Introduction to Markov processes: Markov chains

EM13-Probability and statistics: Courses 15-16

September 2014

Manuel SAMUELIDES[1] Zhigang SU[2]

[1]Professor
Institut Supereur de l'Aeronautique et de l'Espace

[2]Professor
Sino-European Institute of Aviation Engineering
Civil Aviation University of China

Homework–20141104

1. Construction of a Markov chain from a sequence of i.d.d. random variables:

1. Let (X, U) be an independent couple of random vectors. Show the following disjunction principle $E(f(X,U)|X) = \int f(X,u)dP_U(u)$
2. Let (U_n) an i.d.d. sequence of random variables with values in $\mathcal{U}$, X_0 a random variable with values in ε and independent of the (U_n) and $g : \mathcal{U} \times \varepsilon \to \varepsilon$ a measurable function. Set

$$X_n = g(X_{n-1}, U_n)$$

Show that (X_n) is a Markov chain.

2. Finite state space and Markov matrix We define a Markov chain on a four state space $\{1, 2, 3, 4\}$ by the following transition matrix:

$$\Pi = \begin{bmatrix} \frac{1}{3} & a & 0 & 0 \\ b & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{4} & c \\ 0 & \frac{2}{5} & d & \frac{3}{5} \end{bmatrix}$$

1. Compute the necessary values of parameters a, b, c, d.
2. Draw the graph of the Markov chain.
3. Classify according to their dynamics each state.
4. Determine the classes and their nature.
5. Is the Markov chain irreducible (可约)?
6. What is the stationary law of the subchain restricted to states $\{1, 2\}$
7. Compute the law of the first return time T_1 to state 1, and its expectation $E(T_1)$.
8. Compute $p_{11}(n) = \frac{1}{7}[3 + 4(-\frac{1}{6})^n]]$ and $\lim_{n\to\infty} p_{11}(n)$.
9. Use a theorem from the course to check $E(T_1)$ from $\lim_{n\to\infty} p_{11}(n)$.

3. Ehrenfest urn Let a box which is divided into two parts. It contains n balls. At each time a ball is selected randomly and its location is changed. Let X_t be the number of balls in one of the two parts at time t. Shows (X_t) is an ergodic(遍历) Markov chain and compute is invariant law.

4. Random walk on a graph We consider a finite non-oriented graph $\mathcal{G} = (\mathcal{V}, \varepsilon)$ where $\mathcal{V}$ is the set of n vertices and $\varepsilon \subset \mathcal{V} \times \mathcal{V}$ is the set of edges. For each vertex i we consider $\mathcal{V}_\rangle = \{j \in \mathcal{V}$ such that $(j, i) \in \varepsilon$ and $n_i = \sharp(\mathcal{V}_\rangle)$ which is the arity of i.

The graph is connex in the sense that for each couple of vertices there is always a path (sequence of adjacent edges) that connect them.
(Cont.)

A random walk on $\mathcal{G}$ is a Markov chain with transition probability defined by:

$$P(X_{t=1} = j | X_t = i) = \frac{1}{n_i}, \qquad j \in \mathcal{V}_\rangle$$

1. Show that this Markov chain is recurrent positive.
2. Check that $\pi^*(i) = \frac{n_i}{\sum_j n_j}$ is the invariant probability distribution of the chain.
3. Show that this invariant law (equilibrium state in statistical physics) checks the detailed balance property:

$$P(X_t = i, X_{t+1} = j) = P(X_t = j, X_{t+1} = i)$$

4. Show the random walk on a graph is not always ergodic with a very simple example (2 nodes)
5. Show that if at least a node is connected to itself, then the random walk is ergodic.

5. The bankrupt of the player This problem was one of the earliest problem in probability theory in XVIII-th century(AD). We consider two players which are playing one against the other with a constant sum of the two fortunes equal to a. That means that if X_n is the fortune of one player at time n, we have

$$0 < x < a \Rightarrow \begin{cases} P(X_{n+1} = x+1 | X_n = x) = p \\ P(X_{n+1} = x-1 | X_n = x) = 1-p \end{cases}$$

Game is stopping when $X_n = 0$ where our player is bankrupted or when $X_n = a$ where its opponent is bankrupted. Mathematically we shall write

$$P(X_{n+1} = 0 | X_n = 0) = 1$$

and we shall say that 0 and a are absorbing barriers.

