Discrete Time Markov Chains. Applications in Bioinformatics

Javier Campos
Universidad de Zaragoza
http://webdiis.unizar.es/~jcampos/
Outline

• Basic definitions
• Representations
• Multi-step transition probabilities
• Classification of states
• Steady-state behaviour
• Examples
• Applications in Bioinformatics
Basic definitions

• Markov processes: special class of stochastic processes that satisfy the **Markov Property** (MP):
  – Given the state of the process at time $t$, its state at time $t + s$ has probability distribution which is a function of $s$ only.
  – i.e. the future behaviour after $t$ is independent of the behaviour before $t$.
  – Often intuitively reasonable, yet sufficiently “special” to facilitate effective mathematical analysis.
Basic definitions

• We consider Markov processes with discrete state (sample) space.

They are called **Markov chains**.

- If time parameter is discrete \( \{t_0, t_1, t_2...\} \) they are called **Discrete Time Markov Chains** (DTMC).

- If time is continuous \( (t \geq 0, t \in IR) \), they are called **Continuous Time Markov Chains** (CTMC).
Basic definitions

• Let $X = \{X_n\}$, with $n = 0,1,...$; $X_i \in \mathbb{N}$, $i \geq 0$ be a non-negative integer valued Markov chain with discrete time parameter $n$.

Markov Property states that:

$$P(X_{n+1} = j \mid X_0 = x_0, ..., X_n = x_n) = \tag{*}$$

$$= P(X_{n+1} = j \mid X_n = x_n), \text{ for } j,n=0,1...$$
Basic definitions

• Evolution of a DTMC is completely described by its 1-step transition probabilities
  \[ p_{ij}(n) = P(X_{n+1} = j \mid X_n = i) \text{ for } i,j,n \geq 0 \]

• If the conditional probability is invariant with respect to the time origin, the DTMC is said to be time-homogeneous
  \[ p_{ij}(n) = p_{ij} \]
  \[ \sum_{j \in \Omega} p_{ij} = 1, \ \forall i \in \Omega \]
Outline

• Basic definitions
• Representations
• Multi-step transition probabilities
• Classification of states
• Steady-state behaviour
• Examples
• Applications in Bioinformatics
Representations

- State transition diagram
  - Directed graph
    - number of nodes = number of states (if \( \Omega \) finite)
    - An arc from \( i \) to \( j \) if and only if \( p_{ij} > 0 \)

  Telephone line example:
  line is either idle (state 0) or busy (state 1)
Representations

• Transition probability matrix

\[
P = \begin{bmatrix}
    p_{00} & p_{01} & \cdots \\
    p_{10} & p_{11} & \cdots \\
    \vdots & \vdots & \ddots \\
    p_{i0} & p_{ii} & \cdots \\
    \vdots & \vdots & \ddots 
\end{bmatrix}
\]

in which all rows sum to 1

– dimension = number of states in \( \Omega \) if finite, otherwise countably infinite

– conversely, any real matrix \( P \) s.t. \( p_{ij} \geq 0, \sum_j p_{ij} = 1 \) (called a stochastic matrix) defines a MC
Representations

Telephone line example

\[ P = \begin{bmatrix} 0.9 & 0.1 \\ 0.3 & 0.7 \end{bmatrix} \]
Representations

- Example 2: I/O buffer, capacity $M$ records
  New record added in any unit of time with prob. $a$ (if not full).
  Buffer emptied in any unit of time with prob. $b$.
  If both occur in same interval, insertion done first.
  Let $X_n$ be the number of records in buffer at (discrete) time $n$.
  Then, assuming that insertions and emptying are independent of each other and of their own past histories, \( \{X_n \mid n=0,1,...\} \) is a MC with state space \( \{0,1,...,M\} \) and state diagram:
Representations

- The transition probability matrix follows immediately, e.g.:

\[ p_{12} = a(1 - b) = p_{n,n+1}, \quad 0 \leq n \leq M - 1 \]

\[ p_{MM} = 1 - b \]

e tc.
Representations

• Example 3:

A system that can be

- Idle
- Busy
- Waiting for a resource
- Broken
- Repairing

\[
P = \begin{pmatrix}
0.2 & 0.75 & 0.0 & 0.05 & 0.0 \\
0.3 & 0.3 & 0.3 & 0.1 & 0.0 \\
0.0 & 0.55 & 0.4 & 0.05 & 0.0 \\
0.0 & 0.0 & 0.0 & 0.5 & 0.5 \\
0.6 & 0.0 & 0.0 & 0.0 & 0.4
\end{pmatrix}
\]
Representations

