Markov Chain Monte Carlo Methods

- provide an extremely flexible way to sample from many potentially complicated densities
- revolve around the simulation of a class of stochastic processes known as Markov chains
- under a set of easily satisfiable regularity conditions, relatively straightforward algorithms exist for *efficiently* generating Markov chains whose elements can (eventually) be considered a sample from a specified distribution

$$\lim_{M \to \infty} \frac{1}{M} \sum_{m} \phi(X_{B+m}) \to \mathrm{E}[\phi(X)] \text{ wp. } 1$$

A Motivating Problem: - the Travelling Salesman Problem

Suppose that a salesman must visit each of n cities, once only, in some order to be determined.
n! possible routes, so a direct search for the 'best route' not feasible for large n.
Let

$$x_i = i\text{th city visited}$$

Let $x = (x_1 \ldots x_n)$ be a particular route and $c(x) =$ cost of route x. e.g. total distance travelled
Then, we seek: $\arg\min_x c(x)$, where x can take n! possible values.

A Motivating Problem: - the Travelling Salesman Problem

Trick: Define

$$
\begin{aligned}
p_\lambda(x) &= \frac{\exp(-\lambda c(x))}{\sum_x \exp(-\lambda c(x))} \\
&= \text{const} \times \exp(-\lambda c(x))
\end{aligned}
$$

Then $p_\lambda(x)$ is a probability distribution over $\underbrace{1, 2, \ldots, n!}_{\text{each possible route}}$.

Notes:

- If λ is large then:
  large $c(x) \Rightarrow$ low probability associated with this x
  small $c(x) \Rightarrow$ high probability associated with this x.
- As λ increases, only x's which nearly minimise $c(x)$ get any probability.
- $\lambda \to \infty \Rightarrow$ spike at the x which minimizes $c(x)$.

A Motivating Problem: - the Travelling Salesman Problem

In order to find the most likely x, we could simulate from $p_\lambda(x)$. But how do we simulate a “state” from a discrete probability distribution with a large but finite number of states? If $n!$ is large we can't evaluate the normalizing constant so inversion is not an option.

One way of doing this is to design a Markov Chain whose ‘stationary distribution’ matches $p_\lambda(x)$ - we can then simulate a realisation of this stochastic process until it converges to this stationary distribution.

Markov Chains

A Markov Chain is a discrete time stochastic process

$$\{X_t, \quad t = 0, 1, 2, \ldots\}$$

with a finite or countable state space, which satisfies the following properties:

(1) Markov Property

$$\mathrm{P}(X_n = j | X_0 = i_0, X_1 = i_1, \ldots, X_{n-1} = i_{n-1}) = \mathrm{P}(X_n = j | X_{n-1} = i_{n-1})$$

"Future depends on the present, but not on the past"
"Lack of memory property"

(2) Time Homogeneity

$$\mathrm{P}(X_{n+1} = j | X_n = i) \quad \text{is the same for all } n$$

As a result of these two properties, a Markov chain can be completely characterised by: i.e. the distribution of X_n, $n \geq 1$ is completely specified by its initial state, X_0 and its transition probabilities

Markov Chains

For a discrete state space, the transition probabilities are specified through a transition matrix:

$$P = \begin{pmatrix} p_{00} & p_{01} & \cdots & \\ p_{10} & p_{11} & \cdots & \\ \vdots & \vdots & & \\ & & p_{ij} & \\ & & & \end{pmatrix}$$

state space $\{0, 1, 2, \ldots\}$
may be finite or infinite

rows sum to 1

with

$$\begin{aligned} p_{ij} &= \mathrm{P}(X_{n+1} = j | X_n = i) \\ &= \mathrm{P}(X_1 = j | X_0 = i) \end{aligned}$$

by time homogeneity

(Note that, in general, $p_{ij} \neq p_{ji}$).

Behaviour of the process after a given number of iterations

***n*-step transition probabilities**: the probability of moving from state i to state j in n steps.

$$p_{ij}^{(n)} = \mathrm{P}(X_n = j \mid X_0 = i) \quad (p_{ij} = p_{ij}^{(1)})$$

***n*-step transition matrix**, $P^{(n)} = \left(p_{ij}^{(n)}\right)$:

$$\begin{aligned}
p_{ij}^{(n)} &= \mathrm{P}(X_n = j \mid X_0 = i) \\
&= \sum_k \mathrm{P}(X_n = j, X_1 = k \mid X_0 = 1) \\
&= \sum_k \mathrm{P}(X_n = j \mid X_1 = k)\mathrm{P}(X_1 = k \mid X_0 = i) \\
&= \sum_k p_{ik} p_{kj}^{(n-1)} \\
\Rightarrow P^{(n)} &= P^n \quad (n\text{th power of P})
\end{aligned}$$

Initial distribution

Let $\pi_i^{(0)} = \mathrm{P}(X_0 = i)$ and use $\underline{\pi}^{(0)}$ to denote the row vector, i.e. the initial distribution over the discrete state space.

