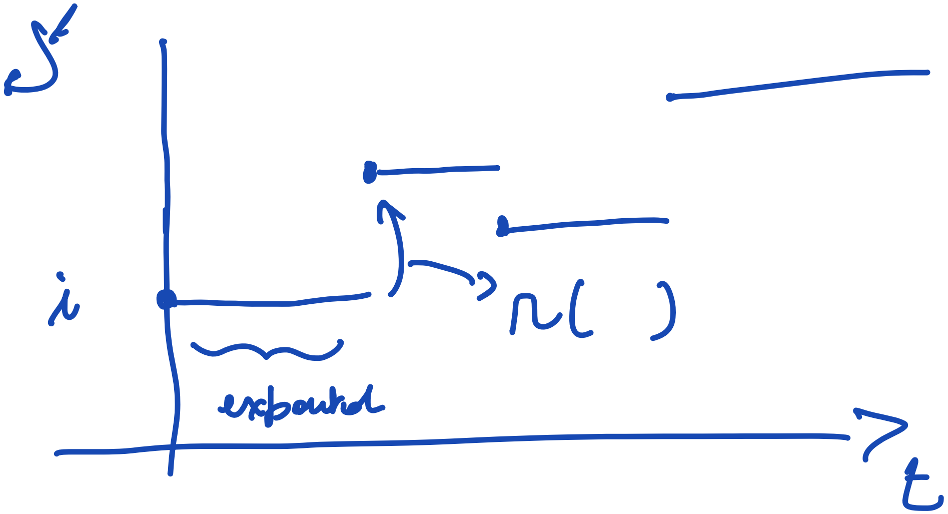
An introduction to stochastic vs. deterministic models of CRNs: 2. Intro to Continuous Time Markov Chains (CTMC)

Enrico Bibbona
enrico.bibbona@polito.it

POLITECNICO DI TORINO
Dipartimento di Scienze Matematiche G. L. Lagrange
ECCELLENZA 2018 · 2022

CTMC in short



Continuous time Markov chains incorporate the structure of a discrete time Markov chain in a continuous time Markov process. Basically, when thinking about a Continuous time Markov chain you should think of a process that "jumps" between different states (exactly as in a discrete time Markov chain), and, before jumping, remains in a given state x for a time τ_x, which is an exponential random variable whose rate depend on x.

A review of the Exponential distribution

We say that $T \sim \exp(\lambda)$ if any of the following holds:

$$f_T(t) = \begin{cases} \lambda e^{-\lambda t} & \text{if } t \geq 0 \\ 0 & \text{if } t < 0 \end{cases} \qquad P(T \leq t) = \begin{cases} 1 - e^{-\lambda t} & \text{if } t \geq 0 \\ 0 & \text{if } t < 0 \end{cases}$$

$$P(T > t) = \begin{cases} e^{-\lambda t} & \text{if } t \geq 0 \\ 1 & \text{if } t < 0 \end{cases}$$

Properties:

- $E[T] = 1/\lambda$, $\mathrm{Var}(T) = 1/\lambda^2$
- Memoryless property: for any $t > s \geq 0$ we have

$$P(T > t + s | T > s) = P(T > t)$$

.

Exponential races

Exponential races: Let $T_1, T_2, \ldots, T_n$ be independent random variables, with $T_i \sim \exp(\lambda_i)$. Let

$$V = \min\{T_1, T_2, \ldots, T_n\} \quad \text{and} \quad I \text{ s.t. } T_I = V.$$

Then:

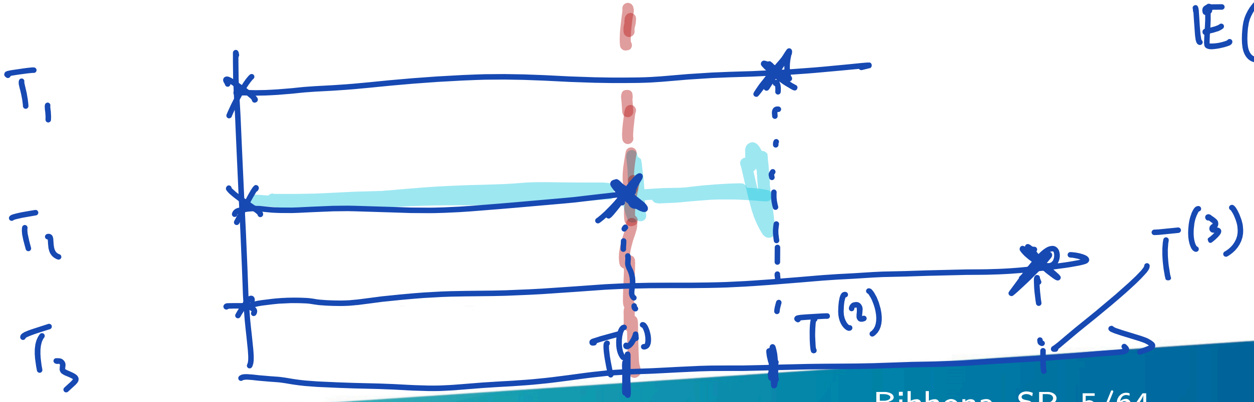
1. $V \sim \exp\left(\sum_{i=1}^n \lambda_i\right)$
2. $P(I = j) = \frac{\lambda_j}{\sum_{i=1}^n \lambda_i}$
3. I and V are independent.

Exponential races and forgetfulness. Exercise

Let T_i, $i \in \{1,2,3\}$, be exponentially distributed with parameters λ_i. Calculate the following quantities.

1. For every realization, sort the three exponentials. Denote by $T^{(1)}$ the minimum, $T^{(3)}$ the maximun and $T^{(2)}$ the intermediate one. Calculate $\mathbb{E}\left[T^{(2)}\right]$.
2. The probability that the second exponential also ranks second $(T^{(2)} = T_2)$

$$\mathbb{E}(T^{(2)}) = \mathbb{E}\left(T^{(1)}\right) + \mathbb{E}\left(T^{(2)} - T^{(1)}\right)$$



$$\mathbb{E}(T^{(1)}) = \frac{1}{\lambda_1+\lambda_2+\lambda_3}$$

$$\mathbb{E}(T^{(2)} - T^{(1)}) =$$ expected time of the winner of a new run when only $T^{(2)}$ and $T^{(3)}$ are participating.

$$\mathbb{E}(T^{(2)} - T^{(1)} \mid T^{(1)} = 1) = \frac{1}{\lambda_2+\lambda_3}$$

$$\mathbb{E}(T^{(2)} - T^{(1)} \mid T^{(1)} = 2) = \frac{1}{\lambda_1+\lambda_3}$$

$$\mathbb{E}(T^{(2)} - T^{(1)} \mid T^{(1)} = 3) = \frac{1}{\lambda_1+\lambda_2}$$

$$\mathbb{E}\left(T^{(2)}-T^{(1)}\right)=\sum_{i=1}^{3}\mathbb{E}\left(T^{(2)}-T^{(1)}\,\middle|\,T^{(1)}=i\right)\cdot\mathbb{P}\left(T^{(1)}=i\right)=$$

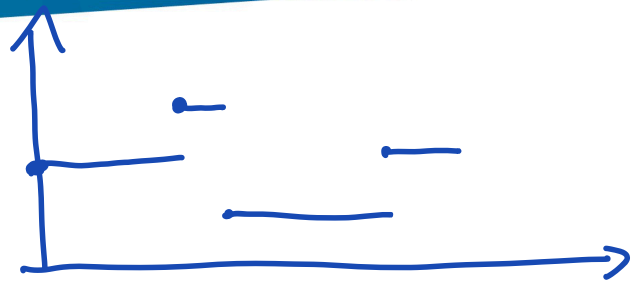
$*$

$$\mathbb{P}\left(T^{(1)}=i\right)=\frac{\lambda_i}{\sum_j \lambda_j}$$

$$*=\frac{1}{\lambda_2+\lambda_3}\cdot\frac{\lambda_1}{\lambda_1+\lambda_2+\lambda_3}+\frac{1}{\lambda_1+\lambda_3}\cdot\frac{\lambda_2}{\lambda_1+\lambda_2+\lambda_3}+$$

$$+\frac{1}{\lambda_1+\lambda_2}\cdot\frac{\lambda_3}{\lambda_1+\lambda_2+\lambda_3}$$

$$\mathbb{E}\left(T^{2}\right)=\frac{1}{\lambda_1+\lambda_2+\lambda_3}+$$



Construction 1

1. We let $(Y_n)_{n\geq 0}$ be a discrete time Markov chain on a state space S. We denote its transition probabilities by

$$r(i,j) = P(Y_{n+1} = j \mid Y_n = i),$$

with the condition that $r(i,i) = 0$ for all $i \in S$.

