.

Welcome to Statistics 217!

Week 1 Tuesday

11 Jan 2011

Announcements

- Do the items on the ToDo sheet!
- Mark office hours sheet (free times $= \cdot$, preferred times $= \odot$)
- Textbook in bookstore, but early readings available on line.
- Reading for next class:

  A. Syllabus

  B. Dialogue on Functions

  C. *Probability Explained*, "How to use this book"

  D. *Probability Explained*, pages 1–11 (through Example 1.2).

- HW 1 due next week (note grace period)
- Handouts: **Syllabus, Calendar, ToDo**

Plan

1. Who, What, and Why
2. Course Logistics
3. Strategies for Math
4. Mathematical Objects, part I

Plan

1. Who, What, and Why
2. Course Logistics
3. Strategies for Math
4. Mathematical Objects, part I

Who

- Who am I?
- Who are you?
  - interests? goals? experience?
  - programming? hardware? math?
  - who sees this material as directly important for downstream work?
  - why are you here?
  - what will you need to do with it?

What

OBJECTIVE: To teach you to develop mathematical and computational models of uncertain or random systems so as to make better predictions and decisions.

By the end of this course, you will be able to

– build effective models of realistic random system;

– use the models to make decisions;

– plan an approach, generate strategies, and solve problems that require probabilistic thinking;

– explain fundamental concepts in probability theory.

Why

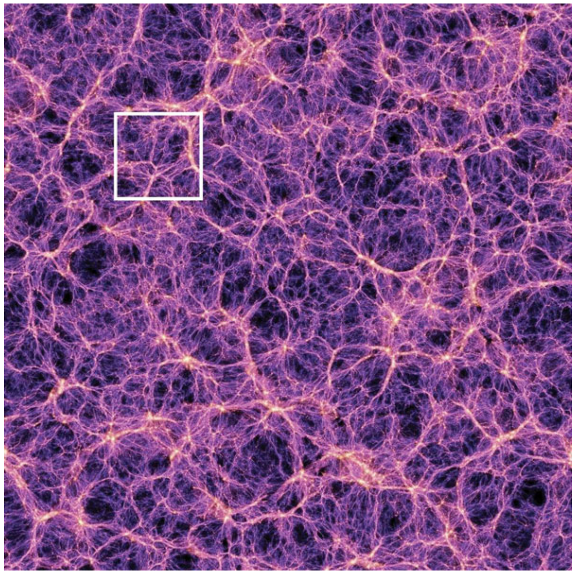
1. We live in an uncertain (and increasingly data rich) world.

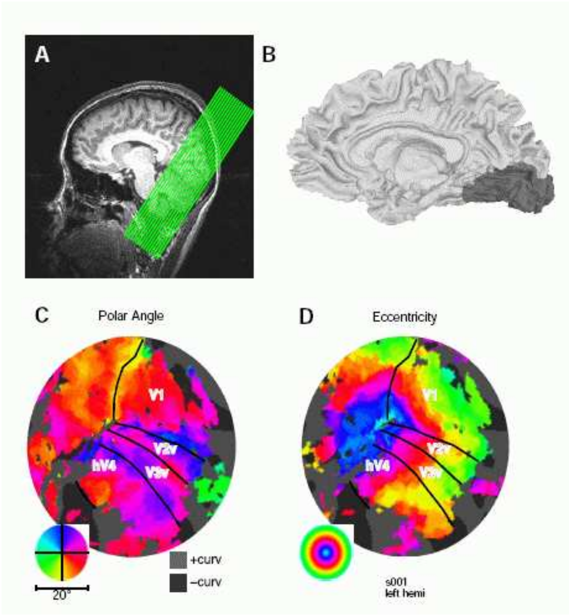
2. We constantly need to make decisions based on uncertain and incomplete information.

