Recap

- Matching problem
- Generalized birthday problem
- Inclusion-exclusion formula
- Countably infinite sample spaces
Countably Infinite Sample Spaces

Example (continued)

• Suppose two players, $A$ and $B$, take turns rolling a die.
  • The first player who rolls a six wins the game.
  • Player $A$ rolls first.

• We can use as a sample space

$$S = \{1, 2, 3, \ldots \}$$
**Example (continued)**

- What is the probability that $A$ wins?
- $A$ wins if the game ends on an odd roll: 1, 3, 5, . . .
- Let $F_n$ be the event that the game ends on roll $n$.
- The $F_n$ are pairwise disjoint.
- It is tempting to write

$$P(A \text{ wins}) = P\left( \bigcup_{k=1}^{\infty} F_{2k-1} \right) = \sum_{k=1}^{\infty} P(F_{2k-1})$$

$$= \sum_{k=1}^{\infty} p_{2k-1}$$

$$= \sum_{k=1}^{\infty} \frac{1}{6} \left( \frac{5}{6} \right)^{(2k-1)-1} = \sum_{k=1}^{\infty} \frac{1}{6} \left( \frac{5}{6} \right)^{2k-2}$$

$$= \sum_{k=1}^{\infty} \frac{1}{6} \left( \frac{25}{36} \right)^{k-1} = \frac{1}{6} \left( \frac{1}{1 - 25/36} \right) = \frac{6}{11}$$
Example (continued)

- This is in fact the right answer.
- But there is a flaw in the argument.
- For infinite sample spaces our definition of probability does not guarantee that

\[ P \left( \bigcup_{k=1}^{\infty} F_{2k-1} \right) = \sum_{k=1}^{\infty} P(F_{2k-1}) \]

for pairwise disjoint \( F_n \).
- This identity does hold in situations like ours where

\[ \sum_{k=1}^{\infty} p_k = 1. \]
**Theorem**

Let $P$ be a probability on a countably infinite sample space $S = \{s_1, s_2, \ldots \}$, let $p_k = P(\{s_k\})$, and suppose that

$$\sum_{k=1}^{\infty} p_k = 1.$$ 

Then for any pairwise disjoint events $A_i \subset S$

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

This property is known as countable additivity.
Theorem

Let $P$ be a countably additive probability, let $A_1 \supset A_2 \supset A_3 \ldots$ be a decreasing sequence of events, and let

$$A = \bigcap_{i=1}^{\infty} A_i.$$ 

Then

$$\lim_{i \to \infty} P(A_i) = P(A).$$

This is called the continuity property of countably additive probability.
• Knowing that a probability is countably additive can make many calculations easier.

• All probabilities on finite sample spaces are countably additive.

• Are there probabilities that are not countably additive?
  • Yes; one example is a uniform probability distribution on the integers.

• For all practical purposes all probabilities used in practice are countably additive.

• As a result, the standard approach is to require countable additivity as part of the definition or a probability.

• The definition we have used so far is then called finitely additive probability.

• There is a school of probability and statistics based entirely on the use of finitely additive probabilities.

• Bruno de Finetti was a major proponent of the finitely additive approach.
Countably Additive Probability

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 $A_1, A_2, \cdots \subset S$ are pairwise disjoint, then

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

Notes

- This definition does not add the countable additivity axiom; it replaces the finite additivity axiom with the countably additive version.
- We need to check that finite additivity still holds.
**Theorem**

Suppose $A, B \subset S$ are disjoint and $P$ is a (countably additive) probability on $S$. Then

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

**Proof.**

Let $C_1 = A$, $C_2 = B$, and $C_i = \emptyset$ for $i = 3, 4, \ldots$. Then the $C_i$ are mutually exclusive, and

$$A \cup B = \bigcup_{i=1}^{\infty} C_i.$$

So by the countable additivity axiom and the fact that $P(\emptyset) = 0$,

$$P(A \cup B) = P\left( \bigcup_{i=1}^{\infty} C_i \right) = \sum_{i=1}^{\infty} P(C_i)$$

$$= P(A) + P(B) + 0 + 0 + \cdots = P(A) + P(B).$$
There is a gap in the proof:

- Our earlier proof that $P(\emptyset) = 0$ used finite additivity!

To fill the hole we need to prove $P(\emptyset) = 0$ using only the three axioms for countably additive probability.

