Discrete time Markov chains

Today:
- Short recap of probability theory
- Markov chain introduction (Markov property)
- Chapmann-Kolmogorov equations
- First step analysis

Next week:
- Random walks
- First step analysis revisited
- Branching processes
- Generating functions

Two weeks from now:
- Classification of states
- Classification of chains
- Discrete time Markov chains - invariant probability distribution

Basic concepts in probability

Sample space $\Omega$ set of all possible outcomes
Outcome $\omega$
Event $A, B$
Complementary event $A^c = \Omega \setminus A$
Union $A \cup B$ outcome in at least one of $A$ or $B$
Intersection $A \cap B$ Outcome is in both $A$ and $B$
(Empty) or impossible event $\emptyset$

Probability axioms and first results

$0 \leq P(A) \leq 1, \quad P(\Omega) = 1$

$P(A \cup B) = P(A) + P(B)$ for $A \cap B = \emptyset$

Leading to
$P(\emptyset) = 0, \quad P(A^c) = 1 - P(A)$

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$
(inclusion-exclusion)
Conditional probability and independence

\[ P(A|B) = \frac{P(A \cap B)}{P(B)} \iff P(A \cap B) = P(A|B)P(B) \]
(multiplication rule)

\[ \bigcup_i B_i = \Omega \quad B_i \cap B_j = \emptyset \quad i \neq j \]

\[ P(A) = \sum_i P(A|B_i)P(B_i) \]  (law of total probability)

Independence:

\[ P(A|B) = P(A|B^c) = P(A) \iff P(A \cap B) = P(A)P(B) \]

Discrete random variables

Mapping from sample space to metric space
(Read: Real space)

Probability mass function

\[ f(x) = P(X = x) = P(\{\omega \mid X(\omega) = x\}) \]

Distribution function

\[ F(x) = P(X \leq t) = P(\{\omega \mid X(\omega) \leq x\}) = \sum_{t \leq x} f(t) \]

Expectation

\[ E(X) = \sum_x xP(X = x), \quad E(g(X)) = \sum_x g(x)P(X = x) = \sum_x g(x)f(x) \]

Joint distribution

\[ f(x_1, x_2) = P(X_1 = x_1, X_2 = x_2), \quad F(x_1, x_2) = P(X_1 \leq x_1, X_2 \leq x_2) \]

\[ f_{X_1}(x_1) = P(X_1 = x_1) = \sum_{x_2} P(X_1 = x_1, X_2 = x_2) = \sum_{x_2} f(x_1, x_2) \]

\[ F_{X_1}(x_1) = \sum_{t \leq x_1, x_2} P(X_1 = t_1, X_2 = x_2) = F(x_1, \infty) \]

Straightforward to extend to \( n \) variables

We can define the joint distribution of \((X_0, X_1)\) through

\[ P(X_0 = x_0, X_1 = x_1) = P(X_0 = x_0)P(X_1 = x_1 | X_0 = x_0) = P(X_0 = x_0)P_{X_0, X_1} \]

Suppose now some stationarity in addition that \( X_2 \) conditioned on \( X_1 \) is independent on \( X_0 \)

\[ P(X_0 = x_0, X_1 = x_1, X_2 = x_2) = \]
\[ P(X_0 = x_0)P(X_1 = x_1 | X_0 = x_0)P(X_2 = x_2 | X_0 = x_0, X_1 = x_1) = \]
\[ P(X_0 = x_0)P(X_1 = x_1 | X_0 = x_0)P(X_2 = x_2 | X_1 = x_1) = \]
\[ P_{X_0}P_{X_1,X_0}P_{X_2,X_1} \]

which generalizes to arbitrary \( n \).
Markov property

\[ P(\{X_n = x_n|H\}) = P(\{X_0 = x_0, X_1 = x_1, X_2 = x_2, \ldots X_{n-1} = x_{n-1}\}) \]

- Generally the next state depends on the current state and the time
- In most applications the chain is assumed to be time homogeneous, i.e. it does not depend on time
- The only parameters needed are \( P(\{X_n = j|X_{n-1} = i\}) = p_{ij} \)
- We collect these parameters in a matrix \( P = \{p_{ij}\} \)
- The joint probability of the first \( n \) occurrences is

\[ P(X_0 = x_0, X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = p_{x_0} P_{x_0,x_1} P_{x_1,x_2} \ldots P_{x_{n-1},x_n} \]

Example 1: Random walk with two reflecting barriers 0 and \( N \)

\[
P = \begin{pmatrix}
1 - p & p & 0 & \ldots & 0 & 0 & 0 \\
q & 1 - p - q & p & \ldots & 0 & 0 & 0 \\
0 & q & 1 - p - q & \ldots & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & \ldots & q & 1 - p - q & p \\
0 & 0 & 0 & \ldots & 0 & q & 1 - q
\end{pmatrix}
\]

