3.0 Lesson Plan

- Answer Questions
- Areas Under Normal Curves
- Probability Definitions
- Drawing from a Box
- Probability Rules
In the last lecture, we learned how to find areas under a standard normal distribution (mean $\mu = 0$, standard deviation $sd = 1$). This required use of the table in the book.

A region under a normal curve corresponds to a proportion of the population. This is because a normal curve can be viewed as the limit of a series of histograms, in which the sample gets large while the bin-size goes to zero.

Thus if a student’s arrival time in minutes for class is represented by a standard normal, then half the time the student arrives before class starts, and 68.27% of the time the student is within $\pm 1$ minute of the start of class.
We now show how to convert a question about an arbitrary normal distribution into an equivalent question about the standard normal, and vice-versa. Thus we can use the table to answer questions about all normal distributions, not just the standard normal.

Let $X$ be a random value from a normal population with mean $\mu$ and standard deviation $\sigma$. We write this as $X \sim N(\mu, \sigma)$.

Define a new random variable $Z = (X - \mu)/\sigma$. Then one can prove that $Z \sim N(0, 1)$.

This is called the $z$-transformation. To go the other way, we convert the standard normal value to an arbitrary normal distribution by solving for $X$. So $X = \mu + Z\sigma$. 
Reggie Jackson, the famous baseball player, has an IQ of 140. What percentage of people are smarter?
Assume that IQs are normally distributed with mean 100 and standard deviation 16.

We want the area under the normal distribution for IQ that lies to the right of 140. By the $z$-transformation, this is equivalent to the area under the standard normal distribution that lies to the right of

$$z = \frac{X - \mu}{\sigma} = \frac{140 - 100}{16} = 2.5.$$  

From the normal table, the area above 2.5 is 0.006. Thus about 0.6% of people are smarter than Reggie Jackson.
Now we go the other way. We find the $X$ value that corresponds to a given percentage.

To join Mensa one must be in the top 2% of the IQ distribution. What score do you need?

In the body of the normal table, look up 2%, or 0.02. That gives the $z$-value of approximately 2.05. So 2% of the area under the standard normal is above 2.05.

Now we use the inverse $z$-transformation:

$$X = \mu + Z\sigma = 100 + (2.05)(16) = 132.8.$$ 

One needs an IQ score of at least 132.8 (i.e., 133) to join.
3.2 Probability Definitions

There are two ways to define the probability of an event.

A frequentist says that the probability of event A (or $P[A]$) is the proportion of times that A occurs in an infinite sequence (or very long run) of separate tries.

Thus

$$P[A] = \lim_{n \to \infty} \frac{\# \text{ times A happens}}{n}.$$  

John Maynard Keynes (1883-1946) commented on this: In the long run, we are all dead.
A **Bayesian** can pick whatever number they prefer for $P[A]$, based on their own personal experience and intuition, provided that number is consistent with all of the other probabilities they choose in life.

A frequentist must define $P[\text{heads in a coin toss}]$ as the limit of the proportion of heads in $n$ tosses.

In contrast, a Bayesian might declare their personal belief about the coin based on symmetry, or knowledge of the integrity of the coin’s owner, or divine inspiration. The Bayesian’s view must:

- conform to all other personal opinions
- change as new data arise according to **Bayes’ Rule**.
Whether one is frequentist or Bayesian, all probabilities must obey Kolmogorov’s Axioms:

- \( 0 \leq P[A] \leq 1 \)
- \( P[ \text{some possible event happens} ] = 1 \) (one of the possible outcomes must occur).
- If \( A \) and \( B \) are incompatible (disjoint) events, then \( P[A \text{ or } B] = P[A] + P[B] \).

Two events \( A \) and \( B \) are **disjoint** if it is impossible for both \( A \) and \( B \) to happen at the same time; e.g., you cannot throw a head and tail on the same toss.

From these three rules, everything else can be derived.

Kolmogorov (1903-1987) was one of the greatest mathematicians of the 20th century. This axiomatization was a trivial accomplishment.
3.3 Drawing From A Box

Drawing tickets from a box is a convenient way to think about random outcomes. A box model requires three pieces of information:

- What are the tickets (numbers) in the box?
- How many draws does one take?
- Are the draws made with replacement?

What is the box model for 50 tosses of a (possibly unfair) coin?

What would be the box model for drawing a random sample of 50 students from STA 111 in order to learn about their GPA?

What would be the box model for drawing a random sample of 50 U.S. voters in order to learn their opinion of gun control laws?
These are the main definitions:

- $A$ and $B$ are **independent** iff $P[A \text{ and } B] = P[A] \times P[B]$.
- $A$ and $B$ are **disjoint** (or incompatible, or mutually exclusive) iff $P[A \text{ and } B] = 0$.
- A set of events $A_1, \ldots, A_n$ is a **finite partition** iff $A_i$ and $A_j$ are incompatible for all $i \neq j$ and $\sum_{i=1}^{n} P[A_i] = 1$.

