Review of Course Outline

Basics

- Class web page:

  http://www.stat.uiowa.edu/~luke/classes/193

  Icon (icon.uiowa.edu) may be used for making grades available.

- Review course outline.

- Exam schedule:
  - Tentative midterm schedule: In class,
    - October 2
    - November 6.
  - The final exam schedule will be announced during the semester.
Review of Course Outline

Reading

- Classes will roughly follow the book outline but may deviate at times.
- Reading assignments will be posted on the web.
- You should read ahead.
Homework

- Working together is OK.
- But try to work on your own as much as you can.
- Your write-up must be your own.
- Working the problems is the important thing, not getting the right answer.
- Few problems worked in class, only simple examples.
- Some homework problems are time-consuming and/or quite difficult.
- Learning how to approach difficult problems is an important skill.
- Solutions should be clear and well written:
  - first drafts are not acceptable;
  - solutions should be described in full and clear sentences
  - use the examples in the text as a style guide.
Prerequisites and Background

- Everyone’s background is different.
- A stronger math background will make things easier.
  - Review your calculus if it has been a while since you have used it.
  - Some math review notes and problems are available at
  - Some useful tutorials are available at
    http://tutorial.math.lamar.edu/.
- A basic mathematical statistics course is a formal prerequisite.
  - This isn’t strictly necessary as this course is self-contained.
  - You may have to work harder at times if you have not had one.
- Do the best you can; don’t worry about how others are doing.
- If you find you are having difficulties please talk to me!
Relation to Other Things

- This course is primarily for Statistics and Biostatistics graduate students.
- I may occasionally make reference to other courses.
- I may make reference to colloquia—you should attend the colloquia.

Ask Questions

- Ask questions if you are confused or think a point needs more discussion.
- If you don’t, I’ll assume you understand everything and will probably speed up!

Problem Session

- The grader will be offering a problem session.
- The grader will contact you by email to identify a time that works for as many of you as possible.
Introduction

What is Statistics About?

• Learning from data
  • You may have a lot of data (Big Data)
  • Or you may have a small amount of data

• Most important data: the data you do not have
  • Location of storm in two hours
  • Data on customers of your competitors

• Another aspect: Making decisions in the presence of uncertainty.
  • ...

---

Examples

- Polling:
  - Poll 1000 people, ask who favors X.
  - Want to know what percent of the population favors X.

- Drug testing:
  - Have a new drug, test on 50 people.
  - Want to know if it would work on others.

- Measurements:
  - Measure the melting point of a substance.
  - How well do we know the melting point?

- Forecasting:
  - Predict tomorrow’s high temperature.
  - How accurate is the prediction?
Common Features

- Unknown “state” that is of interest.
  - population proportion in favor of X
  - true melting point
- Understanding of how the data are produced
  - sampling mechanism
  - measurement process
- Can assess how likely data are given “true” state.
  - model for sampling mechanism
  - measurement error model
- Can use this to assess support for different states given data.
Mathematical Statistics

- This course is about the mathematical framework and tools used for studying and developing statistical models and methods.
- To do this carefully, we need a mathematical framework for
  - formulating how likely things are
  - quantifying variability/uncertainty
- Probability theory provides this framework.
- We will spend most of the semester on probability.
Basics of Probability

- Probability is a framework for
  - quantifying uncertainty
  - reasoning about uncertainty
- Some probability statements you may be familiar with:
  - Roll a die: chance, probability of rolling a 5 is 1/6
  - Weather forecasting:
    - chance, probability of rain tomorrow is 25%.
    - chance, probability, that the high temperature tomorrow is at least 80°F is 75%.
- Probability is a fraction, or a percentage
  - Certain: 1 or 100%.
  - Impossible: 0 or 0%.
- How do we think about numbers in between?
Relative Frequencies

- One way to think about this:
  - Roll a die many times, see how often 5 occurs.
  - Expect each side to come up roughly equally often.
  - Relative frequency of 5 should be about 1/6.

- Probabilities are often thought of as limiting/long run relative frequencies.

- This view is sometimes called “objective.”
Subjective Probability

- Probability can also be used when repeating many times is not possible.
- Probability then provides a subjective quantification of uncertainty.

Example
What is the probability that the Democrats will win the next presidential election?

Example
What is the probability that a particular grant application will be funded?

