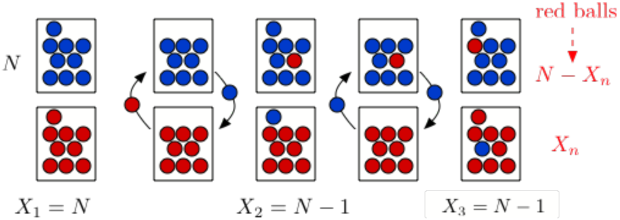
Today

Finish up Conditional Expectation.

Markov Chains.

Application: Mixing



Each step, pick ball from each well-mixed urn. Transfer it to other urn.
Let X_n be the number of red balls in the bottom urn at step n.
What is $E[X_n]$?

Given $X_n = m$, $X_{n+1} = m+1$ w.p. p and $X_{n+1} = m-1$ w.p. q

where $p = (1-m/N)^2$ (B goes up, R down)
and $q = (m/N)^2$ (R goes up, B down).

Thus,
$E[X_{n+1}|X_n] = X_n + p - q = X_n + 1 - 2X_n/N = 1 + \rho X_n,\ \rho := (1-2/N)$.

Mixing

We saw that $E[X_{n+1}|X_n] = 1 + \rho X_n,\ \rho := (1 - 2/N)$.
Does that make sense? Decreases: $X_n > n/2$. Increases: $X_n < n/2$.
Hence,

$$\begin{aligned}
&E[X_{n+1}] = 1 + \rho E[X_n] \\
&E[X_2] = 1 + \rho N; E[X_3] = 1 + \rho(1 + \rho N) = 1 + \rho + \rho^2 N \\
&E[X_4] = 1 + \rho(1 + \rho + \rho^2 N) = 1 + \rho + \rho^2 + \rho^3 N \\
&E[X_n] = 1 + \rho + \cdots + \rho^{n-2} + \rho^{n-1} N.
\end{aligned}$$

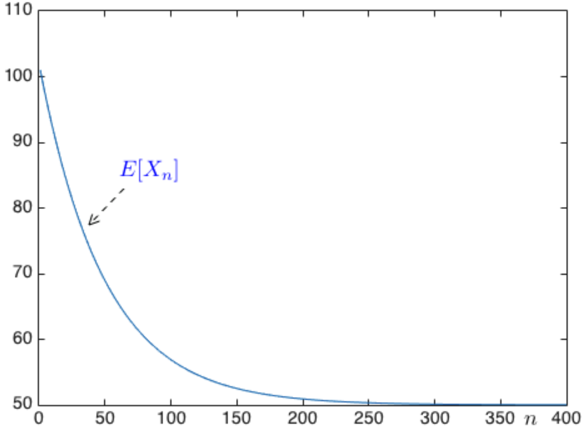
Hence,

$$E[X_n] = \frac{1 - \rho^{n-1}}{1 - \rho} + \rho^{n-1} N, n \geq 1.$$

As $n \to \infty$, goes to $N/2$.
Since $1 - \rho = 2/N$. And $\rho^n \to 0$.

Application: Mixing

Here is the plot.



Application: Going Viral

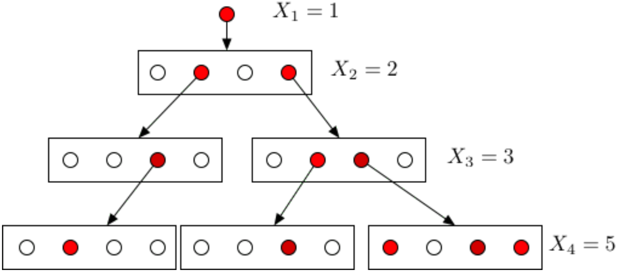
Consider a social network (e.g., Twitter).

You start a rumor (e.g., Rao is bad at making copies).

You have d friends. Each of your friend retweets w.p. p.

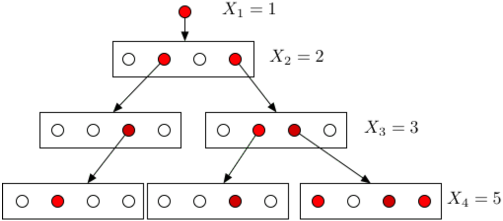
Each of your friends has d friends, etc.

