Today we will

1) Define joint, conditional, and total probabilities.
2) See how Bayes' theorem helps to compute conditional probabilities.
3) Discuss Bernoulli trials

Suggested reading material

section 1.6 Joint, cond., total probabilities
section 1.7 Bayes' theorem
section 1.9 Bernoulli trials

You will find section 1.8 (combinatorics) useful for review; read it if you haven't seen combinatorics in a while.

We will discuss section 1.10 (Poisson law) later in the course.
And section 1.11 is more easily read once you know about the central limit theorem.
joint, conditional, & total probability

Example
A - event that temperature > 50°F.
B - event that it's snowing.
C - event that A ∩ B occurs (it's above 50°F and snowing).

You probably agree that
Pr(A) > 0  [maybe not in August?]
Pr(B) > 0  [barely - we're in Raleigh]
Pr(C) > 0

Pr(C|A) = Pr(A ∩ B) is called the joint probability of events A and B.
The events must be defined over the same probability space; they're subsets of F.

We note that Pr(A) > 0 conditioned on the fact that event B happened.
We write
Pr(A|B) = \frac{Pr(A ∩ B)}{Pr(B)}

this is called conditional probability.

Joint probability can also be written as Pr(A ∩ B) or Pr(A, B).
Active learning assignment

A - event that person has purple eyes.
B - event that person gets A in ECES14.

\[ P(A \cap B) = 0.05 \]
\[ P(A \cap B^c) = 0.01 \]
\[ P(A^c \cap B) = 0.44 \]
\[ P(A^c \cap B^c) = 0.50 \]

(1) Please compute \( P(A) \), \( P(A^c) \), \( P(B) \), \( P(B^c) \).

(2) Please compute \( P(B \mid A) \) and \( P(B \mid A^c) \).

Conclusion - having purple eyes is not merely unique; it's good for your grades!
Two events \( A, B \in F \) are independent if \[ \Pr(A \cap B) = \Pr(A) \cdot \Pr(B) \. \]

Note the definition requires \( \Pr(A), \Pr(B) > 0 \), else independence is not defined.

**Active learning assignment**

Explain why \( A, B \) from previous page are dependent. (not independent)

Note if they were independent, then purple eyes wouldn't help grades 😊

If \( A, B \) are independent, then

\[ \Pr(A \mid B) = \Pr(A), \]
\[ \Pr(B \mid A) = \Pr(B). \]

That is, conditioning each event on the other would have no effect.
Three events are jointly independent if
\[ \Pr(A \cap B \cap C) = \Pr(A) \cdot \Pr(B) \cdot \Pr(C) \]
\[ \Pr(A \cap B) = \Pr(A) \cdot \Pr(B) \]
\[ \Pr(A \cap C) = \Pr(A) \cdot \Pr(C) \]
\[ \Pr(B \cap C) = \Pr(B) \cdot \Pr(C) \]

That is, in addition to *, we require pairwise independence.
Joint independence of more than 3 events described in book. All combinations must be independent.

Useful fact
If \( E_1, E_2 \) independent, then \( E_1, E_2^c \) independent.

Proof: We know \( \Pr(E_1 \cap E_2^c) = \Pr(E_1) \cdot \Pr(E_2^c) \)
Now let's compute \( \Pr(E_1 \cap E_2^c) \).

\[ \Pr(E_1 \cap E_2) = \]
Law of total probability

Let $E_1, E_2, \ldots, E_n$ be $n$ mutually exclusive events such that
\[ \cap_{i=1}^n E_i = \emptyset. \]

Let $F$ be an event defined over the same probability space.

\[
\Pr(F) = \Pr(F \cap E_1) \cdot \Pr(E_1) + \cdots + \Pr(F \cap E_n) \cdot \Pr(E_n)
= \Pr(F \cap E_1) = \Pr(F \cap E_n)
\]

Example

In the active learning exercise, we should have $\Pr(A) = \Pr(A \mid B) \cdot \Pr(B) + \Pr(A \mid B^c) \cdot \Pr(B^c)$.

Let's verify that numerically:
Example (problem 1.32, Stark & Woods)
(Ternary communication channel)

\[ P_r(Y=1|X=1) = 1 - \alpha \]

We also know \( P_r(X=3) = 3 \cdot P_r(X=1) \)
and \( P_r(X=2) = 2 \cdot P_r(X=1) \)

A 1 was observed. What's the conditional probability that 1 was transmitted?

