Thinning

An arrival process, $(N(s))_{s\in[0,\infty)}$ is Poisson with rate λ. Assume that the arrivals can be of k different types, specified by a sequence of iid random variables $\{Y_i\}_{i=1}^{\infty}$, taking values in $\{1,2,3,\dots,k\}$, with probability mass function $P(Y_i = j) = p_j$. Let these random variables be **independent of** $(N(s))_{s\in[0,\infty)}$. Let $N_j(t)$ be the arrivals before time t that are of type j:

$$N_j(t) = \sum_{i=1}^{N(t)} 1_{\{Y_i=j\}}.$$

Theorem

$\{(N_j(t))_{t\in[0,\infty)}\}_j$ *are independent Poisson processes with respective rates* λp_j.

Superposition

Theorem

Suppose that $N_1(t), \ldots, N_k(t)$ are independent Poisson processes with rates $\lambda_1, \ldots, \lambda_k$, respectively. Then

$$N(t) = N_1(t) + \cdots + N_k(t),$$

is a Poisson process with rate $\lambda_1 + \cdots + \lambda_k$.

$$N(t) - N(s) = \underbrace{(N_1(t) - N_1(s))}_{\sim \mathrm{Pois}(\lambda_1(t-s))} + \underbrace{(N_2(t) - N_2(s))}_{\sim \mathrm{Pois}(\lambda_2(t-s))} + \cdots + (N_k(t) - N_k(s))$$

Proof.

We simply check the second definition of Poisson process.

$t_1 < t_2 < t_3 < t_4$

1. $N(0) = N_1(0) + \cdots + N_k(0) = 0.$

   $N(t_4) - N(t_3)$ $\quad$ $N(t_2) - N(t_1)$

2. Independent increments follows from that of the other processes
3. Increments are Poisson since the sum of Poisson independent r.v. is Poisson $N(t) - N(s) \sim \mathrm{Pois}(\lambda_1(t-s) + \lambda_2(t-s) + \cdots + \lambda_k(t-s))$

□

Example

Assume people from city A and people from city B arrive at a stadium according to two independent Poisson processes $(N_A(t))_{t=0}^{\infty}$ and $(N_B(t))_{t=0}^{\infty}$ with rates 10 and 20 people per hour, respectively. The ticket seller only cares about how many tickets are sold, he does not care about the buyer comes from city A or B. That is, the ticket seller only cares about the process $\boxed{N(t) = N_A(t) + N_B(t)}$, which corresponds to the total number of tickets sold by time t. By superposition, $(N(t))_{t\in[0,\infty)}$ is a Poisson process with rate 10+20=30 per hour. Hence, for example we have

$$P(N(2) = 4) = e^{-2\cdot 30}\frac{(2\cdot 30)^4}{4!}$$

$$N(2) \sim \text{Pois}(30\cdot 2)$$

Conditioning

Assume that we have had only one arrival up to time t. What is the distribution of the arrival time T_1, given this information? It is uniform on $[0, t]$. NOT DEPENDS ON λ

$s \in [0,t]$

$$P(\underbrace{T_1 > s}_{N(s)=0} | N(t) = 1) = \frac{P(N(s) = 0, N(t) = 1)}{P(N(t) = 1)} = \frac{P(N(s) = 0, N(t) - N(s) = 1)}{P(N(t) = 1)}$$

independent increments $\to$
$$= \frac{P(N(s) = 0)P(N(t) - N(s) = 1)}{P(N(t) = 1)}$$

$$= \frac{\cancel{e^{-\lambda s}}\,\cancel{e^{-\lambda(t-s)}}\,\frac{\lambda(t-s)}{1!}}{\cancel{e^{-\lambda t}}\,\frac{\lambda t}{1!}} = \frac{\cancel{\lambda}(t-s)}{\cancel{\lambda} t}$$

$$= 1 - \frac{s}{t} = P(U > s),$$

$U \sim \text{Unif}[0, t]$

where U is a uniform random variable on $[0, t]$. Hence, T_1 is uniformly distributed on $[0, t]$ given that $N(t) = 1$! No matter λ.

