

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 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?
 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


 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 - 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
 - $P_a=P1 > P2=P_b$ or
 - $P_a=P2 < P1=P_b$
 - 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 P_a
- Same for ad b
- By repeating a,b,a,b,a,b,a,... forever we will find out P_a and P_b .
- 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:

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

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

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

For the sequence above the running averages are:

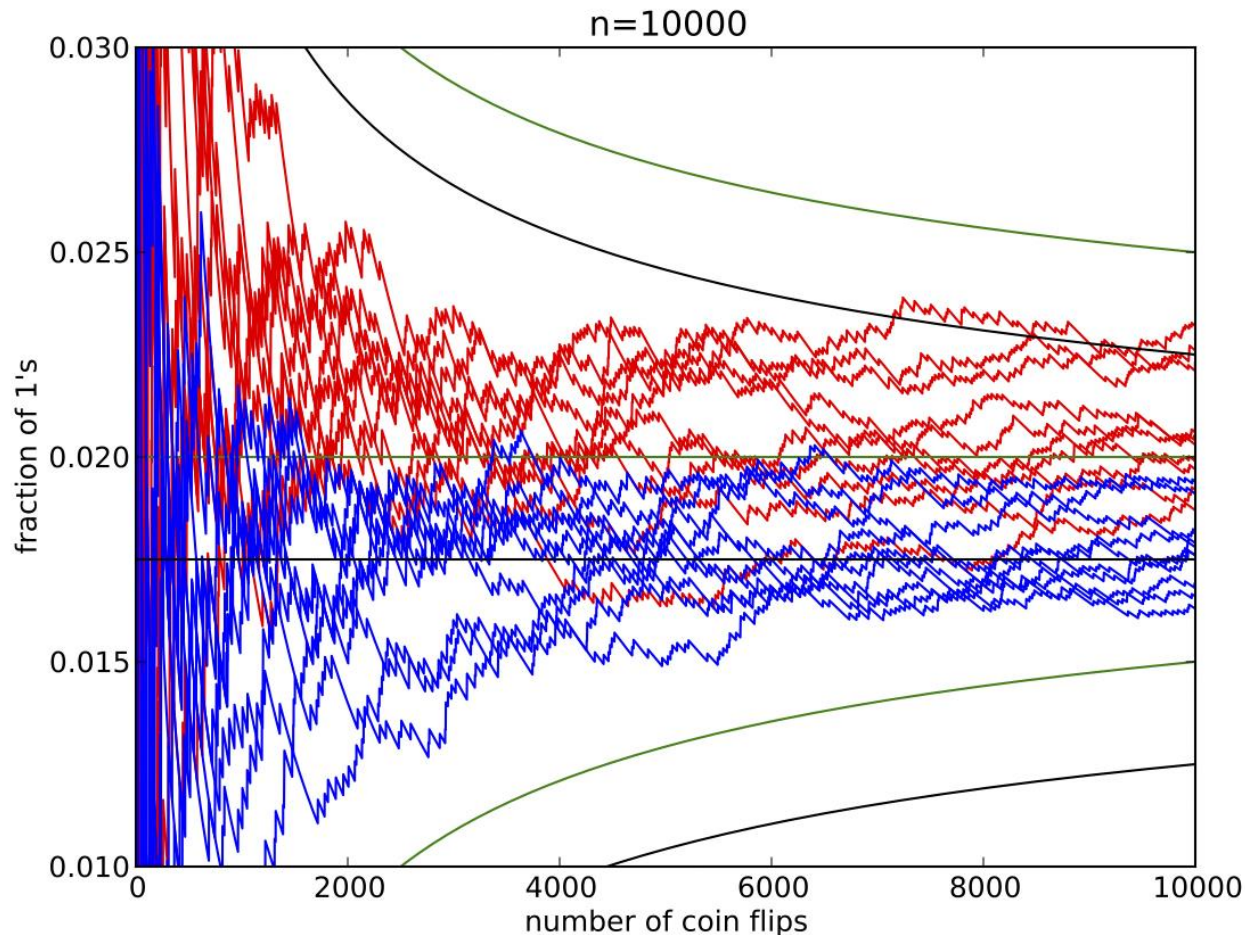
$$\frac{0}{1}, \frac{0}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \frac{1}{6}, \frac{1}{7}, \frac{2}{8}, \dots$$