Does the rumor spread? Does it die out (mercifully)?



In this example, $d = 4$.

Application: Going Viral



Fact: Number of tweets $X = \sum_{n=1}^{\infty} X_n$ where X_n is tweets in level n. Then, $E[X] < \infty$ iff $pd < 1$.

Proof:

Given $X_n = k$, $X_{n+1} = B(kd, p)$. Hence, $E[X_{n+1}|X_n = k] = kpd$.

Thus, $E[X_{n+1}|X_n] = pdX_n$. Consequently, $E[X_n] = (pd)^{n-1}, n \geq 1$.

If $pd < 1$, then $E[X_1 + \cdots + X_n] \leq (1 - pd)^{-1} \implies E[X] \leq (1 - pd)^{-1}$.

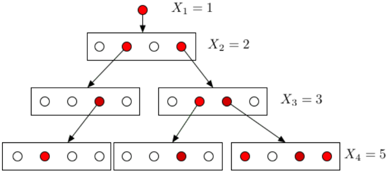
If $pd \geq 1$, then for all C one can find n s.t.

$$E[X] \geq E[X_1 + \cdots + X_n] \geq C.$$

□

In fact, one can show that $pd \geq 1 \implies Pr[X = \infty] > 0$.

Application: Going Viral



An easy extension: Assume that everyone has an independent number D_i of friends with $E[D_i] = d$. Then, the same fact holds.

Why? Given $X_n = k$.
$D_1 = d_1, \ldots, D_k = d_k$ – numbers of friends of these X_n people.
$\Longrightarrow X_{n+1} = B(d_1 + \cdots + d_k, p)$. Hence,

$$E[X_{n+1}|X_n = k, D_1 = d_1, \ldots, D_k = d_k] = p(d_1 + \cdots + d_k).$$

Thus, $E[X_{n+1}|X_n = k, D_1, \ldots, D_k] = p(D_1 + \cdots + D_k)$.

Consequently, $E[X_{n+1}|X_n = k] = E[p(D_1 + \cdots + D_k)] = pdk$.

Finally, $E[X_{n+1}|X_n] = pdX_n$, and $E[X_{n+1}] = pdE[X_n]$.

We conclude as before.

Application: Wald's Identity

Here is an extension of an identity we used in the last slide.

Theorem Wald's Identity

Assume that $X_1, X_2, \dots$ and Z are independent, where

Z takes values in $\{0, 1, 2, \dots\}$

and $E[X_n] = \mu$ for all $n \geq 1$.

Then,

$$E[X_1 + \cdots + X_Z] = \mu E[Z].$$

Proof:

$E[X_1 + \cdots + X_Z | Z = k] = \mu k$.

Thus, $E[X_1 + \cdots + X_Z | Z] = \mu Z$.

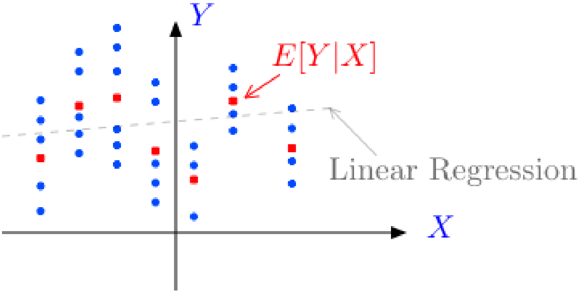
Hence, $E[X_1 + \cdots + X_Z] = E[\mu Z] = \mu E[Z]$. $\square$

CE = MMSE

Theorem
$E[Y|X]$ is the 'best' guess about Y based on X.

