3 Discrete-time Markov chains

In this section we shall study processes of the form

$$(X_n)_{n\in\mathbb{N}_0} \qquad (\mathbb{N}_0 = \{0,1,2,\dots\})$$

on a prob. space $(\Omega, \mathcal{F}, P)$ such that for all $n \in \mathbb{N}_0$

(i) X_n takes values in a countable state space $\mathcal{I}$, i.e. $\mathcal{I} = \{\gamma_1, \gamma_2, \gamma_3, \dots\}$ as e.g. $\mathcal{I} = \mathbb{N}_0$

(ii) X_n is a r.v. in the sense that X_n is a function

$$X_n : \Omega \to \mathcal{I}$$

(Choose e.g. $\mathcal{F} = \mathcal{P}(\Omega)$ class of all subsets of Ω)

See Appl. 6 in Sect. 1 (branching proc.) as an example of such a proc.

3.1 Some basic definitions and concepts

A sequence $(\lambda_i)_{i\in\mathcal{I}}$ is called distribution, if $0 \le \lambda_i < \infty$ for all $i \in \mathcal{I}$ and

$$\sum_{i\in\mathcal{I}} \lambda_i = 1$$

Ex.: $\lambda_i := P(\{X_n = i\})$ for a fixed n, since

$$1 = P(\{X_n \in \mathcal{I}\}) = P\Big(\dot{\bigcup_{i\in\mathcal{I}}} \{X_n = i\}\Big) \overset{\text{Def 2.2(ii)}}{=} \sum_{i\in\mathcal{I}} P(\{X_n = i\}) = \sum_{i\in\mathcal{I}} \lambda_i$$

($\dot{\cup}$: disjoint)

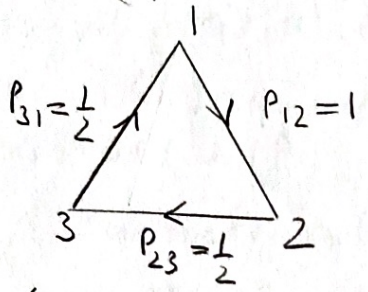
A matrix $P = (p_{ij})_{i,j\in\mathcal{I}}$ is called stochastic $\iff$ $\forall i \in \mathcal{I}$:

$$\lambda_j := p_{ij} \text{ is a distribution}$$

→ Interpretation: p_{ij} the probability for jumping from state i to state j, e.g. $S_n = i$ (S_n stock price at time n) jumps to $S_{n+1} = j$?

Ex.: $P = \begin{pmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & & \\ p_{31} & \cdots & \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}$, $J = \{1,2,3\}$

→ diagram:



Def. 3.1.1 (Conditional probability)
Let A, B events with $P(B) > 0$.
Then

$$P(A|B) := \frac{P(A \cap B)}{P(B)}$$

is called the conditional probability of A given B

Ex.: $A = \{S_{n+1} = j\}$ (stock price at time $n+1$), $B = \{S_n = i\}$

→ $P(\{S_{n+1} = j\} | \{S_n = i\})$ or $P(S_{n+1} = j | S_n = i)$ for short is the prob. that $S_{n+1} = j$ given that $S_n = i$

Rem.: $\tilde{P}(A) := P(A|B)$ defines a probability on $\tilde{\Omega} := B$ $(\subseteq \Omega)$

Def. 3.1.2 (Markov chain)
Let $P = (p_{ij})_{i,j \in J}$ be a stoch. matrix and $\lambda = (\lambda_i)_{i \in J}$ a distr.
A proc. $(X_n)_{n \geq 0}$ is called a Markov chain with initial distribution λ and transition matrix P (or Markov(λ, P) for short) if for all $n \geq 0$, $i_0, \ldots, i_{n+1} \in J$:

(i) $P(X_0 = i_0) = \lambda_{i_0}$

(ii) $P(X_{n+1} = i_{n+1} | X_n = i_n) = p_{i_n i_{n+1}}$

(iii) $P(X_{n+1} = i_{n+1} | X_0 = i_0, \ldots, X_n = i_n) = p_{i_n i_{n+1}}$
$\left(\overset{(ii)}{=} P(X_{n+1} = i_{n+1} | X_n = i_n)\right)$

(i), (ii), (iii) → X_{n+1} just remembers the last position $X_n = i_n$

$\longrightarrow$ $(X_n)_{n\geq 0}$ is a proc. without "memory".

