Markov Processes

A Markov Process is sequence of R.V.s $X^{(t)}$ with $t \in \mathbb{Z}$ or $t \in \mathbb{R}$ s.t. for any $B \subset \mathbb{R}^d$

$$P(X^{(t)} \in B \mid \mathcal{F}_s) = P(X^{(t)} \in B \mid X_s) \quad \forall\ t \geq s$$

where $\mathcal{F}_s$ is an increasing sequence of σ-algebras (a filtration) and X_t is $\mathcal{F}_t$ measurable

$\mathcal{F}_s$ records all information about X up to and including time s

So if X is Markov, knowing the complete history up to time s isn't better (for predicting future behavior) than just knowing $X^{(s)}$

ex) $y(t)$ is the solution

$$\frac{d}{dt} y = b(t, y)$$

y is Markovian

ex] Now y solves $\frac{d^2}{dt^2} y = b(t,y)$

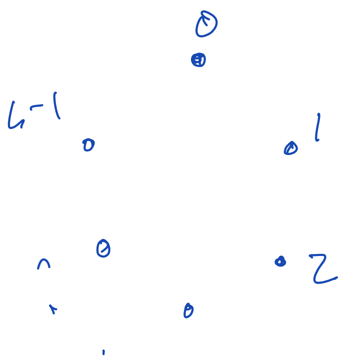
e.g. $m\ddot{y} = f(y)$ Not Markovian

but y also solves

$$\frac{d}{dt}\begin{pmatrix} y \\ v \end{pmatrix} = \begin{pmatrix} v \\ b(t,y) \end{pmatrix}$$ i.e. $v = \frac{d}{dt} y$

ex] $X^{(t+1)} = (X^{(t)} + \xi^{(t)}) \bmod L$ $\xi^{(t)} \in \{-1,+1\}$

and $\xi^{(t)}$ independent of $X^{(0)}, X^{(1)}, \ldots, X^{(t)}$



e.g. $P(\xi^{(t)} = 1) = P(\xi^{(t)} = -1) = 1/2$

How do we encode the evolution of $X^{(t)}$?

Suppose $X \in \{1, 2, \ldots, n\}$ is "time-homogenous"

(i.e. $P(X^{(t)} \in B \mid X^{(s)} = x)$ only depend on $t-s$ and not separately on s and t)

Set $T_{ij} = P(X^{(1)} = j \mid X^{(0)} = i) = P(X^{(t+1)} = j \mid X^{(t)} = i)$

$(T \in \mathbb{R}^{n \times n})$

In fact $T_{ij} \geq 0$ $\quad \sum_j T_{ij} = \sum_j P(X^{(1)} = j \mid X^{(0)} = i) = 1$

T is row-stochastic

Now suppose $P(X^{(0)} = i) = \mu_i$

What is $P(X^{(1)} = j)$?

$$
\begin{aligned}
P(X^{(1)} = j) &= \sum_i P(X^{(1)} = j \text{ and } X^{(0)} = i) \\
&= \sum_i P(X^{(1)} = j \mid X^{(0)} = i) P(X^{(0)} = i) \\
&= \sum_i T_{ij} \mu_i \\
&= (\mu^T T)_j = (T^T \mu)_j
\end{aligned}
$$

$$P(X^{(2)} = j) = (\mu^T T^2)_j$$

Given the current distribution of X we find the distribution of X at future times by mult of the current distribution on the right by T

What if I want to know

$$E[f(X^{(1)}) \mid X^{(0)} = i]$$

$$E[f(X^{(1)}) \mid X^{(0)}=i] = \sum_j f(j)\, P(X^{(1)}=j \mid X^{(0)}=i)$$

$$= \sum_j f(j)\, T_{ij}$$

$$= (Tf)_i$$

To compute expectations of a function of X at a future time conditioned on the current value of X, just multiply the function on the <u>left</u> by T

T is the "transition matrix" of X

For a general Markov process we can define the "transition operator"

$$Tf(x) = E[f(X^{(1)}) \mid X^{(0)} = x]$$

for $f: \mathbb{R}^d \to \mathbb{R}$

Now assume $\int f^2(x)\,dx < \infty$ (i.e L^2)

Consider the inner product $\langle f, g\rangle = \int f(x) g(x)\,dx$

The "adjoint" of T in this inner product is defined by

$$\langle Tf, g\rangle = \langle f, T^* g\rangle \quad \forall f, g$$

Suppose μ is a probability density.

