

# CSE103: Introduction to Probability and Statistics

Prof. Yoav Freund



# Flipping two dice

- 1 dice-What is the probability that it will land on 6 ? Or on 5 ?
- **G,R** dice. What is the probability of green=6 and red=5
- **R,R** dice. What is the probability of red=5 and red=6?
- **R,R** dice. What is the probability of red=5 and red=5 ?
- The difference is because dice of the same color are indistinguishable

# Indistinguishability / Exchangeability

- Two objects are indistinguishable if exchanging them makes no difference.
- What makes two objects indistinguishable?
- Mathematical objects (points, lines) are indistinguishable
  - Physical objects are distinguishable (we can mark them)
- Poker cards are indistinguishable, unless marked (illegal)
- iPhones (of the same model) are indistinguishable
  - iPhone covers make the iPhones distinguishable.
- Dollars are indistinguishable - that is what makes the economy work (compare that to bartering).
- Are fruits indistinguishable? (same DNA)
- Are animals of a species distinguishable?
- Are people distinguishable?

# Probabilities regarding people

- Which of the following is more correct?
  - Each of us is unique, we have our own free will.
  - We belong to groups, our opinions are the opinions of the group.
- When the number of people is large, a very effective way to reason is to think of people as interchangeable:
  - How many children in this district have special needs?
  - How many voters in this county are likely to vote republican?
  - What is the lowest price for a life insurance policy that will not bankrupt the insurance company?
- Are we all the same or are we all different?

# We are individuals - The life of Brian



**HAROLD STEVENS**  
DIAMOND AND JEWELRY STUDIO



SHOP >  
*vibrant*



**U-T**  
San Diego  
**99¢ A WEEK**  
RESOLVE TO BE INFORMED

[LOCAL TOPICS](#) | [SCIENCE & ENVIRONMENT](#) | [SCIENCE](#)

# Can Amazon read consumers' minds?

By **Gary Robbins** 4:31 P.M. JAN. 27, 2014

[PRINT](#) [COMMENTS](#) 0



RECOMMENDED

TRENDING

SDSU's Toledo: 'I'm Monday morning QB t...

Smash and grab robber at Shell station ge...

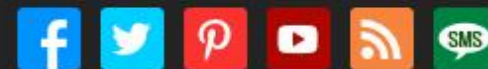
Brandon Flowers added to injury report

Pace of home price appreciation in San Di...

End near for NFL blackout rule?

Recommended by

CONNECT WITH THE U-T



Sign Up for Email Alerts

# The power of predictive analytics

## AMAZON PATENTED AN ANTICIPATORY SHIPPING SYSTEM THAT PREDICTS ORDERS

By Mike Flacy — January 18, 2014

Noticed by the [Wall Street Journal](#), Amazon recently [patented a new system](#) that will help the retailer create predictive models to accurately forecast where an item will need to ship. Calling the system “anticipatory shipping,” Amazon will collectively compile data such as product searches, page visits, wish list items, order history, overall time on page, items left in the shopping cart and return history to pre-ship items to closer warehouse locations or even directly to the eventual recipient. Amazon even plans to measure the length of time that a user’s mouse cursor hovers over an item in order to predict an upcoming purchase.





# What is going on?


- Actual patent: shipping a package without a final destination.
- How does this work:
  - Amazon predicts that >100 people in san diego will order the new iPad on Dec 24 with same-day delivery.
  - Mail 100 iPad's to san diego on dec 20, using UPS ground, with no final address (special agreement with UPS)
  - Fulfill last minute orders by assigning the address while package is in transit.
- This method works well when there are many identical orders from one location/city.
- Supply chain management is a long standing practice, amazon is bringing it to the next level.
- Statistics used to predict how many orders we will have of item X from location Y.





# Ads on my Facebook Page


 Yoav Freund





 Yoav Home 20+




















Yoav Freund








Timeline


About


Photos


Friends 304

More ▾


 What movies have you watched? ✕


 Status

 Photo / Video


 Life Event

What's on your mind?

Sponsored 



**NORDSTROM Men's Jeans**  
shop.nordstrom.com  
Shop tons of different denim brands and styles at NORDSTROM. FREE shipping. FREE returns.



