EENG 311
Information Systems Science II

Lecture Notes, Set A: Course Introduction and Syllabus

Prof. Mike Wakin
Colorado School of Mines

Fall 2016

Overview

This course serves as Part II of the Information Systems Science ("Info Systems") course sequence for EE undergraduates. In Info Systems I (EENG 310 or EENG 388) you learned that **signals** come in a variety of forms:

- voltage, temperature, humidity, pressure recordings
- gas prices, stock prices
- sounds (air pressure)
- images (light intensity)
- electromagnetic radio/TV signals

You also learned to view **systems** as devices that process, or operate on, signals. Examples of systems include:

- control systems
- communication systems, televisions, smart phones
- digital computers
- sensors and transducers

Overview - 2

Info Systems I was dedicated to building up a common mathematical framework for describing signals and systems. This framework involved a collection of transforms

- continuous- and discrete-time Fourier transforms
- z-transform
- Laplace transform (EENG 307)

These transforms are convenient for

- understanding the information carried by a signal
- understanding how a system will interact with a signal (e.g., "filtering")
- designing systems for effective transmission and processing of signals
- etc.

Overview - 3

Our treatment of signals in Info Systems I was entirely **deterministic**. That is, we ignored the important roles that **randomness** can play in information systems:

1. **Many real-world signals are corrupted by random noise.** Sources of noise include thermal noise in electrical circuits, interference from other electronic devices, atmospheric noise (such as lightning), cosmic background noise, etc. Keeping in mind that many types of (non-electrical) data can also be viewed as signals, other sources of noise include errors in data collection (such as in the U.S. census), typographical errors, missing data, etc.
2. **Some real-world signals are themselves randomly generated.** Examples include lottery numbers, counts of defective circuits from an assembly line, meteorological phenomena, incidences of disease, etc. And whether "truly" random or not, some phenomena—such as stock prices—are so difficult to predict that they are often modeled as having some degree of randomness.

Overview: Probability

The study of **probability** provides a framework for dealing with random phenomena.

Sometimes, probability is used to describe the fraction of time a given event might be expected to occur:

- "The probability is 50% that a fair coin flip will land on heads."
- "If two parents carry the gene for cystic fibrosis, the probability is 25% that their child will have cystic fibrosis."

These statements can be interpreted in what is called a **frequentist** sense: if one flips a large number of fair coins, it is indeed true that about 50% of them will land heads.

Overview: Probability - 2

Probability is also used to describe, more subjectively, the level of confidence one has that an event will occur:

- "There is a 21% chance that the Cubs will win the World Series this year."
- "I have a 90% chance of passing this test."

In contrast to the frequentist perspective, this degree-of-belief interpretation of probability is known as the **Bayesian** perspective. Both perspectives are useful!

Overall, probability and randomness are very difficult concepts to define precisely, and philosophical debates remain about exactly what these terms mean. However, it is not necessary to settle these debates in order to use probability as a very handy tool. Indeed, very sophisticated uses of probability underlie many modern technologies, including insurance (actuarial work), Google's web search, artificial intelligence (speech recognition, self-driving cars), and satellite communications.

Overview: Statistics

Whereas probability concerns the randomness *in* data, the study of **statistics** concerns the analysis *of* random data.

- "Given noisy temperature readings from 5 thermometers, what is an estimate of the current room temperature?"
- "Based on a survey of 1000 likely voters, what is the probability that Candidate A will beat Candidate B in an election?"
- "Given that 23 out of 50 patients on a medication showed signs of improvement, but 18 out of 50 patients in a control group showed signs of improvement, is the medication effective?"
- "Who is better 3rd Down quarterback: Tom Brady or Andrew Luck?"

Overview: Statistics - 2

Statistical analysis requires careful processing of the random data, and a proper appreciation of how the randomness of the data affects the reliability of the final conclusions.

A statistician named Nate Silver has become well known for his ability to forecast elections based on polling data. Silver is editor-in-chief of an ESPN-owned website `http://fivethirtyeight.com` which focuses on "data journalism" especially in relation to sports, politics, and popular culture. Silver also wrote a book describing principled approaches for predicting events (such as earthquakes or financial outcomes) based on data, and at the center of this is Bayes' rule (which we will discuss). The title of his book? "The Signal and the Noise."

Overview: In this course

In this class, we will discuss several basic topics in **probability** which are broadly useful, even outside the realm of Info Systems. However, we will also discuss **probabilistic models for signals and noise**, including the ubiquitous "white Gaussian noise" model. We will also touch on the field of **Information Theory**, which quantifies how much information a signal contains based on the randomness of that signal. Information Theory has interesting connections with **data compression**, and we will talk about how to construct an efficient binary compression strategy for random signals.

We will also briefly discuss some basic topics in **statistics** which are broadly useful outside the realm of Info Systems. However, we will mostly focus on **statistical signal processing**, which is the application of statistics in analyzing noisy signals. Statistical signal processing plays an important role in radar and communication systems, where it is important to determine precisely what signal was received (and when), but this decision must be made in the presence of noise. We will discuss **matched filtering** as a procedure for making these decisions, and we will quantify the effectiveness of the matched filter by studying how the **signal-to-noise ratio (SNR)** affects the **bit error rate (BER)** of communication systems.

Overview: In this course - 2

Ultimately we will see the reprise of several topics first introduced in Info Systems I:

- analog and digital signals
- Fourier transforms
- convolution
- filtering
- etc.

