A brief introduction to Markov chains

Introduction to Markov chains

- We have a finite set of states, $S = \{1, 2, \ldots, n\}$
  - $s_t = k$ means that the process, or system, is in state k at time t
- Example:
  - We take as states the kind of weather R (rain), N (nice), and S (snow)

$$\mathbf{P} = \begin{array}{c c} & \begin{array}{ccc} \mathrm{R} & \mathrm{N} & \mathrm{S} \end{array} \\ \begin{array}{c} \mathrm{R} \\ \mathrm{N} \\ \mathrm{S} \end{array} & \begin{pmatrix} 1/2 & 1/4 & 1/4 \\ 1/2 & 0 & 1/2 \\ 1/4 & 1/4 & 1/2 \end{pmatrix} \end{array}$$

Introduction to Markov chains

- Markov chains are models of sequential discrete-time and discrete-state stochastic processes

Introduction to Markov chains

- The entries of this matrix $\mathbf{P}$ are p_{ij} with

$$p_{ij} = \mathrm{P}(s_{t+1} = j | s_t = i)$$

- We assume that the probability of jumping to a state only depends on the current state
  - And not on the past, before this state
  - This is the Markov property

Introduction to Markov chains

- The matrix $\mathbf{P}$ is called the one-step transition probabilities matrix
  - It is a stochastic matrix
  - The row sums are equal to 1
- We also assume that these transition probabilities are stationary
  - That is, independent of time
- Let us now compute $\mathrm{P}(s_{t+2} = j \mid s_t = i)$:

5

Introduction to Markov chains

$$\begin{aligned}
\mathrm{P}(s_{t+2}=j|s_t=i) &= \sum_{k=1}^{n} \mathrm{P}(s_{t+2}=j, s_{t+1}=k|s_t=i) \\
&= \sum_{k=1}^{n} \mathrm{P}(s_{t+2}=j|s_t=i, s_{t+1}=k)\mathrm{P}(s_{t+1}=k|s_t=i) \\
&= \sum_{k=1}^{n} \mathrm{P}(s_{t+2}=j|s_{t+1}=k)\mathrm{P}(s_{t+1}=k|s_t=i) \\
&= \sum_{k=1}^{n} p_{kj}p_{ik} \\
&= [\mathbf{P}^2]_{ij}
\end{aligned}$$

6

Introduction to Markov chains

- The matrix $\mathbf{P}^2$ is the two-steps transition probabilities matrix
- By induction, $\mathbf{P}^\tau$ is the τ-steps transition probabilities matrix containing elements

$$p_{ij}^{(\tau)} = \mathrm{P}(s_{t+\tau} = j | s_t = i)$$

7

Introduction to Markov chains

- If $\mathbf{x}(t)$ is the column vector containing the probability distribution of finding the process in each state of the Markov chain at time step t, we have

$$\begin{aligned} x_i(t) &= \mathrm{P}(s_t = i) \\ &= \sum_{k=1}^{n} \mathrm{P}(s_t = i, s_{t-1} = k) \\ &= \sum_{k=1}^{n} \mathrm{P}(s_t = i | s_{t-1} = k)\mathrm{P}(s_{t-1} = k) \\ &= \sum_{k=1}^{n} p_{ki}\, x_k(t-1) \end{aligned}$$

8

Introduction to Markov chains

- In matrix form, $\mathbf{x}(t) = \mathbf{P}^{\mathrm{T}}\,\mathbf{x}(t-1)$
- Or, in function of the initial distribution,
  $\mathbf{x}(t) = (\mathbf{P}^{\mathrm{T}})^{t}\,\mathbf{x}(0)$
- Now, $x_j(t) = \mathbf{x}(t)^{\mathrm{T}}\,\mathbf{e}_j = \mathbf{x}(0)^{\mathrm{T}}\,\mathbf{P}^{t}\,\mathbf{e}_j$