**Kids' Shoes at DSW**  
dsw.com  
Little ones need new kicks? Get to dsw.com and select stores to shop top brands!

# click-through

- A “click-through” occurs when
  - Surfer reaches a web page.
  - Surfer clicks on an ad - sent to a new page.
  - Advertiser pays web-host company 1 cent.
- Web host wants to place ads that are more likely to get a click through.
- Many factors to consider, but we'll keep things simple:
  - There are two alternative ads: a and b
  - The (unknown) probabilities of click through are  $P_a$  and  $P_b$
  - If  $P_a > P_b \rightarrow$  always display a
  - If  $P_a < P_b \rightarrow$  always display b
- But we don't know!
  - To estimate  $P_a$  and  $P_b$  We alternate a,b,a,b,a,b,a,....

# The law of large numbers

- If we repeat displaying ad a forever, the fraction of times that a click-through occurs converges to the true probability  $P_a$
- Same for ad b
- By repeating a,b,a,b,a,b,a,... forever we will find out which of the two is better.
- Forever is too long - we will never get to use this knowledge.
- In this class we will study the rate at which we approach the limit.
- For now, a few figures.

# The sequence of running averages

1=click through, 0=no click through

Focus on a sequence with one of the two ads:

$$X_1, X_2, X_3, \dots, \frac{1}{4} = 0, 0, 1, 0, 0, 0, 0, 1, \frac{1}{4}$$

The running average is:

$$X_1, \frac{X_1 + X_2}{2}, \frac{X_1 + X_2 + X_3}{3}, \dots, \frac{1}{t} \sum_{i=1}^t X_i, \frac{1}{4}$$

For the sequence above it the running average is:

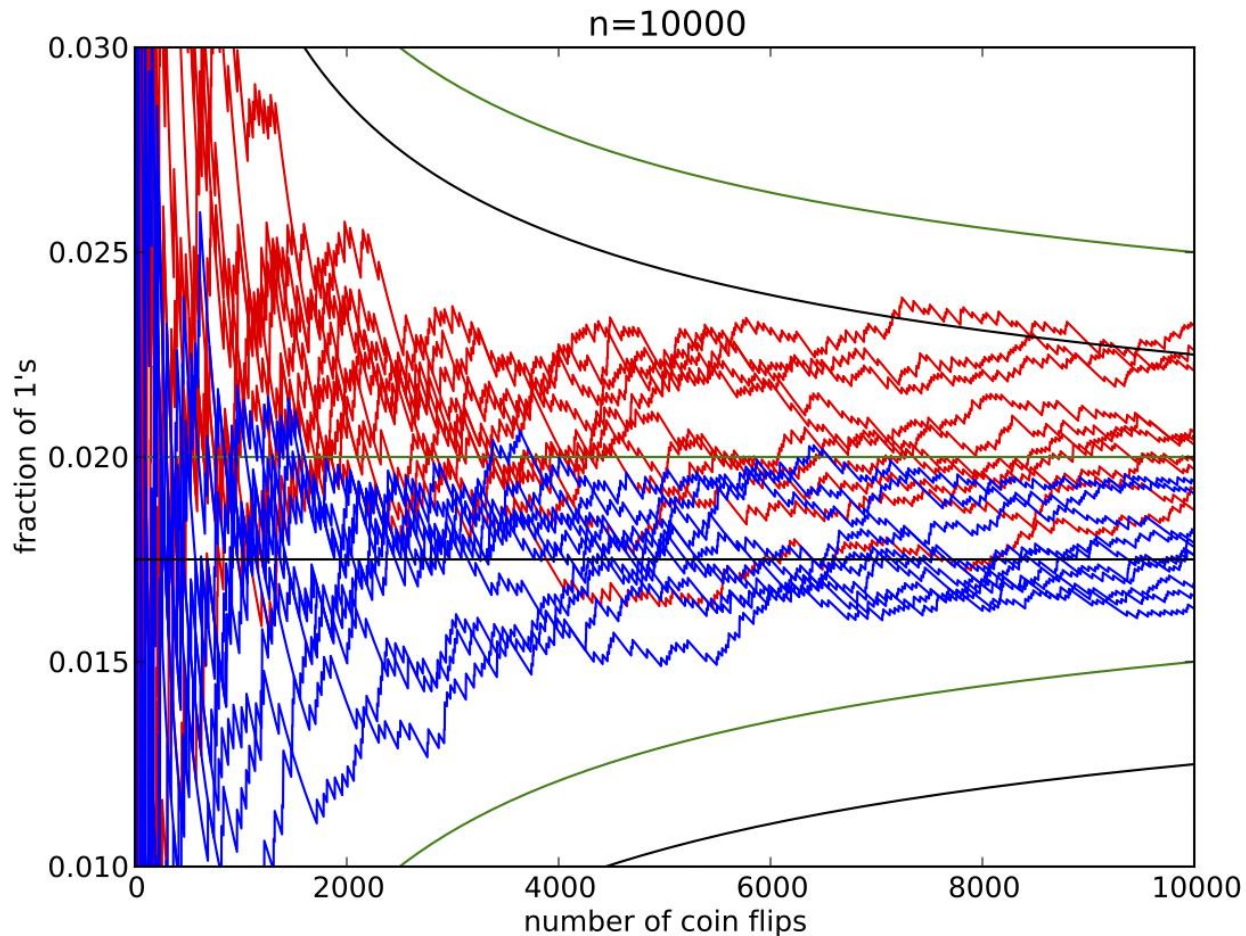
$$0, 0, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \frac{1}{6}, \frac{1}{7}, \frac{2}{8}, \frac{1}{4}$$