What is the distribution at iteration n, $\underline{\pi}^{(n)}$?

$$\begin{aligned}
\pi_i^{(n)} = \mathrm{P}(X_n = i) &= \sum_k \mathrm{P}(X_n = i, X_{n-1} = k) \\
&= \sum_k \mathrm{P}(X_n = i \,|\, X_{n-1} = k)\mathrm{P}(X_{n-1} = k) \\
&= \sum_k \pi_k^{(n-1)} P_{ki} \\
\Rightarrow \underline{\pi}^{(n)} &= \underline{\pi}^{(n-1)} P \\
\Rightarrow \underline{\pi}^{(n)} &= \underline{\pi}^{(0)} P^n
\end{aligned}$$

Thus, $\underline{\pi}^{(n)}$ is fully specified by P and $\underline{\pi}^{(0)}$.

Stationary Distributions

We will consider a subclass of Markov chains: those whose distribution is invariant under multiplication by the transition matrix P.

$$\underline{\pi}^{(n)} = \underline{\pi}^{(n-1)} = \underline{\pi}\ ; \qquad \underline{\pi} = \underline{\pi} P,$$

i.e. we will want the elements of the chain to be identically distributed.

Markov chains with this property are said to have a stationary distribution, $\underline{\pi}$, defined as follows:

<u>Definition</u>

$\underline{\pi}$ is a stationary distribution iff

(i) $\pi_i \geq 0 \qquad \forall i$.

(ii) $\sum_i \pi_i = 1$.

(iii) $\pi_j = \sum_i \pi_i p_{ij} \qquad \forall j \ (\pi = \pi P)$.

Properties of stationary distributions

(i) $\underline{\pi}^{(n)} = \underline{\pi} \Longrightarrow \underline{\pi}^{(m)} = \underline{\pi} \qquad \forall m \geq n \qquad (\underline{\pi}^{(n)} = \underline{\pi}^{(n-1)} P)$.
"process is in equilibrium"

(ii) Suppose state space is <u>finite</u>, and suppose $\underline{\pi}^{(n)}$ converges as $n \to \infty$. Then the limit must be a <u>stationary distribution</u>.

<u>**Proof:**</u>

Suppose $\pi_i^{(n)} \to \pi_i \quad \forall\, i$. Since, $\pi_i^{(n)} = \sum_j \pi_j^{(n-1)} p_{ji}$

$$\lim_{n\to\infty} : \quad \pi_i = \sum_j \pi_j p_{ji}$$

(limit of a finite sum is the sum of the limits) $\Rightarrow \underline{\pi} = \underline{\pi} P$

Finding the stationary distribution

Supposing a stationary distribution exists, it can be found by solving the equations

$$\underline{\pi} = \underline{\pi} P \qquad \sum \pi_i = 1.$$

(Note: $\underline{\pi} \in [0,1]^d \implies d+1$ eqns for d unknowns...
$\implies$ at least one equation in $\underline{\pi} = \underline{\pi} P$ is redundant).

For some $P, \exists$ a <u>unique</u> stationary distribution...

...for some $P, \exists$ <u>more than one</u> stationary distribution...

...and for some $P, \exists$ <u>no</u> stationary distribution when S is infinite.

Finding the stationary distribution

It will be useful to us to establish the conditions under which there exists a stationary distribution.

Furthermore, we note that the existence of a stationary distribution does not, in general, guarantee the chain's convergence to that distribution. - stationary and limiting distributions are distinct, but related concepts.

We will also, therefore, establish the conditions under which a limiting distribution exists for the chain. This will allow us to build Markov chains that converge to a stationary distribution that we specify.

Definition

If a Markov chain converges to its stationary distribution, then this is also referred to as its equilibrium distribution.

Reversibility

We will focus on a (further) subclass of Markov chains - those with the property of reversibility. We will see that this is a sufficient condition to guarantee the existence of a stationary distribution.

Recall our definition of a stationary distribution: we require

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

the probability of being in state j to be equal to the summed probability of being in each of the other states and then moving to state j.

