EC381/MN308
Probability and Some Statistics

Yannis Paschalidis

yannisp@bu.edu, http://ionia.bu.edu/

Dept. of Manufacturing Engineering

Dept. of Electrical and Computer Engineering

Center for Information and Systems Engineering

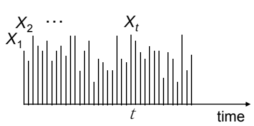
Lecture 21 - Outline

1. **Markov chains**

Chapter 12 Random Function
Stochastic Processes

Random function = stochastic process

= time-indexed sequence of random variables



t = Discrete or continuous. We focus only on discrete time

X_t = Discrete or continuous RVs.

12.1 Probability Models of Random function

$X_1, X_2, \ldots, X_t, \ldots$ are RVs defined on a sample space

Each individual RV is described, as usual, by:

- PMF $P[X_t]$ or PDF $f[X_t]$, CDF $F[X_t]$
- Mean $E[X_t]$ = function of time $\mu_X(t)$
- Variance Var$[X_t]$= function of time $\sigma_X^2(t)$

Each pair of RVs is described, as usual, by:

Joint PMF $P[X_{t1}, X_{t2}]$ or Joint PDF $f[X_{t1}, X_{t2}]$, Joint CDF $F[X_{t1}, X_{t2}]$

Conditional PMF $P[X_{t1} \mid X_{t2}]$ or PDF $f[X_{t1} \mid X_{t2}]$

Correlation $E[X_{t1} X_{t2}]$ Covariance Cov$[X_{t1} X_{t2}]$

Each triplets, quadruplets, or n-tuplets of RVs are similarly described.

The result is a hierarchy of multivariate joint PDFs, PMFs, CDFs, etc, with properties that are generalizations of the bivariate case.

Generalization 1: Joint PDF and CDF

$$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = \frac{\partial^n F_{X_1,\ldots,X_n}(x_1,\ldots,x_n)}{\partial x_1 \ldots \partial x_n}$$

(joint PDF) (joint PDF)

Generalization 2: Independence

Random variables $X_1, \ldots, X_n$ are independent if and only if joint probabilities are products of marginals

- Discrete: $P_{X_1,X_2,\ldots,X_n}(x_1,x_2,\ldots,x_n) = P_{X_1}(x_1)P_{X_2}(x_2)\ldots P_{X_n}(x_n)$
- Continuous: $f_{X_1,X_2,\ldots,X_n}(x_1,x_2,\ldots,x_n) = f_{X_1}(x_1)f_{X_2}(x_2)\ldots f_{X_n}(x_n)$

Generalization 3: Conditional Probabilities

Pairs: Tomorrow's value conditioned on today's value = $P_{X_2|X_1}(x_2 \mid x_1)$

$$P_{X_2,X_1}(x_2,x_1) = P_{X_2|X_1}(x_2 \mid x_1)P_{X_1}(x_1)$$

Triplets: Tomorrow's value conditioned on today's & yesterday's values $P_{X_3|X_2,X_1}(x_3 \mid x_2,x_1)$

$$P_{X_3,X_2,X_1}(x_3,x_2,x_1) = P_{X_3|X_2,X_1}(x_3 \mid x_2,x_1)P_{X_2,X_1}(x_2,x_1)$$
$$= P_{X_3|X_2,X_1}(x_3 \mid x_2,x_1)P_{X_2|X_1}(x_2 \mid x_1)P_{X_1}(x_1)$$

n-tuplets: Future value conditioned on present & past values $P_{X_{n+1}|X_n,\ldots,X_2,X_1}(x_{n+1} \mid x_n,\ldots,x_2,x_1)$

$$P_{X_{n+1},X_n,\ldots,X_1}(x_{n+1},x_n,\ldots,x_1)$$
$$= P_{X_{n+1}|X_n,\ldots,X_1}(x_{n+1} \mid x_n,\ldots,x_1)P_{X_n,X_{n-1},\ldots,X_1}(x_n,x_{n-1},\ldots,x_1)$$
$$= P_{X_{n+1}|X_n,\ldots,X_1}(x_{n+1} \mid x_n,\ldots,x_1)\ldots P_{X_2|X_1}(x_2 \mid x_1)P_{X_1}(x_1)$$