Conditioning

Something more general holds:

Theorem

*Given that $N(t) = n$, the set $\{T_1, T_2, \ldots, T_n\}$ of arrival times in $[0, t]$ is distributed as a set $\{U_1, U_2, \ldots, U_n\}$ of n **independent uniform** random variables on $[0, t]$.*

Corollary

If $s < t$ and $n \geq 0$, then given that $N(t) = n$ the random variable $N(s)$ is distributed as a binomial random variable $\boxed{Bin(n, s/t)}$.

0 s t

$N(t)=n$

$p = \frac{s}{t}$

Proof.

The result follows from the fact that, given that $N(t) = n$, the set of arrival times up to t is distributed as a set of n independent random variables, each one with probability s/t of being smaller than s. □

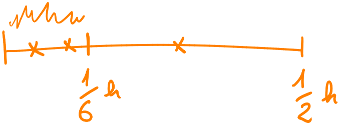
Example

$\frac{1}{2} h$

Assume customers arrive to a shop with rate of 5 customers per hour. Given that exactly 3 customers arrive in the first 30 minutes, what is the probability that exactly 2 customers entered the shop in the first ten minutes? In symbols, what is $P(N(1/6) = 2 | N(1/2) = 3)$?

$$P(N(1/6) = 2 | N(1/2) = 3) = \binom{3}{2} \left(\frac{1/6}{1/2} \right)^2 \left(1 - \frac{1/6}{1/2} \right) = 3 \left(\frac{1}{3} \right)^2 \frac{2}{3} = \frac{2}{9}.$$

$\sim \mathrm{Bin}\left(3, \frac{1/6}{1/2}\right)$

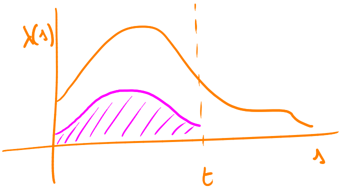


Conditioning and non-homogeneous Poisson process

Theorem

Consider a non-homogeneous Poisson process $\{M(t) : t \geq 0\}$ with rate $\lambda(\cdot)$. Given that $M(t) = n$, the set $\{T_1, T_2, \ldots, T_n\}$ of arrival times in $[0, t]$ is distributed as a set $\{U_1, U_2, \ldots, U_n\}$ of n ***independent*** *random variables on $[0, t]$ with density*

$$\frac{\lambda(\cdot)}{\int_0^t \lambda(s)ds}.$$



3° MODO x SIMULARE

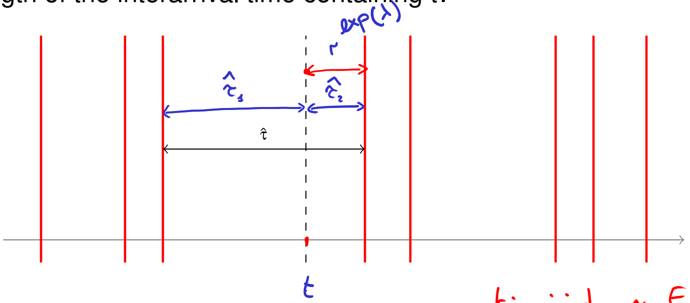
P.P. NON OMOGENEO.