1. Show that all the states x such that $0 < x < a$ are transitory
2. Let T be the term of the game. Show that $P(T = n + p|X_n = x) = P(T = p|X_0 = x)$
3. Set $t_x = E(T|X_0 = x)$, prove that

$$t_x = pt_{x+1} + (1-p)t_{x-1}$$

   with the boundary conditions $t_0 = t_a = 0$.
4. If $p = q = 0.5$, check that $t_x = x(x-a)$
5. Set $p_x = P(X_T = a)$. Write the difference equation that governs p_x and the boundary conditions. Check that $p_x = \frac{x}{a}$ if $p = 0.5$.

Example of random walk

- Basically, a random walk is an increasing sum of independent identically distributed random variables with values in a finite-dimensional vector space:

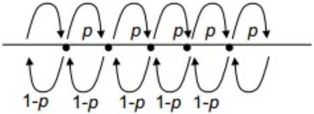
$$Sn = X_1 + \cdots + X_n$$

  or recursively

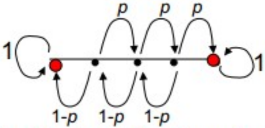
$$S_1 = X_1; \qquad S_n = S_{n-1} + X_n$$

- It is possible to add boundary conditions as absorbing boundary or reflecting boundary as is shown in the following graph.

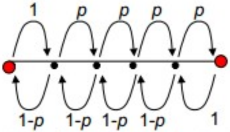
Graphic representations of random walk



(a) without boundary



(b) with absorbing boundaries

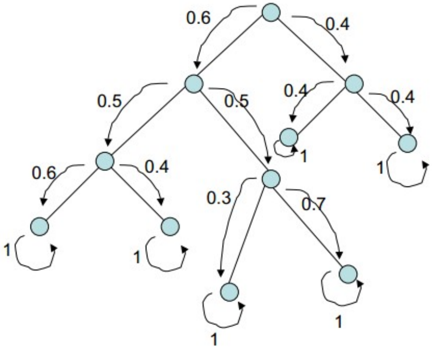


(c) with reflecting boundaries

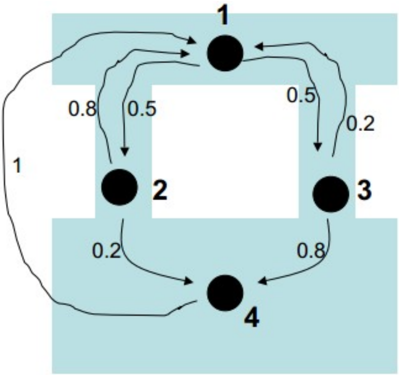
Example of graph transitions

- More generally, we can consider a random exploration on an oriented graph.
- If node i is connected to node j by the link $i \to j$, we endowed this node with a probability p_{ij} of the occurence of this move if the system is in state i
- The sum of all the probabilities from a node has to be equal to one.
- If it is less, then it can be completed by adding the connection $i \to i$
- This representation is currently applied in modelization of reliability, maintenance, networks, web exploration, robotics...

Random walk on a tree



Cartography



Introduction to stochastic processes

- A stochastic process is a measurable application $(t, \omega) \in \mathcal{T} \times \Omega \rightarrow X_t(\omega) \in \epsilon$
- $\mathcal{T}$ is the time set which can be discrete $\mathcal{T} = \mathbb{N}$ or $\mathcal{T} = \mathbb{Z}$ or continuous $\mathcal{T} = [a, b], \mathcal{T} = \mathbb{R}^+$ or $\mathcal{T} = \mathbb{R}$
- ϵ is the state space of the process, it may be discrete (finite or denumerable, or continuous (finite-dimensional vector space, bounded domain, manifold).
- ϵ is endowed with a σ-algebra of measurable events $(\mathcal{P}(\epsilon))$ if ϵ is discrete or the Borel σ-algebra if ϵ is continuous)
- $\Omega \in \epsilon^{\mathcal{T}}$ may be considered as the set of feasible trajectories in the state space. In that case, we have $X_t(\omega) = \omega_t$

