Section 4.1 - Randomness
Section 4.2 - Probability Models

Statistics 104

Autumn 2004

Copyright ©2004 by Mark E. Irwin

Randomness

I am going to roll 2 dice, one yellow and one purple.

- What is the probability that the yellow die comes up 5?
- What is the probability that the purple die comes up 5?
- What is the probability that both dice come up 5?

What is probability?

A mathematical approach to modelling randomness.

Random Phenomenon

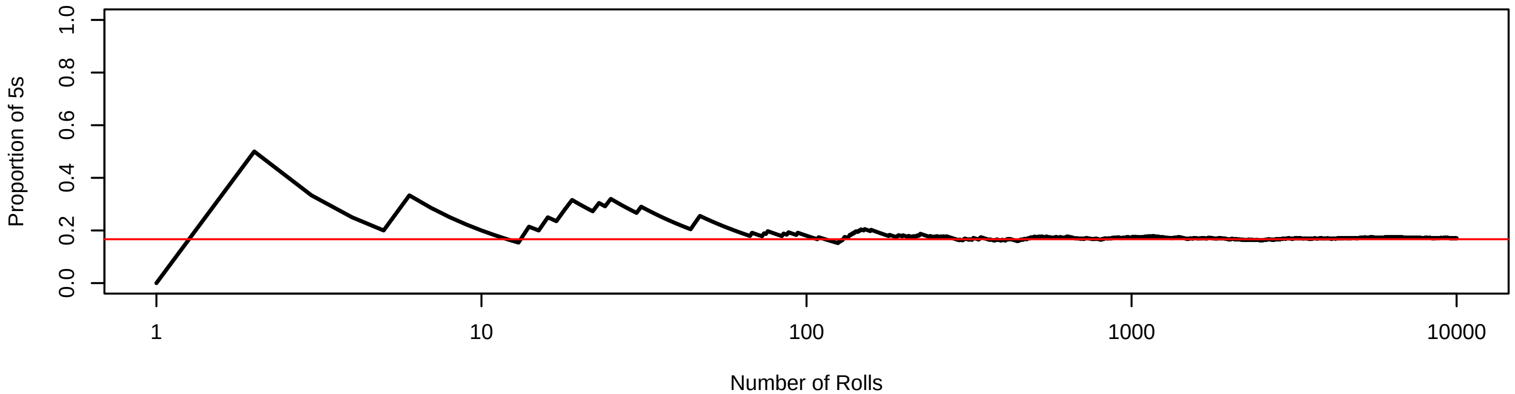
- Outcome in a single trial is uncertain
- Regular distribution of relative frequencies in a large number of repetitions

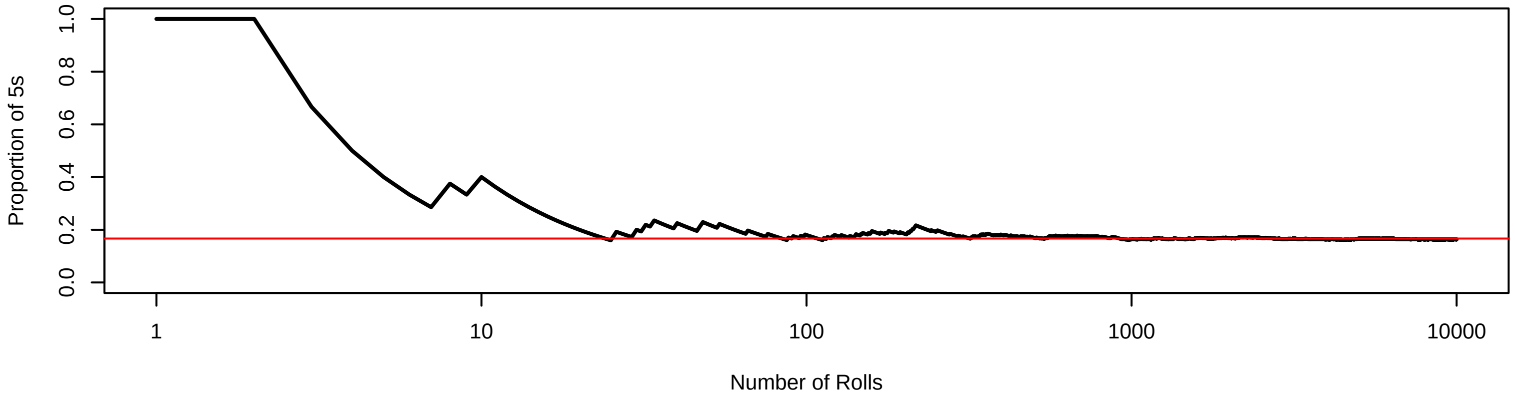
Examples:

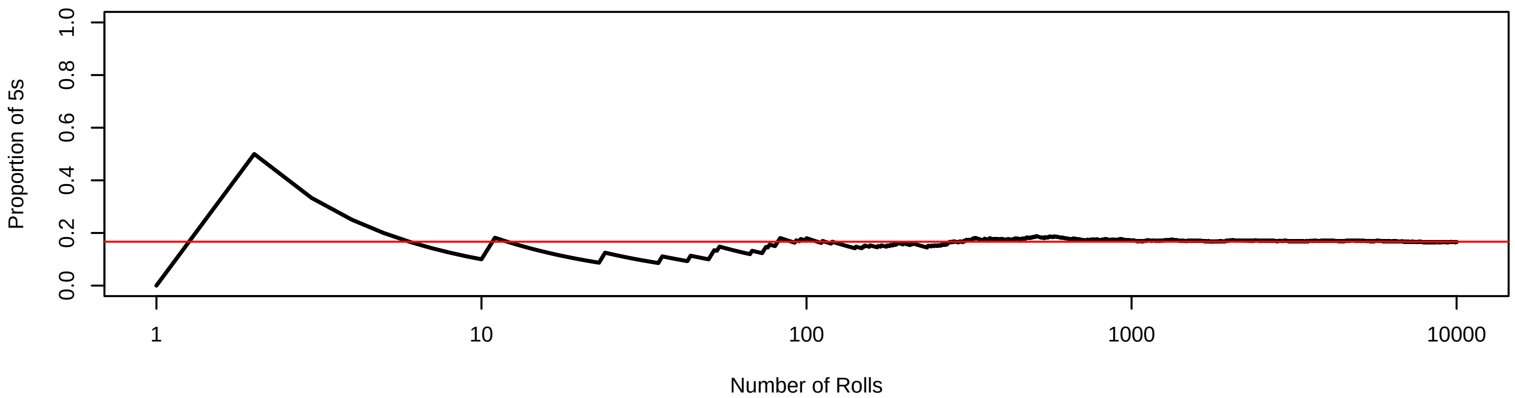
1. Flipping a fair coin - Proportion of heads $\Rightarrow \frac{1}{2}$
2. Rolling a fair 6 sided die - Proportion of 5's $\Rightarrow \frac{1}{6}$

This gives us one way of thinking about probability: Long run frequencies.

Usually you can't really determine probabilities this way, but it is a useful motivation.


Proportion of 5s
0.0
0.2
0.4
0.6
0.8
1.0
1
10
100
1000
10000
Number of Rolls


Proportion of 5s
0.0
0.2
0.4
0.6
0.8
1.0
1
10
100
1000
10000
Number of Rolls


Proportion of 5s
0.0
0.2
0.4
0.6
0.8
1.0
1
10
100
1000
10000
Number of Rolls

This idea of thinking about probability doesn't always work.

Example: What is the chance that the Red Sox beat the Yankees in tonight's game?

Its hard to think about repeated trials here. The conditions change after every game.

However we can consider odds that would lead to a "fair" bet.

Can also compare different possibilities, thinking about how much more likely one possible outcome is than another.