Here is one way:

- Let $A_1 = S$ and let $A_i = \emptyset$ for $i = 2, 3, \ldots$.
- Then the $A_i$ are pairwise disjoint, and

\[
P(S) = P \left( \bigcup_{i=1}^{\infty} A_i \right) = \sum_{i=1}^{\infty} P(A_i) = P(S) + \sum_{i=2}^{\infty} P(\emptyset).
\]

- This is only possible if $P(\emptyset) = 0$. 
• We have already seen that a countably additive probability satisfies a certain continuity property.

• A finitely additive probability that has a continuity property is countably additive:

**Theorem**

Suppose $P$ is a finitely additive probability on a sample space $S$ and that for any events $A_1 \supset A_2 \supset \ldots$ with $\bigcap_{i=1}^{\infty} A_i = \emptyset$

$$\lim_{i \to \infty} P(A_i) = 0.$$  

Then $P$ is a countably additive probability on $S$.

• So, in this sense, a countably additive probability is a finitely additive probability that is also continuous.
Proof.

Outline: Using finite additivity, write

$$P \left( \bigcup_{i=1}^{\infty} A_i \right) = P \left( A_1 \cup A_2 \cup \cdots \cup A_n \cup \bigcup_{i=n+1}^{\infty} A_i \right)$$

$$= P(A_1) + P(A_2) + \cdots + P(A_n) + P \left( \bigcup_{i=n+1}^{\infty} A_i \right).$$

Then argue that continuity implies

$$P \left( \bigcup_{i=n+1}^{\infty} A_i \right) \to 0,$$

as $n \to \infty$. \qed
Building Probability Models

- The main tool we have so far for building probability models is equally likely outcomes.
- A lot can be done with this framework.
  - Much of classical probability is based on equally likely outcomes.
  - Box models used in some introductory text books are another example.
- But more flexible tools are needed for more complex situations.
Example

- About 50% of decided Iowa voters say they will vote Democratic, 50% Republican.
- About 10% of all people are left handed.
- If we select a decided voter at random, what is the chance the person will be a left-handed Democratic voter?
- It seems reasonable to assume
  - about 10% of Iowa decided voters are left handed;
  - about 10% of decided Republican voters are left handed;
  - about 10% of decided Democratic voters are left handed.
- So the probability should be about $0.5 \times 0.1 = 0.05$.
- The classifications of party and handedness seem unrelated, or independent.
- What if instead of handedness the second classification was
  - Male/Female?
  - Urban/Rural?
Independence

Definition
Two events $A$ and $B$ are independent if $P(A \cap B) = P(A)P(B)$.

Notes

• In model building we often assume (at least approximate) independence.
• This allows complex probabilities to be derived from simple ones.
• Like any assumption, you have to think it through carefully before relying on it.
• It is rarely exactly true but often close enough to be useful.
• In some cases independence is derived.
• If $A$ and $B$ are disjoint and have positive probability then they are not independent; they are dependent.
Example

- A remote sensing device can be equipped with one or two power supplies.
- With a single power supply the device fails if the power supply fails.
- With two power supplies the device will only fail once both power supplies have failed.
- Suppose the chance of one power supply failing in a particular time frame is 10%.
- If we assume independent failures, then the chance that both fail is $0.1 \times 0.1 = 0.01$ or 1%.
- Is the independence assumption reasonable?
Example

- A fair die is rolled.
- Let $A$ be the event that an even number is rolled.
- Let $B$ be the event that the number rolled is at most 2.
- Then

\[
P(A) = \frac{3}{6} = \frac{1}{2}
\]

\[
P(B) = \frac{2}{6} = \frac{1}{3}
\]

\[
P(A \cap B) = \frac{1}{6} = \frac{1}{2} \times \frac{1}{3} = P(A)P(B).
\]

- So $A$ and $B$ are independent, even though they relate to the same roll of a die.
- This is an example of derived independence.
If $A$ and $B$ are independent, then

$$P(A \cap B^c) = P(A) - P(A \cap B)$$
$$= P(A) - P(A)P(B)$$
$$= P(A)(1 - P(B))$$
$$= P(A)P(B^c)$$

So $A$ and $B^c$ are also independent.

Similarly, $A^c$ and $B$ are independent, and $A^c$ and $B^c$ are independent.

In words: knowing whether $A$ has or has not occurred tells us nothing about whether $B$ has or has not occurred.
**Definition**

Events $A_1, A_2, \ldots, A_n$ are mutually independent if

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1)P(A_2) \cdots P(A_n)$$

and every sub-collection of $n-1$ of the $A_i$ is mutually independent.