2. After arriving (or starting the process at time zero) at state $i \in S$, we let the amount of time we spend in state i, the *holding time*, be an exponential random variable with rate $\lambda(i)$. We denote this random variable by τ_i.
3. After the holding time τ_i we transition away from state i according to the probabilities associated with the discrete time Markov chain $(Y_n)_{n\geq 0}$.
4. Finally, we assume that all exponential random variables utilized are independent of each other and of the discrete time Markov chain $(Y_n)_{n\geq 0}$.

Construction 2

1. We suppose that the process has just arrived (or is starting) in state $i \in S$. For each $j \in S$ with $r(i,j) > 0$ we place an alarm clock on state j set to go off after an amount of time $\tau_{ij} \sim \text{Exp}(q(i,j))$ where

$$\underline{q(i,j)} = \lambda(i) \cdot r(i,j)$$

All exponential random variables are independent of each other and of all previous random variables.

2. When the first alarm goes off, we move to the state associated with that alarm. Formally, we do the following:
   - We let $\tau_i = \min_j\{\tau_{ij}\}$, and let $y \in S$ be the index of the minimum.
   - We then move to state y after a holding time equal to τ_i.

$q(i,k)$

i

$q(i,j)$

Equivalence

Theorem

The two constructions are equivalent, in the sense that the distribution of the process $(X_t)_{t\geq 0}$ associated with any of the two constructions is the same.

Proof.

This follows from the properties of independent exponential random variables. E.g. the minimum time $\tau_i = \min_j\{\tau_{ij}\}$ is exponential with parameter

$$\sum_{j\in S-\{i\}} q(i,j) = \lambda(i)\sum_{j\in S} r(i,j) = \lambda(i).$$

Moreover, the probability that the minimum is achieved at state y is precisely

$$\frac{q(i,y)}{\sum_j q(i,j)} = \frac{\lambda(i)\cdot r(i,y)}{\lambda(i)} = r(i,y),$$

The time spent in one state and which state is next are independent r. v. $\square$

CTMC

$\lambda(i)$ HOLDING RATES

$R(i,j)$ JUMP CHAIN

$$\sum_j Q(i,j) = \sum_j \lambda(i) R(i,j) = \lambda(i) \sum_j R(i,j) = \lambda(i)$$

$$q(i,j) = \lambda(i) \cdot R(i,j)$$

$$\downarrow$$

$$Q$$

$$P(X_t = j \mid X_0 = i) = P_t(i,j)$$

$$P_t = \begin{pmatrix} \cdot \end{pmatrix}$$

$$\boxed{Q = P_t'(0)}$$

Definitions

ON THE DIAGONAL I PUT − THE SUM OF THE ROW ⇒ if i take the sum of each row of Q it gives 0

$$Q \begin{pmatrix} -\lambda_1 & & & q_{ij} \\ & -\lambda_2 & & \\ & & \ddots & \\ & & & -\lambda_n \end{pmatrix}$$

The quantities introduced introduce so far are very important and they have a name: $q(i,j)$ are called the transition rates of the process.

It was helpful to put the transition probabilities of a discrete time Markov chain into matrix form. We do the same here. We define the matrix Q to have entries

$$Q(i,j) = \begin{cases} q(i,j) & i \neq j \\ -\sum_{\ell \neq i} q(i,\ell) & i = j \end{cases}$$

for all i, j in S. Note that the diagonal is the *negative* of the sum of the other terms in its row (this choice will turn out to be clever later on). This matrix is also called *transition rate matrix* or *generator*.

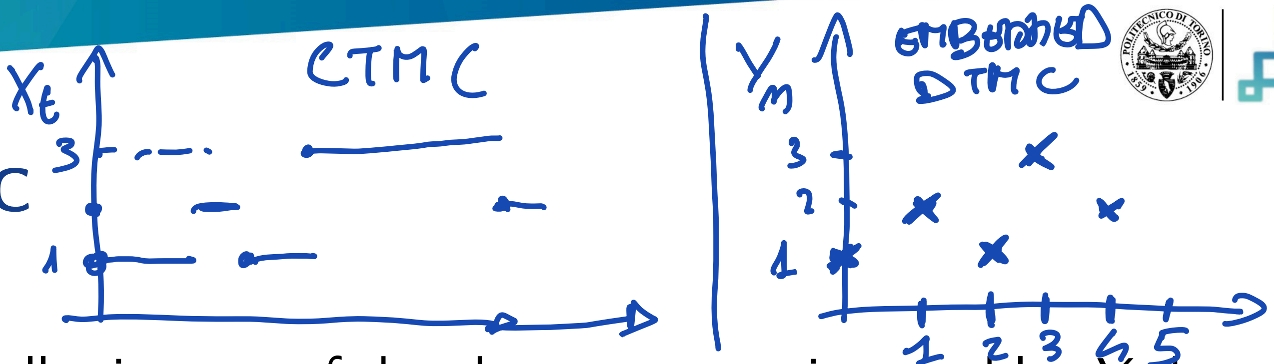
Remarks

1. The values $q(i,j)$ are *not* probabilities. Specifically, they can take values larger than 1. Still they are non-negative
2. The row sums of Q are zero.
3. Such matrices (row sums equal to zero, non-negative off-diagonal entries) are sometimes termed *generator matrices* in Linear Algebra.
4. Why did we choose to put the q's in the matrix and not the values $r(i,j)$? The reason is that from the q's we can recover the values $r(i,j)$ and the values $\lambda(i)$. Specifically,

$$\lambda(i) = \sum_{j \neq i} q(i,j) \qquad r(i,j) = \frac{q(i,j)}{\lambda(i)} = \frac{q(i,j)}{\sum_{\ell \neq i} q(i,\ell)}.$$

From the matrix Q we derive all the quantities of interest of our process.

The embedded DTMC

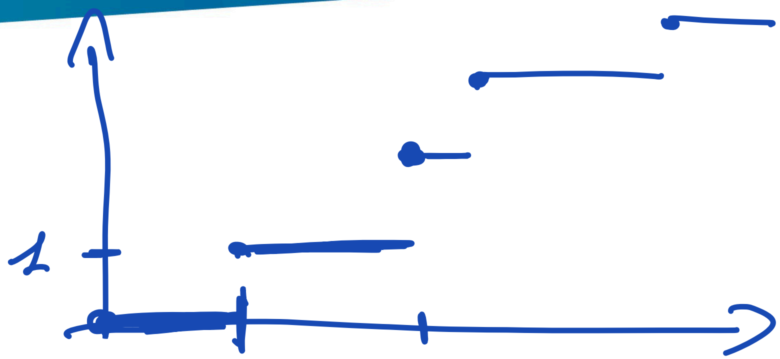


Let $(X_t)_{t\geq 0}$ be a process following one of the above constructions and let Y_n be the discrete time Markov chain giving the (ordered) sequence of states visited by $(X_t)_{t\geq 0}$ (I.e., the DTMC given in Construction 1.) Y_n is called *the embedded discrete time Markov chain* associated with X_t. Moreover, the transition probabilities of Y_n are given by the values $r(i,j)$. That is

$$\begin{aligned} P(Y_{n+1} = j \mid Y_n = i) &= P(\text{next state visited by } X \text{ is } j \mid X \text{ is currently in state } i) \\ &= r(i,j). \end{aligned}$$

Note that the embedded discrete time Markov chain keeps track of the "changes of state" of the continuous time Markov chain. Therefore it never stays put in the same state: that is why transition probabilities $r(i,i)$ are set to 0 for every state i.

Examples: Poisson process



Let $N(t)$ be a rate λ Poisson process. Note that this process follows the above construction with

$$r(i, i+1) = 1, \text{ and } \tau_i \sim \text{Exp}(\lambda).$$

The Q matrix is below as well as the transition matrix P for the embedded DTMC

$$q(i,j) = \lambda(i) \cdot R(i,j) = \lambda\, \delta_{j,i+1} \quad \forall j \neq i$$

$$Q = \begin{pmatrix} -\lambda & \lambda & 0 & 0 & \cdots \\ 0 & -\lambda & \lambda & 0 & \cdots \\ 0 & 0 & -\lambda & \lambda & \cdots \\ \vdots & \vdots & & \ddots & \ddots \end{pmatrix} \qquad R = \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots \\ 0 & 0 & 1 & 0 & \cdots \\ 0 & 0 & 0 & 1 & \cdots \\ \vdots & \vdots & \vdots & & \ddots \end{pmatrix}.$$

Note that we could have $\lambda > 1$.

