Stochastic Processes and Stochastic Calculus - 3
Markov Chains

Dario Trevisan

Università degli Studi di Pisa

San Miniato - 13 September 2016

Overview

A “lake model” of employment flows

We propose a simple model for

- flows between unemployment and employment
- employment and unemployment rates at “equilibrium”

A “lake model” of employment flows

We propose a simple model for

- flows between unemployment and employment
- employment and unemployment rates at “equilibrium”

People can be

1. Employed, working and not looking for alternatives

A “lake model” of employment flows

We propose a simple model for

- flows between unemployment and employment
- employment and unemployment rates at “equilibrium”

People can be

1. Employed, working and not looking for alternatives
2. Unemployed and looking for a job

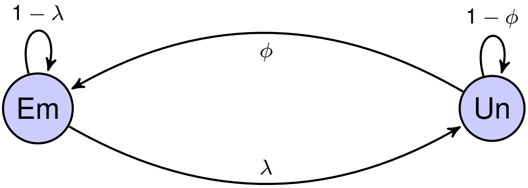
A “lake model” of employment flows

We propose a simple model for

- flows between unemployment and employment
- employment and unemployment rates at “equilibrium”

People can be

1. Employed, working and not looking for alternatives
2. Unemployed and looking for a job



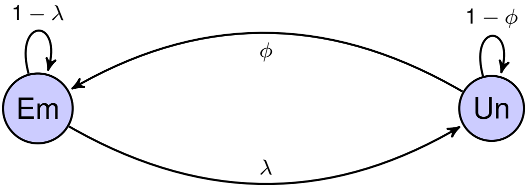
A “lake model” of employment flows

We propose a simple model for

- flows between unemployment and employment
- employment and unemployment rates at “equilibrium”

People can be

1. Employed, working and not looking for alternatives
2. Unemployed and looking for a job



Fix a unit of time (e.g. one month) and introduce two parameters

$$\lambda := \text{probability that a worker loses his/her job in a month}$$

$$\phi := \text{probability that an unemployed one finds a job in a month}$$

We call $\lambda, \phi \in [0, 1]$ transition probabilities.

A more precise description of the model

We assume that λ and ϕ do not depend

A more precise description of the model

We assume that λ and ϕ do not depend

- on the person $\rightarrow$ "typical" person

A more precise description of the model

We assume that λ and ϕ do not depend

- on the person $\rightarrow$ "typical" person
- nor on time $\rightarrow$ time-homogeneous

A more precise description of the model

We assume that λ and ϕ do not depend

- on the person $\rightarrow$ "typical" person
- nor on time $\rightarrow$ time-homogeneous

For $n = 0, 1, 2, \ldots$, let X_n denote the (random) state of employment of such "typical" person.

A more precise description of the model

We assume that λ and ϕ do not depend

- on the person $\rightarrow$ "typical" person
- nor on time $\rightarrow$ time-homogeneous

For $n = 0, 1, 2, \ldots$, let X_n denote the (random) state of employment of such "typical" person.

We have

$$\begin{array}{ll} \mathbb{P}(X_{n+1} = E | X_n = E) = 1 - \lambda & \mathbb{P}(X_{n+1} = U | X_n = E) = \lambda \\ \mathbb{P}(X_{n+1} = E | X_n = U) = \phi & \mathbb{P}(X_{n+1} = U | X_n = U) = 1 - \phi. \end{array}$$

A more precise description of the model

We assume that λ and ϕ do not depend

- on the person $\rightarrow$ "typical" person
- nor on time $\rightarrow$ time-homogeneous

For $n = 0, 1, 2, \ldots$, let X_n denote the (random) state of employment of such "typical" person.

We have

$$\begin{array}{ll} \mathbb{P}(X_{n+1} = E | X_n = E) = 1 - \lambda & \mathbb{P}(X_{n+1} = U | X_n = E) = \lambda \\ \mathbb{P}(X_{n+1} = E | X_n = U) = \phi & \mathbb{P}(X_{n+1} = U | X_n = U) = 1 - \phi. \end{array}$$

Problem

Are these assumptions sufficient to specify uniquely our model?

A more precise description of the model

We assume that λ and ϕ do not depend

- on the person $\rightarrow$ "typical" person
- nor on time $\rightarrow$ time-homogeneous

For $n = 0, 1, 2, \ldots$, let X_n denote the (random) state of employment of such "typical" person.

We have

$$\begin{array}{ll} \mathbb{P}(X_{n+1} = E | X_n = E) = 1 - \lambda & \mathbb{P}(X_{n+1} = U | X_n = E) = \lambda \\ \mathbb{P}(X_{n+1} = E | X_n = U) = \phi & \mathbb{P}(X_{n+1} = U | X_n = U) = 1 - \phi. \end{array}$$

Problem

Are these assumptions sufficient to specify uniquely our model?

NO. For example, we do not know the initial state X_0.

The Markov assumption

The Markov assumption

Problem

What about the probability of finding a new job, knowing that he/she has been unemployed for the last two months?

The Markov assumption

Problem

What about the probability of finding a new job, knowing that he/she has been unemployed for the last two months? How to compute

$$\mathbb{P}(X_{n+1} = E | X_n = U, X_{n-1} = U)?$$

Problem

What about the probability of finding a new job, knowing that he/she has been unemployed for the last two months? How to compute

$$\mathbb{P}(X_{n+1} = E | X_n = U, X_{n-1} = U)?$$

We may say that two months without work are not a big deal, so we could assume

$$\mathbb{P}(X_{n+1} = E | X_n = U, X_{n-1} = U) = \phi$$

Problem

What about the probability of finding a new job, knowing that he/she has been unemployed for the last two months? How to compute

$$\mathbb{P}(X_{n+1} = E | X_n = U, X_{n-1} = U)?$$

We may say that two months without work are not a big deal, so we could assume

$$\mathbb{P}(X_{n+1} = E | X_n = U, X_{n-1} = U) = \phi$$

But what about three or more? One year?

The Markov assumption

Problem

What about the probability of finding a new job, knowing that he/she has been unemployed for the last two months? How to compute

$$\mathbb{P}(X_{n+1} = E | X_n = U, X_{n-1} = U)?$$

We may say that two months without work are not a big deal, so we could assume

$$\mathbb{P}(X_{n+1} = E | X_n = U, X_{n-1} = U) = \phi$$

But what about three or more? One year? At this stage we introduce the following assumption:

Markov property

At any time, regardless of the information about the past months, the next-month state of employment depends uniquely on the present one, i.e.

$$\mathbb{P}(X_{n+1} = E | X_n = U, X_{n-1} = \cdot, X_{n-2} = \cdot, \ldots, X_{n-k} = \cdot) = \mathbb{P}(X_{n+1} = E | X_n = U)$$

$$\mathbb{P}(X_{n+1} = U | X_n = E, X_{n-1} = \cdot, X_{n-2} = \cdot, \ldots, X_{n-k} = \cdot) = \mathbb{P}(X_{n+1} = U | X_n = E).$$

The Markov assumption

How can we justify the Markov property assumption?