Reversibility

It is straightforward to see that $\pi_j = \sum_i \pi_i p_{ij}$ will follow if we make the following, simpler restriction on P:

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

$$\text{summing over } i \qquad \Rightarrow \qquad \pi_j = \sum_i \pi_i p_{ij}$$

This is known as the **detailed balance equation**, and a Markov chain that satisfies this property will have a stationary distribution by design.

We now consider the properties of periodicity and reducibility for Markov chains in general, and show that these are sufficient to guarantee convergence of a Markov chain to its invariant distribution $\underline{\pi}$ (assuming its existence).

Irreducibility

Definition- Classification of States

For a given Markov chain, state j communicates with state i $(i \leftrightarrow j)$ if there exists a finite path of non-zero probability from i to j and back again:

$$p_{ij}^{(n)} > 0 \qquad \text{and} \qquad p_{ji}^{(m)} > 0 \qquad \text{for some} \qquad n, m \geq 0.$$

The state space can be divided into disjoint classes s.t.

$$i \leftrightarrow j \Leftrightarrow i, j \quad \text{are in the same class;}$$

Note:

1. $p_{ij} = 0 \;\; \forall \;\; j \neq i \Leftrightarrow i$ is an absorbing state.
2. $p_{ij} = 0 \;\; \forall \;\; i \in C, \;\; \forall \;\; j \notin C \Leftrightarrow C$ is a closed class.

To determine classes, need only look at the structure of zero and non-zero elements of P.

Example:

state space $\{1, 2, 3, 4\}$.

$$P = \begin{pmatrix} + & + & 0 & 0 \\ + & + & 0 & 0 \\ 0 & + & + & + \\ 0 & 0 & 0 & + \end{pmatrix}$$

"$+$" represents a probability > 0

Classes:

$\{1, 2\}$	$\{3\}$	$\{4\}$
closed	open	absorbing state

Irreducibility

Definition- Irreducible Chains

If the state space of a given Markov chain consists of a single class (necessarily closed) then the chain is said to be irreducible.

Note - Finite Irreducible Chains

If a Markov chain is defined on a finite state space and is irreducible, then there exists a unique stationary distribution satisfying

$$\underline{\pi} = \underline{\pi} P \qquad \text{and} \qquad \sum \pi_i = 1.$$

Furthermore,

$$\pi_i > 0 \qquad \forall i$$

i.e. for finite state spaces, we do not require the chain to be reversible if it is irreducible - unique stationary distribution is already guaranteed

Periodicity

The period of state i is the greatest common divisor of

$$\{n : p_{ii}^{(n)} > 0\}$$

Periodicity

Definition - Aperiodicity

- A state is said to be **aperiodic** if its period is 1.
  ($p_{ii} > 0 \quad \Rightarrow \quad$ state i is aperiodic)
- A Markov chain $\{X_t\}$ is **aperiodic** if all states in the corresponding state space are aperiodic.

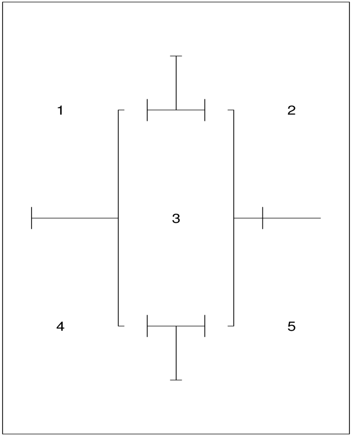
If $i \leftrightarrow j$ then i and j have the same period. Thus, all states of a class have the same period. So...if $\{X_t\}$ is irreducible, all states share the same periodicity

Suppose $\{X_t\}$ is an aperiodic, irreducible Markov chain with finite state space and stationary distribution $\underline{\pi}$, then

$$\mathrm{P}(X_n = i) \to \pi_i \quad \text{as} \quad n \to \infty \qquad \forall i \in S$$

Note that, if the chain is irreducible, we need only show that a single state is aperiodic.

Mouse in maze (see Exercises 8)...



In any room, mouse selects at random any of the possible exits, each equally likely.

$$P = \begin{pmatrix} 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \\ \frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{1}{3} \\ \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{1}{3} \\ 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \end{pmatrix}$$

Key property: The MC is irreducible

<u>State 1</u> $\{n : p_{11} > 0\} = \{2, 3, 4, 5, \ldots\}$, so period is 1.

State 1 aperiodic $\Rightarrow$ all states aperiodic.