Examples: ABC

$q(i,j) = \lambda(i) \cdot r(i,j)$

Consider the 3 state model with $S = \{A, B, C\}$

$\lambda(i) = \sum_j q(i,j)$

$$Q = \begin{pmatrix} -3 & 1 & 2 \\ 0 & -4 & 4 \\ 2 & 6 & -8 \end{pmatrix}.$$

$\lambda(A) = 1 + 2 = 3$

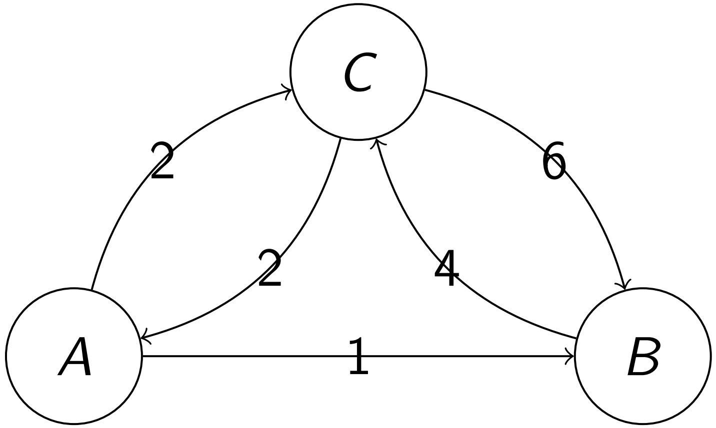
?

$R = \begin{pmatrix} 0 & 1/3 & 2/3 \\ 0 & 0 & 1 \\ 1/4 & 3/4 & 0 \end{pmatrix}$

We can draw the situation by means of a *transition graph*:

$\mathbb{E}\left(\begin{smallmatrix}\text{time} \\ \text{spent} \\ \text{in } A\end{smallmatrix}\right) = \frac{1}{3}$

$r(A,B) = \frac{q(A,B)}{\lambda(A)} = \frac{1}{3}$



$q(i,j) = \lambda(i)\, r(i,j)$

$\lambda(B) = 4$

where the numbers on top of the arrows describe the transition rates.

$r(B,A) = \frac{Q(B,A)}{\lambda(B)} = 0$

$r(i,j) = \frac{q(i,j)}{\lambda(i)} = \frac{q(i,j)}{\sum_{k \neq i} q(i,k)}$

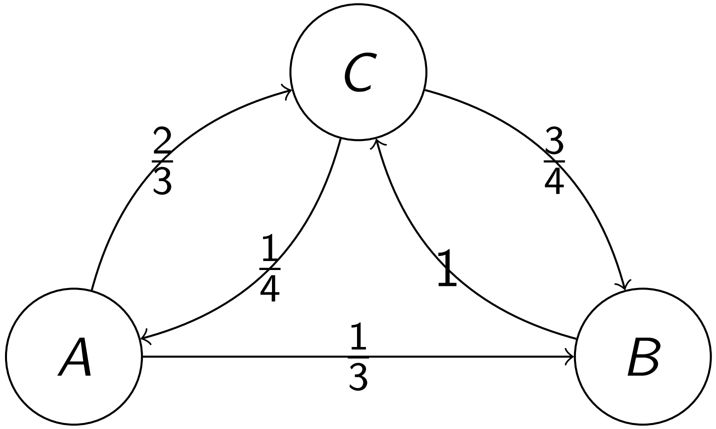
$\lambda(C) = q(C,B) + q(C,A) = 8 \qquad r(C,A) = \dfrac{q(C,A)}{\lambda(C)} = \dfrac{2}{8} = \dfrac{1}{4}$

Examples: ABC

The embedded chain has the same state space and transition matrix

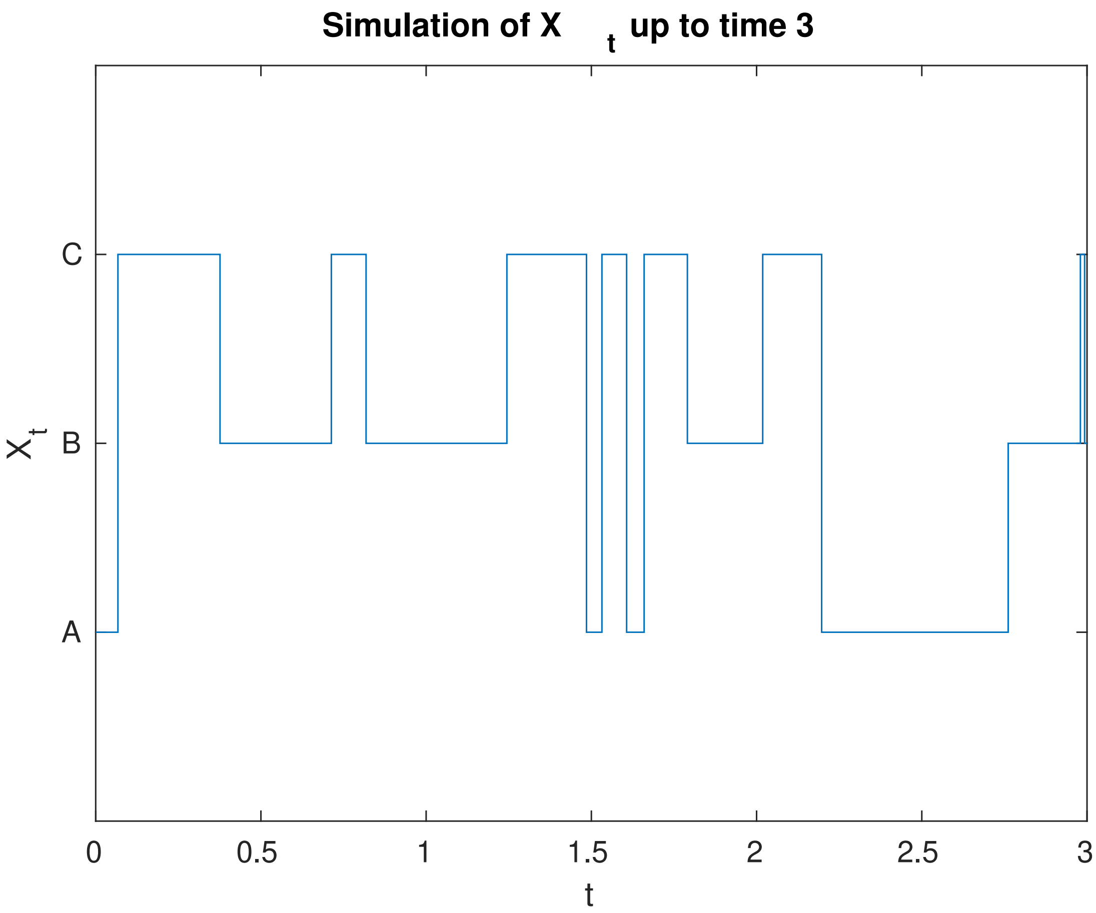
$$R = \begin{pmatrix} 0 & \frac{1}{3} & \frac{2}{3} \\ 0 & 0 & 1 \\ \frac{1}{4} & \frac{3}{4} & 0 \end{pmatrix}$$

For comparison, the transition graph of the embedded DTMS:

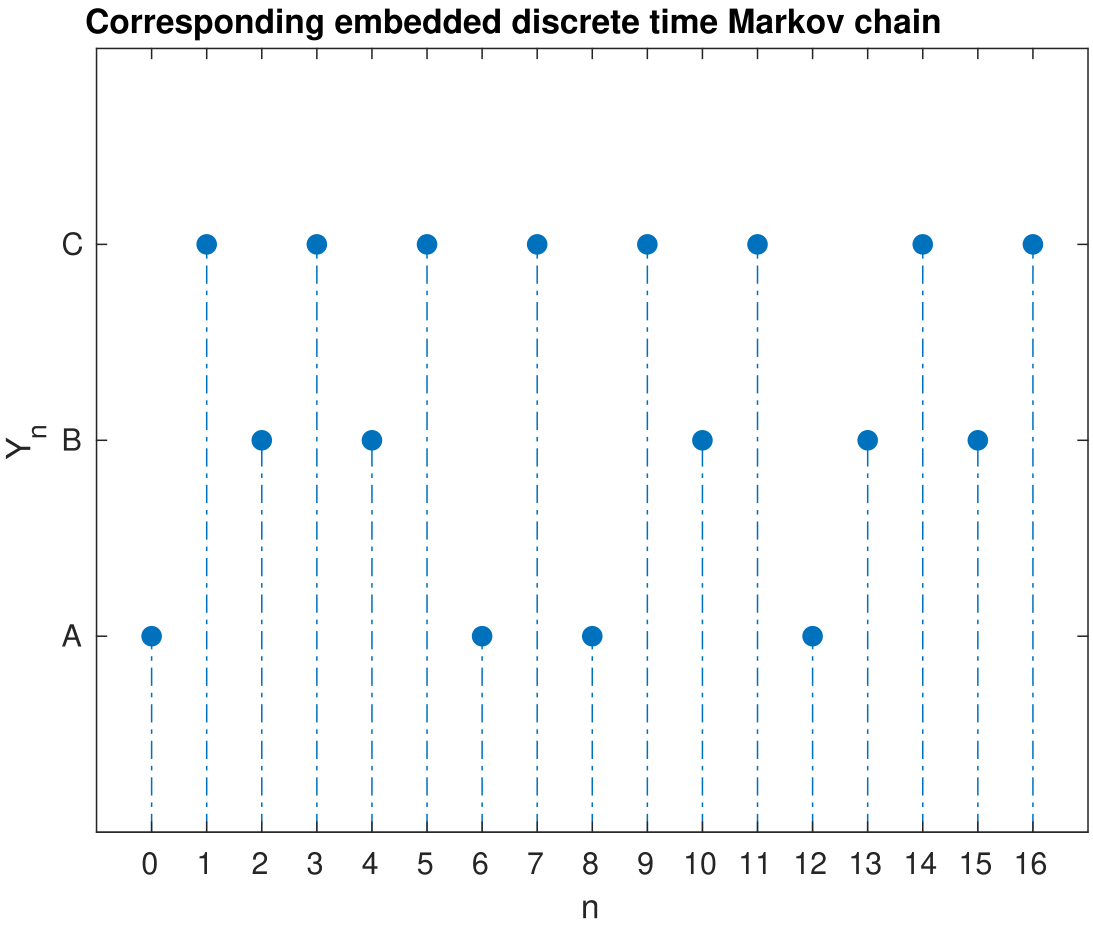


The holding times in states A, B, and C are exponentially distributed with parameters $3, 4$ and 8, respectively.

Examples: ABC



Examples: ABC



M/M/s queue

In queueing theory, the notation M/M/s means:

- the arrivals are Markovian, meaning that customers arrive according to a Poisson process: the time between two arrivals is $\sim \text{Exp}(\lambda)$
- the service times are Markovian, meaning that they are independent and exponentially distributed with a common rate μ;
- there are s servers that works in parallel.

To make it more concrete you can think at a bank with s teller stations.
Try to guess the generator.

M/M/s queue

In queueing theory, the notation M/M/s means:

- the arrivals are Markovian, meaning that customers arrive according to a Poisson process: the time between two arrivals is $\sim \text{Exp}(\lambda)$
- the service times are Markovian, meaning that they are independent and exponentially distributed with a common rate μ;
- there are s servers that works in parallel.

To make it more concrete you can think at a bank with s teller stations.
Try to guess the generator.
The answer is

$$q(n, n-1) = \begin{cases} 0 & \text{if } n = 0 \\ n\mu & \text{if } 1 \leq n \leq s \\ s\mu & \text{if } n \geq s \end{cases} \qquad q(n, n+1) = \lambda$$

The Markov property

We say that a continuous process $(X_t)_{t\in[0,\infty)}$ has the Markov property if for **any** $s, t > 0$ and **any**

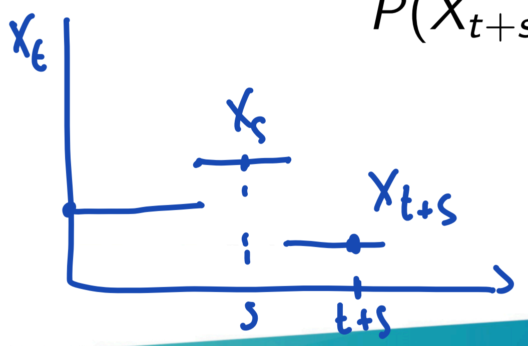
$$0 \leq s_0 < s_1 < \cdots < s_n < s$$

and states $i_0, i_1, \ldots, i_n, i, j$, we have

$$P(X_{t+s} = j \mid X_s = i, X_{s_n} = i_n, \ldots, X_{s_0} = i_0) = P(X_{t+s} = j \mid X_s = i).$$

Moreover, as in the discrete time case, we will only consider time homogeneous cases, that is we always assume that

$$P(X_{t+s} = j \mid X_s = i) = P(X_t = j \mid X_0 = i_0).$$



DTMCs enjoy the Markov property

The idea is the following.

- If you know that $X_s = i$, then the residual holding time at i from time s on will still be $\sim \text{Exp}(\lambda(i))$ by the memoryless property.
- Then by construction the process will jump with a target state that is independent from the holding time and reiterate this behavior, following the original construction.

Therefore, after time s the process will start anew from state i with the same law that it would have had if i had been the initial state, irrespectively of all the history of the process before time s.

Markov Prop + Discrete state space $\implies$ CTMC

the proof mainly boils down to prove that

- the holding times of $(X_t)_{t\in[0,\infty)}$ must be exponentially distributed
- the jumps directions are independent on the holding times

so $(X_t)_{t\in[0,\infty)}$ follows Construction ??).
Now, we suppose $X_0 = i$ and let τ_i denote the time we transition away from state i. Our goal is to prove that τ_i has the memoryless property, and is therefore exponentially distributed.
This follows from the fact that if a continuous random variable has the memoryless property, then it is exponentially distributed.

Markov Prop + Discrete state space $\implies$ DTMC

Let $s, t \geq 0$ and τ_i be the holding time at i

$$
\begin{aligned}
&P_i(\tau_i > s + t \mid \tau_i > s) \\
&\quad = P(X_r = i \text{ for } r \in [0, s+t] \mid X_r = i \text{ for } r \in [0, s]) \\
&\quad = P(X_r = i \text{ for } r \in [s, s+t] \mid X_r = i \text{ for } r \in [0, s]) \quad (P(A \cap B \mid A) = P(B|A)) \\
&\quad = P(X_r = i \text{ for } r \in [s, s+t] \mid X_s = i) \qquad \text{(Markov property)} \\
&\quad = P(X_r = i \text{ for } r \in [0, t] \mid X_0 = i) \qquad \text{(time homogeneity)} \\
&\quad = P_i(\tau_i > t).
\end{aligned}
$$

Finally, the jump direction of $(X_t)_{t \in [0,\infty)}$ cannot depend on the past, hence it cannot depend on the holding time in one state nor on the directions of the previous jumps: they have to follow a DTMC (the embedded DTMC).

in DISCR TIME $p^{(2)}(i,j) = \sum_k p(i,k)\cdot p(k,j)$

Transition probabilities

In the context of discrete time Markov chains, we defined the *n*-step transition probabilities $p^{(n)}(i,j)$. Similarly, we define the *transition probability* for a continuous time Markov chain as

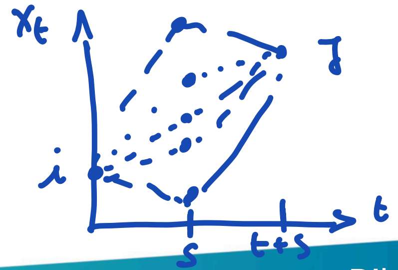
$$p_t(i,j) = P(X_t = j | X_0 = i).$$

We still have the analogue of the DT Chapman-Kolmogorov equation

Theorem (Chapman-Kolmogorov)

Consider a continuous time Markov chain $(X_t)_{t\geq 0}$. Then, for any real $s, t > 0$

$$p_{s+t}(i,j) = \sum_k p_s(i,k)p_t(k,j).$$



$\Rightarrow P_{s+t} = P_s \cdot P_t = P_t \cdot P_s$

Transition probabilities

The structure of the proof is exactly the same as in the discrete time case. We have

$$
\begin{aligned}
P(X_{s+t} = j | X_0 = i) &= \sum_k P(X_{s+t} = j | X_s = k, X_0 = i) P(X_s = k | X_0 = i) \\
&= \sum_k P(X_{s+t} = j | X_s = k) P(X_s = k | X_0 = i) \qquad \text{(by Markov pr.)} \\
&= \sum_k P(X_t = j | X_0 = k) P(X_s = k | X_0 = i) \quad \text{(by homogeneity)} \\
&= \sum_k p_t(k, j) p_s(i, k).
\end{aligned}
$$