Generalization 4: Correlation Matrix and Correlation Function

The correlation between pairs of RVs at times t_1 and t_2 is the function function $R_X(t_1, t_2) = E[X_{t1} X_{t2}]$

If t is discrete $R_X(t_1, t_2) = E[X_{t1} X_{t2}]$ is an array that can be cast as a matrix

$$R_{\mathbf{X}} = E\left[(\mathbf{X})(\mathbf{X})^T\right] \qquad \mathbf{X} = [X_1, \ldots, X_n]^T$$

$$= E\left[\begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}\begin{pmatrix} X_1 & \cdots & X_n \end{pmatrix}\right]$$

$$= \begin{pmatrix} E[X_1]^2 & E[X_1X_2] & \cdots & E[X_1X_n] \\ E[X_2X_1] & E\left[X_2^2\right] & \cdots & E[X_2X_n] \\ \vdots & \vdots & \ddots & \vdots \\ E[X_nX_1] & E[X_nX_2] & \cdots & E\left[X_n^2\right] \end{pmatrix}$$

Generalization 6: Covariance Matrix

$$\sum_{\mathbf{X}} = E\left[(\mathbf{X}-\boldsymbol{\mu}_X)(\mathbf{X}-\boldsymbol{\mu}_X)^T\right]$$

$$= E\left[\begin{pmatrix} (X_1-\mu_{X_1})^2 & (X_1-\mu_{X_1})(X_2-\mu_{X_2}) & \cdots & (X_1-\mu_{X_1})(X_n-\mu_{X_n}) \\ (X_1-\mu_{X_1})(X_2-\mu_{X_2}) & (X_2-\mu_{X_2})^2 & \cdots & (X_n-\mu_{X_n})(X_2-\mu_{X_2}) \\ \vdots & \vdots & \ddots & \vdots \\ (X_1-\mu_{X_1})(X_n-\mu_{X_n}) & (X_n-\mu_{X_n})(X_2-\mu_{X_2}) & \cdots & (X_n-\mu_{X_n})^2 \end{pmatrix}\right]$$

$$= \begin{pmatrix} E\left[X_1-\mu_{X_1}\right]^2 & E\left[(X_1-\mu_{X_1})(X_2-\mu_{X_2})\right] & \cdots & E\left[(X_1-\mu_{X_1})(X_n-\mu_{X_n})\right] \\ E\left[(X_1-\mu_{X_1})(X_2-\mu_{X_2})\right] & E\left[(X_2-\mu_{X_2})^2\right] & \cdots & E\left[(X_n-\mu_{X_n})(X_2-\mu_{X_2})\right] \\ \vdots & \vdots & \ddots & \vdots \\ E\left[(X_1-\mu_{X_1})(X_n-\mu_{X_n})\right] & E\left[(X_n-\mu_{X_n})(X_2-\mu_{X_2})\right] & \cdots & E\left[(X_n-\mu_{X_n})^2\right] \end{pmatrix}$$

$$= \begin{pmatrix} \sigma_{X_1}^2 & \mathrm{Cov}[X_1, X_2] & \cdots & \mathrm{Cov}[X_1, X_n] \\ \mathrm{Cov}[X_2, X_1] & \sigma_{X_2}^2 & \cdots & \mathrm{Cov}[X_2, X_n] \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}[X_n, X_1] & \mathrm{Cov}[X_n, X_2] & \cdots & \sigma_{X_n}^2 \end{pmatrix}$$

$$E[\mathbf{X}] = E\left[\begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}\right] = \begin{pmatrix} E[X_1] \\ \vdots \\ E[X_n] \end{pmatrix} \equiv \mu_X$$

Generalization 7: Gaussian Random Function

n random variables $\mathbf{X} = [X_1, \ldots, X_n]^T$ are jointly continuous Gaussian if their joint PDF is

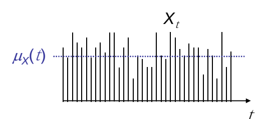
$$f_{\mathbf{X}}(\mathbf{x}) = Ce^{-Q(\mathbf{x}-\mu_{\mathbf{X}})}$$

–Quadratic exponent: $Q(\mathbf{x}) = \dfrac{1}{2}\mathbf{x}^T \sum_X^{-1} \mathbf{x}$