These ideas lead to personal, or subjective, probability.

We often need to consider both.

Probability Models

Let's get back to the original question

What is the probability of rolling a 5?

Need to consider all the possible outcomes

- What type of die is it? 4, 6, 8, 12, 20 sided?
- What are the numbers on each face?
- Is there other information about the die which might affect my belief about the chance of a 5 coming up?

Question: What is the chance of getting a head when flipping a coin?

What can happen?

- Head
- Tail
- You have a double headed coin
- Lands on edge
- Falls down sewer
- Caught by bird passing overhead

Sample Space:

The set of all possible outcomes of a random phenomenon. Often denoted by S.

Examples:

- Rolling a 4-sided die: $S = \{1, 2, 3, 4\}$. (Assuming that weird things can't happen)

- Flipping a coin twice: $S = \{HH, HT, TH, TT\}$ (order matters)

Sometimes you are interesting in more than one of the possible outcomes, for example, are both flips the same (HH or TT). This is an example of ...

Event:

An **event** is an outcome or a set of outcomes of a random phenomenon (a subset of the sample space).

Example: Roll a 4 sided die twice

$$S = \left\{ \begin{array}{cccc} (1,1) & (1,2) & (1,3) & (1,4) \\ (2,1) & (2,2) & (2,3) & (2,4) \\ (3,1) & (3,2) & (3,3) & (3,4) \\ (4,1) & (4,2) & (4,3) & (4,4) \end{array} \right\}$$

Possible events of interest

1. Both rolls are the same - $A = \{(1,1),(2,2),(3,3),(4,4)\}$

2. Sum is less than 4 - $B = \{(1,1),(1,2),(2,1)\}$

3. Sum is 8 - $C = \{(4,4)\}$

4. First roll is less than second roll - $D = \{(1,2),(1,3),(1,4),(2,3),(2,4),(3,4)\}$

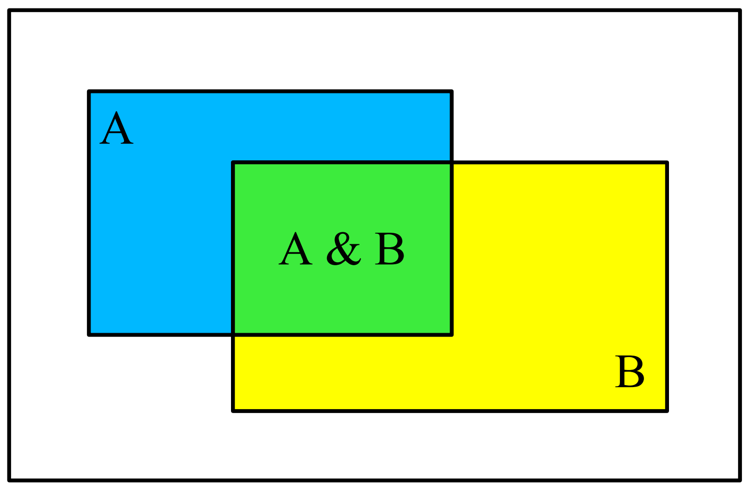
Need to be able to define probabilities for all possible events, e.g. $P[\text{Both rolls the same}] = P[A]$ or $P[\text{sum } < 4] = P[B]$

This can be difficult since their are lots of different events. For the case where two 4-sided dice are rolled, there are 65536 $(= 2^{16})$ different possible events to be considered.

However it can be done by defining probabilities on a small set of events and building up from that.