The Markov assumption

How can we justify the Markov property assumption?

- The model becomes very tractable analytically
- It fits real data (?) – we can make predictions
- Given only the data λ and ϕ (from real world), this is the “fairest” model that one can think.

The Markov assumption

How can we justify the Markov property assumption?

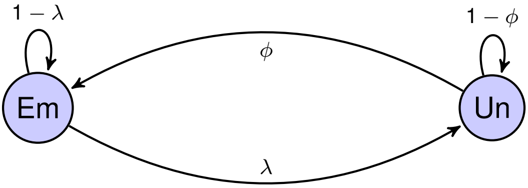
- The model becomes very tractable analytically
- It fits real data (?) – we can make predictions
- Given only the data λ and ϕ (from real world), this is the “fairest” model that one can think.

A more realistic assumption would be e.g.

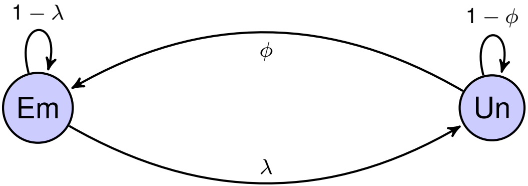
$$\mathbb{P}(X_{n+1} = E | X_n = U, X_{n-1} = U, X_{n-2} = U, \ldots, X_{n-k} = U) \leq \mathbb{P}(X_{n+1} = E | X_n = U)$$

but how to make it quantitative?

Examples



Examples



$$\begin{aligned}\mathbb{P}(X_2 = E, X_1 = U | X_0 = E) &= \mathbb{P}(X_2 = E | X_1 = U, X_0 = E)\mathbb{P}(X_1 = U | X_0 = E) \\ &= \mathbb{P}(X_2 = E | X_1 = U)\mathbb{P}(X_1 = U | X_0 = E) \\ &= \phi\lambda\end{aligned}$$

$$\mathbb{P}(X_3 = E, X_2 = U, X_1 = U | X_0 = E) = \phi(1 - \phi)\lambda$$

$$\mathbb{P}(X_3 = E | X_0 = E) = ?$$

The Markov property

Let us provide a general definition.

Let us provide a general definition.

Definition (Markov process)

Given a filtration $(\mathcal{F}_t)_{t\in T}$, a process $X_t : \Omega \to E$ is Markov if

$$\text{for all } s \leq t \in T,\ A \subseteq E \quad \mathbb{P}(X_t \in A|\mathcal{F}_s) = \mathbb{P}(X_t \in A|X_s)$$

The Markov property

Let us provide a general definition.

Definition (Markov process)

Given a filtration $(\mathcal{F}_t)_{t \in T}$, a process $X_t : \Omega \to E$ is Markov if

$$\text{for all } s \leq t \in T, A \subseteq E \quad \mathbb{P}(X_t \in A | \mathcal{F}_s) = \mathbb{P}(X_t \in A | X_s)$$

- Usually one has $\mathcal{F}_t = \sigma(X_r : r \leq t)$, the natural filtration of X_t.

Markov Chains - 1

Markov Chains - 1

To generalize the example above:

$\{E, U\}$ states of employment $\Rightarrow$ $\mathcal{S}$ set of states (finite or countable)

To generalize the example above:

$\{E, U\}$ states of employment $\Rightarrow$ $\mathcal{S}$ set of states (finite or countable)

Recall

$$\begin{array}{ll} \mathbb{P}(X_{n+1} = E | X_n = E) = 1 - \lambda & \mathbb{P}(X_{n+1} = U | X_n = E) = \lambda \\ \mathbb{P}(X_{n+1} = E | X_n = U) = \phi & \mathbb{P}(X_{n+1} = U | X_n = U) = 1 - \phi \end{array}$$

The transition probabilities are $|\mathcal{S}|^2$ numbers that we write matrix notation:

$$\begin{pmatrix} 1-\lambda & \lambda \\ \phi & 1-\phi \end{pmatrix} \Rightarrow Q = (p_{ij})_{i,j \in \mathcal{S}} \quad p_{ij} = \mathbb{P}(X_{n+1} = j | X_n = i) \quad \text{for } i, j \in \mathcal{S}$$

To generalize the example above:

$\{E, U\}$ states of employment $\Rightarrow$ $\mathcal{S}$ set of states (finite or countable)

Recall

$$\mathbb{P}(X_{n+1} = E | X_n = E) = 1 - \lambda \quad \mathbb{P}(X_{n+1} = U | X_n = E) = \lambda$$
$$\mathbb{P}(X_{n+1} = E | X_n = U) = \phi \quad \mathbb{P}(X_{n+1} = U | X_n = U) = 1 - \phi$$

The transition probabilities are $|\mathcal{S}|^2$ numbers that we write matrix notation:

$$\begin{pmatrix} 1-\lambda & \lambda \\ \phi & 1-\phi \end{pmatrix} \Rightarrow Q = (p_{ij})_{i,j \in \mathcal{S}} \quad p_{ij} = \mathbb{P}(X_{n+1} = j | X_n = i) \quad \text{for } i, j \in \mathcal{S}$$

Remark (Q is a stochastic matrix)

We must have $p_{ij} \in [0, 1]$ for $i, j \in \mathcal{S}$ and

$$\sum_{j \in \mathcal{S}} p_{ij} = \sum_{j \in \mathcal{S}} \mathbb{P}(X_{n+1} = j | X_n = i) = \mathbb{P}(X_{n+1} \in \mathcal{S} | X_n = i) = 1$$

Definiton (Markov chain)

Given

Definiton (Markov chain)

Given

- a finite or countable $\mathcal{S}$ (set of states)

Definiton (Markov chain)

Given

- a finite or countable $\mathcal{S}$ (set of states)
- a stochastic matrix $Q = (p_{ij})_{i,j\in\mathcal{S}}$ (transition probabilities)

Definiton (Markov chain)

Given

- a finite or countable $\mathcal{S}$ (set of states)
- a stochastic matrix $Q = (p_{ij})_{i,j \in \mathcal{S}}$ (transition probabilities)
- a filtration $(\mathcal{F})_{n \in \mathbb{N}}$

Markov Chains - 2

Definiton (Markov chain)

Given

- a finite or countable $\mathcal{S}$ (set of states)
- a stochastic matrix $Q = (p_{ij})_{i,j\in\mathcal{S}}$ (transition probabilities)
- a filtration $(\mathcal{F})_{n\in\mathbb{N}}$

we say that a process $X_n : \Omega \to \mathcal{S}$ ($n \in \mathbb{N}$) is a markov chain if for all $n \in \mathbb{N}$,

$$\mathbb{P}(X_{n+1} = j|\mathcal{F}_n) = \mathbb{P}(X_{n+1} = j|X_n) \quad \forall j \in \mathcal{S} \quad \text{(Markov property)}$$

$$\mathbb{P}(X_{n+1} = j|X_n = i) = p_{ij} \quad \forall i, j \in \mathcal{S} \quad \text{(transition probability)}$$

Markov Chains - 2

Definiton (Markov chain)

