INTRODUCTION TO PROBABILITY

COVERAGE

- Section 2.1 & 2.2 .... Basic Definitions
- Section 2.3 .... Set Notation and Manipulation
- Section 2.4 & 2.5 .... Sample Spaces and Events

> **NOTE: It is important that you (at least briefly) read through the corresponding sections of the textbook before coming to class.**

LEARNING GOALS

By the end of this set of notes, students should be able to:

- Understand the differences between the four types of probability (**subjective**, **empirical**, and **theoretical**).
- Appropriately use the basics of set notation including **intersection**, **union**, **complement**, **distributive laws**, and **DeMorgan's laws**.
- Illustrate (and in simple cases count or find probability) **sample spaces**, **outcomes**, and **events** in the context of any experiment. Note that this involves being able to read a word problem and identify the important pieces of information. It also involves learning some notation.
- Utilize **tables** and/or **Venn diagrams** to represent sample spaces for a variety of word problems.
- Apply basic probability rules including the axioms of probability and the complement rule.

WHAT IS PROBABILITY AND WHY IS IT IMPORTANT?

> As a field of study, probability is the study of "chance". It may be applied to any "experiment" where multiple outcomes are possible. It is used in an attempt to quantify the "likelihood" of particular outcomes occurring within the context of an experiment.

> In practice, probability provides the basis for statistics – we use understanding of probability to help us draw conclusions about real-life questions based on data. In this course, we will begin with a study of probability for the first ~8~ weeks. We will then segue into an introductory study of statistics. Hopefully you will be able to draw connections between the two areas!

ACTIVITY

There are four approaches that can be taken with respect to a discussion of probability. This brief activity will illustrate the four approaches.

Consider the experiment of randomly drawing a single card from a shuffled deck of playing cards. Question: What is the "probability" that the card will be a heart?

1. **Subjective Probability**: Basically an educated guess – for the example, on a continuum from 0 to 1, how likely do you "think" it is that you will draw a heart?

2. **Empirical Probability**: A guess based on data collection - what do collected "data" suggest about the probability of drawing a heart?

3. **Theoretical Probability**: An exact assessment of the probability based on mathematical argument. For our example...

In most cases, we won't have the luxury of considering all three probabilistic viewpoints. Some comments:

1. Subjective probabilities are both easiest to obtain and typically ***least useful***. Why?

2. Theoretical probabilities are the ***most useful*** but may be difficult to obtain. Again, why?

3. Empirical probabilities are the most commonly used in practice when theoretical probabilities are difficult or impossible.

For the first part of the course, we focus almost entirely on ***theoretical probability***. Then in the second part, we will consider its applications using the ideas of ***empirical probability***.

IMPORTANT DEFINITIONS

Random Experiment: Any process or activity that produces a result or "outcome" that cannot be predicted with absolute certainty.

Outcome: Any uniquely defined individual possible result for an experiment (the book also calls this a "simple event").

Sample Space (S): The collection of all possible outcomes for a random experiment. When this collection is finite or countable, we call the sample space **discrete**.

Event (A): subset of the sample space; a collection of one or more outcomes that (generally) can be described by some specific criteria. Capital letters at the beginning of the alphabet are used to denote events.

Probability of an event P(A): roughly speaking the fraction of experiments in which the event would occur if the experiment could be repeated infinitely often.

Example: Toss a 10-sided die (sides labeled from 0 to 9).

- What is the sample space?

- Identify three different events.

- Find the probability of the events you identified.

Key Point: If every outcome in the sample space is equally likely (a highly desirable property of S), then the probability of any event A may be calculated by P(A) = #A / #S (where #A is the number of outcomes in A and #S is the number of outcomes in S).

OTHER IMPORTANT REPRESENTATIONS OF SAMPLE SPACES

Often sample spaces can be LARGE. We do not want to enumerate all of the outcomes in a list! We will still need to count them! Depending on what attributes are important to us, we often group outcomes together and describe them in **Tree Diagrams**, **Tables** and/or **Venn Diagrams**.

Tree Diagrams can be useful for making sure you identify all of the elements of a sample space. They are most useful when the experiment consists of a multi-step process.

Example: Toss a die and a coin.

Tables are often used when we have groups of people that are being categorized on two different attributes.

Example: The faculty of a mathematics department consist of 13 men and 11 women. 6 of the men favor the new general education requirements, while 9 of the women favor the new general education requirements. The 5 o'clock news will select one person at random to interview.

There are 24 "people" in our sample space, but we don't need to itemize them. We only need to know how many fall into each "category". We can do this using a table....

Venn Diagrams are most useful when the categories we want to use can overlap (not the case above!).

Example: The 24 faculty have voted on each of two general education proposals. 16 faculty voted in favor of proposal A, 12 faculty voted in favor of proposal B, and 3 faculty were not in favor of either.

SET THEORY

Note: If this part is not review, or if you find that you feel uncomfortable with the set notation, then please make sure you spend a fair amount of time doing additional problems from Section 2.3! Venn Diagrams can be useful for visualization.