Example
What is the probability that tomorrow’s high temperature will be at least 80°F?

Example
What is the probability that the high temperature one month ago was at least 80°F?
Are there any objective probabilities?

- No one actually rolls a die many times to determine them.
  - Almost no one: The Swiss astronomer Robert Wolf rolled dice 100,000 times and recorded the results.
- Objective probabilities are usually based on statements like:
  - Assume all six sides are equally likely (or: the die is “fair”)
  - I believe all six sides are equally likely

Regardless of interpretation, the mathematics are the same.
**Example**

- Yesterday’s forecast high temperature for today was $72^\circ$F.
- What is the chance that the actual high is at least $77^\circ$F.
- A simple approach: Look at prediction errors over recent months.

- An estimate of the probability is
  
  ```r
  > mean(true - pred >= 5)
  [1] 0.07142857
  ```

- The errors provide a framework for replication/repeatability.
Some Terminology

- Probability is used to quantify uncertainty about things that may or may not happen.
- We call these things *events*.
  - The event that it rains tomorrow.
  - The event that the unemployment rate next month is below 7.5%.
- We often deal with numerical quantities where the value is not known:
  - Tomorrow’s high temperature.
  - Next month’s unemployment rate.
  - The amount of rainfall that will be recorded tomorrow.
  - The number of fatal car accidents there will be this year in Iowa.
  - The number of axles on the next vehicle to enter a bridge.
- These are called *random variables*. 
Random Variables and Distributions

• It is useful to think about
  • the possible values of a random variable
  • how likely these possible values are
• This is called the \textit{probability distribution} of the random variable.
• A random variable is called \textit{discrete} if it has a finite or countable set of possible values.
  • number of traffic fatalities
• A random variable that has a continuous range of possible values, with no values particularly distinguished, is called \textit{continuous}.
  • measurements of temperature, time, weight are usually thought of as continuous.
  • sometimes these are \textit{discretized} by rounding.
• There are combinations: a continuous range and one or more special values
  • rainfall: none, or some continuous positive amount
Visualizing and Summarizing Distributions

- Discrete distributions can be shown as a table or a probability histogram:

<table>
<thead>
<tr>
<th>Axles</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.75</td>
</tr>
<tr>
<td>3</td>
<td>0.13</td>
</tr>
<tr>
<td>4</td>
<td>0.08</td>
</tr>
<tr>
<td>5</td>
<td>0.04</td>
</tr>
</tbody>
</table>

- Continuous distributions are usually shown as a probability density
Visualizing and Summarizing Distributions

- Center, or location, of a distribution can be summarized by
  - average, mean
  - median

- Spread, variability of a distribution can be summarized by
  - inter-quartile range
  - mean or median absolute deviation from the center
  - the standard deviation, or root-mean-square deviation from the mean.
Some Observed Phenomena

- As an experiment is repeated the relative frequency of an event seems to converge.

- An average of several independently repeated measurements is more accurate than a single measurement.
Some Observed Phenomena

- Often we are interested in the *joint distribution* of several random variables.
- Galton’s data on heights of fathers and their sons:

![scatter plot](image)

- There is an association: taller parents tend to have taller children.
Some Observed Phenomena

- Looking at taller fathers with height around 71 inches:
  
  ```
  > g71 <- subset(galton, 70 <= parent & parent <= 72)
  > with(g71, mean(parent))
  [1] 70.87472
  > with(g71, mean(child))
  [1] 69.7864
  ```

- For shorter fathers with height around 65 inches:
  
  ```
  > g65 <- subset(galton, 64 <= parent & parent <= 66)
  > with(g65, mean(parent))
  [1] 65.13953
  > with(g65, mean(child))
  [1] 66.40778
  ```

- This is sometimes called the *regression effect*, or *regression towards the mean*. 
Recap

- Probability: quantifying uncertainty, variability
- Events and random variables
- Probability distributions
- Types of random variables
- Some observed phenomena
Some Probability Questions

• If a die is rolled 10 times, what is the probability that at least four rolls are sixes?
• If 1000 Iowa households are sampled and asked to report their household income, how likely is it that the sample median income will be within $1,000 of the population median income?
• Given the forecasts available, what is the chance that the maximum temperature over the next four days will be below 80°F?
• Suppose 10% of the population has a disease. A test for the disease has a 5% error rate. If a person tests positive, how likely is it that the person has the disease?
• Probabilities of basic events are often easy to determine or to estimate from data.
  • chance of a six on a single roll of a die
  • population prevalence of a disease

• We need a set of tools to help us compute probabilities of more complicated events from simpler ones.