These are the main rules:

- $P[\text{ not } A] = 1 - P[A]$: **(complement rule)**
- $P[A \text{ given that } B \text{ occurs }] \equiv P[A \mid B] = P[A \text{ and } B] / P[B]$: **(conditional probability)**
Let $A$ be the event that you draw a heart from a standard deck. Let $B$ be the event that you draw a red card.

- **Independence:** These are not independent, since 
  $$\Pr[A \text{ and } B] = 1/4 \neq 1/4 \times 1/2 = \Pr[A] \times \Pr[B].$$

- **Disjoint:** These are not incompatible; a card can be both red and a hear.

- **Partition:** Define the events $A_1$, a ♠; $A_2$, a ♦; $A_3$, a ♥; $A_4$ a ♣. This is one finite partition (among many) for a draw from a standard deck.

- **Complement:** The probability you don’t draw a heart is the 1 - 
  $$\Pr[\text{ heart }] = 1 - 1/4 = 3/4.$$  

- **Inclusive Or:** The probability of a red card or a heart is 
  $$\Pr[A \text{ or } B] = 1/4 + 1/2 - \Pr[\text{ red heart }] = 1/2.$$  

- **Conditional:** The probability of a heart given that the card is red is 
  $$\Pr[A \mid B] = \Pr[A \text{ and } B]/\Pr[B] = (1/4)/(1/2) = 1/2.$$
Each year, the U.S. DoD holds a draft lottery. Tickets with birthdays are placed in a box, and randomly drawn without replacement. This determines the order in which 18 year-old men would be drafted.

What is the probability that the first two draws are both from December? (Assume it is not a leap year.)

A. There are 365*364=132860 ways to draw two dates. Of these, 31*30=930 are in December. Thus the chance is 930/132860 = 0.007.

B. Let $A$ be the event that the second draw is in December, and $B$ be the event that the first is in December. Then

$$\Pr[A \text{ and } B] = \Pr[A \mid B]\Pr[B] = \frac{30}{364} \times \frac{31}{365} = 0.007.$$
1. You make two draws, without replacement, from a standard 52 card deck. What is the probability that the first card is a king \((B)\) and the second card is a queen \((A)\)?

\[
\mathbb{P}[A \text{ and } B] = \mathbb{P}[B] \times P[A|B] = \frac{4}{52} \times \frac{4}{51} = 0.00603.
\]

2. You roll a die. What is the probability of a six \((A)\), given that the result is greater than or equal to 3 \((B)\)?

\[
\mathbb{P}[A|B] = \frac{\mathbb{P}[A \text{ and } B]}{\mathbb{P}[B]} = \frac{(1/6)}{(4/6)} = 0.25.
\]

Note that here the conditioning is not sequential in time.
3. Roll a die twice. What is the probability of a six on the second throw (A) if the first throw was a six (B)?

Define: \( A = \) second throw a six; \( B = \) first throw a six. Then

\[
P[A|B] = \frac{P[A \text{ and } B]}{P[B]} = \frac{1/36}{1/6} = 1/6.
\]

Note: Independence of rolls is an assumption.

If \( P[A|B] = P[A] \), then \( A \) is independent of \( B \). This is equivalent to the previous definition. (Prove that if \( A \) is independent of \( B \), then \( B \) is independent of \( A \).)

When events are independent, the occurrence of one does not affect the chance of the occurrence of the other.
4. Consider a single draw from a deck of cards. Let $A$ be the event that the card is a king, and $B$ be the event that the card is a queen. What is the probability of $A$ or $B$? Clearly

$$P[A \text{ or } B] = P[A] + P[B] - P[A \text{ and } B] = \frac{4}{52} + \frac{4}{52} - 0 = \frac{2}{13}.$$ 

5. Consider a single draw from a deck of cards. Let $A$ be the event that the card is a king, and $B$ be the event that the card is red. What is the probability of $A$ or $B$?

$$P[A \text{ or } B] = P[A] + P[B] - P[A \text{ and } B] = \frac{4}{52} + \frac{26}{52} - \frac{2}{52} = \frac{7}{13}.$$
6. Consider **two** draws from a deck of cards **without** replacement. Let \( A \) be the event that the first card is a king, and \( B \) be the event that the second card is a king. What is the probability of \( A \) or \( B \)?

One approach is to find the probabilities of (King, non-King), (non-King, King), and (King, King). These events are disjoint, so summing the probabilities gives the answer.

But it is easier to recall the **complement rule** and find the probability of no Kings, then subtract this from 1.

\[
P[A \text{ or } B] = 1 - P[\text{ not } A \text{ and not } B]
\]
\[
= 1 - P[\text{ not } B | \text{ not } A] \times P[\text{ not } A]
\]
\[
= 1 - \frac{47}{51} \times \frac{48}{52} = 0.149.
\]

Hard calculations can often be simplified if one knows the tricks.