- $M(t) \sim Pois\left(\int_0^t \lambda(s)ds\right)$
- DISTRIBUIRE ARRIVI NON SEMPRE FACILE

The inspection paradox

The lenght of a pinned interarrival interval

Consider a Poisson process with rate λ. Fix a time t different from 0, and let $\hat{\tau}$ be the length of the interarrival time containing t.



t_i i.i.d. $\sim \mathrm{Exp}(\lambda)$

What is $\mathbb{E}[\hat{\tau}]$? $\neq \frac{1}{\lambda}$

$$E[\hat{\tau}] = E[\hat{\tau}_1] + E[\hat{\tau}_2] = E[\hat{\tau}_1] + \frac{1}{\lambda} > \frac{1}{\lambda}$$

The inspection paradox

In general, whenever we have a renewal process with interarrival intervals of length t_i, if we fix a large enough t and define by $\hat{\tau}$ the length of the interarrival time containing t, then

$$\mathbb{E}[\hat{\tau}] > \mathbb{E}[t_i].$$

A MENO CHE I t_i
NON SIANO TUTTI
DETERMINISTICI

Example

Let's make a trivial example: let $(t_i)_{i=1}^{\infty}$ be i.i.d. with

$$t_i = \begin{cases} 1/9 & \text{with probability} 9/10 \\ 10 & \text{with probability} 1/10. \end{cases}$$

Then,

$$\mathbb{E}[t_i] = \frac{1}{9} \cdot \frac{9}{10} + \cancel{10} \cdot \frac{1}{\cancel{10}} = \frac{11}{10}.$$

However, on average the short intervals cover $1/11$ of the whole time (you can use the theory of renewal reward processes to prove it), so we expect our fixed time t is in a long interval with much more probability! As a consequence, if t is large, we expect

$$\mathbb{E}[\hat{\tau}] \approx \frac{1}{9} \cdot \frac{1}{11} + 10 \cdot \frac{10}{11} > \mathbb{E}[t_i].$$

The paradox is everywhere!

Assume that there are two classes, one with 30 students and another one with 10 students. then, the average size of the classes is 20. $(\mathbb{E}(t_i))$

However if you interview a randomly chosen student t, the *perceived* average size is larger: if A is the student's answer then

$$\mathbb{E}[A] = 30 \cdot \frac{30}{40} + 10 \cdot \frac{10}{40} = 25 > 20.$$

interviewing a student is similar to fix a time point t and ask it: how large is the interarrival time containing you? Picking a time point pertaining to a longer interarrival interval is more likely!

When transport companies complain that the trains or airplanes or buses are mostly empty, and passengers complain that they are often uncomfortably crowded, they may be both right! Simply, more people experience the crowd.

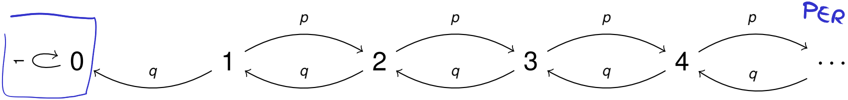
Discrete-time Markov chains

DTMC

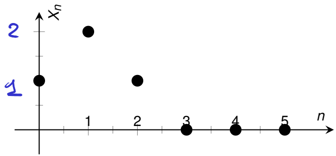
Gambling

Consider a gambler that plays a game until he has money available:

p VITTORIA
$q = 1 - p$ PERDITA



A possible time evolution of the process, given that $X_0 = 1$ is



$x > 0$

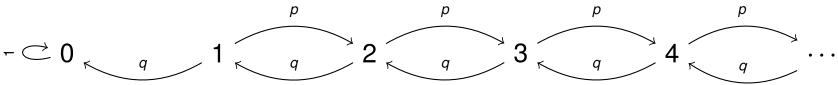
$$P(X_{n+1} = x+1 \mid X_n = x) = p$$

$$P(X_{n+1} = x-1 \mid X_n = x) = q = 1 - p$$

$$P(X_{n+1} = x+1 \mid \underline{X_n = x}, \cancel{X_{n-1} = x-1}, \cancel{X_{n-2} = x}) = p$$

Gambling

Consider a gambler that plays a game until he has money available:



It is clear that for any $i \in \mathbb{Z}_{>0}$

$$
\begin{aligned}
P(X_{n+1} = i+1 | \underline{X_n = i}, \cancel{X_{n-1} = i_{n-1}, \dots, X_0 = i_0})
&= P(X_{n+1} = i+1 | X_n = i) \\
&= p.
\end{aligned}
$$

This is the **Markov property**.