• Time spent in a state:

\[ P = \begin{bmatrix} 0.9 & 0.1 \\ 0.3 & 0.7 \end{bmatrix} \]

\( T_0 = \text{random variable \textit{“time spent in state 0”}} \)

\[ P(T_0=1) = (1-p_{00}) \]
\[ P(T_0=2) = p_{00} (1-p_{00}) \]
\[ P(T_0=3) = p_{00}^2 (1-p_{00}) \]
\[ \vdots \]
\[ P(T_0=n+1) = p_{00}^n (1-p_{00}) \]

→ Geometrically distributed random variable

Is the discrete analogue of exponential distribution → memoryless
Outline

• Basic definitions
• Representations
• Multi-step transition probabilities
• Classification of states
• Steady-state behaviour
• Examples
• Applications in Bioinformatics
Multi-step transition probabilities

• Let the 2-step transition probability be

\[ p_{ij}^{(2)} = P(X_{n+2} = j \mid X_n = i) \]

\[ = \sum_{k \in \Omega} P(X_{n+1} = k, X_{n+2} = j \mid X_n = i) \text{ by law of tot. prob.} \]

\[ = \sum_{k \in \Omega} P(X_{n+2} = j \mid X_n = i, X_{n+1} = k) P(X_{n+1} = k \mid X_n = i) \]

\[ = \sum_{k \in \Omega} P(X_{n+2} = j \mid X_{n+1} = k) P(X_{n+1} = k \mid X_n = i) \text{ by MP} \]

\[ = \sum_{k \in \Omega} p_{ik} p_{kj} \]

\[ = \left( P^2 \right)_{ij} \]
Multi-step transition probabilities

• Similarly, the $n$-step transition probability

\[ p_{ij}^{(n)} = P(X_n = j \mid X_0 = i) \]

\[ = \sum_{k \in \Omega} P(X_n = j \mid X_{n-1} = k, X_0 = i)P(X_{n-1} = k \mid X_0 = i) \]

\[ = \sum_{k \in \Omega} p_{ik}^{(n-1)} p_{kj} \]

In matrix form:

\[ P^{(n)} = P^{(n-1)} \times P \]

If $n=2$:

\[ P^{(2)} = P^{(1)} \times P = P^2 \]

And in general: \[ P^{(n)} = P^n \] i.e. \[ p_{ij}^{(n)} = (P^n)_{ij} \]
Multi-step transition probabilities

- A more general version of previous equations
  - Chapman-Kolmogorov equations

\[ p_{ij}^{(n+m)} = \sum_{k \in \Omega} p_{ik}^{(n)} p_{kj}^{(m)} \]

Because

\[ p_{ij}^{(n+m)} = \sum_{k \in \Omega} p(X_{n+m} = j \mid X_n = k, X_0 = i) p(X_n = k \mid X_0 = i) \]

Thus

\[ P^{(n+m)} = P^{(n)} \times P^{(m)} = P^n \times P^m \]
Multi-step transition probabilities

- Computation of transient distribution
  - Probabilistic behaviour of the Markov chain over any finite period time, given the initial state

\[ P(X_n = j \mid X_0 = i) = p_{ij}^{(n)} = \left(P^n\right)_{ij} \]

- E.g., in the example of the I/O buffer with capacity of \( M \) records, the average number of records in the buffer at time 50 is

\[ E(X_{50} \mid X_0 = 0) = \sum_{j=1}^{\min(M,50)} j q_{0j}^{(50)} \]
Multi-step transition probabilities

• Computation of transient distribution
  – $n$th-step distribution:
    \[
    \pi_j(n) = P(X_n = j) = \sum_{i \in \Omega} P(X_0 = i) P(X_n = j \mid X_0 = i)
    \]
    \[
    = \sum_{i \in \Omega} \pi_i(0) p_{ij}^{(n)} = \sum_{i \in \Omega} \pi_i(0) (P^n)_{ij}
    \]
  – in matrix form:
    \[
    \pi(n)^\mathsf{T} = \pi(0)^\mathsf{T} P^n
    \]
  – Problem: computationally expensive!
Outline

• Basic definitions
• Representations
• Multi-step transition probabilities
• Classification of states
• Steady-state behaviour
• Examples
• Applications in Bioinformatics
Classification of states

- State $j$ is **accessible** from state $i$ (written $i \rightarrow j$) if
  \[ p_{ij}^{(n)} > 0, \text{ for some } n \]

- A state $i$ is said to **communicate** with state $j$ (written $i \leftrightarrow j$) if $i$ is accessible from $j$ and $j$ is accessible from $i$