From transition probabilities to transition rates

Let τ_i be the holding time in state i. We know that $\tau_i \sim \text{Exp}(\lambda(i))$. Consider a small interval of time $[0, h]$. Given that $X_0 = i$, the probability that a jump occurs in $[0, h]$ is

$$P_i(\tau_i < h) = 1 - e^{-\lambda(i)h} = 1 - \left(1 - \lambda(i)h + \frac{(\lambda(i)h)^2}{2} + \dots\right) = \lambda(i)h + O(h^2),$$

From the above calculation, in particular we have

$$\lim_{h \to 0} \frac{P_i(\tau_i < h)}{h} = \lambda(i) + \lim_{h \to 0} O(h) = \lambda(i).$$

From transition probabilities to transition rates

The probability of two or more jumps in $[0, h]$ is $O(h^2)$. Here is why: let Y_0 and Y_1 denote the initial and the second visited states of the chain, respectively:

$$\begin{aligned} P(2 \text{ or more jumps in}[0,h] | Y_0 = i, Y_1 = j) &= P(\tau_i + \tau_j < h) \\ &\leq P(\tau_i < h, \tau_j < h) \\ &= P(\tau_i < h)P(\tau_j < h) \quad \text{(by independence)} \\ &= \lambda(i)\lambda(j)h^2 + O(h^3) = O(h^2) \end{aligned}$$

$$P(2 \text{ or more jumps in}[0,h] | Y_0 = i) =$$

$$= \sum_j P(2 \text{ or more jumps in}[0,h] | Y_0 = i, Y_1 = j) \cdot P(Y_1 = j | Y_0 = i)$$

$$= \sum_j O(h^2) P(Y_1 = j | Y_0 = i) = O(h^2)$$

The last equality holds due to the dominated convergence theorem.

How many jumps in $[0, h]$ with h small

Take home message:

1. the probability of one jump in one interval $[0, h]$ goes to zero *as* h
2. he probability of two or more jumps in $[0, h]$ goes to zero *as* h^2.

From transition probabilities to transition rates

$P_t'(\cdot)\Big|_{t=0} = Q$

The transition rates are basically the *derivatives* of the transition probabilities calculated at 0: for $i \neq j$ we have

$$\frac{dp_t(i,j)}{dt}(0) = \lim_{h \to 0} \frac{p_h(i,j) - p_0(i,j)}{h} = \lim_{h \to 0} \frac{p_h(i,j)}{h} = q(i,j).$$

$p_h(i,j) \overset{h \to 0}{\approx} q(i,j) \cdot h$

Moreover, if $i = j$ we have

$$\frac{dp_t(i,i)}{dt}(0) = \lim_{h \to 0} \frac{p_h(i,i) - p_0(i,i)}{h} = \lim_{h \to 0} \frac{\left(1 - \sum_{j \neq i} p_h(i,j)\right) - 1}{h}$$

$$= \lim_{h \to 0} \frac{-\sum_{j \neq i} p_h(i,j)}{h} = -\sum_{j \neq i} q(i,j) = Q(i,i) = -\lambda(i)$$

From transition probabilities to transition rates

Moreover, for any $i, j \in S$ with $i \neq j$ we have

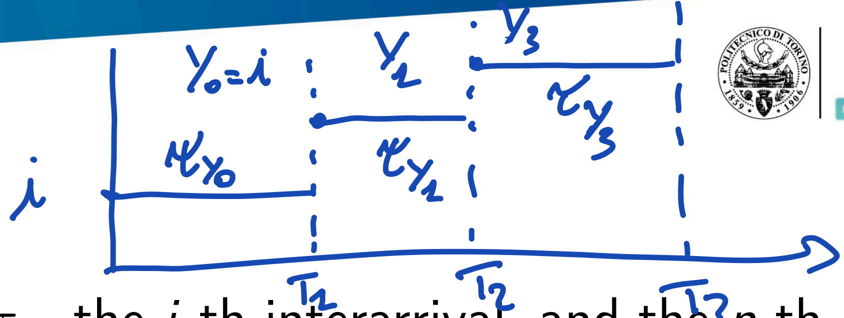
$$\lim_{h \to 0} \frac{p_h(i,j)}{h} = q(i,j).$$

Indeed

$$\begin{aligned}
\lim_{h \to 0} \frac{p_h(i,j)}{h} &= \lim_{h \to 0} \frac{P(X_h = j | Y_0 = X_0 = i)}{h} \\
&= \lim_{h \to 0} \frac{P(X_h = j, \text{only one jump in } [0,h] | Y_0 = X_0 = i)}{h} \\
&\quad + \frac{P(X_h = j, \text{two or more jumps in } [0,h] | Y_0 = X_0 = i)}{h} \\
&= \lim_{h \to 0} \left(\frac{P(\tau_i < h, Y_1 = j, \tau_j > h - \tau_i | Y_0 = X_0 = i)}{h} + \frac{O(h^2)}{h} \right) \\
\text{independence} \rightarrow &= \lim_{h \to 0} \left(\frac{P(\tau_i < h)}{h} P(Y_1 = j | Y_0 = i) P(\tau_j > h - \tau_i | \tau_i < h) + \frac{O(h^2)}{h} \right) \\
&= \lambda(i) \cdot r(i,j) \cdot 1 = q(i,j).
\end{aligned}$$

Explosions

Y_i is the state visited by X_t at its i-th jump



Costruction 1 has a problem: we denote by τ_{Y_i} the i-th interarrival, and the n-th arrival time $T_n = \tau_{Y_1} + \tau_{Y_2} + \cdots + \tau_{Y_n}$. If

$$T_n \to \infty, \text{ as } n \to \infty,$$

then the process is defined for all $t \geq 0$. However, what happens if the limit is finite? Then the process is defined only up to

$$T_\infty = \lim_{n\to\infty} T_n = \sum_{i=1}^{\overset{n}{\cancel{\infty}}} \tau_{Y_n}.$$

This is called an **explosion**. In this case the process has infinitely many jumps in a finite interval. These pathological behavior creates several technical troubles and we assume that all MC we are dealing with are non explosive. Giving sufficient conditions for non-explosiveness is sometimes a challenging problem, though.

Kolmogorov's equations (from trans. rates to trans. probs.)

If S is finite, the transition probabilities $p_t(i,j)$ form a true matrix that we denote P_t. Remember that $Q = P'_0$ and that by the Ch-K eqns. $\boxed{P_{t+h} = P_t P_h = P_h P_t}$ Therefore

$P_{t+h} = P_t \cdot P_h = P_h \cdot P_t$

$$P'(t) = \lim_{h\to 0} \frac{P_t P_h - P_t}{h} = P(t) \cdot \lim_{h\to 0} \frac{P_h - I}{h} = P_t Q \quad \Longrightarrow \quad P'(t) = P_t Q$$

$$P'(t) = \lim_{h\to 0} \frac{P_h P_t - P_t}{h} = \cdot \lim_{h\to 0} \frac{P_h - I}{h} P_t = Q P_t \quad \Longrightarrow \quad P'(t) = Q P_t$$

that are called *forward* and *backword* Kolmogorov eqns. In components they read

$$\begin{cases} p'_t(i,j) = \sum_{k\neq j} p_t(i,k) q(k,j) - p_t(i,j)\lambda(j) \\ p'_t(i,j) = \sum_{k\neq i} q(i,k) p_t(k,j) - \lambda(i) p_t(i,j) \end{cases}$$

Kolmogorov's equations (from trans. rates to trans. probs.)

- If S is infinite, but the chain is non explosive the equations in components are still valid and the series there appearing are convergent.
- In the explosive case, there might be troubles

Explosive chains are treated in details in Chung, *Markov Chains with Stationary Transition Probabilities*, Springer 1967.

Kolmogorov's equations (from trans. rates to trans. probs.)

We focus on Kolmogorov's backward eqn: $P_t' = QP_t$. In dimension 1 it reads $p'(t) = qp(t)$ with q a constant. We have the solution $p(t) = p(0)e^{qt}$. Something similar happens in higher dimension: we can define the *exponential matrix*

$$e^{Qt} = \sum_{k=0}^{\infty} t^k \frac{Q^k}{k!},$$

where $Q^0 = I$. We have that $e^{Q0} = I = P_0$.