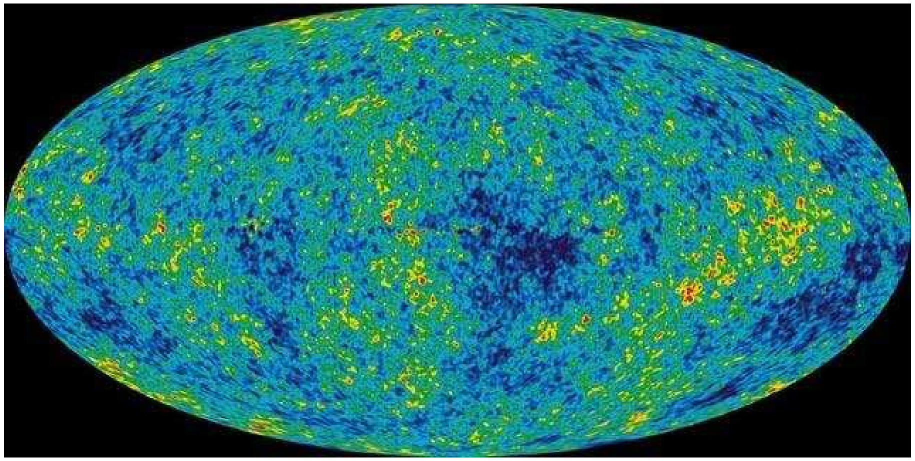
3. We are rather bad at it, with intuition alone.

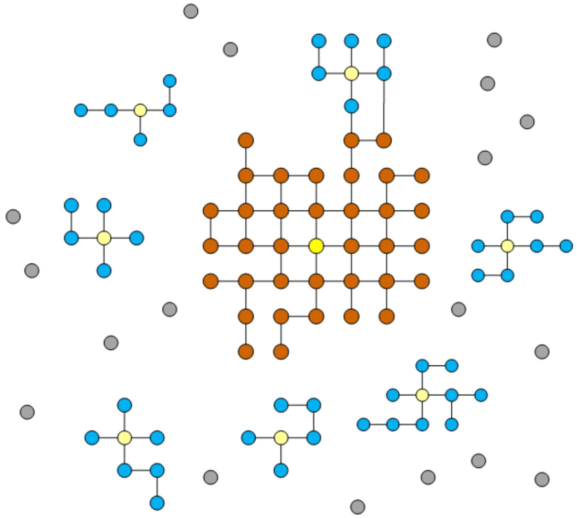
"Say ... what's a mountain goat doing way up here in a cloud bank?"

Why




A
B
C
Polar Angle
D
Eccentricity
V1
V2v
V3v
hV4
+curv
-curv
20°
s001
left hemi





Why

Applications in EE:

- In modern digital systems, probability models are used to analyze and optimize performance for low-power consumption.
- In multicore systems, probabilistic criteria determine the size of the buffers used to store data (typically packets) in transit.
- Deep submicron technologies used to manufacture high performance chips make parameter variations unavoidable, and so various probability-based metrics are needed to quantify the yield and system lifetime.
- In communication systems, probability calculations are fundamental to system analysis and optimization.

Applications in CS:

- Analysis of algorithms
- Design and optimization of real-time networks and queues
- Machine learning and data mining techniques
- Randomized algorithms and heuristics for search and optimization problems
- Data privacy and cryptography