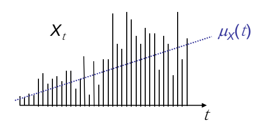
–Constant: $C = \dfrac{1}{(2\pi)^{n/2} \det\left(\sum_{\mathbf{X}}\right)^{1/2}}$

New Concepts

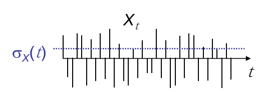
Stationarity (Time Invariance)



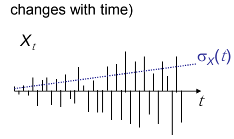
Stationary



Nonstationary
(distribution, e.g., mean, changes with time)



Stationary



Nonstationary
(distribution, e.g., variance, changes with time)

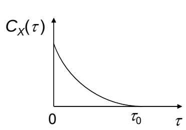
Correlation Function of a Stationary Stochastic Process

$R_X(t_1, t_2) = E[X_{t1} X_{t2}]$ is a function of only the time difference (lag) $t_2 - t_1$, i.e.

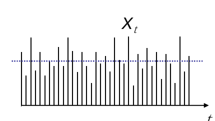
$R_X(t, t + \tau)$ is independent of t, and is usually written as $R_X(\tau)$.

Covariance function $C_X(\tau) = R_X(\tau) - \mu_X$

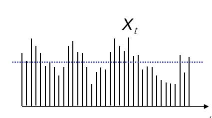
The width of the function $C_X(\tau)$ is regarded as the memory time of the stochastic process and is called the correlation time.



Values of the stochastic process at times separated by a time > τ_0 are uncorrelated.
Values at times separated by time << τ_0 are highly correlated



Stationary stochastic process
With short correlation time



Stationary stochastic process
With longer correlation time

12.2 Markov (Stochastic) Processes

A discrete-time stochastic process for which

$P\left[x_{t+1} \middle| x_t, \ldots, x_1\right] = P\left[x_{t+1} \middle| x_t\right]$

i.e., future depends only on present state, is called Markov

For a Markov process

$P\left[x_t, \ldots, x_1, x_0\right] = P\left[x_t \middle| x_{t-1}\right] P\left[x_{t-2} \middle| x_{t-3}\right] \cdots P\left[x_2 \middle| x_1\right] P\left[x_1 \middle| x_0\right] P\left[x_0\right]$

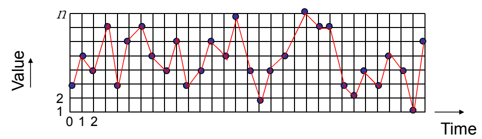
i.e., the n-tuplet joint PDF is dependent only on the pair joint PDF.

Knowledge of $P\left[x_{t+1} \middle| x_t\right]$, called the transition probability for all times n, and the initial PDF $P[x_0]$ is sufficient to know all statistical properties of the process

If a Markov process is also stationary, its transition probability is independent of time t

Markov Chains

Consider a stationary Markov process X_t
Assume discrete time and discrete values (states)



The transition probability $P[X_t|X_{t-1}]$ is an n x n matrix $\boldsymbol{P}$ with elements:

$P_{ij} = P[X_t = j \mid X_{t-1} = i]$ = Probability that state j follows state i
Can allow infinite n.

The matrix P has "special" structure: every row sums to 1 (stochastic matrix)

$$P_{ij} \geq 0, \ \forall i, j, \qquad \sum_j P_{ij} = 1, \ \forall j$$

Example: Speech Generation

Sounds on a sentence are not chained randomly

- We don't pronounce "GRXZZT"
- We tend to follow consonants with vowel sounds, but choice of vowel varies, depending on context
- Sounds are followed randomly by other sounds
  - "t" followed by s, a, e, i, o, u. The specific sound following "t" is random, with probability distribution P("sound"|"t")
  - Similar models developed for other fundamental sounds (phonemes)
  - Probabilities estimated from speech recordings (SLLN!)

