Summary: so far.

Markov Chains

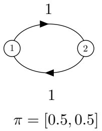
- Markov Chain: $Pr[X_{n+1} = j | X_0, \ldots, X_n = i] = P(i, j)$
- $\pi_n = \pi_0 P^n$
- π is invariant iff $\pi P = \pi$
- Irreducible $\Rightarrow$ one and only one invariant distribution π
- Irreducible $\Rightarrow$ fraction of time in state i approaches $\pi(i)$
- Irreducible + Aperiodic $\Rightarrow \pi_n \to \pi$.
- Calculation: $\pi = [0, 0 \ldots, 1] Q^{-1}$ where Q is $P - I$ plus all 1's..
  $\pi P = \pi \implies \pi(P - I) = 0.$
  And $\sum_i \pi(i) = 1$ is all ones.

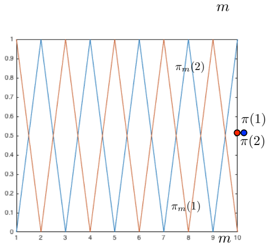
Convergence of π_n

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

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

Non Example: periodic chain



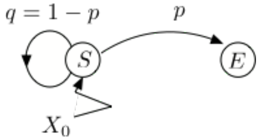


Note: $d > 1$, can d-color Markov Chain
vertices in color $i \pmod d$ point to $i+1 \pmod d$.

$\implies$ decompose invariant distribution into d distributions on d colors.

First Passage Time - Example 1. Poll

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.

What is correct?

(A) $\beta(S)$ is at least 1.
(B) From S, in one step, go to S with prob. $q = 1 - p$
(C) From S, in one step, go to E with prob. p.
(D) If you go back to S, you are back at S.
(E) $\beta(S) = 1 + q \times \beta(S) + p \times 0$.
(F) Wait: this is $G(p)$.

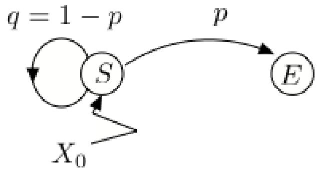
All are correct. (D) is the "Markov property." Only know where you are.

Hitting Time - Example 1

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

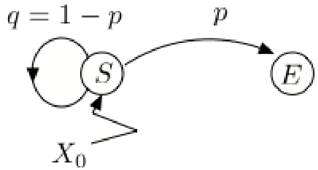
Let's define a Markov chain:

- $X_0 = S$ (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)



Hitting Time - Example 1

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



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

Then,

$$\beta(S) = 1 + q \times \beta(S) + p \times 0.$$

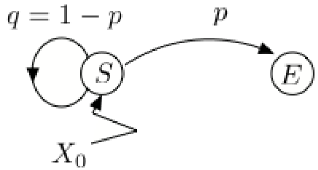
(See next slide.) Hence,

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

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 in expectation?



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

$$\beta(S) = 1 + q \times \beta(S) + p \times 0.$$

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." $E[N'] = E[N] = \beta(S)$.
Hence, taking expectation,

$$\beta(S) = E[N] = 1 + (1 - p) \times E[N'] + p \times 0 = 1 + q \times \beta(S) + p \times 0.$$

Hitting Time - Example 2

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

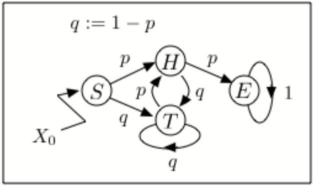
$$H\ T\ H\ T\ T\ T\ H\ T\ H\ T\ H\ T\ T\ H\ T\ H\ H$$

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

Hitting 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

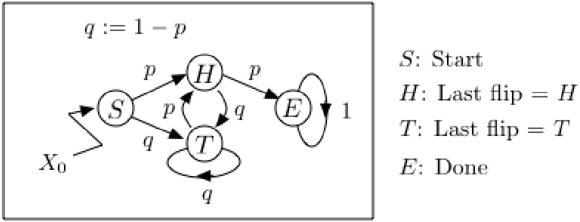
Which one is correct?
(A) $\beta(S) = 1 + p \times \beta(H) + q \times \beta(T)$
(B) $\beta(S) = p \times \beta(H) + q \times \beta(T)$
(C) $\beta(S) = \beta(S) + q \times \beta(T) + p \times \beta(H)$.