Combining Events

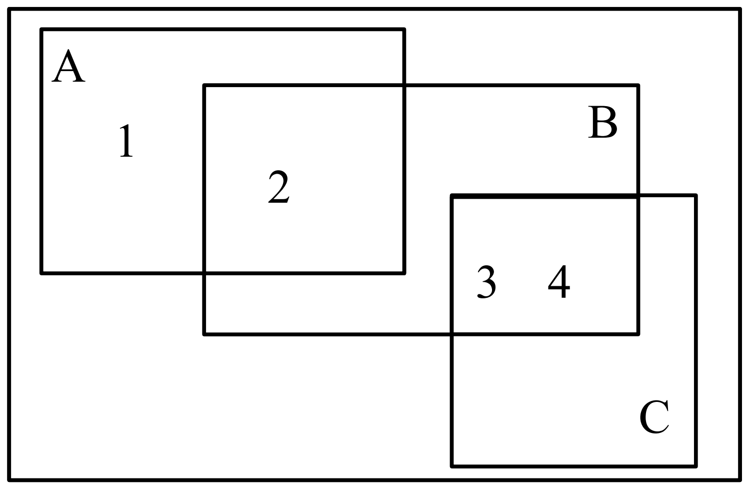
- Intersection (and): The set of all outcomes that occur in all of the sets



$$A \text{ and } B = B \text{ and } A \qquad \text{(Commutative)}$$

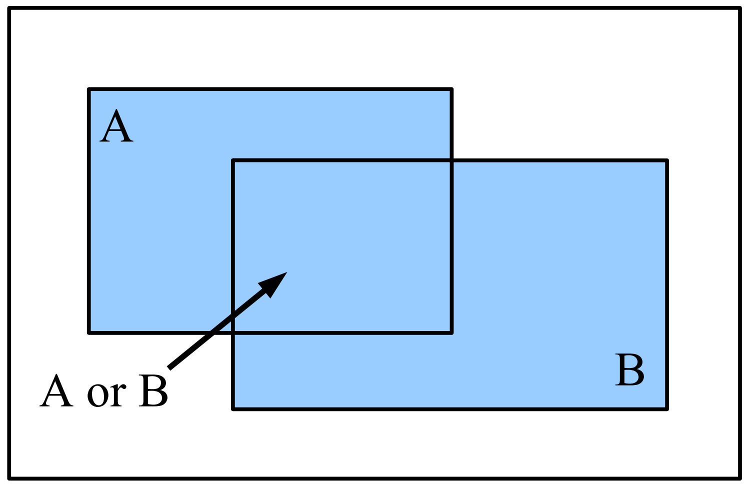
If $A = \{1, 2\}, B = \{2, 3, 4\}$ and $C = \{3, 4\}$, then

$$\begin{aligned} A \text{ and } B &= \{2\} \\ A \text{ and } C &= \phi \qquad \text{(Empty Set)} \\ B \text{ and } C &= \{3, 4\} \end{aligned}$$



If A and $B = \phi$, the events A and B are said to be **disjoint**.

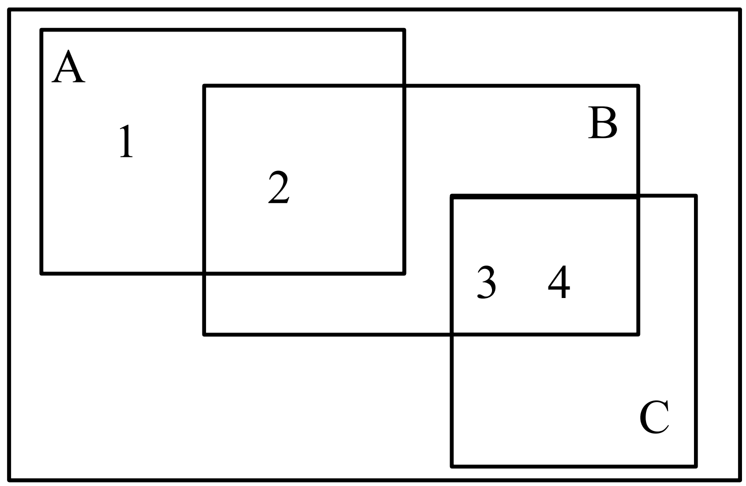
- Union (or): The set of all outcomes that occur in at least one of the sets