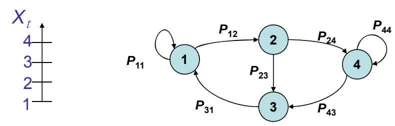
Example: Text Generation

Words are not chained randomly from a dictionary

- Grammar rules, sentence construction, etc.
- Words followed by other words obey certain probabilities

Markov Chain as a graph

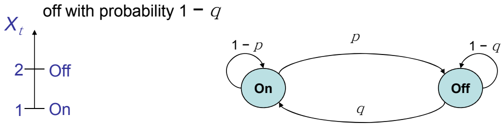
- Nodes represent the states (values of X_t)
- Arcs are transition probabilities



- Key property: Sum of probabilities on arcs starting at a single node must equal 1

Two-State (On-Off) System

- When on, system fails at the next time with probability p, stays on with probability $1 - p$
- When off, system is turned on at next time with probability q, left off with probability $1 - q$

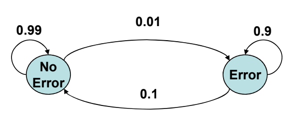


- Markov chain matrix: $\boldsymbol{P} = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}$

Example: Quiz 12.1

Wireless packet channel has clustered errors

- Prob. error following another error is 0.9
- Prob. no error following another no error is 0.99
- Sketch Markov chain, compute matrix $\boldsymbol{P}$

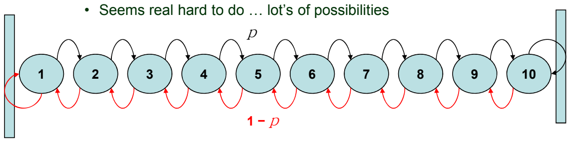


$$\boldsymbol{P} = \begin{bmatrix} 0.99 & 0.01 \\ 0.1 & 0.9 \end{bmatrix}$$

Example: One step forward, one step backward

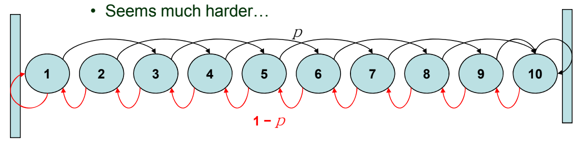
Professor blindfolded in corridor, length 10 steps

- Initially placed in middle (position 5)
- With probability p, Prof. takes one step to right
- With probability $(1 - p)$, Prof. takes one step to left
- If Prof. hits wall, says ouch but does not move…
- What is probability that Prof. hits right wall at time 10? (one step per time…)
  - Seems real hard to do … lot's of possibilities



Example: Two step forward, one step backward

- Same blindfolded Prof. in same corridor
  - Initially placed in middle (position 5)
  - With probability p, Professor takes TWO steps to right
  - With probability $(1 - p)$, Prof. takes one step to left
  - If Prof. hits wall, says ouch and stops…
  - What is probability that Prof. hits right wall at time 10? (one attempted move per time…)
    - Seems much harder…



Example: "Drunken" Knight's tour

– Chess board
– White knight in starting right position
– At each move, randomly pick one of the "legal" moves, with equal probability

Compute the expected number of moves until the knight visits every square!

Markov chain with finite state (n = 64) and simple transition.
Much harder problem!!

Computing with Markov Chains

A Markov chain is completely characterized by:
transition probabilities $P[X_{t+1}|X_t]$ & initial probability $P[X_0]$.
Other joint, conditional, or marginal probabilities may be determined from these two quantities.

Compute $P[X_{t+2}|X_t]$

Conditional probability definition: $P[X_{t+2}, X_{t+1} \mid X_t] = P[X_{t+2}|X_{t+1}, X_t]\, P[X_{t+1}|X_t]$

Markov Property: $P[X_{t+2}|X_{t+1}, X_t] = P[X_{t+2}|X_{t+1}]$

Marginalization

$$P(X_{t+2}|X_t) = \sum_{X_{t+1}} P(X_{t+2}, X_{t+1}|X_t)$$
$$= \sum_{X_{t+1}} P(X_{t+2}|X_{t+1}) P(X_{t+1}|X_t)$$

↑ Sum over values of unwanted RV

If X_t is discrete with finite number of states n

$$P(X_{t+2} = j|X_t = i) = \sum_{k=1}^{n} P(X_{t+2} = j|X_{t+1} = k) P(X_{t+1} = k|X_t = i)$$
$$= \underbrace{\sum_{k=1}^{n} P_{ik}(t) P_{kj}(t+1)}_{\text{matrix multiplication!}}$$

- ➢ $P(X_{t+1} = k \mid X_t = i)$ is the ik-th entry of P(t)
- ➢ $P(X_{t+2} = j \mid X_{t+1} = k)$ is the kj-th entry of P(t+1)
- ➢ $P(X_{t+2} = j \mid X_t = i)$ is the ij-th entry of P(t)P(t+1)

