2.8 & 31.7
MURPHY.

RECITATION – 04/21
MARKOV CHAINS

a) Modeling tool for sequences.

b) Inference tool (as ex MCMC) – by using markov chains to defining a stochastic system.

MODELING TOOL:

Consider a sequence of random variables, $X_1, \dots X_T$. The joint probability is:

$$P(X_1, \dots X_T) = P(X_1) \cdot P(X_2|X_1) \cdot P(X_3|X_2, X_1) \dots$$

$$= P(X_1) \prod_{t=2}^{T} P(X_t | X_{1:t-1})$$

\# of parameters grow with t.

Markovian Assumption:

First order: $P(X_1, \dots X_T) = P(X_1) P(X_2|X_1) P(X_3|X_2) \dots$

Second order: $P(X_1, \dots X_T) = P(X_1) \cdot P(X_2|X_1) \cdot P(X_3|X_2, X_1) \cdot P(X_4|X_3, X_2$

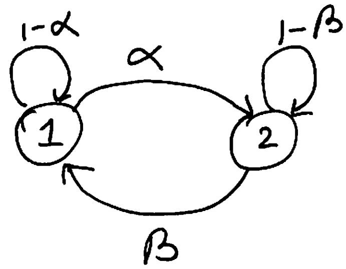
Note: however we can always define a new random variable $Y_t = \{X_t, X_{t+1}\}$

$$P(Y_{t+1}|Y_t) = P(X_{t+2}| X_{t+1}, X_t)$$

Thus without loss of generality we can just focus on first order

If X_t is discrete, a markov chain is completely defined by the transition matrix. (Assuming the chain is time invariant).

Consider:



$$A = \begin{bmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{bmatrix}.$$

$$A_{ij} = P(X_{t+1}=j \mid X_t = i)$$

We can further consider the n step transition

$$P(X_{t+n}=j \mid X_t = i) = \sum_{k=1}^{K} P(X_{t+m}=k \mid X_t = i) \cdot P(X_{t+n}=j \mid X_{t+m}=k)$$

$$\Downarrow$$

$$A_{ij}(n) = \sum_{k=1}^{K} A_{ik}(m) \cdot A_{kj}(n-m).$$

$$A(n) = A(m)A(n-m)$$

$$= AA \cdots A$$

$$= A^n$$

For some $n > 0$ if $A_{ij}(n) \geq 0 \Rightarrow$ state j is reachable from state i

$$\Pi_t(j) = P(X_t = j)$$

$$= \sum_{i=1}^{K} P(X_{t-1} = i) \cdot P(X_t = j \mid X_{t-1} = i)$$

$$= \sum_{i=1}^{K} \Pi_{t-1}(i)\, A_{ij}$$

In matrix notation Let $\Pi_t = \{\Pi_t(1) \dots \Pi_t(K)\} \in R^{1 \times K}$

$$\Pi_t = \Pi_{t-1} A$$

After repeating/iterating if we reach a stage when

$\Pi_t = \Pi_{t-1}$ ie $\Pi = \Pi A$ we say that the chain has reached it's <u>stationary distribution</u>.

Note* $\Pi \perp\!\!\!\perp \Pi_0$

ASIDE

Why Do we care about the stationary distribution??

→ Intractable integral. For instance GPC; interested in prediction $P(Y_* = 1 \mid X_*, Y, f, K) = \int P(Y_* \mid f_*)\, P(f_* \mid f, Y, K)\, df_*$

→ Can't compute the integral directly and we made approximated $P(f_* \mid f, Y, X_*, K) \approx \mathcal{N}(\cdot\, \hat{f}, \Sigma)$ and computed the integral.

→ Alternatively, we could evaluate the function at various f_*, and ~~add up the~~ compute the integral numerically

→ doesn't really scale to high dimensions.

→ What we really want is to sample from the true posterior distribution and use these samples to evaluate the required moments.

→ Setting up a Markov chain whose stationary distribution = posterior distribution, lets us do so.

$$\pi = \pi A$$

$$\Rightarrow \quad A^T \pi^T = \pi^T \Rightarrow A^T v = v$$

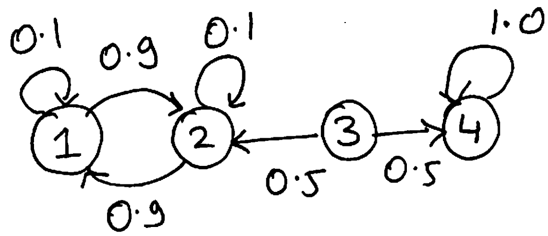
Thus given a markov chain; it's stationary distribution can be found by finding the eigenvector of A which has a eigenvalue $= 1$.

From PERRON-FROBENIUS Theorem,

$\Rightarrow$ for a row stochastic matrix the eigen value 1 is dominant. ie $\lambda_i = 1$ & $\lambda_{j : j \neq i} < \infty \quad \forall j$

$\Rightarrow$ Thus we only need to find the dominant eigenvector; efficient tools for this exist ex "Power Method"

When do unique stationary distributions exist?



⇒ If we start at node 4 we are stuck. (0, 0, 0, 1)

⇒ Starting at node 3 we could end up in either (0.5, 0.5, 0, 0)
or (0, 0, 0, 1)

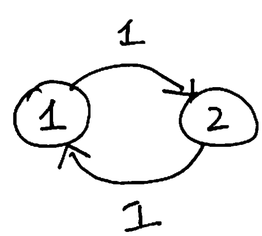
Thus the above chain has no unique stationary distribution.

But both connected components do.

IRREDUCIBILITY ⇒ All states in the chain must be reachable

APERIODICITY

⇒



if we start at state 1 we can be in state 2 only in the {2, 4, 6 . . .} time step

Period of 2 ⇒ 1

~~Stationary~~ Limiting distribution would depend on which state w start. Thus no stationary distribution.

If the period of all states in the chain is 1 the chain is <u>aperiodic</u>.

Intuitively, this means we can visit a state i at irregular times.

$\Rightarrow$ An irreducible and aperiodic finite markov chain ~~is~~ has a unique stationary distribution.

\# of states is finite.

<u>A quick test</u> :— Chain is singly connected and each state has a self loop. (Sufficient not necessary)

<u>DETAILED BALANCE test</u> :—

$$P(X_{t-1}=i)\,P(X_t=j\,|\,X_{t-1}=i) = P(X_{t-1}=j)\,P(X_t=i\,|\,X_{t-1}=j)$$

$$\Rightarrow \quad \pi_i A_{ij} = \pi_j A_{ji}$$

detailed balance equation.

If a markov chain satisfies the detailed balance equation wrt π then π is a stationary distribution of A.

$$\sum_i P(X_{t-1}=i)\,P(X_t=j\,|\,X_{t-1}=i) = \sum_i P(X_{t-1}=j)\,P(X_t=i\,|\,X_{t-1}=j)$$

$$= \pi_j$$

Given a desired π, a transition matrix A can be constructed to satisfy the detailed balance equations.