CMPSCI 240: Reasoning about Uncertainty

Lecture 1: Introduction

Andrew McGregor

University of Massachusetts

Last Compiled: January 24, 2017

Introduction	Course Logistics	Sets and Elements

Outline

1 Introduction

2 Course Logistics

3 Sets and Elements

What's this course about...

- Course description: Development of mathematical reasoning skills for problems that involve uncertainty.
 - Uncertainly could arise because we don't have all the facts or because there's inherent randomness.
 - Almost everything we learn is subject to some amount of uncertainty.
- **Examples**: Card games and gambling, spam filters, stock market, collecting coupons, game theory, communication and compression...
- Along the way: Get practice making mathematical arguments and using mathematical concepts.

Introduction	Course Logistics	Sets and Elements
0000000		

Probability

- Probability is the most commonly accepted, well-developed and successful approach for reasoning about uncertainty.
 - Basic idea is to represent the degree of certainty using real numbers between 0 and 1. For example, we could express our belief that it's likely to snow tonight by saying "the probability of snow is 0.9".
- Probability theory provides mathematical rules specifying how probabilities can be manipulated to reason about uncertainty correctly and make good decisions in the face of uncertainty.
 - Our intuition can fail if we don't use these rules carefully.

Medical Testing and Diagnosis



You go to your doctor because you haven't been feeling well. The doctor orders some tests and you get lab results in the mail showing that that you tested positive for a deadly disease that affects 1 in 10000 people. The test correctly identifies positive cases 99% of the time and correctly identifies negative cases 95% of the time. How worried should you be?

Games of Chance and Gambling



You're in Las Vegas for spring break. One afternoon in the casino you watch incredulously as black comes up 10 times in a row on the roulette wheel. You think the streak can't possibly continue so you put \$100 on red. Is this a good bet?

What if you're observing a slot machine that is guaranteed to pay out a certain amount per day but hasn't paid out anything yet?

Planning and Routing



You study all night and wake up 15 minutes before your exam. You could take a 10 minute bus ride but you have to catch a bus first: each minute, a bus to school passes your house with probability 1/5. Your other choice is to bike to the University from home, but it will take you 20 minutes. Should you wait for the bus or hop on your bike?

ntroduction	
0000000000	

Hand Written Digit Recognition

A friend writes their cell number down on a scrap of paper for you. You go to call them later but you're not sure what the second number is. Is the number a 1 or a 7?



ntroduction	
000000000	

Hand Written Digit Recognition

A friend writes their cell number down on a scrap of paper for you. You go to call them later but you're not sure what the second number is. Is the number a 1 or a 7? How about now?

413-545-2744

Probability and Computer Science

Probability can be applied to day-to-day situations, but it also forms a core component of many of the most exciting areas of modern computer science and computational statistics including:

- Natural language processing and speech recognition
- Computer vision and robotics
- Machine learning and planning
- Information retrieval and search
- Information theory and data compression
- Algorithms and Complexity
- Game theory and mechanism design
- Computational biology
- Social network analysis

...

Probability and Industry

A correspondingly wide range of high tech companies actively use probability and reasoning under uncertainty to develop products.



Course Logistics

Outline



2 Course Logistics

3 Sets and Elements

Course Logistics	Sets and Elements
•0000000	

Course Contents

- Part I: Sample Space and Probability
 - 1 Sample space, events, and probability laws
 - 2 Conditional probability, Bayes' theorem, independence
 - 3 Counting (i.e., combinations, permutations, partitions)
- Part II: Random Variables
 - 1 Discrete and continuous random variables
 - 2 Expectation, mean, and variance
 - 3 Correlation and covariance
 - 4 Concentration results and limit theorems
- Part III: Applications: Inference, Markov Chains,
 - Inference and Bayesian Networks: reasoning about unobserved quantities from available evidence, e.g., spam filtering
 - 2 Markov Chains: reasoning about random processes that evolve over time, e.g., Google's PageRank algorithm
- Part IV: More Applications: Game Theory and Information Theory
 Game Theory: how to you randomized strategies when playing games
 Information Theory: how to communicate when there's random noise

