Welcome to
STAT 230
Spring 2022!

James Adcock

Office: M3 3019

E-mail: jradcock@uwaterloo.ca

"A statistician is someone who can have one foot in a pail of ice cold water and another foot in a pail of boiling hot water, and say that, on average, the water temperature is fine."

Top ten reasons to be a statistician:

10. Estimating parameters is easier than dealing with real life.

9. Statisticians are significant.

8. I always wanted to learn the entire Greek alphabet.

7. The probability a statistician major will get a job is > 0.9999.

6. If I flunk out, I can always transfer to Engineering.

5. We do it with confidence, frequency, and variability.

4. You never have to be right – only close.

3. We're normal and everyone else is skewed.

2. The regression line looks better than the unemployment line.

1. No one knows what we do so we are always right.

http://www.amstat.org/

What is Statistics?

Statistics is the science of understanding data and making decisions in the face of variability and uncertainty.

- When used correctly, statistics can tell us about trends observed in the past and can be useful in predicting what **may** happen in the future.

Some Applications of Statistics:

1. **Insurance**

- You know that in order to drive your car you are required by law to have car insurance. If you have a mortgage on your house, you must have it insured as well. The rate that an insurance company charges you is based on statistics from all drivers or homeowners in your area.

2. **Consumer Goods**

- Wal-Mart, a worldwide leading retailer, keeps track of everything they sell and uses statistics to calculate what to ship to each store, and when. From analyzing their vast store of information, for example, Wal-Mart decided that people buy strawberry Pop Tarts when a hurricane is predicted in Florida! So they ship this product to Florida stores based upon the weather forecast.
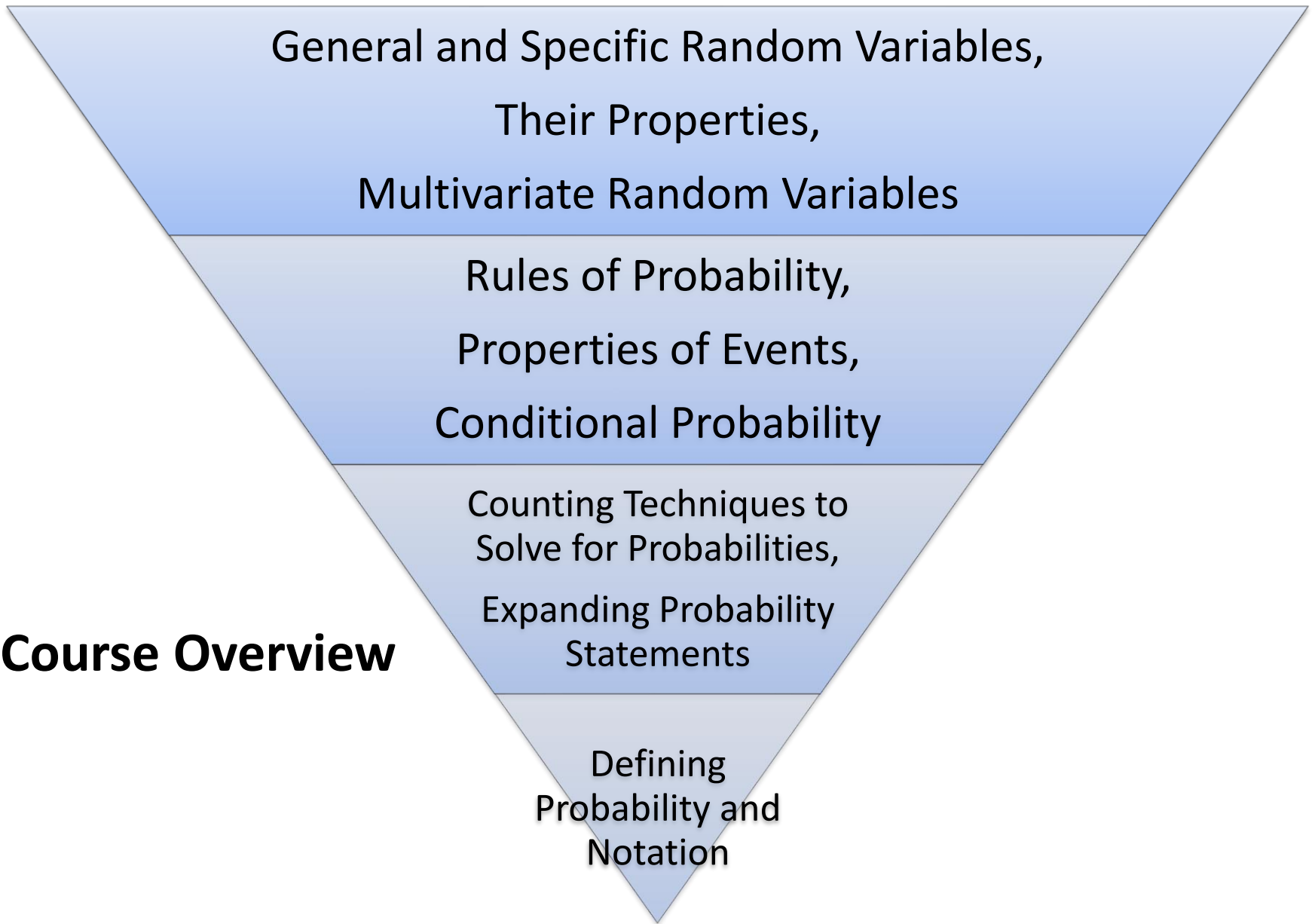
3. Medical Studies