• We will often need to make modeling assumptions in the process.

• A formal mathematical framework for probability can help us with this.
Mathematical Formulation of Probability

• A typical probability statement:
  \textit{The probability that it will rain today is 25\%}

• A natural notation is $P(\text{Rain}) = 1/4$.

• Probability is a \textit{function}
  \[ P : \mathcal{B} \rightarrow \mathcal{R} \]
  with some domain $\mathcal{B}$ and range $\mathcal{R}$.

• The range is the unit interval $\mathcal{R} = [0, 1]$.

• What is the domain $\mathcal{B}$?

• What are the key properties of $P$? E.g.
  \[ P(A) = 1 - P(\text{not } A) \]

• This means: if $A$ is in the domain $\mathcal{B}$, so is “not $A$”
**Primitive concept: Experiment**

An experiment is a procedure that can result in one and only one of several possible outcomes.

**Examples**

- roll a die, see which face is up
- see if it rains tomorrow
- poll 1000 people, ask how many favor X
**Definition**

The set of all possible outcomes of an experiment is called the *sample space* of the experiment.

Common notations: $S$, $\Omega$.

**Definition**

An *event* is a subset of the sample space.

**Definition**

A *random variable* is a real-valued function defined on the sample space.
Examples

- Roll a die, record the number of eyes.
  - $S = \{1, 2, 3, 4, 5, 6\}$
  - $X = X(s) =$ number of eyes $= s$.
  - $A =$ event that the number of eyes is even $= \{2, 4, 6\}$.

- A coin is flipped twice and the number of heads is recorded.
  - One possible sample space is $S = \{HH, HT, TH, TT\}$.
  - $A =$ event that the two results are different $= \{HT, TH\}$.
  - The random variable $X =$ number of heads is
    \[
    \begin{array}{c|cccc}
      s & HH & HT & TH & TT \\
      \hline
      X(s) & 2 & 1 & 1 & 0
    \end{array}
    \]

- An alternative formulation:
  - $S = \{0, 1, 2\}$.
  - $A = \{1\}$.
  - $X(s) = s$.

- There are often several choices for the sample space.
Examples (continued)

- Poll 1000, ask how many favor a particular issue.
- One possible formulation:
  - Sample space: $S = \{0, 1, \ldots, 1000\}$.
  - Number in favor: $X(s) = s$.
- Alternative:
  - $S = \{\text{all possible samples of 1000 from the population}\}$.
  - $X(s)$ is the number in favor in sample $s$.
- The first sample space is much smaller.
- The second is easier to use in formulating a probability model.
- The random variable $X$ for the second can be seen as a mapping from the second to the first.
Examples (continued)

- Flip a coin until the first head occurs:
  - \( S = \{H, TH, TTH, TTTTH, \ldots \} \).
  - \( X = \text{number of flips until first head} \).
    
    \[
    \begin{array}{c|cccc}
    s & H & TH & TTH & \ldots \\
    \hline
    X(s) & 1 & 2 & 3 & \ldots \\
    \end{array}
    \]
  - \( Y = \text{number of tails until first head} = X - 1 \).
  - This sample space is countably infinite.

- Measure the melting point of a substance:
  \[
  S = [\text{low}, \text{high}] \\
  S = \mathbb{R} = (-\infty, \infty)
  \]
  - These are continuous ranges and are uncountable sets.
Some Terminology for Events and Outcomes

- Suppose $A$ is an event, a subset of a sample space $S$.
- The experiment is performed and an outcome $s$ is observed.
- We say that the event $A$ has *occurred* if $s \in A$.
- If $s \notin A$ then $A$ has *not occurred*. 
Example

- Experiment: Roll die,

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

- Events:

\[ A = \text{Even Roll} = \{2, 4, 6\} \]
\[ B = \text{Roll} \leq 3 = \{1, 2, 3\} \]

- Observe \( s = 4 \).
  - \( A \) has occurred.
  - \( B \) has not occurred.

