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 ?
- Dice of the same color are
  indistinguishable or interchangeable

Indistinguishability / Exchangeability

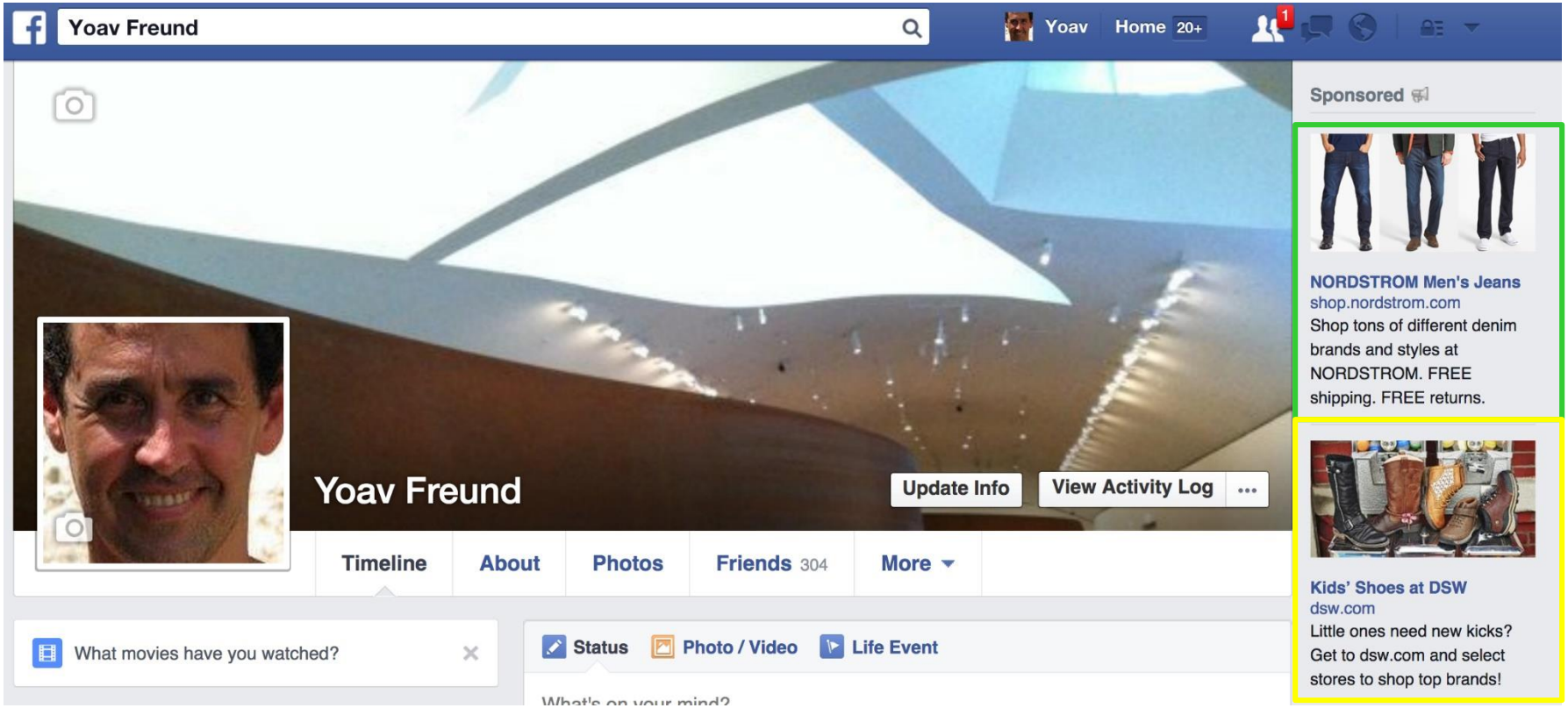
- Two object are indistinguishable if exchanging them makes not difference.
- What makes two objects indistinguishable?
- Mathematical objects (points, lines) are indistinguishable
  - Physical objects are distinguishable (we can mark them)
- Poker cards are indistinguishable, unless marked (illegal)
- 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 fruit indistinguishable? (same DNA)
- Are animals of a species distinguishable?
- Are people distinguishable?

