Course Staff

- **Instructor:** Prof. Erik Sudderth
  *Research Interests: Statistical machine learning, computer vision, AI, …*

- **Graduate TA:** Zhile Ren
  *Brown CS PhD student. Will lead course recitations.*

- **Head Undergraduate TA:** Nick Catoni

- **Undergraduate TAs:** Uthsav Chitra, John Joe Friedmann, Sarah Grace, Justin Semonsen, Quynh Tran, Keshav Vemuri
Why Probability?

Most advanced computer applications today involve randomization.

- Secured web connections are *probabilistically secured*
- Web search engines need *statistical inference* to determine which pages are most relevant to ambiguous queries
- Computer games would be boring without *randomization*
- Spam filters, recommendation systems, and web advertising are constructed via *(statistical) machine learning*
- Efficient data structures are often *randomized* (e.g., hashing)
- Wireless communications: split bandwidth via *random codes*
- Computational finance: Must model market *uncertainty*
- Computational biology: DNA sequencing, clinical trials, …
- Robotics: navigating new environments, human interaction, …
Overview of Course Topics

I. Probability Models
II. Discrete Random Variables
III. Continuous Random Variables
IV. Normal Distributions
V. Limit Theorems
VI. Markov Chains
VII. Monte Carlo Methods
VIII. Bayesian Statistical Inference
IX. Frequentist Statistical Inference
**Sample spaces:** When a random event happens, what is the set of all possible outcomes? May be discrete or continuous.

**Conditioning:** Suppose I observe some data. How does my probability model change?

**Independence:** Is there any relationship between pairs of variables in my model? Would data provide knowledge?
Suppose I toss a coin 10 times.
The number of tosses that come up heads, rather than tails, is an example of a *discrete random variable*.

A *probability mass function* gives the (non-negative) probability of each possible outcome. These probabilities sum to one.
Joint, Marginal, & Conditional Distributions

**Joint Distribution:** Probability of each possible outcome.

**Marginal Distribution:** If some variables are not observed and not relevant, how do I remove them from the model?

**Conditional Distribution:** What if I observe some data?
Continuous Random Variables

Model processes or data which are encoded as real numbers: *temperature*, *commodity price*, *DNA expression level*, *light on camera sensor*, …
Summaries: Mean, median, mode, variance, standard deviation, …
Central Limit Theorem

- In a large population, how likely is a person to be much taller than average?
- How likely is a request on my web server to be much larger than average?

Monte Carlo Methods

Hurricane Sandy made landfall in New Jersey on October 29, 2012.

Weather Wisdom, Boston.com
Markov Chains

Markov Property: Conditioned on the present, past & future are independent

- Building block for modeling random processes that evolve and change over time.
- What is the long-term behavior of some process? What is the probability of reaching a good state? A bad state?
- Allows agents to reason about future consequences of actions.

\[ p(x) = p(x_1)p(x_2 | x_1)p(x_3 | x_2)p(x_4 | x_3) \cdots = p(x_1) \prod_{t=2}^{T} p(x_t | x_{t-1}) \]
The approach to solve the SLAM problem is addressed in this framework. The goal is to estimate the robot state at current time: $t$ and to build up a map within an unknown environment while at the same time preserving the robot's ability to know its environment.

$p(x_t, m \mid z_{1:t}, u_{1:t})$

- $x_t =$ State of the robot at time $t$
- $m =$ Map of the environment
- $z_{1:t} =$ Sensor inputs from time 1 to $t$
- $u_{1:t} =$ Control inputs from time 1 to $t$

**Raw odometry (controls)**

**True trajectory (GPS)**

**Inferred trajectory & landmarks**
Markov Chains for Web Search: PageRank

Wikipedia
From Probability to Statistics

- In probability theory we compute the probability that 20 independent flips of a fair (unbiased) coin give the sequence $HTTHTHTHTTHHTHTHHTTT$

- In statistics we ask: Given that we observed the sequence $HTTHTHTHTTHHTHTHHTTT$
  
  what is the likelihood that the coin is fair (unbiased)?
The Frequentist Model: The probability of an outcome in a trial is the frequency of that outcome in a long sequence of identical and independent such trials (limiting frequency).
From Probability to Statistics

The Frequentist Model: The probability of an outcome in a trial is the frequency of that outcome in a long sequence of identical and independent such trials (limiting frequency).

The Bayesian Belief Model: Based on all the information we have seen so far, the probability is our best estimate for the chance of a particular outcome.