- Scientists must show a statistically valid rate of effectiveness before any drug can be prescribed. Statistics are behind every medical study you hear about.

4. Quality Testing

- Companies make thousands of products every day and each company must make sure that a good quality item is sold. However, a company cannot test each and every item that they ship to the consumer. Instead, the company uses statistics to test just a few, called a sample, of what they make. If the sample passes quality tests, then the company assumes that all the items made in the group, called a batch, are good.

Course Overview

General and Specific Random Variables,

Their Properties,

Multivariate Random Variables

Rules of Probability,

Properties of Events,

Conditional Probability

Counting Techniques to
Solve for Probabilities,

Expanding Probability
Statements

Defining
Probability and
Notation

Introduction to Probability: Chapter 1

Chapter Outcomes:

- Introduce the concept of probability

- Definitions of probability

- Probability is used to model **uncertainty** and **variability**.

- This uncertainty can occur due to a mixture of at least two factors:

1. *Variability in populations* made up of animate or inanimate objects (i.e. people naturally vary in size, weight, height, etc.)

2. *Variability in processes or phenomena* (i.e. fluctuations in stock prices or exchange rates, weather, etc.)

- Variability and uncertainty in a system often make it more difficult to plan or make decisions without suitable tools.

- We can *never eliminate* this variability and uncertainty, but we can usually describe, quantify, and deal with it using the **theory of probability**.

- **Some Examples of Probability:** Compare and contrast these probability statements.

1. To check whether a company is manufacturing bulbs properly, 150 bulbs are selected randomly from a certain big lot for examination. After the examination, it is found 80 bulbs out of 150 are defective (P(defective) = 80/150).

2. A die is rolled 60 times on the table, and every time a 5 appears it is recorded. How many rolls will it take? Can we calculate how many rolls would we "expect" it to take (i.e. the average number of rolls)?

3. Knowing that a die has 6 faces, each equally likely to turn up on any roll, we can conclude that there is a 1/6 chance of obtaining a 5.

4. Knowing that a fair coin has two faces, each equally likely to turn up when flipped, we can conclude that there is a 50% chance of obtaining a head.

5. I missed my alarm this morning and woke up late, hence I have an 80% chance of arriving late to class.

6. I have taken data management in high school and so there is a 70% chance that I will do well in this STAT course.

Defining Probability:

- To start: An **experiment** is a situation involving chance or uncertainty that leads to results called outcomes.