(A) Expected time from S to E.

$$\beta(S) = Pr[H]E[\beta(S)|H] + Pr[T]E[\beta(S)|T]$$
$$\beta(S) = p(1 + \beta(H)) + q(1 + \beta(T))$$
$$\beta(S) = 1 + p\beta(H) + q\beta(T)$$

Hitting 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:



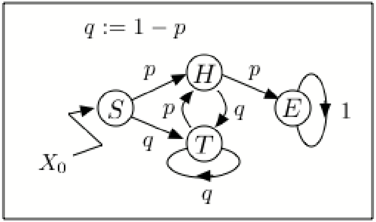
Let $\beta(i)$ be the average time from state i until the MC hits state E.

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

$$\beta(S) = 1 + p \times \beta(H) + q \times \beta(T)$$
$$\beta(H) = 1 + p \times 0 + q \times \beta(T)$$
$$\beta(T) = 1 + p \times \beta(H) + q \times \beta(T).$$

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

Hitting 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

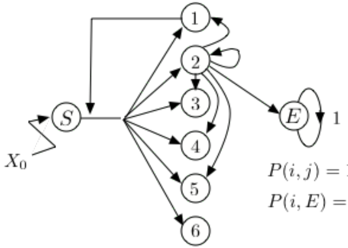
$$E[N(T)] = 1 + pE[N(H)] + qE[N'(T)],$$

i.e.,

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

Hitting 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.

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

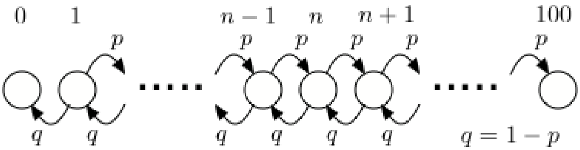
Symmetry: $\beta(2) = \cdots = \beta(6) =: \gamma$. Also, $\beta(1) = \beta(S)$. Thus,

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

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

Here before There - 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?



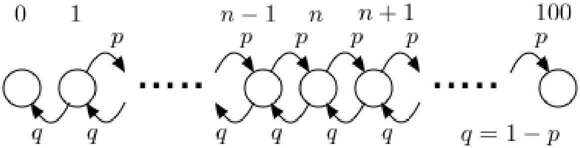
Let $\alpha(n) = Pr[\text{reach 100 before 0}|\text{at } n]$ for $n = 0, 1, \ldots, 100$.

Which equations are correct?
(A) $\alpha(0) = 0$
(B) $\alpha(0) = 1$.
(C) $\alpha(100) = 1$.
(D) $\alpha(n) = 1 + p\alpha(n+1) + q\alpha(n-1), 0 < n < 100$.
(E) $\alpha(n) = p\alpha(n+1) + q\alpha(n-1), 0 < n < 100$.

(B) is incorrect, 0 is bad.
(D) is incorrect. Confuses expected hitting time with *A* before *B*.

Here before There - 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?



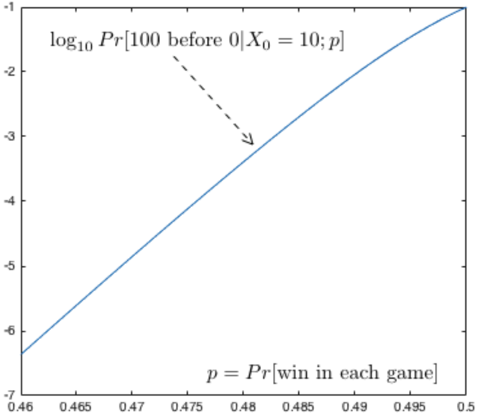
Let $\alpha(n) = Pr[\text{reach 100 before 0}|\text{at } n]$ for $n = 0, 1, \ldots, 100$.

