of exponentially distributed random variables is an exponentially distributed random variable. Therefore, the sojourn time in state i is exponentially distributed with parameter $\sum_{j \neq i} q_{ij} = -q_{ii}$.
- So, given that the CTMC is in state i the probability that the transition to some $\hat{j}$ is made is equal to

$$\frac{q_{i\hat{j}}}{\sum_{j \neq i} q_{ij}}$$

Chapman-Kolmogorov Equations

The continuous-time analogue of the Chapman-Kolmogorov equations seen in the discrete-time case are:

$$\mathbb{P}(X(t) = j \mid X(s) = i) = p_{ij}(s,t) = \sum_k p_{ik}(s,u)p_{kj}(u,t),$$

for any i and j and $s \leq u \leq t$.
For a homogeneous CTMC, they simplify as follows:

$$p_{ij}(t+\Delta t) = \sum_k p_{ik}(t)p_{kj}(\Delta t) = \sum_{k \neq j} p_{ik}(t)p_{kj}(\Delta t) + p_{ij}(t)p_{jj}(\Delta t).$$

Therefore,

$$\frac{p_{ij}(t+\Delta t) - p_{ij}(t)}{\Delta t} = \sum_{k \neq j} p_{ik}(t)\frac{p_{kj}(\Delta t)}{\Delta t} + p_{ij}(t)\frac{(p_{jj}(\Delta t) - 1)}{\Delta t}.$$

Taking the limit $\Delta t \to 0$,

$$\frac{dp_{ij}(t)}{dt} = \sum_{k \neq j} p_{ik}(t)q_{kj} + p_{ij}q_{jj} = \sum_{\text{all } k} p_{ik}(t)q_{kj}.$$

Chapman-Kolmogorov Equations

$$\frac{dp_{ij}(t)}{dt} = \sum_{k \neq j} p_{ik}(t) q_{kj} + p_{ij} q_{jj} = \sum_{\text{all } k} p_{ik}(t) q_{kj}.$$

In matrix notation,

$$\frac{dP(t)}{dt} = P(t)Q.$$

These equations are called the Kolmogorov forward equations.

The solution is given by:

$$P(t) = e^{Qt} = I + \sum_{i=1}^{\infty} \frac{Q^n t^n}{n!}.$$

Using $p_{ij}(t + \Delta t) = \sum_k p_{ik}(\Delta t) p_{kj}(t)$ instead of $p_{ij}(t + \Delta t) = \sum_k p_{ik}(t) p_{kj}(\Delta t)$ leads to the Kolmogorov backward equations:

$$\frac{dP(t)}{dt} = QP(t)$$

Transient Distribution

We wish to find $\pi_i(t) := \mathbb{P}\big(X(t) = i\big)$ for all t and all states i for a homogeneous CTMC.

$$\begin{aligned}
\pi_i(t+\Delta t) &= \mathbb{P}\big(X(t+\Delta t) = i\big) \\
&= \mathbb{P}\big(X(t+\Delta t) = i \mid X(t) = 0\big)\mathbb{P}\big(X(t) = 0\big) \\
&\qquad + \mathbb{P}\big(X(t+\Delta t) = i \mid X(t) = 1\big)\mathbb{P}\big(X(t) = 1\big) + \dots \\
&\qquad + \mathbb{P}\big(X(t+\Delta t) = i \mid X(t) = k\big)\mathbb{P}\big(X(t) = k\big) + \dots \\
&= p_{ii}(\Delta t)\pi_i(t) + \sum_{k \neq i} p_{ki}(\Delta t)\pi_k(t).
\end{aligned}$$

$$\pi_i(t+\Delta t) - \pi_i(t) = \big(p_{ii}(\Delta t) - 1\big)\pi_i(t) + \sum_{k \neq i} p_{ki}(\Delta t)\pi_k(t)$$

$$\frac{\pi_i(t+\Delta t) - \pi_i(t)}{\Delta t} = \frac{\big(p_{ii}(\Delta t) - 1\big)}{\Delta t}\pi_i(t) + \sum_{k \neq i} \frac{p_{ki}(\Delta t)}{\Delta t}\pi_k(t)$$