Probabilities regarding people

- Which of the following is more correct?
  1. Each of us is unique, we have our own free will.
  2. 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:
  1. How many children in this district have special needs?
  2. How many voters in San Diego county are likely to vote republican?
  3. Is the compulsory seat belt law beneficial to society?
- Are we all the same or are we all different?

We are individuals – The life of Brian

Ads on my Facebook Page



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 – 10 dollars).
- 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 probabilities of click through are P1=0.02 (2%)
    and P2=0.0175 (1.75%)
  - But we don't know whether
    - Pa=P1 > P2=Pb or
    - Pa=P2 < P1=Pb
  - We need to find out by experiment

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 Pa
- Same for ad b
- By repeating a,b,a,b,a,b,a,… forever we will find out Pa and Pb.
- This is called "the law of large numbers"
- But we cannot wait forever!
- How many times do we need to display each ad in order to find out which has click-through rate of 0.02 and which has click through rate of 0.0175?
- This is a typical statistics question.
- We will give increasingly more accurate answers to this questions throughout the quarter.
- For now, consider some examples.

The sequence of running averages

Notation: $X_t = 1$: Click Through $X_t = 0$: No Click Through

Suppose we focus on the sequence corresponding to one of the two ads:

$$X_1, X_2, \ldots, X_n, \ldots = 0,0,1,0,0,0,0,1, \ldots$$

The running averages (Fractions of 1's) are:

$$X_1, \frac{X_1+X_2}{2}, \frac{X_1+X_2+X_3}{3}, \frac{X_1+X_2+X_3+X_4}{4}, \ldots$$

For the sequence above the running averages are:

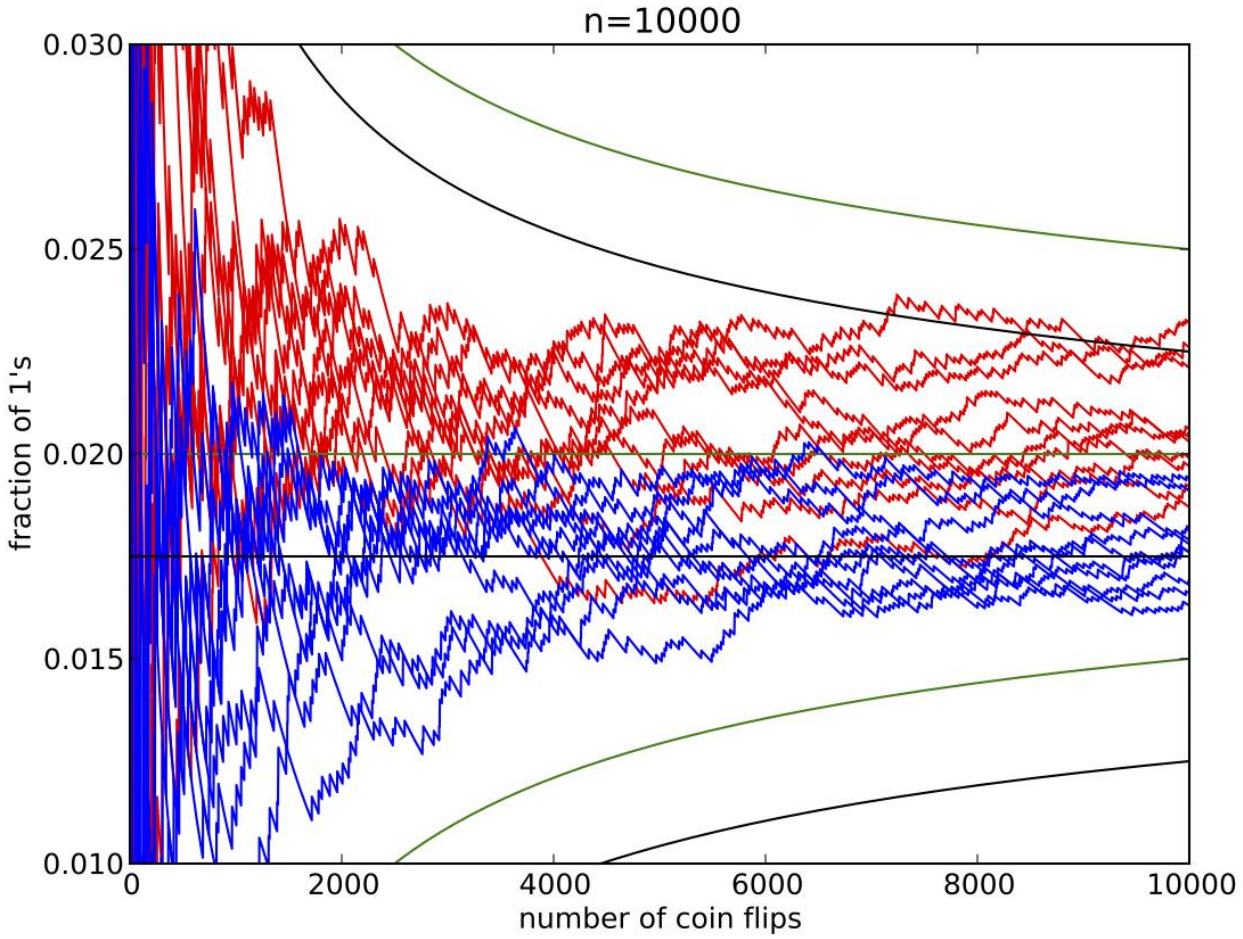
$$\frac{0}{1}, \frac{0}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \frac{1}{6}, \frac{1}{7}, \frac{2}{8}, \ldots$$