Rem.: If $(X_n)_{0\leq n\leq N}$ is a finite seq. of r.v.'s satisfying (i), (ii), (iii) for $n = 0, \dots, N-1$ then $(X_n)_{0\leq n\leq N}$ is also called Markov (λ, P)

$\longrightarrow$ Th. 3.1.3 (Characterization of finite Markov chains)

$(X_n)_{0\leq n\leq N}$ Markov (λ, P) $\Longleftrightarrow$

$\forall\, i_0, \dots, i_N \in I$:

$$P(X_0 = i_0, X_1 = i_1, \dots, X_N = i_N) = \lambda_{i_0} p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{N-1} i_N} \qquad (3.1.1)$$

Proof: "$\Longrightarrow$": $P(X_0 = i_0, X_1 = i_1, \dots, X_N = i_N)$

$= P(X_0 = i_0) \cdot \frac{P(X_0 = i_0, X_1 = i_1)}{P(X_0 = i_0)} \cdot \frac{P(X_0 = i_0, X_1 = i_1, X_2 = i_2)}{P(X_0 = i_0, X_1 = i_1)} \cdots \frac{P(X_0 = i_0, X_1 = i_1, \dots, X_N = i_N)}{P(X_0 = i_0, X_1 = i_1, \dots, X_{N-1} = i_{N-1})}$

Def. 3.1.1

$P(X_0 = i_0) \cdot P(X_1 = i_1 \mid X_0 = i_0) \cdot P(X_2 = i_2 \mid X_0 = i_0, X_1 = i_1) \cdots P(X_N = i_N \mid X_0 = i_0, \dots, X_{N-1} = i_{N-1})$

Def. 3.1.2 (Markov chain)

$\lambda_{i_0} p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{N-1} i_N}$

"$\Longleftarrow$": We have to verify Def. 3.1.2.

We know that

$$P(X_0 = i_0, \dots, X_N = i_N) = \lambda_{i_0} p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{N-1} i_N} \qquad (*)$$

for all $i_0, i_1, \dots, i_N \in I$

$(*) \Longrightarrow \underbrace{\sum_{i_N \in I} P(X_0 = i_0, X_1 = i_1, \dots, X_N = i_N)}_{=: A}$

$= \underbrace{\sum_{i_N \in I} \lambda_{i_0} p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{N-1} i_N}}_{=: B}$

Note: $\bigcup_{i_N \in I} \{X_0 = i_0, \dots, X_{N-1} = i_{N-1}, X_N = i_N\} =$ (disjoint)

(21) $= \{X_0 = i_0, \dots, X_{N-1} = i_{N-1}\}$

$\Rightarrow A \overset{\text{Def. 2.2 (ii)}}{=} P\left(\bigcup_{i_N \in I} \{X_0 = i_0, \dots, X_N = i_N\}\right)$

$= P(\{X_0 = i_0, \dots, X_{i_{N-1}} = i_{N-1}\})$

On the other hand

$B = \lambda_{i_0} \cdot p_{i_0 i_1} \cdot p_{i_1 i_2} \cdots p_{i_{N-2} i_{N-1}} \underbrace{\sum_{i_N \in I} p_{i_{N-1} i_N}}_{=1 \text{ since } P=(p_{ij}) \text{ stoch. matrix}}$

$\Rightarrow P(X_0 = i_0, \dots, X_{N-1} = i_{N-1})$

$= \lambda_{i_0} \cdot p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{N-2} i_{N-1}}$

Using the same procedure applied to $i_{N-1}, \dots, i_1$ we get

$P(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) = \lambda_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} i_n}$ (**)

for all $n = 0, \dots, N$

$\Rightarrow \forall\, n = 0, \dots, N-1$:

$P(X_{n+1} = i_{n+1} \mid X_0 = i_0, \dots, X_n = i_n)$

$\overset{\text{Def. 3.1.1}}{=} \dfrac{P(X_0 = i_0, \dots, X_n = i_n, X_{n+1} = i_{n+1})}{P(X_0 = i_0, \dots, X_n = i_n)}$