Probabilistic model

- Let $(t_1, \ldots, t_n) \in \mathcal{T}^n$ be a finite set of times, then the law of $(X_{t_1}, \ldots, X_{t_n})$ is said a finite probability distribution of the process.
- All the stochastic calculus (微积分) about the process are based on the finite probability distributions which define completely the probability law on Ω (Kolmogorov theorem)
- For $t \in \mathcal{T}$, we define the σ-algebra $\mathcal{F}_t$ of the past at time t as generated by all the random variables X_s for $s \leq t$.
- In advanced control modelling we can consider the sub-σ-algebra of the information available at time t. A strategy is always made of control actions that are measurable for the tribe of available information.
- In the same way, we can define the future at time t.

Definition of a Markov process

Definition

A Markov process(X_t) is a stochastic process such that the conditional law of any future event at time t, given the past at time t is equal to its conditional probability by the present state

$$\forall (t_0 = t < t_1, \ldots, t_n) \in \mathcal{T}^{n+1},$$

$$\mathbb{P}(X_{t_1} \in A_1, \ldots, X_{t_n} \in A_n | \mathcal{F}_t) = \mathbb{P}(X_{t_1} \in A_1, \ldots, X_{t_n} \in A_n | X_t)$$

Markov Chain

A Markov process with discrete time-set is called a Markov chain

Manuel SAMUELIDES, Zhigang SU

中国民航大学

A simple characterization

Theorem

A stochastic process is a Markov process if and only if

$$\forall (t_0 = t < t_1) \in \mathcal{T}^2, \quad \mathbb{P}(X_{t_1} \in A | \mathcal{F}_t) = \mathbb{P}(X_{t_1} \in A | X_t)$$

Proof By recurrence over n. Suppose the property is true till n. Consider

$$\begin{aligned}
&\mathbb{P}(X_{t_1} \in A_1, \ldots, X_{t_n} \in A_n, X_{t_{n+1}} \in A_{n+1} | \mathcal{F}_t) \\
&= \mathbb{P}(X_{t_{n+1}} \in A_{n+1} | X_{t_n} \in A_n) \mathbb{P}(X_{t_1} \in A_1, \ldots, X_{t_n} \in A_n | \mathcal{F}_t) \\
&= \mathbb{P}(X_{t_{n+1}} \in A_{n+1} | X_{t_n} \in A_n) \mathbb{P}(X_{t_1} \in A_1, \ldots, X_{t_n} \in A_n | X_t) \\
&= \mathbb{P}(X_{t_1} \in A_1, \ldots, X_{t_n} \in A_n, X_{t_{n+1}} \in A_{n+1} | X_t)
\end{aligned}$$

Composition of Markov kernels

- From the previous theorem, we infer that the probability law of a Markov process with time set $\mathcal{T} = \mathbb{N}$ or $\mathcal{T} = \mathbb{R}^+$ is completely defined by the law of initial time $\mathbb{P}_{X_0}$ and the stochastic kernel of probability transition from X_s to X_t for any $(s,t) \in \mathcal{T} \times \mathcal{T}$ such that $s < t$
- Such kernel checks: $\forall 0 < s < t < u$

$$\int dP_{X_u}(z|X_t = y)dP_{X_t}(y|X_s = x) = dP_{X_u}(z|X_s = x)$$

Proof

$$\begin{aligned}
&\iint g(x,z)dP_{X_u}(z|X_t = y)dP_{X_t}(y|X_s = x)\\
=&\iint g(x,z)dP_{X_u}(z|X_t = y, X_s = x)dP_{X_t}(y|X_s = x)\\
=&\iint g(x,z)dP_{X_u,X_t}(z,y)|X_s = x)\\
=&\int g(x,z)dP_{X_u}(z|X_s = x)
\end{aligned}$$