Key Definitions & Rules

1. **Subset**: A is a subset of B (denoted $A \subset B$) if every outcome from A is also found in B.

2. **Complement**: The complement of a set A is the set of all outcomes not found in A. Complementation is denoted $\overline{A}$ in our text. You are welcome to use A^c if you prefer that notation instead. THINK: Complement == NOT

3. **Union**: A union B (denoted $A \cup B$) is the collection of outcomes found in either A or B (or perhaps both). THINK: Union == OR

4. **Intersection**: A intersect B ($A \cap B$) is the collection of outcomes found in both A and B. THINK: Intersection == AND

5. **Disjoint / Mutually Exclusive**: Two sets A and B are disjoint (or mutually exclusive) if they have no outcomes in common. That is, if $A \cap B = \emptyset$.

6. **Distributive Laws**

   $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ $\quad\quad$ $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

7. **DeMorgan Laws**

   $\overline{(A \cap B)} = \overline{A} \cup \overline{B}$ $\quad\quad$ $\overline{(A \cup B)} = \overline{A} \cap \overline{B}$

Example: Toss a 10-sided die (sides labeled from 0 to 9). Let A be the event an even number is rolled. Let B be the event that the outcome of the roll is at least 6.

$A =$	$P(A) =$
$B =$	$P(B) =$
$A \cap B =$	$P(A \cap B) =$
$A \cup B =$	$P(A \cup B) =$
$\overline{A} =$	$P(\overline{A}) =$
$\overline{A} \cap B =$	$P(\overline{A} \cap B) =$
$\overline{A \cap B} =$	$P(\overline{A \cap B}) =$

Example: Restaurants that wish to operate in Fort Mitchell may obtain certifications from the city planning commission (A), the health and safety board (B), and the liquor commission (C). Records show that only 5% of potential candidates are certified by all three boards. 20% are certified by A and B, 28% by A and C, and 21% by B and C. 50% are certified by A and 57% are certified by B. Additionally 3% of restaurants have gone rogue and are not certified by any commission.

a. Find the probability that a randomly chosen restaurant is certified by the liquor commission and may serve liquor.
b. Find the probability that a randomly chosen restaurant is certified by A and B but not by C.
c. Find the probability that a randomly chosen restaurant is certified by at least one organization.
d. Find the probability that a randomly chosen restaurant is certified by exactly one organization.

AXIOMS & RULES OF PROBABILITY

Axioms of Probability (Definition 2.6)

Suppose that S is a sample space. To any event $A \subset S$ we assign a number P(A), called the probability of A, so that the following axioms are always true:

1. $P(A) \geq 0$
2. $P(S) = 1$
3. For any sequence of pairwise mutually exclusive events $A_1, A_2, A_3, \ldots$, the probability of the union of the sequence is the sum of the individual event probabilities:

$$P\left(A_1 \cup A_2 \cup A_3 \cup \cdots\right) = \sum P\left(A_i\right)$$

From these axioms we may verify some other simple rules:

Probability of the Null Event:

$P(A) \leq 1$:

Subset Rule:

Complement Rule:

Additive Rule (Theorem 2.6):

PROBABILITY (TRUTHS & MISCONCEPTIONS)

WHAT IS TRUE?

1. If we do an experiment, the probability that some outcome from the sample space will occur is _____________. In the same context, if an outcome has a large probability, it is pretty likely to occur.

2. If we do an experiment, the probability that some outcome not in the sample space will occur is ____________. In the same context, if an outcome has a very small probability, we'd be quite surprised to see that outcome occur in a single experiment.

3. Events that have larger probabilities will tend to occur more often as an experiment is repeated.

4. The only guarantee in a probabilistic situation is **variability**. Consider rolling dice...

   http://highered.mcgraw-hill.com/sites/dl/free/0072868244/124727/Probability.html

5. Variability does decrease the more an experiment is repeated; as the number of trials gets larger, the results (empirical probabilities) will get closer to the theoretical probabilities. Consider spinning the wheel of fortune...

   http://illuminations.nctm.org/ActivityDetail.aspx?ID=79

6. Events that are very unlikely WILL occur if we repeat the experiment enough times. More wheel...

CLEARING UP MISCONCEPTIONS

7. In a fixed number of trials, it is VERY UNLIKELY that outcomes will happen in exact proportion to their theoretical probabilities. Roll dice...

   Question to ponder: Which of the following would be more likely – six dice being rolled to show each number once? Or sixty dice being rolled to show each number 10 times?

8. If we repeat an experiment several times, and event A doesn't happen – this doesn't make event A more likely to happen on the next repetition. This concept is known as **independence**. For example, flip a coin...

9. That an event has positive probability does not guarantee the event will happen in any given fixed number of trials. For example, pocket aces in Texas Hold'em...

