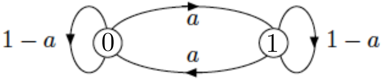
CS70: Markov Chains.

$\boxed{\text{Markov Chains 1}}$

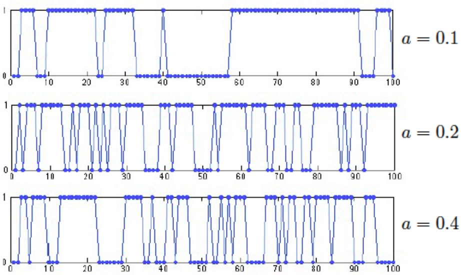
1. Examples
2. Definition
3. First Passage Time

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.

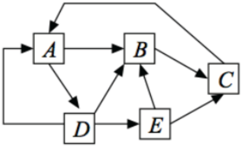


Let's simulate the Markov chain:

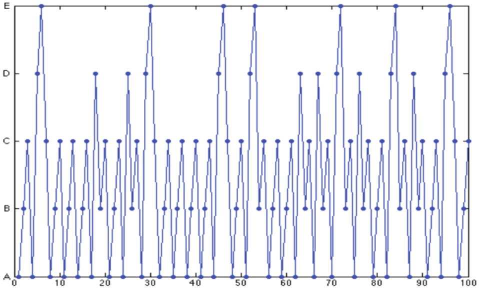


Five-State Markov Chain

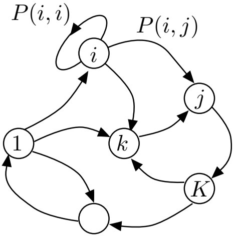
At each step, the MC follows one of the outgoing arrows of the current state, with equal probabilities.



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

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

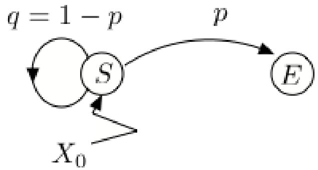
  $Pr[X_{n+1} = j \mid X_0, \ldots, X_n = i] = P(i,j), i, j \in \mathscr{X}$.

First Passage Time - Example 1

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

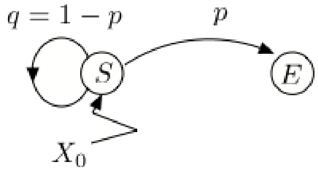
Let's define a Markov chain:

- $X_0 = S$ (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,

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

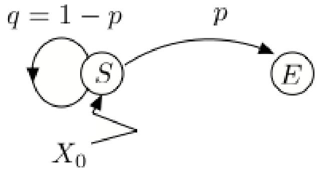
(See next slide.) Hence,

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

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?

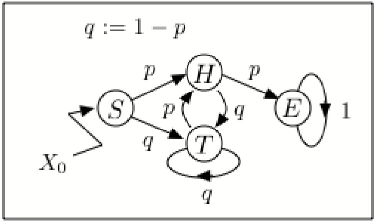
$$H\ T\ H\ T\ T\ T\ H\ T\ H\ T\ H\ T\ T\ H\ T\ H\ H$$

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

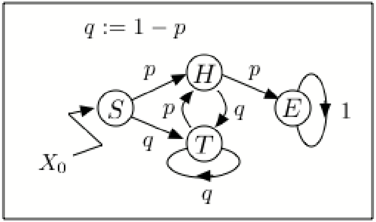
Let $\beta(i)$ be the average time from state i until the MC hits state E.

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

$$\begin{aligned}
\beta(S) &= 1 + p\beta(H) + q\beta(T) \\
\beta(H) &= 1 + p0 + q\beta(T) \\
\beta(T) &= 1 + p\beta(H) + q\beta(T).
\end{aligned}$$

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



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

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

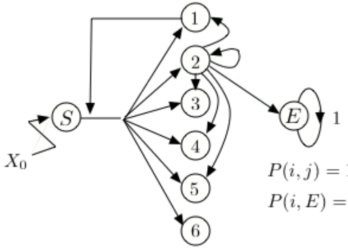
$$E[N(T)] = 1 + pE[N(H)] + qE[N'(T)],$$

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.