- Observe \( s = 1 \).
  - \( A \) has not occurred.
  - \( B \) has occurred.
Logical Operations on Events

- More complicated events can be built up from simpler ones.
- Suppose $A$ and $B$ are events.
- We can consider the event that either $A$, or $B$, or both occurs.
- We can consider the event that $A$ and $B$ both occur.
- Or we can consider the event that $A$ does not occur.
- These logical operations correspond to simple set operations.
Simple Set Operations

- Most of this should be very familiar.

- Union, inclusive ‘or’
  \[
  A \cup B = \{ x \in S : x \in A \text{ or } x \in B \text{ or both} \}
  \]

- Intersection, ‘and’
  \[
  A \cap B = \{ x \in S : x \in A \text{ and } x \in B \}
  \]

- Complement, ‘not’
  \[
  A^c = \{ x \in S : x \notin A \}
  \]

- Empty set, impossible event: \( \emptyset = \{ \} \).
Mutual Exclusion, Two Events

- $A$ and $B$ are mutually exclusive, or disjoint, if $A \cap B = \emptyset$.
- Mutually exclusive events cannot both occur.

Example

Roll a die,

$$A = \text{roll is even} = \{2, 4, 6\}$$
$$B = \text{roll is odd} = \{1, 3, 5\}$$
$$C = \text{roll } \leq 3 = \{1, 2, 3\}$$

- $A$ and $B$ are mutually exclusive
- $A$ and $C$ are not
Mutually Exclusive, or Pairwise Disjoint, Collections

- \{A_\alpha\} are mutually exclusive, or pairwise disjoint, if \(A_\alpha \cap A_\beta = \emptyset\) for all \(\alpha \neq \beta\).
- If \{A_\alpha\} are mutually exclusive, then at most one of them can occur.

Example

A coin is flipped until the first head occurs. Let

\[A_i = \text{the first head occurs on the } i\text{-th flip}\]

The \(A_i\) are mutually exclusive.
Rules for Combining Events

For any events $A$, $B$, $C$ on a sample space $S$:

**Commutative Laws**

$A \cup B = B \cup A$

$A \cap B = B \cap A$

**Associative Laws**

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

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

**Distributive Laws**

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

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

**De Morgan’s Laws**

$(A \cup B)^c = A^c \cap B^c$

$(A \cap B)^c = A^c \cup B^c$
Recap

- Sample spaces and events
- Logical operations on events
- Basic set theory
Verifying Set Identities

Suppose we want to verify \((A \cup B)^c = A^c \cap B^c\).

*Informal Approach: Use Venn diagrams*
Formal Approach: Double inclusion argument

• Suppose $x \in (A \cup B)^c$. Then $x \notin A$ and $x \notin B$, i.e. $x \in A^c$ and $x \in B^c$. So $x \in A^c \cap B^c$. Since $x$ was arbitrary, this holds for any $x \in (A \cup B)^c$, and thus $(A \cup B)^c \subset A^c \cap B^c$.

• Now suppose $x \in A^c \cap B^c$. Then $x \in A^c$, i.e. $x \notin A$, and similarly $x \notin B$. So $x \notin A \cup B$, i.e. $x \in (A \cup B)^c$. So $A^c \cap B^c \subset (A \cup B)^c$.

• Together these two containments imply that $(A \cup B)^c = A^c \cap B^c$.

• You can usually avoid working at this level by using higher level results like De Morgan’s laws.
Formal Approach: Using Already Established Identities

For any sets $A$, $B$, and $C$

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

Proof.

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

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

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

$= A \cup (B \cap C)$$
Unions, Intersections of Collections

Because of associativity, unions and intersections extend to collections of events.

\[ \bigcup_{\alpha \in \mathcal{I}} A_\alpha = \{ x : x \in A_\alpha \text{ for some } \alpha \in \mathcal{I} \} \]

\[ \bigcap_{\alpha \in \mathcal{I}} A_\alpha = \{ x : x \in A_\alpha \text{ for all } \alpha \in \mathcal{I} \} \]
**Example**

Suppose

\[ I = (0, \infty) = \{x \in \mathbb{R} : x > 0\} \]

\[ A_\alpha = [0, \alpha] = \{x \in \mathbb{R} : 0 \leq x \leq \alpha\} \]