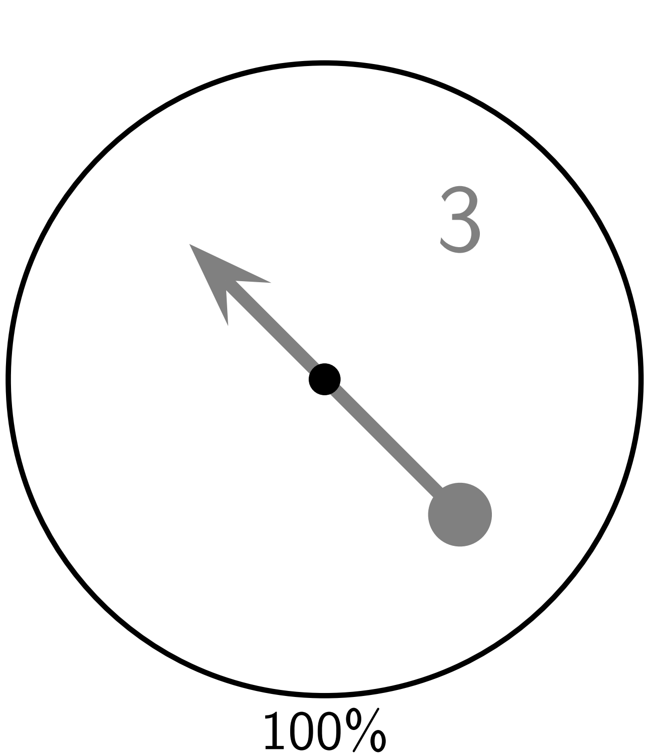
Why

More generally, the following is just a sampling of application areas:

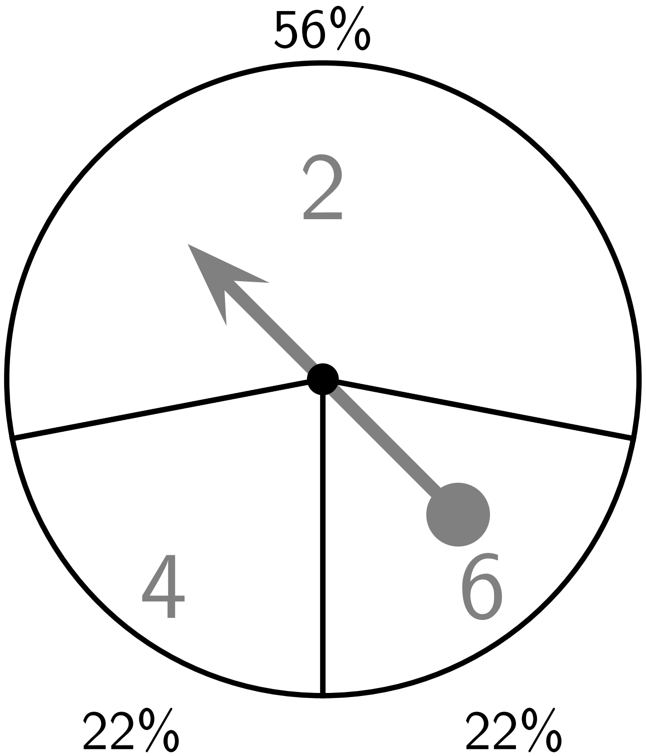
- Signal Processing
- Circuit Design (noise!)
- Simulation, Modeling, and Design
- Robotics and Stochastic Control
- Algorithm Analysis and Design
- Quantum Computation
- Statistics and Machine Learning
- Population Genetics and Evolutionary Modeling
- Forecasting, Optimization, Quality Control
- Econometrics and Game Theory

Why

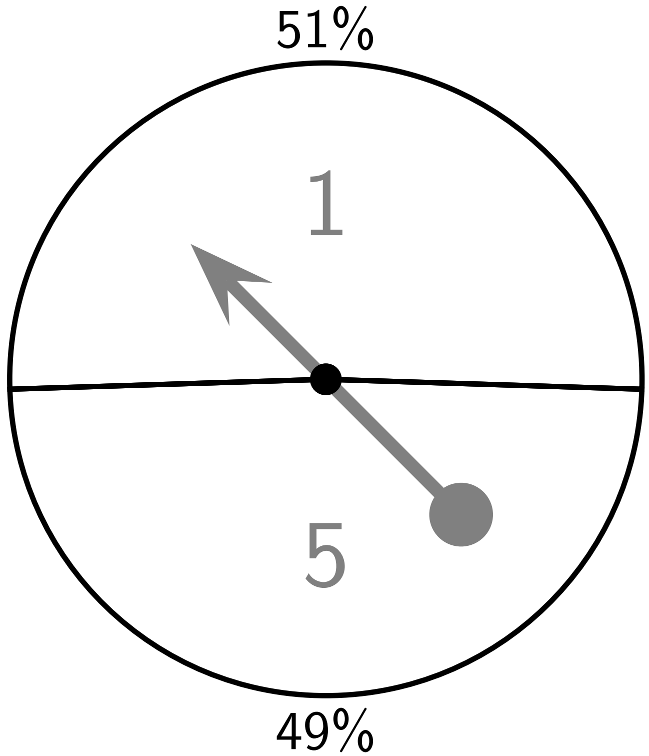
A



B



C



Plan

1. Who, What, and Why
2. Course Logistics
3. Strategies for Math
4. Mathematical Objects, part I

About This Course

- If you know how to compute an average, then already understand the primary operation we will use.