By induction,

$P[X_{t+k+1} = j \mid X_t = i]$ = ij element of matrix $P(t).\ P(t+1) \dots P(t+k)$

Chapman-Kolmogorov equation

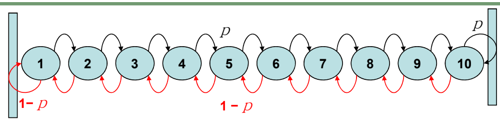
For any Markov chain with a finite number of states and a stationary transition probability matrix $\mathbf{P}$, if $P(k)$ is the transition matrix for k steps, i.e., $P_{ij}(k) \equiv P[X_{t+k} = j|X_t = i]$ then

$$P(k) = \mathbf{P}^k$$

The k-step transition matrix equals the 1-step transition matrix raised to the power k

$$P_{ij}(k+m) = \sum_{x=1}^{n} P_{ix}(k) P_{xj}(m)$$
$$= \sum_{x=1}^{n} P_{ix}(m) P_{xj}(k)$$
$$= (P^k P^m)_{ij}$$

Example: One step forward, one step backward



Initially placed at position 5; one step to right with probability $p = ½$; one step to left with probability $1-p$; if wall is hit, stay put.
What is probability of being in position 10 at time $t = 10$?

This is a stationary Markov chain with $n = 10$ *states*

Transition matrix

$P_{ij} = P[X_t = j \mid X_{t-1} = i]$

$$P = \begin{pmatrix} 1-p & p & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1-p & 0 & p & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1-p & 0 & p & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1-p & 0 & p & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1-p & 0 & p & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1-p & 0 & p & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1-p & 0 & p & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1-p & 0 & p & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1-p & 0 & p \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1-p & p \end{pmatrix}$$

- At $t = 0$, $P[X_0 = 5] = 1$
- Probability of being in position 10 at time $t = 10$ is:
  $P[X_{10} = 10] = (\boldsymbol{P}^{10})_{5,10}$
- Go to MATLAB, compute $\boldsymbol{P}^{10}$, read the (5,10) element:
  $(\boldsymbol{P}^{10})_{5,10} = 0.0439$
- Probability of hitting the wall at $t = 11$ is ½ of this:
  $P[X_{11} = 11] = P[X_{10} = 10] \times p = 0.02195$

What if we did not know where we started?

- $P(X_0)$ is a probability distribution
  - Represent it as a column vector

$$P(X_0) \equiv \underline{p}(0) = \begin{pmatrix} P(X_0 = 1 \\ P(X_0 = 2) \\ \vdots \\ P(X_0 = n) \end{pmatrix}$$

- How does this vector evolve over time?

Evolution of Probability Vectors

- Computing $P(X_1) = \underline{p}(1)$

$$\begin{aligned} P(X_1 = j) &= \sum_{i=1}^{n} P(X_1 = j, X_0 = i) \ (\textit{marginalize}) \\ &= \sum_{i=1}^{n} P(X_1 = j \mid X_0 = i) P(X_0) = i) \ (\textit{conditional def.}) \\ &= \sum_{i=1}^{n} \underline{p}(0)_i P_{ij} \\ &= (P^T \underline{p}(0))_j = (\underline{p}(0)^T P)_j \end{aligned}$$

➔The j-th entry of $\underline{p}(1)$ is the j-th element of the vector $P^T\ \underline{p}(0)$

Superscript T indicates transpose of a matrix or vector

General Form

- In general, it is easy to compute the evolution of probability vectors
  - Recursion

$$\underline{p}(t) = P^T \underline{p}(t-1), t = 1, \ldots,$$
$$\underline{p}(0) \text{ given}$$

  - Linear system!
- Key: Since P is a stochastic matrix, this conserves probability

$$\sum_{i=1}^{n} \underline{p}(t)_i = 1$$

Steady State Behavior of Homogeneous Markov Chains

- If it is a linear system, what happens as t → ∞?
  - Under many (but not all…) conditions, this will go to a limit!