Theorem

Let $(X_t)_{t=0}^{\infty}$ be a non-explosive CTMC. Then, for any $t \geq 0$

$$P_t = e^{Qt}.$$

Techniques are available to calculate e^{Qt} as a function of t, most symbolic computation software can do that.

Absorbing states

we excluded $r(i,i) \neq 0$

but we have to restore it only for absorbing states

A problem with Construction 1 is that in the presence of absorbing states we should redefine the embedded CTMC. Indeed after the system reaches an absorbing state, there is no *next state*. To be consistent we set the following agreement

- if i is absorbing, then $q(i,j) = 0$ for any j, meaning that to exit i (towards any state j) we need to wait a degenerate exponential time with an infinite mean.
- $\lambda(i) = \sum_{j \neq i} q(i,j) = 0$ and the whole raw of the generator Q corresponding to the absorbing state i has zero entries
- an absorbing state i in a DTMC is such that the transition matrix $\boxed{p(i,j) = \delta_{i,j}}$. Analogously for the embedded DTMC, we set $\boxed{r(i,j) = \delta_{i,j}}$ for all absorbing states i. Therefore for absorbing states (and only for them) the transition matrix of the embedded DTMC has diagonal entries equal to 1 (and therefore non-vanishing)!

$r(i,i) = 1$

Stationary distributions and measures

Let $(X_t)_{t=0}^{\infty}$ be a CTMC. We say that π is a stationary distribution if it is a probability distribution, and if for all $j \in S$ and all $t \geq 0$.

$$P(X_t = j \mid X_0 \sim \pi) = \sum_i \pi(i) p_t(i,j) = \pi(j)$$

Theorem

Let $(X_t)_{t=0}^{\infty}$ be a ***non-explosive*** *CTMC. Then, π is a stationary distribution if and only if $\pi Q = 0$ (and π sums up to 1).*

Proof:

$\Rightarrow$ Since π is stationary, $\pi = \pi P_t$. Taking derivative we have $\pi P_t' = 0$ and applying the forward Kolmogorov eqn. we conclude $\pi P_t' = \pi P_t Q = \pi Q = 0$. We skipped the necessary technical details requires if S is infinite, but the result still holds.

$\Leftarrow$ we have $\pi P_t' = \pi Q P_t$ by the backward eqn. If $\pi Q = 0$, then $\pi P_t' = 0$. Therefore πP_t is constant and $\pi P_t = \pi P_0 = \pi$

Finite irreducible chains

The CTMC $(X_t)_{t\geq 0}$ is **irreducible** if so is the embedded DTMC. That is, if for any two states i, j, we can get from i to j in a finite number of steps.

For a finite irreducible chain we can find a stationary distribution if we can solve the linear system

$$\begin{cases} \pi Q = 0 \\ \pi \cdot e = 0 \end{cases}$$

Where e is a vector whose length is the same as that of π (the cardinality of S) and all entries are 1.

Example

$$(\pi_1, \pi_2 \pi_3)\begin{pmatrix} -1/3 & 1/3 & 0 \\ 0 & -1/4 & 1/4 \\ 1 & 0 & -1 \end{pmatrix} = (0\ 0\ 0)$$

Then, assume that the transition rate matrix is

$$Q = \begin{bmatrix} -1/3 & 1/3 & 0 \\ 0 & -1/4 & 1/4 \\ 1 & 0 & -1 \end{bmatrix}$$

$$\begin{cases} -\frac{1}{3}\pi_1 + \pi_3 = 0 \\ \frac{1}{3}\pi_1 - \frac{1}{4}\pi_2 = 0 \\ \frac{1}{4}\pi_2 - \pi_3 = 0 \\ \pi_1 + \pi_2 + \pi_3 = 1 \end{cases}$$

Then, we can find the stationary distribution by solving $\pi Q = 0$, with the additional constraint that π sums to 1. No matter the strategy we use, in this case we obtain

$$\pi = \left(\frac{3}{8}, \frac{1}{2}, \frac{1}{8}\right).$$

Stationary measures CTMS vs. embedded DTMC

Cosider a two state CTMC with transition rate matrix

$$Q = \begin{pmatrix} -\alpha & \alpha \\ \beta & -\beta \end{pmatrix}$$

with $\alpha, \beta \geq 0$ and $\alpha + \beta > 0$. Then, $\pi = \left(\frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta}\right)$ is a stationary distribution of the continuous time Markov chain. The corresponding embedded DTMC has transition matrix

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

and its stationary distribution is $\tilde{\pi} = \left(\frac{1}{2}, \frac{1}{2}\right)$. Certainly the two differ. Is there any connections in general?

Stationary measures CTMS vs. embedded DTMC

Theorem

Consider a continuous time Markov chain with no absorbing states. Then μ is a stationary measure of the embedded discrete time Markov chain if and only if γ, defined entry-wise by

$$\gamma(j) = \frac{\mu(j)}{\lambda(j)},$$

(where $\lambda(j)$ is as usual the rate of the holding time in j), satisfies $\gamma Q = 0$.

Proof

Since no absorbing states, $r(j,j) = 0$ for all j. μ is a stationary measure of the embedded chain iff $\mu P = \mu$, iff (since $\mu(j) = \gamma(j)\lambda(j)$),

$$\sum_{i \neq j} \gamma(i)\lambda(i)r(i,j) = \gamma(j)\lambda(j)$$

for all $j \in S$. Using that $r(i,j) = q(i,j)/\lambda(i)$ we have that stationarity of μ is equivalent to

$$\sum_{i \neq j} \gamma(i)q(i,j) = \gamma(j)\lambda(j)$$

for all $j \in S$ and, the latter can be rewritten as

$$\left(\sum_{i \neq j} \gamma(i)Q(i,j) \right) + \gamma(j)Q(j,j) = 0$$

and it holds for any $j \in S$ if and only if $\gamma Q = 0$, concluding the proof.

Stationary measures CTMS vs. embedded DTMC

Why is the previous theorem important?

If the continuous time Markov chain is non-explosive and $\sum \gamma(i)$ is finite, then we can normalize γ to get a stationary distribution! We can also go in the opposite direction and calculate the stationary distributions (if any!) of the embedded discrete time Markov chain from those of the continuous time Markov chain.

Example

Consider again the 2-state CTMC with transition rate matrix

$$Q = \begin{pmatrix} -\alpha & \alpha \\ \beta & -\beta \end{pmatrix}$$

The stationary distribution of the embedded DTMC is $\tilde{\pi} = \left(\frac{1}{2}, \frac{1}{2}\right)$. How can we find the stationary distr. of the CTMC, without solving $\pi Q = 0$? Define

$$\gamma = \left(\frac{\tilde{\pi}(1)}{\lambda(1)}, \frac{\tilde{\pi}(2)}{\lambda(2)}\right) = \left(\frac{1}{2\alpha}, \frac{1}{2\beta}\right).$$

We know that $\gamma Q = 0$. However, γ is not a distribution. Define $M = \gamma(1) + \gamma(2) = \frac{\alpha+\beta}{2\alpha\beta}$ and let $\pi = \frac{1}{M}\gamma = \left(\frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta}\right)$. Then, π is a stationary distribution for the CTMC since $\pi Q = 0$ and $\pi(1) + \pi(2) = 1$ and the chain is not explosive.

Stationary measures of finite CTMS vs. embedded DTMC

Corollary

An irreducible CTMC with finitely many states has a ***unique*** *stationary distribution* π*. Moreover, for every state* j *we have* $\pi(j) > 0$*.*

The embedded DTMC is irreducible and has finitely many states. Hence, it has a *unique* stationary distribution $\tilde{\pi}$. Moreover, $\tilde{\pi}(j) > 0$ for any j. In order to prove the statement, we first show that π exists and then that is it unique:

Stationary measures CTMS vs. embedded DTMC

Existence: consider the vector γ defined by $\gamma(j) = \frac{\tilde{\pi}(j)}{\lambda(j)}$ for all j. Then we know that $\gamma Q = 0$. Moreover, $\sum_j \gamma(j) = M < \infty$ because the state space is finite. Hence, $\pi = \frac{1}{M}\gamma$ satisfies

$$\pi Q = \frac{1}{M}\gamma Q = 0 \quad \text{and} \quad \sum_j \pi(j) = \frac{1}{M}\sum_j \gamma(j) = 1.$$

Hence, since the continuous time Markov chain is finite and therefore non-explosive, π is a stationary distribution of the continuous time Markov chain. Moreover, for any state j

$$\pi(j) = \frac{1}{M}\gamma(j) = \frac{\mu(j)}{M\lambda(j)} > 0.$$

Stationary measures CTMS vs. embedded DTMC

Uniqueness: Let π and π' be two stationary distributions of the continuous time Markov chain. We want to prove that necessarily $\pi = \pi'$. Let

$$C = \sum_j \pi(j)\lambda(j) \quad \text{and} \quad C' = \sum_j \pi'(j)\lambda(j)$$

Since $\tilde{\pi}$ is unique, we have that for any state j

$$\tilde{\pi}(j) = \frac{\pi(j)\lambda(j)}{C} = \frac{\pi'(j)\lambda(j)}{C'},$$

which implies that for any state j, $\pi(j) = \frac{C'}{C}\pi'(j)$. If are able to show that $C'/C = 1$ we are done. But this follows from

$$1 = \sum_j \pi(j) = \sum_j \frac{C'}{C}\pi'(j) = \frac{C'}{C}\sum_j \pi'(j) = \frac{C'}{C}$$

so the proof is concluded.