The law of large numbers implies that, if the probability of click-through Is a constant, the running average sequence converges over the long term

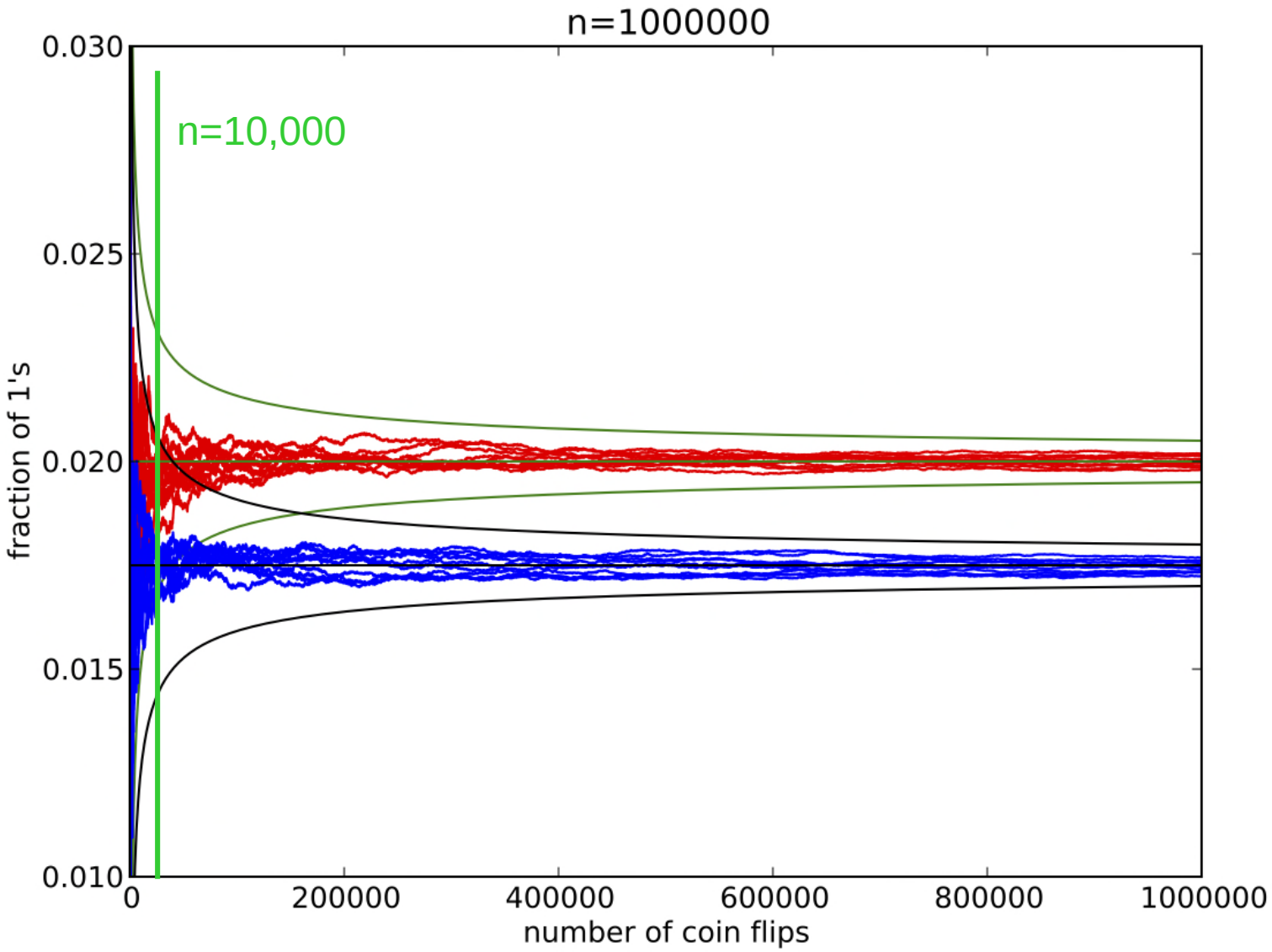