$\overset{(**)}{=} \dfrac{\lambda_0 \cdot p_{i_0 i_1} \cdots p_{i_{n-1} i_n} \cdot p_{i_n i_{n+1}}}{\lambda_0 \cdot p_{i_0 i_1} \cdots p_{i_{n-1} i_n}} = p_{i_n i_{n+1}}$ ($\Rightarrow$ Def. 3.1.2 (iii))

Similarly, we use (**) to show that

$P(X_{n+1} = i_{n+1} \mid X_n = i_n) = p_{i_n i_{n+1}}$ ($\Rightarrow$ Def. 3.1.2 (ii))

$P(X_0 = i_0) = \lambda_{i_0}$ by (**) for $n = 0$ ($\Rightarrow$ Def 3.1.2 (i))

$\Rightarrow$ Proof

convention: Let $X_0, \dots, X_m \in \mathcal{I}$ be r.v.'s
An event of the type
$$A := \{X_0 = i_0, \dots, X_m = i_m\}$$
is called an elementary event
We say an event B is determined by $X_0, \dots, X_m$ if
$$B = \dot{\bigcup_{k \geq 1}} A_k \quad (3.1.2)$$
($\dot{\cup}$: disjoint)
where $A_k, k \geq 1$ are elementary events
For all $i \in \mathcal{I}$ def. the distr. $\delta_i = (\delta_{ij})_{j \in \mathcal{I}}$ by
$$\delta_{ij} = \begin{cases} 1, & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \quad (3.1.3)$$

Th. 3.1.4 (Markov property)
Let $(X_n)_{n \geq 0}$ be Markov(λ, P). Then the time-shifted proc. $\tilde{X}_n := X_{n+m}, n \geq 0$ is Markov(δ_i, P) with respect to the (conditional) probability $\tilde{P}(B) := P(B \mid X_m = i)$
Further the Markov chain $(\tilde{X}_n)_{n \geq 0}$ is independent of events A determined by $X_0, \dots, X_m$.

Proof: Let A be determined by $X_0, \dots, X_m$.
We have to verify that
$$\tilde{P}(\{\tilde{X}_0 = i_{m_0}, \dots, \tilde{X}_n = i_{m+n}\} \cap A) \quad (3.1.4)$$
($\tilde{X}_0 \hat{=} X_m$, $\tilde{X}_n \hat{=} X_{m+n}$)
$$= \tilde{\lambda}_{i_0} \cdot \tilde{P}_{i_0 i_1} \cdots \tilde{P}_{i_{n-1} i_n} \cdot \tilde{P}(A)$$
where $\tilde{\lambda}_{i_0}$: init. distr. of $\tilde{X}_n$, $\tilde{P}_{i_{j-1} i_j}$: transition prob. of $\tilde{X}_n$
$\tilde{\lambda}_{i_0} = \delta_{i i_m}$ and $\tilde{P}_{i_{j-1} i_j} := P_{i_{m+j-1} i_{m+j}}, i_j := i_{m+j}$
Because of Def. 2.2 (ii) applied to $\tilde{P}$ and (3.1.2) we can assume w.l.o.g that
$$A = \{X_0 = i_0, \dots, X_m = i_m\} \quad (*)$$
$$\Rightarrow \tilde{P}(\{\tilde{X}_0 = i_m, \dots, \tilde{X}_n = i_{m+n}\} \cap A)$$
$$\overset{\text{def.}}{=} P(\{X_m = i_m, \dots, X_{m+n} = i_{m+n}\} \cap A \mid X_m = i)$$
$$\overset{\text{def. of cond. prob. and } (*)}{=} P(X_0 = i_0, \dots, X_{m+n} = i_{m+n} \text{ and } i = i_m) \Big/ P(X_m = i)$$

Th. 3.1.3 $(\delta_{i_0 i_m} \cdot \underbrace{\lambda_{i_0} p_{i_0 i_1} \cdots p_{i_{m-1} i_m}}_{\overset{\text{Th. 3.1.3}}{=} P(X_0 = i_0, \ldots, X_m = i_m)} \cdot p_{i_m i_{m+1}} \cdots p_{i_{m+n-1} i_{m+n}}) / P(X_m = i)$