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

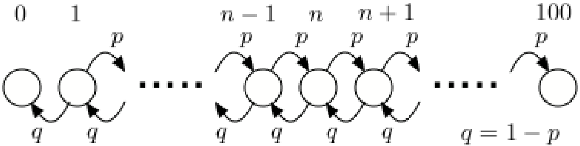
Symmetry: $\beta(2) = \cdots = \beta(6) =: \gamma$. Also, $\beta(1) = \beta(S)$. Thus,

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

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

First Passage Time - A before B

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?



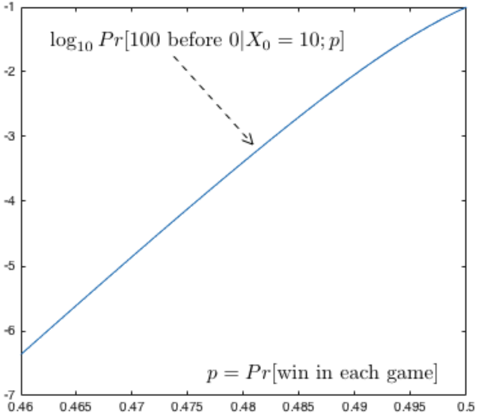
Let $\alpha(n)$ be the probability of reaching 100 before 0, starting from n, for $n = 0, 1, \ldots, 100$.

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

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

First Passage Time - A before B

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?



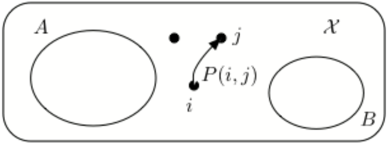
Less than 1 in a 1000. Morale of example: Money in Vegas stays in Vegas.

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

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

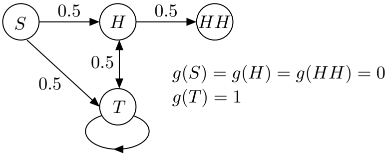
Then

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

Example

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

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



FSE:

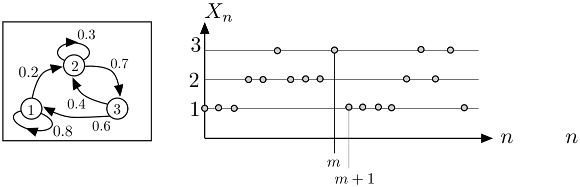
$$
\begin{aligned}
\gamma(S) &= 0 + 0.5\gamma(H) + 0.5\gamma(T) \\
\gamma(H) &= 0 + 0.5\gamma(HH) + 0.5\gamma(T) \\
\gamma(T) &= 1 + 0.5\gamma(H) + 0.5\gamma(T) \\
\gamma(HH) &= 0.
\end{aligned}
$$

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

Recap

- Markov Chain:
  - Finite set $\mathscr{X}$; π_0; $P = \{P(i,j), i,j \in \mathscr{X}\}$;
  - $Pr[X_0 = i] = \pi_0(i), i \in \mathscr{X}$
  - $Pr[X_{n+1} = j \mid X_0, \ldots, X_n = i] = P(i,j), i,j \in \mathscr{X}, n \geq 0$.
  - Note:
    $Pr[X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n] = \pi_0(i_0)P(i_0,i_1)\cdots P(i_{n-1},i_n)$.
- First Passage Time:
  - $A \cap B = \emptyset; \beta(i) = E[T_A|X_0 = i]; \alpha(i) = P[T_A < T_B|X_0 = i]$
  - $\beta(i) = 1 + \sum_j P(i,j)\beta(j)$;
  - $\alpha(i) = \sum_j P(i,j)\alpha(j)$. $\alpha(A) = 1, \alpha(B) = 0$.

Distribution of X_n



Recall π_n is a distribution over states for X_n.

Stationary distribution: $\pi = \pi P$.

Distribution over states is the same before/after transition.

probability entering i: $\sum_{i,j} P(j,i)\pi(j)$.

probability leaving i: π_i.

are Equal!

Distribution same after one step.