From the law of large numbers we know that the sequence converges to the true probability. The question is how fast?

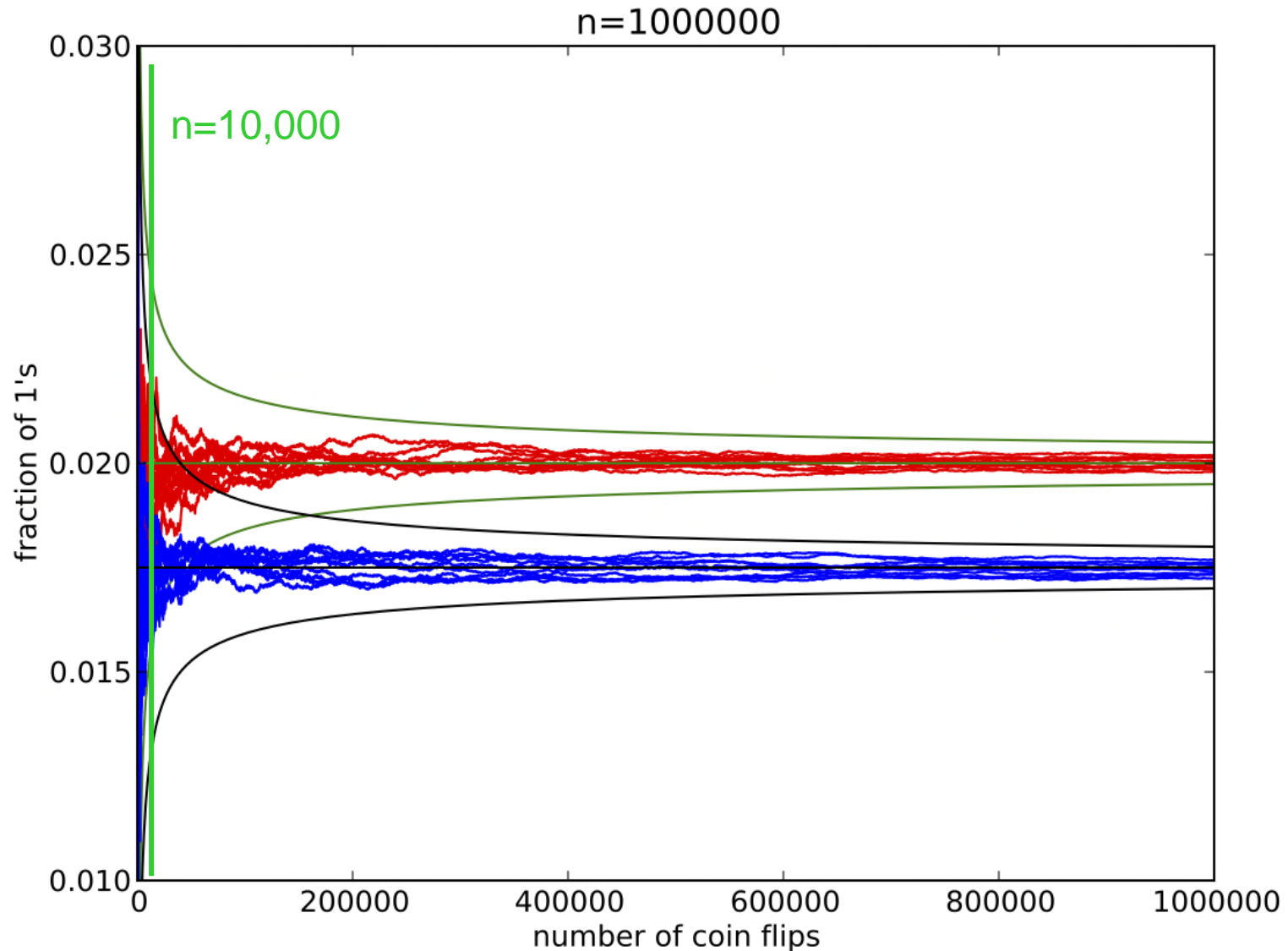
# Running averages after 10,000 trials

Each jagged line is the running average for one sequence

The smooth green and black curves define the “envelope” of likely sequences





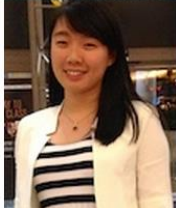


# Running average after 1,000,000 trials



# Class web site: <http://cse103.github.io>

## Instruction Staff

|                     |                     |   | Office hours                          | Office Location            | E-mail address   |
|---------------------|---------------------|---|---------------------------------------|----------------------------|--|
| Instructor:         | Yoav Freund         |    | Tuesday 10AM - Noon.                  | Office #4126, CSE building |  |
| Teaching Assistant: | Akshay Balsubramani |    | Monday 4PM-5PM, Thursday 10AM-11AM    | Office #4142, CSE building | <a href="mailto:abalsubr@ucsd.edu">abalsubr@ucsd.edu</a> |
| Teaching Assistant: | Ibrahim Awwal       |    | Tuesday 12PM-1PM, Friday 10AM-11AM    | Office #4223, CSE building | <a href="mailto:iawwal@ucsd.edu">iawwal@ucsd.edu</a>     |
| Teaching Assistant: | Sanjeev Rao         |   | Wednesday 11AM-12PM, Friday 11AM-12PM | Office #4223, CSE building | <a href="mailto:sjrao@ucsd.edu">sjrao@ucsd.edu</a>       |
| Tutor:              | Janet Zhen Zhai     |  | Monday 10AM-11AM, Wednesday 10AM-11AM | Office #4223, CSE building | <a href="mailto:zzhai@ucsd.edu">zzhai@ucsd.edu</a>       |

To contact the teaching staff send a private message through [Piazza](#).



# Webwork

- Weekly Assignments run from wed-noon to wed-noon
- Assignments are 60% of the final grade
- 3 lowest grades are dropped before taking the average.
- Collaborate to understand the problem and how to solve it but be sure to solve it yourself!
- Problems are often randomized – your problem is likely slightly different than that of your friend.

## Week1: Problem 3

Prev Up Next

This set is visible to students.

(1 pt) Reorganized/Orientation/prob03.pg

### Typing in Your Answers

Here are the standard symbols that WeBWork, along with most other computer software, uses for arithmetic operations:

| Symbol  | Meaning        | Example                   |
|---------|----------------|---------------------------|
| +       | Addition       | $3+4 = 7$                 |
| -       | Subtraction    | $3-4 = -1$                |
| *       | Multiplication | $3*4 = 12$                |
| /       | Division       | $3/4 = .75$               |
| ^ or ** | Exponentiation | $3^4 = 81$ Or $3**4 = 81$ |