General Information

- Instructor: Andrew McGregor (www.cs.umass.edu/~mcgregor)
- Lectures: Tu/Th, 11:30am–12:45pm, Thompson Hall Room 102
- Discussion Sections: Please keep to the section your registered:
 - **1** Section AA: Wed 12:20-1:10pm, Engineering Laboratory 323
 - 2 Section AB: Wed 10:10-11:00am, Tobin Hall 204
 - 3 Section AC: Wed 9:05-9:55am, Engineering Laboratory 323
 - 4 Section AD: Wed 1:25-2:15pm, Engineering Laboratory 306
- Teaching assistants: Raj Maity and Emma Tosch

• Office Hours:

- **1** Raj: Monday 4-5pm (TBA)
- 2 Andrew: Tuesday 3-4pm (CS 334)
- 3 Emma: Thursday 2:30-3:30pm (TBA)

Website and Communication

• Website: The official course webpages are

```
people.cs.umass.edu/~mcgregor/courses/CS240S17/
```

Quizzes and grades will be posted at

moodle.umass.edu

Discussions will happen at piazza.com and homework submissions will be via gradescope.com.

 Outgoing email: Course email to the address Moodle has on file for you. Check it regularly.

Textbooks

- Required textbook: Introduction to Probability, 2nd Edition by Dimitri P. Bertsekas and John N. Tsitsiklis
- Supplemental textbooks:
 - Introduction to Probability by Charles M. Grinstead and J. Laurie Snell. Available free:

www.math.dartmouth.edu/~prob/prob.pdf

 Introduction to Probability, Statistics, and Random Processes by Hossein Pishro-Nik. Available free:

http://www.probabilitycourse.com

Some application topics will require extra reading that will be posted.

Grade Breakdown

- Participation (10%): Discussion assignments, Piazza, and in class contributions.
- Homework (20%): Homework and weekly online quiz.
- Midterm 1 (15%): Focuses on Part I.
- Midterm 2 (15%): Focuses on Part II.
- Midterm 3 (15%): Focuses on Part III.
- Final (25%): Covers entire course.

Submission Policy

- Written Assignments: Assignments must be uploaded to Gradescope by the specified deadline. If you handwrite your solutions and scan them, make sure they are legible; if we can't read your answer, it's hard to give it marks!
- Online Quizzes: Every week, there'll be a online quiz posted on Friday. The quiz should take about 30 minutes but you're allowed two hours. The cut-off for doing the quiz is 8pm Monday.
- Late Policy: Marks for any homework that is late by 5 minutes up to 24 hours, will be halved. However, you are allowed to submit one homework up to 24 hours late without penalty; don't use this lifeline prematurely! Homework later than 24 hours receives no credit. The cut-off for quizzes is strict but we'll drop the lowest quiz mark.

Academic Honesty

- Homework: You may discuss homework with other students but the writeup and code must be your own. There'll be formal action if cheating is suspected. You must list your collaborators and any printed or online sources at the top of each assignment.
- Online Quizzes: Should be done entirely on your own although it's fine to consult the book and slides as you do the quiz. Again, there'll be formal action if cheating is suspected.
- Discussions: Groups for the discussion section exercises will be assigned randomly at the start of each session. You must complete the discussion session exercise with your assigned group.
- **Exams:** Closed book. Cheating will result in an F in the course.
- If it doubt whether something is allowed, ask...

How to be Successful in this Class

This class is not easy! It's a very mathematical course but you also have to develop intuition for why the math makes sense in the context of real-world problems. We'll try to help as much as possible but to succeed:

- **1** Come to class and your discussion section.
- 2 If there are readings, complete them *before* class.
- **3** Stay on top of the assignments and submit them on time.
- 4 Check the course website and the email address Moodle has for you frequently for news and updates.
- **5** Get help early and often. Ask questions in class, come to office hours, post to the discussion forums or send email.