Specifically, it is the function $g(X)$ of X that

$$\text{minimizes } E[(Y - g(X))^2].$$



CE = MMSE

Theorem CE = MMSE

$g(X) := E[Y|X]$ is the function of X that minimizes $E[(Y - g(X))^2]$.
Proof:
Let $h(X)$ be any function of X. Then

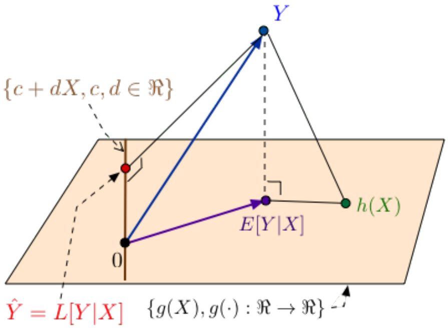
$$\begin{aligned} E[(Y-h(X))^2] &= E[(Y-g(X)+g(X)-h(X))^2] \\ &= E[(Y-g(X))^2]+E[(g(X)-h(X))^2] \\ &\qquad +2E[(Y-g(X))(g(X)-h(X))]. \end{aligned}$$

But,

$$E[(Y-g(X))(g(X)-h(X))] = 0 \text{ by the projection property.}$$

Thus, $E[(Y-h(X))^2] \geq E[(Y-g(X))^2]$. □

$E[Y|X]$ and $L[Y|X]$ as projections



$L[Y|X]$ is the projection of Y on $\{a+bX, a, b \in \Re\}$: LLSE
$E[Y|X]$ is the projection of Y on $\{g(X), g(\cdot) : \Re \to \Re\}$: MMSE.

Functions of X are linear subspace?
Vector $(g(X(\omega_1), \ldots, g(X(\omega_\Omega))$.
Coordinates ω and ω' with $X(\omega) = X(\omega')$
have same value: $v_\omega = v_{\omega'}$.
Linear constraints! Linear Subspace.

Summary

Conditional Expectation

- Definition: $E[Y|X] := \sum_y yPr[Y = y|X = x]$
- Properties: Linearity, $Y - E[Y|X] \perp h(X)$; $E[E[Y|X]] = E[Y]$
- Some Applications:
  - Calculating $E[Y|X]$
  - Diluting
  - Mixing
  - Rumors
  - Wald
- MMSE: $E[Y|X]$ minimizes $E[(Y - g(X))^2]$ over all $g(\cdot)$

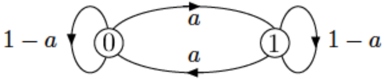
CS70: Markov Chains.

Markov Chains 1

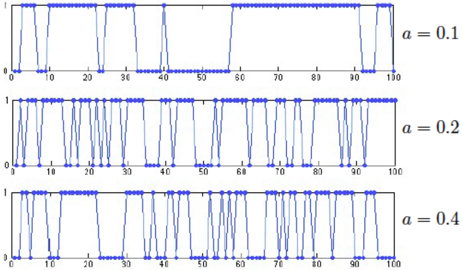
1. Examples
2. Definition
3. First Passage Time

Two-State Markov Chain

Here is a symmetric two-state Markov chain. It describes a random motion in $\{0, 1\}$. Here, *a* is the probability that the state changes in the next step.

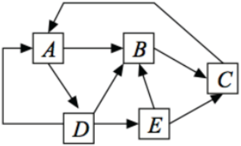


Let's simulate the Markov chain:

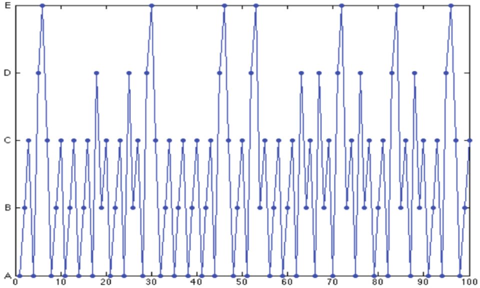


Five-State Markov Chain

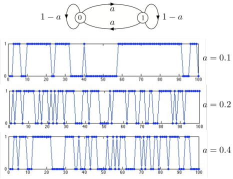
At each step, the MC follows one of the outgoing arrows of the current state, with equal probabilities.



Let's simulate the Markov chain:



Finite Markov Chain: Definition



- A finite set of states: $\mathscr{X} = \{1, 2, \ldots, K\}$
- A probability distribution π_0 on $\mathscr{X}$: $\pi_0(i) \geq 0, \sum_i \pi_0(i) = 1$
- Transition probabilities: $P(i,j)$ for $i,j \in \mathscr{X}$

  $P(i,j) \geq 0, \forall i,j; \sum_j P(i,j) = 1, \forall i$
- $\{X_n, n \geq 0\}$ is defined so that

  $$Pr[X_0 = i] = \pi_0(i), i \in \mathscr{X} \text{ (initial distribution)}$$

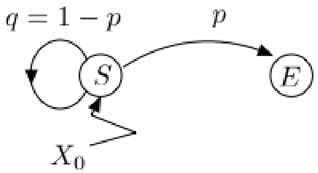
  $$Pr[X_{n+1} = j \mid X_0, \ldots, X_n = i] = P(i,j), i,j \in \mathscr{X}.$$

First Passage Time - Example 1

Let's flip a coin with $Pr[H] = p$ until we get H. How many flips, on average?

