ECE 313 – Section G (R. Iyer)

- **Meeting time and Place:**
  - 11:00am – 12.20pm, Mondays and Wednesdays
  - 0216 Siebel Center

- **Instructor:**
  - Professor Ravi K. Iyer
  - Office: 255 Coordinated Science Lab (Phone: 333-9732)
  - Email: rkiyer@illinois.edu
  - Office Hours: 12:30pm – 2:00pm, Mondays and Wednesdays
- **Teaching Assistants:**

Phuong Cao **Office: 249 Coordinated Science Lab (CSL)** Alternately, Phuong is at 246 CSL **Email: pcao3@illinois.edu** **Office Hours:** **4:00pm – 5:00pm Tuesdays** **5:00pm – 6:00pm Thursdays**	**Krishnakant Saboo** **Office: 249 Coordinated Science Lab** **Email: ksaboo2@illinois.edu** **Office Hours: 4:00pm – 5:00pm, Mondays and Wednesdays**

More about Section G

- **Text Book:**
  - Sheldon Ross, *A First Course in Probability*, 9th edition, Pearson, 2012. a pdf downloadable free edition is available
- **Further Reading and Sample Problems:**
  - Sheldon M. Ross, *Introduction to Probability Models*, 10th Edition, Academic Press, 2010 (Chapters 1-5).

- **Class Website:**
  - https://courses.engr.illinois.edu/ece313/sp2017/sectionG
  - Please check the class web site for all announcements regularly.
  - A detailed class schedule including topics covered and reading for each class is on the class website
  - Lecture notes will be posted on the class web site weekly.
  - Ask questions on Piazza page for the course.

Grading Policies

- **Homeworks will be handed out or posted on the class website weekly on Wednesdays and will be due the following Wednesdays before the class.**
  - **No late homework will be accepted.**
  - **No Make up for In Class Activity**
  - **Lowest scores dropped for final grades**

- **There will be two mini-projects and a final project to provide hands-on experience with applications of probability in analysis of computer systems.**
  - **Project descriptions and due dates will be posted on the class web site under Student Projects.**
  - **Please check the Resources for tutorials and hints related to projects.**