Sometimes WeBWork will insist that you calculate the value of an expression as a single number before you enter it. For example, calculate the value of  $6(-3 + 4) - (6 - 5)$  and enter it in the following blank. (Here you have to enter a single integer; the question is testing whether you can do the operations correctly.)

$$6(-3 + 4) - (6 - 5) = \text{[input box]}$$

Most often you will not have to simplify your answer, but can let WeBWork do this for you. The following blanks are all expecting the value 16. Try entering it several different ways, such as  $7+9$ ,  $18-2$ ,  $8*2$ ,  $32/2$ , and  $4^2$ . Note: pressing the "Tab" key on your keyboard will move you from one answer box to the next.

$$16 = \text{[input box]} \text{ or } \text{[input box]} \text{ or } \text{[input box]} \text{ or } \text{[input box]} \text{ or } \text{[input box]}$$

WeBWork also understands that quantities written next to each other are supposed to be multiplied. For example, you can enter  $(9)(7)$  instead of  $63$ . Most often this is used when one quantity is a number and the other is a variable or function. For instance,  $2 \sin(30)$  or  $2x^2 + 5x + 12$  or  $2 \sin(x)$ . The following blank is expecting the value 128, so

# Piazza

- **Main purpose:** discussion of Webwork assignment
- Search and read previous postings before
- Use it for everything! Any problem/question/idea/complaint
- 3 lowest grades

The screenshot shows the Piazza web application interface. The browser address bar displays the URL <https://piazza.com/class/hzpd9rb1z9f3go?cid=1>. The page header includes the Piazza logo, navigation links for CSE 103, Q & A, Resources, Statistics, and Manage Class, and a user profile for Yoav Freund. Below the header, there are tabs for polls, homework (hw1, hw2, hw3, hw4), and a sidebar with filters for Unread, Updated, Unresolved, and Following. The main content area is titled 'note' and features a 'Welcome to Piazza!' message. The message explains that Piazza is a Q&A platform designed to get great answers from classmates and instructors fast. It includes a list of tips for getting started, such as asking questions, editing questions and answers wiki-style, and adding followup comments or further questions. The sidebar on the left shows a list of pinned posts, including 'Search for Teammates!', 'Introduce Piazza to your stu...', 'Get familiar with Piazza', and 'Tips & Tricks for a successf...'. Each post has a 'Private' label, a date of 9/5/14, and a '1' icon indicating the number of answers.

← → ↻ <https://piazza.com/class/hzpd9rb1z9f3go?cid=1> [Icons: Print, Download, Share, etc.]

Apps Recent changes - Se My Page AWS, Python & noteb Hadoop WebWork Googles TWiki UCSD banks Services » Other Bookmarks

**PIAZZA** CSE 103 ▾ Q & A Resources Statistics Manage Class Yoav Freund

polls hw1 hw2 hw3 hw4

Unread Updated Unresolved Following ⚙

**New Post** 🔍 Search or add a post...

▼ PINNED

- Private **Search for Teammates!** 9/5/14 1
- ▼ WEEK 8/31 - 9/6
- Private **Introduce Piazza to your stu...** 9/5/14 1
- Private **Get familiar with Piazza** 9/5/14
- Private **Tips & Tricks for a successf...** 9/5/14 1

**Welcome to Piazza!** 9/5/14

Piazza is a Q&A platform designed to get you great answers from classmates and instructors fast. We've put together thi

**note** ☆ 3 view

## Welcome to Piazza!

Piazza is a Q&A platform designed to get you great answers from classmates and instructors fast. We've put together this list of tips you might find handy as you get started:

- 1. Ask questions!**

The best way to get answers is to ask questions! Ask questions on Piazza rather than emailing your teaching staff so everyone can benefit from the response (and so you can get answers from classmates who are up as late as you are).

- 2. Edit questions and answers wiki-style.**

Think of Piazza as a Q&A wiki for your class. Every question has just a single **students' answer** that students can edit collectively (and a single **instructors' answer** for instructors).

- 3. Add a followup to comment or ask further questions.**

To comment on or ask further questions about a post, start a **followup discussion**. Mark it resolved when the issue has been addressed, and add any relevant information back into the Q&A above.