Running averages after 10,000 trials

Each jagged line is the running average for one sequence

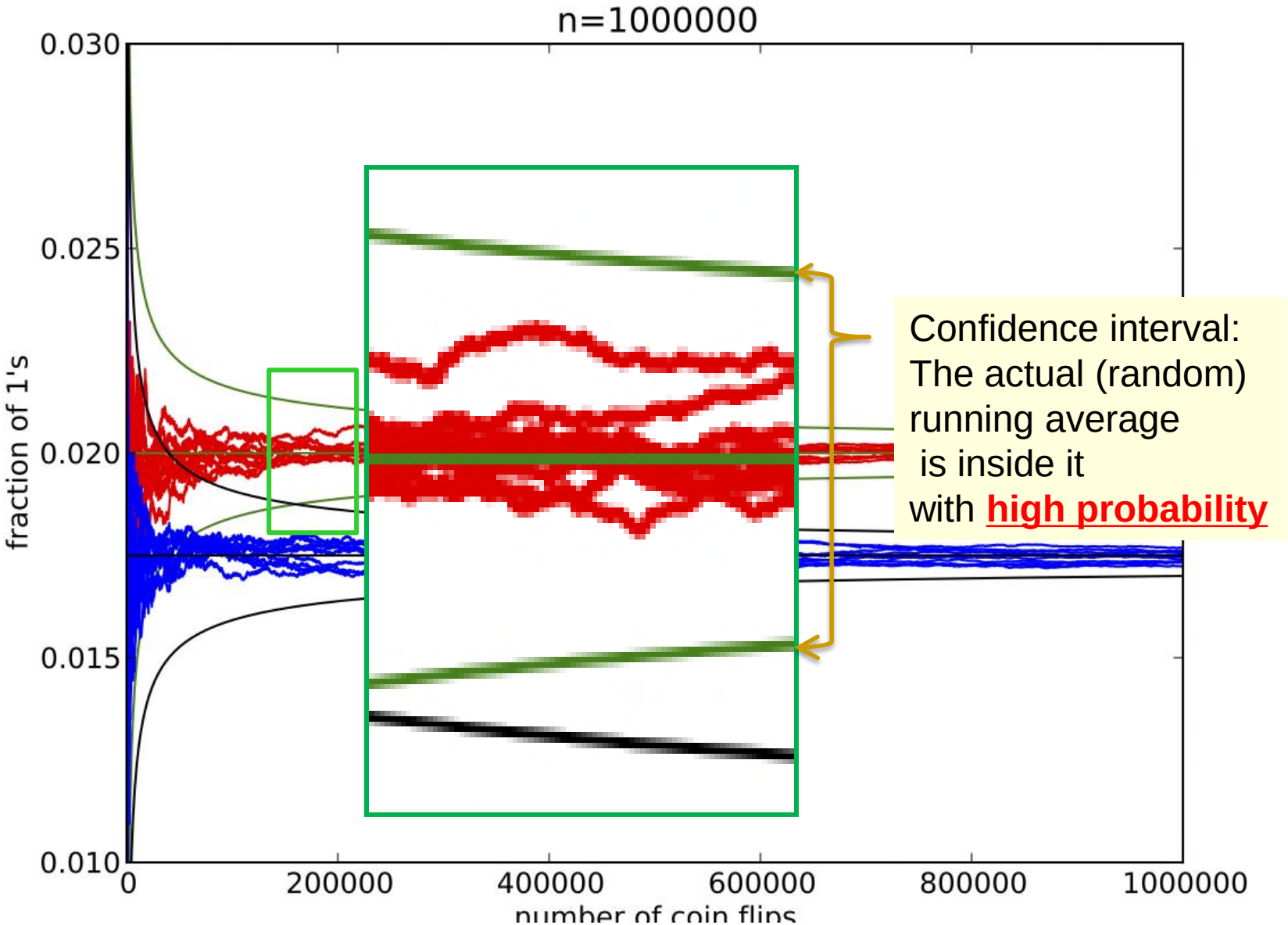
The smooth green and black curves define the “envelope” of likely sequences