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

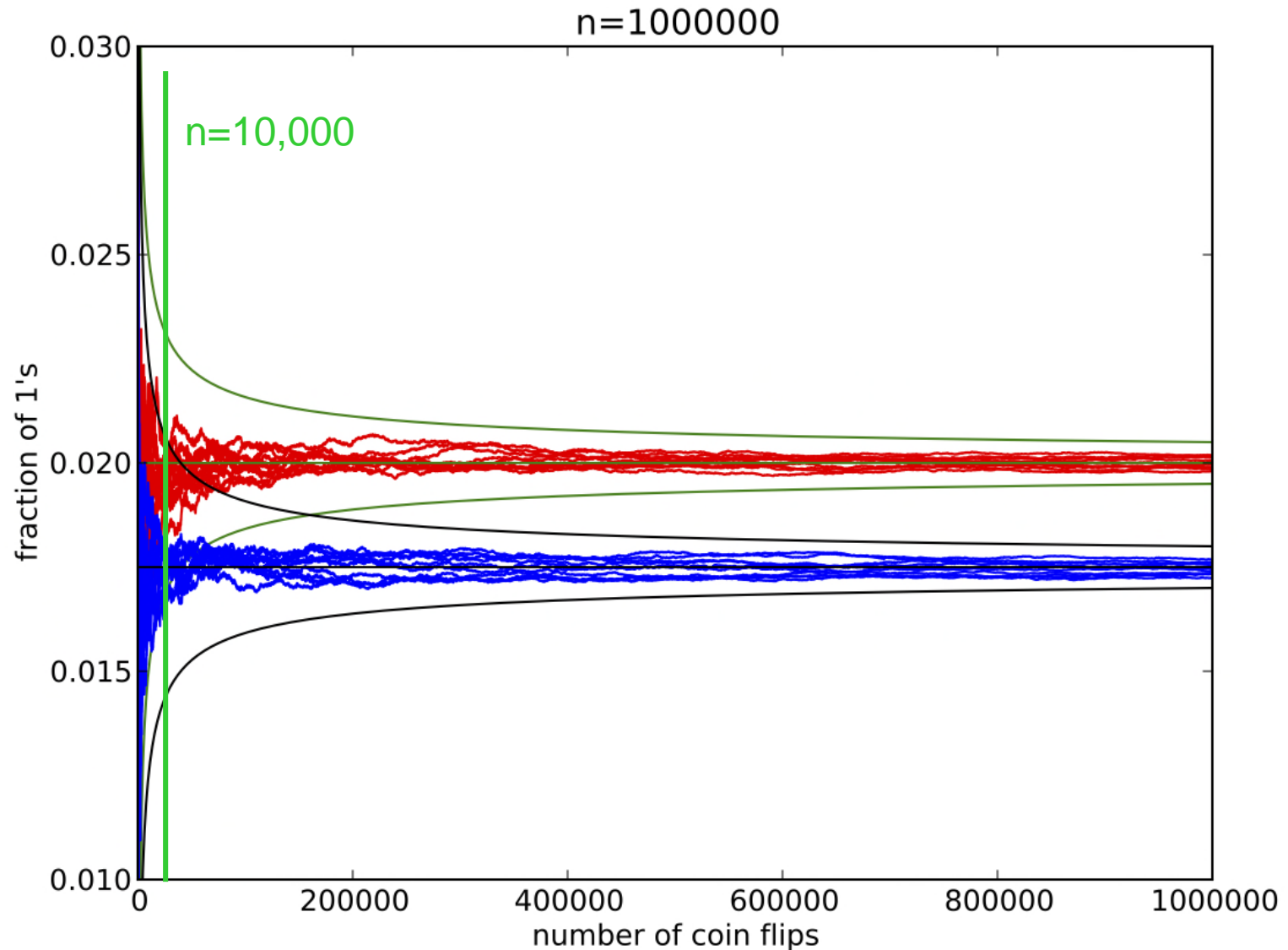
Running averages after 10,000 trials