Transient Distribution

$$\frac{\pi_i(t+\Delta t) - \pi_i(t)}{\Delta t} = \frac{\big(p_{ii}(\Delta t) - 1\big)}{\Delta t}\pi_i(t) + \sum_{k \neq i} \frac{p_{ki}(\Delta t)}{\Delta t}\pi_k(t)$$

$$\frac{d\pi_i(t)}{dt} = q_{ii}\pi_i(t) + \sum_{k \neq i} q_{ki}\pi_k(t) = \sum_{k} q_{ki}\pi_k(t)$$

In matrix notation the transient distribution can be written as:

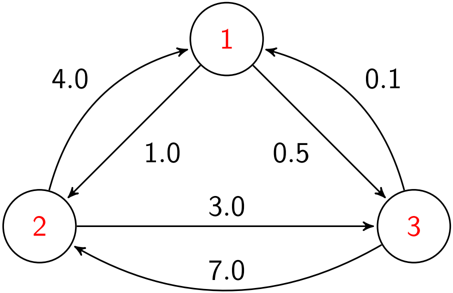
$$\frac{d\pi(t)}{dt} = \pi(t)Q.$$

This is a system of first-order coupled differential equations which has solution

$$\pi(t) = \pi(0)e^{Qt} = \pi(0)\left(I + \sum_{i=1}^{\infty} \frac{Q^i t^i}{i!}\right),$$

where $\pi(0)$ is some initial distribution for the CTMC.

Numerical Example



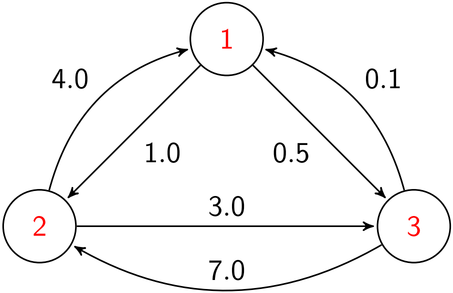
$$Q = \begin{bmatrix} -1.5 & 1.0 & 0.5 \\ 4.0 & -7.0 & 3.0 \\ 0.1 & 7.0 & -7.1 \end{bmatrix}$$

$$\pi(0) = \begin{bmatrix} 1/3 & 1/3 & 1/3 \end{bmatrix}$$

$$\pi(t) = \pi(0)e^{Qt}$$

$$\pi(0.1) = \begin{bmatrix} 0.4152 & 0.3347 & 0.2501 \end{bmatrix} \pi(0.2) = \begin{bmatrix} 0.4796 & 0.3119 & 0.2084 \end{bmatrix} \pi(0.$$

Numerical Example



$$Q = \begin{bmatrix} -1.5 & 1.0 & 0.5 \\ 4.0 & -7.0 & 3.0 \\ 0.1 & 7.0 & -7.1 \end{bmatrix}$$

$$\pi(0) = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$$

$$\pi(t) = \pi(0)e^{Qt}$$

$$\pi(0.1) = \begin{bmatrix} 0.8772 & 0.0794 & 0.0434 \end{bmatrix} \pi(0.2) = \begin{bmatrix} 0.7954 & 0.1301 & 0.0745 \end{bmatrix} \pi(0.5$$

Limiting Distribution

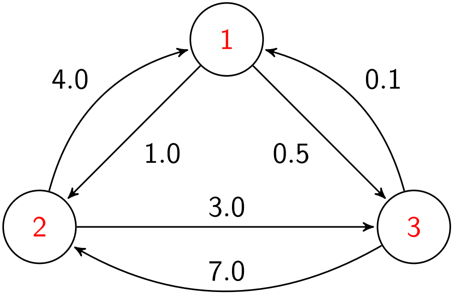
When the limit exists, and all its components are strictly positive, and when the limit is independent from the initial probability distribution $\pi(0)$ then the limit is unique and is called the *steady-state distribution* of the CTMC.

The chain reaches a condition in which the rate of change of the probability vector is zero, i.e.

$$\frac{d\pi(t)}{dt} = \pi Q = 0.$$

If the CTMC is finite and irreducible, the steady-state distribution exists.

