Markov Chains

Eivind Coward
IM, NTNU
Eivind.Coward@ntnu.no

Topics and textbook

- Today's main topic: Markov Chains
- Copies of relevant parts found on the course homepage
- Covered in Ch. 4 and 10.1-10.2, but my presentation (and notation) is slightly different.
- Markov Chains are
  - a useful statistical modeling tool in many fields, including bioinformatics
  - necessary background for Hidden Markov Models (next week’s topic)
- Requires knowledge of basic discrete probability theory
  - assumed known, but a brief review will be given.
  - Covered in Ewens and Grant, Ch. 1.2-1.5, 1.12 and 2.1

Outline

1. Review of basic probability theory
2. Stochastic processes
3. Markov chains
4. Transition probabilities
5. Stationary distribution
6. Periodicity, reducibility, and absorbing states
7. Reversibility
8. Higher order Markov Chains

Review: Random events

- Informally: Something that will happen with a given probability
- Ex.: We roll two dice. “The sum is 4” is an event.
- More formally: An event is a subset $A$ of the set $\Omega$ of all elementary events, denoted the sample space.
- Ex.: $\Omega$ represents the outcome of roll of a pair of dice $\Omega = \{(1,1),(1,2),\ldots,(6,6)\}$ (36 elementary events) $A = \{(1,3),(2,2),(3,1)\}$

Review: random variables

- Formally: A function that assigns real values to random events in a sample space.
- Informally: variable that takes a “random value”.
- The value is drawn from a probability distribution.
- We will consider only discrete random variables, (finite or countable range)
- Probability function $P_X(x) = Pr(X = x)$
- Note: $\sum_{x \in \text{Range}(X)} P_X(x) = 1$

Review: Conditional probability

- Given two events $A$ and $B$
  - Ex: $A = "\text{sum of two dice is 4}\”$ $B = "\text{first die shows 3}\”$
  - What is $Pr(A)$ and $Pr(B)$?
  - What is $Pr(A \& B)$, also written $Pr(AB)$?
  - What is the probability of $A$ given that $B$ occurs, written $Pr(A \mid B)$?
  - We define the conditional probability of $A$ given $B$ as $Pr(A \mid B) = \frac{Pr(AB)}{Pr(B)}$
Events can be stated as values of random variables

\( X_1 \) = result of first die,
\( X_2 \) = result of second die,
\( Y = X_1 + X_2 \)

Ex: A = "sum of two dice is 4"
B = "first die shows 3"

\( \Pr(A) = \Pr(Y = 4) \)
\( \Pr(B) = \Pr(X_1 = 3) \)
\( \Pr(AB) = \Pr(Y = 4, X_1 = 3) \)

Conditional probability:

\[
\Pr(A \mid B) = \frac{\Pr(Y = 4, X_1 = 3)}{\Pr(X_1 = 3)}
\]

Review: Independent events

Intuitively, two events \( A \) and \( B \) are independent if knowledge of one does not affect the probability of the other.

That is:
\( \Pr(A \mid B) = \Pr(A) \)
\( \Pr(B \mid A) = \Pr(B) \)

Each of the above relations is equivalent to
\( \Pr(AB) = \Pr(A) \Pr(B) \)

We take this as the definition that the events \( A \) and \( B \) are independent.

Independence of three or more events: See exercise 4.

Independent random variables

Same intuitive meaning for random variables as for events: Two random variables \( X \) and \( Y \) are independent if knowledge of the value of one does not affect the probability distribution of the other.

For all values \( x \) in the range of \( X \) and \( y \) in the range of \( Y \):
\( \Pr(X = x, Y = y) = \Pr(X = x) \Pr(Y = y) \)

That is, for each \( x, y \) the corresponding events are independent.

Generalizes to three or more variables in the same way as events.

Other familiar concepts (assumed to be known)

- The (cumulative) distribution function (cdf):
\( F_X(x) = \Pr(X \leq x) \).

- The mean (expected value) of a random variable
\( \mu = E(X) \)

- The variance of a random variable
\( \sigma^2 = \text{Var}(X) \)

- Some important discrete distributions:
uniform, binomial, geometric, Poisson

- Basic probability theory of multiple variables:
joint probability, independence, covariance and correlation

Stochastic process

A stochastic process is a family of random variables \( \{X_t\} \) for all \( t \) in some index set \( T \).

\( t \) is often associated with "time", but this is not always the case.

Ex: \( X_t \) = particle position at time \( t \)
\( X_t \) = integer representing state of a system at time \( t \)
\( X_t \) = base (1=A,2=C,3=G,4=T) in position \( t \) in a DNA sequence

\( T \) can be discrete or continuous.

There is typically some dependence between the random variables.

Markov process

- Special kind of stochastic process \( \{X_t\} \) satisfying the Markov property (or memoryless property):
Given the value of \( X_t \), the "future" values \( X_s \) for \( s > t \) are not influenced by the "past" values \( X_u \) for \( u < t \).

- The range of \( X \) ("state space") \( S \) can be discrete or continuous
- The "time" \( T \) can be discrete or continuous
- A Markov process with discrete state space is called a Markov chain.
  - Continuous time Markov Chains are important in genetics and evolutionary modelling.
  - Discrete time Markov chains are used in these and many other biological modelling applications, often related to sequences.
  - We will restrict ourselves further to the case where \( S \) is finite, the finite discrete time Markov chain.
Markov property for the discrete case

- Finite state space: $S = \{1, 2, \ldots, N\}$
- Discrete time: $T = \{0, 1, 2, 3, \ldots\}$
- **Markov property:**
  - State probabilities for time $t+1$ given state at time $t$ not dependent on previous history:
    \[ \Pr(X_{t+1} = j | X_t = i_0, X_1 = i_1, \ldots, X_t = i_t) = \Pr(X_{t+1} = j | X_t = i) \equiv p_{ij}^{t+1} \]
  - Additionally, we assume that this probability is independent of $t$:
    \[ p_{ij}^{t+1} = p_{ij} \]
    This is called **time homogeneity** (or **stationary** transition probabilities).