The Markov property

$$X_0 \to X_1 \to X_2 \to X_3 \to \dots$$

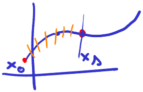
Markov processes are stochastic processes such that the probability distribution of the future observations is completely determined by the present state (or distribution), regardless of any knowledge of the *past* history.

The interest in MP follows two main lines of thought

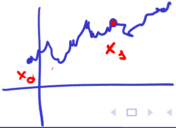
- (math) The original idea by Markov, is that the basic theorems (LLN, CLT...) that holds for a sequence of $\boxed{\text{i.i.d.}}$ random variables may be extended to some sequence of dependent r.v. The easiest kind of dependence that could be introduced is that of the Markov property
- (physics and other applications) It is natural to assume that the dynamics of a physical system that starts with some initial condition (ic) is only determined by such ic, without any need for further informations. cf. ODE

ODE

$$\begin{cases} x(0) = x_0 \\ \dot{x} = f(x) \end{cases}$$



$$\begin{cases} x(0) = x_1 \\ \dot{x} = f(x) \end{cases}$$



The Markov property

Definition

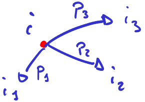
A stochastic process $(X_i)_{i \in I}$ has the Markov property if for any $i, j \in I$ with $j \geq i$

$$X_j \,|\, \mathcal{F}_i \quad \sim \quad X_j \,|\, X_i$$

where $(\mathcal{F}_i)_{i \in I}$ is the natural filtration.

$\mathcal{F}_i = \sigma(X_u,\ u \leq i)$

Markov Property in discrete time



Definition

Let $(X_n)_{n=0}^{\infty}$ be a discrete time stochastic process with a discrete state space S. $(X_n)_{n=0}^{\infty}$ is a Discrete-time Markov chain (**DTMC**) if for any $j, i, i_{n-1}, \ldots, i_0 \in S$,

$$P(X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = P(X_{n+1} = j \mid X_n = i) = p(i,j).$$

$p(i,j)$ are called the (one step) transition probabilities (also called transition matrix)

Note that the transition probabilities are not time (n) dependent. We restrict to this so-called *temporally homogeneous* case.

Transition probabilities

If the state space S is finite (say of cardinality k), the transition probabilities can be organized into a $\boxed{k \times k}$ matrix

$$P = \begin{pmatrix} p(1,1) & \cdots & p(1,k) \\ \vdots & & \vdots \\ p(k,1) & \cdots & p(k,k) \end{pmatrix}$$

with non-negative entries such that the row sums are 1:

$$P(i \to *) = \sum_{j \in S} p(i,j) = 1.$$

SOMMA i-ESIMA RIGA

Matrices of this kind are called *stochastic matrices*.
If S is countably infinite you can still think at a matrix with infinite dimensions.

$$P_{ij} = p(i,j)$$

Consequence 1: joint prob of a trajectory

We can apply the Markov property to compute the probability of observing a trajectory $X_0, X_1 \cdots X_m$. We have

$$
\begin{aligned}
&P(X_n = i_n, X_{n-1} = i_{n-1}, \cdots, X_0 = i_0) = \\
&= P(X_n = i_n \mid X_{n-1} = i_{n-1} \cancel{, \cdots, X_0 = i_0}) \cdot P(X_{n-1} = i_{n-1}, \cdots, X_0 = i_0) =
\end{aligned}
$$

by the Markov property

$$
\begin{aligned}
&= p(i_{n-1}, i_n) \cdot \\
&\quad \cdot P(X_{n-1} = i_{n-1} \mid X_{n-2} = i_{n-2} \cancel{, \cdots, X_0 = i_0}) \cdot P(X_{n-2} = i_{n-2}, \cdots, X_0 = i_0) \\
&= p(i_{n-1}, i_n) \cdot p(i_{n-1}, i_{n-2}) \cdots p(i_0, i_1) \underbrace{\boxed{P(X_0 = i_0)}}_{\alpha(i_0)}
\end{aligned}
$$

To compute the probability of a trajectory it is sufficient to know the transition matrix and the initial distribution $\alpha(i) = P(X_0 = i)$.