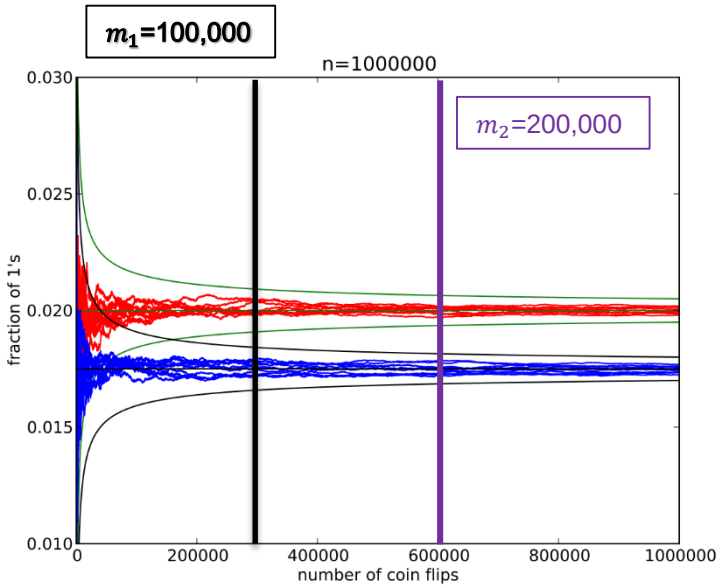
Running average after 1,000,000 trials



Confidence Intervals



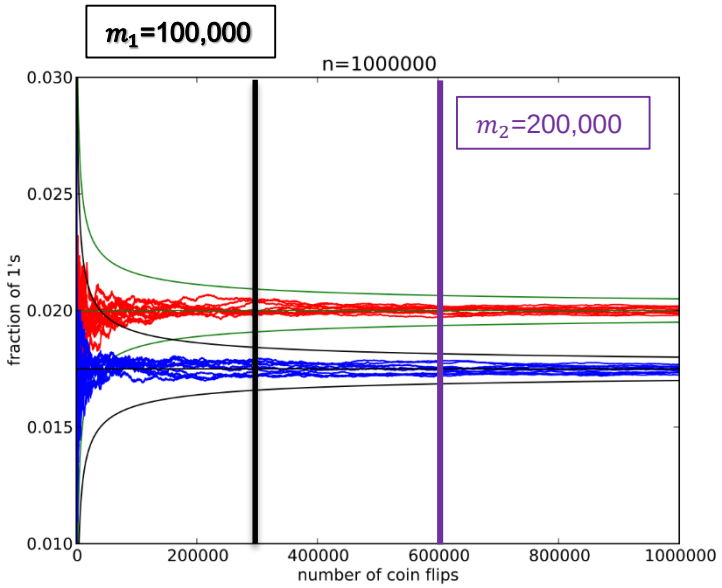
Length of confidence interval



- Consider repeating the experiment 100,000 vs. 200,000 times.
- Doubling the number of experiments decreases the length of the confidence interval. (Keeping confidence level fixed)
- By how much?