Example 2: Random walk with one reflecting barrier at 0

\[
P = \begin{pmatrix}
1 - p & p & 0 & 0 & 0 & \ldots \\
q & 1 - p - q & p & 0 & 0 & \ldots \\
0 & q & 1 - p - q & p & 0 & \ldots \\
0 & 0 & q & 1 - p - q & p & \ldots \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \ldots & 0 & q & 1 - q
\end{pmatrix}
\]

A profuse number of applications

- Storage/inventory models
- Telecommunications systems
- Biological models
- \( X_n \), the value attained at time \( n \)
- \( X_n \) could be
  - The number of cars in stock
  - The number of days since last rainfall
  - The number of passengers booked for a flight
- See textbook for further examples
Example 3: Random walk with two absorbing barriers

\[ P = \begin{bmatrix}
1 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\
q & 1 - p - q & p & 0 & 0 & \cdots & 0 & 0 & 0 \\
0 & q & 1 - p - q & p & 0 & \cdots & 0 & 0 & 0 \\
0 & 0 & q & 1 - p - q & p & \cdots & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \cdots & q & 1 - p - q & p \\
0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 1
\end{bmatrix} \]

The matrix can be finite (if the Markov chain is finite)

\[ P = \begin{bmatrix}
p_{1,1} & p_{1,2} & p_{1,3} & \cdots & p_{1,n} \\
p_{2,1} & p_{2,2} & p_{2,3} & \cdots & p_{2,n} \\
p_{3,1} & p_{3,2} & p_{3,3} & \cdots & p_{3,n} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
p_{n,1} & p_{n,2} & p_{n,3} & \cdots & p_{n,n}
\end{bmatrix} \]

Two reflecting/absorbing barriers

or infinite (if the Markov chain is infinite)

\[ P = \begin{bmatrix}
p_{1,1} & p_{1,2} & p_{1,3} & \cdots & p_{1,n} & \cdots \\
p_{2,1} & p_{2,2} & p_{2,3} & \cdots & p_{2,n} & \cdots \\
p_{3,1} & p_{3,2} & p_{3,3} & \cdots & p_{3,n} & \cdots \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
p_{n,1} & p_{n,2} & p_{n,3} & \cdots & p_{n,n} & \cdots \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots 
\end{bmatrix} \]

At most one barrier

The matrix \( P \) can be interpreted as

\[ \begin{align*}
\text{the engine that drives the process} \\
\text{the statistical descriptor of the quantitative behaviour} \\
a collection of discrete probability distributions
\end{align*} \]

- For each \( i \) we have a conditional distribution
- What is the probability of the next state being \( j \) knowing that the current state is \( i \)
  \[ P_{ij} = \Pr(X_n = j | X_{n-1} = i) \]
- \( \sum_j P_{ij} = 1 \)
- We say that \( P \) is a stochastic matrix
More definitions and the first properties

- We have defined rules for the behaviour from one value and onwards.
- Boundary conditions specify e.g. behaviour of $X_0$:
  - $X_0$ could be certain $X_0 = a$
  - or random $P(X_0 = i) = p_i$
- Once again we collect the possibly infinite many parameters in a vector $p$.

$n$ step transition probabilities

- the probability of being in $j$ at the $n$'th transition having started in $i$.
- Once again collected in a matrix $P^{(n)} = \{P_{ij}^{(n)}\}$.
- The rows of $P^{(n)}$ can be interpreted like the rows of $P$.
- We can define a new Markov chain on a larger time scale $(P^n)$.

Small example

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

$$P^{(2)} = \begin{bmatrix}
(1 - p)^2 + pq & (1 - p)p & p^2 & 0 \\
q(1 - p) & 2qp & 0 & p^2 \\
q^2 & 0 & 2qp & p(1 - q) \\
0 & q^2 & (1 - q)q & (1 - q)^2 + qp
\end{bmatrix}$$

Chapmann Kolmogorov equations

- There is a generalisation of the example above.
- Suppose we start in $i$ at time 0 and wants to get to $j$ at time $n + m$.
- At some intermediate time $n$ we must be in some state $k$.
- We apply the law of total probability

  $$P(B) = \sum_k P(B|A_k) P(A_k)$$

  $$P(X_{n+m} = j | X_0 = i)$$

  $$= \sum_k P(X_{n+m} = j | X_0 = i, X_n = k) P(X_n = k | X_0 = i).$$
\[
\sum_k P(X_{n+m} = j|X_0 = i, X_n = k) P(X_n = k|X_0 = i)
\]
by the Markov property we get
\[
\sum_k P(X_{n+m} = j|X_n = k) P(X_n = k|X_0 = i)
= \sum_k P^{(m)}_{kj} P^{(n)} = \sum_k P^{(n)}_{ik} P^{(m)}_{kj}
\]
which in matrix formulation is
\[
P^{(n+m)} = P^{(n)} P^{(m)} = P^{n+m}
\]

