ECE 3530/CS 3130: Engineering Probability & Stats
Lecture Notes Spring 2015

Lecture Notes Handout #1

1 Intro to Probability and Statistics

Randomness is all around us; in many engineering and scientific applications.

- Communications: A transmitted signal is attenuated, corrupted, and is received with added noise. The attenuation, channel-induced corruption, and noise can be considered to be random.
- Controls: A controls system includes measuring the state of an engineered system and using actuators (e.g., motors) to react to that data. The measurement and the actuation are not exact; both are considered random.
- Manufacturing: There are process variations that are considered random. Manufacturing imprecision causes products to differ; some will even not work, or meet specs, in a way that can be considered random. If cost was no concern, you might measure precisely every single product; but if not, you don't know for sure whether each product is"good".
- Economics: Securities prices change in a random fashion; we don't know in advance how it will change.
- Imaging: Medical images (e.g., CT scans) try to use as little energy as possible due to possible health effects. These images are corrupted by noise. For example, a system may be counting photons which arrive randomly over time, and the number which arrive in a given time period would be considered random.
- Biology: The spread of an infection; the growth of cells; genetic inheritance; can all be considered as random.
- Internet: The spread of an idea (via social network), the sequence of clicks while surfing the web, the arrival (or dropping) of packets at a destination or at a router, the delay between send and receive of a packet, downtime of a server, can all be considered as random. The randomness of passwords and keys are critical to security.
- Algorithms: Many algorithms incorporate randomness on purpose in order to compute a result or to store data.

In all of these applications, we have what we call random variables. These are things which vary across time in an unpredictable manner. Sometimes, these things we truly could never determine beforehand. For example, thermal noise in a receiver is truly unpredictable. In other cases, perhaps we could have determined if we had taken the effort. For example, whether or not a machine is going to fail today could have been determined by a maintenance checkup at the start of the day. But in this case, if we do not perform this checkup, we can consider whether or not the machine fails today as a random variable, simply because it appears random to us.

The study of probability is all about taking random variables and quantifying what can be known about them. Probability is a set of tools which take random variables and output deterministic numbers which answer particular questions. So while the underlying variable or process may be random, we as engineers are able to 'measure' them.

For example:

- The expected value is a tool which tells us, if we observed lots of realizations of the random variable, what its average value would be.
- Probability of the random variable being in an interval or set of values quantifies how often we should expect the random value to fall in that interval or set.
- The variance is a tool which tells us how much we should expect it to vary.
- The correlation or covariance (between two random variables) tells us how closely two random variables follow each other.

The study of probability is simply the study of random variables, to provide tools which allow us to find deterministic measures of what can be known about the random variables.

The study of statistics is about learning about something from noisy data. Statistics is a key element of science, that is, how to test a hypothesis, or estimate an unknown value. As engineers, we build systems that then we need to understand. We do a lot of testing of systems to characterize them. We essentially are scientists of engineered systems, and as such, need to understand one of its key building blocks.

1.1 Probability is Not Intuitive

We are often mislead by our intuition when it comes to understanding probability. Thus a mathematical framework for determining probabilities is critical. To do well in this course you must follow the frameworks we set up to answer questions about probability.

Example: Rare Condition Tests
One example of how our intuition can betray us is in the understanding of tests for rare medical conditions. As an example, consider a common

test for genetic abnormalities, the "nuchal translucency" (NT) test, on a 12-week old fetus. (I'm taking these numbers from the internet and I'm not a doctor, so please take this with a grain of salt.) About 1 in 1000 fetuses have such an abnormality. The NT test correctly detects 62% of fetuses that have the abnormality. But, it has a false positive rate of 5%, that is, 5% of normal fetuses are detected by the NT test as positive for the abnormality. Given that a fetus has tested positive in the NT test, what is the probability that it is, in fact, abnormal?

Here's my solution, using Bayes' Law: **1.2%**. That is, even after the positive on the NT test, it is *highly unlikely* that the fetus has a genetic abnormality. True, it is 12 times higher than before the test came back with a positive, but because the condition is so rare, and the false positive rate is reasonably high, the probability of the fetus having the abnormality is still very very low. We will learn Bayes' Law and how to use it in the first segment of this course.

$$P\left[A|NT+\right] = \frac{P\left[NT+|A\right]P\left[A\right]}{P\left[NT+|A\right]P\left[A\right] + P\left[NT+|A'\right]P\left[A'\right]} = \frac{(0.001)(0.62)}{(0.001)(0.62)+(0.999)(0.05)} = 0.012$$

Example: Conjunction Fallacy
Linda is 31 years old, single, outspoken and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice. Before entering the workforce, Linda spent time in the Peace Corps helping women gain access to health care in and around the Congo region of Africa. Which is most likely?

1. Linda is a bank teller.
2. Linda is a bank teller and writes for a feminism blog.

This question is adapted from a question posed to U. of British Columbia undergraduates in 1983 by Amos Tversky and Daniel Kahneman. They found 85% of students rated (2) as more likely than (1). However, if we define A = the event that Linda is a bank teller, and B = the event that Linda writes for a feminism blog, then (1) has probability $P\left[A\right]$, and (2) has probability $P\left[A \cap B\right]$. The laws of probability, as we will see, mandate that $P\left[A \cap B\right] \leq P\left[A\right]$, with equality only if all bank tellers write for feminism blogs. In short, Tversky and Kahneman noticed that people are swayed by a narrative – the more detail, the more believable it seems.

Example: Probability of Failure in Disasters
Consider a hypothetical nuclear reactor located near an ocean that will leak radiation only if there is both a severe earthquake, and simultaneously a flood in the reactor facility. The probability of an severe earthquake is, on any given day, 1/10,000, and the probability of a flood in the facility is 1/5,000. What is the probability the nuclear reactor will leak radiation?

There is a tendency, without thinking about the events in question, to assume they are independent, and thus the probability $P\left[A \cap B\right] =$

$P[A]\,P[B]$. In this case would lead to a probability of $\frac{1}{5\times10^7}$, or one in 50 million. That would be, about once every 130,000 years. Based on this analysis, an engineer might say, this facility will never fail. However, it would be more accurate to realize the events are closely related, that a tsunami can be both a severe earthquake, and cause major flooding at the same time. The rate of leaks would be close to the probability of a major tsunami, something likely to happen in an average person's lifetime.

In summary, intuition is not a good way to approach problems in probability and statistics. It is important to follow the methods we will present in this course. It is also important to have lots of practice – repetition will help to train us on the proper procedure for analysis.