10. And be careful: The probability of an event may not always match your intuition!

    http://www-stat.stanford.edu/~susan/surprise/Birthday.html

ADDITIONAL EXAMPLES (YOU WILL NEED ADDITIONAL PAPER)

1. Two of four applicants (Amy, Frank, Mary, & Tom) are to be chosen for identical positions at Bank One.

   a. Write out the sample space for possible selection of applicants.
   b. Let A denote the set of selections containing at least one male. How many elements are in A?
   c. Let B denote the set of selections containing two females. How many elements are in B?
   d. Describe the set $A \cap B$? $A \cup B$? Explain the significance of this result.

2. Cars enter an intersection and may either go left, right, or straight.

   a. Write out the sample space for an experiment in which two cars enter the intersection.
   b. Discuss whether you believe the outcomes in S are equally likely.

3. The math department at a school has 107 students. Of these, there are 53 females – 28 of whom are MathEd majors, 16 of whom are Stat majors, and the rest of whom are straight Math. Of the males, 16 are MathEd. There are 40 straight Math majors in the department.

   a. Find the probability that a randomly selected student from the department majors in Statistics.
   b. Find the probability that a randomly selected student from the department is male and majors in MathEd.
   c. Find the probability that a randomly selected student from the department is either female or majors in statistics (or both).
   d. Find the probability that a randomly selected ***MathEd*** student is female.

4. I have three marbles in a bag. They are colored red, blue, and green, respectively. An experiment consists of drawing out two marbles **with replacement**. What are the possible outcomes and their associated probabilities? What is the probability of drawing out a red marble at least once?

   What happens if I draw **without replacement**? What happens if I add a 2^{nd} red marble to the bag?

5. Football players can be classified as to whether they play offense or defense. But in high school, many players do both. In a recent survey of 568 football players, there were in fact 110 that played both offense and defense. There were 327 offensive players in the group. And there were 28 players that only participate on "special teams" – that is they play neither offense nor defense.

   a. What is the probability that a randomly selected player plays defense?
   b. What is the probability that a randomly selected ***defensive*** player also plays offense?

ASSIGNMENT

- HW 2A (Online) – to be completed by the deadline on IMath.
- HW 2A (Written) – to be completed by the deadline announced in class.
- For extra practice you may wish to try the following problems from the text: 2.2, 2.3, 2.6, 2.8, 2.11, 2.14, 2.19, 2.21, 2.28

HW 2A (WRITTEN)

Please submit a set of solutions for the following problems (due-date to be announced in class). Note that you should explain and provide mathematical detail as necessary to support your answers in order to achieve full credit. In general, a sampling of 2-3 problems from each written assignment will be graded for credit (the rest will be graded for completion).

1. Give an appropriate sample space for each of the following experiments. Then indicate whether you should assume the elements of your sample space are equally likely. Note: When there are fewer than 20 elements, you should simply list them in set notation. For larger numbers you will want to use some other representation.

    a. Flip three coins (a quarter, nickel, and dime).
    b. Draw a card from a standard deck and observe its rank (the number or symbol on the card).
    c. In the lottery's pick-4, four balls are drawn with each draw coming from the digits 0-9. If you match the sequence of numbers drawn, you win.
    d. A chair-person and a secretary will be chosen at random from a committee of 4 members.
    e. Roll a 4-sided die (labeled 0,1,2,3) and then flip N coins where N is the number that showed on the die.
    f. Throw a bean-bag high into the air and record the distance away at which it lands.

2. Two marbles are drawn from a bag containing five marbles labeled 1,2,3,4,5. Note that marbles are not replaced in between draws and the order in which marbles are drawn is not important. *List all elements of S in set notation. Then list the following events in set notation and calculate their probabilities.*

    A = The "1" and "2" have been drawn.
    B = The "1" is the first number drawn.
    C = The "2" has been drawn.
    D = Neither the "1" nor "2" nor "3" were drawn.
    E = Either the "3" was drawn or the minimum number on the dice was "5".

3. Textbook Problem #2.6. Additionally, find the probabilities of all events in question.

4. (Similar to Problem #2.30) Three imported wines are to be ranked from lowest to highest by a purported wine expert. That is, one wine will be identified as best, another as 2^{nd} best, etc.

    a. List the sample space for the rankings.
    b. Assume the "expert" really knows nothing and simply assigns ranks at random. What is the probability that the rankings are correct?
    c. Repeat (a) and (b) supposing there are 4 wines. Briefly discuss how and why the probability from part (b) changes.
    d. **(Extra credit)**: Calculate the probabilities that the fake "expert" would rank groups of 5, 6, and 7 wines correctly.

5. Textbook Problems #2.5(a) and #2.21. Note: This method of rewriting A in terms of B is called a ***partition***.

Note: No part of any problem on this assignment is designed to take more than 3-5 minutes. So if you find that you are struggling and spending 10+ minutes on any part, please make sure to ask questions (email, office hours, beginning of class). This advice applies to HW 2A Online as well!