Markov chains

Based largely on Grimmet & Stirzaker

Random (or Stochastic) process

A random process X is a collection of random variables, indexed by a set T, $\{X_t : t \in T\}$

T can be discrete or continuous (discrete or continuous process)

Discrete Markov Chain

$X_1, X_2, X_3, ...$ is a discrete Markov chain if it satisfies the Markov property.

Markov Condition:

$$\mathbb{P}(X_n = j | X_0 = x_0, \ldots, X_{n-1} = x_{n-1}) = \mathbb{P}(X_n = j | X_{n-1} = x_{n-1})$$

Process frequently referred to as *'memoryless'* (future independent of the past, given the present)

Examples of Markov chains

- Gambler's fortune
  Suppose a gambler repeatedly plays a game in which he has some (fixed) probability of winning. When he wins he adds x to his accumulated winnings. When he loses he subtracts y. The gambler's fortune can be described by a Markov chain.
- Gene frequencies over time
  Mutations often spread in populations of organisms largely by chance. Suppose a certain mutation is present in a proportion, a, of individuals in a human population. The behaviour of a across successive generations is an example of a discrete Markov chain.

Examples of Markov chains

- Gambler's fortune
  Suppose a gambler repeatedly plays a game in which he has some (fixed) probability of winning. When he wins he adds x to his accumulated winnings. When he loses he subtracts y. The gambler's fortune can be described by a Markov chain.
- Gene frequencies over time
  Mutations often spread in populations of organisms largely by chance. Suppose a certain mutation is present in a proportion, a, of individuals in a human population. The behaviour of a across successive generations is an example of a discrete Markov chain.

Homogeneous Markov chain

A Markov chain is homogeneous if

$$\mathbb{P}(X_{n+1} = j | X_n = i) = \mathbb{P}(X_1 = j | X_0 = i) \ \forall\ n, i, j$$

The square matrix $\mathbf{P}$ with $p_{ij} \equiv \mathbb{P}(X_{n+1} = j | X_n = i)$ is called the transition probability matrix.

Note $\mathbf{P}$ is a *stochastic matrix* so

$$p_{ij} \geq 0 \ \forall\ i, j$$

$$\sum_j p_{ij} = 1 \ \forall\ i, j$$

Chapman-Kolmogorov equations

Let $\mathbf{P}_n$ be the transition matrix corresponding to n steps of the Markov chain then $\mathbf{P}_n = \mathbf{P}^n$

Proof (by induction):

$\mathbf{P}_1 = \mathbf{P}^1$

$$p_{ij}(n+1) \quad = \quad \mathbb{P}(X_{n+1} = j | X_0 = i)$$

Chapman-Kolmogorov equations

Let $\mathbf{P}_n$ be the transition matrix corresponding to n steps of the Markov chain then $\mathbf{P}_n = \mathbf{P}^n$

Proof (by induction):

$$\mathbf{P}_1 = \mathbf{P}^1$$

$$\begin{aligned}
p_{ij}(n+1) &= \mathbb{P}(X_{n+1} = j | X_0 = i) \\
&= \sum_{k \in S} \mathbb{P}(X_{n+1} = j, X_n = k | X_0 = i) \\
&= \sum_{k \in S} \mathbb{P}(X_{n+1} = j | X_0 = i, X_n = k) \mathbb{P}(X_n = k | X_0 = i) \\
&= \sum_{k \in S} \mathbb{P}(X_{n+1} = j | X_n = k) \mathbb{P}(X_n = k | X_0 = i) \\
&= \sum_{k \in S} p_{kj} p_{ik}(n)
\end{aligned}$$

Distribution over states after n steps of the chain

Let $\mu_i^{(n)} = \mathbb{P}(X_n = i)$ so that $\boldsymbol{\mu}^{(n)}$ is a vector describing the distribution of the states after n steps, then

$$\mu_j^{(n)} = \sum_i \mathbb{P}(X_n = j | X_0 = i)\mathbb{P}(X_0 = i) \ldots \textit{marginalizing over } X_0$$

Markov chain completely specified by $\boldsymbol{\mu}$ and **P** (the vector of initial

state frequencies and transition probability matrix)

Classifying states

A state is persistent if the chain returns to it with probability 1 - i.e. $\mathbb{P}(X_n = i, \text{ for some } n \geq 1 | X_0 = i) = 1$ (otherwise transient)

A persistent state is called null if its mean recurrance time is ∞ - iff $p_{ii} \to 0$ as $n \to \infty$

A state i communicates ($i \to j$) with a state, j if $p_{ij}(m) > 0$ for some $m > 0$.

States i, j intercommunicate ($i \leftrightarrow j$) if $i \to j$ and $j \to i$.