Consequence 2: Chapman-Kolmogorov eqs. and n-step transitions

A natural question is how to calculate transition probabilities in two steps, e.g.

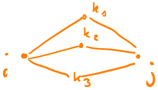
2-STEP TRANSITION PROBABILITY

$$p^{(2)}(i,j) = P(X_{n+2} = j \mid X_n = i) = P(X_2 = j \mid X_0 = i)$$

Again, conditioning on $X_{n+1} = k$ and applying the law of total probabilities, the Markov property, and temporal homogeneity we have

$$\begin{aligned}
p^{(2)}(i,j) =& P(X_{n+2} = j \mid X_n = i) = \\
\xrightarrow{\text{TEO PROB. TOT.}} =& \sum_{k \in S} \underbrace{P(X_{n+2} = j \mid X_{n+1} = k, \cancel{X_n = i})}_{p(k,j)} \cdot \underbrace{P(X_{n+1} = k \mid X_n = i)}_{p(i,k)} \\
=& \boxed{\sum_{k \in S} p(i,k)p(k,j)} = (P^2)_{ij}
\end{aligned}$$

this formula has a nice geometric interpretation, as the sum of the probabilities of all the paths that leads from i to j in two steps.



Consequence 2: Chapman Kolmogorov and n-step transitions

Please note that the formula above can be read by saying that the two steps transition probability matrix $p^{(2)}$ is noting but the matrix product between the transition matrix with itself (the second power of the transition matrix p)

$$p^{(2)} = p \cdot p = p^2$$

$$P^{(2)}_{ij} = p^{(2)}(i,j)$$

More in general, the have that the following Chapman Kolmogorov equations hold

$$p^{(m+n)} = p^{(m)} \cdot p^{(n)} = p^{m+n}$$

that, componentwise, means that

$$p^{(m+n)}(i,j) = \sum_{k \in S} p^{(m)}(i,k)p^{(n)}(k,j)$$

$$\shortparallel$$

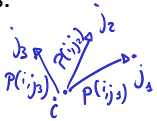
$$P(X_{m+n} = j \mid X_0 = i) = \sum_{k \in S} \underbrace{P(X_{m+n} = j \mid X_m = k, \cancel{X_0 = i})}_{p^{(n)}(k,j)} \underbrace{P(X_m = k \mid X_0 = i)}_{p^{(m)}(i,k)}$$

Summary

A Markov chain is fully specified if we know:

- the state space (somehow implicit in the next requirements)
- the transition probability matrix $p(i, j)$
- the initial distribution $\boxed{\alpha}$

Another way of specifying the chain is through its transition graph that is a graph whose nodes are the states, and the (weighted) edges are arrows that connect two states if the corresponding transition probability is positive. Weights are the transition probabilities.



$\alpha(i) = P(X_0 = i)$

$\alpha = (\alpha(1), \alpha(2), \ldots, \alpha(7))$

Examples

$p(2,2) = 0.8 \quad p(2,1) = 0.2$

$p(1,1) = 0.6 \quad p(1,2) = 0.4$

Ex. 1: Weather forecast

Let us define the following Markov model of weather conditions. Let the state be either rainy (state 1) or sunny (state 2). Let the probability that a sunny day is followed by another sunny day be 0.8, and the probability that a rainy day is followed by another rainy day be 0.6.