- **Homework: 15%**

  In-Class Activity and class participation: 15%

  2 Mini Projects + Final Project: 20% (2 * 5% + 10%)

  Midterm Exam: 20%

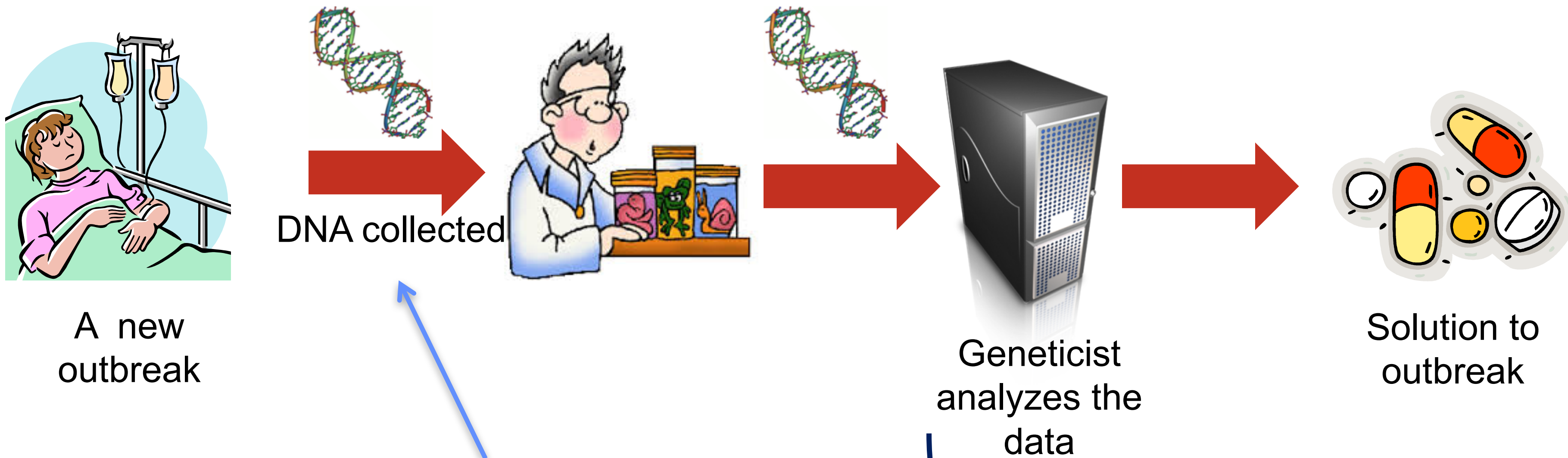
  Final: 30%

Homework 0

- Write a summary (no more than ¾ of a page) of your experience with an application of probability to a real-life situation (e.g., an engineering problem.
  - How was probability used to model the phenomena/situation?
  - How was it measured?
  - Did you perceive any useful outcomes or interesting insights?

- Be concise and use readable English. Use a font size of 11pt, 1.5 line spacing and margins of 1 inch. Otherwise your homework will be returned without a grade.
- Turn in your single page write-up at the beginning of the class on **Wednesday, Jan 25th**

Computational Genomics



Sequencing Technologies: Faster and Cheaper

Biologist's Desires: Performance & Accuracy

- **~ days – weeks**
- **~ 100's of Gb data**
- **~ Varying degrees of accuracy**

Computational Genomics: Basis for Personalized Medicene

Future of Personalized medicine:

Compare known and patient DNA sequences to detect mutations

Known Reference DNA Sequence: A A C T G A C

Patient DNA Sequence: A A ? T G A C

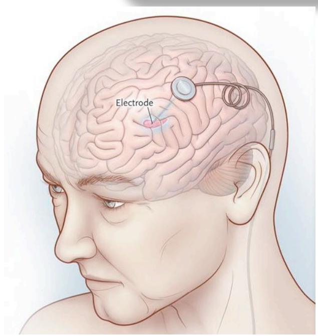
Due to error artifacts in biology tools, certain DNA bases (molecules) cannot be correctly determined

Probability theory is used to infer most-likely bases (A,C,T or G) of DNA sequences

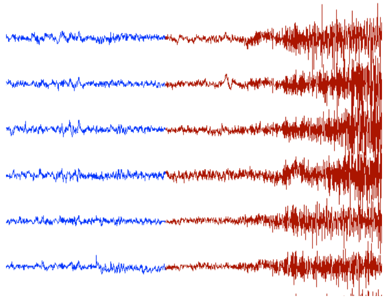
Probability (position 3 = C, given patient DNA and reference DNA)

Why do we need this? If the ? is A or T or G, likely the patient has a certain disorder, else he/she is may be fine