Given

- a finite or countable $\mathcal{S}$ (set of states)
- a stochastic matrix $Q = (p_{ij})_{i,j\in\mathcal{S}}$ (transition probabilities)
- a filtration $(\mathcal{F})_{n\in\mathbb{N}}$

we say that a process $X_n : \Omega \to \mathcal{S}$ ($n \in \mathbb{N}$) is a markov chain if for all $n \in \mathbb{N}$,

$$\mathbb{P}(X_{n+1} = j|\mathcal{F}_n) = \mathbb{P}(X_{n+1} = j|X_n) \quad \forall j \in \mathcal{S} \quad \text{(Markov property)}$$

$$\mathbb{P}(X_{n+1} = j|X_n = i) = p_{ij} \quad \forall i, j \in \mathcal{S} \quad \text{(transition probability)}$$

The two conditions above yield

$$\mathbb{P}(X_{n+1} = j|X_n = i, X_{n-1} = \cdot, X_{n-2} = \cdot, \dots, X_{n-k} = \cdot) = p_{ij}$$

Graphical representation

We associate

- a node to each state $i \in \mathcal{S}$
- a weighted arrow to each transition probability p_{ij} (no arrow if $p_{ij} = 0$).

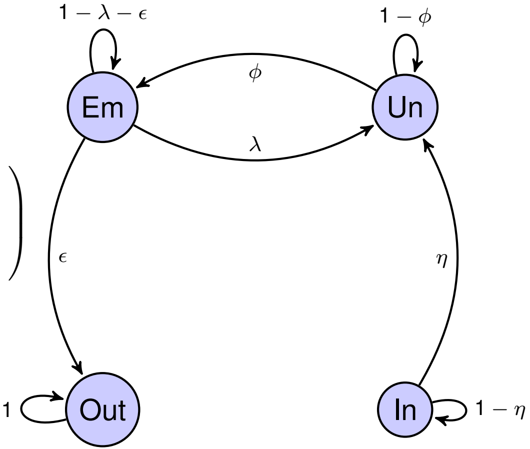
Graphical representation

We associate

- a node to each state $i \in \mathcal{S}$
- a weighted arrow to each transition probability p_{ij} (no arrow if $p_{ij} = 0$).

$$\mathcal{S} = (I, U, E, O)$$

$$Q = \begin{pmatrix} 1-\eta & \eta & 0 & 0 \\ 0 & \phi & 1-\phi & 0 \\ 0 & \lambda & 1-\lambda-\epsilon & \epsilon \\ 0 & 0 & 0 & 1 \end{pmatrix}$$



The following theorem ensures that Markov chains exist.

The Kolmogorov theorem

The following theorem ensures that Markov chains exist.

Theorem

Given a finite or countable $\mathcal{S}$, a stochastic matrix $Q = (p_{ij})_{i,j\in\mathcal{S}}$ and a probability measure μ_0 on $\mathcal{S}$.

∃) there exists a Markov chain $(X_n)_{n\in\mathbb{N}}$ (with respect to the natural filtration) with transition probability Q and law of X_0 equal to μ_0.

!) Such a chain X is unique in law, i.e. given any Markov chain Y with transition probability Q and law of X_0 equal to μ_0, one has

$$X = Y \quad \text{in law.}$$

The Kolmogorov theorem

The following theorem ensures that Markov chains exist.

Theorem

Given a finite or countable $\mathcal{S}$, a stochastic matrix $Q = (p_{ij})_{i,j\in\mathcal{S}}$ and a probability measure μ_0 on $\mathcal{S}$.

∃) there exists a Markov chain $(X_n)_{n\in\mathbb{N}}$ (with respect to the natural filtration) with transition probability Q and law of X_0 equal to μ_0.

!) Such a chain X is unique in law, i.e. given any Markov chain Y with transition probability Q and law of X_0 equal to μ_0, one has

$$X = Y \quad \text{in law.}$$

The condition $X = Y$ in law means that for every $n \geq 1$, $A_0, \dots, A_n \subseteq \mathcal{S}$

$$\mathbb{P}(X_0 \in A_0, X_1 \in A_1, \dots, X_n \in A_n) = \mathbb{P}(Y_0 \in A_0, Y_1 \in A_1, \dots, Y_n \in A_n)$$

Marginal laws

Marginal laws

How to compute the quantity

$$\mathbb{P}(X_0 \in A_0, X_1 \in A_1, \dots, X_n \in A_n)?$$

Marginal laws

How to compute the quantity

$$\mathbb{P}(X_0 \in A_0, X_1 \in A_1, \ldots, X_n \in A_n)?$$

The simplest case is that of 1-marginals, i.e. the law of X_n. For example

$$P(X_1 = j) = \sum_{i \in \mathcal{S}} P(X_1 = j | X_0 = i) P(X_0 = i) = \sum_{i \in \mathcal{S}} p_{ij} \mu_i^0$$

Marginal laws

How to compute the quantity

$$\mathbb{P}(X_0 \in A_0, X_1 \in A_1, \ldots, X_n \in A_n)?$$

The simplest case is that of 1-marginals, i.e. the law of X_n. For example

$$P(X_1 = j) = \sum_{i \in \mathcal{S}} P(X_1 = j | X_0 = i) P(X_0 = i) = \sum_{i \in \mathcal{S}} p_{ij} \mu_i^0$$

Recall the matrix-vector products for column vectors v or row vectors r

$$(Qv)_i = \sum_j p_{ij} v_j, \quad (rQ)_i = \sum_i p_{ij} r_i$$

Marginal laws

How to compute the quantity

$$\mathbb{P}(X_0 \in A_0, X_1 \in A_1, \ldots, X_n \in A_n)?$$

The simplest case is that of 1-marginals, i.e. the law of X_n. For example

$$P(X_1 = j) = \sum_{i \in \mathcal{S}} P(X_1 = j | X_0 = i) P(X_0 = i) = \sum_{i \in \mathcal{S}} p_{ij} \mu_i^0$$

Recall the matrix-vector products for column vectors v or row vectors r

$$(Qv)_i = \sum_j p_{ij} v_j, \quad (rQ)_i = \sum_i p_{ij} r_i$$

If we identify μ^0 and $P(X_1 = \cdot)$ with row vectors, we have

$$P(X_1 = \cdot) = \mu^0 Q$$

Marginal laws

How to compute the quantity

$$\mathbb{P}(X_0 \in A_0, X_1 \in A_1, \ldots, X_n \in A_n)?$$

The simplest case is that of 1-marginals, i.e. the law of X_n. For example

$$P(X_1 = j) = \sum_{i \in \mathcal{S}} P(X_1 = j | X_0 = i) P(X_0 = i) = \sum_{i \in \mathcal{S}} p_{ij} \mu_i^0$$