Outline

1 Introduction

2 Course Logistics

3 Sets and Elements

Sets and Elements

- A set S is a container for a collection of objects called **elements**.

• |S| denotes the size of a set, the number of elements it contains. • Ex: $|S| = \left| \left\{ \underbrace{0}, \underbrace{0}, \underbrace{0}, \underbrace{0} \right\} \right| = 3$

Sets Membership

The notation $x \in S$ asserts that object x is contained in set S. We say x is an element of S.

• Ex: If
$$S = \left\{ \bigcup, \bigcup, \bigcup, \right\}$$
, then $\bigcup \in S$.

■ The notation x ∉ S asserts that object x is not contained in set S. We say x is not an element of set S.

• Ex: If
$$S = \left\{ \bigcirc, \bigcirc, \bigcirc, \oslash \right\}$$
, then $\not \in S$.

Equality of Sets

■ The notation S = T means that set S is equal to T. Two sets are equal if every element of S is a member of T and every element of T is a member of S. Otherwise, S ≠ T.



Introduction 00000000	Course Logistics 00000000	Sets and Elements

Subsets

• The assertion $\mathcal{T} \subseteq S$ means \mathcal{T} is a subset of S. \mathcal{T} is a subset of S if all the elements in \mathcal{T} are also in S.



Course Logistics	Sets and Elements
	000000000000000000000000000000000000000
	Course Logistics 00000000

Strict Subsets

• The assertion $\mathcal{T} \subset \mathcal{S}$ means \mathcal{T} is a strict subset of \mathcal{S} . \mathcal{T} is a strict subset of \mathcal{S} if all the elements in \mathcal{T} are also in \mathcal{S} , but $\mathcal{S} \neq \mathcal{T}$.



Creating Subsets By Filtering

• Let F(x) be a logical formula that takes objects x and returns a value in the set $\{true, false\}$. We indicate the set of objects such that F(x) is *true* using the notation $\{x|F(x)\}$.

• Let
$$S = \left\{ \bigotimes_{x \in S} , \bigotimes_{x \in S} , \bigotimes_{x \in S} \right\}$$

Then $\{x | x \in S \text{ and } x \text{ is green}\} = \left\{ \bigotimes_{x \in S} , \bigotimes_{x \in S} \right\}$
• Let $S = \left\{ \bullet, \bullet, \bullet, \bullet, \bullet, \bullet, \bullet, \bullet, \bullet, \bullet \right\}$
Then $\{x | x \in S \text{ and } x \text{ is even}\} = \left\{ \bullet, \bullet, \bullet, \bullet, \bullet \right\}$

The Universal Set

- The universal set Ω is the set of all possible objects of interest in a given context.
- **Ex:** Coin flip:



Ex: Dice roll:
$$\left\{ \bullet, \bullet, \bullet, \bullet, \bullet, \bullet, \bullet, \bullet, \bullet \right\}$$

Course Logistics	Sets and Elements
	000000000000000000000000000000000000000

The Power Set

■ The power set of a set S is denoted 2^S or P(S). It is the set of all subsets of S. For finite sets, |2^S| = 2^{|S|}.



	Sets and Elements
	0000000000000000

Set Complement

The notation S^c indicates the complement of set S with respect to the universal set Ω. S^c contains all elements of Ω that are not in S:

$$S^c = \{x \,|\, x \in \Omega \text{ and } x \notin S\}$$
.