Limit distribution, example

Consider a 2-state continuous time Markov chain with state space $S = \{1, 2\}$ and transition rate matrix

$$Q = \begin{pmatrix} -\alpha & \alpha \\ \beta & -\beta \end{pmatrix},$$

with $\alpha, \beta > 0$. We have seen that the stationary distribution for this model is given by

$$\pi = \left(\frac{\beta}{\alpha + \beta}, \frac{\alpha}{\alpha + \beta} \right).$$

The transition probabilities are given by

$$P_t = e^{Qt} = \begin{pmatrix} \frac{\beta}{\alpha+\beta} + \frac{\alpha}{\alpha+\beta} e^{-(\alpha+\beta)t} & \frac{\alpha}{\alpha+\beta} \left(1 - e^{-(\alpha+\beta)t}\right) \\ \frac{\beta}{\alpha+\beta} \left(1 - e^{-(\alpha+\beta)t}\right) & \frac{\alpha}{\alpha+\beta} + \frac{\beta}{\alpha+\beta} e^{-(\alpha+\beta)t} \end{pmatrix}$$

Limit distribution, example

it follows that

$$P_t \xrightarrow[t\to\infty]{} \begin{pmatrix} \frac{\beta}{\alpha+\beta} & \frac{\alpha}{\alpha+\beta} \\ \frac{\beta}{\alpha+\beta} & \frac{\alpha}{\alpha+\beta} \end{pmatrix} = \begin{pmatrix} \pi(1) & \pi(2) \\ \pi(1) & \pi(2) \end{pmatrix}$$

Specifically,

$$\lim_{t\to\infty} p_t(i,1) = \pi(1) \quad \text{for all } i \in \{1,2\}$$
$$\lim_{t\to\infty} p_t(i,2) = \pi(2) \quad \text{for all } i \in \{1,2\}$$

Hence, the transition probabilities converge to the stationary distribution!
REMARK: the embedded DTMC is periodic, therefore it does not have a limit distribution.

Limit distribution

For CTMCs the notion of periodicity does not exist! We have

Lemma

*If X_t is irreducible, then $p_t(i,j) > 0$ for any states i,j and **for any** $t > 0$.*

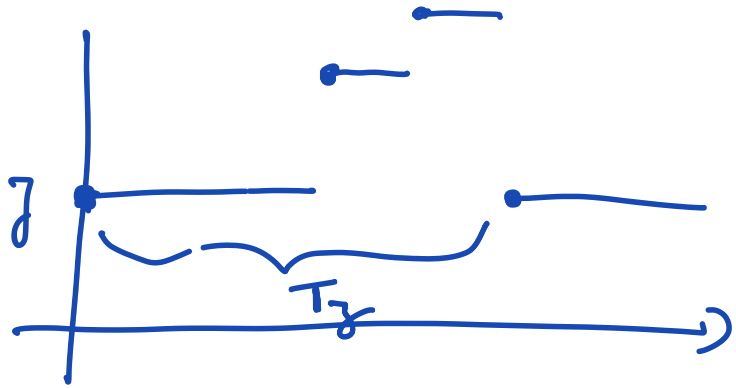
Since periodicity was the only thing preventing an irreducible DTMC from converging, we expect that convergence will occur more often in the continuous time setting. This is true, except for the case of explosions (that do not exist in DTMC).

Theorem

If a CTMC $(X_t)_{t=0}^{\infty}$ is irreducible and a has stationary distribution π, then for any $i,j \in S$, we have $\lim_{t\to\infty} p_t(i,j) = \pi(j)$.

Remember that to prove the existence of a stationary distribution γ it is not enough to check that $\gamma Q = 0$ and $\sum_i \gamma(i) = 1$, but we also need to know that the chain is not explosive!

Return times



In continuous time if $X_0 = j$, then X_t will stay in state j for a positive amount of time. To account for this fact, the concept of return time needs to be redefined as follows

$$T_j^c = \min\{t \geq 0 : X_t \text{ enters in } j \text{ from another state}\}$$
$$= \min\{t \geq 0 : X_t = j \text{ and } X_s \neq j \text{ for some } 0 \leq s < t\}$$

Remark: with the given definition the return time to an absorbing state is infinite.

Return times and recurrence

In continuous time if $X_0 = j$, then X_t will stay in state j for a positive amount of time. To account for this fact, the concept of return time needs to be redefined as follows

$$\begin{aligned} T_j^c &= \min\{t \geq 0 : X_t \text{ enters in } j \text{ from another state}\} \\ &= \min\{t \geq 0 : X_t = j \text{ and } X_s \neq j \text{ for some } 0 \leq s < t\} \end{aligned}$$

Remark: with the given definition the return time to an absorbing state is infinite.

Definition

A state j of a CTMC is recurrent if either $P_j(T_j^c < \infty) = 1$ or if j is absorbing.

Return times and recurrence

The return time T_j^c of a CTMC is related to the return time T_j^d of its embedded DTMC as follows. Let us define $\tau_{Y_{n-1}}$ the $n-th$ waiting time of the CTMC, that is the time elapsed between the $(n-1)$th and the nth jump. We have $\sim \text{Exp}(\lambda(Y_{n-1}))$, and hence it is a.s. finite. We have

$$T_j^c = \sum_{n=1}^{T_j^d} \tau_{Y_{n-1}},$$

therefore T_j^c is finite if and only if T_j^d is finite.

Theorem

Let $(X_t)_{t=0}^{\infty}$ be a non-explosive continuous time Markov chain. Then, j is a transient (resp. recurrent) state for the continuous time Markov chain $(X_t)_{t=0}^{\infty}$ if and only if it is a transient (resp. recurrent) state for the embedded discrete time Markov chain $(Y_n)_{n=0}^{\infty}$.

Positive and null recurrence

Let $(X_t)_{t=0}^{\infty}$ be a CTMC, and let j be a **recurrent** state. We say that

- j is positive recurrent if $E_j[T_j^c] < \infty$ or if j is absorbing;
- j is null recurrent if it is not positive recurrent: that is, if $E_j[T_j^c] = \infty$ and j is not absorbing;

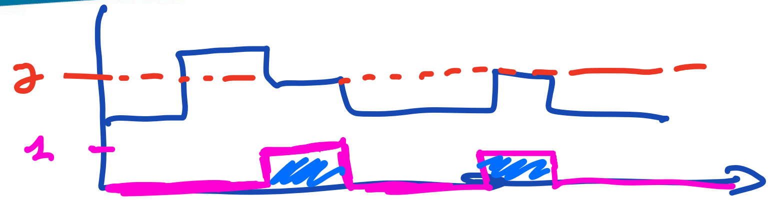
As for discrete time Markov chains, the following holds:

Theorem

Let $(X_t)_{t=0}^{\infty}$ be an irreducible continuous time Markov chain with state space S. Then the following are equivalent:

1. *every state in S is positive recurrent.*
2. *some state in S is positive recurrent.*
3. *There exists a stationary distribution π.*

Asymptotic frequency



The asymptotic freq. of state j is defined as the fraction of time spent therein:

$$\lim_{t\to\infty} \frac{N_t(j)}{t} = \lim_{t\to\infty} \frac{1}{t} \boxed{\int_0^t 1_{\{j\}}(X_s)ds.}$$

counts how much time do I spent at state j before time t

Theorem

Let $(X_t)_{t=0}^{\infty}$ be an irreducible CTMC, with arbitrary initial distribution. To avoid dealing with absorbing states (trivial), assume that the state space has at least 2 states. Then, for any $j \in S$,

$$\lim_{t\to\infty} \frac{N_t(j)}{t} = \frac{1}{E_j[T_j^c]\lambda(j)} \qquad a.s.$$

Theorem

Let $(X_t)_{t=0}^{\infty}$ be an irreducible CTMC. We assume that S contains at least two states. If π is a stationary distribution, then

$$\pi(j) = \frac{1}{\lambda(j)E_j[T_j^c]}.$$

Asymptotic reward

Theorem

Let $(X_t)_{t=0}^{\infty}$ be an irreducible CTMC, with arbitrary initial distribution. Assume there exists a stationary distribution π. Then, for any bounded function $f : S \to \mathbb{R}$ we have

$$\lim_{t\to\infty} \frac{1}{t} \int_0^t f(X_s)ds = \sum_{j\in S} f(j)\pi(j) \qquad a.s.$$

That is the time average converges to the space average, computed over the stationary distribution π.