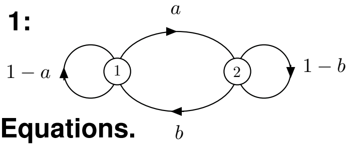
Questions? Does one exist? Is it unique?

If it exists and is unique. Then what?

Sometimes the distribution as $n \to \infty$

Stationary: Example

Example 1:



$$P = \begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix}$$

Balance Equations.

$$\begin{aligned}
\pi P = \pi \quad &\Leftrightarrow \quad [\pi(1), \pi(2)] \begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix} = [\pi(1), \pi(2)] \\
&\Leftrightarrow \quad \pi(1)(1-a) + \pi(2)b = \pi(1) \text{ and } \pi(1)a + \pi(2)(1-b) = \pi(2) \\
&\Leftrightarrow \quad \pi(1)a = \pi(2)b.
\end{aligned}$$

These equations are redundant! We have to add an equation: $\pi(1) + \pi(2) = 1$. Then we find

$$\pi = [\frac{b}{a+b}, \frac{a}{a+b}].$$

Stationary distributions: Example 2

$$P = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\pi P = \pi \Leftrightarrow [\pi(1), \pi(2)] \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = [\pi(1), \pi(2)] \Leftrightarrow \pi(1) = \pi(1) \text{ and } \pi(2) = \pi(2).$$

Every distribution is invariant for this Markov chain. This is obvious, since $X_n = X_0$ for all n. Hence, $Pr[X_n = i] = Pr[X_0 = i], \forall (i, n)$.

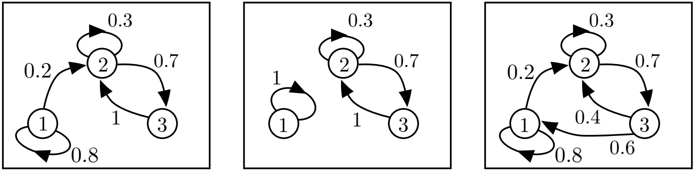
Discussion.
We have seen a chain with one stationary,
and a chain with many.

When is here just one?

Irreducibility.

Definition A Markov chain is irreducible if it can go from every state i to every state j (possibly in multiple steps).

Examples:



[A] [B] [C]

[A] is not irreducible. It cannot go from (2) to (1).

[B] is not irreducible. It cannot go from (2) to (1).

[C] is irreducible. It can go from every i to every j.

If you consider the graph with arrows when $P(i,j) > 0$, irreducible means that there is a single connected component.

Existence and uniqueness of Invariant Distribution

Theorem A finite irreducible Markov chain has one and only one invariant distribution.

That is, there is a unique positive vector $\pi = [\pi(1), \ldots, \pi(K)]$ such that $\pi P = \pi$ and $\sum_k \pi(k) = 1$.

Ok. Now.
Only one stationary distribution if irreducible (or connected.)

Long Term Fraction of Time in States

Theorem Let X_n be an irreducible Markov chain with invariant distribution π.

Then, for all i,

$$\frac{1}{n}\sum_{m=0}^{n-1} 1\{X_m = i\} \to \pi(i), \text{ as } n \to \infty.$$

The left-hand side is the fraction of time that $X_m = i$ during steps $0, 1, \ldots, n-1$. Thus, this fraction of time approaches $\pi(i)$.

Proof: Lecture note 21 gives a plausibility argument.

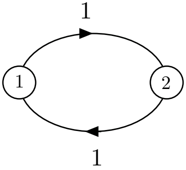
$\square$

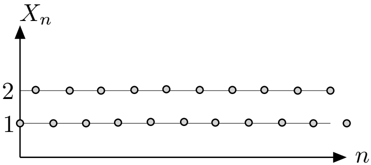
Long Term Fraction of Time in States

Theorem Let X_n be an irreducible Markov chain with invariant distribution π. Then, for all i, $\frac{1}{n}\sum_{m=0}^{n-1} 1\{X_m = i\} \to \pi(i)$, as $n \to \infty$.