Computational Neuroscience



Patient monitoring using implanted sensors



Data collection

Deep analytics

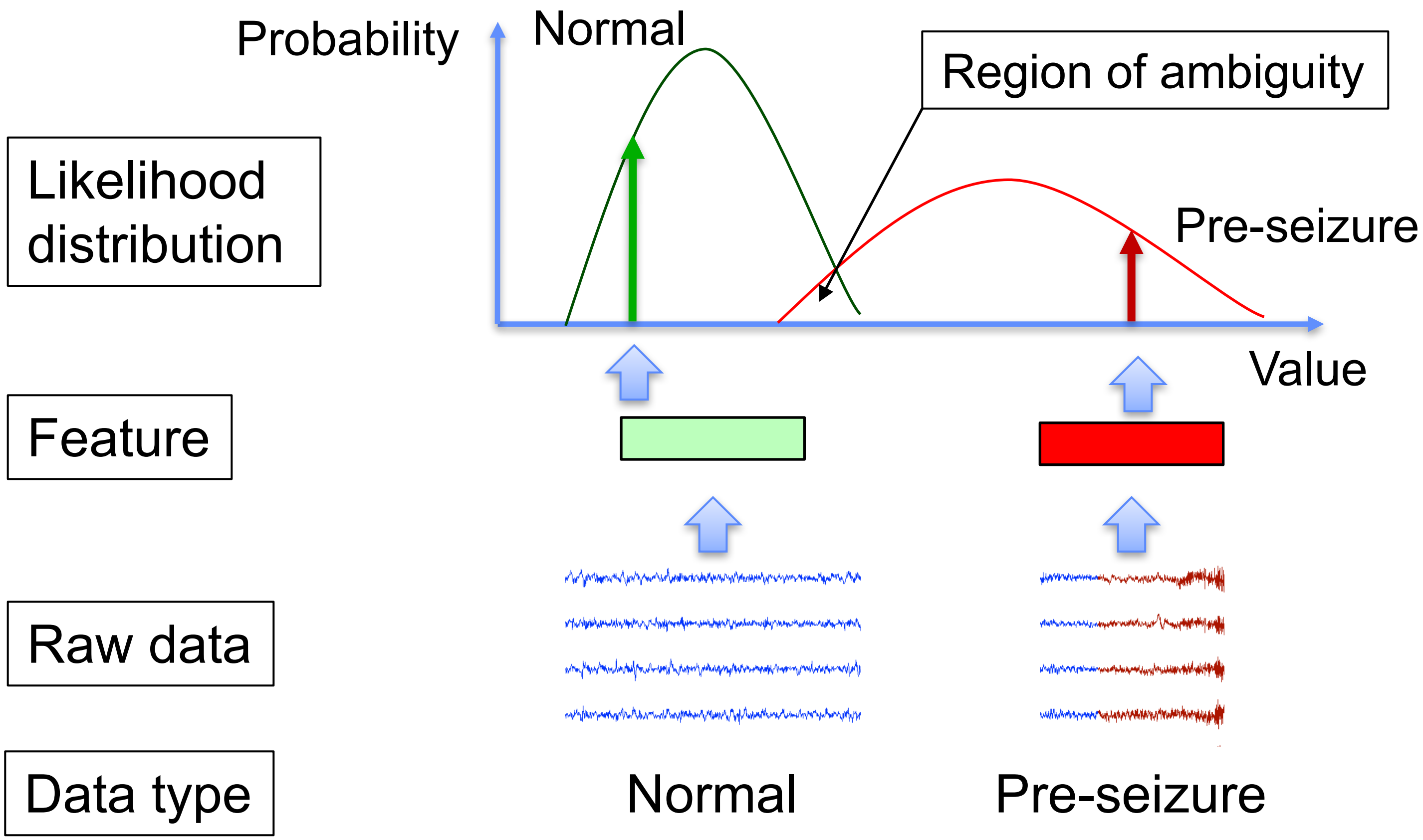
New insights and hypotheses

Expert Knowledge

Individualized patient care

Probability on EEG data

Goal: Being able to predict seizures ahead of time



Probability Theory Basics:

- Sample spaces and events
- Probability axioms
- Combinatorial problems

Conditional Probability & Independence of Events:

- Bayes rule
- Theorem of total probability
- Bayes formula
- Independence vs. mutual exclusivity

Random Variables:

- Discrete and continuous
- Cumulative distribution function (CDF)
- Probability mass function (PMF)
- Probability density function (PDF)

Discrete Random Variables:

- Bernoulli and binomial
- Geometric and modified geometric
- Poisson

Continuous Random Variables:

- Uniform
- Gaussian (normal)
- Exponential

Exponential Distributions:

- Memory-less property
- Relation to Poisson
- Phase-type exponentials

Expectations

Independence of Random variables

Functions of RVs

Joint and conditional densities

Binary hypothesis testing

- Maximum likelihood
- Maximum A-Posteriori

Limit Theorems:

- Law of large numbers
- Central limit theorem

Inequalities: Markov and Chebyshev

Example Applications we will come Across

- Reliability assessment
- System evaluation
- Software design
- Failure data analysis
- Genomics computing
- Health monitoring
- Security monitoring …..