- A set of states $C$ such that each pair of states in $C$ communicates is a **communicating class**

- A communicating class is **closed** if the probability of leaving the class is zero (no state out of $C$ is accessible from states in $C$)

- A Markov chain is **irreducible** if the state space is a communicating class

- State $i$ is an **absorbing** state if there is no state reachable from $i$
Classification of states

• Periodicity:
  – A state $i$ has **period** $k$ if any return to state $i$ must occur in some multiple of $k$ time steps.

$$k = \gcd\{n : P(X_n = i \mid X_0 = i) > 0\}$$

  – If $k = 1$, then the state is **aperiodic**; otherwise ($k>1$), the state is **periodic with period** $k$.
  – It can be shown that every state in a communicating class must have the same period.
  – An irreducible Markov chain is **aperiodic** if its states are aperiodic.
Classification of states

• Recurrence
  
  – A state \( i \) is **transient** if, given that we start in state \( i \), there is a non-zero probability that we will never return back to \( i \).
    
      • Formally, next return time to state \( i \) ("hitting time"):
        \[
        T_i = \min\{n : X_n = i \mid X_0 = i\}
        \]

      • State \( i \) is transient if \( P(T_i < \infty) < 1 \) (i.e. \( P(T_i = \infty) > 0 \))

  – If a state \( i \) is not transient (it has finite hitting time with probability 1), then it is said to be **recurrent**.

  – Let \( M_i \) be the expected (average) return time, \( M_i = E[T_i] \)
    
      • Then, state \( i \) is **positive recurrent** if \( M_i \) is finite; otherwise, state \( i \) is **null recurrent**.

  – It can be shown that a state is recurrent iff \( \sum_{n=0}^{\infty} p_{ii}^{(n)} = \infty \)
Classification of states

• In a finite DTMC:
  – All states belonging to a closed class are positive recurrent.
  – All states not belonging to a closed class are transient.
  – There are not null recurrent states.

• In an irreducible DTMC:
  – Either all states are transient or recurrent
    (in case of finite DTMC, all are positive recurrent).

• Ergodicity:
  – A state $i$ is said to be **ergodic** if it is aperiodic and positive recurrent (finite average return time).
  – If all states in a DTMC are ergodic, the chain is said to be ergodic.
Outline

- Basic definitions
- Representations
- Multi-step transition probabilities
- Classification of states
- Steady-state behaviour
- Examples
- Applications in Bioinformatics
Steady-state behaviour

• Transient behaviour: computationally expensive
• Easier and maybe more interesting to determine the limit or steady-state distribution

\[ \pi_j = \lim_{n \to \infty} \pi_j(n) \]

In vector form

\[ \pi = \lim_{n \to \infty} \pi(n) \]

– Does it exist?
– Is it unique?
– Is it independent of the initial state?
Steady-state behaviour

• If limit distribution exists… we know how to compute it!

\[
\pi(n+1) = \pi(0)P^{n+1} = \pi(n)P
\]

\[
\lim_{n \to \infty} \pi(n+1) = \lim_{n \to \infty} \pi(n)P
\]

i.e., it must be equal to the stationary distribution, the solution of:

\[
\begin{align*}
\pi^T P &= \pi^T & \rightarrow & \text{balance equations} \\
\pi^T e &= 1 & \rightarrow & \text{normalizing equat.}
\end{align*}
\]

where \( e = (1,1,\ldots,1)^T \), and the initial distribution does not affect the limit distribution
Steady-state behaviour

• Other interpretation:
  – The solution of balance equations can be seen as the proportion of time that the process enters in each state in the long run
    • Let $N_j(n)$ be the number of visits of the process to the state $j$ until instant $n$
    • The **occupation distribution** can be defined as
      \[
      \pi_j = \lim_{n \to \infty} \frac{E[N_j(n)]}{n + 1}
      \]
    • Of course, its inverse is the mean interval between visits, or **mean return time** ($1/\pi_j$)
  – If the occupation distribution exists, it verifies
    \[
    \pi^T P = \pi^T ; \quad \pi^T e = 1
    \]
Steady-state behaviour

• But,
  – Does limit distribution exist?
  – Is it unique?
  – Is it independent of the initial state?

We know some cases where the answer is no

We know some cases where the answer is yes
Steady-state behaviour

• If a unique limit distribution exists, all rows of $P^n$ must be equal in the limit, in this way the distribution of $X_n$ does not depend on the initial distribution.

