OUTLINE

① Probability Spaces & Probability Functions
   ① Example: Rolling a Die
   ② Conditional Probability
   ③ Independent Events

② Bayesian Statistics
STATISTICS ≠ PROBABILITY THEORY

**Probability theory**: mathematical theory that describes uncertainty.

**Statistics**: set of techniques for extracting useful information from data.
PROBABILITY SPACE

A probability space has three components:

1. A sample space $\Omega$ which is the set of all possible outcomes of the random process modeled by the probability space;
2. A family of sets $F$ representing the allowable events, where each set in $F$ is a subset of the sample space $\Omega$;
3. A probability function $\Pr : F \rightarrow \mathbb{R}$ satisfying the definition below:

An element of $\Omega$ is a simple event. In a discrete probability space, we use $F = 2^\Omega$
A **probability function** is any \( \Pr : F \rightarrow R \) that satisfies the following conditions:

1. For any event \( E \), \( 0 \leq \Pr(E) \leq 1 \);
2. \( \Pr(\Omega) = 1 \);
3. For any finite or countably infinite sequence of pairwise mutually disjoint events \( E_1, E_2, E_3, \ldots \)

\[
\Pr \bigcup_{i=1}^{\infty} E_i = \sum_{i=1}^{\infty} \Pr(E_i)
\]

The probability of an event is the sum of the probabilities of its simple events.
EXAMPLE: TOSSING A (FAIR) COIN

\[ F = \{ H, T \} \]

\[ F = 2 \times 2 = 4 \text{ Events} \]

\[ F = \{ \emptyset, \{ H \}, \{ T \}, \{ H, T \} \} \]

\[ \Pr(\emptyset) = 0 \]

\[ \Pr(\{ H \}) = 0.5 \]

\[ \Pr(\{ T \}) = 0.5 \]

\[ \Pr(\{ H, T \}) = 1 \]
EXAMPLE: ROLLING A DIE

\[ W = \{1, 2, 3, 4, 5, 6\} \]

\[ F = 2^6 \text{ Events} \]

\[ \Pr(\{\ \}) = 0 \]

\[ \Pr(\{1\}) = \Pr(\{2\}) = \Pr(\{3\}) = \Pr(\{4\}) = \Pr(\{5\}) = \Pr(\{6\}) = \frac{1}{6} \]

\[ \Pr(\{1, 2\}) = \Pr(\{1, 3\}) = \Pr(\{1, 4\}) = \Pr(\{1, 5\}) = \Pr(\{1, 6\}) = \frac{2}{6} \]

...
The conditional probability that event E1 occurs given that event E2 occurs is:

$$\Pr(E_1 \mid E_2) = \frac{\Pr(E_1 \cap E_2)}{\Pr(E_2)}$$

The conditional probability is only well-defined if $\Pr(E_2) > 0$

By conditioning on E2 we restrict the sample space to set E2.

Thus we are interested in $\Pr(E_1 \cap E_2)$ normalized by $\Pr(E_2)$. 
EXAMPLE: CONDITIONAL PROBABILITY

We have two coins: A is a fair coin, B has probability 2/3 to come up as HEAD. We chose a coin at random and got HEAD.

What is the probability that we chose coin A?

① \(E_1\) = the event “chose coin A”
② \(E_2\) = the event “outcome is HEAD”

Conditional probability that we chose coin A given that the outcome is HEAD is denoted: \(Pr(E_1 | E_2)\)
EXAMPLE: CONDITIONAL PROBABILITY

Define a **sample space** of **ordered** pairs: \((\text{coin}, \text{outcome})\)

The same space has four points:

1. \{\((A, h), (A, t), (B, h), (B, t)\)\}
2. \(\Pr((A, h)) = \Pr((A, t)) = \frac{1}{4}\)
3. \(\Pr((B, h)) = (\frac{1}{2})(\frac{2}{3}) = \frac{1}{3}\)
4. \(\Pr((B, t)) = \frac{1}{2} \times \frac{1}{3} = \frac{1}{6}\)

Define 2 events:

1. \(E_1 = \text{“chose coin } A\”\)
2. \(E_2 = \text{“outcome is HEAD”}\)

\[
\Pr(E_1 \mid E_2) = \frac{\Pr(E_1 \cap E_2)}{\Pr(E_2)} = \frac{\frac{1}{4}}{\frac{1}{4} + \frac{1}{3}} = \frac{3}{7}
\]
Two events $E$ and $F$ are independent if and only if:

$$\Pr(E_1 \cap E_2) = \Pr(E)\Pr(F)$$

More generally, events $E_1, E_2, \ldots, E_k$ are mutually independent if and only if for any subset $I \subseteq [1, k]$:

$$\Pr\left[ \bigcap_{i \in I} E_i \right] = \Pr(E_i)$$
EXAMPLE: TOSSING A (FAIR) COIN TWICE

\[ = \{HH, HT, TH, TT\} \]

\[ F = 2 \quad = 2^4 \text{ Events} \]

\[ P(H) = 0.5 \]
\[ P(T) = 0.5 \]

\[ P(HT) = 0.25 \]
\[ P(HH) = 0.25 \]
\[ P(TH) = 0.25 \]
\[ P(TT) = 0.25 \]

\[ P(\{HH, HT\}) = P(\{HT, TT\}) = 0.5 \]
A fair coin was tossed 10 times and always ended up on **HEAD**. What is the likelihood that it will end up **TAIL** next?
A fair coin was tossed 10 times and always ended up on **HEAD**. What is the likelihood that it will end up **TAIL** next?

The prior observations don’t affect the likelihood. → 1/2
Stan has two kids. One of his kids is a boy. What is the likelihood that the other one is also a boy?

\[ E_1 = \text{Two Boys (BB)} \]
\[ E_2 = \text{At least one kid is a boy (B)} \]
Stan has two kids. One of his kids is a boy. What is the likelihood that the other one is also a boy?

\[ E_1 = \text{Two Boys (BB)} \]
\[ E_2 = \text{At least one kid is a boy (B)} \]

\[ P(BB) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4} \]

\[ P(B) = 1, \quad P(GG) = 1 \quad \frac{1}{2} \times \frac{1}{2} = \frac{3}{4} \]

\[ P(BB|B) = \frac{P(BB \cap B)}{P(B)} = \frac{1/4}{3/4} = \frac{1}{3} \]
BAYESIAN STATISTICS
**LAW OF TOTAL PROBABILITY**

**Theorem: Law of Total Probability**

Let $E_1, E_2, \ldots, E_n$ be mutually disjoint events in a sample space $\Omega$, and $\bigcup_{i=1}^{n} E_i = \Omega$.

Then: $\Pr(B) = \sum_{i=1}^{n} \Pr(B \mid E_i) \Pr(E_i)$.
BAYES’ LAW

Theorem: Bayes’ Law

Let \( E_1, E_2, \ldots, E_n \) be mutually disjoint events in a sample space \( \Omega \), and

\[
\bigg\{ E_i \bigg\} = \bigcup_{i=1}^{n} E_i
\]

Then:

\[
Pr(E_j \mid B) = \frac{Pr(B \cap E_i)}{Pr(B)} \cdot \frac{Pr(B \mid E_j)Pr(E_j)}{\sum_{i=1}^{n} Pr(B \mid E_i)Pr(E_i)}
\]

Conditional Probability: \( Pr(A \mid B) = \frac{Pr(A \cap B)}{Pr(B)} \)

Law of Total Probability: \( Pr(B) = \sum_{j=1}^{n} Pr(B \mid E_j)Pr(E_j) \)
BAYES’ LAW

**Likelihood**
Probability of collecting this data when our hypothesis is true

**Prior**
The probability of the hypothesis being true before collecting data

**Posterior**
The probability of our hypothesis being true given the data collected

**Marginal**
What is the probability of collecting this data under all possible hypotheses?

\[
P(H|D) = \frac{P(D|H) P(H)}{P(D)}
\]
APPLICATION: FINDING A BIASED COIN

- We are given three coins. 2 coins are fair, and the 3rd is biased (landing heads with probability \(\frac{2}{3}\)). We need to identify the biased coin.
- We flip each of the coins. The first and second come up heads, and the third comes up tails.
- What is the probability that the first coin was the biased one?
APPLICATION: FINDING A BIASED COIN