Then

$$\langle Tf, \mu\rangle = \int E[f(X^{(1)}) \mid X^{(0)} = x]\,\mu(x)\,dx$$

$$= \int\int f(y)\, P(X^{(1)} \in dy \mid X^{(0)} = x)\,\mu(x)\,dx$$

$$= \int f(y) \underbrace{\left(\int P(X^{(1)} \in dy \mid X^{(0)} = x)\,\mu(x)\,dx\right)}_{\text{the distribution of } X^{(1)} \text{ if } X^{(0)} \sim \mu} dy$$

drawn from

by definition of adjoint

$$T^*\mu = \int P(X^{(1)} \in dy \mid X^{(0)} = x)\,\mu(x)\,dx$$

(note this might not have a density)

We usually write νT

For any t (real or discrete) we can also define

$$T^t f(x) = E\left[f(X^{(t)}) \mid X^{(0)} = x \right]$$

Can we think of t as a power in T^t

$$
\begin{aligned}
T^t T^s f(x) &= E\left[T^s f(X^{(t)}) \mid X^{(0)} = x \right] \\
&= E\left[E\left[f(X^{(t+s)}) \mid X^{(t)} \right] \mid X^{(0)} = x \right] \\
&= E\left[E\left[f(X^{(t+s)}) \mid \mathcal{F}_t \right] \mid X^{(0)} = x \right] \\
&= E\left[f(X^{(t+s)}) \mid X^{(0)} = x \right] = T^{t+s} f(x)
\end{aligned}
$$

If $\quad Tf(x) = \int f(y)\, p(y \mid x)\, dy$

then p is called transition probability density
or kernel

Ergodicity

In many contexts (including Monte Carlo)
we want to know if

$$\bar{f}_N = \frac{1}{N} \sum_{k=1}^{N} f(X^{(k)}) \longrightarrow \pi[f]$$

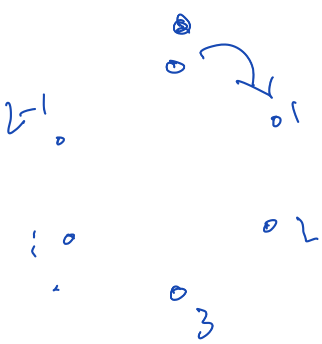
where π is some probability distribution

We may also want to know if the distribution of $X^{(t)}$ converges to some distribution π

the distribution of $X^{(t)}$ does not have to converge for $\bar{f}_N \to \pi[f]$

ex) $P[X^{(t+1)} = (i+1) \bmod L \mid X^{(t)} = i] = 1$

$X^{(0)} = L-1$



$$\bar{f}_N = \frac{1}{N} \sum_{k=0}^{N-1} f(k \bmod L)$$

$$= \frac{1}{N}\left(\left\lfloor \frac{N}{L} \right\rfloor \sum_{i=0}^{L-1} f(i) + \sum_{i=0}^{N \bmod L} f(i) \right)$$

$$= \frac{1}{L} \sum_{i=0}^{L-1} f(i) + \mathcal{O}\left(\frac{L}{N}\right)$$

$$= \pi[f] + \mathcal{O}\left(\frac{L}{N}\right)$$

So $\pi_i = \frac{1}{L}$

But $P[X^{(t)} = i] = \begin{cases} 1 & \text{if } i = (t-1) \bmod L \\ 0 & \text{otherwise} \end{cases}$

which does not converge

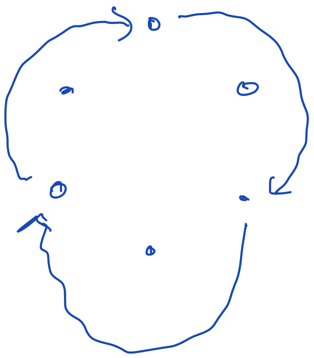
If we want the distribution of $X^{(t)}$ to converge to π then π should be invariant

$$\pi T = \pi$$

Given a particular π it is easy to construct T so that $\pi T = \pi$
(This is the setting of MCMC)

Given T it is usually impossible to identify π so that $\pi T = \pi$

ex) $P[X^{(t+1)} = (i+2) \bmod L \mid X^{(t)} = i] = 1$



$\pi_\alpha = \alpha \pi_e + (1-\alpha)\pi_o$ all invariant $\alpha \in [0,1]$

$\pi_e \propto \begin{cases} 1 & \text{on even sites} \\ 0 & \text{on odd sites} \end{cases}$

$\pi_o \propto \begin{cases} 1 & \text{on odd} \\ 0 & \text{on even} \end{cases}$

ex) $P[X^{(t+1)} = X^{(t)}] = P[X^{(t+1)} = X^{(t)} \pm 1] = 1/3$

$X^{(t)} \in \mathbb{Z}$

$$T = \begin{bmatrix} & \cdot & \cdot & \cdot & & \\ 0, \cdots 0, & \frac{1}{3}, & \frac{1}{3}, & \frac{1}{3}, & 0, \cdots, 0 \\ & & \cdot & \cdot & \cdot & \end{bmatrix}$$

$$v = \begin{bmatrix} \vdots \\ 1 \\ 1 \\ 1 \\ \vdots \end{bmatrix}$$

$v^T T = v^T$

But v isn't finite

A necessary condition to have a unique invariant distribution is that

$\forall$ B and any x, $\exists$ m

so that $P(X^{(m)} \in B \mid X^{(0)} = x) > 0$

the chain is "irreducible"

On a finite state space irreducible is the same as the eigenvalue $\lambda = 1$ of T having a 1-dimensional eigenspace.