The probability of \(X_n\)

> The behaviour of the process itself - \(X_n\)
> The behaviour conditional on \(X_0 = i\) is known (\(P^{(n)}_{ij}\))
> Define \(P(X_n = j) = p^{(n)}_j\)
> with \(p^{(n)} = \{p^{(n)}_j\}\) we find
\[
p^{(n)} = p P^{(n)} = p P^n
\]

Small example - revisited

\[
P = \begin{pmatrix}
1 - p & p & 0 & 0 \\
q & 0 & p & 0 \\
0 & q & 0 & p \\
0 & 0 & q & 1 - q
\end{pmatrix}
\]
with \(p = \left(\frac{1}{3}, 0, 0, \frac{2}{3}\right)\) we get
\[
p^{(1)} = \begin{pmatrix}
1 - p & p & 0 & 0 \\
q & 0 & p & 0 \\
0 & q & 0 & p \\
0 & 0 & q & 1 - q
\end{pmatrix} = \begin{pmatrix}
\frac{1 - p}{3} & \frac{p}{3} & \frac{2q}{3} & \frac{2(1 - q)}{3}
\end{pmatrix}
\]

\[
P^2 = \begin{pmatrix}
(1 - p)^2 + pq & (1 - p)p & p^2 & 0 \\
q(1 - p) & 2qp & 0 & p^2 \\
q^2 & 2qp & 0 & (1 - q) \\
0 & q^2 & (1 - q)q & (1 - q)^2 + qp
\end{pmatrix}
\]
First step analysis - setup

Consider the transition probability matrix

\[
P = \begin{pmatrix}
1 & 0 & 0 \\
\alpha & \beta & \gamma \\
0 & 0 & 1
\end{pmatrix}
\]

Define

\[
T = \min \{ n \geq 0 : X_n = 0 \text{ or } X_n = 2 \}
\]

and

\[
u = P(X_T = 0 | X_0 = 1) \quad v = E(T | X_0 = 1)
\]

First step analysis - absorption probability

\[
u = P(X_T = 0 | X_0 = 1)
\]

\[
u = \sum_{k=0}^{2} P(X_1 = k | X_0 = 1) P(X_T = 0 | X_0 = 1, X_1 = k)
\]

\[
u = \sum_{k=0}^{2} P(X_1 = k | X_0 = 1) P(X_T = 0 | X_1 = k)
\]

\[
u = P_{1,0} \cdot 1 + P_{1,1} \cdot u + P_{1,2} \cdot 0.
\]

And we find

\[
u = \frac{P_{1,0}}{1 - P_{1,1}} = \frac{\alpha}{\alpha + \gamma}
\]

First step analysis - time to absorption

\[
u = \frac{1}{1 - P_{1,1}} = \frac{1}{1 - \beta}
\]
More than one transient state

\[
P = \begin{pmatrix}
1 & 0 & 0 & 0 \\
P_{1,0} & P_{1,1} & P_{1,2} & P_{1,3} \\
P_{2,0} & P_{2,1} & P_{2,2} & P_{2,3} \\
0 & 0 & 0 & 1
\end{pmatrix}
\]

- Here we will need conditional probabilities
  \[ u_i = \mathbb{P}(X_T = 0 | X_0 = i) \]
- and conditional expectations \[ v_i = \mathbb{E}(T | X_0 = i) \]

Leading to

\[
u_1 = P_{1,0} + P_{1,1} u_1 + P_{1,2} u_2 \\
u_2 = P_{2,0} + P_{2,1} u_1 + P_{2,2} u_2
\]

and

\[
v_1 = 1 + P_{1,1} v_1 + P_{1,2} v_2 \\
v_2 = 1 + P_{2,1} v_1 + P_{2,2} v_2
\]

General finite state Markov chain

\[
P = \begin{pmatrix}
Q & R \\
0 & I
\end{pmatrix}
\]

General absorbing Markov chain

\[ T = \min \{ n \geq 0, X_n \geq r \} \]

In state \( j \) we accumulate reward \( g(j) \), \( w_i \) is expected total reward conditioned on start in state \( i \)

\[
w_i = \mathbb{E} \left( \sum_{n=0}^{T-1} g(X_n) | X_0 = i \right)
\]

leading to

\[
w_i = g(i) + \sum_j P_{i,j} w_j
\]
Special cases of general absorbing Markov chain

- $g(i) = 1$ expected time to absorption ($\nu_i$)
- $g(i) = \delta_{ik}$ expected visits to state $k$ before absorption