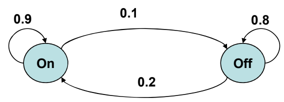
$$\lim_{t \to \infty} \underline{p}(t) = \underline{\pi}$$

  - If so, this limit will be a probability distribution because all terms in the sequence are also probability vectors
    - Nonnegative entries
    - Sums to 1
  - The limit may or may not depend on the initial probability distribution $\underline{p}(0)$

Example

- Two state, on-off system
  - p = 0.1, q = 0.2
  - Initially on: $\underline{p}(0) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$
  - Transition matrix:

$$P = \begin{pmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{pmatrix}$$



  - MATLAB: $(P^T)^n \underline{p}(0)$

$$\underline{p}(5) = \begin{pmatrix} 0.7227 \\ 0.2773 \end{pmatrix} \quad \underline{p}(9) = \begin{pmatrix} 0.6801 \\ 0.3199 \end{pmatrix} \quad \underline{p}(15) = \begin{pmatrix} 0.6667 \\ 0.3333 \end{pmatrix}$$

Can we find the limit directly?

- The limit must satisfy $P^T\underline{\pi} = \underline{\pi}$
  - $\underline{\pi}$ is a "stationary" probability vector
  - Not necessarily unique
- A bit of linear algebra:
  - $\underline{\pi}$ is an eigenvector of P^T, corresponding to eigenvalue 1
  - All stationary probability vectors are eigenvectors of P^T corresponding to eigenvalue 1
  - If the eigenvalue 1 is repeated, there can be multiple stationary probability vectors

Back to Example

- Find the eigenvalues and eigenvectors of P^T:
  - MATLAB: [V,D] = eig(P')

$$V = \begin{pmatrix} 0.8944 & -0.7071 \\ 0.4472 & 0.7071 \end{pmatrix} \qquad D = \begin{pmatrix} 1.0000 & 0 \\ 0 & 0.7000 \end{pmatrix}$$

  - Eigenvalue corresponding to 1 is what we want, but have to scale it so it sums up to 1

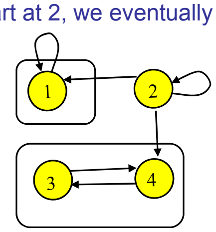
$$\pi = \frac{1}{0.8944 + 0.4472} \begin{pmatrix} 0.8944 \\ 0.4472 \end{pmatrix} = \begin{pmatrix} 0.6667 \\ 0.3333 \end{pmatrix}$$

Good Markov Chains

- Desirable properties
  - Converge to a stationary distribution
  - Which is unique
  - And every state has nonzero limiting probability
- Would like to identify conditions when this occurs

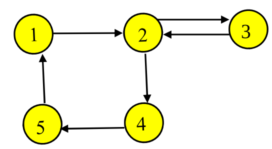
Bad Case

- If we start at 1, we never leave 1
- If we start at 3, we never reach 2
- If we start at 2, we eventually leave 2 and never return



Bad Case 2

- Periodic states
  - If we start at 1, we can only return to A after an even number of time steps
    - Probability of being in 1 is zero for all odd times!
    - ➔ Can't have a limit



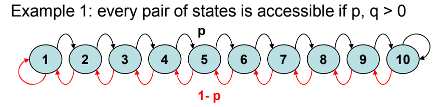
Recognizing Structure of Markov Chain

- We can identify conditions for the following:
  - The Markov chain will have at least one stationary distribution
  - The Markov chain will have a unique stationary distribution
  - The stationary distribution will have nonzero probability at a given state
- Entirely from the diagram of the Markov chain!
  - Just have to classify states

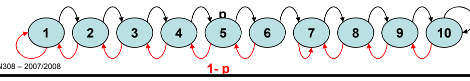
A bunch of definitions…

- State j is accessible from state k if $P_{ij}(t) > 0$ for some t
  - Graph: there is a directed path with nonzero probability from i to j in the graph

Example 1: every pair of states is accessible if p, q > 0

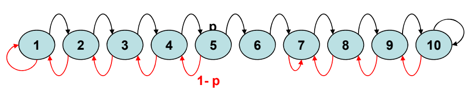


Example 2: 7 accessible from 1, but 1 not accessible from 7



Communicating Class

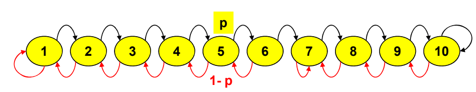
- States i and j communicate if i is accessible from j and j is accessible from i
  - Convention: every state communicates with itself: j communicates with j
  - Note: communicate is a transitive relationship: i communicates with j, j communicates with k → i communicates with k
- A communicating class is a non-empty set of states that communicate with each other
  - Communicating classes {1,2,3,4,5}, {6}, {7,8,9,10}