Answer

\[
P_r(X=1) \cdot (1 + 2 + 3) = 1
\]

\[
\Rightarrow P_r(X=1) = \frac{1}{1 + 2 + 3} = \frac{1}{6}
\]

\[
\Rightarrow P_r(X=2) = \frac{2}{6}, \quad P_r(X=3) = \frac{3}{6}
\]

\[
P_r(X=1|Y=1) = \frac{P_r(X=1) \cap (Y=1)}{P_r(Y=1)}
\]

\[
= \frac{P_r(X=1) \cdot P_r(Y=1|X=1)}{P_r(X=1) \cdot P_r(Y=1|X=1) + P_r(X=2) \cdot P_r(Y=1|X=2) + P_r(X=3) \cdot P_r(Y=1|X=3)}
\]

\[
= \frac{\frac{1}{6} (1 - \alpha)}{\frac{1}{6} (1 - \alpha) + \frac{1}{5} \beta + \frac{1}{2} \gamma}
\]

Bayes' theorem = definitions + total expectation
Bayes Theorem

Thomas Bayes (1702 - 1761)
(no painting of him was made during his lifetime)

He was elected to the Royal Society after publishing a book on fluxions (calculus).
Notes of his that contained some derivations in probability were published after his death.

Theorem

Let $E_i$, $i = 1, \ldots, n$ be disjoint and exhaustive events defined on probability space $\mathcal{P}$. Let $F$ be another event such that $\Pr(F) > 0$ and $\Pr(E_i) > 0$ for all $i$. Then

$$
\Pr(E_j | F) = \frac{\Pr(F \cap E_j) \Pr(E_j)}{\sum_{i=1}^{n} \Pr(F \cap E_i) \Pr(E_i)}
$$

Proof:

* the denominator comes from law of total probability $\Pr(F)$
* numerator is $\Pr(E_j \cap F)$
* $\Pr(E_j \cap F)$ is definition of conditional probability $\square$
The last example that we saw already looked like Bayes' rule. How about another example?

Active learning assignment
(adapted from Stark & Woods, edition 4, p. 37)
Pr(protein | disease) = 0.9
Pr(protein | no disease) = 0.36
Pr(disease) = 0.1
Is testing for this protein a "good" test?
Another Example

* An electrical circuit connects the input and output when there is some path with connectivity.
* Component \( C_i \) fails with probability \( P_i \).
* Failures are independent.
* We want to compute \( \Pr(\text{connectivity}) \).

Approach 1 - Compute probabilities of all 25 combinations and add up probabilities of good outcomes.

Approach 2. (Divide and conquer)

The crucial component is \( C_3 \).

\[
\Pr(\text{connect}) = \Pr(\text{connect} | C_3 \text{ good}) \cdot (1 - P_3) \\
+ \Pr(\text{connect} | C_3 \text{ bad}) \cdot P_3
\]

and computing the two conditional probabilities isn't too hard.
Bernoulli trials
(named after Jacob Bernoulli, 1654-1705.)

We begin with a single trial that has a binary outcome, e.g., flipping a coin. If we denote the possible outcomes by \( s \) (success) and \( f \) (failure), \( S = \{s, f\} \) and \( \mathcal{F} = \{\emptyset, S, \{s\}, \{f\}\} \).

With \( n \) Bernoulli trials,
\[
S_n = S \times S \times \cdots \times S,
\]
n times

this is a cartesian product taken \( n \) times.

\( S_n \) has \( 2^n \) outcomes,
\[
S_n = \{a_1, \ldots, a_m\},
\]
where \( m = 2^n \),
and
\[
a_i = z_{i1} z_{i2} \cdots z_{in},
\]
and
\[
z_{ij} \in \{s, f\}, \quad j \in \{1, \ldots, n\}.
\]

Consider \( \alpha \) with \( k \) successes and \( n-k \) failures.
\[
\Pr(\alpha) = p^k (1-p)^{n-k}.
\]
Example (Ex. 1.9-1 in edition 4, p. 49) we toss a coin 3 times. Each time \( \Pr(H) = p \).

Consider the events with two heads and one tail:

\[ E_1 = \{ HHT \}, \quad E_2 = \{ HTH \}, \quad E_3 = \{ THH \}. \]

Let \( F = E_1 \cup E_2 \cup E_3 \), the event of getting (somehow) two heads and one tail. Each of these events, \( E_1 \), \( E_2 \), and \( E_3 \), has probability \( p^2 (1-p) \). Therefore, \( \Pr(F) = 3 \cdot p^2 (1-p) \).

More generally, there are \( \binom{n}{k} \) outcomes with \( k \) successes and \( n-k \) failures (alternately \( n-k \) successes and \( k \) failures), where

\[
\binom{n}{k} = \frac{n!}{k! (n-k)!}
\]

is the "choose" function from combinatorics (see section 1.8 for a review).

Seeing that each such outcome has probability \( p^k (1-p)^{n-k} \), we have

\[
\Pr(\text{ \( k \) successes}) = \binom{n}{k} p^k (1-p)^{n-k}.
\]
Active learning exercise

A random student fills out random answers on a true/false test. The probability of getting an answer right is 50%.

1. There are 5 questions. What is the probability of getting exactly 3 right?

2. At least 3?

3. Enumerate the student's results writing Right or Wrong. For example RWRRR means only question #2 was wrong. Out of $2^5 = 32$ outcomes, write down all those such that less than 3 were right.