- An **outcome** is the result of a single trial (attempt) of an experiment.

- An **event** is one or more outcomes of an experiment.

- The set of **ALL** possible distinct outcomes in a random experiment is called the **sample space, *S***.

- Probability is the measure of how likely an event is.

- The Classical definition:

The probability of an event is

$$\frac{\text{\# of ways the event can occur}}{\text{\# of outcomes in } S}$$

Here all the points in *S* are **equally likely**

Example: rolling a die

S = {1, 2, 3, 4, 5, 6}

If a die is rolled once, the number 2 can be observed in exactly 1 out of 6 ways.

Question: Does this mean that in 6 rolls, a one will or must occur exactly once?

- The <u>Relative Frequency</u> definition:

- Here probabilities are assigned on the basis of experimentation or historical data.

The probability of an event is the (limiting) proportion (or fraction) of times the event occurs when the experiment is repeated a large number of times under the exact same conditions.

- Disadvantages?

- The Subjective Probability definition:

The probability of an event is based on how confident the person making the statement is that the event will occur. Usually based on prior knowledge (belief) or available information.

• Disadvantages?
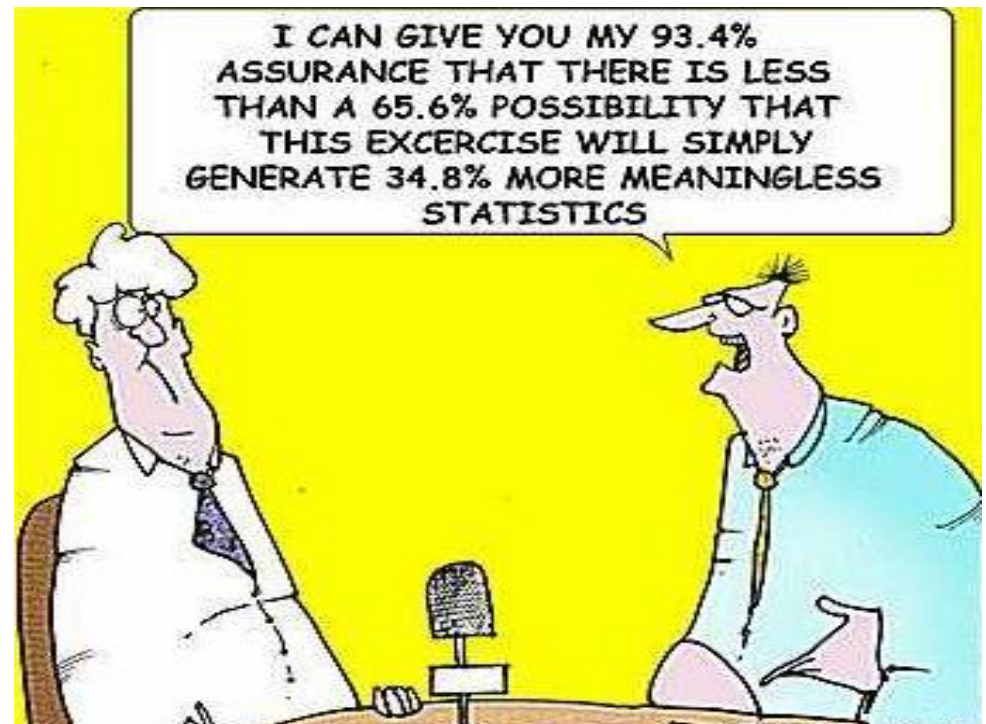
- To avoid difficulties in providing a satisfactory definition of probability, we opt to use a *formal*, *mathematical* approach.

- A **Probability Model** consists of:
1. a **sample space**, *S*, of our random experiment
2. a **set of events** (subset of the sample space) to which probabilities can be assigned
3. a **mechanism** for assigning probabilities to the events