Example 1:

$$P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \qquad \pi P = \pi \Rightarrow \pi = [1/2, 1/2]$$



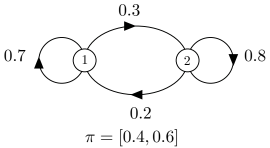


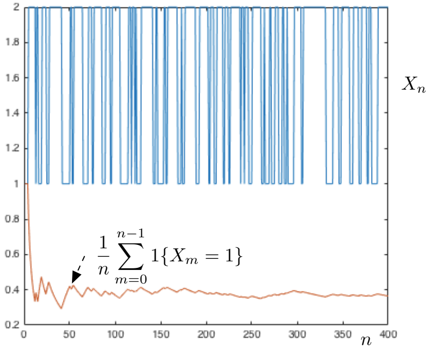
The fraction of time in state 1 converges to $1/2$, which is $\pi(1)$.

Long Term Fraction of Time in States

Theorem Let X_n be an irreducible Markov chain with invariant distribution π. Then, for all i, $\frac{1}{n}\sum_{m=0}^{n-1} 1\{X_m = i\} \to \pi(i)$, as $n \to \infty$.

Example 2:

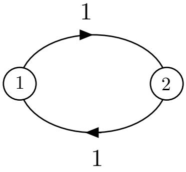




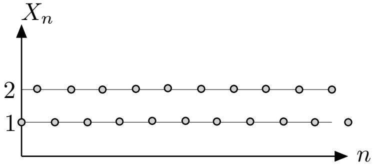
Convergence to Invariant Distribution

Question: Assume that the MC is irreducible. Does π_n approach the unique invariant distribution π?

Answer: Not necessarily. Here is an example:



$$P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \qquad \pi P = \pi \Rightarrow \pi = [1/2, 1/2]$$



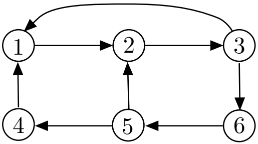
Assume $X_0 = 1$. Then $X_1 = 2, X_2 = 1, X_3 = 2, \ldots$.

Thus, if $\pi_0 = [1,0]$, $\pi_1 = [0,1], \pi_2 = [1,0], \pi_3 = [0,1]$, etc.

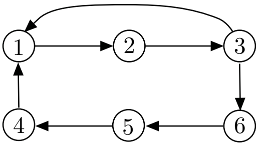
Hence, π_n does not converge to $\pi = [1/2, 1/2]$.
Notice, all cycles or closed walks have even length.

Periodicity

Definition: Periodicity is gcd of the lengths of all closed walks in irreducible chain.   Previous example: 2.
Definition If periodicity is 1, Markov chain is said to be aperiodic. Otherwise, it is periodic.
Example



$[A]$



$[B]$

[A]: Closed walks of length 3 and length 4 $\implies$ periodicity = 1.

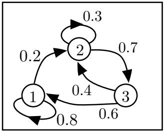
[B]: All closed walks multiple of 3 $\implies$ periodicity =2.

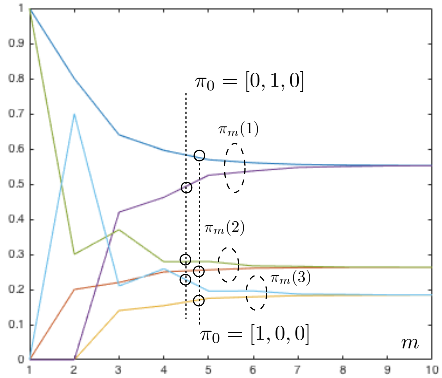
Convergence of π_n

Theorem Let X_n be an irreducible and aperiodic Markov chain with invariant distribution π. Then, for all $i \in \mathscr{X}$,

$$\pi_n(i) \to \pi(i), \text{ as } n \to \infty.$$

Example



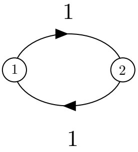


Convergence of π_n

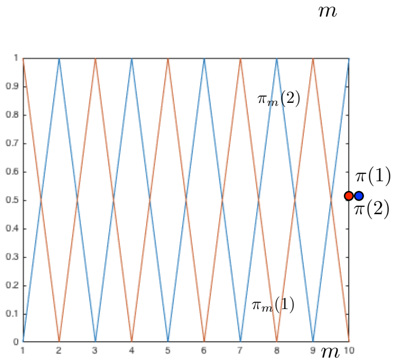
Theorem Let X_n be an irreducible and aperiodic Markov chain with invariant distribution π. Then, for all $i \in \mathscr{X}$,