**Notes**

- Three events $A$, $B$, and $C$ are mutually independent if

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

- It is possible to have the first three but not the fourth;
- it is possible to have the fourth but not the first three.
Example (Remote sensing power supplies, continued)

- Suppose we can use $n$ power supplies in our remote sensing device.
- The device will only fail once all power supplies have failed.
- Suppose the chance of a particular power supply failing is again 10%.
- Suppose failures of these devices are mutually independent.
- Then the chance of all power supplies failing is $(0.1)^n$.
- By making $n$ large we can make the chance of failure as small as we like.
- Is the independence assumption likely to be reasonable for very large $n$?
Examples

• A fair coin is tossed independently \( n \) times.
  - All \( 2^n \) outcomes are equally likely.

• A fair die is rolled independently \( n \) times.
  - All \( 6^n \) outcomes are equally likely.

• A biased coin with probability \( p \) of heads is tossed \( n \) times.
  - The probability of a particular outcome with \( k \) heads and \( n - k \) tails is
    \[
    p^k(1 - p)^{n-k}.
    \]
  - The probability of \( k \) heads is
    \[
    \binom{n}{k} p^k(1 - p)^{n-k}.
    \]
  - This is an important basic model for binary data.
Recap

- Countably additive probability
- Independence
Example

- Reliability block diagrams are sometimes used to analyze reliability of complex systems.

- Failure modes are usually assumed independent.
- Let

\[
F = \{\text{whole system fails}\}
\]
\[
F_i = \{\text{Component } i \text{ fails}\}
\]
\[
p_i = P(F_i).
\]
Example (continued)

- We can often break larger systems into smaller sub-systems.
- The system of components 1 and 2 is a series system:

\[
\begin{array}{c}
A \\
\hline
1 \\
\hline
2 \\
\hline
B
\end{array}
\]

- The sub-system fails if either 1 or 2 fails
- The sub-system failure event is \(F_{12} = F_1 \cup F_2\).
- So the sub-system failure probability is

\[
P(F_{12}) = P(F_1 \cup F_2) = P(F_1) + P(F_2) - P(F_1 \cap F_2) = p_1 + p_2 - p_1p_2
\]

- Similarly for the second sub-system \(P(F_{34}) = p_3 + p_4 - p_3p_4.$$
Example (continued)

- The whole system is a *parallel system* of these two sub-systems.

- The whole system only fails if both sub-systems fail, \( F = F_{12} \cap F_{34} \).

- So the failure probability for the whole system is

\[
P(F) = P(F_{12} \cap F_{34}) = P(F_{12})P(F_{34}) = (p_1 + p_2 - p_1 p_2)(p_3 + p_4 - p_3 p_4)
\]

- The process of generating system failure probabilities from reliability block diagrams can be automated.
Update Probabilities: Conditional Probability

**Example**

- Suppose I roll a die. The probability of an even roll is 1/2.
- Suppose I tell you the roll is $\leq 3$. Does this change the probability that the roll is even?
- Using relative frequencies, we can look at cases where the roll is $\leq 3$:

  $$\text{new } P(\text{even}) \approx \frac{\#(\text{even and } \leq 3)}{\#(\text{roll } \leq 3)} = \frac{\text{prop(\text{even and } \leq 3})}{\text{prop(\leq 3})}$$

  $$\approx \frac{P(\text{even and } \leq 3)}{P(\leq 3)} = \frac{1/6}{1/2} = 1/3$$

- You adjust your probability to a new sample space that only allows rolls $\leq 3$. 

**Definition**

For events $A$ and $B$ with $P(B) > 0$, the *conditional probability* of $A$ given that $B$ has occurred is

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

- If $A$ and $B$ are independent then $P(A|B) = P(A)$.
- Sometimes we compute conditional probabilities.
- Other times we use them to build up more complicated probabilities:
  - using the multiplication formula:
    \[ P(A \cap B) = P(A|B)P(B) \]
  - using the Law of Total Probability:
    \[
    P(A) = P(A \cap B) + P(A \cap B^c) \\
    = P(A|B)P(B) + P(A|B^c)P(B^c)
    \]
Example

- About half of potential Iowa voters are male and half female
- About 40% of female potential voters would vote Republican
- About 60% of male potential voters would vote republican
- If a voter is selected at random then

\[ P(F) = 0.5 \quad P(M) = 0.5 \]
\[ P(R|F) = 0.4 \quad P(R|M) = 0.6 \]

- The probability of selecting a Republican potential voter is

\[ P(R \cap F) = P(R|F)P(F) = 0.4 \times 0.5 = 0.2 \]
\[ P(R \cap M) = P(R|M)P(M) = 0.6 \times 0.5 = 0.3 \]
\[ P(R) = P(R \cap F) + P(R \cap M) = 0.2 + 0.3 = 0.5 \]

- Mosaic plots can be used to illustrate joint probabilities of two or more classifications.
Example

• 10% of a population has a disease.