Then

\[ \bigcup_{\alpha \in I} A_\alpha = [0, \infty) \]

\[ \bigcap_{\alpha \in I} A_\alpha = \{0\} \]
De Morgan’s Laws for Collections

De Morgan’s laws still hold:

\[
\left( \bigcup_{\alpha \in I} A_\alpha \right)^c = \bigcap_{\alpha \in I} A_\alpha^c \\
\left( \bigcap_{\alpha \in I} A_\alpha \right)^c = \bigcup_{\alpha \in I} A_\alpha^c
\]
Partitions

A collection $\{A_\alpha\}$ is a partition of a set $B$ if

1. the $A_\alpha$ are pairwise disjoint
2. $\bigcup_\alpha A_\alpha = B$

The sets in a partition are said to be

- mutually exclusive, and
- collectively exhaustive.
A Simple Definition of Probability

In simple cases we can define probabilities for all subsets of $S$.

**Definition**

A probability, or probability function, is a real-valued function defined on all subsets of a sample space $S$ such that

(i) $P(A) \geq 0$ for all $A \subset S$

(ii) $P(S) = 1$

(iii) If $A, B \subset S$ are disjoint, then

$$P(A \cup B) = P(A) + P(B).$$
Notes

- These properties are called *Axioms* of probability
- Axiom (iii) is the *Additivity Axiom*
- The three properties make sense for relative frequencies.
- Many other properties can be motivated by relative frequencies.
- We could add these to the definition.
- But that would be wasteful: All follow from the definition we have.
Consequences of the Definition

Complements

\[ P(A^c) = 1 - P(A). \]

Proof.

Since \( S = A \cup A^c \), by Axiom (ii)

\[ P(A \cup A^c) = P(S) = 1 \]

Since \( A \cap A^c = \emptyset \), by Axiom (iii)

\[ P(A \cup A^c) = P(A) + P(A^c). \]

So \( P(A) + P(A^c) = 1 \) and thus \( P(A^c) = 1 - P(A) \). □
Upper Bound

\[ P(A) \leq 1 \]

\textbf{Proof.}

\[ 1 = P(S) = P(A) + P(A^c) \geq P(A) \]

since \( P(A^c) \geq 0 \) by Axiom (i).

Empty Set

\[ P(\emptyset) = 0 \]

\textbf{Proof.}

Since \( \emptyset = S^c \),

\[ P(\emptyset) = P(S^c) = 1 - P(S) = 1 - 1 = 0. \]
Set Difference

\[ P(A \setminus B) = P(A \cap B^c) = P(A) - P(A \cap B) \]

Proof.

Write \( A \) as a disjoint union,

\[ A = (A \cap B) \cup (A \cap B^c) \]

Then

\[ P(A) = P(A \cap B^c) + P(A \cap B) \]
Non-Disjoint Unions

\[ P(A \cup B) = P(A) + P(B) - P(A \cap B) \]

**Proof.**

Decompose the events into disjoint pieces:

\[
\begin{align*}
A &= (A \cap B) \cup (A \cap B^c) \\
B &= (A \cap B) \cup (A^c \cap B) \\
A \cup B &= (A \cap B) \cup (A \cap B^c) \cup (A^c \cap B)
\end{align*}
\]

Then

\[
\begin{align*}
P(A \cup B) &= P(A \cap B) + P(A \cap B^c) + P(A^c \cap B) \\
&= P(A \cap B) + P(A \cap B^c) + P(A \cap B) + P(A^c \cap B) - P(A \cap B) \\
&= P(A) + P(B) - P(A \cap B)
\end{align*}
\]
Union of $n$ Mutually Exclusive Events

Suppose $A_1, \ldots, A_n$ are mutually exclusive. Then

$$P \left( \bigcup_{i=1}^{n} A_i \right) = \sum_{i=1}^{n} P(A_i)$$
Proof.

The proof is by induction. The claim is true by Axiom (iii) for $n = 2$. Suppose the claim holds for $n$. We need to show it holds for $n + 1$ mutually exclusive events $A_1, \ldots, A_{n+1}$. Let $B_n = \bigcup_{i=1}^{n} A_i$. Then $B_n$ and $A_{n+1}$ are disjoint, $\bigcup_{i=1}^{n+1} A_i = B_n \cup A_{n+1}$, and

$$P\left(\bigcup_{i=1}^{n+1} A_i\right) = P(B_n \cup A_{n+1})$$

$$= P(B_n) + P(A_{n+1})$$

$$= \sum_{i=1}^{n} P(A_i) + P(A_{n+1})$$

$$= \sum_{i=1}^{n+1} P(A_i)$$

So the claim holds for $n + 1$ as well.
Boole’s Inequality
For any \( n \) events \( A_1, \ldots, A_n \)
\[
P \left( \bigcup_{i=1}^{n} A_i \right) \leq \sum_{i=1}^{n} P(A_i)
\]