Course Outline

I. Probability Theory Basics

- Probability theory, models and their uses, examples
- Definitions: sample space, elements, events
- Algebra of events (union, intersections, laws/axioms)
- Probability axioms and other useful relationships
- Basic procedure for problem solving and an example
- Combinatorial problems
- Introduction to measurements

- Mini Project 1:Heath Monitoring - Probability Basics, Conditional Probability

Course Outline

II. Conditional Probability and Independence of Events

- Definitions of conditional problems, multiplication rule
- Theorem of total probability and Bayes' Formula
- Independent events and associated rules
- Bernoulli Trials
- Reliability evaluation applications:
  - Multi-version programming
  - Series systems
  - Parallel redundancy
  - Series-parallel system evaluation
  - Non-series-parallel systems
  - Triple Modular Redundant (TMR) system with voter

Course Outline

III. Random Variables (Discrete)

- Introduction to random variables
- Cumulative distribution function (CDF)
- Probability mass function (PMF)
- Important discrete random variables and their distributions:
  - Bernoulli and Binomial
  - Geometric and modified geometric
  - Poisson

IV. Random Variables (Continuous)

- Probability density function (PDF)
- Important continuous random variables and their distributions:
  - Uniform
  - Gaussian (Normal)
  - Exponential

- Mini Project 2

Course Outline

V. Exponential Distribution

- Memory-less property
- Relationship to Poisson distribution and examples
- Phase-type exponential distributions:
  - Hypo-exponentials
  - K-stage Erlang
  - Gamma
  - Hyper-exponential
- Applications to reliability evaluation

Course Outline

VI. Expectations

- Moments: Mean and variance
- Mean and variance of important random variables
- Conditional expectation
- Expectation of function of random variables
- Covariance and correlation
- Reliability evaluation applications:
  - Mean time to failure
  - Failure rates
  - Hazard function
  - Failure Data Analysis

- **Final Project**

Course Outline

VII. Joint and Conditional Density Functions

- Joint CDFs and PDFs
- Jointly Gaussian random variables
- Functions of many random variables
- Independent random variables

VIII. Binary Hypothesis Testing

IX. Inequalities and Limit Theorems

- Markov Inequality
- Chebyshev's inequality
- Law of large numbers
- The Central Limit Theorem

Introduction: Some Basic Concepts

ECE 313

Probability with Engineering Applications

Lecture 1

Professor Ravi K. Iyer

University of Illinois

Motivation

- Examples of the need for probability and statistics in computer engineering
  - Tools to analyze computer systems and the algorithms that execute on them (worst case vs. average behavior)
    - distribution of inputs
    - arrival patterns of tasks
    - computer resources (CPU, memory, I/O, network)
    - distributions of resource requirements of jobs
    - effect of hardware/software failures (random phenomena, environmental effects, software failures, aging)
  - Applications in healthcare, transportation systems, smart power grid
- Probability is important for evaluating the system design and its performance using measures like:
  - Throughput, response time, reliability, availability

Motivation (cont.)

- Popular evaluation techniques
  - performance benchmarks - (SPEC, LINPACK
  - parallel benchmarks (NAS, Livermore Loops, PARKBENCH)
  - Possible benchmarks for Security, Resilience, and Fault Tolerance
- Experimental analysis
  - Need to specify various probability distributions
  - where do these distributions come from?
  - collect data during actual operations
  - measurements using Hardware and Software monitors
  - Analyze and interpret this data to estimate relevant parameters/distributions
  - Mathematical Statistics ==> provides the tool

Introduction

- Probability theory studies phenomena for which the future behavior cannot be predicted in a deterministic fashion – ***Random Phenomena***
- A "Probabilistic Model" is used to abstract the real-word problems or situations
  - the model consists of a list of possible inputs and their outcomes and an assignment of their respective probabilities
  - the model is only as good as the underlying assumptions (probability assignments and distributions)
  - the model must be validated against actual measurements on the real phenomena

Introduction
Example 1

Example 1:

Predicting the number of job arrivals at a computer center in a fixed time interval (0, t)

- a common model is to assume that the number of job arrivals in this period has a particular distribution, such as Poisson distribution
- a complex physical situation is represented by a simple model with a single unknown parameter - the average job arrival rate λ

Introduction
Example 2

Example 2:

- Consider a computer system with automatic error recovery capability.
- Model: probability of successful recovery is c (Coverage) and probability of an abortive error recovery is unsuccessful) is (1- c).
- Estimate parameter c in this model: we observe N errors out of which n are successfully recovered.