A set C of states is closed if $p_{ij} = 0 \; \forall \; i \in C, j \notin C$

C is irreducible if $i \leftrightarrow j \; \forall \; i, j \in C$

Stationary distributions

The vector $\boldsymbol{\pi}$ is a stationary distribution of the chain if

1. $\pi_j \geq 0 \; \forall \; j$ and $\sum_j \pi_j = 1$
2. $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}$

   (i.e. $\pi_j = \sum_i \pi_i p_{ij}$)

Existence of stationary distribution

1. An irreducible Markov chain has a stationary distribution $\boldsymbol{\pi}$ iff all states are persistent (and non-null)
2. In such a case $\boldsymbol{\pi}$ is unique
3. $\pi_i =$ mean recurrence time of state i

After a large number of steps the probability of state j 'settles down' to π_j, independent of the starting state i - i.e. the columns of **P** become identical

Exercise: Set up a valid (finite $|S|$) probability transition matrix with $p_{ij} > 0 \; \forall \; i, j$ use a computer to work out $\mathbf{P}^n$ for some large n to convince yourself of this

Begin and end states

The probabilty of a specific chain $i_0, \ldots, i_n$ of states of $\{\mathbf{X}\}$ is

$$\pi_{i_0} p_{i_0,i_1} p_{i_1,i_2} \ldots p_{i_{n-1},i_n}$$

In certain applications the probability of beginning in state i may not be π_i. Introduce a *begin* state, $\mathcal{B}$, such that $p_{\mathcal{B},i}$ is the probability of the chain beginning in state i

Introduce *end* state, $\mathcal{E}$ to model finite chains with a given length distribution and with probability of chain termination dependent on the current state.

Reversibility

If $X_n : 0 \leq n \leq N$ is an irreducible non-null persistent Markov chain with transition matrix $\mathbf{P}$ and stationary distribution $\boldsymbol{\pi}$, define the reverse chain as $Y_n = X_{N-n}$ for $0 \leq n \leq N$

Then Y is a Markov chain with $\mathbb{P}(Y_{n+1} = j | Y_n = i) = (\frac{\pi_j}{\pi_i})p_{ji}$

Proof:

Reversibility II

Let $\mathbf{P}$ be the transition matrix of an irreducible Markov chain, X. If there exists $\boldsymbol{\pi}$ such that $\pi_i p_{ij} = \pi_j p_{ji}$ for all $i, j \in S$ and $\sum_j \pi_j = 1$ then $\boldsymbol{\pi}$ is the stationary distribution of the chain and X is reversible.

Proof:

$$\sum_i \pi_i p_{ij} = \sum_i \pi_j p_{ji} = \pi_j \sum_i p_{ji} = \pi_j$$

Markov Chain Monte Carlo

In many cases it is useful to use Markov chains for simulation. Frequently, we would like to set up a Markov chain with with a pre-specified equilibrium frequency vectory, $\boldsymbol{\pi}$

Consider the arbitrary (and usually simple) stochastic matrix $\mathbf{Q}$ and let

$$p_{ij} = q_{ij} \min\left(1, \frac{\pi_j q_{ji}}{\pi_i q_{ij}}\right) \quad \text{for } i \neq j; \quad (p_{ii} \text{ s.t. } \sum_j p_{ij} = 1)$$

Then $\mathbf{P}$ is a Markov transition probability matrix with equilibrium frequency vector $\boldsymbol{\pi}$

Markov Chain Monte Carlo

Q is referred to as the proposal matrix and $a_{ij} = \min\left(1, \frac{\pi_j q_{ji}}{\pi_i q_{ij}}\right)$ are the 'acceptance probabilities'

In practice: with the chain in state i propose a move to state j according to q_{ij}. Accept the move to state j with probability a_{ij}, otherwise remain in state i.

For efficiency the acceptance probabilities can be multiplied by t_{ij}, where (t_{ij}) is a symmetric matrix with $t_{ij} > 0$ and $a_{ij}t_{ij} \leq 1$ for all i, j

Example: application of Markov Chain Monte Carlo in Bayesian statistics

Bayes' Rule:

$$\mathbb{P}(\theta|D) = \frac{\mathbb{P}(D|\theta)\mathbb{P}(\theta)}{\mathbb{P}(D)}$$

Frequently $\mathbb{P}(D)$ is very hard to calculate.

Consider discrete θ. We set up a Markov chain with equilibrium frequency,$\boldsymbol{\pi}$, such that $\mathbb{P}(\theta_i|D) = \pi_i$. Simulation of the chain requires only ratios of posterior probabilities (i.e. $\frac{\mathbb{P}(\theta_j|D)}{\mathbb{P}(\theta_i|D)}$), resulting in the cancellation of $\mathbb{P}(D)$ terms.

Gibbs sampling

A method to sample from the joint distribution of two or more variables by keeping all variables fixed except one (at a time).

