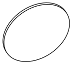
C2. Probability

Notes originally prepared by Judy-anne Osborn and Pierre Portal.
Editing, expansion and additions by Malcolm Brooks and AP.

Text Reference (Epp)	3ed: Sections	6.7-9
	4ed: Sections	9.7-9
	5ed: Sections	9.7-9

(Only the last part of §9 on ‘independence’ is relevant for this course)

A thought experiment:



- Toss a coin.
- What are the possible outcomes?
- ‘Heads’ or ‘Tails’
- What is the probability of ‘Heads’?
- We say it is

$$\mathbb{P}(\text{Heads}) = \frac{1}{2}. \qquad \textit{Why?}$$

Methods of assigning probabilities

Method 1: Use relative frequencies (empirical experiment)

Method 2: Use a model (combination of prior knowledge, guessing, deduction)

- Eg. assume equally likely outcomes

Relative Frequencies

- For example, one may carry out the experiment of tossing a coin ten times and noting the results.

- My experiment gave

T,T,H,H,T,T,T,H,T,H

- Slightly fewer than half the coin-tosses resulted in 'H' (for 'Heads').

- A 'longer run' may give different (better?) results.

- There is much more to be said on 'relative frequencies', but for this course we will focus on making 'models'.

A model for coin tossing

- Observe a real coin: it has two sides – ‘Heads’ and ‘Tails’.
- Actually there is a third ‘side’: the rim.
- Because the rim is small we decide to ignore this possibility.
- We simplify on purpose, to make the model *tractable*.
- The ‘Heads’ and ‘Tails’ sides are so similar physically that we make an assumption:

  equal likelihood

  for

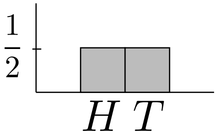
  Heads or Tails.

A model for coin tossing: equal likelihood

The two possibilities are just as likely as each other.

$$\mathbb{P}(\text{Heads}) = \frac{1}{2} \qquad \mathbb{P}(\text{Tails}) = \frac{1}{2}$$

We can represent this situation graphically as



Fix some terminology

An **Experiment** observes a phenomenon that has one or more possible **outcomes**.

The **Sample space** of an experiment is the *set* of possible outcomes of the experiment.

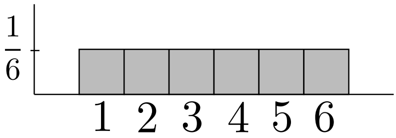
An **Event** is any *subset* of the sample space.

The probability of an event E is denoted by $\mathbb{P}(\boldsymbol{E})$.

Example:

- Experiment: single toss of a standard die, noting upper face's number.
- One possible outcome would be '3'
- Sample space: $\{1, 2, 3, 4, 5, 6\}$

An equal likelihood model for die-tossing



What's an event? Any subset of the sample space.
Eg. an event is

$$\boxed{\{3,6\}}$$

$=$ *the set of numbers divisible by 3 in sample space* $\{1,2,3,4,5,6\}$.

$$\boxed{\mathbb{P}(\{3,6\}) = \frac{|\{3,6\}|}{|\{1,2,3,4,5,6\}|} = \frac{2}{6} = \frac{1}{3}}$$

Equal Likelihood for finite Sample Spaces

Generalising from the previous example we have:

- Let S be a finite sample space in which all outcomes are equally likely.
- Let E be an event in S.
- Then the probability of the event E is

$$\mathbb{P}(E) = \frac{|E|}{|S|}$$

where $|E|$ is the number of outcomes in E, and
$|S|$ is the number of outcomes in S.

Properties of Probability

For any finite sample space:

- $\mathbb{P}(E)$ is a real number between 0 and 1, i.e.
$$0 \leq \mathbb{P}(E) \leq 1.$$
- Probability of the complement is one minus the probability of the event, i.e.
$$\mathbb{P}(\text{not } E) = 1 - \mathbb{P}(E).$$
- The sum of the probabilities of all outcomes in the sample space is 1.
- '$\mathbb{P}(E) = 1$' implies E is certain to occur.*
- '$\mathbb{P}(E) = 0$' implies E is impossible.*

*For infinite sets, this isn't necessarily true. 'Measure theory' explains why.

Previous example of tossing a die:

Probability of event of ‘getting a number exactly divisible by 3’ is one third, which satisfies

$$0 \leq \frac{1}{3} \leq 1.$$

Probability of ‘not E’ is

$$\mathbb{P}(\text{number not divisible by 3}) = 1 - \frac{1}{3} = \frac{2}{3}.$$

The sum of the probabilities of all outcomes is

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

The Sum and Product Rules for Probability

The Sum Rule

Sum Rule: If events $E_1, ..., E_n$ are mutually disjoint, i.e. $E_i \cap E_j = \emptyset$ for all $i \neq j$, then

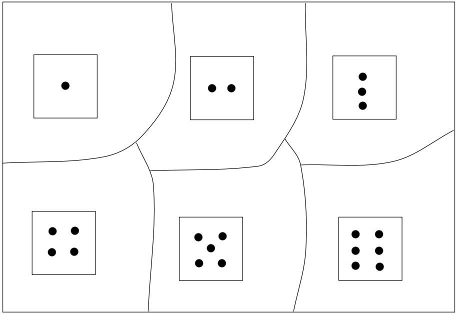
$$\mathbb{P}(E_1 \cup ... \cup E_n) = \mathbb{P}(E_1) + ... + \mathbb{P}(E_n).$$

Disjoint events exclude one another in the sense that they cannot happen at the same time.

Sum Rule for probability: another die-tossing example

What is the probability that the outcome from a single toss of a die is an odd number?

The six possible outcomes are all disjoint (cannot occur simultaneously).



Thus the sum rule applies.

- We assign equal probabilities to each of these disjoint events (Why?)
- Six possible outcomes in total $\rightarrow$ each has probability $\frac{1}{6}$ of occurring.
- The probability that the die lands with an odd number up is

$$
\begin{aligned}
& \Pr\left(\text{⚀}\right) + \Pr\left(\text{⚂}\right) + \Pr\left(\text{⚄}\right) \\
= \quad & \frac{1}{6} + \frac{1}{6} + \frac{1}{6} \\
= \quad & \frac{1}{2}
\end{aligned}
$$

by the sum rule.

Sum Rule for probability. Example: Non-zero numbers

Let $R_n = \{-n, .., -2, -1, 0, 1, 2, ..., n\}$.

What is the probability that a number chosen at random from R_n is non-zero?

We assume that we are equally likely to choose any element of R_n.

- The probability that the number is negative is $\frac{|R_n^-|}{|R_n|} = \frac{n}{2n+1}$.
- The probability that the number is positive is $\frac{|R_n^+|}{|R_n|} = \frac{n}{2n+1}$.

