Administrata

The course syllabus is available at http://thirteen-01.stat.iastate.edu/wiki/Stat430. You must use the username **Stat430** and the password **2010group**.

The course will use the following grading scheme.

- **Project**: 20%
- **Homework**: 40%
- **Midterm**: 15%
- **Final**: 25%

Review of Statistical Methodology

The following iterative methodology highlights the fundamentals of statistics.

1. Pose a question, identify the population of interest
2. State your hypothesis
3. Design and experiment, collect data
4. Analyze the data

Example

Here is an example about potential Stat 430 students. We will not cover the last step today, just the first 3.

1. Within each class, is there a difference in knowledge of the prerequisite material? Does this change from semester to semester?
2. I hypothesize that there are differences in each class, but that it does not change significantly from semester to semester.
3. Take a sample of the population and administer the survey. The sample is today's class attendees, and the survey consists of two questions–does the student know conditional probability and does the student know Bayes' rule. The survey is administered by raising hands and counts are taken.

We make several assumptions with our hypothesis and these should be listed.

- The sample is a random sample of the population (maybe, maybe not, but it's probably the best we can do in this case).
- Individuals answer independently (definitely not in this case–we would want to administer the test differently for this).
- The process by which each student answers yes or no is identical (probably not).

Now let's try to be a bit more mathematically rigorous with our hypothesis.

2. Let p_{1c}, p_{2c} be the probabilities that a student from last year's class and this year's class (respectively) knows conditional probability. Let p_{1b}, p_{2b} be the probabilities that a student from last year's class and this year's class (respectively) knows Bayes' rule. My hypothesis is can thus be written as follows: $p_{2b} \neq p_{2c} \wedge p_{1b} = p_{2b} \wedge p_{1c} = p_{2c}$.

Review of Probability

Review of some definitions.

- A random experiment is a sequence of happenings where all possible outcomes are known
- The sample space of an experiment is the set of all possible outcomes, usually denoted as Ω
- An outcome is an element of the sample space, in other words, some $\omega \in \Omega$
- An even is a collection of outcomes, usually denoted using upper-case letters.

Important Definition: Probability Measure

For some event E in sample space Ω, a probability measure is a function $P\colon E \to [0,1]$ such that the following are true.

1. $P(\Omega) = 1$
2. $P(E) \geq 0$ for all $E \in \Omega$
3. For disjoint events $E, F \in \Omega, P(E \cup F) = P(E) + P(F)$

There are some immediate implications from these properties.

- $P(\overline{E}) = 1 - P(E)$
- $P(\emptyset) = 0$
- $A \subset B \Rightarrow P(A) \leq P(B)$
- Addition law: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Counting Method

How do we assign a probability to an event? The counting method works in cases where there are a finite number of outcomes in the sample space and when they are all equally likely. Using this method, the probability of an event E with a sample space size N is as follows.

$$P(E) = \sum_{\omega \in E} P(\omega) = \sum_{\omega \in E} \frac{1}{N} = \frac{1}{N} \sum_{\omega \in E} = \frac{1}{N}|E|$$

Conditional Probability

The following equation summarized conditional probability. Essentially, we want to know the probability of A given B, or in other words, the probability of A occurring given that we know B has or will occur. If we consider a visual diagram, then A and B are areas in the sample space and $A \cap B$ is the overlap (what we are looking for). The denominator of the equation simply re-normalizes the probability since now our sample space is restricted to B (because we know B has or will occur).

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Multiplication Rule

These are the first two examples of the multiplication rule that can be repeated *ad nauseum*.

$$P(A \cap B) = P(A|B)P(B)$$

$$P(A \cap B \cap C) = P(A|BC)P(B|C)P(C)$$

Law of Total Probability

$$P(A) = \sum_{i=1}^{|B|} P(A|B_i)P(B_i)$$

Bayes' Rule

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{i=1}^{|B|} P(A|B_i)P(B_i)} = \frac{P(A|B_i)P(B_i)}{P(A)}$$