Recall the matrix-vector products for column vectors v or row vectors r

$$(Qv)_i = \sum_j p_{ij} v_j, \quad (rQ)_i = \sum_i p_{ij} r_i$$

If we identify μ^0 and $P(X_1 = \cdot)$ with row vectors, we have

$$P(X_1 = \cdot) = \mu^0 Q$$

Proposition (marginal laws)

$$P(X_n = \cdot) = \mu^0 (Q^n), \quad \text{for any } n \geq 0.$$

A geometric interpretation of Q^n

A geometric interpretation of Q^n

By definition of power of a matrix

$$(Q^n)_{ij} = \sum_k p_{ik}(Q^{n-1})_{kj} = \sum_{k_1, k_2, \ldots k_{(n-1)}} p_{ik_1} p_{k_1 k_2} \cdot p_{k_{(n-1)} j}$$

A geometric interpretation of Q^n

By definition of power of a matrix

$$(Q^n)_{ij} = \sum_k p_{ik}(Q^{n-1})_{kj} = \sum_{k_1,k_2,\ldots k_{(n-1)}} p_{ik_1} p_{k_1 k_2} \cdot p_{k_{(n-1)}j}$$

1. Each choice of $k_1, \ldots, k_{(n-1)}$ defines the path of ‘length” n

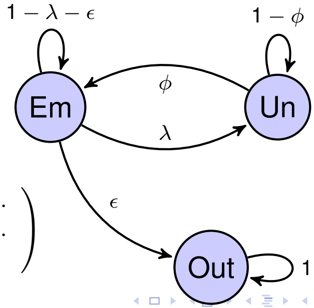
$$i \to k_1 \to k_2 \to \ldots \to k_{(n-1)} \to j.$$

2. Each path has total weight given by the product of the weights.
3. We sum the weights over all the paths of length n joining i to j.

$$\mathcal{S} = (U, E, O)$$

$$Q = \begin{pmatrix} 1-\phi & \phi & 0 \\ \lambda & 1-\lambda-\epsilon & \epsilon \\ 0 & 0 & 1 \end{pmatrix}$$

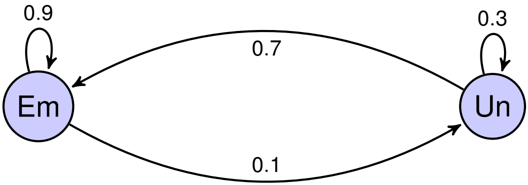
$$Q^2 = \begin{pmatrix} \lambda\phi & \ldots & \ldots \\ \lambda(1-\phi)+\lambda(1-\lambda-\epsilon) & \ldots & \ldots \\ 0 & 0 & 1 \end{pmatrix}$$



Long time behavior

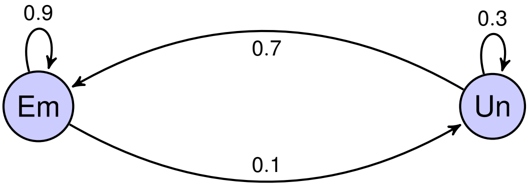
Long time behavior

Consider the 2-states model, with $\lambda = 0.1$ and $\phi = 0.3$.



Long time behavior

Consider the 2-states model, with $\lambda = 0.1$ and $\phi = 0.3$.

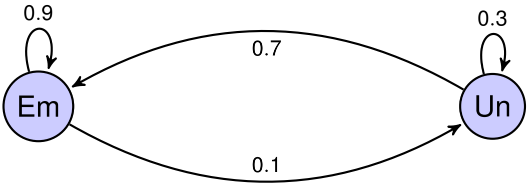


Problem

What can we say about X_n, for large n?

Long time behavior

Consider the 2-states model, with $\lambda = 0.1$ and $\phi = 0.3$.



Problem

What can we say about X_n, for large n?

$\mathbb{P}(X_n = E)$ will be close to 1. How to compute it?

$$Q = \begin{pmatrix} 0.9 & 0.1 \\ 0.7 & 0.3 \end{pmatrix} \quad Q^2 = \begin{pmatrix} 0.88 & 0.12 \\ 0.84 & 0.16 \end{pmatrix} \quad Q^4 \sim \begin{pmatrix} 0.87 & 0.13 \\ 0.87 & 0.13 \end{pmatrix}$$

Invariant distributions

Do the laws of $(X_n)_{n \geq 1}$ converges as $n \to \infty$ towards some limit law $\bar{\mu}$ on $\mathcal{S}$?

Problem

What are relevant properties of $\bar{\mu}$?

Do the laws of $(X_n)_{n\geq 1}$ converges as $n \to \infty$ towards some limit law $\bar{\mu}$ on $\mathcal{S}$?

Problem

What are relevant properties of $\bar{\mu}$?

Since the law of X_n is $\mu^0(Q^n)$, we should find

$$\bar{\mu} = \lim_{n\to\infty} \mu^0(Q^n)$$

but also

$$\bar{\mu}Q = \lim_{n\to\infty} \mu^0(Q^n)Q = \lim_{n\to\infty} \mu^0(Q^{n+1}) = \bar{\mu}$$

Invariant distributions

Do the laws of $(X_n)_{n\geq 1}$ converges as $n \to \infty$ towards some limit law $\bar{\mu}$ on $\mathcal{S}$?

Problem

What are relevant properties of $\bar{\mu}$?

Since the law of X_n is $\mu^0(Q^n)$, we should find

$$\bar{\mu} = \lim_{n\to\infty} \mu^0(Q^n)$$

but also

$$\bar{\mu}Q = \lim_{n\to\infty} \mu^0(Q^n)Q = \lim_{n\to\infty} \mu^0(Q^{n+1}) = \bar{\mu}$$

We obtain that $\bar{\mu}$ is invariant (also called stationary), i.e.

$$\bar{\mu}Q = \bar{\mu}.$$

Do the laws of $(X_n)_{n\geq 1}$ converges as $n \to \infty$ towards some limit law $\bar{\mu}$ on $\mathcal{S}$?

Problem

What are relevant properties of $\bar{\mu}$?

Since the law of X_n is $\mu^0(Q^n)$, we should find

$$\bar{\mu} = \lim_{n\to\infty} \mu^0(Q^n)$$

but also

$$\bar{\mu}Q = \lim_{n\to\infty} \mu^0(Q^n)Q = \lim_{n\to\infty} \mu^0(Q^{n+1}) = \bar{\mu}$$

We obtain that $\bar{\mu}$ is invariant (also called stationary), i.e.

$$\bar{\mu}Q = \bar{\mu}.$$

In linear algebra notation, $\bar{\mu}$ is a row eigenvector of Q with eigenvalue 1.

Existence of Invariant distributions

Existence of Invariant distributions

Theorem (Perron-Frobenius)

If the number of states $\mathcal{S}$ is finite, there exists at least one invariant distribution:

$$\bar{\mu}Q = \bar{\mu}.$$

Existence of Invariant distributions

Theorem (Perron-Frobenius)

If the number of states $\mathcal{S}$ is finite, there exists at least one invariant distribution:

$$\bar{\mu}Q = \bar{\mu}.$$

Proof.