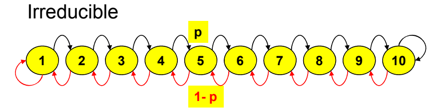
Irreducible Markov Chains

- A Markov chain is irreducible if all of its states belong to a single communicating class
  - Every state communicates with every other state

Not irreducible



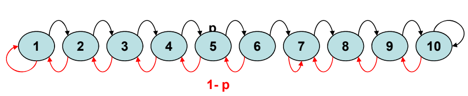
Irreducible



Transient States

- A state j is transient if there exists a second state k so that j communicates with k, but k does not communicate with j
  - Nontransient states are called recurrent
  - All states in a communicating class are either transient or recurrent

Example: States {1,2,3,4,5,6} are transient
States {7,8,9,10} are recurrent

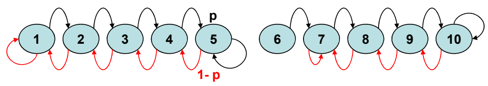


Transient vs Recurrent States

- Transient states: if there is a stationary probability distribution $\underline{\pi}$, then $\pi_j = 0$ if j is transient
- Every Markov chain has at least one recurrent communicating class
  - If there is a stationary probability distribution $\underline{\pi}$, then $\underline{\pi}$ will be non-zero only on recurrent communicating classes
- Multiple recurrent classes → stationary distribution depends on initial probability vector

Example

- Illustration of multiple stationary distributions

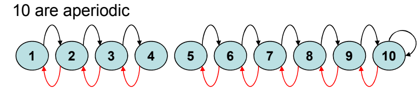


  - Two recurrent classes
  - One transient state {6}
  - If initial probability is at 1, then stationary distribution has $\underline{\pi}_7 = 0$
  - If initial probability is at 10, then stationary distribution has $\underline{\pi}_1 = 0$

So, When Do Limits Exist?

- Period of a state:
  - The period d of a state j is the greatest common divisor of the length of all cycles from j to itself
    - Equivalent but obscure definition: Period d is the largest integer such that $P(n)_{jj} = 0$ if n not divisible by d
    - If period is 1, then state is aperiodic
  - A Markov chain is periodic if all states have the same period d > 1

  Example: Periods of states {1,2,3,4} is 2. States 5, 6, 7, 8, 9, 10 are aperiodic

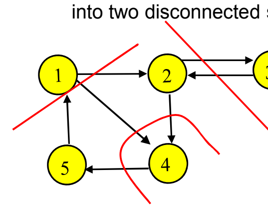


Finally, What are Good Markov Chains?

- Markov Chain irreducible, aperiodic ➔ Good!
  - There is a unique stationary probability distribution
  - Every state has positive probability in the limit
  - Every initial condition leads to the same unique stationary distribution
- Such a Markov chain is called ergodic
  - The stationary distribution is the unique eigenvector of P^T corresponding to eigenvalue 1
    - Normalized to sum to 1

Do we need to solve eigenvalue problems to compute $\underline{\pi}$?

- Graphical characterization: Probability balance
  - Given any cut, the net flow of probability across that cut must be zero at equilibrium
  - Cut: set of arcs that when removed, separates nodes into two disconnected sets



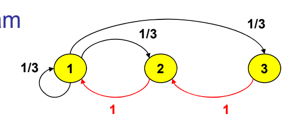
$$P_{14}\pi_1 + P_{12}\pi_2 = P_{51}\pi_5$$
$$P_{24}\pi_2 + P14\pi_1 = P_{45}\pi_4$$
$$P_{23}\pi_2 = P_{32}\pi_3$$

Quiz 12.5

- Renumber the states 1 to 3
- N random variable, values in {1,2,3}, discrete uniform
- Markov chain: at each step, get an independent sample of N.
  - If at state 1, the next state is N
  - If at another state j, the next state is j -1

Quiz 12.5 cont.

- Diagram



  - 1 Recurrent class {1,2,3}
  - Aperiodic (cycle of length 1)

  ➔ Ergodic!