$$\pi_n(i) \to \pi(i), \text{ as } n \to \infty.$$

Example



$\pi = [0.5, 0.5]$



Summary

Markov Chains

- Markov Chain: $Pr[X_{n+1} = j | X_0, \ldots, X_n = i] = P(i,j)$
- FSE: $\beta(i) = 1 + \sum_j P(i,j)\beta(j); \alpha(i) = \sum_j P(i,j)\alpha(j)$.
- $\pi_n = \pi_0 P^n$
- π is invariant iff $\pi P = \pi$
- Irreducible $\Rightarrow$ one and only one invariant distribution π
- Irreducible $\Rightarrow$ fraction of time in state i approaches $\pi(i)$
- Irreducible + Aperiodic $\Rightarrow \pi_n \to \pi$.
- Calculating π: One finds $\pi = [0, 0 \ldots, 1]Q^{-1}$ where $Q = \cdots$.

Confirmation Bias: An experiment

There are two bags.

One with 60% red balls and 40% blue balls;
the other with the opposite fractions.

One selects one of the two bags.

As one draws balls one at time, one asks people to declare whether they think one draws from the first or second bag.

Surprisingly, people tend to be reinforced in their original belief, even when the evidence accumulates against it.

Report Data not Opinion!

A bag with 60% red, 40% blue or vice versa.

Each person pulls ball, reports opinion on which bag:
Says “majority blue” or “majority red.”

Does not say what color their ball is.

What happens if first two get blue balls?

Third hears two blue, so says blue, whatever she sees.
Plus Induction.

Everyone says blue...forever ...and ever.

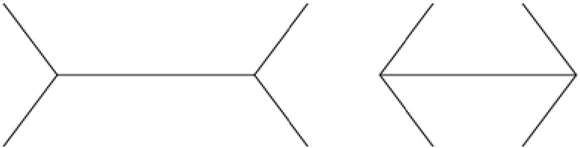
Problem: Each person reported honest opinion rather than data!

Being Rational: ‘Thinking, Fast and Slow’

In this book, Daniel Kahneman discusses examples of our irrationality.

Here are a few examples:

- A judge rolls a die in the morning.
  In the afternoon, he has to sentence a criminal.
  Statistically, morning roll high $\implies$ sentence is high.
- People tend to be more convinced by articles printed in Times Roman instead of Computer Modern Sans Serif.
- Perception illusions: Which horizontal line is longer?



It is difficult to think clearly!

What to Remember?

Professor, what should I remember about probability from this course?

I mean, after the final.

Here is what the prof. remembers:

- Given the uncertainty around us, understand some probability.
- One key idea - what we learn from observations: the role of the prior; Bayes' rule; Estimation; confidence intervals... quantifying our degree of certainty.
- This clear thinking invites us to question vague statements, and to convert them into precise ideas.

Power of course.

Fight Ignorance.

Tools for Rationality.

Think it through
verify what you know and you concluded
carefully and completely and simply.

What's Next?

Professor, I loved this course so much!
I want to learn more about discrete math and probability!

Funny you should ask! How about

- CS170: Efficient Algorithms and Intractable Problems a.k.a. Introduction to CS Theory: Graphs, Dynamic Programming, Complexity.
- EE126: Probability in EECS: An Application-Driven Course: PageRank, Digital Links, Tracking, Speech Recognition, Planning, etc. Hands on labs with python experiments (GPS, Shazam, ...).
- CS188: Artificial Intelligence: Hidden Markov Chains, Bayes Networks, Neural Networks.
- CS189: Introduction to Machine Learning: Regression, Neural Networks, Learning, etc. Programming experiments with real-world applications.
- EE121: Digital Communication: Coding for communication and storage.
- EE223: Stochastic Control.
- EE229A: Information Theory; EE229B: Coding Theory.