Choose any μ^0 and consider the averages

$$\bar{\mu}^n := \frac{1}{n}\left(\mu^0 + \mu^0 Q + \mu^0 Q^2 + \ldots, + \mu^0 Q^{n-1}\right).$$

Existence of Invariant distributions

Theorem (Perron-Frobenius)

If the number of states $\mathcal{S}$ is finite, there exists at least one invariant distribution:

$$\bar{\mu} Q = \bar{\mu}.$$

Proof.

Choose any μ^0 and consider the averages

$$\bar{\mu}^n := \frac{1}{n}\left(\mu^0 + \mu^0 Q + \mu^0 Q^2 + \ldots, + \mu^0 Q^{n-1}\right).$$

These are “almost invariant”:

$$\bar{\mu}^n - \bar{\mu}^n Q = \frac{1}{n}\left(\mu^0 - Q^n \mu^0\right) \to 0$$

Existence of Invariant distributions

Theorem (Perron-Frobenius)

If the number of states $\mathcal{S}$ is finite, there exists at least one invariant distribution:

$$\bar{\mu}Q = \bar{\mu}.$$

Proof.

Choose any μ^0 and consider the averages

$$\bar{\mu}^n := \frac{1}{n}\left(\mu^0 + \mu^0 Q + \mu^0 Q^2 + \ldots, + \mu^0 Q^{n-1}\right).$$

These are "almost invariant":

$$\bar{\mu}^n - \bar{\mu}^n Q = \frac{1}{n}\left(\mu^0 - Q^n \mu^0\right) \to 0$$

One can prove that some limit point exists and it is invariant. □

Existence of Invariant distributions

Theorem (Perron-Frobenius)

If the number of states $\mathcal{S}$ is finite, there exists at least one invariant distribution:

$$\bar{\mu}Q = \bar{\mu}.$$

Proof.

Choose any μ^0 and consider the averages

$$\bar{\mu}^n := \frac{1}{n}\left(\mu^0 + \mu^0 Q + \mu^0 Q^2 + \dots, + \mu^0 Q^{n-1}\right).$$

These are "almost invariant":

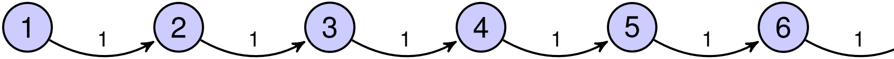
$$\bar{\mu}^n - \bar{\mu}^n Q = \frac{1}{n}\left(\mu^0 - Q^n \mu^0\right) \to 0$$

One can prove that some limit point exists and it is invariant. □

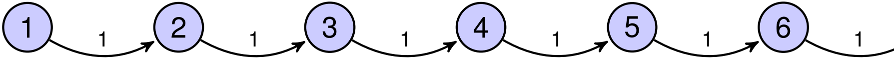
Notice that the proof gives a "way" to find $\bar{\mu}$, essentially by taking powers Q^n and averaging.

Beware: the existence result could be false if the number of states is infinite!

Beware: the existence result could be false if the number of states is infinite!
Let $\mathcal{S} = \mathbb{N}$

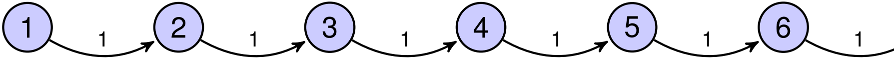


Beware: the existence result could be false if the number of states is infinite!
Let $\mathcal{S} = \mathbb{N}$



$$P(X_{n+1} = i + 1 | X_n = i) = 1$$

Beware: the existence result could be false if the number of states is infinite!
Let $\mathcal{S} = \mathbb{N}$



$$P(X_{n+1} = i + 1 | X_n = i) = 1$$

- We have $P(X_{n+1} = i) = 0$, for every $i \leq n$
- Any invariant probability satisifes $\bar{\mu}_i = 0$ for every i

Convergence of the laws

If the number of states is finite, we know that some $\bar{\mu}$ exists.

Convergence of the laws

If the number of states is finite, we know that some $\bar{\mu}$ exists.

Problem

Is it true that, for $j \in \mathcal{S}$,

$$P(X_n = j) \to \bar{\mu}(j), \quad \text{as } n \to \infty?$$

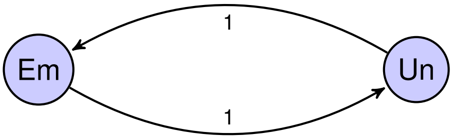
If the number of states is finite, we know that some $\bar{\mu}$ exists.

Problem

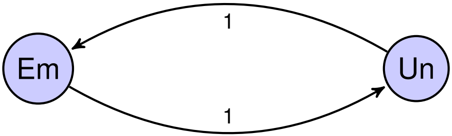
Is it true that, for $j \in \mathcal{S}$,

$$P(X_n = j) \to \bar{\mu}(j), \quad \text{as } n \to \infty?$$

In general the answer is NO. Let us consider one example.



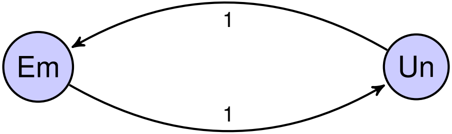
Why? If $P(X_0 = E) = 1$, we have $P(X_n = E) = 1$ or 0 alternatively...



Why? If $P(X_0 = E) = 1$, we have $P(X_n = E) = 1$ or 0 alternatively. . . If we write the transition matrix

$$Q = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad Q^2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad Q^3 = Q = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

We see there is no limit as $n \to \infty$.



Why? If $P(X_0 = E) = 1$, we have $P(X_n = E) = 1$ or 0 alternatively... If we write the transition matrix

$$Q = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad Q^2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad Q^3 = Q = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

We see there is no limit as $n \to \infty$.

Question: what are the invariant distributions for Q?

Communicating states

Communicating states

Definition (Communicating states)

Given states $i, j \in \mathcal{S}$, we say that $i \rightsquigarrow j$ if there is $n \in \mathbb{N}$ such that

$$(Q^n)_{ij} = P(X_n = j \mid X_0 = i) > 0.$$

Communicating states

Definition (Communicating states)

Given states $i, j \in \mathcal{S}$, we say that $i \rightsquigarrow j$ if there is $n \in \mathbb{N}$ such that

$$(Q^n)_{ij} = P(X_n = j \,|\, X_0 = i) > 0.$$

How to see if $i \rightsquigarrow j$?

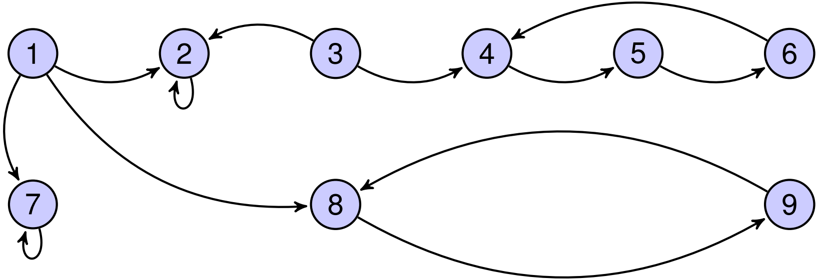
Communicating states

Definition (Communicating states)

Given states $i, j \in \mathcal{S}$, we say that $i \rightsquigarrow j$ if there is $n \in \mathbb{N}$ such that

$$(Q^n)_{ij} = P(X_n = j \,|\, X_0 = i) > 0.$$

How to see if $i \rightsquigarrow j$? just follow the arrows.