Each jagged line is the running average for one sequence

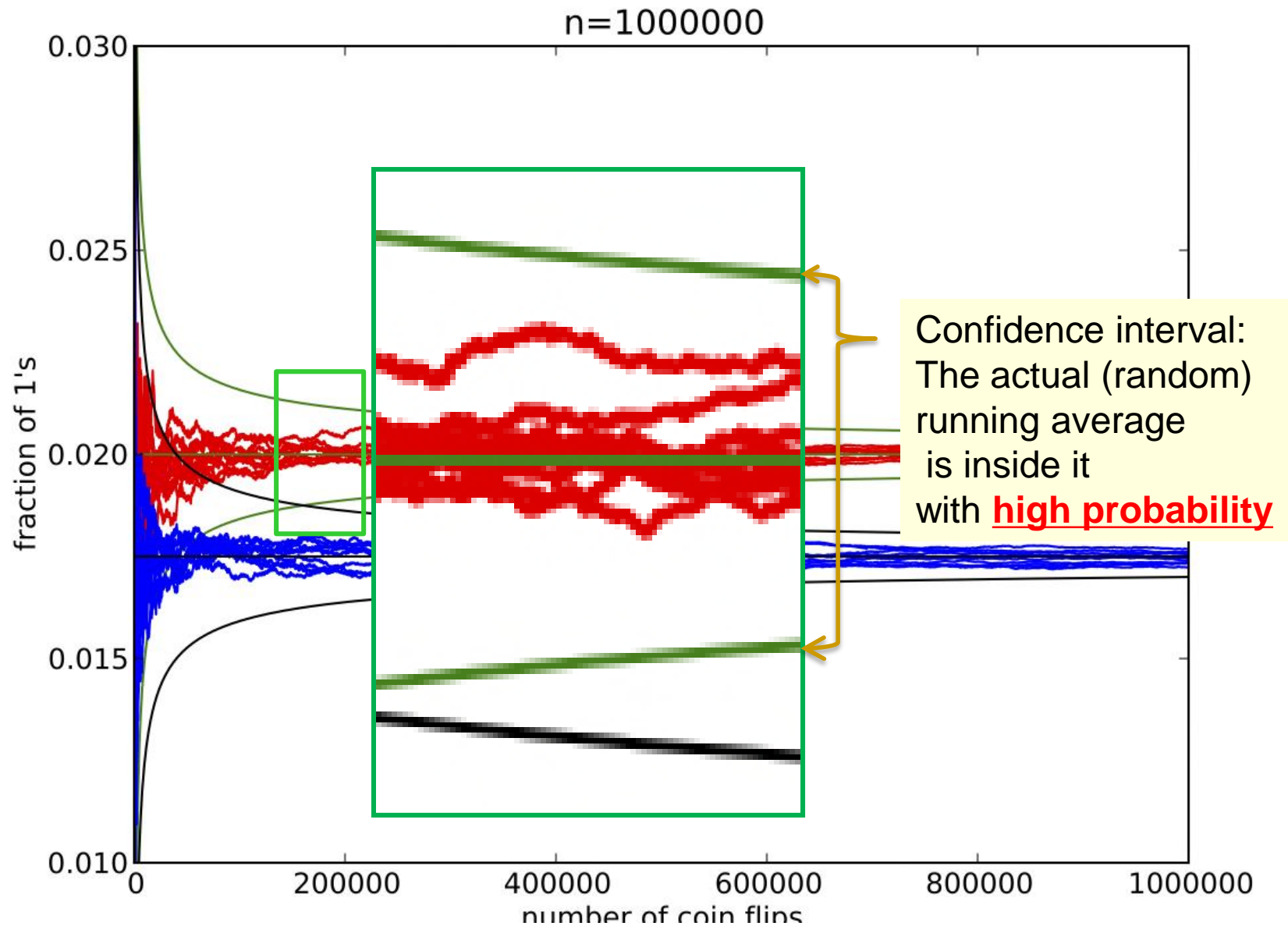
The smooth green and black curves define the “envelope” of likely sequences