Numerical Example

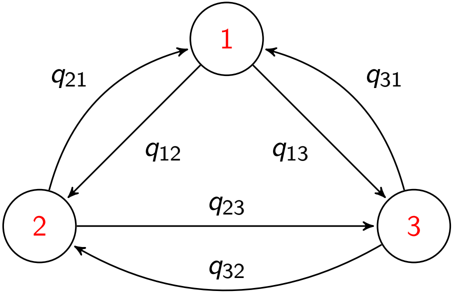


$$Q = \begin{bmatrix} -1.5 & 1.0 & 0.5 \\ 4.0 & -7.0 & 3.0 \\ 0.1 & 7.0 & -7.1 \end{bmatrix}$$

We observe that

$$\begin{bmatrix} 0.6266 & 0.2314 & 0.1419 \end{bmatrix} \cdot Q = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}$$

Global Balance Equations of a CTMC



$$Q = \begin{bmatrix} q_{11} & q_{12} & q_{13} \\ q_{21} & q_{22} & q_{23} \\ q_{31} & q_{32} & q_{33} \end{bmatrix}$$

$$\begin{bmatrix} \pi_1 & \pi_2 & \pi_3 \end{bmatrix} Q = 0 \Rightarrow \begin{cases} \pi_1 q_{11} + \pi_2 q_{21} + \pi_3 q_{31} = 0 \\ \pi_1 q_{12} + \pi_2 q_{22} + \pi_3 q_{32} = 0 \\ \pi_1 q_{13} + \pi_2 q_{23} + \pi_3 q_{33} = 0 \end{cases}$$

$$\Rightarrow \begin{cases} \pi_1 q_{12} + \pi_1 q_{13} = \pi_2 q_{21} + \pi_3 q_{31} \\ \pi_2 q_{21} + \pi_2 q_{23} = \pi_1 q_{12} + \pi_3 q_{32} \\ \pi_3 q_{31} + \pi_3 q_{32} = \pi_1 q_{13} + \pi_2 q_{23} \end{cases}$$

flux out = flux in!

Stationary Distribution

Analogously to a DTMC, a stationary distribution is nonzero vector π such that $\pi Q = 0, \sum_i \pi_i = 1$ and $\pi_i \geq 0$ for all i.

One of the resulting equations is dependent linearly on the others. If the replacement of an equation with the normalising condition makes the coefficient matrix non-singular, then the stationary distribution is unique.

A stationary distribution is unique if the CTMC is irreducible and finite. In this case, it is identical to the steady-state distribution of the Markov chain.

Solution Routine

For a finite, irreducible CTMC, in order to obtain the stationary distribution we need to solve the following system of equations:

$$\pi Q = 0, \qquad \text{with } \sum_i \pi_i = 1.$$

Given the singularity of Q, let $\hat{Q}$ be the coefficient matrix Q where one column is replaced with a vector of ones. The problem is equivalent to solving:

$$\pi \hat{Q} = \begin{bmatrix} 0 & 0 & \dots & 1 \end{bmatrix}$$

In practice, we (i.e., software tools) can solve systems in the form $Ax = b$, with x unknown, therefore the consider the transposed equations instead:

$$\hat{Q}^T \pi^T = \begin{bmatrix} 0 \\ 0 \\ \dots \\ 1 \end{bmatrix} \Rightarrow \pi^T = (\hat{Q}^T)^{-1} \begin{bmatrix} 0 \\ 0 \\ \dots \\ 1 \end{bmatrix}$$

Solution Routine

Essentially, the solution methods is the same for both DTMCs and CTMCs. The difference is that in a DTMC, the coefficient matrix is formed by replacing one column of $P - I$ with a vector of ones, whereas in CTMC we replace a column of Q.

Therefore, the numerical solvers available for linear systems of equations in the form $Ax = b$ can be used for both.

References

For more details on the numerical solution of Markov chains (and linear systems in general) see:

- W. J. Stewart. *Introduction to the Numerical Solution of Markov Chains*. Princeton University Press, 1994.
- R. Barrett, M. Berry, T. F. Chan, J. Demmel, J. Donato, J. Dongarra, V. Eijkhout, R. Pozo, C. Romine and H. Van der Vorst. *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods, 2nd Edition*. SIAM, 1994.
  Also available at `www.netlib.org/templates/templates.pdf`