ECE 3530 / CS 3130: Probability and Statistics
Lecture Notes

Neal Patwari, Spring 2014

Lecture Notes Handout #1

1 Intro to Probability and Statistics

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

- communications,
- controls,
- manufacturing,
- economics,
- imaging,
- biology,
- the Internet,
- algorithms,
- systems engineering.

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 sequenced by continuous or discrete time (or space), which represent the temporal (or spatial) variation of a random variable. This class is all about quantifying *what can be known* about random variables and random events.

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 do a lot of testing of complicated systems to determine whether or not they're working. We also build sensors that record noisy observations and report some value to the user or to a system.

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.

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

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 $P[A]$, and (2) has probability $P[A \cap B]$. The laws of probability, as we will see, mandate that $P[A \cap B] \leq P[A]$, 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 $P[A \cap B] = P[A]\,P[B]$. In this case would lead to a probability of $\frac{1}{5\times 10^7}$, 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.

2 Events as Sets

All probability is defined on sets. In probability, we call these sets *events*. A set is a collection of elements. In probability, we call these *outcomes*. The words we use for events is slightly different than what you did when you learned sets in a math class, see Table 2.

Def'n: *Event*
A collection of outcomes. Order doesn't matter, and there are no duplicates.

Set Theory	**Probability Theory**	**Probability Symbol**
universe	sample space (certain event)	S
element	outcome (sample point)	s
set	event	E
disjoint sets	disjoint events	$E_1 \cap E_2 = \emptyset$
null set	null event	$\emptyset$

Table 1: Set Terminology vs. Probability Terminology

2.1 Introduction

There are different ways to define an event (set):

- List them: $A = \{0, 5, 10, 15, \ldots\}$; $B = \{Tails, Heads\}$.
- As an interval: $[0, 1]$, $[0, 1)$, $(0, 1)$, $(a, b]$. Be careful: the notation overlaps with that for coordinates!
- An existing event set name: $\mathbb{N}$, $\mathbb{R}^2$, $\mathbb{R}^n$.
- By rule: $C = \{x \in \mathbb{R} | x \geq 0\}$, $D = \{(x, y) \in \mathbb{R}^2 | x^2 + y^2 < R^2\}$.

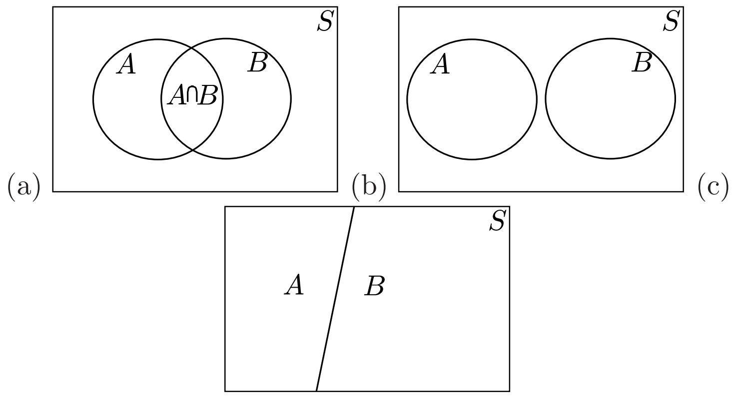

Figure 1: Venn diagrams of sample space S and two sets A and B which (a) are not disjoint, (b) are disjoint, and (c) form a partition of S.

2.2 Venn Diagrams

Venn diagrams can be used to pictorially show whether or not there is overlap between two or more sets. They are a good tool for helping remember some of the laws of probability. We don't use them in proofs of the laws of probability, however, they're particularly good to develop intuition.

2.3 Important Events

Here's an important event: $\emptyset = \{\}$, the *null event* or the *empty set*.

Here's the opposite: S is used to represent the set of everything possible in a given context, the *sample space*.

- $S = B$ above for the flip of a coin.
- $S = \{1, 2, 3, 4, 5, 6\}$ for the roll of a (6-sided) die.
- $S = \{Adenine, Cytosine, Guanine, Thymine\}$ for the nucleotide found at a particular place in a strand of DNA.
- $S = C$, *i.e.*, non-negative real numbers, for your driving speed (maybe when the cop pulls you over).

2.4 Operating on Events

We can operate on one or more events:

- Complement: $A^c = \{x \in S | x \notin A\}$. We must know the sample space S!
- Union: $A \cup B = \{x | x \in A \,\text{or}\, x \in B\}$. Merges two events together.

- Intersection: $A \cap B = \{x | x \in A \,\text{and}\, x \in B\}$. Limits to outcomes common to both events.