$$P = \begin{pmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{pmatrix}$$

Calculate the probability that if today is sunny the day after tomorrow will be sunny again. $p^{(2)}(2,2)$

$$P^2 = \begin{pmatrix} * & * \\ * & \boxed{*} \end{pmatrix} \leftarrow p^{(2)}(2,2)$$

Calculate the probability of observing 4 sunny days in a raw.

$$P(X_0=2, X_1=2, X_2=2, X_3=2) = \alpha(2)\cdot 0.8\cdot 0.8\cdot 0.8 = (0.8)^3\cdot\alpha(2)$$

Given an initial distribution such that of $P(X_0 = 2) = 0.6$, calculate the distribution of X_1.

$\alpha(2) = 0.6 \qquad \alpha = (0.4, 0.6)$

$$P(X_1=1) = P(X_1=1 \mid X_0=1)P(X_0=1) + P(X_1=1 \mid X_0=2)P(X_0=2)$$

$$= 0.6\cdot 0.4 + 0.2\cdot 0.6 = (\alpha\cdot P)_1$$

$$P(X_1=j) = \sum_i p(i,j)\cdot\alpha(i) = \sum_i \alpha(i)\cdot p(i,j) = (\alpha\cdot P)_j$$

$$n-1 \xleftarrow{q} n \xrightarrow{p} n+1$$

Ex. 2: Random walk in 1d

Consider the discrete time Markov chain on $S = \mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$ where for some $0 < p < 1$ we have

$$p(i, i+1) = p, \quad p(i, i-1) = q, \quad \text{with} \quad \boxed{q = 1 - p.}$$

This chain is very important.It is referred to as *one-dimensional random walk*.

Calculate $p^{(k)}(0,0)$ for any integer k.

$$P \leadsto P^k \quad \text{leggo } (P^k)_{(0,0)}$$

$$P^2_{ij} = \sum_{k \in S} p(i,k)\, p(k,j)$$

SERIE! QUINDI UN LIMITE

$$p^{(k)}(0,0) = P(\text{VITTORIE} = \text{SCONFITTE})$$

$$= \begin{cases} 0 & k \text{ DISPARI} \\ P\left(\text{Bin}(k,p) = \frac{k}{2}\right) & k \text{ PARI} \end{cases}$$

Classification of states, recurrence and transience

Let T_y be the time of the first visit to y, without counting X_0.

$y \quad x_1 \quad x_2 \quad y$
$0 \quad 1 \quad 2 \quad 3$

$$T_y = \min\{n \geq 1 : X_n = y\} \quad \text{V.A.}$$

$T_y = 3$

T_y is called **hitting time of** y, and if the chain starts in $X_0 = y$ **return time to** y.
T_y is a random variable expressing how many steps are needed to visit y.
Let

$$\rho_{xy} = P_x(T_y < \infty) = P(\text{we will visit } y | X_0 = x),$$

(EXCLUDING TIME 0)

$P_x(\cdot) = P(\cdot | X_0 = x)$

$E_x[\cdot] = E[\cdot | X_0 = x]$

be the probability of visiting y in a finite time if we start at y.

There are two distinct types of states:

- y is recurrent if $\rho_{yy} = 1$;
- y is transient if $\rho_{yy} < 1$.

Classification of states, recurrence and transience

The names *recurrent* and *transient* are better justified by the following properties:

- Recurrent states are visited infinitely many times.(IF WE START THERE!)
- on the contrary, the number of visits to any transient state is finite. Therefore you will always find a large enough time such that after that time a transient state is never visited any more.

How to prove it?

Return times

Define

$$T_y^{(1)} = T_y$$

and for $k \geq 2$

$$T_y^{(k)} = \begin{cases} \min\{n > T_y^{(k-1)} : X_n = y\} & \text{if } T_y^{(k-1)} < \infty \\ \infty & \text{if } T_y^{(k-1)} = \infty. \end{cases}$$

So, $T_y^{(k)}$ is the time of the $k-$th visit to y (without considering X_0).