Let $E_i$ be the event that the $ith$ coin flip is the biased one and let $B$ be the event that the three coin flips came up HEADS, HEADS, and TAILS. Before we flip the coins we have $\Pr(E_i) = 1/3$ for $i=1,...,3$, thus

$$\Pr(B \mid E_1) = \frac{2}{3} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{6}$$

and

$$\Pr(B \mid E_3) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{12}$$

Applying Bayes' Law we have...

$$\Pr(E_1 \mid B) = \frac{\Pr(B \mid E_1)(\Pr(E_1)}{\sum_{i=1}^{3} \Pr(B \mid E_i)\Pr(E_i)} = \frac{2}{5}$$

The outcome of the 3 coin flips increases the probability that the first coin is the biased one from $1/3$ to $2/5$. 
Stan has two kids. One of his kids is a boy. What is the likelihood that the other one is also a boy?

$E_1 =$ Two Boys (BB)

$E_2 =$ At least one kid is a boy (B)

$P(\text{BB}) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$

$P(\text{B}) = 1$ $P(\text{GG}) = 1$ $\frac{1}{2} \cdot \frac{1}{2} = \frac{3}{4}$

$P(\text{BB}|\text{B}) = \frac{P(\text{BB} \cap \text{B})}{P(\text{B})} = \frac{1/4}{3/4} = \frac{1}{3}$

$P(\text{BB}|\text{B}) = \frac{P(\text{B}|\text{BB}) \cdot P(\text{BB})}{P(\text{B})} = \frac{1 \cdot 1/4}{3/4} = \frac{1}{3}$
IN CLASS EXERCISES: DRUG TEST

0.4% of the Rhode Island population use Marijuana*

Drug Test: The test will produce 99% true positive results for drug users and 99% true negative results for non-drug users.

If a randomly selected individual is tested positive, what is the probability he or she is a user?

\[
P(User|+) = \frac{P(+|User)P(User)}{P(+)}
\]

\[
= \frac{P(+|User)P(User)}{P(+|User)P(User) + P(+|!User)P(!User)}
\]

\[
= \frac{0.99 \times 0.004}{0.99 \times 0.004 + 0.01 \times 0.996}
\]

\[
= 28.4\%
\]
SPAM FILTERING WITH NAÏVE BAYES

\[
P(\text{spam} | \text{words}) = \frac{P(\text{spam}) P(\text{words} | \text{spam})}{P(\text{words})}
\]

\[
P(\text{spam} | \text{viagra, rich, ..., friend}) = \frac{P(\text{spam}) P(\text{viagra, rich, ..., friend} | \text{spam})}{P(\text{viagra, rich, ..., friend})}
\]

\[
P(\text{spam} | \text{words}) \quad \frac{P(\text{spam}) P(\text{viagra} | \text{spam}) P(\text{rich} | \text{spam}) \cdots P(\text{friend} | \text{spam})}{P(\text{viagra, rich, ..., friend})}
\]
WARM UP QUESTION

- Assume a statistical test has a chance of 1% (P=0.01) to be wrong
- How many test can you run before the likelihood of being wrong at least once is 50% or more?

\[
P(T_1) = 1 \quad P(F_1) = 0.99
\]
\[
P(T_1T_2) = 0.99 \quad 0.99 \quad 0.98
\]
\[
P(F_1F_2,F_1T_2,T_1F_2) = 1 \quad P(T_1T_2) = 0.02
\]

\[
P(\text{Being Wrong At Least Once}) =
1 \quad P(T_1T_2 \sqcup T_n \quad T_n) = 1 \quad P(T)^n
\]
\[
n = \log(0.5) / \log(0.99) \quad 68.96 \quad 69
\]