$$\alpha(0) = 0; \alpha(100) = 1.$$
$$\alpha(n) = Pr[n+1|n]Pr[100 < 0|n+1] + Pr[n-1|n]Pr[100 < 0|n-1]$$
$$\alpha(n) = p\alpha(n+1) + q\alpha(n-1), \qquad 0 < n < 100.$$

$$\Rightarrow \alpha(n) = \frac{1-\rho^n}{1-\rho^{100}} \text{ with } \rho = qp^{-1}. \text{ (See LN 22)}$$

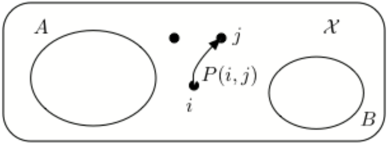
Here before There - A before B

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. Moral 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

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

For $\beta(i) = E[T_A \mid X_0 = i]$, first step equations are:

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

For $\alpha(i) = Pr[T_A < T_B \mid X_0 = i], i \in \mathscr{X}$,, first step equations are:

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

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 reward function.

Define

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

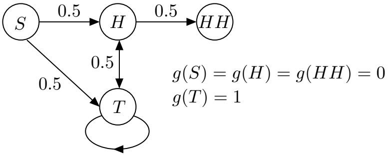
Then

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

Example

Flip a fair coin until you get two consecutive Hs.

What is the expected number of Ts 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 $\gamma(S) = 2.5$.

Recap

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

Summary

Markov Chains

- Markov Chain: $Pr[X_{n+1} = j | X_0, \ldots, X_n = i] = P(i,j)$
- $\pi_n = \pi_0 P^n$
- 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 $\Rightarrow \pi_n \rightarrow \pi$.
- Calculation: $\pi = [0, 0 \ldots, 1] Q^{-1}$ where Q is $P - I$ plus all 1's..
- FSE for Hitting Time: $\beta(i) = 1 + \sum_j P(i,j) \beta(j)$
- FSE for A befure B: $\alpha(i) = \sum_j P(i,j) \alpha(j)$.

CS70: Lecture 27

1. Continuous Probability
2. Normal Distribution
3. Central Limit Theorem
4. Confidence Intervals
5. Wrapup.

Continuous Probability

1. pdf: $Pr[X \in (x, x+\delta]] \approx f_X(x)\delta$.
2. CDF: $Pr[X \leq x] = F_X(x) = \int_{-\infty}^{x} f_X(y)dy$.
3. $U[a,b]$, $Expo(\lambda)$, target.
4. Expectation: $E[X] = \int_{-\infty}^{\infty} x f_X(x) dx$.
5. Variance: $var[X] = E[(X - E[X])^2] = E[X^2] - E[X]^2$.
6. Variance of Sum of Independent RVs: If X_n are pairwise independent, $var[X_1 + \cdots + X_n] = var[X_1] + \cdots + var[X_n]$
7. Joint Density function:
   $Pr[X \in [x, x+\delta], Y \in [y, y+\delta]] \approx f_{X,Y}(x,y)\delta^2$.
8. Conditional Density:
   $Pr[X \in [x, x+\delta] | Y = y] \approx f_{X|Y}(x,y)dx = \frac{f_{X,Y}(x,y)}{f_Y(y)}dx$.

Summary

1. Bayes' Rule: Replace $\{X = x\}$ by $\{X \in (x, x + \varepsilon)\}$.
2. Gaussian: $\mathcal{N}(\mu, \sigma^2) : f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{1}{2}(\frac{(x-\mu)^2}{\sigma})}$ "bell curve"
3. CLT: X_n i.i.d. $\implies \frac{A_n - \mu}{\sigma/\sqrt{n}} \to \mathcal{N}(0,1)$
4. CI: $[A_n - 2\frac{\sigma}{\sqrt{n}}, A_n + 2\frac{\sigma}{\sqrt{n}}] =$ 95%-CI for μ.

CS70: Wrapping Up.