Recurrence

Let y be recurrent:

$$P_y(T_y^{(1)} < \infty) = \rho_{yy} = 1$$
$$P_y(T_y^{(2)} < \infty) = \rho_{yy} \cdot \rho_{yy} = 1$$
$$P_y(T_y^{(3)} < \infty) = \rho_{yy} \cdot \rho_{yy} \cdot \rho_{yy} = 1$$
$$\vdots$$
$$P_y(T_y^{(k)} < \infty) = (\rho_{yy})^k = 1$$

Recurrence

Let y be recurrent:

$$P_y(T_y^{(1)} < \infty) = 1$$
$$P_y(T_y^{(2)} < \infty) = 1$$
$$P_y(T_y^{(3)} < \infty) = 1$$

$$P_y(\overbrace{T_y^{(k)} = \infty, \text{ for some } k}^{\#\text{VISITE A } y \text{ FINITO}}) = P_y\left(\bigcup_{k=1}^{\infty}\{T_y^{(k)} = \infty\}\right) \overset{\text{CONTINUITÀ}}{=} \lim_{k\to\infty} \overbrace{P_y(T_y^{(k)} = \infty)}^{1-P_y(T_y^{(k)}<\infty) \;=\; 1-1} = 0.$$

Hence

$$P_y(y \text{ is visited infinitely many times}) = 1 - P_y(T_y^{(k)} = \infty, \text{ for some } k) = 1.$$

Let y be transient:

$$P_y(T_y^{(1)} < \infty) = \rho_{yy} < 1$$
$$P_y(T_y^{(2)} < \infty) = \rho_{yy} \cdot \rho_{yy} = (\rho_{yy})^2$$
$$P_y(T_y^{(3)} < \infty) = (\rho_{yy})^3$$

$$P_y(T_y^{(k)} < \infty) = (\rho_{yy})^k \xrightarrow[k \to \infty]{} 0$$

Transience

Let y be transient:

$$P_y(T_y^{(1)} < \infty) = \rho_{yy} < 1$$
$$P_y(T_y^{(2)} < \infty) = \rho_{yy}^2$$
$$P_y(T_y^{(3)} < \infty) = \rho_{yy}^3$$

Hence, for all $k \geq 1$

$$\underbrace{P_y(y \text{ is visited infinitely many times})}_{=0} \leq P_y(\overbrace{T_y^{(k)} < \infty}^{\text{AT LEAST } k \text{ TIMES}}) = \rho_{yy}^k \xrightarrow[k \to +\infty]{} 0$$

Regeneration

$P(\text{VISITARE DI NUOVO } y \mid X_{\tau_y^{(1)}} = y)$

$= P(\text{VISITARE } y \mid X_0 = y)$

FUNZIONERÀ!??

So, we have

$$P(X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = P(X_1 = j \mid X_0 = i) = \underline{p(i,j)}.$$

Is it true that for a random variable τ

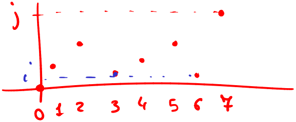
$$P(X_{\tau+1} = j \mid X_\tau = i, X_{\tau-1} = i_{\tau-1}, \ldots, X_0 = i_0) = P(X_1 = j \mid X_0 = i) = p(i,j)?$$

NON SEMPRE

Regeneration

$\tau = 6$

$P(X_{\tau+1} = j \mid X_\tau = i) =$



Let τ be defined as the first time n such that $X_{n+1} = j$. In formulas,

$$\boxed{\tau = \min\{n \geq 0 : X_{n+1} = j\}.}$$

Calculate

$$P(X_{\tau+1} = j \mid X_\tau = i) = 1 \neq P(X_1 = j \mid X_0 = i)$$

$$P(X_{n+1} = j \mid X_n = i) = P(X_1 = j \mid X_0 = i) = p(i,j)$$