• Example

\[
P = \begin{bmatrix}
0 & 1 & 0 \\
0.1 & 0 & 0.9 \\
0 & 1 & 0
\end{bmatrix}
\]

\[
P^n = \begin{bmatrix}
0.1 & 0 & 0.9 \\
0 & 1 & 0 \\
0.1 & 0 & 0.9
\end{bmatrix}
\]

\[
P^{2n-1} = \begin{bmatrix}
0 & 1 & 0 \\
0.1 & 0 & 0.9 \\
0 & 1 & 0
\end{bmatrix}
\]

If $a$ is the initial distribution, then the distribution of $X_n$, $n \geq 1$ is:

\[
(0.1(a_1+a_3), \ a_2, \ 0.9(a_1+a_3)), \text{ if } n \text{ is odd}
\]

\[
(0.1a_2, \ a_1+a_3, \ 0.9a_2), \text{ if } n \text{ is even}
\]

Thus, the DTMC has no limit distribution.

If balance and normalization equations are solved, we get a unique solution $\pi = (0.05, 0.5, 0.45)$.

This means: if $\pi$ is assumed as initial distribution, then $\pi$ is also the distribution for $X_n$, for all $n$. 
Steady-state behaviour

- Example: limit and stationary distributions may be non unique

\[ P = \begin{bmatrix} 0.2 & 0.8 & 0 \\ 0.1 & 0.9 & 0 \\ 0 & 0 & 1 \end{bmatrix} \]

Then,

\[ \lim_{n \to \infty} P^n = \begin{bmatrix} 0.1111 & 0.8889 & 0 \\ 0.1111 & 0.8889 & 0 \\ 0 & 0 & 1 \end{bmatrix} \]

Limit distribution exists, but it is not unique since it depends on the initial distribution: if \( a \) is the initial distribution

\[ \pi = (0.111(a_1+a_2), 0.8889(a_1+a_2), a_3) \]

is a limit distribution for \( X_n \), and it is also a stationary distribution.
Steady-state behaviour

- Finite & irreducible DTMC ⇒ there exists a unique stationary distribution

- Finite & irreducible DTMC ⇒ there exists a unique occupation distribution, and it is equal to the stationary distribution

- Finite, irreducible & aperiodic DTMC ⇒ it has a unique limit distribution, and it is equal to the stationary distribution

- Positive recurrent & aperiodic DTMC ⇒ there exists limit distribution
  - If in addition DTMC is irreducible, the limit distribution is independent of the initial probability

- Irreducible, positive recurrent & periodic DTMC with period \( d \) ⇒ \( \lim_{n \to \infty} p_{ij}^{(nd)} = dp_j \)

- An irreducible & aperiodic DTMC is positive recurrent ⇔ there exists a unique solution of balance equation

- Irreducible, aperiodic & null recurrent DTMC ⇒ \( \lim_{n \to \infty} p_{ij}^{(n)} = 0 \)
Outline

• Basic definitions
• Representations
• Multi-step transition probabilities
• Classification of states
• Steady-state behaviour
• Examples
• Applications in Bioinformatics
Examples

• Random walks
  – A *random walk* is a random process consisting of a sequence of discrete steps of fixed length.
  – *One-dimensional nearest-neighbour* random walk

• In this case, the reachable states are integer (or natural) numbers and at each step, the process jumps to the nearest-neighbour to the right with probability \( p \) or to the nearest-neighbour to the left with probability \( 1 - p \).

• A random walk with state space equal to \( \mathbb{Z} \) (integer numbers) can be transient or null recurrent depending on \( p \).
  – If \( p = 1/2 \) then it is null recurrent and in other cases it is transient. Look at the *book by S. Ross* (Example 4.3d).

• A random walk with reflectant barrier at 0 (i.e., state space equal to non negative integers) can be transient, null recurrent or positive recurrent. Look here (local copy).
Examples

- A processor has certain tasks to perform
  - State transition diagram. Possible states:
    - idle (no task to do)
    - busy (working on a task)
    - waiting (stopped for some resource)
    - broken (no longer operational)
    - repair (fixing the failure)
Examples

– Transition probability matrix representation

\[
P = \begin{bmatrix}
0.2 & 0.75 & 0.0 & 0.05 & 0.0 \\
0.3 & 0.3 & 0.3 & 0.1 & 0.0 \\
0.0 & 0.55 & 0.4 & 0.05 & 0.0 \\
0.0 & 0.0 & 0.0 & 0.5 & 0.5 \\
0.6 & 0.0 & 0.0 & 0.0 & 0.4
\end{bmatrix}
\]
Examples