Random Thoughts

Confusing Statistics: Simpson's Paradox

College	F. Appl.	F. Adm.	% F. Adm.	M. Appl.	M. Adm.	% M. Adm.
A	980	490	50%	200	80	40%
B	20	20	100%	800	720	90%
Total	1000	510	51%	1000	800	80%

Applications/admissions by two [sic] genders two colleges of a university.

Male admission rate 80% but female 51%!

But admission rate is larger for female students in both colleges....

More female applicants to college that admits fewer students.

Side note: average high school GPA is higher for female students.

Other example: On-time arrival for airlines.
If "hub" in chicago, that's a problem overall.

GPA: stronger students take harder classes. Maybe.

More on Confusing Statistics

Statistics are often confusing:

- The average household annual income in the US is $\$72k$.
  Yet, the median is $\$52k$.
- The false alarm rate for prostate cancer is only 1%.
  Still only 1 person in $8,000$ has that cancer. Prior.
  $\Longrightarrow$ there are 80 false alarms for each actual case.
- The Texas sharpshooter fallacy:
  Shoot a barn. Paint target cluster. I am sharpshooter!
  People living close to power lines.
  You find clusters of cancers!
  Also find such clusters when looking at people eating kale!
- False causation. Vaccines cause autism.
  Both vaccination and autism rates increased....
- Beware of statistics reported in the media!

Confirmation Bias

Confirmation bias: tendency to search for, interpret, and recall information in a way that confirms one's beliefs or hypotheses, while giving less consideration to alternative possibilities.

Confirmation biases contribute to overconfidence in personal beliefs and can maintain or strengthen beliefs in the face of contrary evidence.

Three aspects:

- Biased search for information.
  E.g., facebook friends effect, ignoring inconvenient articles.
- Biased interpretation.
  E.g., valuing confirming versus contrary evidence.
- Biased memory.
  E.g., remember facts that confirm beliefs and forget others.

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 say blue, so says blue, whatever she sees.
Plus Induction.

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

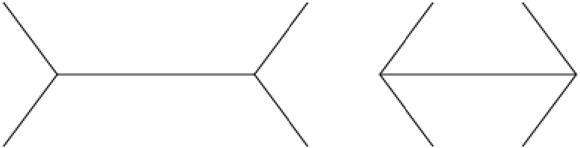
Problem: Each person reported honest opinion rather than data!

Being Rational: ‘Thinking, Fast and Slow’

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

Here are a few examples:

- Experiment: A judge rolls a die in the morning.
  In the afternoon, he assigns a sentence to a criminal (based on folder).
  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: Which horizontal line is longer?



It is difficult to think clearly!

Really?

Judges at Lousiana give longer sentences when LSU gives upset losses.

Judges give larger sentences when hungry.

Let's check google and google scholar.

Uh oh.

Certainty is the enemy?
Unless you work hard! You have the internet.
You have your intellect.
...and (most important) your integrity.

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.

Parting Thoughts

You have learned a lot in this course!

Proofs, Graphs, Mod(p), RSA, Reed-Solomon, Decidability, Probability, ... ,

how to handle stress, how to sleep less, how to keep smiling, ...

Difficult course? Perhaps. Mind expanding! I believe!!

Useful? You bet!

Arguments, reasoning.

What you know: slope, limit.
  Plus: definition.
yields calculus.
  Minimization, optimization, …..

Knowing how to program plus some syntax (google) gives the ability to program.

Knowing how to reason plus some definition gives calculus.

Discrete Math: basics are counting, how many, when are two sets the same size?

Probability: division.

...plus reasoning.

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.

Final Thoughts

More precisely: Some thoughts about the final ....

How to study for the final?

- Lecture Slides; Notes; Discussion Problems; HW
- Approximate Coverage: Probability 2/3, Discrete Math: 1/3.
- Every question topic covered in at least two places. Most will be covered in all places.

Finally....

Thanks for taking the course!

Thanks to the CS70 Staff:

- The Terrific Tutors
- The Rigorous Readers
- The Thrilling TAs
- The Amazing Assistants

Good studying!!!