(a) by 2 (b) by $\sqrt{2}$ (c) by 4

Level of confidence



- Again, Consider repeating the experiment 100,000 vs. 200,000 times.
- Suppose we keep the length of the interval fixed.
- In this case the confidence increases.
- Suppose the confidence of 100,000 is 90%, what is the confidence of 200,000?

(a) 92% (b) 95% (c) 99%

90% = 0.9 = 1-0.1
99% = 0.99 = 1-0.01 = $1 - 0.1^2$

Goal of this course

- The last example hints at how probability and statistics can be used to deal with uncertainty.
- At the end of the course you would be able to explain why sqrt(2) and 99% are the correct answers.
- Along the way we will explore games of chance, poker, Combinatorics, Bayesian inference, Hypothesis testing, randomized algorithms and many other subjects.
- The main components of the course are:
  - Lectures
  - Webwork (60% of final grade)
  - Office hours, Discussion sections, and Piazza.
  - Final exam (40% of final grade).

Webwork

- Weekly Assignments run from thu 5pm to thu 5pm
- 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.)

$6(-3+4)-(6-5)=$ [ ]

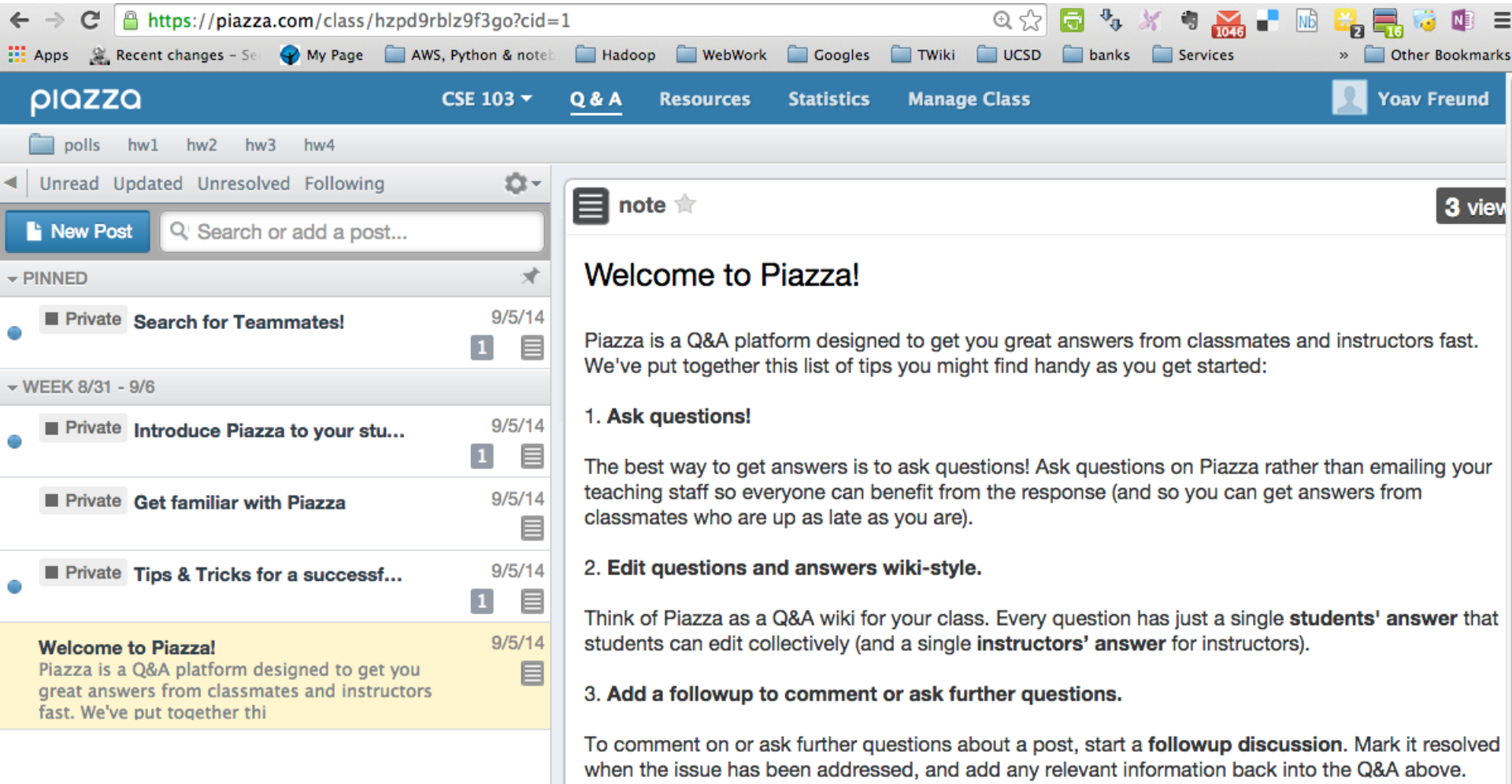
Most often you will not have to simplify your answer, but can let WeBWorK do this for you. The following blanks are all expecting the value 16. Try entering it several different ways, such as `7+9`, `18-2`, `8*2`, `32/2`, and `4^2`. Note: pressing the "Tab" key on your keyboard will move you from one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