This is a sensible requirement because the distribution $\mu_i(t) = P(X_t = i)$ evolves by power iteration

$$\mu^T(t+1) = \mu^T(t)\, T$$

so it can't converge to $\pi^T = \pi^T T$ unless π is unique.

Power iteration also won't converge if there are other eigenvalues $\lambda \neq 1$ w/ $|\lambda| = 1$.

If $\lambda \neq 1 \Rightarrow |\lambda| < 1$ then the chain is "aperiodic"

For finite state chain irreducible + aperiodic $\Longrightarrow$ ergodic.

A little bit about convergence of $P(X^{(t)} \in B)$ to $\pi(B)$

where $\pi T = \pi$

Coupling: We want to show that

$$\| \eta T^t - \pi \| \longrightarrow 0 \quad \text{in some norm}$$

(recall $\eta T^t(A) = P(X^{(t)} \in A)$ if $X^{(0)} \sim \eta$)

We'll do this by showing that T is a contraction

i.e. if η and ν are two initial prob distributions then

$$(*) \qquad \| \eta T - \nu T \| \leq (1-\alpha) \| \eta - \nu \| \quad \text{for } \alpha \in (0,1)$$

If so then

$$\begin{aligned} \| \eta T^t - \pi \| &= \| \eta T^t - \pi T^t \| \\ &\leq (1-\alpha) \| \eta T^{t-1} - \pi T^{t-1} \| \qquad (\text{by } (*)) \\ &\leq \cdots \leq (1-\alpha)^t \| \eta - \pi \| \longrightarrow 0 \quad \text{as } t \to \infty \end{aligned}$$

We'll use a very specific norm

$$\| \eta - \nu \| = \min_{\substack{X \sim \eta \\ Y \sim \nu}} P(X \neq Y)$$

The minimization is over the joint distribution of X and Y, but the marginal distribution of X has to be η and the marginal distribution of Y has to be μ.

This is called the total variation norm
It can also be written as

$$\|\eta - \mu\| = 2 \sup_A |\eta(A) - \mu(A)|$$

Let's assume that for some $\alpha > 0$ and prob dist ν

$$P[X^{(1)} \in A \mid X^{(0)} = x] \geq \alpha \nu(A) \qquad \forall x$$

"Doeblin condition"

We can write

$$\forall A \qquad P[X^{(1)} \in A \mid X^{(0)} = x] = \alpha \nu(A) + (1-\alpha) \underbrace{\left(\frac{P[X^{(1)} \in A \mid X^{(0)} = x] - \alpha \nu(A)}{1-\alpha} \right)}_{Q_x(A) \text{ is a prob distribution}}$$

Here's a recipe to generate $X^{(1)}$ from $X^{(0)}$

1. Select $Z \sim \text{Bernoulli}(\alpha)$
2. If $Z = 1$ draw $X^{(1)}$ from ν
   If $Z = 0$ draw $X^{(1)}$ from Q_x with $x = X^{(0)}$

If $X^{(0)} \sim \eta$ then $X^{(1)} \sim \eta T$

Now select $Y^{(0)} \sim \nu$ so that $P(X^{(0)} \neq Y^{(0)}) = \|\eta - \nu\|$

Generate $Y^{(1)}$ by

1. If $X^{(0)} = Y^{(0)}$ then set $Y^{(1)} = X^{(1)}$

otherwise 2. If $Z=1$ set $Y^{(1)} = X^{(1)}$

If $Z=0$ draw $Y^{(1)}$ from Q_x with $x = Y^{(0)}$

note $Y^{(1)} \sim \nu T$

$$P(X^{(1)} \neq Y^{(1)}) \leq (1-\alpha) P(X^{(0)} \neq Y^{(0)})$$
$$= (1-\alpha) \|\eta - \nu\|$$

so $\|\eta T - \nu T\| \leq (1-\alpha) \|\eta - \nu\|$

For an aperiodic, irreducible Markov chain we can choose m so that
$$(T^m)_{ij} > 0$$

let $\alpha = n \left(\min_{i,j} (T^m)_{ij} \right)$ and $v(i) = \frac{1}{n}$

then $(T^m)_{ij} \geq \alpha\, v(j) \quad \forall\, i,j$