- Properties: finite state space, irreducible, aperiodic

\[ \lim_{n \to \infty} \pi(n) \]

\[ \pi^T P = \pi^T \]

\[ \pi^T e = 1 \]

\[ \pi = (0.2155, 0.3804, 0.1902, 0.1167, 0.0972)^T \]

⇒ it has a unique limit distribution, and it is equal to the stationary distribution

- Solution:
Examples

\[ \pi = (0.2155, 0.3804, 0.1902, 0.1167, 0.0972)^T \]

- Compute performance indices:
  - Availability: \( P(\text{idle + busy + wait}) = 0.7861 \) (in other words, 78.61% of the time)
  - So, not available: 21.39% of the time
  - Working time: \( P(\text{busy + wait}) = 0.57 \)
Outline

• Basic definitions
• Representations
• Multi-step transition probabilities
• Classification of states
• Steady-state behaviour
• Examples
• Applications in Bioinformatics
Simple model of virus mutation

- Suppose a virus can exist in \( N \) different strains and in each generation either stays the same, or with probability \( \alpha \) mutates to another strain, which is chosen at random.
- What is the probability that the strain in the \( n \)th generation is the same as that in the \( O \)th?

\( N \)-state chain, with \( N \times N \) transition matrix \( P \) given by

\[
P_{ii} = 1 - \alpha
\]
\[
P_{ij} = \alpha / (N - 1), \text{ for } i \neq j.
\]

Answer: computing \( p_{11}^{(n)} \).

Much simpler: exploiting the symmetry present in the mutation rules…
Simple model of virus mutation

- At any time a transition is made from the initial state to another with probability $\alpha$, and a transition from another state to the initial state with probability $\alpha/(N - 1)$ \(\rightarrow\) two-state chain

\[
P^{n+1} = P^n P \Rightarrow p_{11}^{(n+1)} = p_{12}^{(n)} \beta + p_{11}^{(n)} (1 - \alpha), \text{ with } \beta = \alpha/(N - 1).
\]

Since $p_{11}^{(n)} + p_{12}^{(n)} = 1$ then we get the recurrence

\[
p_{11}^{(n+1)} = (1 - \alpha - \beta) p_{11}^{(n)} + \beta, \text{ and } p_{11}^{(0)} = 1.
\]

That has a unique solution

\[
p_{11}^{(n)} = \begin{cases} \frac{\beta}{\alpha + \beta} + \frac{\alpha}{\alpha + \beta} (1 - \alpha - \beta)^n & \text{for } \alpha + \beta > 0 \\ 1 & \text{for } \alpha + \beta = 0. \end{cases}
\]

Then, in our case $p_{11}^{(n)} = \frac{1}{N} + \left(1 - \frac{1}{N}\right) \left(1 - \frac{\alpha N}{N - 1}\right)^n$. 
Life behaviour model exercise

• An octopus, called Paul, is trained to choose object A from a pair of objects A, B, by being given repeated trials in which it is shown both and is rewarded with food if it chooses A.

• Modelling its mind states…
  – state 1: it cannot remember which object is rewarded
    → it is equally likely to choose either
  – state 2: it remembers and chooses A but may forget again
  – state 3: it remembers and chooses A and never forgets

• Modelling its evolution in time
  – After each trial it may change its state of mind according to the transition matrix

\[
\begin{align*}
\text{state 1} & \quad 1/2 & \quad 1/2 & \quad 0 \\
\text{state 2} & \quad 1/2 & \quad 1/12 & \quad 5/12 \\
\text{state 3} & \quad 0 & \quad 0 & \quad 1
\end{align*}
\]
Life behaviour model exercise

• Questions: It is in state 1 before the first trial,

  – What is the probability that it is in state 1 just before the \((n+1)th\) trial?
    • Answer.
  – What is the probability \(P_{n+1}(A)\) that it chooses A on the \((n+1)th\) trial?
    • Answer.
  – Someone suggests that the record of successive choices (a sequence of As and Bs) might arise from a two-state Markov chain with constant transition probabilities.

Discuss, with reference to the value of \(P_{n+1}(A)\) that you have found, whether this is possible.

• Answer.
Other examples

• Population genetics, Norris’ book:
  – Wright-Fisher model
  – Moran model

• Epidemic models, Linda Allen’s book:
  – SIS Epidemic Model
  – Chain Binomial Epidemic Models

(you can read it at http://webdiis.unizar.es/asignaturas/SPN/?page_id=104)
Many other applications…

- Birth-death processes used to model the evolution of populations
- Markov models appear also in chemical and biochemical kinetics
- Are the basis for Hidden Markov Models
- Et cetera…