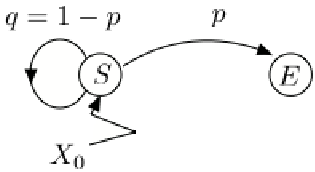
Let's define a Markov chain:

- $X_0 = S$ (start)
- $X_n = S$ for $n \geq 1$, if last flip was T and no H yet
- $X_n = E$ for $n \geq 1$, if we already got H (end)



First Passage Time - Example 1

Let's flip a coin with $Pr[H] = p$ until we get H. How many flips, on average?



Let $\beta(S)$ be the average time until E, starting from S.

Then,

$$\beta(S) = 1 + q\beta(S) + p0.$$

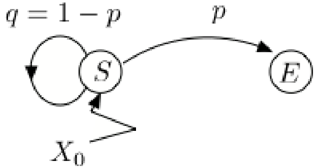
(See next slide.) Hence,

$$p\beta(S) = 1, \text{ so that } \beta(S) = 1/p.$$

Note: Time until E is $G(p)$.
The mean of $G(p)$ is $1/p$!!!

First Passage Time - Example 1

Let’s flip a coin with $Pr[H] = p$ until we get H. How many flips, on average?



Let $\beta(S)$ be the average time until E.
Then,

$$\beta(S) = 1 + q\beta(S) + p0.$$

Justification: N – number of steps until E, starting from S.
N' – number of steps until E, after the second visit to S.
And $Z = 1\{\text{first flip } = H\}$. Then,

$$N = 1 + (1 - Z) \times N' + Z \times 0.$$

Z and N' are independent. Also, $E[N'] = E[N] = \beta(S)$.
Hence, taking expectation,

$$\beta(S) = E[N] = 1 + (1 - p)E[N'] + p0 = 1 + q\beta(S) + p0.$$

First Passage Time - Example 2

Let's flip a coin with $Pr[H] = p$ until we get two consecutive Hs. How many flips, on average?

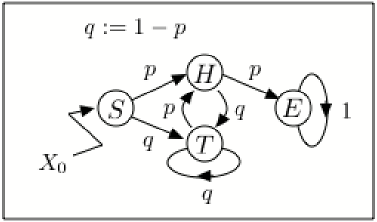
$$H\ T\ H\ T\ T\ T\ H\ T\ H\ T\ H\ T\ T\ H\ T\ H\ H$$

Let's define a Markov chain:

- $X_0 = S$ (start)
- $X_n = E$, if we already got two consecutive Hs (end)
- $X_n = T$, if last flip was T and we are not done
- $X_n = H$, if last flip was H and we are not done

First Passage Time - Example 2

Let's flip a coin with $Pr[H] = p$ until we get two consecutive Hs. How many flips, on average? Here is a picture:



S: Start
H: Last flip $= H$
T: Last flip $= T$
E: Done

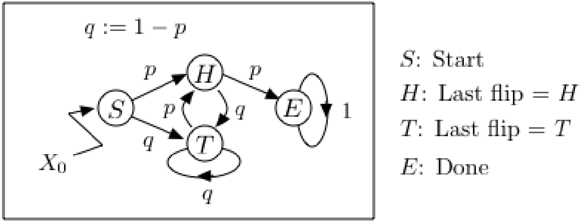
Let $\beta(i)$ be the average time from state i until the MC hits state E.

We claim that (these are called the first step equations)

$$\beta(S) = 1 + p\beta(H) + q\beta(T)$$
$$\beta(H) = 1 + p0 + q\beta(T)$$
$$\beta(T) = 1 + p\beta(H) + q\beta(T).$$

Solving, we find $\beta(S) = 2 + 3qp^{-1} + q^2p^{-2}$. (E.g., $\beta(S) = 6$ if $p = 1/2$.)