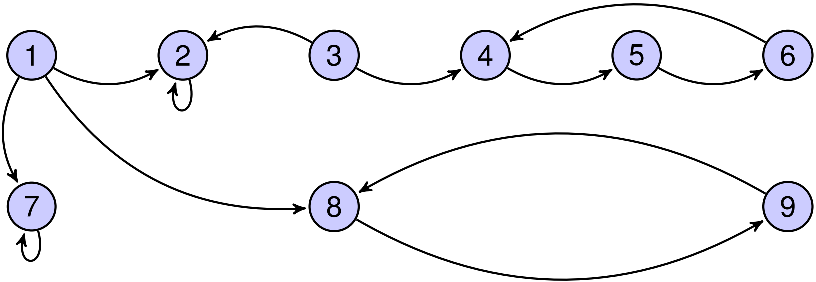


Definition (Communicating states)

Given states $i, j \in \mathcal{S}$, we say that $i \rightsquigarrow j$ if there is $n \in \mathbb{N}$ such that

$$(Q^n)_{ij} = P(X_n = j \mid X_0 = i) > 0.$$

How to see if $i \rightsquigarrow j$? just follow the arrows.



Notice that $1 \rightsquigarrow 2$ but not $2 \rightsquigarrow 1$.

Irreducible chains

Irreducible chains

Irreducible transition probability

We say that the transition matrix Q is irreducible if

$$\text{for every } i, j \in \mathcal{S},\ i \rightsquigarrow j.$$

Irreducible chains

Irreducible transition probability

We say that the transition matrix Q is irreducible if

$$\text{for every } i, j \in \mathcal{S},\ i \rightsquigarrow j.$$

- The example

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

  is irreducible.

Irreducible chains

Irreducible transition probability

We say that the transition matrix Q is irreducible if

$$\text{for every } i, j \in \mathcal{S},\ i \rightsquigarrow j.$$

- The example

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

  is irreducible.
- in terms of the transition matrix, this means that for every i, j there exists $m \in \mathbb{N}$ such that $(Q^m)_{ij} > 0$.

Irreducible chains

Irreducible transition probability

We say that the transition matrix Q is irreducible if

$$\text{for every } i, j \in \mathcal{S},\ i \rightsquigarrow j.$$

- The example

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

  is irreducible.
- in terms of the transition matrix, this means that for every i, j there exists $m \in \mathbb{N}$ such that $(Q^m)_{ij} > 0$.
- The name "irreducible" comes from the fact that any Markov chain can be decomposed into smaller chains which are irreducible, plus some "remainder" (called transitory states).

Regularity

Regularity

For our problem of convergence towards invariant distributions, we need something more than irreducible chains.
The problem with

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

is irreducible but not “random” (it is deterministic).

Regularity

For our problem of convergence towards invariant distributions, we need something more than irreducible chains.
The problem with

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

is irreducible but not "random" (it is deterministic).

Regular transition probability

We say that the transition matrix Q is regular if there exists $m \in \mathbb{N}$ such that Q^m has strictly positive entries, i.e.

$$(Q^m)_{ij} > 0 \quad \text{for every } i, j \in \mathcal{S}.$$

Regularity

For our problem of convergence towards invariant distributions, we need something more than irreducible chains.
The problem with

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

is irreducible but not “random” (it is deterministic).

Regular transition probability

We say that the transition matrix Q is regular if there exists $m \in \mathbb{N}$ such that Q^m has strictly positive entries, i.e.

$$(Q^m)_{ij} > 0 \quad \text{for every } i, j \in \mathcal{S}.$$

- Notice regular $\to$ irreducible, but m does not depend upon $i, j \in \mathcal{S}$.

Regularity

For our problem of convergence towards invariant distributions, we need something more than irreducible chains.
The problem with

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

is irreducible but not “random” (it is deterministic).

Regular transition probability

We say that the transition matrix Q is regular if there exists $m \in \mathbb{N}$ such that Q^m has strictly positive entries, i.e.

$$(Q^m)_{ij} > 0 \quad \text{for every } i, j \in \mathcal{S}.$$

- Notice regular $\rightarrow$ irreducible, but m does not depend upon $i, j \in \mathcal{S}$.
- The example

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

  is not regular.

In terms of the Markov chain, regularity means that there exists $m \in \mathbb{N}$ such that

$$P(X_m = j | X_0 = i) > 0 \quad \text{for every } i, j \in \mathcal{S},$$

In terms of the Markov chain, regularity means that there exists $m \in \mathbb{N}$ such that

$$P(X_m = j | X_0 = i) > 0 \quad \text{for every } i, j \in \mathcal{S},$$

This means that after some time m we are “very uncertain” about the actual position, we cannot exclude of being at any state.

In terms of the Markov chain, regularity means that there exists $m \in \mathbb{N}$ such that

$$P(X_m = j | X_0 = i) > 0 \quad \text{for every } i, j \in \mathcal{S},$$

This means that after some time m we are "very uncertain" about the actual position, we cannot exclude of being at any state.

Notice that if Q^m has strictly positive entries, then Q^{m+1} also is strictly positive.

$$\begin{aligned} P(X_{m+1} = j | X_0 = i) &= \sum_k P(X_{m+1} = j | X_1 = k, X_0 = i) P(X_1 = k | X_0 = i) \\ &= \sum_k P(X_{m+1} = j | X_1 = k) P(X_1 = k | X_0 = i) > 0 \end{aligned}$$

In terms of the Markov chain, regularity means that there exists $m \in \mathbb{N}$ such that

$$P(X_m = j | X_0 = i) > 0 \quad \text{for every } i, j \in \mathcal{S},$$

This means that after some time m we are "very uncertain" about the actual position, we cannot exclude of being at any state.

Notice that if Q^m has strictly positive entries, then Q^{m+1} also is strictly positive.

$$\begin{aligned} P(X_{m+1} = j | X_0 = i) &= \sum_k P(X_{m+1} = j | X_1 = k, X_0 = i) P(X_1 = k | X_0 = i) \\ &= \sum_k P(X_{m+1} = j | X_1 = k) P(X_1 = k | X_0 = i) > 0 \end{aligned}$$

because

$$\sum_k P(X_1 = k | X_0 = i) = 1.$$

The importance of being regular

The importance of being regular

Theorem (Markov)

If Q is a regular transition probability (on a finite state space $\mathcal{S}$) then there is a unique invariant distribution $\bar{\mu}$ and

$$\lim_{n\to\infty} (Q^n)_{ij} = \bar{\mu}(j), \quad \text{for every } i, j \in \mathcal{S}.$$

The importance of being regular

Theorem (Markov)

If Q is a regular transition probability (on a finite state space $\mathcal{S}$) then there is a unique invariant distribution $\bar{\mu}$ and

$$\lim_{n\to\infty} (Q^n)_{ij} = \bar{\mu}(j), \quad \text{for every } i, j \in \mathcal{S}.$$

In terms of Markov chain, this answer positively to the problem

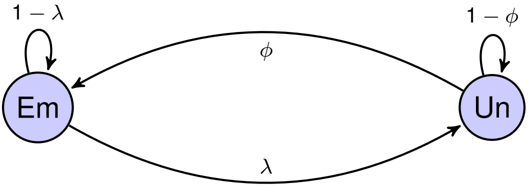
$$P(X_n = j) \to \bar{\mu}(j), \quad \text{as } n \to \infty,$$

for every $j \in \mathcal{S}$, whatever the initial law of X_0 was.