DO NOT use addition to represent the union, and DO NOT use multiplication to represent the intersection. An example of why this is confusing:

$$\{1\} + \{1\} = \{1\}.$$

This leads to one of the most common written mistakes – exchanging unions and plusses when calculating probabilities. Don't write $P[A] + P[B]$ when you really mean $P[A \cup B]$. Don't add sets and numbers: for example, if A and B are sets, don't write $P[A] + B$.

2.5 Disjoint Sets

Def'n: *Disjoint*
Two events A_1 and A_2 are disjoint if $A_1 \cap A_2 = \emptyset$.

Def'n: *Mutually exclusive*
Multiple events $A_1, A_2, A_3, \ldots$ are mutually exclusive if every pair of events is disjoint.

Some disjoint events: $\{1, 2\}$ and $\{6\}$; A and A^c; A and $\emptyset$.

3 Probability

You're familiar with functions, like $f(x) = x^2$, which assign a number output to each number input. Probability assigns a number output to each event input.

3.1 How to Assign Probabilities to Events

As long as we follow three intuitive rules (*axioms*) our assignment can be called a *probability model.*

Axiom 1: For any event A, $P[A] \geq 0$.

Axiom 2: $P[S] = 1$.

Axiom 3: For any two disjoint events A_1 and A_2,

$$P[A_1 \cup A_2] = P[A_1] + P[A_2].$$

The final axiom in the Walpole book is more complicated:

Axiom 3: If $A_1, A_2, A_3, ...$ is a sequence of mutually exclusive events, then

$$P[A_1 \cup A_2 \cup A_3 \cup \cdots] = P[A_1] + P[A_2] + P[A_3] + \cdots$$

However, one may prove the more complicated Walpole Axiom 3 from the three axioms given above. For details, see [Billingsley 1986].

Example: DNA Measurement

Consider the DNA experiment above. We measure from a strand of DNA its first nucleotide. Let's assume that each nucleotide is equally likely. Using axiom 3,

$$P[\{a, c, g, t\}] = P[\{a\}] + P[\{c\}] + P[\{g\}] + P[\{t\}]$$

But since $P[\{a, c, g, t\}] = P[S]$, by Axiom 2, the LHS is equal to 1. Also, we have assumed that each nucleotide is equally likely, so

$$1 = 4P[\{a\}]$$

So $P[\{a\}] = 1/4$.

Def'n: *Discrete Uniform Probability Law*
In general, for event A in a discrete sample space S composed of equally likely outcomes,

$$P[A] = \frac{|A|}{|S|}$$

3.2 Properties of Probability Models

1. $P[A^c] = 1 - P[A]$. Proof:
   First, note that $A \cup A^c = S$ from above. Thus

   $$P[A \cup A^c] = P[S]$$

   Since $A \cap A^c = \emptyset$ from above, these two events are disjoint.

   $$P[A] + P[A^c] = P[S]$$

   Finally from Axiom 2,

   $$P[A] + P[A^c] = 1$$

   And we have proven what was given.

   Note that this implies that $P[S^c] = 1 - P[S]$, and from axiom 2, $P[\emptyset] = 1 - 1 = 0$.

2. For any events E and F (not necessarily disjoint),

   $$P[E \cup F] = P[E] + P[F] - P[E \cap F]$$

   Essentially, by adding $P[E] + P[F]$ we double-count the area of overlap. Look at the Venn diagram in Figure 1(a). The $-P[E \cap F]$ term corrects for this. Proof: Do on your own using these four steps:

(a) Show $P[A] = P[A \cap B] + P[A \cap B^c]$.

(b) Same thing but exchange A and B.

(c) Show $P[A \cup B] = P[A \cap B] + P[A \cap B^c] + P[A^c \cap B]$.

(d) Combine and cancel.

3. If $A \subset B$, then $P[A] \leq P[B]$. Proof:
   Let $B = (A \cap B) \cup (A^c \cap B)$. These two events are disjoint since $A \cap B \cap A^c \cap B = \emptyset \cap B = \emptyset$. Thus:

   $$P[B] = P[A \cap B] + P[A^c \cap B] = P[A] + P[A^c \cap B] \geq P[A]$$

   Note $P[A \cap B] = P[A]$ since $A \subset B$, and the inequality in the final step is due to the Axiom 1.

4. $P[A \cup B] \leq P[A] + P[B]$. Proof: directly from 2 and Axiom 1. This is called the *Union Bound*. **On your own:** Find intersection bounds for $P[A \cap B]$ from 2.