9

Introduction to Markov chains

- Thus, when starting from state i,
  $\mathbf{x}(0) = \mathbf{e}_i$, and

  $x_j(t \mid s_0 = i) = x_{j|i}(t) = (\mathbf{e}_i)^{\mathrm{T}}\,\mathbf{P}^{t}\,\mathbf{e}_j$
- It is the probability of observing the process in state j at time t when starting from state i at time $t = 0$

10

Introduction to Markov chains

$$\mathbf{P}^1 = \begin{array}{c} \\ \text{Rain} \\ \text{Nice} \\ \text{Snow} \end{array} \begin{array}{c} \begin{array}{ccc} \text{Rain} & \text{Nice} & \text{Snow} \end{array} \\ \begin{pmatrix} .500 & .250 & .250 \\ .500 & .000 & .500 \\ .250 & .250 & .500 \end{pmatrix} \end{array}$$

$$\mathbf{P}^2 = \begin{array}{c} \\ \text{Rain} \\ \text{Nice} \\ \text{Snow} \end{array} \begin{array}{c} \begin{array}{ccc} \text{Rain} & \text{Nice} & \text{Snow} \end{array} \\ \begin{pmatrix} .438 & .188 & .375 \\ .375 & .250 & .375 \\ .375 & .188 & .438 \end{pmatrix} \end{array}$$

$$\mathbf{P}^3 = \begin{array}{c} \\ \text{Rain} \\ \text{Nice} \\ \text{Snow} \end{array} \begin{array}{c} \begin{array}{ccc} \text{Rain} & \text{Nice} & \text{Snow} \end{array} \\ \begin{pmatrix} .406 & .203 & .391 \\ .406 & .188 & .406 \\ .391 & .203 & .406 \end{pmatrix} \end{array}$$

$$\mathbf{P}^4 = \begin{array}{c} \\ \text{Rain} \\ \text{Nice} \\ \text{Snow} \end{array} \begin{array}{c} \begin{array}{ccc} \text{Rain} & \text{Nice} & \text{Snow} \end{array} \\ \begin{pmatrix} .402 & .199 & .398 \\ .398 & .203 & .398 \\ .398 & .199 & .402 \end{pmatrix} \end{array}$$

$$\mathbf{P}^5 = \begin{array}{c} \\ \text{Rain} \\ \text{Nice} \\ \text{Snow} \end{array} \begin{array}{c} \begin{array}{ccc} \text{Rain} & \text{Nice} & \text{Snow} \end{array} \\ \begin{pmatrix} .400 & .200 & .399 \\ .400 & .199 & .400 \\ .399 & .200 & .400 \end{pmatrix} \end{array}$$

$$\mathbf{P}^6 = \begin{array}{c} \\ \text{Rain} \\ \text{Nice} \\ \text{Snow} \end{array} \begin{array}{c} \begin{array}{ccc} \text{Rain} & \text{Nice} & \text{Snow} \end{array} \\ \begin{pmatrix} .400 & .200 & .400 \\ .400 & .200 & .400 \\ .400 & .200 & .400 \end{pmatrix} \end{array}$$

11

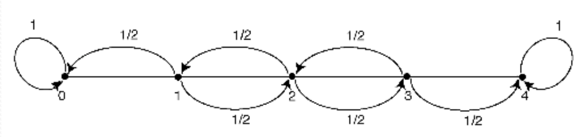
Introduction to Markov chains

- Let us now introduce absorbing Markov chains
- A state i of a Markov chain is called absorbing if it is impossible to leave it, $p_{ii} = 1$
- An absorbing Markov chain is a Markov chain containing absorbing states
  - The other states being called transient (TR)

12

Introduction to Markov chains

- Let us take an example: the drunkard' walk (from Grinstead and Snell)



$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{ccccc} 0 & 1 & 2 & 3 & 4 \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array}$$

13

Introduction to Markov chains

- The transition matrix can be put in canonical form:

$$\mathbf{P} = \begin{array}{c} \\ \text{TR.} \\ \\ \text{ABS.} \end{array} \begin{array}{c} \begin{array}{cc} \text{TR.} & \text{ABS.} \end{array} \\ \left(\begin{array}{c|c} \mathbf{Q} & \mathbf{R} \\ \hline \mathbf{0} & \mathbf{I} \end{array}\right) \end{array}$$