A simple characterization

Definition

A Markov process is said time-homogeneous or with stationary transition if for any $s < t$,

$$dP_{X_t}(y|X_s = x) = dP_{X_{t-s}}(y|X_0 = x)$$

Its dynamic is defined by a one-parameter semi-group of probabilistic kernels $t \in \mathcal{T} \to dP_t(y|x)$ checking

$$\int_{\epsilon} dP_s(z|y)dP_t(y|x) = dP_{s+t}(z|x)$$

This equation is called the Chapman-Kolmogorov equations

Examples of Markov semi-groups

When $\mathcal{T} = \mathbb{N}$, we speak about Markov chains. In this course we shall focus on Markov chains with discrete state space. Main examples of Markov semi-groups are

- $\mathcal{T} = \mathbb{N}$, ϵ finite, $P_1(j|i) = p_{ij}$ any Markov matrix.
- $\mathcal{T} = \mathbb{N}$, $\epsilon = \mathbb{N}$, $P_t(j|i) = \mathcal{B}(t, p)$, binomial law for discrete random walk.
- $\mathcal{T} = \mathbb{N}$, $\epsilon = \mathbb{N}$, $P_t = \mathcal{P}(\lambda t)$ Poisson process
- $\mathcal{T} = \mathbb{N}$, $\epsilon = \mathbb{R}$, $P_t = \mathcal{N}(mt, \sigma^2 t)$ Gaussian process

Notations

- From now on, we focus on Markov chains with discrete time $\mathcal{T} = \mathbb{N}$. The states are labelled by integers and letters such as $i, j, k, \ldots$
- The Markov semi-group is defined by its generator for time unit which is a finite or infinite stochastic matrix $\mathbf{P} = (p_{ij}) = \mathbf{P}(j|i)$ such as $\forall (i,j) \geq 0$ and $\forall i$, $\sum_j p_{ij} = 1$.
- $\mathbf{P}$ is called the transition matrix of the chain.
- The matrix $\mathbf{P}^n$ gives the probability of $X_n = j$ given $X_0 = i$. Its general term is noted $p_{ij}^{(n)}$.

Matrix formalism

- The matrix formalism is very useful when the state space is finite and may be extended when it is denumerable.
- The probability distributions on the state space are represented by line vectors with positive coordinates with sum 1. Such vectors are called stochastic vectors.
- The transition matrix is a square matrix, the lines of which are stochastic vectors. Such a matrix is called a stochastic matrix.
- The product of a stochastic vector by a stochastic matrix is a stochastic vector.
- The product of two stochastic matrixes, any power of a stochastic matrix are stochastic matrixes.

Basic matrix manipulation

- If a stochastic vector π is the probability distribution of the chain at time t and if $\mathbf{P}$ is the transition matrix of the chain, then πP is the probability distribution at time $t+1$:

$$\mathbb{P}(X_{t+1} = j) = \sum_i \mathbb{P}(X_{t+1} = j | X_t = i)\mathbb{P}(X_t = i)$$

- With the same notations $\pi \mathbf{P}^n$ is the probability distribution at time $t+n$
- Invariant stochastic vectors such that $\pi \mathbf{P} = \pi$ represent invariant probability distribution by the Markov dynamics.

Examples of bounded random walks

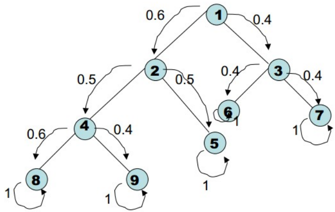
- Random walk with absorbing boundaries

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

- Random walk with reflecting boundaries

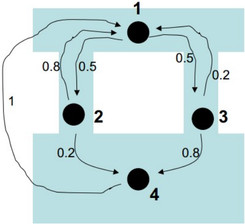
$$\mathbf{P} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1-p & 0 & p & 0 & 0 \\ 0 & 1-p & 0 & p & 0 \\ 0 & 0 & 1-p & 0 & p \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