Therefore the probability of a number chosen at random from the set $\{-n, .., -2, -1, 0, 1, 2, ..., n\}$ being non-zero is:

$$\begin{aligned} &\mathbb{P}(\text{the number is negative}) + \mathbb{P}(\text{the number is positive}) \\ =&\frac{n}{2n+1} + \frac{n}{2n+1} = \frac{2n}{2n+1}. \end{aligned}$$

The Product Rule

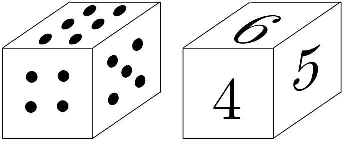
- Product Rule: If events $E_1, ..., E_n$ are 'independent' of each other; then the probability of composite event 'E_1 and E_2 and ... and E_n' is

$$\mathbb{P}(E_1 \cap E_2 \cap \ldots \cap E_n) = \mathbb{P}(E_1) \times \mathbb{P}(E_2) \times \ldots \mathbb{P}(E_n).$$

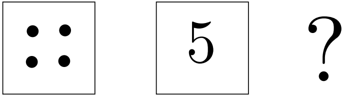
- To see what we mean by 'independent', consider a procedure that can be broken down into successive tasks, each of which could be done in a number of ways. If the choice of the way to do any one task had no influence on the choice of ways to do any other of the tasks, then the tasks would be independent.

- A formal definition of independence will be given later.

Product Rule probability example: Tossing two dice



- What is the probability that the outcome from tossing a pair of dice is '4' for the first die and '5' for the second die i.e.



- We assume that the outcomes for each die are **independent**, i.e that they don't influence one another at all.

- Hence the product rule applies.

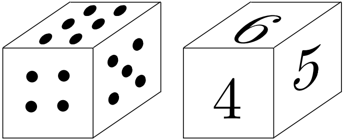
$$
\begin{aligned}
& \Pr\left(\boxed{\text{⚃}}\ \ \boxed{5}\right) \\
= \ & \Pr\left(\boxed{\text{⚃}}\right) \times \Pr\left(\boxed{5}\right) \\
= \ & \frac{1}{6} \times \frac{1}{6} \\
= \ & \frac{1}{36}
\end{aligned}
$$

by the Product Rule.

An example of the Sum and Product Rules used together

- Often we combine use of the Sum and Product rules in one problem.

- For example, *what is the probability of getting an odd total when tossing a pair of dice?*



- To obtain an odd total, either
  - the first die must give odd and the second die even; or
  - the first die must give even and the second die odd.

- These two possibilities are **disjoint**, so the sum rule applies:

  $\mathbb{P}(\text{odd total}) = \mathbb{P}(\text{1st odd, 2nd even}) + \mathbb{P}(\text{1st even, 2nd odd})$

- But now consider $\mathbb{P}(\text{1st odd, 2nd even})$. The events

  "1st odd" and "2nd even"

  are **independent** of each other; they don't affect each other.

- Hence the product rule applies to this part of the problem:

$$\mathbb{P}(\text{1st odd, 2nd even}) = \mathbb{P}(\text{1st odd}) \times \mathbb{P}(\text{2nd even})$$
$$= \frac{3}{6} \times \frac{3}{6} = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

- Similarly, $\quad \mathbb{P}(\text{1st even, 2nd odd}) = \frac{1}{4}$

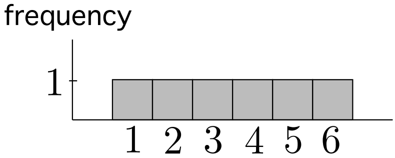
- Putting it all together,

$$\mathbb{P}(\text{odd total}) = \mathbb{P}(\text{1st odd, 2nd even}) + \mathbb{P}(\text{1st even, 2nd odd})$$
$$= \frac{1}{4} + \frac{1}{4} = \frac{1}{2}.$$

Density and Distribution

Frequency Histograms

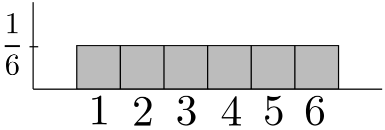
- One way to visualize all possible outcomes of an experiment together is to draw a frequency histogram.
- We have already seen some simple examples, like tossing a die with equally likely possible outcomes: $1, 2, 3, 4, 5, 6$:



Probability Density Functions

- The Probability Density Function (or just Density) is obtained from a Frequency Histogram by normalizing. We divide the vertical axis by the total number of outcomes.

- Continuing the die-tossing example, we have

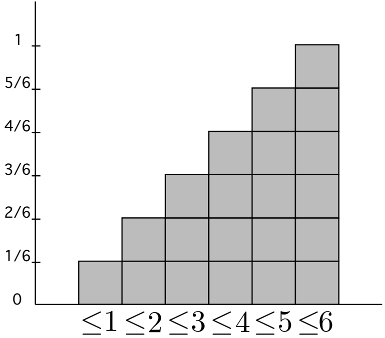


What is the area under the curve? *Why?*

Cumulative Probability Distribution Functions

- The Cumulative Probability Distribution Function (or Distribution) is obtained from the Density Function by graphing cumulative totals.

- Continuing the die-tossing example, we have

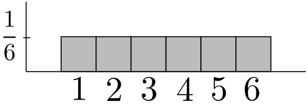


- We will only use of cumulative distributions when looking up probability values in tables or online.

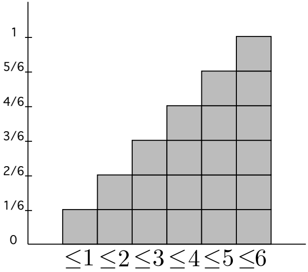
Uniform Distribution

- When every event has the same probability the resulting densities and distributions are called 'uniform'. Examples:

- Uniform density:



- Uniform distribution:



Tossing two coins:

- Some more interesting densities and distributions are obtained by considering events which combine several outcomes.