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

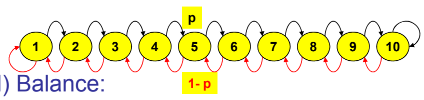
Using cuts, get these simple equations:

$$\pi_1/3 = \pi_3$$
$$2\pi_1/3 = \pi_2$$

$$\Rightarrow \quad \underline{\pi} = \begin{pmatrix} 1/2 \\ 1/3 \\ 1/6 \end{pmatrix}$$

Harder example

- p = 0.4



- (Detailed) Balance:

$$0.4\pi_1 = 0.6\pi_2;\ \ 0.4\pi_2 = 0.6\pi_3;\ \ \ldots\ \ 0.4\pi_9 = 0.6\pi_{10}$$

$$\pi_{i+1} = \frac{2}{3}\pi_i,\ \ i = 1,\ldots,9$$

$$\Rightarrow \pi_i = (\frac{2}{3})^{i-1}\pi_1, i = 1,\ldots,10$$

- Couple with fact that probabilities sum to 1:

$$\sum_{i=1}^{10}(\frac{2}{3})^{i-1}\pi_1 = \frac{1-(2/3)^{10}}{1/3}\pi_1 = 1$$

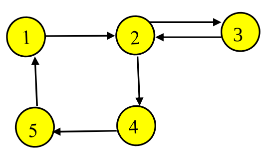
$$\Rightarrow \pi_1 = .3392; \pi_2 = 0.2261; \pi_3 = 0.1508; \ldots; \pi_{10} = 0.0088$$

So what can we say about "Bad" Markov Chains in the limit?

- Finite state, irreducible ➔ there is a *unique stationary distribution*
  - Unique eigenvector of P^T corresponding to eigenvalue 1, normalized to sum to 1
- If Markov chain is periodic, and starts at this distribution, then it will be stationary…
  - Otherwise, the Markov chain will eventually oscillate between several distributions at a period equal to the Markov chain period

Illustration

- Period 2 chain
  - $P_{23}=P_{24} = 0.5$

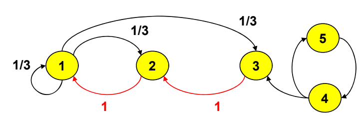


- Probabilities oscillate between these two:

$$\underline{\pi}^1 = \begin{pmatrix}1/3\\0\\1/3\\1/3\\0\end{pmatrix};\qquad \underline{\pi}^2 = \begin{pmatrix}0\\2/3\\0\\0\\1/3\end{pmatrix};$$

Limits of "Bad" Markov Chains

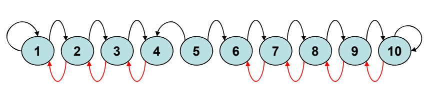
- If Markov chain has transient states, their stationary probability distribution will be zero!
  - An aperiodic Markov chain with a single recurrent class will have its stationary distribution distributed only on that recurrent class
  - Can ignore transient states→ Smaller chain…



$$\underline{\pi} = \begin{pmatrix}1/2\\1/3\\1/6\\0\\0\end{pmatrix}$$

Limits of "Bad" Markov Chains

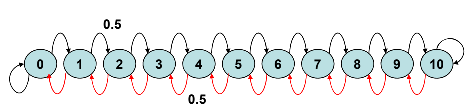
- If there are multiple recurrent classes,
  - There will be multiple stationary distributions
  - The matrix P will have multiple eigenvalues equal to 1
  - The limit distribution will depend on the initial distribution



Useful Example: Reflected Random Walk

$$X_t \in \{0,1,\ldots,10\},\quad X_0 = 0$$

$$P(X_{t+1}|X_t) = \begin{cases}0.5 & X_{t+1} = X_t + 1, X_t \in \{0,\ldots,9\}\\ 0.5 & X_{t+1} = X_t - 1, X_t \in \{1,\ldots,10\}\\ 0.5 & X_{t+1} = X_t, X_t = 10\\ 0.5 & X_{t+1} = X_t, X_t = 0\end{cases}$$



Reflected Random Walk

$$P = \begin{pmatrix} 0.5 & 0.5 & 0 & \cdots & 0 \\ 0.5 & 0 & 0.5 & \ddots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0.5 & 0.5 \end{pmatrix}$$

Detailed balance:

$$\begin{aligned} \pi_0 &= \pi_1 \\ \pi_9 &= \pi_{10} \\ \pi_i &= \pi_{i+1}, i = 1, \ldots, 9 \\ \Rightarrow \pi_i &= \frac{1}{11} \end{aligned}$$