*If a rational person is willing to bet 2 to 1 that New England will defeat Seattle in the Super Bowl, then she estimates the probability of that event to be at least 2/3.*
Course Syllabus

Probability Models  sample spaces, axioms of probability, sets and counting, conditioning, Bayes’ rule, independence

Discrete Random Variables  probability mass functions, expectation and moments, conditioning and independence, functions of random variables, conditional expectation

Continuous Random Variables  probability density functions and cumulative distributions, expectation and moments, conditioning and independence, functions and derived distributions

Normal Distributions  properties, covariance and correlation, bivariate distributions

Limit Theorems  Markov and Chebyshev inequalities, weak law of large numbers, convergence in probability, central limit theorem

Discrete-time Markov Chains  classification of states, steady-state behavior and equilibrium distributions, absorption probabilities

Monte Carlo Methods  pseudo-random number generation, Monte Carlo integration, Markov chain Monte Carlo (MCMC), Metropolis algorithm

Bayesian Statistical Inference  posterior distributions, hypothesis testing, maximum a posteriori (MAP) estimation, least mean squares estimation

Frequentist Statistical Inference  maximum likelihood parameter estimation, Neyman-Pearson hypothesis testing, significance tests, linear regression
Administrative Information
Course Prerequisites

Not formally enforced, but we will assume comfort with:

Calculus

- Two semesters of college-level calculus
- AP Calculus BC exam, or Brown MATH 0100/0170
- Topics: limits, basic derivatives & chain rule, basic integrals & fundamental theorem of calculus, sequences & series, …

Programming
Example: Buffon’s Needle

A surface is ruled with parallel lines, which are at distance \( d \) from each other. Suppose that we throw a needle of length \( l < d \) on the surface at random. What is the probability that the needle will intersect one of the lines?

\[
f_{X,\Theta}(x, \theta) = \begin{cases} 
  \frac{4}{\pi d} & \text{if } x \in [0, d/2] \text{ and } \theta \in [0, \pi/2], \\
  0 & \text{otherwise.}
\end{cases}
\]

\[
P(X \leq (l/2) \sin \Theta) = \int_{x \leq (l/2) \sin \Theta} \int f_{X,\Theta}(x, \theta) \, dx \, d\theta
\]

\[
= \frac{4}{\pi d} \int_0^{\pi/2} \int_0^{(l/2) \sin \theta} dx \, d\theta
\]

\[
= \frac{4}{\pi d} \int_0^{\pi/2} \frac{l}{2} \sin \theta \, d\theta
\]

\[
= \frac{2l}{\pi d} \left[-\cos \theta\right]_0^{\pi/2}
\]

\[
= \frac{2l}{\pi d}.
\]
Course Prerequisites

Not formally enforced, but we will assume comfort with:

Calculus

- Two semesters of college-level calculus
- AP Calculus BC exam, or Brown MATH 0100/0170
- Topics: limits, basic derivatives & chain rule, basic integrals & fundamental theorem of calculus, sequences & series, …

Programming

- Any single-semester programming course: CS4, CS15, CS17, CS19, etc.
- Or, other experience that gives comfort with writing simple functions
- All support code will be provided in Matlab, and we strongly recommend that you use Matlab (easy to learn if you’ve seen Java, Python, etc.)
- Topics: functions, loops & flow control, arrays, …
Course Textbook

- Supplemental readings (online) for a few advanced topics
Course Instruction

Lectures
- Tuesdays & Thursdays from 2:30-3:50pm. CIT 368 (for now).

Recitations
- An *optional* review and problem-solving session led by graduate TA.
- Time and location TBD. Likely Tuesday or Wednesday at 5:30pm.
- Begins week of February 2nd with a *Matlab tutorial*.

Office Hours
- Prof. Sudderth: Tuesday & Wednesday, 4:00-5:00pm, CIT 509. 
  *Begins on Wednesday, January 28.*
- Graduate TA and UTAs will hold hours on Sunday-Thursday evenings. 
  *Begins on February 1, after homework 1 is distributed.*
Course Evaluation

Homeworks: 50%
- Probabilistic derivations, calculations, and reasoning (i.e., math)
- Usually one Matlab problem: Monte Carlo, statistical data analysis, etc.
- Ten equally weighted assignments (we will not drop any scores)
- Submitted electronically by Thursdays at midnight, out for one week

Midterm Exam: 20%
- During normal lecture period on Thursday, March 12.
- No rescheduling except in extraordinary, unexpected circumstances.

Final Exam: 30%
- Scheduled by registrar on Wednesday, May 6 from 9:00am-12:00pm.
- Final lecture on Thursday, April 30.
Course Enrollment

Why does registration require an override?

- There were 69 students pre-registered for CS145.
- CIT368 holds 65 students with wall closed (distractions when open).
- Some students have dropped, but others want to add.
- Can we get a different, larger lecture hall? Possibly, but no guarantee.

How can I register?

- We will distribute an online survey early next week, to gauge the total amount of student interest in enrolling in CIT368.
- Based on this survey and the Banner registration at that point, we will explore options to let most or all interested students register.
- Please, do not individually e-mail Prof. Sudderth for an override.
  For fairness, we will handle registrations of all interested students jointly.
Questions?