Example of a random walk on a tree



$$P = \begin{bmatrix} 0 & 0.6 & 0.4 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.5 & 0.5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.2 & 0 & 0 & 0.4 & 0.4 & 0 & 0 \\ 0 & 0 & 0.2 & 0 & 0 & 0 & 0 & 0.6 & 0.4 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Example of cartography



$$P = \begin{bmatrix} 0 & 0.5 & 0.4 & 0 \\ 0.8 & 0 & 0 & 0.2 \\ 0.2 & 0 & 0 & 0.8 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Partition into classes (分类)

Definition

We say that state j is accessible (可达) from state i if there is a path of non null probability which goes from i to j. If j is accessible from i and i is accessible form j, we say that i and j belong to **the same class (同类)**.

Example

- In the examples of infinite random walk and random walk with reflecting boundaries, all the states belong to the same class.
- In the example of random walk with absorbing boundaries, each state of the boundary is alone in its class.

Time of first passage (首次到达时间)

Definition

The time of first passage T_{ij} in state j from i is defined as

$$(T_{ij} = t) = (x_0 = i) \cap (X_1 \neq j) \cdots \cap (X_{n-1} \neq j) \cap (X_n = j)$$

- We have

$$Pr(T_{ij} = 1) = p_{ij}, \qquad Pr(T_{ij} = t+1) = \sum_{k \neq j} Pr(T_{ik} = t) p_{kj}$$

- So we have

$$\mathbf{E}(T_{ij}) = 1 + \sum_{k \neq j} p_{kj} \mathbf{E}(T_{ik})$$

Recurrent (常返) and transient (暂留)

Recurrent (常返) and transient (暂留)

- States i which check $\mathbb{P}(T_{ii} < \infty) = 1$ are called recurrent, other states are called transient.
- If recurrent state i checks $\mathbb{E}(T_{ii}) < \infty$, state i is said positive recurrent (正常返),
- if $\mathbb{E}(T_{ii}) = \infty$, state i is said null recurrent (零常返)
- All the states of one class are of the same type.
- If the state is transient, the number of returns follows a geometric law with finite expectation.

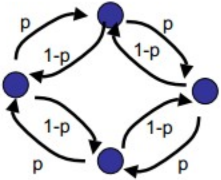
Periodicity and aperiodicity

Definition

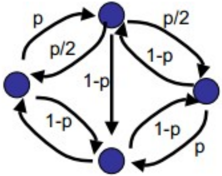
We say that a recurrent state is periodic of period d if it can only reoccur at times that are multiple of d. Otherwise, we say that this state is aperiodic.

- It can be shown that positive recurrence, null recurrence, periodicity and aperiodicity are properties of class (see exercises)
- In the following examples, show the properties pf the chain, write transition matrix, compute the eigenvalues and the stationary distribution. (For numerical applications $p = 0.4$ can be taken)

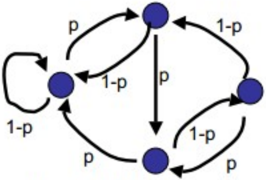
Examples



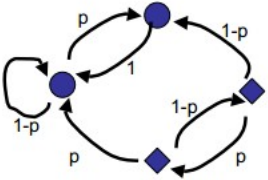
Irreducible recurrent
Period 2 Markov chain



Ergodic Markov chain



Ergodic Markov chain



2 classes: 1 transitory &1 ergodic

Stationary distribution

Definition

Let $X_t \in \epsilon$ be a Markov chain with a transition probability kernel $P : x \in \epsilon \to P_x = P(\cdot|x)$. A probability law π on ϵ is said a stationary distribution if $\int_\epsilon P_x(\cdot)d\pi(x) = \pi(\cdot)$, i.e. if π is the probability law of X_0,it is the probability law of the state of the process X_t at any time $t \in \mathbb{N}$.

- Markov chains have not always stationary distribution, notably null recurrent or transient irreducible chains as the random walk or the binomial counting process have no stationary distribution.
- Positive recurrent Markov chains have a unique invariant measure π^* which checks $\forall x \in \epsilon, \pi^*(x) = \frac{1}{\mathbb{E}(T_{xx})}$

Ergodic Markov chain

Definition

Anergodic Markov chain (X_t) is defined as an irreducible, aperiodic, positive recurrent Markov chain.