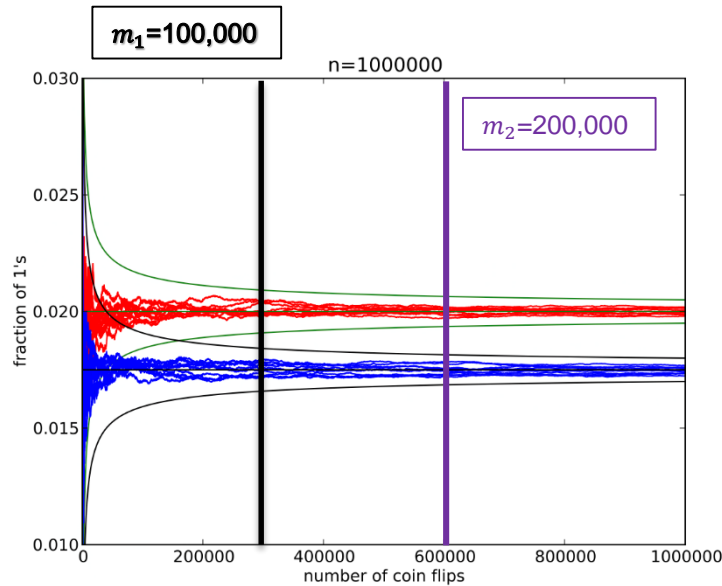
Running average after 1,000,000 trials



Confidence Intervals



Length of confidence interval



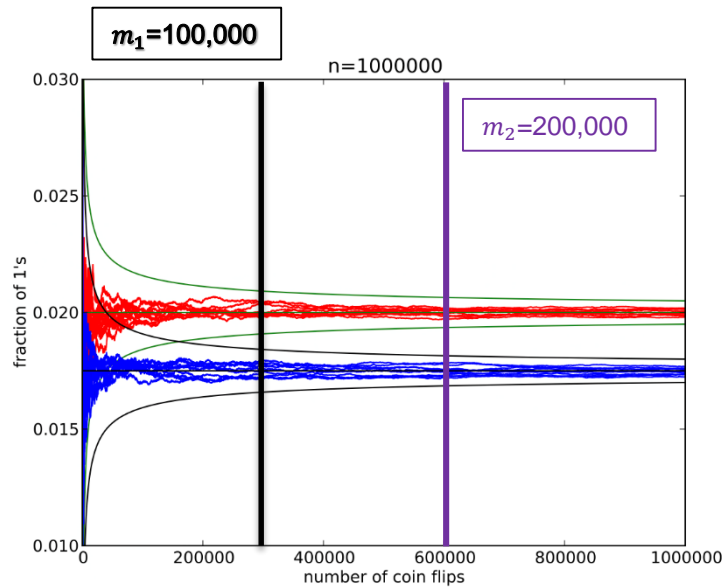
- Consider repeating the experiment 100,000 vs. 200,000 times.
- Doubling the number of experiments decreases the length of the confidence interval. (Keeping confidence level fixed)
- By how much?

(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.)

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

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 120, 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 website interface for a CSE 103 class. The browser address bar shows the URL <https://piazza.com/class/hzpd9rb1z9f3go?cid=1>. The page has a blue header with 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 displays a list of pinned posts, including a welcome message and several private posts. The right sidebar shows a 'note' section with a star icon and a '3 view' button. The main content area also features a 'Welcome to Piazza!' section with a list of tips for getting started.

PIOZZA 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

AppsRecent changes - SetMy PageAWS, Python & notebHadoopWebWorkGooglesTWikiUCSDbanksServices»Other Bookmarks

Yoav FreundHelp

Home

★ Introduction to Probability and Statistics




★ Tensors

Admin Controls

Add fileNew folderInvite UsersUsersSectionsPropertiesSpreadsheetDownload as .xls

Contents of Introduction to Probability and Statistics

Sort by ▾ date added ▲ name

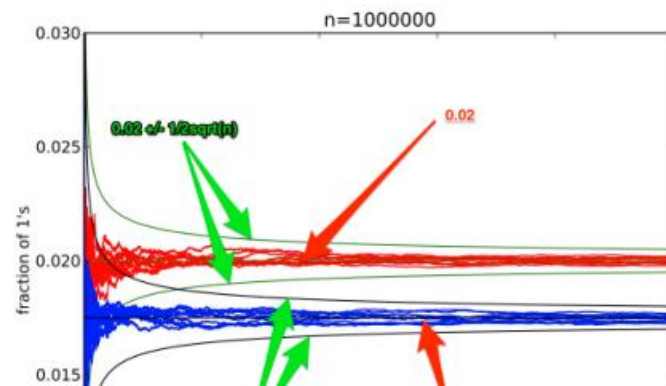
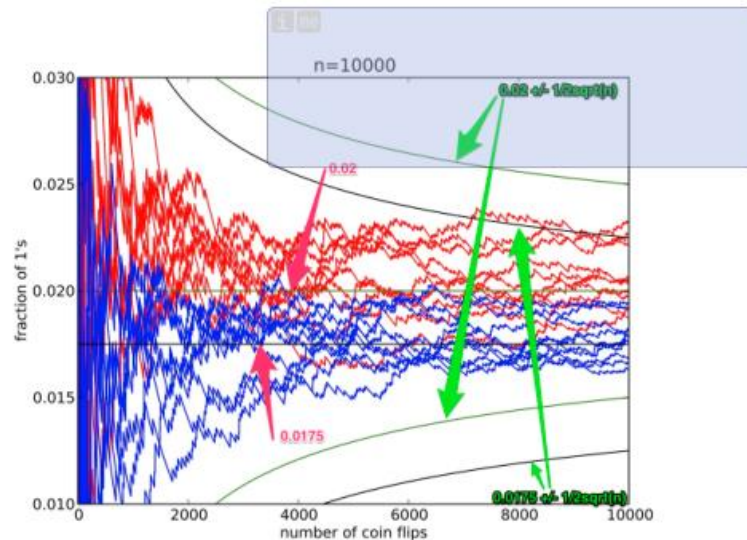
Name	Assignment	Download PDF	Stats	Actions
 GrinsteadSnell.pdf	No	original		Actions
 ProbabilityStatistics.pdf	No	original annotated	me 3unread 0all 4	Actions
 thinkstats.pdf	No	original		Actions

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 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!