Proof.

We have (the second equality holds for the bounded convergence theorem)

$$\lim_{t\to\infty} \frac{1}{t} \int_0^t f(X_s)ds = \lim_{t\to\infty} \sum_{j\in S} \frac{N_t(j)}{t} f(j) = \sum_{j\in S} \lim_{t\to\infty} \frac{N_t(j)}{t} f(j) = \sum_{j\in S} f(j)\pi(j),$$

Reaction networks

Chemical Reaction Networks (CRNs) are popular mathematical models of several phenomena in system biology, epidemiology, population dynamics, telecommunications, chemistry. In such models, individuals are identical units (e.g. molecules), classified into several groups (e.g. chemical species) and interact through the so-called reactions. A reaction means, for example, an individual eating one of another group, or dying, or reproducing, as well as a protein binding with the RNA to regulate gene expression. Reactions (as it happens in chemistry) are organised in a graph.

Example

An epidemic SI model. This is a more complex system. There are two population and the state is identified by the couple (S, I) that denotes the numbers of susceptible S and infected I. In any given state 6 transitions may occur.

reaction	descr.	rate	increment
$\emptyset \to S$	immigration	λ_1	(+1,0)
$\emptyset \to I$	immigration	λ_2	(0,+1)
$S \to \emptyset$	emigration/death	$\lambda_3 S$	(-1,0)
$I \to \emptyset$	death/emigration	$\lambda_4 I$	(0,-1)
$S + I \to 2I$	infection	$\lambda_5 SI$	(-1,+1)
$I \to S$	recovery	$\lambda_6 I$	(+1,-1)

The reaction graph has two connected components and will be drawn on the blackboard

Example

Prey-predator.
This is a more complex system. There are two population and the state is identified by the couple (p, P) that denotes the numbers of preys p and predators P. In any given state 4 transitions may occur

reaction	descr.	rate	increment
$p \to 2p$	prey reproduction	$\lambda_1 p$	(+1,0)
$p + P \to P$	prey consumption	$\lambda_2 pP$	(-1,0)
$p + P \to p + 2P$	predator reproduction	$\lambda_3 pP$	(0,+1)
$P \to \emptyset$	predator death	$\lambda_4 P$	(0,-1)

Reaction networks

A reaction network is a triple $\{\mathcal{S}, \mathcal{C}, \mathcal{R}\}$ such that

1. $\mathcal{S} = \{S_1, \cdots, S_d\}$ is the set of species of cardinality d where d is finite.
2. $\mathcal{C}$ is the set of complexes, consisting of some nonnegative integer linear combination of the species.
3. $\mathcal{R}$ is a finite set of ordered couples of complexes which is defined by the stoichiometric equations (1).

A reaction network of K chemical reactions is specified by stoichiometric equations

$$\sum_{i=1}^{d} y_{ki} S_i \to \sum_{i=1}^{d} y'_{ki} S_i, \quad \text{shortly} \quad y_k \to y'_k, \quad k = 1, \ldots, K \tag{1}$$

meaning that the reaction consumes $\sum y_{ki} S_i$ to produce $\sum y'_{ki} S_i$ where y_{ki}, y'_{ki} are nonnegative integers.

Stochastic Models of Reaction networks

The state of a reaction network model is specified by the vector

$$s = (s_1 \cdots s_d)$$

which counts how many molecules of each species are present in a given instant. Thus $S \subset \mathbb{N}^d$.

The occurrence of the reaction $y \to y'$ causes a state change (a jump) from s to $s + y' - y$. This is an event of a Markov chain process, whose rates have to be specified. An exponential race between the different reaction takes place, and such race decides which reaction take place first.

Stochastic Models of Reaction networks

The network follows the mass-action kinetics if the rate of reaction k in state s can be written in the form

$$q(s, s+y'-y) = \lambda_{y\to y'}(s) = \kappa_{y\to y'} \frac{s!}{(s-y)!} \mathbb{1}_{\{s\geq y\}} \tag{2}$$

$\lambda_{y\to y'}(s) > 0$ if and only if $s \geq y$ component-wise. $\lambda_{y\to y'}(s)$ is the transition propensity for reaction $y \to y'$.

Examples with species A,B,C:

$$\begin{array}{lll} A \to B & \text{has rate} & \lambda_{A\to B}(a,b,c) = \kappa_{A\to B}\, a \\ A+B \to C & \text{has rate} & \lambda_{A\to B}(a,b,c) = \kappa_{A+B\to C}\, ab \\ 2A \to C & \text{has rate} & \lambda_{2A\to C}(a,b,c) = \kappa_{2A\to C}\, a(a-1) \end{array}$$

Example

Observe that the following reaction network

$$\emptyset \underset{\lambda}{\overset{\mu}{\leftrightarrows}} Q$$

coincide with a M/M/∞ queue

Density dependent families of Markov chains

A family of Markov chains $X^V(t)$ with state space $S^V \subseteq \mathbb{N}^r_+$, is density dependent iff

- in each configuration only a finite number of state changes is possible
- the initial condition $X^V(0) = V \cdot x_0$
- for every possible state change l, there exists a continuous positive function $f_l : \mathbb{R}^r \to \mathbb{R}$ such that the instantaneous transition rate can be written as

$$q^V(k, k+l) = V f_l\left(\frac{k}{V}\right), \quad l \neq 0.$$

Remark: mass-action rates are density dependent if the constants are scaled:

$A \to B$ has rate $\lambda_{A\to B}(a,b,c) = \kappa_{A\to B}\, a = V\kappa_{A\to B}\frac{a}{V}$

$A + B \to C$ has rate $\lambda_{A\to B}(a,b,c) = \kappa'_{A+B\to C}\, ab = V\kappa_{A+B\to C}\frac{a}{V}\frac{b}{V}$

$2A \to C$ has rate $\lambda_{2A\to C}(a,b,c) = \kappa'_{2A\to C}\, a(a-1) = V\kappa_{2A\to C}\frac{a}{V}\frac{a-1}{V}$

Density processes

For any density dependent family $X^V(t)$ it makes sense to consider the density processes

$$Z^V(t) = \frac{X^V}{V}(t)$$

Its states are concentrations $x = \frac{k}{V}$ and belong to $\mathbb{Q}_+^r \subset \mathbb{R}_+^r$. Its rates are

$$\boxed{Q^V_{x,x+\frac{l}{V}} = Vf_l\left(x\right), \quad l \neq 0.}$$

thus if $V \to \infty$ the events of the density process becomes very frequent and cause small state changes.

Kurtz deterministic continuous approximations

Kurtz proved in the '70-'80 that the density process converges a.s. to the deterministic solution $x(t)$ of the following d-dimensional ODE system

$$\dot{x} = F(x(t)) = \sum_{l \in C} l f_l(x(t)) \tag{3}$$

under rather general assumptions. The proof relies on the previously introduced LLN for the Poisson process.

Kurtz deterministic continuous approximations

The precise statement follows

Theorem

Suppose that for each compact $K \subset E$,

$$\sum_{l} |l| \sup_{x \in K} f_l(x) < \infty$$

and the function F defined in (3) *is Lipshitz continuous in K, suppose that for each n $X^{[V]}(t)$ is DD with initial conditions $X^{[V]}(0)$ such that* $\lim_{n \to \infty} X^{[V]}(0) = x_0$ *and $x(t)$ solves* (3) *with initial condition $x(0) = x_0$, then for every $t \geq 0$,*

$$\lim_{n \to \infty} \sup_{s \leq t} |X^{[V]}(s) - x(s)| = 0 \qquad a.s.$$