- For example, tossing two coins. A neat way to list all possible outcomes is to expand

$$
\begin{aligned}
& (T+H)(T+H) \\
= \; & TT + TH + HT + HH
\end{aligned}
$$

- *What is the sample space?*

$$\boxed{\{TT, TH, HT, HH\}}$$

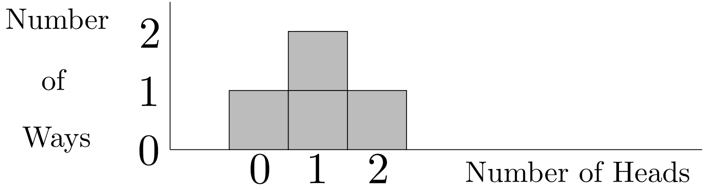
Tossing two coins:

- Now consider events:
  - E_0: 'No heads'
  - E_1: 'exactly 1 Head'
  - E_2: 'exactly 2 Heads'

- *Which subsets of the sample space correspond to these events?*

$$E_0 = \{TT\}, \qquad E_1 = \{TH, HT\}, \qquad E_2 = \{HH\}$$

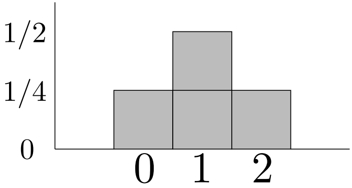
Frequency Histogram:

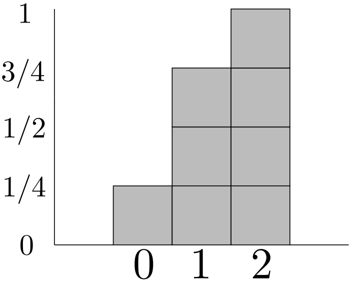


Two fair coins: Density and Distribution Functions

- Assuming a fair coin (equally likely outcomes), divide out by the size of the sample space to get density function. Then take the cumulative sum of the density to get the distribution:

Density $\longrightarrow$ **Distribution**





Class Example - Three Fair Coins

Q: *For tossing three fair coins, what is the sample space?*

A: All possible outcomes are given by $(T+H)(T+H)(T+H)$
$= TTT + TTH + THT + THH + HTT + HTH + HHT + HHH$.
S.Space $= \{TTT, TTH, THT, THH, HTT, HTH, HHT, HHH\}$.

Q: *What subsets of the sample space correspond to the events E_0:'0 heads', 'E_1:'exactly 1 Head', E_2:'exactly 2 Heads' and E_3:'exactly 3 Heads' ?*

A: $E_0 = \{TTT\}$; $E_1 = \{TTH, THT, HTT\}$;
$E_2 = \{THH, HTH, HHT\}$; $E_3 = \{HHH\}$.

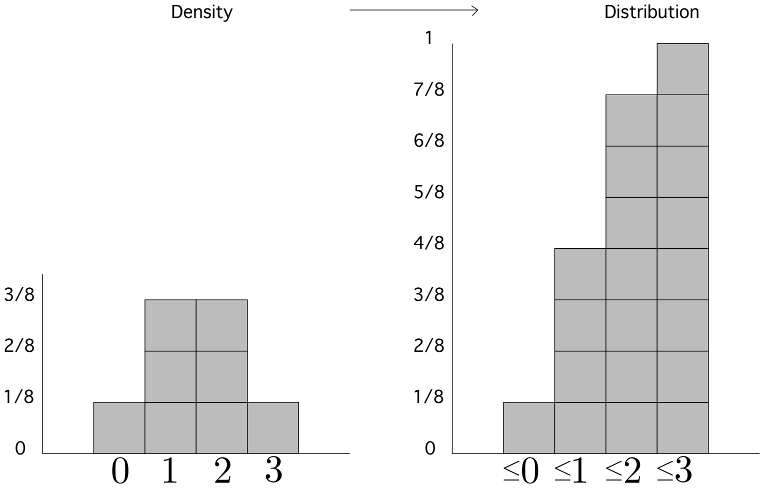
Q: *What is the size of the sample space?* A: 8

Q: What do we divide frequencies by (on the Frequency Histogram) to get the Density Function? A: 8

Q: What are the probabilities of the four events E_0, E_1, E_2, E_3?

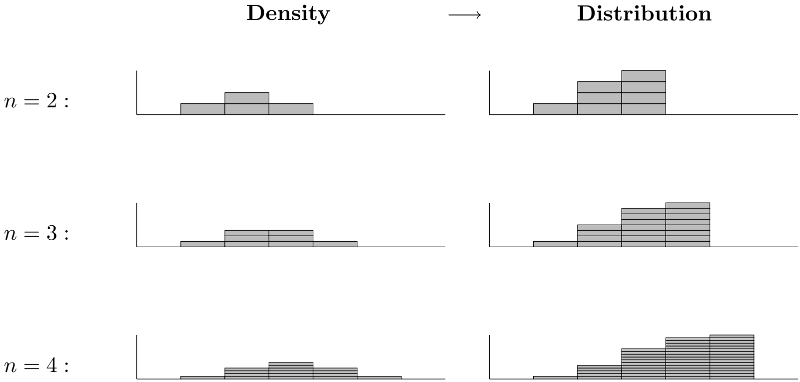
A: $\mathbb{P}(E_0) = \dfrac{1}{8}$; $\mathbb{P}(E_1) = \dfrac{3}{8}$; $\mathbb{P}(E_2) = \dfrac{3}{8}$; $\mathbb{P}(E_3) = \dfrac{1}{8}$.

Three Fair Coins: Density and Distribution Functions



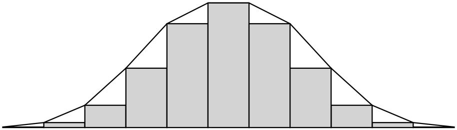
Binomial Probability Distributions

The family of functions that come from coin-tossing are all examples of binomial densities/distributions:



Bell-like curves for large n

As n gets larger and larger these **binomial probability density functions** get closer and closer to the famous Bell Curve:



which is the so-called **'Normal' Probability Density Function**.

Pascal's Triangle
and
Coin Tossing

Pascal's Triangle

$$
\begin{array}{ccccccccc}
 & & & & 1 & & & & \\
 & & & 1 & & 1 & & & \\
 & & 1 & & 2 & & 1 & & \\
 & 1 & & 3 & & 3 & & 1 & \\
1 & & 4 & & 6 & & 4 & & 1
\end{array}
$$

Pascal's Triangle

- Frequencies in Coin-Tossing are numbers in Pascal's Triangle

$$\begin{array}{ccccccccc}
 & & & & 1 & & & & \\
 & & & 1 & & 1 & & & \\
 & & 1 & & 2 & & 1 & & \\
 & 1 & & 3 & & 3 & & 1 & \\
1 & & 4 & & 6 & & 4 & & 1
\end{array}$$

- Each row is generated by expanding a binomial, eg:

$$(y+x)^4 = y^4 + 4y^3x + 6y^2x^2 + 4yx^3 + x^4.$$

Pascal's Triangle

- We've seen these numbers before in 'combinations': $\binom{n}{k}$:

$$
\begin{array}{ccccccccc}
 & & & & \binom{0}{0} & & & & \\
 & & & \binom{1}{0} & & \binom{1}{1} & & & \\
 & & \binom{2}{0} & + & \binom{2}{1} & & \binom{2}{2} & & \\
 & \binom{3}{0} & & \binom{3}{1} & + & \binom{3}{2} & & \binom{3}{3} & \\
\binom{4}{0} & & \binom{4}{1} & & \binom{4}{2} & & \binom{4}{3} & & \binom{4}{4}
\end{array}
$$

The Binomial Theorem

- The Binomial Theorem states that

$$(y+x)^n = \binom{n}{0} y^n x^0 + \binom{n}{1} y^{n-1} x^1 + \cdots + \binom{n}{n} y^0 x^n$$

and gives the rows of Pascal's Triangle in its coefficients.