16 = [ ] or [ ] or [ ] or [ ] or [ ]

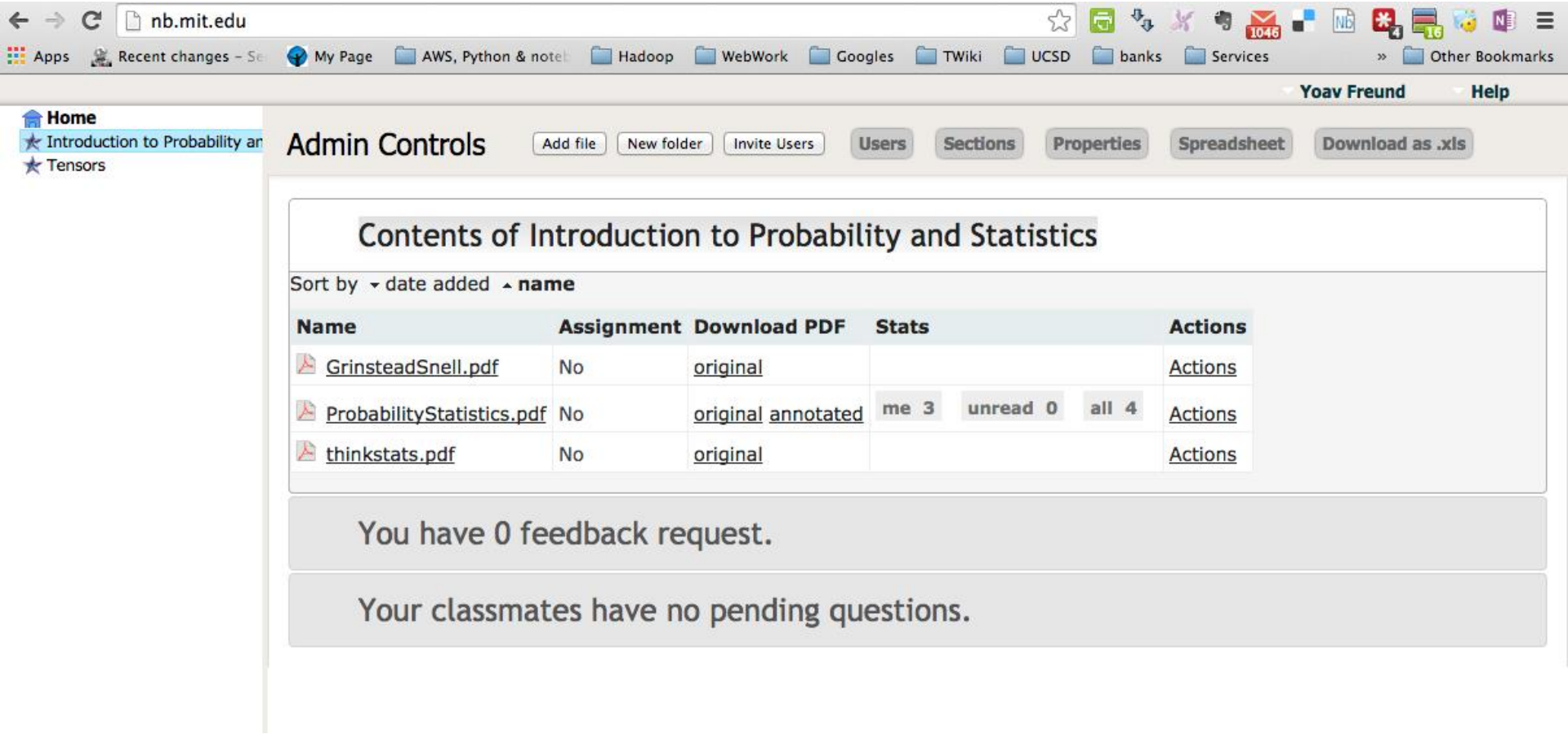
WeBWorK also understands that quantities written next to each other are supposed to be multiplied. For example, you can enter `(9)(7)` instead of `63`. Most often this is used when