• A test has the properties

\[ P(+) = 0.96 \] sensitivity
\[ P(+) = 0.95 \] specificity

\[ P(+) = P(\cap D) + P(\cap D^c) \]
\[ = P(+)P(D) + P(+)P(D^c) \]
\[ = P(+)P(D) + (1 - P(-)P(D^c))P(D^c) \]
\[ = 0.96 \times 0.1 + (1 - 0.95) \times 0.9 \]
\[ = 0.096 + 0.045 = 0.141 \]
Bayes Theorem

It is often useful to reverse the calculation:

- Given that a randomly chosen person has a positive result, what is $P(D)$, the probability that the person has the disease?

$$P(D|+) = \frac{P(D \cap +)}{P(+)} = \frac{0.096}{0.141} = 0.678$$

- This calculation is very easy to do from scratch—just rewrite the probability you want in terms of the inputs you have.

- But this calculation is very important, so it has a formal theorem attached to it.
Theorem (Bayes Theorem)

Let the events $B_1, B_2, \ldots$ partition $S$, i.e.

- the $B_i$ are pairwise disjoint, and
- they are collectively exhaustive: $\bigcup B_i = S$.

Then for any event $A$

$$P(B_i|A) = \frac{P(A \cap B_i)}{P(A)} = \frac{P(A|B_i)P(B_i)}{\sum_j P(A|B_j)P(B_j)}$$

Notes

- Bayes theorem is the foundation to the Bayesian approach to drawing conclusions from data.
- The denominator follows from the law of total probability.
Example

• Another reliability block diagram:

```
A  1  2  B
  3
  5
  4
```

• One possible approach is to condition on whether component 5 has failed or not.
Example (continued)

- If component 5 has failed, then the system is like the one we analyzed previously; so

\[ P(F|F_5) = (p_1 + p_2 - p_1 p_2)(p_3 + p_4 - p_3 p_4). \]

- If component 5 has not failed, then the system is a series of two parallel systems, so

\[ P(F|F_5^c) = p_1 p_3 + p_2 p_4 - p_1 p_3 p_2 p_4 \]

- The system failure probability is then

\[ P(F) = P(F|F_5)p_5 + P(F|F_5^c)(1 - p_5) \]
Example

- The counting multiplication rule can often be replaced by a conditioning argument.
- In one approach to the matching problem we needed to compute $P(A_{i_1} \cap \cdots \cap A_{i_k})$, with $A_i$ the event that person $i$ receives their own hat.
- $P(A_{i_1}) = \frac{1}{n}$.
- $P(A_{i_1} \cap A_{i_2}) = P(A_{i_1})P(A_{i_2}|A_{i_1}) = \frac{1}{n} \times \frac{1}{n-1} = \frac{1}{n(n-1)}$.
- $P(A_{i_1} \cap A_{i_2} \cap A_{i_3}) = P(A_{i_1} \cap A_{i_2})P(A_{i_3}|A_{i_1} \cap A_{i_2}) = \frac{1}{n(n-1)} \times \frac{1}{n-2}$.
Example

• In analyzing processes based on a sequence of independent trials it is often useful to condition on the result of the first trial.

• Suppose a die is rolled until a six occurs.

• Let $E$ be the event that a six eventually occurs.

• Let $A$ be the event that the first roll is a six.

• Then

$$P(E) = P(E|A)P(A) + P(E|A^c)P(A^c) = P(E|A)\frac{1}{6} + P(E|A^c)\frac{5}{6}.$$ 

• Now $P(E|A) = 1$.

• If $A$ does not occur, then the process starts over independently, and $P(E|A^c) = P(E)$.

• So $P(E)$ satisfies

$$P(E) = \frac{1}{6} + P(E)\frac{5}{6}.$$ 

• So $P(E) = 1$. 