Ergodicity

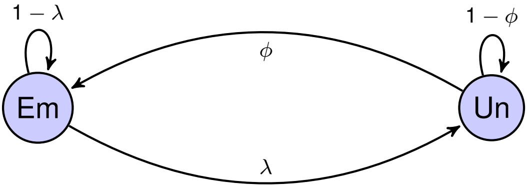
Ergodicity

Going back to our model of employment flows,



Ergodicity

Going back to our model of employment flows,

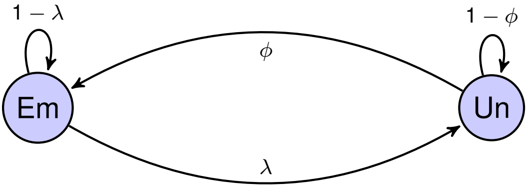


Problem

What is the “typical” time spent looking for a new job (over a long period)?

Ergodicity

Going back to our model of employment flows,



Problem

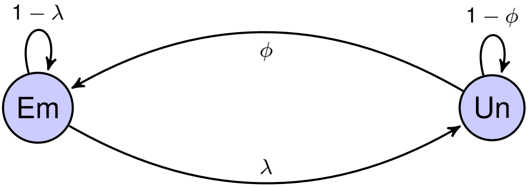
What is the “typical” time spent looking for a new job (over a long period)?

Answer to this question uses the “ergodicity” principle from physics:

time averages on long period $=$ space averages for large numbers of indivituals

Ergodicity

Going back to our model of employment flows,



Problem

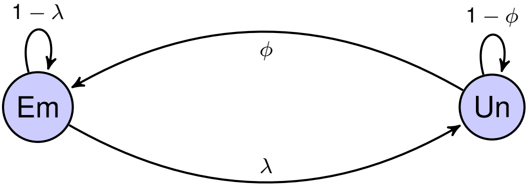
What is the “typical” time spent looking for a new job (over a long period)?

Answer to this question uses the “ergodicity” principle from physics:

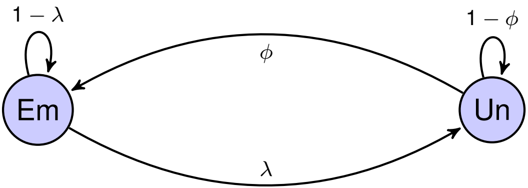
time averages on long period $=$ space averages for large numbers of indivituals

Beware: this holds for systems at equilibrium!

Ergodicity



Ergodicity

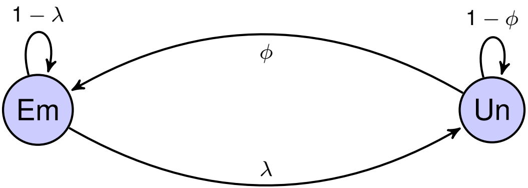


In our model, the large number of individuals is encoded in the law of X_n, hence the space average at equilibrium is just

$$\bar{\mu}(E),$$

with respect to the invariant distribution $\bar{\mu}$.

Ergodicity



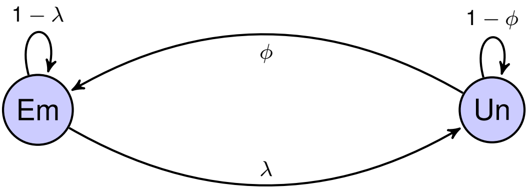
In our model, the large number of individuals is encoded in the law of X_n, hence the space average at equilibrium is just

$$\bar{\mu}(E),$$

with respect to the invariant distribution $\bar{\mu}$.
We compute the time average of being unemployed:

$$\frac{\sharp \{k \in \{0, 1, \ldots, n-1\} \mid X_k = E\}}{n}.$$

Ergodicity



In our model, the large number of individuals is encoded in the law of X_n, hence the space average at equilibrium is just

$$\bar{\mu}(E),$$

with respect to the invariant distribution $\bar{\mu}$.
We compute the time average of being unemployed:

$$\frac{\sharp\{k \in \{0,1,\dots,n-1\} \mid X_k = E\}}{n}.$$

Then the ergodic principle reads

$$\lim_{n\to\infty} \frac{\sharp\{k \in \{0,1,\dots,n-1\} \mid X_k = E\}}{n} = \bar{\mu}(E).$$

The ergodic principle is actually a theorem in the setting of Markov chains.

The ergodic principle is actually a theorem in the setting of Markov chains.

Ergodic theorem

Let $(X_n)_{n\geq 1}$, be a irreducible Markov chain on a finite state space $\mathcal{S}$. Then

$$\lim_{n\to\infty} \frac{\sharp\{k \in \{0, 1, \ldots, n-1\} \mid X_k = j\}}{n} = \bar{\mu}(j), \quad \text{for every } j \in \mathcal{S}.$$

The ergodic principle is actually a theorem in the setting of Markov chains.

Ergodic theorem

Let $(X_n)_{n \geq 1}$, be a irreducible Markov chain on a finite state space $\mathcal{S}$. Then

$$\lim_{n \to \infty} \frac{\sharp \{k \in \{0, 1, \dots, n-1\} \mid X_k = j\}}{n} = \bar{\mu}(j), \quad \text{for every } j \in \mathcal{S}.$$

Of course, this theorem applies also to regular Markov chains.

Problem

How to modify our model if, instead of measuring time intervals in months we have better precision, e.g. weeks (1 month $\sim$ 4 weeks).

Problem

How to modify our model if, instead of measuring time intervals in months we have better precision, e.g. weeks (1 month $\sim$ 4 weeks).

We can split the time $\{0, 1, 2, \dots\}$ into

$$\{0, 1/4, 2/4, 3/4, 1, 5/4, 6/4, 7/4, \dots\}$$

and build a process $Y_0, Y_{1/4}, Y_{2/4}, Y_{3/4}, \dots$.
If we want to have a Markov chain "close" to the original one, we should have $Q_X \sim Q_Y^4$, i.e. $Q_Y = Q_X^{1/4}$.

Problem

How to modify our model if, instead of measuring time intervals in months we have better precision, e.g. weeks (1 month $\sim$ 4 weeks).