Markov chains

- A (finite-state, discrete time, time homogeneous) **Markov chain** is a discrete stochastic process $\{X_t\}$ ($t = 0, 1, 2, \ldots$) defined by
  1. A finite set of **states** $S = \{1, 2, \ldots, N\}$ (for convenience often just labelled $1, 2, \ldots, N$).
  2. An $N \times N$ **transition probability matrix** $P = (p_{ij})$ where $p_{ij} = \Pr(X_{t+1} = j | X_t = i)$.
  3. An **initial probability distribution** $\pi = \Pr(X_0 = i)$
    - **Markov property** (memoryless): Probability of being in a state depends only on previous state, not on past history.
    - Time homogeneity: the probability is independent of “time” $t$. 

Example: weather Markov chain

- Let $X_t$ be the weather at day $t$, with 3 states: $1 =$ sunny, $2 =$ cloudy, $3 =$ rainy
- How many transition probabilities are there?
- Are they all independent?

\[
P = \begin{pmatrix}
  p_{11} & p_{12} & p_{13} \\
  p_{21} & p_{22} & p_{23} \\
  p_{31} & p_{32} & p_{33}
\end{pmatrix} = \begin{pmatrix}
  0.5 & 0.4 & 0.1 \\
  0.3 & 0.4 & 0.3 \\
  0.1 & 0.2 & 0.7
\end{pmatrix}
\]

- Initial probability distribution $\pi = (\pi_1, \pi_2, \pi_3) = (0.0, 1.0)$

State diagram

$n$-step transition probability

- One-step transition probabilities (given):
  \[ p_{ij} = \Pr(X_{t+1} = j | X_t = i) \]
- $n$-step transition probabilities (to be computed):
  \[ p^{(n)}_{ij} = \Pr(X_{t+n} = j | X_t = i) \]
- Turns out that these can be computed by matrix multiplication:
  \[ P^{(n)} = P^n \]
Stationary distribution

- Given any state probability distribution row vector $\pi(t) = (\pi(t)_i)$, the state probability in the next step is $\pi(t+1) = \pi(t)P$
- A stationary distribution is a state distribution $\phi = (\phi_i)$ such that $\phi = \phi P$

Periodic Markov chains

- A Markov chain is periodic if some states are reachable only in $P$, $2P$, $3P$, … steps after their first visit, for an integer $P > 1$.
- Simplest example:
  
  ![Diagram of a periodic Markov chain with states 1, 2, and 3 connected in a cycle](image)
  
  $\rho_{12} = 1$
  
  $\rho_{21} = 1$

- A Markov Chain that is not periodic, is called aperiodic.
- Most practical Markov Chains are aperiodic.

Reducibility and absorbing states

- A Markov chain is irreducible if every state can be reached from every other state in a finite number of steps.
- Otherwise, the Markov Chain is reducible (can be separated into components).
- An important kind of reducible Markov Chain is one containing absorbing states, which can never be exited when entered.

Long run behaviour

- Important theorem for “nice behaving” Markov chains:
  
  A finite state, aperiodic, irreducible Markov chain has a unique stationary distribution $\phi$, which is also a limiting distribution:

  $P(n)$ → $\phi$
  
  and $\pi P(n)$ → $\phi$ as $n → \infty$, independent of $\pi$.

Regularity

How can we check whether a Markov chain is aperiodic and irreducible?

**Definition:**
A finite state Markov chain is called regular if there exists $k \geq 1$ for which all elements of the $k$-th order transition matrix $P^k$ are nonzero.

**Example:**
Weather Markov chain: OK for $k=1$.

**Theorem:**
A finite state Markov chain is aperiodic and irreducible if and only if it is regular.

Example 2: DNA sequence

- $X_t = DNA$ base at position $t$ in the sequence
- 4 states: 1=A, 2=C, 3=G, 4=T
- Example of MC where $t$ is not time
- Many similar models used in bioinformatics
- What is the Markov property? Is it valid?
Higher order Markov chains

- Sometimes we need models with longer memory, relaxing the Markov condition.
- A second order Markov chain is defined by the following transition probabilities:
  \[ p_{ijk} = \Pr(X_{n+2} = k | X_n = i, X_{n+1} = j) \]
- In a similar way, a k-th order Markov chain depends on the k previous states.
- A higher order Markov chain can be reformulated as a standard (first order) Markov chain with more states.

Reversibility

- Time can be reversed:
- **Theorem:**
  Let \((X_t), t = 0, 1, \ldots, T\) be a finite, aperiodic Markov chain with transition probabilities \(p_{ij}\) and stationary distribution \(\phi\).
  Let \(X'_t = X_{T-t}\). Then \((X'_t), t = 0, 1, \ldots, T\) is also a finite, aperiodic, irreducible Markov chain with transition probabilities
  \[ p'_{ij} = \frac{\phi_i p_{ij}}{\phi_j} \]
  and the same stationary distribution \(\phi\).
- **Definition:** A Markov chain is called reversible if \(p'_{ij} = p_{ij}\)
- This is equivalent to the detailed balance equation:
  \[ \phi_i p_{ij} = \phi_j p_{ji} \]