- $\mathbf{Q}$ is the transition matrix between transient states

14

Introduction to Markov chains

- $\mathbf{R}$ is the transition matrix between transient and absorbing states
- Both $\mathbf{Q}$ and $\mathbf{R}$ are sub-stochastic
  - Their row sums are ≤ 1 and at least one row sum is < 1

15

Introduction to Markov chains

- Now, $\mathbf{P}^t$ can be computed as

$$\mathbf{P}^t = \begin{array}{c c} & \begin{array}{cc} \text{TR.} & \text{ABS.} \end{array} \\ \begin{array}{c} \text{TR.} \\ \text{ABS.} \end{array} & \left(\begin{array}{c|c} \mathbf{Q}^t & \cdots \\ \hline \mathbf{0} & \mathbf{I} \end{array}\right) \end{array}$$

- Since $\mathbf{Q}$ is sub-stochastic, it can be shown that:

$$\mathbf{Q}^t \to \mathbf{0} \text{ as } t \to \infty$$

16

Introduction to Markov chains

- The matrix

$$\mathbf{N} = \mathbf{I} + \mathbf{Q} + \mathbf{Q}^2 + \mathbf{Q}^3 + \cdots = (\mathbf{I} - \mathbf{Q})^{-1}$$

  - is called the fundamental matrix of the absorbing Markov chain

- Let us interpret the elements $n_{ij} = [\mathbf{N}]_{ij}$ of the fundamental matrix, where i, j are transient states

17

Introduction to Markov chains

- Recall that since i, j are transient states,
  - we have $x_j(t \mid s_0 = i) = x_{j|i}(t) = (\mathbf{e}_i)^{\mathrm{T}} \mathbf{P}^t \mathbf{e}_j = (\mathbf{e}_i)^{\mathrm{T}} \mathbf{Q}^t \mathbf{e}_j$
- Thus, entry i, j of the matrix $\mathbf{N}$ for transient states, n_{ij}, is:

$$\begin{aligned} n_{ij} &= \mathbf{e}_i^{\mathrm{T}} \mathbf{N} \mathbf{e}_j \\ &= \mathbf{e}_i^{\mathrm{T}} \left(\sum_{t=0}^{\infty} \mathbf{Q}^t \right) \mathbf{e}_j \\ &= \sum_{t=0}^{\infty} \left(\mathbf{e}_i^{\mathrm{T}} \mathbf{Q}^t \mathbf{e}_j \right) \\ &= \sum_{t=0}^{\infty} x_{j|i}(t) \end{aligned}$$

18

Introduction to Markov chains

- Thus, element n_{ij} contains the expected number of passages, or visits, through transient state j when starting from transient state i
- The expected number of visits (and therefore steps) before being absorbed when starting from each state is

$$\mathbf{n} = \mathbf{N}\mathbf{e}$$

19

Introduction to Markov chains

- For the drunkard's walk,

$$\mathbf{P} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 0 \\ 4 \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 0 & 4 \end{array} \\ \left(\begin{array}{ccc|cc} 0 & 1/2 & 0 & 1/2 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 1/2 \\ \hline 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{array} \right) \end{array} .$$

20

Introduction to Markov chains

$$\mathbf{Q} = \begin{pmatrix} 0 & 1/2 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1/2 & 0 \end{pmatrix}$$

$$\mathbf{I} - \mathbf{Q} = \begin{pmatrix} 1 & -1/2 & 0 \\ -1/2 & 1 & -1/2 \\ 0 & -1/2 & 1 \end{pmatrix}$$

$$\mathbf{N} = (\mathbf{I} - \mathbf{Q})^{-1} = \begin{array}{c c} & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \begin{pmatrix} 3/2 & 1 & 1/2 \\ 1 & 2 & 1 \\ 1/2 & 1 & 3/2 \end{pmatrix} \end{array}$$