We can split the time $\{0, 1, 2, \ldots\}$ into

$$\{0, 1/4, 2/4, 3/4, 1, 5/4, 6/4, 7/4, \ldots\}$$

and build a process $Y_0, Y_{1/4}, Y_{2/4}, Y_{3/4}, \ldots$.
If we want to have a Markov chain "close" to the original one, we should have $Q_X \sim Q_Y^4$, i.e. $Q_Y = Q_X^{1/4}$.
If instead of 4 weeks we split into days, we should have the transition probability $Q_X^{1/30}$.

Problem

How to modify our model if, instead of measuring time intervals in months we have better precision, e.g. weeks (1 month $\sim$ 4 weeks).

We can split the time $\{0, 1, 2, \ldots\}$ into

$$\{0, 1/4, 2/4, 3/4, 1, 5/4, 6/4, 7/4, \ldots\}$$

and build a process $Y_0, Y_{1/4}, Y_{2/4}, Y_{3/4}, \ldots$.
If we want to have a Markov chain "close" to the original one, we should have $Q_X \sim Q_Y^4$, i.e. $Q_Y = Q_X^{1/4}$.
If instead of 4 weeks we split into days, we should have the transition probability $Q_X^{1/30}$.
What happens if we split times into infinitely many small intervals, i.e. continous times?

Continuous time Markov chains

Problem

How to modify our model if, instead of measuring time intervals in months we have better precision, e.g. weeks (1 month $\sim$ 4 weeks).

We can split the time $\{0, 1, 2, \dots\}$ into

$$\{0, 1/4, 2/4, 3/4, 1, 5/4, 6/4, 7/4, \dots\}$$

and build a process $Y_0, Y_{1/4}, Y_{2/4}, Y_{3/4}, \dots$.
If we want to have a Markov chain "close" to the original one, we should have $Q_X \sim Q_Y^4$, i.e. $Q_Y = Q_X^{1/4}$.
If instead of 4 weeks we split into days, we should have the transition probability $Q_X^{1/30}$.
What happens if we split times into infinitely many small intervals, i.e. continous times?
We have $Q^{1/\infty} = Q^0 = Id$, but imagine, for large n,

$$Q^{1/n} \sim Id + \frac{1}{n} R + \text{smaller terms.}$$

We describe the chain by means of the matrix R (called transition rate matrix).

Recalling the formal expansion

$$Q^{1/n} \sim Id + \frac{1}{n}R + \text{smaller terms.}$$

Assuming $Q^{1/n}$ to be a stochastic matrix, then we must have

$$R_{ij} \geq 0, \text{for } i \neq j, \quad R_{ii} \leq 0$$

and

$$\sum_j R_{ij} = 0.$$

Recalling the formal expansion

$$Q^{1/n} \sim Id + \frac{1}{n} R + \text{smaller terms.}$$

Assuming $Q^{1/n}$ to be a stochastic matrix, then we must have

$$R_{ij} \geq 0, \text{for } i \neq j, \quad R_{ii} \leq 0$$

and

$$\sum_j R_{ij} = 0.$$

We interpret R_{ij} as the rate at which we jump from i to j.

Actually, we can describe the continuous Markov chain as follows:

1. When the particle is on the state $i \in \mathcal{S}$, take independent "alarm clocks" A_j, one for every $j \in \mathcal{S}$ with exponential laws

$$A_j \sim \mathcal{E}(r_{ij})$$

Actually, we can describe the continuous Markov chain as follows:

1. When the particle is on the state $i \in \mathcal{S}$, take independent "alarm clocks" A_j, one for every $j \in \mathcal{S}$ with exponential laws

$$A_j \sim \mathcal{E}(r_{ij})$$

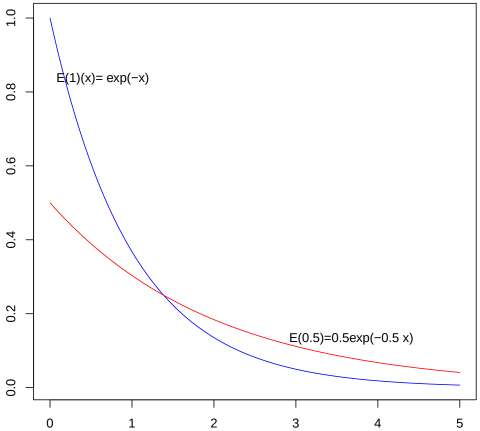
2. When the first clock rings, say A_j jump to state j and repeat the previous point.

Actually, we can describe the continuous Markov chain as follows:

1. When the particle is on the state $i \in \mathcal{S}$, take independent "alarm clocks" A_j, one for every $j \in \mathcal{S}$ with exponential laws

$$A_j \sim \mathcal{E}(r_{ij})$$

2. When the first clock rings, say A_j jump to state j and repeat the previous point.



Poisson process

Poisson process

An important example of such continuous time jump process has state space

$$\mathcal{S} = \mathbb{N} = \{0, 1, \ldots\}$$

and transition rates

$$R_{ij} = \begin{cases} 1 & \text{if } j = i + 1 \\ -1 & \text{if } j = i \\ 0 & \text{otherwise.} \end{cases}$$

Poisson process

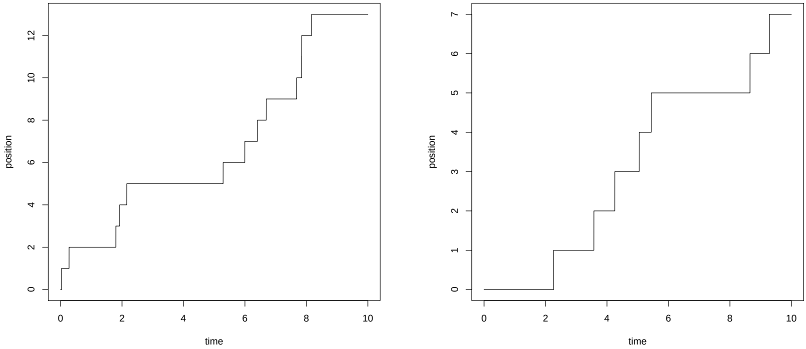
An important example of such continuous time jump process has state space

$$\mathcal{S} = \mathbb{N} = \{0, 1, \ldots\}$$

and transition rates

$$R_{ij} = \begin{cases} 1 & \text{if } j = i + 1 \\ -1 & \text{if } j = i \\ 0 & \text{otherwise.} \end{cases}$$

Two realizations of a Poisson process:



We denote $(N_t)_{t\in[0,\infty)}$ the Poisson process (N = number of jumps).

We denote $(N_t)_{t \in [0,\infty)}$ the Poisson process (N = number of jumps).
It is possible to prove that

- N is a Markov process (obvious)
- For every $t \in [0, \infty)$, N_t has Poisson law

$$P(N_t = k) = \frac{e^{-k}}{k!}, \quad k \in \mathbb{N}.$$

  In particular,

$$\mathbb{E}\left[N_t\right] = t.$$

- the increments are independent, e.g. for $s < t < u < v$,

  $N_t - N_s$ and $N_v - N_u$ are independent random variables.

One can also change the intensity of jumps (not the size), by changing the transition rates: for $\lambda > 0$,

$$R_{ij}^{\lambda} = \lambda R_{ij}$$