Idea of Proof of Binomial Theorem:

$$(y+x)(y+x)(y+x)$$

$$= yyy + yyx + yxy + yxx + xyy + xyx + xxy + xxx$$

$$= \underbrace{yyy}_{\binom{3}{0}\ x\text{'s}} + \underbrace{yyx + yxy + xyy}_{\binom{3}{1}\ x\text{'s}} + \underbrace{yxx + xyx + xxy}_{\binom{3}{2}\ x\text{'s}} + \underbrace{xxx}_{\binom{3}{3}\ x\text{'s}}$$

$$\square\square\square \quad\; \square\square\boxtimes \;\; \square\boxtimes\square \;\; \boxtimes\square\square \quad\; \square\boxtimes\boxtimes \;\; \boxtimes\square\boxtimes \;\; \boxtimes\boxtimes\square \quad\; \boxtimes\boxtimes\boxtimes$$

Tossing n Coins

- Let's toss a (fair) coin n times ($n \in \mathbb{N}$.)
- As earlier, let E_k denote the event of obtaining k heads.
- Then
  - E_0 can occur in $\binom{n}{0} = 1$ way, *i.e* $|E_0| = \binom{n}{0}$
  - E_1 can occur in $\binom{n}{1} = n$ ways, *i.e* $|E_1| = \binom{n}{1}$
  - E_2 can occur in $\binom{n}{2}$ ways, *i.e* $|E_2| = \binom{n}{2}$

    $\vdots$
  - E_n can occur in $\binom{n}{n}$ ways, *i.e* $|E_n| = \binom{n}{n}$

  What is the total size of the sample space?

- I.e. what is

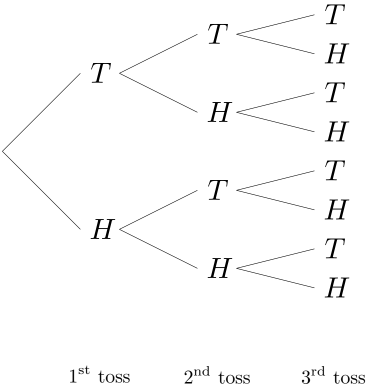
$$\binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \dots \binom{n}{n}?$$

- The binomial theorem gives a neat way to find the sum.
- Set $x = y = 1$ then
$$\binom{n}{0}1^n1^0 + \binom{n}{1}1^{n-1}1^1 + \cdots + \binom{n}{n}1^01^n = (1+1)^n$$
so that $$\boxed{\binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{n} = 2^n.}$$
- We can also get 2^n by by observing that there are two possible outcomes for each toss, and so $2 \times 2 \times \cdots \times 2 = 2^n$ possible outcomes for n tosses.
- So, for example, the probability of obtaining exactly three heads from six tosses of a fair coin is
$$\frac{\binom{6}{3}}{2^6} = \frac{\frac{6 \cdot 5 \cdot 4}{3 \cdot 2 \cdot 1}}{64} = \frac{20}{64} = {}^5\!/_{16}.$$
- Probabilities like these can be looked up in tables rather than calculated. Examples will be found in worksheet and assignment questions.

Tree Diagrams, Fair and Unfair Coins, and the General Binomial Distribution

A tree representation of Coin-tossing

- Another way to list all the outcomes of an event is to draw a Tree Diagram of the Possibilities

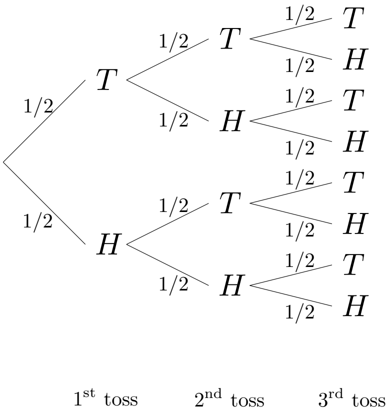


Outcome:

TTT
TTH
THT
THH
HTT
HTH
HHT
HHH

Three tosses of a fair coin

- This allows us to deal with fair coins, as before:



Outcome:

$\Pr(TTT) = 1/8$

$\Pr(TTH) = 1/8$

$\Pr(THT) = 1/8$

$\Pr(THH) = 1/8$

$\Pr(HTT) = 1/8$

$\Pr(HTH) = 1/8$

$\Pr(HHT) = 1/8$

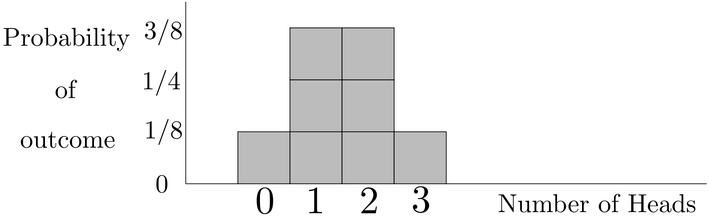
$\Pr(HHH) = 1/8$

Three tosses of a fair coin

Collecting possibilities from the tree and using the sum rule gives

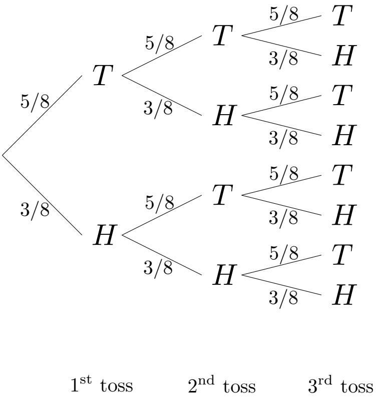
$$\boxed{\mathbb{P}(0\text{heads}) = \tfrac{1}{8},\ \ \mathbb{P}(1\text{head}) = \tfrac{3}{8},\ \ \mathbb{P}(2\text{heads}) = \tfrac{3}{8},\ \ \mathbb{P}(3\text{heads}) = \tfrac{1}{8}}$$

– the same density function as before; $n = 3$ and $p = \frac{1}{2}$ (fair coin):



Three tosses of an unfair coin

- But we can also deal with an unfair coin – not equal likelihood:



Outcome:

$\Pr(TTT) = 125/512$

$\Pr(TTH) = 75/512$

$\Pr(THT) = 75/512$

$\Pr(THH) = 45/512$

$\Pr(HTT) = 75/512$

$\Pr(HTH) = 45/512$

$\Pr(HHT) = 45/512$

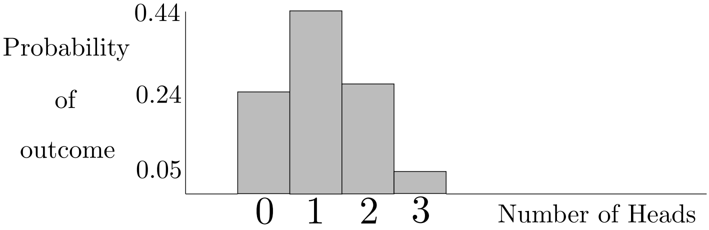
$\Pr(HHH) = 27/512$

Three tosses of an unfair coin

Collecting possibilities from the tree and using the sum rule gives

$$\boxed{\mathbb{P}(0\text{heads}) = \tfrac{125}{512},\ \mathbb{P}(1\text{head}) = \tfrac{225}{512},\ \mathbb{P}(2\text{heads}) = \tfrac{135}{512},\ \mathbb{P}(3\text{heads}) = \tfrac{27}{512}}$$

The unfair coin with $n = 3$ tosses and probability $p = 3/8$ of heads on a single toss, gives a non-symmetric binomial density function:



The general binomial density function for n trials (*e.g.* tosses) with probability p of a success (*e.g.* head) on each trial is given by

$$\boxed{\mathbb{P}(k \text{ successes}) = \binom{n}{k} p^k (1-p)^{n-k}.}$$

Review of
Probability Density Functions
with
More Challenging Examples

Probability for equally likely outcomes (Review)

For a finite non-empty set S and $E \subseteq S$. the **probability of E for equally likely outcomes** is the number

$$\boxed{\mathbb{P}(E) = \frac{|E|}{|S|}}$$

Vocabulary: S is called the **sample space**. E is called an **event**. An element $s \in S$ is called an **outcome**.

Examples:

(1) Model for a (fair) coin toss. $S = \{\text{H,T}\}$.

$\{\text{H}\}$ is the event 'coin shows a Head'. $\mathbb{P}(\{\text{H}\}) = \dfrac{|\{\text{H}\}|}{|\{\text{H,T}\}|} = \dfrac{1}{2}$.

The probability for equally likely outcomes is such that:

$$\boxed{\mathbb{P}(\emptyset) = 0,\ \mathbb{P}(\{\text{H}\}) = \tfrac{1}{2},\ \mathbb{P}(\{\text{T}\}) = \tfrac{1}{2},\ \mathbb{P}(\{\text{H,T}\}) = 1.}$$

(2) Model for a (balanced) die roll. $S = \{1, 2, 3, 4, 5, 6\}$.

5 is an outcome. $\{3, 5\}$ is an event. $\mathbb{P}(\{3, 5\} = \dfrac{|\{3, 5\}|}{|S|} = \dfrac{2}{6} = \dfrac{1}{3}$.

General probabilities on finite sets

Let $\mathbb{Q}_+ = \{q \in \mathbb{Q} \,;\; q \geq 0\}$, (the set of non-negative rational numbers).

A (general) **probability density function** on a finite set (sample space) S is any function

$$\boxed{\mathbb{P} : S \to \mathbb{Q}_+ \text{ with } \textstyle\sum_{s \in S} \mathbb{P}(s) = 1.}$$

We call $\mathbb{P}(s)$ the **probability of** s, so the probabilities sum to 1.

For any subset (event) $E \subseteq S$ the **probability of** E**,** $\mathbb{P}(\boldsymbol{E})$ is

$$\boxed{\mathbb{P}(E) = \textstyle\sum_{s \in E} \mathbb{P}(s).}$$

Example:

For $S = \{s_1, \ldots, s_n\}$ define $\mathbb{P} : S \to \mathbb{Q}_+$ by $\mathbb{P}(s_j) = \frac{1}{n}$, $j = 1, \ldots, n$.

Then $\quad \mathbb{P}(\{\omega_1, \ldots, \omega_m\}) = \sum_{j=1}^{m} \mathbb{P}(\omega_j) = \sum_{j=1}^{m} \frac{1}{n} = \dfrac{m}{n} = \dfrac{|\{\omega_1, \ldots, \omega_n\}|}{|S|}.$

So $\mathbb{P}$ is the probability of equally likely outcomes.

Another general probability example

In a group of 10 students, 5 are studying computer science, 2 are studying art history, and 3 are studying mathematics. We pick a student from this group and ask what her/his major is.

The sample space is $S = \{M, A, C\}$, where M stand for maths, A for art history, and C for computer science.

The probability density function $\mathbb{P} : S \to \mathbb{Q}_+$ is given by

$$\mathbb{P}(M) = \frac{3}{10}, \quad \mathbb{P}(A) = \frac{2}{10}, \quad \mathbb{P}(C) = \frac{5}{10}.$$

The associated event probabilities are

$$\mathbb{P}(\emptyset) = 0\ , \quad \mathbb{P}(\{M\}) = \tfrac{3}{10}, \quad \mathbb{P}(\{A\}) = \tfrac{2}{10}, \quad \mathbb{P}(\{C\}) = \tfrac{5}{10},$$
$$\mathbb{P}(\{M, A\}) = \tfrac{5}{10}, \quad \mathbb{P}(\{A, C\}) = \tfrac{7}{10}, \quad \mathbb{P}(\{M, C\}) = \tfrac{8}{10}, \quad \mathbb{P}(S) = 1\ .$$

The Monty Hall problem

A game: Three doors, with a prize behind one of them.
The contestant chooses one door.
The host, who knows where the prize is, opens one of the other two doors, revealing that the prize is not there.

The contestant can then change her/his choice, or not.

Should she/he change doors?

Before the host opens the door:
Sample space is $S = \{d_1, d_2, d_3\}$ with $\mathbb{P}(d_1) = \mathbb{P}(d_2) = \mathbb{P}(d_3) = \frac{1}{3}$.
The contestant chooses d_1 and has a $\frac{1}{3}$ chance of winning.

The host then opens door 2:
The sample space is now $S' = \{d_1, d_3\}$, but $\mathbb{P}(d_1)$ remains at $\frac{1}{3}$ (since the prize hasn't moved) so $\mathbb{P}(d_3) = \frac{2}{3}$ since the sum of the probabilities must be 1.
Thus the prize is twice as likely to be behind door 3 as behind door 1.

So the contestant should change doors.

Properties of probabilities

Theorem: Let S be a finite set and $\mathbb{P} : S \to \mathbb{Q}_+$ a probability density function. Then:

(i) $\forall E \in \mathcal{P}(S) \quad 0 \leq \mathbb{P}(E) \leq 1$.

(ii) For $E, F \in \mathcal{P}(S)$ with $E \cap F = \emptyset$, $\quad \mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(F)$.