First Passage Time - Example 2



Let us justify the first step equation for $\beta(T)$. The others are similar.

$N(T)$ – number of steps, starting from T until the MC hits E.
$N(H)$ – be defined similarly.
$N'(T)$ – number of steps after the second visit to T until MC hits E.

$$N(T) = 1 + Z \times N(H) + (1 - Z) \times N'(T)$$

where $Z = 1\{\text{first flip in } T \text{ is } H\}$. Since Z and $N(H)$ are independent, and Z and $N'(T)$ are independent, taking expectations, we get

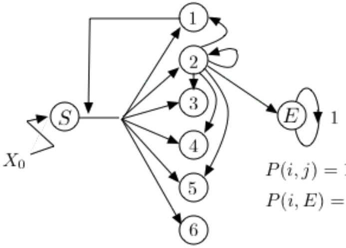
$$E[N(T)] = 1 + pE[N(H)] + qE[N'(T)],$$

i.e.,

$$\beta(T) = 1 + p\beta(H) + q\beta(T).$$

First Passage Time - Example 3

You roll a balanced six-sided die until the sum of the last two rolls is 8. How many times do you have to roll the die, on average?



S = Start; E = Done

i = Last roll is i, not done

$P(S, j) = 1/6, j = 1, \ldots, 6$

$P(1, j) = 1/6, j = 1, \ldots, 6$

$P(i, j) = 1/6, i = 2, \ldots, 6; 8 - i \neq j \in \{1, \ldots, 6\}$

$P(i, E) = 1/6, i = 2, \ldots, 6$

The arrows out of $3, \ldots, 6$ (not shown) are similar to those out of 2.

$$\beta(S) = 1 + \frac{1}{6}\sum_{j=1}^{6} \beta(j); \beta(1) = 1 + \frac{1}{6}\sum_{j=1}^{6} \beta(j); \beta(i) = 1 + \frac{1}{6}\sum_{j=1,\ldots,6; j \neq 8-i} \beta(j), i = 2, \ldots, 6.$$

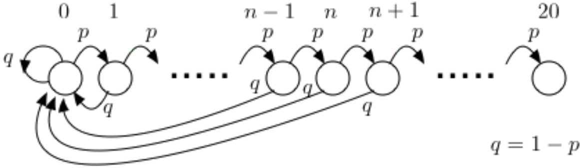
Symmetry: $\beta(2) = \cdots = \beta(6) =: \gamma$. Also, $\beta(1) = \beta(S)$. Thus,

$$\beta(S) = 1 + (5/6)\gamma + \beta(S)/6; \quad \gamma = 1 + (4/6)\gamma + (1/6)\beta(S).$$

$$\Rightarrow \cdots \beta(S) = 8.4.$$

First Passage Time - Example 4

You try to go up a ladder that has 20 rungs.
Each step, succeed or go up one rung with probability $p = 0.9$.
Otherwise, you fall back to the ground. Bummer.
Time steps to reach the top of the ladder, on average?



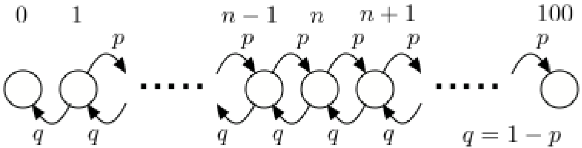
$$\beta(n) = 1 + p\beta(n+1) + q\beta(0), 0 \leq n < 19$$
$$\beta(19) = 1 + p0 + q\beta(0)$$

$$\Rightarrow \beta(0) = \frac{p^{-20} - 1}{1 - p} \approx 72.$$

See Lecture Note 24 for algebra.

First Passage Time - Example 5

Game of "heads or tails" using coin with 'heads' probability $p < 0.5$.
Start with \$10.
Each step, flip yields 'heads', earn \$1. Otherwise, lose \$1.
What is the probability that you reach \$100 before \$0?