As it is positive recurrent, it has a unique stationary distribution π^*. The following law of large numbers is also valid for infinite Markov chain and explains the term ergodic.

Theorem

For any initial probability π_0 on ϵ, for any $f \in L^1(\epsilon, \pi^*)$, $\frac{f(X_0)+f(X_1)+\cdots+f(X_n)}{n+1}$ converges almost surely, in probability and in law towards $\mathbb{E}_{\pi^*}(f(X))$

This theorem is admitted.

Spectral properties of a Markov matrix

Manuel SAMUELIDES, Zhigang SU

Proposition

Any Markov matrix P have 1 as eigenvalue and spectral radius.

Proof Quite simple

$$\sum_j \left|\sum_i \pi_i P_{ij}\right| \leq \sum_i |\pi_i| \sum_i P_{ij} \leq \sum_i |\pi_i|$$

We need to go further and to extract positive eigenvectors. It is possible form the Perron-Frobenius theorem. We prefer to restrict to a more specific class of Markov matrix.

Spectral theorem of a regular Markov matrix

Definition

Aregular Markov matrixis a (finite) Markov matrix P such that there exists a power P^n of P where all the components are strictly positive.

Theorem

A regular Markov matrix admits a unique invariant stochastic vector (positive coordinates) π_∞. Moreover when $n \to \infty$ P^n converges to a matrix where all the lines are equal to π_∞

This theorem is admitted.

Definition of stopping times

Definition

A stopping time $T \in T$ of a stochastic process $(X_t)_{t \in \mathcal{T}}$ is a random variable which takes its values in the time set and such that $\forall t \in \mathcal{T}, (T \leq t) \in \mathcal{F}_t$

- Typically, the first (or successive) access time to a subset, boudary... is a stopping time. For instance T_{ij} is a stopping time.
- When the time set is discrete, it enough to check that $\forall t, (T = t) \in \mathcal{F}_t$. The information contained in the past of t is enough to decide that $T = t$.

Strong Markov property

The Markov property is also checked for stopping times by a Markov process and is calledstrong Markov property

Theorem

Let T a stopping time for an homogeneous Markov process (X_t).
Then

$$P(X_{T+t} \in A | X_T = x_0, X_{T-1} = x_{-1}, \ldots, X_{T-n} = x_{-n})$$
$$= P(X_t \in A | X_0 = x_0)$$

Proof of the strong Markov property

$$
\begin{aligned}
& P(X_{T+t} \in A | X_T = x_0, X_{T-1} = x_{-1}, \ldots, X_{T-n} = x_{-n}) \\
= & \sum_p P(X_{T+t} \in A \cap (T = P) | X_T = x_0, X_{T-1} = x_{-1}, \ldots, X_{T-n} = x_{-n}) \\
= & \sum_p P(X_{T+t} \in A | T = P, X_T = x_0, X_{T-1} = x_{-1}, \ldots, X_{T-n} = x_{-n}) \\
& P(T = P | X_T = x_0, X_{T-1} = x_{-1}, \ldots, X_{T-n} = x_{-n}) \\
= & P(X_p \in A | X_0 = x_0)
\end{aligned}
$$

Corollary

Let (X_t) be a Markov process and T a stopping time, then the process $Y_t = X_{T \wedge t}$ is a Markov process which is adapted at the filtration $\mathcal{F}_t of (X_t)$, i.e. for any $t Y_t$ is $\mathcal{F}_t$-measurable.

Summary

- We have defined the model of Markov process which is the stochastic version of the state representation of dynamic systems in engineering.
- We have defined the model of Markov process which is the stochastic version of the state representation of dynamic systems in engineering.
- In the case of finite state space, it amounts to linear algebra computation.
- A classification of Markov dynamics has been produced with regards to large time limit.
- The hypothesis for a large time stationary limit independent of inital conditions have been investigated.