(iii) For *any* $E, F \in \mathcal{P}(S)$, $\quad \mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(F) - \mathbb{P}(E \cap F)$.

(iv) $\mathbb{P}(E^c) = 1 - \mathbb{P}(E) \quad \forall E \in \mathcal{P}(S)$.

The birthday problem

In a group of 50 people, what is the probability that two people have the same birthday (assuming all birthdays are equally likely)?

Let $D = \{1, ..., 365\}$ represent the days of the year.

Then the sample space is $S = D^{50} = \{(a_n)_{n=1,..,50} \subseteq D\}$, representing all sequences of 50 birthdays.

We have $|S| = |D^{50}| = |D|^{50} = 365^{50}$.

The event E of 'two persons have the same birthday' is $E = \{(a_n)_{n=1,..,50} \subseteq D \; ; \; \exists j, k \in \{1, ..., 50\} \; a_j = a_k\}$.

The complementary event 'everybody has a different birthday' is $E^c = \{(a_n)_{n=1,..,50} \subseteq D \; ; \; \forall j, k \in \{1, ..., 50\} \; j \neq k \implies a_j \neq a_k\}$.

Thus

$$\mathbb{P}(E^c) = \frac{|E^c|}{|S|} = \frac{P(365, 50)}{365^{50}} = \frac{365!}{365^{50} \times 315!} \sim 0.03.$$

$$\mathbb{P}(E) = 1 - \mathbb{P}(E^c) \sim 0.97.$$

There is a 97% chance that two people will have the same birthday.

Random Variables, Expected Values and Independence

Random variables and Expected values

A (simple) **random variable** on a sample space S is any function $X : S \to \mathbb{Q}$. (More generally, $S \to \mathbb{Q}^m$ but we will stick to $m = 1$.)

Note: We will denote the event 'the random variable X is equal to a' by just $\{X = a\}$ instead of the more formal $\{s \in S \; ; \; X(s) = a\}$.

Example:

$S = \{\text{H,T}\}^3 =$ set of outcomes of tossing three coins.
$X((a, b, c)) =$ number of H's amongst a, b, c.
$\{X = 2\} = \{\text{HHT,HTH,THH}\}$.

Relative to a probability density function $\mathbb{P} : S \to \mathbb{Q}_+$ the **expected value** $\mathbb{E}(X)$ of a random variable X is defined by

$$\boxed{\mathbb{E}(X) = \sum_{s \in S} \mathbb{P}(s) X(s) = \sum_{a \in \text{Range}(X)} \mathbb{P}(\{X = a\}) a}$$

Example(cont.):

$$\mathbb{E}(X) = (\tfrac{1}{8})0 + (\tfrac{3}{8})1 + (\tfrac{3}{8})2 + (\tfrac{1}{8})3 = \tfrac{12}{8} = 1.5.$$

Thus the expected value of X is just the mean (average) number of heads obtained when three coins are tossed.

Die roll example of expected value

Game: \$ 2 to play. Roll a die. Win \$10 if you get a 6. Play many games. Should you expect to make or lose money? How much?

$S = \{1, 2, 3, 4, 5, 6\}$.

$\mathbb{P} : S \to \mathbb{Q}_+$ given by $p(j) = \frac{1}{6} \quad \forall j \in \{1, ..., 6\}$.

$\mathbb{P} : \mathcal{P}(S) \to \mathbb{Q}_+$ given by $\mathbb{P}(E) = \frac{|E|}{6}$ (equally likely outcomes).

$X : S \to \mathbb{Q}$ defined by $X(j) = \begin{cases} \text{10-2=8} & \text{if } j = 6, \\ -2 & \text{otherwise.} \end{cases}$

X is your gain (or loss), which is a random variable.

$$\mathbb{E}(X) = \sum_{j=1}^{6} \frac{1}{6} \times X(j) = 5\left(\frac{1}{6} \times -2\right) + \left(\frac{1}{6} \times 8\right) = \frac{-2}{6} = -\frac{1}{3}.$$

If you play this game 30 times, you should expect to *loose* $30(\frac{1}{3}) = 10$ dollars.

Independent Events

For a sample space S with probability density function $\mathbb{P} : S \to \mathbb{Q}_+$, $E, F \in \mathcal{P}(S)$ are called **independent events** when

$$\boxed{\mathbb{P}(E \cap F) = \mathbb{P}(E) \times \mathbb{P}(F)}$$

Illustration:

Toss two coins:
$S = \{\text{H,T}\}^2 = \{\text{HH,HT,TH,TT}\}$ with equally likely outcomes.

- $E = \{\text{HH,HT}\}$ (1st coin gives Head), $\mathbb{P}(E) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$,
  $F = \{\text{HT,TT}\}$ (2nd coin gives Tail), $\mathbb{P}(F) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$.
  E, F are independent (as we would expext) since
  $$\mathbb{P}(E \cap F) = \mathbb{P}(\{\text{HT}\}) = \frac{1}{4} = \frac{1}{2} \times \frac{1}{2} = \mathbb{P}(E) \times \mathbb{P}(F).$$

- $G = \{\text{HT,TH,HH}\}$ (at least one Head), $\mathbb{P}(G) = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4}$,
  $K = \{\text{TH,HT,TT}\}$ (at least one Tail), $\mathbb{P}(K) = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4}$.
  G, K are **not** independent (again as we would expect) since
  $$\mathbb{P}(G \cap K) = \mathbb{P}(\{\text{HT,TH}\}) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2} \neq \frac{9}{16} = \frac{3}{4} \times \frac{3}{4} = \mathbb{P}(G) \times \mathbb{P}(K).$$

Independent random variables

For a sample space S with probability density function $\mathbb{P} : S \to \mathbb{Q}_+$, $X, Y : S \to \mathbb{Q}$ are called **independent random variables** when

$$\boxed{\begin{array}{l}\forall a \in \text{Range}(X) \;\; \forall b \in \text{Range}(Y) \\ \{X = a\}, \{Y = b\} \text{ are independent.}\end{array}}$$

Illustration:

The last illustration showed that, when tossing two coins, getting T on the second toss is independent of getting H on the first.

Using random variables we get the expected result that *any* results of the two tosses are independent:

Let X, Y be number of heads (0 or 1) on the 1st, 2nd toss respectively.

Then for *any* $a, b \in \{0, 1\}$:

$\mathbb{P}(\{X = a\}) = \mathbb{P}(\{Y = b\}) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$ and $\mathbb{P}(\{X = a\} \cap \{Y = b\}) = \frac{1}{4}$,

and hence the events $\{X = a\}, \{Y = b\}$ are independent because

$$\mathbb{P}(\{X = a\} \cap \{Y = b\}) = \frac{1}{4} = \frac{1}{2} \times \frac{1}{2} = \mathbb{P}(\{X = a\}) \times \mathbb{P}(\{Y = b\}).$$

Thus, by the above definition, X, Y are independent.