$= P(A)$

$= \underbrace{\delta_{i_0 i_m} p_{i_m i_{m+1}} \cdots p_{i_{m+n-1} i_{m+n}}}_{\overset{\text{def}}{=} \tilde{\lambda}_{i_0} \tilde{p}_{i_0 i_1} \cdots \tilde{p}_{i_{n-1} i_n}} \cdot \underbrace{P(A)/P(X_m = i)}_{= \tilde{P}(A)}$

$\Rightarrow$ (3.1.4)

If $A = \{X_m = i\}$ in (3.1.4) $\overset{\text{Th. 3.1.3}}{\Longrightarrow}$ $(\tilde{X}_j)_{0 \le j \le n}$ finite Markov chain for all n (w.r.t. $\tilde{P}$)

$\Rightarrow (\tilde{X}_j)_{j \ge 0}$ Markov chain

The second statement of the Theorem also follows from (3.1.4) $\Longrightarrow$ proof

$(X_n)_{n \ge 0}$ Markov chain

$\longrightarrow$ Problem: $P(X_n = j)$?

Notation: $\lambda = (\lambda_i)_{i \in I}$ distr., $P = (p_{ij})_{i,j \in I}$ stoch. matrix

Def. the new distr. λP by

$$(\lambda P)_j := \sum_{i \in I} \lambda_i p_{ij} \qquad (3.1.5)$$

(j-th comp. of λP)

Def. the new matrix P^2 by

$$(P^2)_{ik} := \sum_{j \in I} p_{ij} p_{jk} \qquad (3.1.6)$$

(entry of P^2)

$\longrightarrow$ usual matrix multiplication if e.g. $I = \{1, \ldots, N\}$

P^n def. similarly ($P^0 := (\delta_{ij})_{i,j \in I}$)

Def. the prob. P_i by

$$P_i(A) := P(A \mid X_0 = i) \qquad (3.1.7)$$

Th. 3.1.5 ((Computation of $P(X_n = j)$))

$(X_n)_{n \ge 0}$ Markov(λ, P) $\Longrightarrow$

(i) $P(X_n = j) = (\lambda P^n)_j$

(ii) $P_i(X_n = j) = P(X_{n+m} = j \mid X_m = i) = p_{ij}^{(n)}$,

where

$$p_{ij}^{(n)} := (P^n)_{ij}$$

is called the n-step transition probability from i to j

Proof: (i) Note that

(24) $\{X_n = j\} = \dot{\bigcup}_{i_0,\dots,i_{n-1} \in I} \{X_0 = i_0, \dots, X_{n-1} = i_{n-1}, X_n = j\}$ (disjoint)

Def. 2.2.(ii) $\Rightarrow$ $P(X_n = j) = \sum_{i_0,\dots,i_{n-1} \in I} P(X_0 = i_0, \dots, X_{n-1} = i_{n-1}, X_n = j)$

Th. 3.1.3 $= \sum_{i_0,\dots,i_{n-1} \in I} \lambda_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} j} \overset{(3.1.5),(3.1.6)}{=} (\lambda p^n)_j$

(ii) Apply (i) to the Markov chain $\widetilde{X}_n := X_{m+n}$ (Th. 3.1.4) with $\lambda = \delta_i$ (see (3.1.3)) $\Rightarrow$ proof.

Ex. 3.1.6 (Virus mutation)

Denote by $1, \dots, N$ N diff. strains (stammer) of a virus! Consider a virus, which belongs to strain 1 in generation $n=0$, i.e. $X_0 = 1$

Assumption: The virus mutates from strain i in gen. n to another strain j in gen. $n+1$ with prob.

$$p_{ij} = \frac{\alpha}{N-1} \quad (i \neq j) \qquad (\geq 0)$$

$$(\Rightarrow \quad p_{ii} = 1 - (N-1) \cdot \frac{\alpha}{(N-1)} = 1 - \alpha)$$

Denote by X_n the strain of the virus in gen. n with $X_0 = 1$

$\rightarrow$ Possible model for $(X_n)_{n \geq 0}$: Markov (δ_1, P) with P as above

$\rightarrow$ Problem: $P(X_n = 1)$? $\overset{\text{Th. 3.1.5}}{=} p_{11}^{(n)}$ (See Exerc. 3)