  (Of course, we will take this idea in new directions.)

- Layers: terminology/notation, mechanics, meaning, logic, modeling

- The role of mathematics . . .

- The hard truth

- My goal. . .

Philosophy I: What I care about

1. Read

2. Think (and talk) about the material

3. Practice

Philosophy II: What I want

1. Ask questions
2. Attend class (and office hours)
3. Do the assigned reading before class
4. Make a good effort on the homework assignments.
5. Have fun

Philosophy III: Read your way to success

The textbook is organized differently than most other books on the topic. It is intended to be read.

Reading well is the surest path to success in this course.

If you read the assignments, you will understand the material better and do better on the homeworks and exams.

You should expect to read about 6 pages per day (that's per day, not per class).

Tips:

- Read, don't skim (Active reading.)
- Record questions as you have them and come talk to me
- Stick to the same notation and conventions.
- Use the problem-solving procedure.
- Try the supplementary exercises as part of your study.

Logistics I: Design and Timing

Three Week Cycle for Each Topic

Week n: Topic Introduced in Class

Week $n+1$: Homework on Topic Due

Week $n+2$: Mini-Test on Topic

This cycle is repeated for every major topic throughout the semester as detailed on the course Calendar.

Note: Calendar updates will be announced on the home page. This should not happen too often, but be sure to check when they do.

Weekly Structure

Tuesday

– Lecture developing new material

Thursday

– Problem and Case Study Workshop

– Mini-Test (∼20 minutes, 1 problem)

Reading assignments are timed to this structure. Classes assume you are up to date with the readings.

Logistics II: Clickers

- A tool for rapid practice with the ideas that provides immediate feedback.
- Helps improve teaching and learning in the class.
- Required! ($\sim$ 10% of the final course grade)
- You can obtain from the bookstore. You need to pick them up at the information desk.

  Registration instructions are given on the home page.

  Resell value to the bookstore typically $>$ 50% cost.
- **Bring them to every class.**

  You need one by Tuesday at the latest, but bring it on Thursday if you have it.

Logistics III: Other Issues

- Office Hours (be sure to mark the sheet being passed around)
- Class Notes
- Homework

$$\text{homework grade} = \min\left(100\frac{2p+3m}{2.35n}, 120\right)$$

- Mini-tests
- Grading

15%	Class participation, Attendance, and Clickers
25%	Homework
40%	Mini-Tests
20%	Final Exam

- Questions?

FAQ

1. Can you explain the three week cycle?

   Briefly: Introduction → Practice → Assessment

2. Why do we need the clickers?

   Interactivity and feedback that's just not possible any other way.

3. Can I really get an overall score above 100% on homeworks?

   Yes, up to 120% depending on how many problems you master.

4. What is mastery anyway?

   Check the syllabus for details. But basically it means you really got the key ideas. Minor typos or arithmetic errors will not keep you from mastery.

5. Why not just give us the formulas and let us at it?

   Let me just say that—

6. And isn't this all rather verbose for a "simple" subject?

   Verily, yet verbosity has value if used validly and can vivify as well as enervate.

Plan

1. Who, What, and Why
2. Course Logistics
3. Strategies for Math
4. Mathematical Objects, part I

Dealing with the Mathematics

- Mathematics is a convenient and powerful language for expressing patterns and logical relationships.

  It is the basic language we use for constructing stochastic models.

- One challenge in learning probability is mastering the various mathematical manipulations, terminology, and notation.

- But if you can read a schematic, decode digital logic, program a computer, or understand a scientific/engineering model – then you can handle the mathematical notation and terminology.