Introduction
Example 2 (cont.)

- A reasonable estimate of c ==> “Relative Frequency” n/N. we expect this ratio to converge to c in the limit $N \rightarrow \infty$.
- Note that this limit is a limit in a probabilistic sense, i.e.,

$$\lim_{N \rightarrow \infty} P\left(\left| \frac{n}{N} - c \right| > \varepsilon\right) = 0$$

Introduction
Example 2 (cont.)

- This is different from the usual mathematical limit: $\lim_{N \to \infty} \frac{n}{N}$
- Thus, given any small $\varepsilon > 0$, it is not possible to find a value M such that (as would be required for the mathematical limit).

$$\left| \frac{n}{N} - c \right| < \varepsilon \qquad for\ all\ N > M$$

Detecting Malicious Internet-of-Things (IoT) devices

In Oct 2016, Cloudflare was hit by Application-layer DDoS attacks, which crossed 1 million HTTP requests per second (link)

52,467 unique IP addresses participated in the attacks. Many of them are linked to IoT devices such as surveillance cameras with weak authentications.

Using default passwords, attackers logged as admin and send malicious requests, which have a larger HTTP request size compared to benign requests.

We calculate the probabilities of an IP address being malicious or benign given the observation on the HTTP requests:

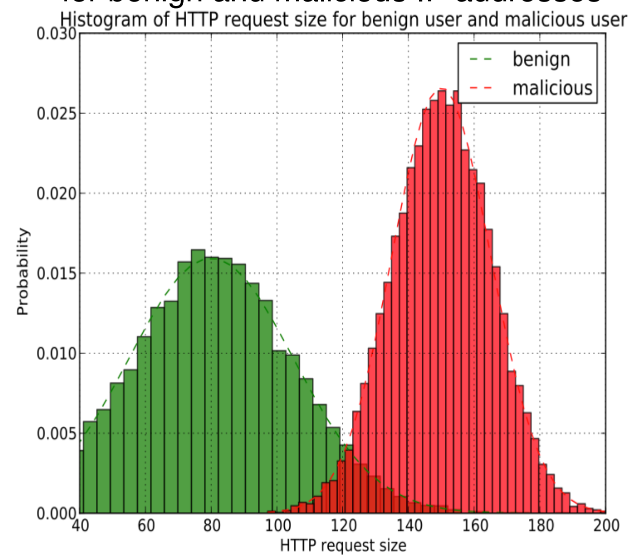
P(IP = malicious, given the HTTP request)

For example using past measurements from the System, we have:

P(IP = malicious | request size > 150 bytes) = 90%

P(IP = benign | request size > 150 bytes) = 10%

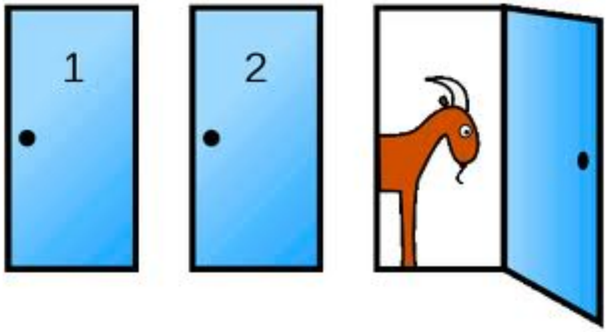
Histogram of request size (in bytes) for benign and malicious IP addresses