We sample from the conditional distribution
$P(x_i|x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$

Prove detailed balance holds with this sampling scheme (sufficient to show that the procedure samples correctly from P)

i.e. show

$$P(x_1, \ldots, x_n) P(x_i'|x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) =$$
$$P(x_1, \ldots, x_{i-1}, x_i', x_{i+1}, \ldots, x_n) P(x_i|x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$$

Continuous-time Chains

Definition 1: Markov Property
Let X be a series of random variables taking values in a countable space, S, and indexed by $[0, \infty]$. X satisfies the Markov Property iff $\mathbb{P}(X(t_n) = j | X(t_1) = i_1, \ldots, X(t_{n-1}) = i_{n-1}) = \mathbb{P}(X(t_n) = j | X(t_{n-1}) = i_{n-1})\ \forall\ j, i_1 \ldots i_{n-1} \in S$ and $t_1 < t_2 < \ldots < t_n$

Definition 2: Homogeneous
Let $p_{ij}(s, t) = \mathbb{P}(X(t) = j | X(s) = i)$ for $s \leq t$, then if $p_{ij}(s, t) = p_{ij}(0, t - s)\ \forall\ i, j, s, t$, the process X is referred to as *homogeneous*

Notation: $p_{ij}(s, t) \equiv p_{ij}(t - s)$ and
$\mathbf{P}_t$ is the matrix, indexed by t with entries $p_{ij}(t)$

Transition Probability Matrix for continuous-time chains

1. $\mathbf{P}_0 = \mathbb{I}$
2. $\mathbf{P}_t$ has non-negative entries with rows summing to 1 (a *stochastic* matrix)
3. $\mathbf{P}_{s+t} = \mathbf{P}_s \mathbf{P}_t$ (Chapman-Kolmogorov equations)

Proof of 3

$$
\begin{aligned}
p_{ij}(s+t) &= \mathbb{P}(X(s+t) = j | X(0) = i) \\
&= \sum_k \mathbb{P}(X(s+t) = j | X(s) = k, X(0) = i)\mathbb{P}(X(s) = k | X(0) = i) \\
&= \sum_k p_{ik}(s) p_{kj}(t)
\end{aligned}
$$

Generator matrix

$\mathbf{P}_t$ is called *standard* if $\mathbf{P}_t \to \mathbb{I}$ as $t \to 0$

In such cases $p_{ij}(h)$ is approximately linear, for small h, i.e. there exists constants g_{ij} such that for $i \neq j$ $p_{ij} \approx g_{ij} h$ and $p_{ii} \approx 1 + g_{ii} h$

Basic characteristics of $\mathbf{G}$

1. $g_{ij} \geq 0 \ \forall \ i \neq j$ and $g_{ii} \leq 0 \ \forall \ i$
2. Since $\sum_j p_{ij}(h) = 1 \approx 1 + \sum_j g_{ij} h \implies \sum_j g_{ij} = 0$

Relating **P** and **G**

$$
\begin{aligned}
p_{ij}(t+h) &= \sum_k p_{ik}(t) p_{kj}(h) \\
&\approx p_{ij}(t)(1+g_{jj}h) + \sum_{k:k\neq j} p_{ik}(t) g_{kj} h \\
&= p_{ij}(t) + h \sum_k p_{ik}(t) g_{kj}
\end{aligned}
$$

$$
\Longrightarrow \ \frac{1}{h}\left[p_{ij}(t+h) - p_{ij}(t)\right] \approx \sum_k p_{ik}(t) g_{kj} = (\mathbf{P}_t \mathbf{G})_{ij}
$$

or

$$
p'_{ij}(t) = \sum_k p_{ik} g_{kj}(t) \text{ or } \mathbf{P}'{}_t = \mathbf{P}_t \mathbf{G} \ldots \textit{Forward equations}
$$

. . .

Similarly

$$\mathbf{P}_t' = \mathbf{G}\mathbf{P}_t \ldots \textit{Backward equations}$$

For $\mathbf{P}_0 = \mathbb{I}$ these equations often have a unique solution

$$\mathbf{P}_t = \sum_{n=0}^{\infty} \frac{t^n}{n!}\mathbf{G}^n$$

Usually written $\mathbf{P}_t = e^{t\mathbf{G}}$

Some properties of **G**

1. Waiting time to next transition is exponentially distributed with parameterr $-\lambda$
2. If the chain is in state *i* and a transition occurs, the probality of transition to state *j* is $-g_{ij}/g_{ii}$
3. Given a continuous-time chain with generator matrix, **G**, the discrete chain formed from the $-g_{ij}/g_{ii}$ is referred to as the *jump chain* of *X*

Further properties of continuous-time chains

1. Classification of states: For any pair, i, j either $p_{ij}(t) = 0\ \forall\ t$ or $p_{ij}(t) > 0\ \forall\ t$
2. $\boldsymbol{\pi}$ is a stationary distribution if $\pi_j \geq 0$; $\sum_j \pi_j = 1$ and $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}_t\ \forall\ t \geq 0$
3. $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}_t\ \forall\ t \geq 0$ iff $\boldsymbol{\pi}\mathbf{G} = \mathbf{0}$