- We will develop a variety of strategies to make this much easier, including the following . . . .

Strategy #1: Think Like A Compiler

- Like a programming language, mathematics has a very specific syntax and semantics that we can exploit to make sense of expressions.
- Each object has a definite type and can only be compared to or assigned to objects of the same type.

  Each operator that we use acts on objects of a specific type in a specific way.
- For example, suppose (x, y) represents a coordinate on a map. What is wrong with the following?

$$(x, y) = 2$$

  If (u, v, w) represents a point in space, what is wrong with the following?

$$(x, y) = (u, v, w)$$

- The main types of objects we will use are numbers, vectors, sets, and functions.

  Understanding these objects and the operators that act on them will take us a long way.

Strategy #2: Explain it to yourself

A good way to test (and refine) your understanding of a mathematical expression is to explain it to yourself in colloquial language. Connect the expressions to their meaning in the larger context

As simple example, suppose you were taking a class with ten homework scores $h_1, \ldots, h_{10}$; three exam scores e_1, e_2, e_3; and a final exam score f. Then, your final grade g might be

$$g = \frac{1}{4} \cdot \frac{h_1 + h_2 + \cdots + h_{10}}{10} + \frac{2}{5} \cdot \frac{e_1 + e_2 + e_3}{3} + \frac{7}{20} \cdot f.$$

How would you explain this to yourself?

We will see many better examples where this can help. Probability is full of them.

Strategy #3: Try a Concrete Special Case

When in doubt, simplify. Reduce to a simpler case to get a feel for an expression.

Particularly useful is plugging in some arbitrary numbers to help you see a pattern.

For example, suppose we have functions p_k for $k \in \{1, \ldots, 100\}$ that satisfy

$$p_k(k + r) = \frac{1}{3} 2^{-|r|}, \qquad \text{integer } r.$$

What sense can we make of this? Can we find a simpler expression for p_k?

Plan

1. Who, What, and Why
2. Course Logistics
3. Strategies for Math
4. Mathematical Objects, part I

Numbers

- We mostly use two kinds: real numbers ($\mathbb{R}$) and integers ($\mathbb{Z}$).

  Unless otherwise specified, assume a variable is real valued. To be more specific, we add qualifiers, e.g., "x is a non-negative real number" or "j is a positive integer."

- Ranges of Real Numbers and Integers
  - Intervals: All real numbers between specified bounds, with endpoints (brackets) or without endpoints (parentheses)

    $[0, 1]$, $(-1, 1]$, $[1, \infty)$, $(-2, 2)$, $(17, \infty)$, $[a, b]$
  - Increments: List of consecutive integers, with endpoints (brackets) or without endpoints (parentheses)

    $[1 \,..\, 10]$, $(10 \,..\, 20]$, $[0 \,..\, 100)$, $(-17 \,..\, 17)$, $[12 \,..\, \infty)$, $[i \,..\, j]$

- We also use vectors of numbers, which I will discuss next time.

Variables

- Variables are symbolic names for an unknown or unspecified object.

  Non-random variables are always named with lowercase greek or roman letters, possibly with subscripts or superscripts attached.

  Whenever possible, the letters used for the names will have some mnemonic or historical meaning.

- Idioms for defining variables: "Let x be ..." or "Let $0 \leq p \leq 1$ be ..." or "Let x_k be ... for k satisfying ..." or "Suppose n is a ..." or "Define u to be ..."

- Local variables – variables defined in an enclosing scope that have no meaning outside it

```
x = 17
print x
def foo(u):
    x = u - 10
    y = u - 20
    return x*y

foo(10)
print x, y
```

Variables Example

A bucket contains n balls, b of which are blue and $r = n - b$ of which are red.

We select s balls from the bucket at random. If k of the selected balls are blue, then with probability $p > 1/2$, we receive d_k dollars and otherwise, we receive $f(k^2/2)$ dollars, where

$$d_i = \begin{cases} 2i(i-1) & \text{if } i \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

$$f(u) = ue^{-2u}.$$

Meaning? Strategies? Local Variables? Constraints?