Syllabus

Instructor

- Prof. Mike Wakin, 303-273-3607, mwakin@mines.edu
- Office: Brown 327E
- Office hours: Monday 1:30-2:30, Wednesday 2:30-3:30, Friday Noon-1:00, or by appointment

Prerequisite

- Information Systems Science I (EENG 310 or EENG 388)

Textbooks

- *Introduction to Probability, 2nd Edition* by D. P. Bertsekas and J. N. Tsitsiklis (ISBN: 978-1-886529-23-6).
- *Fundamentals of Electrical Engineering I* by D. H. Johnson, available for free on OpenStax CNX (`http://cnx.org/content/col10040`).

Syllabus - 2

Schedule (approximate)

- Introduction to probability (B & T Chapter 1) – 2 weeks
  - independence, conditional probability, Bayes' rule, basic combinatorics
- Discrete random variables (B & T Chapter 2) – 2 weeks
  - probability mass functions (PMFs), expectation and variance, joint PMFs, independence
- Information theory and coding (Johnson Chapter 6) – 2 weeks
  - entropy, source coding, Huffman coding, channel capacity
- Continuous random variables (B & T Chapters 3, 4, 5) – 3 weeks
  - probability density functions (PDFs), conditioning, joint PDFs, independence, covariance and correlation, Bayes' rule, central limit theorem
- Random signals and matched filtering – 2 weeks
  - random vectors, random processes, power spectral density, matched filtering, hypothesis testing
- Digital communication over analog channels (Johnson Chapter 6) – 1 week
  - digital receivers, matched filtering, bit error rate, signal-to-noise ratio
- Parameter estimation (B & T Chapter 9) – 1 week
  - sample mean, sample variance, confidence intervals

Syllabus - 3

Course description

- This course covers signals and noise in electrical systems. Topics covered include information theory, signal to noise ratio, random variables, probability density functions, statistics, noise, matched filters, coding and entropy, power spectral density, and bit error rate. Applications are taken from radar, communications systems, and signal processing.

Students with disabilities

- The Colorado School of Mines is committed to ensuring the full participation of all students in its programs, including students with disabilities. If you are registered with Disability Support Services (DSS) and I have received your letter of accommodations, please contact me at your earliest convenience so we can discuss your needs in this course. For questions or other inquiries regarding disabilities, I encourage you to visit `disabilities.mines.edu` for more information.

Syllabus - 4

Honor code

- All students are expected to abide by the Mines honor code, both in letter and in spirit. This policy may be found in the Undergraduate Bulletin (`http://bulletin.mines.edu/policiesandprocedures/`) and applies to all aspects of the course.
- **Violations of the honor code will be taken very seriously.** The consequences for academic dishonesty at the Colorado School of Mines are severe and can lead to expulsion.
- Students are encouraged to work together on homework problems to develop an understanding of the material. However, each student must:
  - generate and turn in his/her own individual solutions that reflect his/her own individual level of understanding (this includes computer programs), and
  - cite any external resources used to answer a homework question (outside of the textbook, lecture notes, and other students in the course).

  In addition, students may not:
  - copy homework solutions from another student or from any other source, or
  - consult homework or exam solutions from previous offerings of this course.

Syllabus - 5

- Violation of these policies will be considered a violation of the honor code. Questions about these policies should be raised before handing in a homework assignment.

Absenteeism

- Attendance in class is required unless the student has an official excused absence.

Homeworks

- Homework will be due approximately once per week.
- I encourage working in groups on the homework, but the final work turned in must be your own.

MATLAB

- MATLAB will be required for some homework assignments. MATLAB is available in the campus computer labs.

Syllabus - 6

Online exercises

- Approximately once per week, we will devote the class period to solving problems using an online system called OpenStax Tutor (OST). There will be a set of required OST exercises that must be completed in class. These problems will be online and the OST website will only be open for the class period and possibly for a short time afterwards. Thus, you will need to **bring your laptop to class on these days**. The exercises are not supposed to be a test; they are meant to help solidify your understanding of the material. In total, they are worth 15% of your course grade.
- Half of the credit will be based on the actual scores you receive on these assignments. The other half will essentially serve as a participation grade: at the end of each class I will check everyone's progress, and you will receive full credit as long as you are in class and have been making a good faith effort at completing the assignment. To receive credit, you must complete the entire assignment **and view the required feedback**.

Syllabus - 7

Exams

- Each exam will have an in-class portion and a take-home portion.
- The in-class portion of each exam will be closed book and closed notes. However,
  - On Midterm Exam 1, you may bring 1 sheet of 8.5" $\times$11" paper with your own handwriting on both sides.
  - On Midterm Exam 2, you may bring 2 sheets of 8.5" $\times$11" paper, each with your own handwriting on both sides.

  You must hand in these sheets of paper with your exam; however, they will be returned to you.
- The take-home portion of each exam will be open book, but you must do the work entirely on your own.

Projects

- Toward the end of the semester, students will work in small groups on project reports investigating a topic related to probability/statistics, and possibly related to electrical engineering.

Syllabus - 8

Grading scale

- Homework: 20%
- Online exercises: 15%
- Two midterm exams: 22.5% each
- Group projects: 20%
- Scores may be standardized before computing the final average if the means and standard deviations vary. Letter grades will be assigned using a curve, but the lower cutoff for A- will be no higher than 90%, and 80% for B-, etc.

Other resources

- MIT OpenCourseware (`http://bit.ly/NA4aLP`). This course uses the B & T textbook, and will match our B & T coverage pretty closely. There are alternative lecture notes and homework assignments (with solutions).
- Harvard Statistics 110: Probability (`http://bit.ly/NhMdxb`). A similar course with video lectures available through iTunes.

Syllabus - 9

- Khan Academy lectures on probability (`http://bit.ly/PaBJmu`). Some fantastic (shortish) lectures available for many of the topics in this course. Many of them are centered on working specific problems.

Acknowledgements

- Lecture notes were prepared with contributions from M. Davenport, J. Romberg, C. Rozell, and R. Haupt.