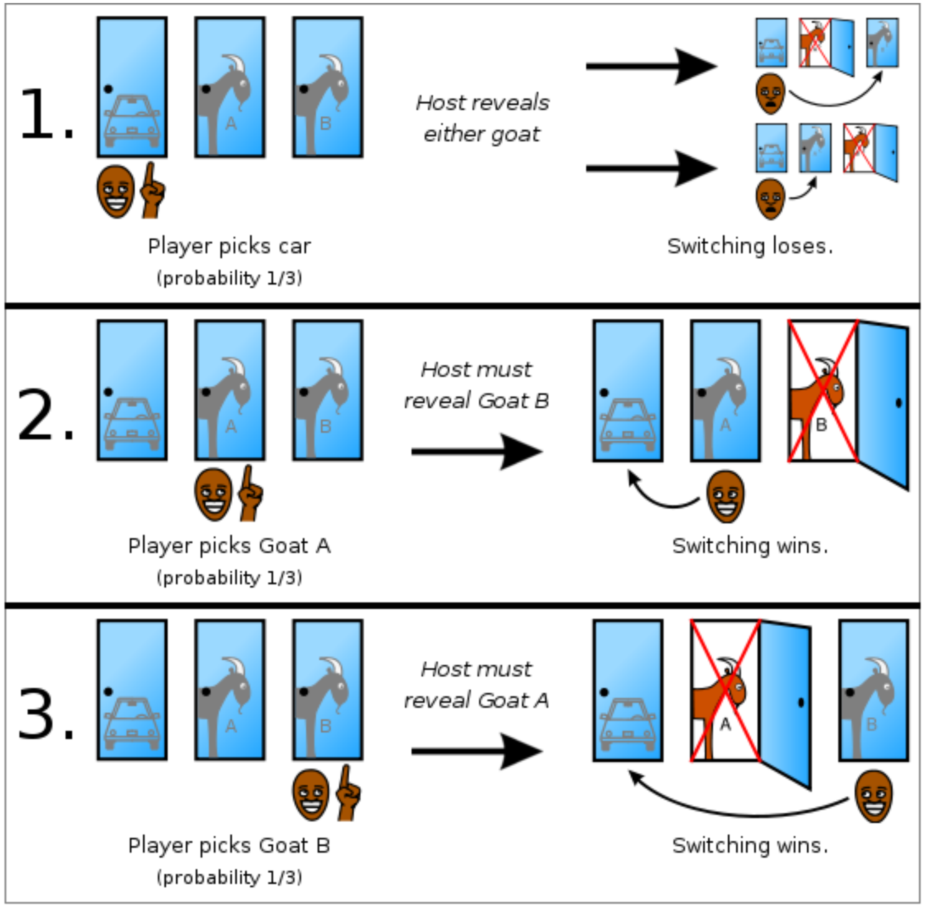
Monty Hall Problem

- The "Let's Make a Deal" game.
- Popular game show in the 1970's.

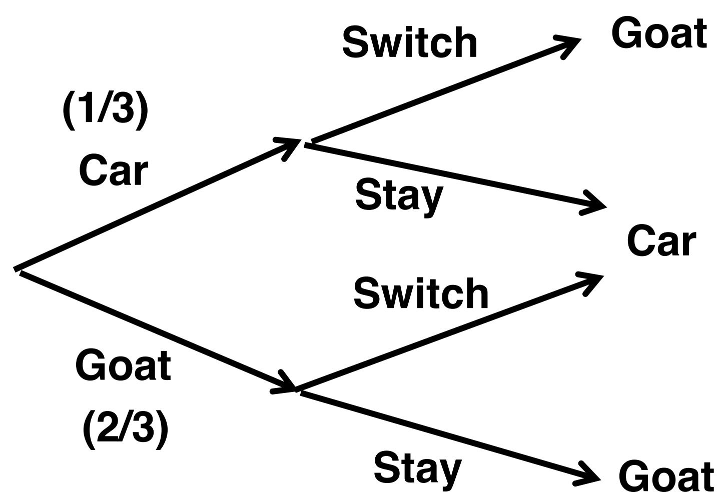
- Suppose you're given the choice of three doors: Behind one door is a car; behind the others, goats. You pick a door, say No. 1, and the host, who knows what's behind the doors, opens another door, say No. 3, which has a goat. He then says to you, "Do you want to pick door No. 2?" Is it to your advantage to switch your choice?



Monty Hall Problem



Probability Tree Diagram



More Reading at: http://en.wikipedia.org/wiki/Monty_Hall_problem