21

Introduction to Markov chains

$$\begin{aligned} \mathbf{n} = \mathbf{N}\mathbf{e} &= \begin{pmatrix} 3/2 & 1 & 1/2 \\ 1 & 2 & 1 \\ 1/2 & 1 & 3/2 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} 3 \\ 4 \\ 3 \end{pmatrix} \end{aligned}$$

22

Introduction to Markov chains

- We can also compute absorption probabilities from each starting state
- We compute the probability of being absorbed by absorbing state j given that we started in transient state i by

$$b_{ij} = \sum_{t=0}^{\infty} \sum_{\substack{k=1 \\ k \in \mathrm{TR}}}^{n_{tr}} x_{k|i}(t)\, r_{kj}$$

where n_{tr} is the number of transient states and the sum over k is taken on the set of transient states (TR) only

23

Introduction to Markov chains

- Indeed, since j is absorbing,

$$\begin{aligned}
b_{ij}(t) &= \sum_{t=0}^{\infty} \mathrm{P}(s_t \in \mathrm{TR}, s_{t+1} = j | s_0 = i) \\
&= \sum_{t=0}^{\infty} \sum_{k \in \mathrm{TR}} \mathrm{P}(s_t = k, s_{t+1} = j | s_0 = i) \\
&= \sum_{t=0}^{\infty} \sum_{\substack{k=1 \\ k \in \mathrm{TR}}}^{n_{tr}} \mathrm{P}(s_t = k, s_{t+1} = j | s_0 = i) \\
&= \sum_{t=0}^{\infty} \sum_{\substack{k=1 \\ k \in \mathrm{TR}}}^{n_{tr}} \mathrm{P}(s_{t+1} = j | s_t = k, s_0 = i) \mathrm{P}(s_t = k | s_0 = i) \\
&= \sum_{t=0}^{\infty} \sum_{\substack{k=1 \\ k \in \mathrm{TR}}}^{n_{tr}} \underbrace{\mathrm{P}(s_{t+1} = j | s_t = k)}_{r_{kj}} \underbrace{\mathrm{P}(s_t = k | s_0 = i)}_{x_{k|i}(t)} \\
&= \sum_{t=0}^{\infty} \sum_{\substack{k=1 \\ k \in \mathrm{TR}}}^{n_{tr}} x_{k|i}(t)\, r_{kj}
\end{aligned}$$

24

Introduction to Markov chains

- The formula states that the probability of reaching absorbing node *j* at time (*t*+1) is given by
  - the probability of passing through any state *k* at *t* and then jumping to state *j* from *k* at (*t*+1)
- The absorption probability is then given by taking the sum over all possible time steps

25

Introduction to Markov chains

- Let us compute this quantity

$$
\begin{aligned}
b_{ij} &= \sum_{t=0}^{\infty} \sum_{\substack{k=1 \\ k \in \mathrm{TR}}}^{n_{tr}} x_{k|i}(t)\, r_{kj} \\
&= \sum_{\substack{k=1 \\ k \in \mathrm{TR}}}^{n_{tr}} \left[\sum_{t=0}^{\infty} \mathbf{e}_i^{\mathrm{T}} \mathbf{Q}^t \mathbf{e}_k \right] r_{kj} \\
&= \sum_{\substack{k=1 \\ k \in \mathrm{TR}}}^{n_{tr}} n_{ik}\, r_{kj} \\
&= [\mathbf{NR}]_{ij}
\end{aligned}
$$

26

Introduction to Markov chains

- The absorption probabilities are put in the $\mathbf{B}$ matrix
- Let us reconsider the drunkard's example

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

27

Introduction to Markov chains

$$\mathbf{B} = \mathbf{NR} = \begin{pmatrix} 3/2 & 1 & 1/2 \\ 1 & 2 & 1 \\ 1/2 & 1 & 3/2 \end{pmatrix} \begin{pmatrix} 1/2 & 0 \\ 0 & 0 \\ 0 & 1/2 \end{pmatrix}$$

$$= \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{cc} 0 & 4 \end{array} \\ \begin{pmatrix} 3/4 & 1/4 \\ 1/2 & 1/2 \\ 1/4 & 3/4 \end{pmatrix} \end{array}$$