• Let:
$$\Omega = \left\{ \bigcirc, \bigcirc, \bigcirc, \oslash, \circ \bigcirc, \circ \bigcirc, \circ \bigcirc, \circ \bigcirc \right\}$$

Let: $S = \left\{ \bigcirc, \bigcirc, \circ \bigcirc, \circ \bigcirc \right\}$
Then: $S^c = \left\{ \bigcirc, \bigcirc, \circ \bigcirc, \circ \bigcirc \right\}$

The notation $S \cup T$ indicates a set containing the union of the elements in set S and the elements in set T:

$$\mathcal{S} \cup \mathcal{T} = \{x \, | \, x \in \mathcal{S} \text{ or } x \in \mathcal{T}\}$$
 .

• Let:
$$S = \left\{ \bigotimes, \bigotimes \right\}$$
 and $T = \left\{ \bigotimes, \bigotimes \right\}$
Then: $S \cup T = \left\{ \bigotimes, \bigotimes, \bigotimes, \bigotimes, \bigotimes \right\}$

	Sets and Elements
	0000000000000000
Set Intersection	

■ The notation S ∩ T indicates a set containing the intersection of the elements contained in S and the elements contained in set T:

$$\mathcal{S} \cap \mathcal{T} = \{x \, | \, x \in \mathcal{S} \text{ and } x \in \mathcal{T}\}$$

• Let:
$$S = \left\{ \bigcirc, \bigcirc \right\}$$
 and $T = \left\{ \bigcirc, \oslash \right\}$
Then: $S \cap T = \left\{ \bigcirc \right\}$

Introduction 000000000	Course Logistics 00000000	Sets and Elements

Disjoint Sets

Sets $S_1, ..., S_N$ are mutually disjoint if for every i, j with $i \neq j$:

 $\mathcal{S}_i \cap \mathcal{S}_j = \emptyset$

Let:

$$S_1 = \left\{ \bigcirc, \bigcirc \right\} , S_2 = \left\{ \bigcirc, \bigcirc \right\} , S_3 = \left\{ \bigcirc, \bigcirc \right\} \right\}$$

Then: $\mathcal{S}_1 \cap \mathcal{S}_2 = \{\}, \ \mathcal{S}_1 \cap \mathcal{S}_3 = \{\}, \ \mathcal{S}_2 \cap \mathcal{S}_3 = \{\}$

So: S_1, S_2, S_3 are mutually disjoint.

Course Logistics	Sets and Elements
	000000000000000000000000000000000000000

Partitions

Sets $S_1, ..., S_N$ are a partition of set S if $S_1, ..., S_N$ are mutually disjoint and $S_1 \cup \cdots \cup S_N = S$.

Let:
$$S = \left\{ \underbrace{\circ}, \underbrace{\circ}$$

Then: $S_1 \cap S_2 = \{\}$, $S_1 \cap S_3 = \{\}$, $S_2 \cap S_3 = \{\}$ And: $S_1 \cup S_2 \cup S_3 = S$ So: S_1, S_2, S_3 are a partition of S.

The Algebra of Sets

Using the above definitions, we can show that:

Intersection Commutativity Union Commutativity Intersection Associativity Union Associativity Intersection Distributivity Union Distributivity

 $S \cap T = T \cap S$ $S \cup T = T \cup S$ $S \cap (T \cap U) = (S \cap T) \cap U$ $S \cup (T \cup U) = (S \cup T) \cup U$ $S \cap (T \cup U) = (S \cap T) \cup (S \cap U)$ $S \cup (T \cap U) = (S \cup T) \cap (S \cup U)$

Course Logistics

Key Notation

Set	$\{a, b, c\}$
Set Size	$ \mathcal{S} $
Set Membership	$x \in S, x \notin S$
Empty Set	{}, Ø
Set Equality	$\mathcal{T}=\mathcal{S}$
Subset	$\mathcal{T}\subseteq\mathcal{S}$
Strict Subset	$\mathcal{T}\subset\mathcal{S}$
Set Filtering	$\{x F(x)\}$
Universal Set	Ω
Set Compliment	Sc
Set Union	$\mathcal{T}\cup\mathcal{S}$
Set Intersection	$\mathcal{T}\cap\mathcal{S}$