Bonferroni’s Inequality
For any \( n \) events \( A_1, \ldots, A_n \)
\[
P \left( \bigcap_{i=1}^{n} A_i \right) \geq 1 - \sum_{i=1}^{n} P(A_i^c)
\]
This follows from Boole’s inequality and DeMorgan’s laws.
Finite Sample Spaces

Suppose we have a finite sample space \( S = \{s_1, \ldots, s_N\} \).

- Often we specify a probability by specifying values \( p_i = P(\{s_i\}) \).
- The \( p_i \) have to satisfy \( p_i \geq 0 \) for \( i = 1, \ldots, N \), and

\[
\sum_{i=1}^{N} p_i = 1
\]

- The probability of a general event \( A \subset B \) is then

\[
P(A) = \sum_{i: s_i \in A} p_i
\]
Equally Likely Outcomes

- Suppose all outcomes are *equally likely*, i.e.

\[ p_1 = p_2 = \cdots = p_N \]

- Since \( \sum_{i=1}^{N} p_i = 1 \), this means

\[ p_i = \frac{1}{N} \]

for all \( i = 1, \ldots, N \).

- The probability of a general event \( A \) is then

\[
P(A) = \sum_{i:s_i \in A} \frac{1}{N} = \frac{\# \text{ elements of } A}{\# \text{ elements of } S}
\]

- So computing probabilities reduces to counting outcomes.
The earliest formal mathematical treatment of probability dealt with gambling games and equally likely outcomes.

Choosing a sample space in which outcomes can be assumed equally likely, when this makes sense, often simplifies a problem.

Games of chance form useful practice examples, but also often have practical implications for statistical modeling.
Example

- A coin is tossed three times and the number of heads is recorded.
- One possible sample space is $S_N = \{0, 1, 2, 3\}$.
- It is hard to justify thinking of these outcomes as equally likely.
- Another possible sample space is

$$S = \{x_1x_2x_3 : x_i \in \{H, T\}\}$$

$$= \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$$

- If the coin is fair then it makes sense to view these 8 outcomes as equally likely.
Example (continued)

- The number of heads for each element of this sample space is

<table>
<thead>
<tr>
<th>s</th>
<th>HHH</th>
<th>HHT</th>
<th>HTH</th>
<th>HTT</th>
<th>THH</th>
<th>THT</th>
<th>TTH</th>
<th>TTT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Num Heads</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

- The probabilities of different numbers of heads are then

\[
P(k \text{ heads}) = \begin{pmatrix} \frac{1}{8} & \frac{3}{8} & \frac{3}{8} & \frac{1}{8} \end{pmatrix}
\]

- Suppose we want to calculate the probability of 4 heads in 10 coin flips.

- The same principle works, but there are \(2^{10} = 1024\) outcomes in the sample space.

- We need a more systematic approach.
Example

- A die is rolled four times. What is the probability that at least one six is rolled?
- This is said to be part of a problem that spurred the involvement of Pascal and Fermat in developing a mathematical framework for probability.
- A possible sample space is
  \[ S = \{(x_1, x_2, x_3, x_4) : x_i \in \{1, 2, 3, 4, 5, 6\}\} \]
- If the die is *fair* then it is reasonable to consider these outcomes equally likely.
- We need to find the number of outcomes in the sample space.
- Then we need to think about the event of rolling at least one six.
Basic Rules of Counting

- The science of counting is called *combinatorics*.
- There are books and semester-long courses on the subject.
- We will only need some fairly basic results and principles.

**Addition Rule**
If \( A \cap B = \emptyset \), \( A \) has \( m \) elements and \( B \) has \( n \) elements, then \( A \cup B \) has \( n + m \) elements.

**Multiplication Rule**
Suppose a task can be broken into two sub-tasks.
- The first sub-task can be done in \( m \) ways.
- The second can be done in \( n \) ways, no matter how the first was done.
Then the entire task can be done in \( m \times n \) ways.