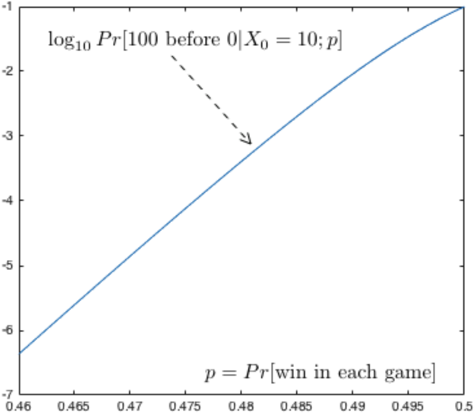
Let $\alpha(n)$ be the probability of reaching 100 before 0, starting from n, for $n = 0, 1, \ldots, 100$.

$$\alpha(0) = 0; \alpha(100) = 1.$$
$$\alpha(n) = p\alpha(n+1) + q\alpha(n-1), 0 < n < 100.$$

$$\Rightarrow \alpha(n) = \frac{1-\rho^n}{1-\rho^{100}} \text{ with } \rho = qp^{-1}. \text{ (See LN 24)}$$

First Passage Time - Example 5

Game of "heads or tails" using coin with 'heads' probability $p = .48$.
Start with \$10.
Each step, flip yields 'heads', earn \$1. Otherwise, lose \$1.
What is the probability that you reach \$100 before \$0?



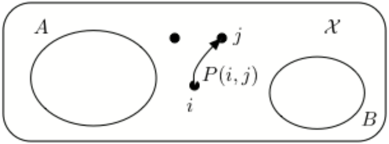
Less than 1 in a 1000. Morale of example: Money in Vegas stays in Vegas.

First Step Equations

Let X_n be a MC on $\mathscr{X}$ and $A, B \subset \mathscr{X}$ with $A \cap B = \emptyset$. Define

$$T_A = \min\{n \geq 0 \mid X_n \in A\} \text{ and } T_B = \min\{n \geq 0 \mid X_n \in B\}.$$

Let $\beta(i) = E[T_A \mid X_0 = i]$ and $\alpha(i) = Pr[T_A < T_B \mid X_0 = i], i \in \mathscr{X}$.



The FSE are

$$\beta(i) = 0, i \in A$$
$$\beta(i) = 1 + \sum_j P(i,j)\beta(j), i \notin A$$
$$\alpha(i) = 1, i \in A$$
$$\alpha(i) = 0, i \in B$$
$$\alpha(i) = \sum_j P(i,j)\alpha(j), i \notin A \cup B.$$

Accumulating Rewards

Let X_n be a Markov chain on $\mathscr{X}$ with P. Let $A \subset \mathscr{X}$

Let also $g : \mathscr{X} \to \Re$ be some function.

Define

$$\gamma(i) = E[\sum_{n=0}^{T_A} g(X_n) | X_0 = i], i \in \mathscr{X}.$$

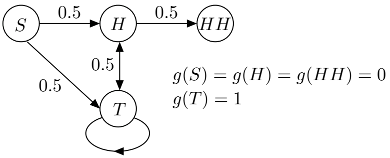
Then

$$\gamma(i) = \begin{cases} g(i), & \text{if } i \in A \\ g(i) + \sum_j P(i,j)\gamma(j), & \text{otherwise.} \end{cases}$$

Example

Flip a fair coin until you get two consecutive *H*s.

What is the expected number of *T*s that you see?



FSE:

$$\gamma(S) = 0 + 0.5\gamma(H) + 0.5\gamma(T)$$
$$\gamma(H) = 0 + 0.5\gamma(HH) + 0.5\gamma(T)$$
$$\gamma(T) = 1 + 0.5\gamma(H) + 0.5\gamma(T)$$
$$\gamma(HH) = 0.$$

Solving, we find $\gamma(S) = 2.5$.

Summary

Markov Chains

1. $Pr[X_{n+1} = j \mid X_0, \ldots, X_n = i] = P(i,j), i,j \in \mathscr{X}$
2. $T_A = \min\{n \geq 0 \mid X_n \in A\}$
3. $\alpha(i) = Pr[T_A < T_B | X_0 = i] \Rightarrow FSE$
4. $\beta(i) = E[T_A | X_0 = i] \Rightarrow FSE$
5. $\gamma(i) = E[\sum_{n=0}^{T_A} g(X_n) | X_0 = i] \Rightarrow FSE$.