# Nota-Benne 1

nb.mit.edu

Apps Recent changes - Set My Page AWS, Python & note Hadoop WebWork Googles TWiki UCSD banks Services » Other Bookmarks

Yoav Freund Help

Home

★ Introduction to Probability and Statistics




★ Tensors

Admin Controls

Add file New folder Invite Users Users Sections Properties Spreadsheet Download as .xls

Contents of Introduction to Probability and Statistics

Sort by ▾ date added ▲ name

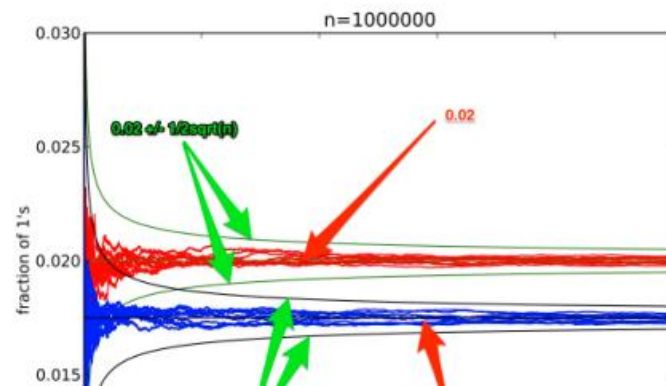
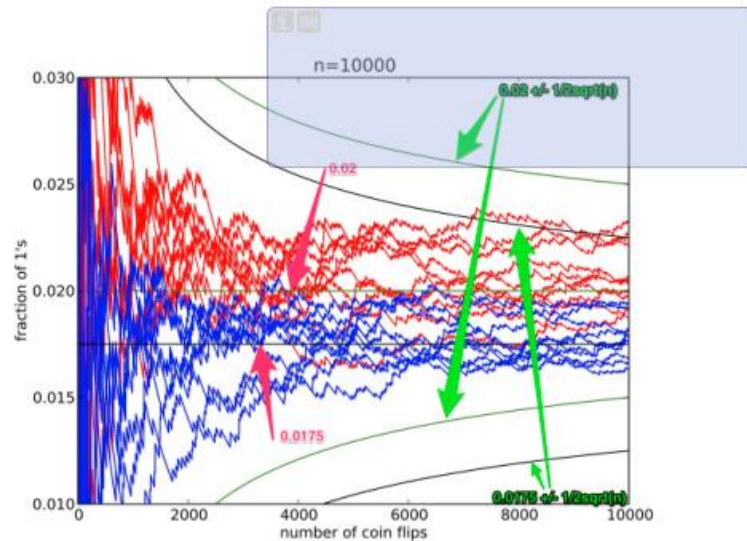
| Name  | Assignment | Download PDF                                       | Stats               | Actions                 |
|---|------------|--|---------------------|-------------------------|
|  <a href="#">GrinsteadSnell.pdf</a>        | No         | <a href="#">original</a>                           |                     | <a href="#">Actions</a> |
|  <a href="#">ProbabilityStatistics.pdf</a> | No         | <a href="#">original</a> <a href="#">annotated</a> | me 3 unread 0 all 4 | <a href="#">Actions</a> |
|  <a href="#">thinkstats.pdf</a>            | No         | <a href="#">original</a>                           |                     | <a href="#">Actions</a> |

You have 0 feedback request.

Your classmates have no pending questions.

# Nota Benne 2

## 1.5. MONTE-CARLO SIMULATIONS



11

3 threads

me 3

★ 0

? 0

1 thread on page 11

1 i me What is the difference between the green and the red arro...

3 threads on page 29

2 i me I think you mean 3 shared, rather than 5 shared...

1 i me seen only by the player who received them. They are re...

1 i me Delete

? + 0 - replies requested

What is the difference between the green and the red arrows?

Yoav Freund i me - 02:42PM

# For Monday

- Make sure you have accounts on:
  - Webwork (ID@ucsd.edu, password=PID)
  - Piazza
  - Nota Benne
- Read chapter 1 of the class notes, comment in places that you don't understand.
- Start on webwork week1 assignment: is due next wed at noon!
- Post your questions on Piazza
- See you on Monday!