Independent random variables — Example

Toss a regular fair die. $S = \{1, \ldots, 6\}$, $\mathbb{P}(i) = \frac{1}{6}$, $i = 1, \ldots, 6$.
Let $X, Y : S \to \mathbb{Q}$ be random variables as follows:

Table 1: Definition of X and Y

s	1	2	3	4	5	6
$s \bmod 2 = X(s)$	1	0	1	0	1	0
$s \bmod 3 = Y(s)$	1	2	0	1	2	0

Table 2: Probabilities

a	0	1	2
$\mathbb{P}(\{X = a\})$	$\frac{1}{2}$	$\frac{1}{2}$	0
$\mathbb{P}(\{Y = a\})$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

The columns in Table 1 are all different and cover all possible combinations of values of X, Y. This ensures that each pair of values $(X,Y) = (a,b)$ relates to a unique s (Table 3), and hence has probability $\mathbb{P}(s)$ $(= \frac{1}{6})$.

Table 3: s

$b =$	0	1	2
$a = 0$	6	4	2
$a = 1$	3	1	5

Using Table 2 it now follows that, for any $a \in \{0, 1\}$ $b \in \{0, 1, 2\}$ the events $\{X = a\}$,$\{Y = b\}$ are independent because

$$\mathbb{P}(\{X = a\} \cap \{Y = b\}) = \frac{1}{6} = \frac{1}{2} \times \frac{1}{3} = \mathbb{P}(\{X = a\}) \times \mathbb{P}(\{Y = b\}).$$

Thus, by definition, the random variables X, Y are independent.

Non-independent random variables — Example

Let's modify the previous example just a little:

Toss a regular fair die. $S = \{1, \dots, 6\}$, $\mathbb{P}(i) = \frac{1}{6}$, $i = 1, \dots, 6$.
Let $Y, Z : S \to \mathbb{Q}$ be random variables as follows:

Table 1: Definition of Y and Z

s	1	2	3	4	5	6
$s \bmod 3 = Y(s)$	1	2	0	1	2	0
$s \bmod 4 = Z(s)$	1	2	3	0	1	2

Table 2: Probabilities

a	0	1	2	3
$\mathbb{P}(\{Y = a\})$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	0
$\mathbb{P}(\{Z = a\})$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{6}$

Notice that, in the Table 1, many potential columns are not present.
For example there is no value of s for which $Y(s) = 0$ and $Z(s) = 0$.
Using this, and also Table 2, we now have

$$\mathbb{P}(\{Y = 0\} \cap \{Z = 0\}) = 0 \neq \frac{1}{3} \times \frac{1}{6} = \mathbb{P}(\{Y = 0\}) \times \mathbb{P}(\{Z = 0\}),$$

and so the events $\{Y = 0\}$, $\{Z = 0\}$ are not independent.

It follows that the random variables Y, Z are not independent.

Challenge: Are the random variables X, Z independent?

Conditional Probability and Bayes' Theorem

Reference: §9.9 of our optional text

Conditional Probability

(Theme: Use all of the information you have.)

Definition

Consider a probability experiment with sample space S. If $A, B \subseteq S$ and $\mathbb{P}(A) \neq 0$, then the **conditional probability of** B **given** A, denoted $\mathbb{P}(B|A)$, is

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

An example

P: I toss two fair coins but only I can see the outcome. You ask "Did they both come up tails?" I say "No."

What is the probability that both coins came up heads?

A: The probability experiment is to toss two fair coins.

An outcome will be recorded as a two letter string using only H's and T's, with the first letter recording the result of tossing the first coin and the second letter the result of tossing the second coin. For example, the outcome HT records that the first coin come up 'heads' and the second coin came up 'tails'.

The sample space is the set $S = \{HH, HT, TH, TT\}$

example (cont.)

Since the coins are 'fair', it is reasonable to suppose that outcomes are equally likely. We then have that $\mathbb{P}(E) = \frac{|E|}{|S|} = \frac{|E|}{4}$ for each event $E \subseteq S$.

Let A be the event that we did not have both coins coming up tails; that is, $A = \{HH, HT, TH\}$. Let B be the event that both coins came up heads; that is, $B = \{HH\}$. We compute

$$\begin{aligned}
&\text{The probability that both coins came up heads given that they}\\
&\text{ did not both come up tails}\\
=&\mathbb{P}(B|A) \qquad \text{(translating into notation)}\\
=&\frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} \qquad \text{(defn of conditional prob.)}\\
=&\frac{\mathbb{P}(\{HH\})}{\mathbb{P}(\{HH, HT, TH\})}\\
=&\frac{1/4}{3/4} = \frac{1}{3}.
\end{aligned}$$

$\square$

An example

P: A pair of fair 6-sided dice, one red and one blue, are rolled. What is the probability that the sum of the numbers showing face up is 8, given that both of the numbers are even?

Proof: The probability experiment is to roll a pair of fair 6-sided dice, one red and one blue.

An outcome will be recorded as an element of $\{1,2,3,4,5,6\} \times \{1,2,3,4,5,6\}$, with the first element recording the result of the red die and the second digit the result of rolling the blue die. For example, the outcome $(2,4)$ records that we rolled a 2 on the red die and a 4 on the blue die.

The sample space is the set $S = \{1,2,3,4,5,6\} \times \{1,2,3,4,5,6\}$. By the product rule for counting, $|S| = 6^2 = 36$.

example (cont.)

Since the dice are 'fair', it is reasonable to suppose that outcomes are equally likely. We then have that $\mathbb{P}(E) = \frac{|E|}{|S|} = \frac{|E|}{36}$ for each event $E \subseteq S$.

Let B be the event that the sum of the numbers showing face up is 8; that is, $\mathrm{B} = \{(2,6),(3,5),(4,4),(5,3),(6,2)\}$. Let A be the event that both of the numbers rolled are even; that is,

$$A = \{(2,2),(2,4),(2,6),(4,2),(4,4),(4,6),(6,2),(6,4),(6,6)\}.$$

example (cont.)

We compute:

$$
\begin{aligned}
&\text{The probability that the sum of the numbers showing face up is 8}\\
&\text{given that both numbers are even}\\
=&\mathbb{P}(B|A) \qquad \text{(translating into notation)}\\
=&\frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} \qquad \text{(defn of conditional prob.)}\\
=&\frac{\mathbb{P}(\{(2,6),(4,4),(6,2)\})}{\mathbb{P}(\{(2,2),(2,4),(2,6),(4,2),(4,4),(4,6),(6,2),(6,4),(6,6)\})}\\
=&\frac{3/36}{9/36}\\
&= \frac{1}{3}.
\end{aligned}
$$

$\square$

A lemma

Lemma

For any probability experiment with sample space S, and for any events $A, B \subseteq S$, if $\mathbb{P}(A) \neq 0$ then

$$\mathbb{P}(A \cap B) = \mathbb{P}(B|A)\mathbb{P}(A).$$

Proof.