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



Nota-Benne 1


nb.mit.edu
Apps
Recent changes – Se
My Page
AWS, Python & notel
Hadoop
WebWork
Googles
TWiki
UCSD
banks
Services
Other Bookmarks
Yoav Freund
Help
Home
Introduction to Probability ar
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
GrinsteadSnell.pdf
No
original
Actions
ProbabilityStatistics.pdf
No
original annotated
me 3
unread 0
all 4
Actions
thinkstats.pdf
No
original
Actions
You have 0 feedback request.
Your classmates have no pending questions.

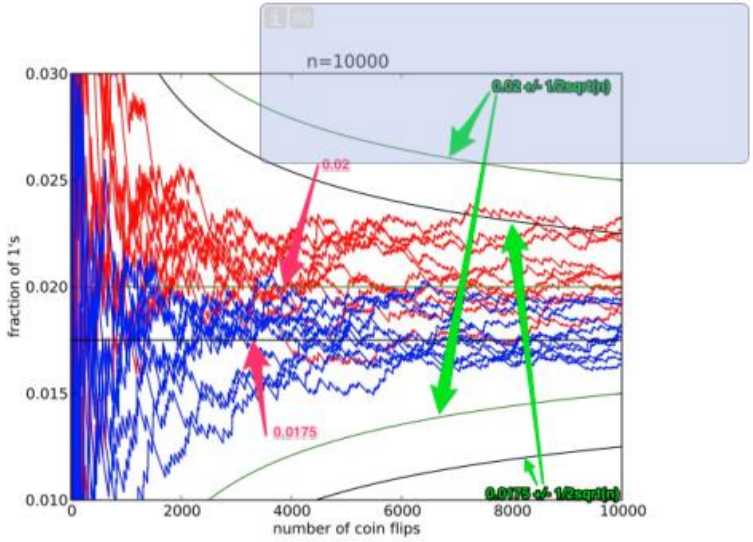
Nota Benne 2

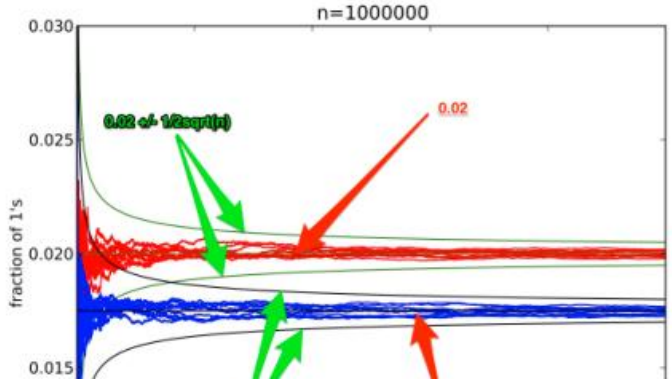
nb.mit.edu/f/16559

Apps | Recent changes – Se | My Page | AWS, Python & notet | Hadoop | WebWork | Googles | TWiki | UCSD | banks | Services | Other Bookmarks

Yoav Freund | Help

1.5. MONTE-CARLO SIMULATIONS 11





3 threads | me 3 | 0 | 0

1 thread on page 11

1 i me What is the diffence 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 diffence between the green and the red arrows?

Yoav Freund i me – 02:42PM

For next class

- 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 thu at 5pm!
- Post your questions on Piazza

- See you on Tue!