28

Introduction to Markov chains

- Some additional definitions
  - If, in a Markov chain, it is possible to go to every state from each state, the Markov chain is called irreducible
  - Moreover, the Markov chain is called regular if some power of the transition matrix has only positive (non-0) elements

29

Introduction to Markov chains

- It can further be shown that the powers $\mathbf{P}^t$ of a regular transition matrix tend to a matrix with all rows the same

$$\lim_{t\to\infty} \mathbf{P}^t = \begin{pmatrix} \boldsymbol{\pi}^{\mathrm{T}} \\ \boldsymbol{\pi}^{\mathrm{T}} \\ \vdots \\ \boldsymbol{\pi}^{\mathrm{T}} \end{pmatrix}$$

30

Introduction to Markov chains

- Moreover, the limiting probability distribution of states is independent of the initial state:

$$\lim_{t\to\infty} \mathbf{x}(t) = \boldsymbol{\pi}$$

31

Introduction to Markov chains

- The stationary vector π is the left eigenvector of $\mathbf{P}$, corresponding to eigenvalue 1 and normalized to a probability vector:

$$\begin{aligned}
\boldsymbol{\pi} &= \lim_{t\to\infty} \mathbf{x}(t) \\
&= \lim_{t\to\infty} \mathbf{P}^{\mathrm{T}}\mathbf{x}(t-1) \\
&= \mathbf{P}^{\mathrm{T}} \lim_{t\to\infty} \mathbf{x}(t-1) \\
&= \mathbf{P}^{\mathrm{T}}\boldsymbol{\pi}
\end{aligned}$$

32

Introduction to Markov chains

- It provides the probability of finding the process in each state on the long run
- One can prove that this vector is unique for a regular Markov chain

33

Introduction to Markov chains

- Notice that the fundamental matrix for absorbing chains can be generalized
- To regular chains
  - It, for instance, allows to compute the average first-passage times in matrix form
  - See, for instance, Grinstead and Snell

34

Application to marketing

Application to marketing

- Suppose we have the following model
  - We have a number n of customer clusters or segments
  - Based, for instance, on RFM (Recency, Frequency, Monetary value)
- Each cluster is a state of a Markov chain
- The last (nth) cluster corresponds to lost customers
  - It is absorbing and generates no benefit

Application to marketing

- Each month, we observe the movements from cluster to cluster
- Transition probabilities are estimated
  - by counting the observed frequencies of jumping from one state to another in the past
  - This provides the entries of the transition probabilities matrix

37

Application to marketing

- Suppose also there is an average profit, m_i per month, associated to each customer in state i
  - which could be negative
- There is also a discounting factor:
  $0 < \gamma < 1$
- The expected profit on an infinite time horizon can be computed

38

Application to marketing

- It is given by

$$\overline{m} = \sum_{t=0}^{\infty} \gamma^t \sum_{i=1}^{n} x_i(t)\, m_i$$

  - It provides the expected profit on a infinite horizon

39

Application to marketing

- Which finally provides

$$\begin{aligned}
\overline{m} &= \sum_{t=0}^{\infty} \gamma^t \sum_{i=1}^{n} x_i(t)\, m_i \\
&= \sum_{t=0}^{\infty} \gamma^t \mathbf{m}^{\mathrm{T}} (\mathbf{P}^{\mathrm{T}})^t \mathbf{x}(0) \\
&= \sum_{t=0}^{\infty} \gamma^t \mathbf{x}^{\mathrm{T}}(0) \mathbf{P}^t \mathbf{m} \\
&= \mathbf{x}^{\mathrm{T}}(0) (\sum_{t=0}^{\infty} \gamma^t \mathbf{P}^t) \mathbf{m} \\
&= \mathbf{x}^{\mathrm{T}}(0) (\mathbf{I} - \gamma \mathbf{P})^{-1} \mathbf{m}
\end{aligned}$$

40

Application to marketing

- This is an example of the computation of the lifetime value of a customer
- Which is the expected profit provided by the customer until it leaves the company