Consider a probability experiment with sample space S. Let $A, B \subseteq S$. Suppose that $\mathbb{P}(A) \neq 0$. Since $\mathbb{P}(A) \neq 0$, the conditional probability $\mathbb{P}(B|A)$ is defined. The definition gives

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

Multiplying both sides by $\mathbb{P}(A)$ gives $\mathbb{P}(B|A)\mathbb{P}(A) = \mathbb{P}(A \cap B)$. □

Bayes' Theorem

Theorem (Bayes' Theorem)

For any probability experiment with sample space S, for any $n \in \mathbb{N}$, for any partition $\{B_1, B_2, \ldots, B_n\}$ of S and for any event $A \subseteq S$, if $\mathbb{P}(A) \neq 0$ and for all $i \in \{1, 2, \ldots, n\}$ we have $\mathbb{P}(B_i) \neq 0$, then for all $k \in \{1, 2, \ldots, n\}$ we have

$$\mathbb{P}(B_k|A) = \frac{\mathbb{P}(A|B_k)\mathbb{P}(B_k)}{\sum_{i=1}^{n} \mathbb{P}(A|B_i)\mathbb{P}(B_i)}$$

Proof

Consider a probability experiment with sample space S. Let $n \in \mathbb{N}$, let $\{B_1, B_2, \ldots, B_n\}$ be a partition of S and let $A \subseteq S$. Suppose that $\mathbb{P}(A) \neq 0$ and for all $i \in \{1, 2, \ldots, n\}$ we have $\mathbb{P}(B_i) \neq 0$. Let $k \in \{1, 2, \ldots, n\}$. Now

$$\mathbb{P}(B_k|A) = \frac{\mathbb{P}(B_k \cap A)}{\mathbb{P}(A)} \qquad \text{(By defn of } \mathbb{P}(B_k|A))$$

$$= \frac{\mathbb{P}(A|B_k)\mathbb{P}(B_k)}{\mathbb{P}(A)}$$

(Applying the lemma, which is OK because $\mathbb{P}(B_k) \neq 0$

$$= \frac{\mathbb{P}(A|B_k)\mathbb{P}(B_k)}{\mathbb{P}(A \cap S)} \qquad \text{(Because } A \cap S = A)$$

$$= \frac{\mathbb{P}(A|B_k)\mathbb{P}(B_k)}{\mathbb{P}(A \cap (B_1 \cup B_2 \cup \cdots \cup B_n))}$$

(Because $\{B_1, \ldots, B_n\}$ is a partition of S we have $S = B_1 \cup B_2 \cup \cdots \cup B_n$)

Proof (cont.)

$$= \frac{\mathbb{P}(A|B_k)\mathbb{P}(B_k)}{\mathbb{P}((A \cap B_1) \cup (A \cap B_2) \cup \cdots \cup (A \cap B_n))} \quad \text{($\cap$ distributes over $\cup$)}$$

$$= \frac{\mathbb{P}(A|B_k)\mathbb{P}(B_k)}{\mathbb{P}(A \cap B_1) + \mathbb{P}(A \cap B_1) + \cdots + \mathbb{P}(A \cap B_n)} \quad \text{(Applying the sum rule}$$

which is OK because $B_1, \ldots, B_n$ are mutually disjoint)

$$= \frac{\mathbb{P}(A|B_k)\mathbb{P}(B_k)}{\mathbb{P}(A|B_1)\mathbb{P}(B_1) + \mathbb{P}(A|B_2)\mathbb{P}(B_2) \cdots + \mathbb{P}(A|B_n)\mathbb{P}(B_n)}$$

(Applying the lemma n times, which is OK because $\mathbb{P}(B_i) \neq 0$

for $i \in \{1, 2, \ldots, n\}$)

$$= \frac{\mathbb{P}(A|B_k)\mathbb{P}(B_k)}{\sum_{i=1}^{n} \mathbb{P}(A|B_i)\mathbb{P}(B_i)} \quad \text{(Using Σ notation)} \qquad \square$$

Example 9.9.3 Applying Bayes' Theorem

Most medical tests occasionally produce incorrect results, called false positives and false negatives. When a test is designed to determine whether a patient has a certain disease, a **false positive** result indicates that a patient has the disease when the patient does not have it. A **false negative** result indicates that a patient does not have the disease when the patient does have it.

When large-scale health screenings are performed for diseases with relatively low incidence, those who develop the screening procedures have to balance several considerations: the per-person cost of the screening, follow-up costs for further testing of false positives, and the possibility that people who have the disease will develop unwarranted confidence in the state of their health.

Consider a medical test that screens for a disease found in 5 people in 1,000. Suppose that the false positive rate is 3% and the false negative rate is 1%. Then 99% of the time a person who has the condition tests positive for it, and 97% of the time a person who does not have the condition tests negative for it. (See exercise 4 at the end of this section.)

a. What is the probability that a randomly chosen person who tests positive for the disease actually has the disease?

b. What is the probability that a randomly chosen person who tests negative for the disease does not in fact have the disease?

Solution Consider a person chosen at random from among those screened. Let A be the event that the person tests positive for the disease, B_1 the event that the person actually has the disease, and B_2 the event that the person does not have the disease. Then

$$P(A|B_1) = 0.99, \quad P(A^c|B_1) = 0.01, \quad P(A^c|B_2) = 0.97, \quad \text{and} \quad P(A|B_2) = 0.03.$$

Also, because 5 people in 1,000 have the disease,

$$P(B_1) = 0.005 \quad \text{and} \quad P(B_2) = 0.995.$$

a. By Bayes' theorem,

$$\begin{aligned} P(B_1|A) &= \frac{P(A|B_1)P(B_1)}{P(A|B_1)P(B_1) + P(A|B_2)P(B_2)} \\ &= \frac{(0.99)(0.005)}{(0.99)(0.005) + (0.03)(0.995)} \\ &\cong 0.1422 \cong 14.2\%. \end{aligned}$$

Thus the probability that a person with a positive test result actually has the disease is approximately 14.2%.

b. By Bayes' theorem,

$$P(B_2|A^c) = \frac{P(A^c|B_2)P(B_2)}{P(A^c|B_1)P(B_1) + P(A^c|B_2)P(B_2)}$$

$$= \frac{(0.97)(0.995)}{(0.01)(0.005) + (0.97)(0.995)}$$

$$\cong 0.999948 \cong 99.995\%.$$

Thus the probability that a person with a negative test result does not have the disease is approximately 99.995%.