$$A \text{ or } B = B \text{ or } A \qquad \text{(Commutative)}$$

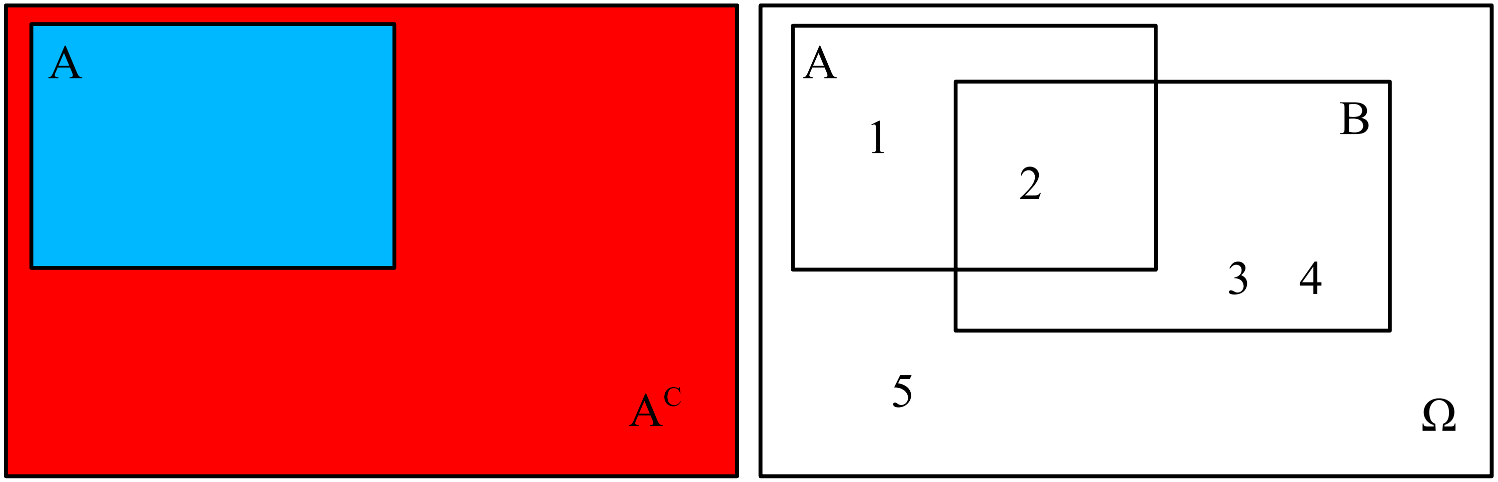
If $A = \{1, 2\}, B = \{2, 3, 4\}$ and $C = \{3, 4\}$, then

$$\begin{aligned} A \text{ or } B &= \{1, 2, 3, 4\} \\ A \text{ or } C &= \{1, 2, 3, 4\} \\ B \text{ or } C &= \{2, 3, 4\} \end{aligned}$$



Note that "or" means one, or the other, or both (not exclusive or).

- Complement: The set of outcomes that don't occur in the event



If $A = \{1, 2\}, B = \{2, 3, 4\}$ and $S = \{1, 2, 3, 4, 5\}$, then

$$\begin{aligned} A^c &= \{3, 4, 5\} \\ B^c &= \{1, 5\} \\ S^c &= \phi \end{aligned}$$

Basic rules of probability

1. The probability of an event $A, P[A]$ is a number between 0 and 1.

$$0 \leq P[A] \leq 1$$

2. The collection S of all possible outcomes has probability 1.

$$P[S] = 1$$

3. The probability than an event does not occur is 1 minus the probability that the event does occur

$$P[A^c] = 1 - P[A]$$

(Complement Rule)

4. If two events have no outcomes in common (they are disjoint), the probability that one or other occurs is the sum of their individuals probabilities

$$P[A \text{ or } B] = P[A] + P[B]$$

   Note that this can be extended to more than 2 events.
   (Addition rule)

Any assignment of probabilities to a set of events must satisfy these 4 rules

Aside: Actually all probabilities rules are based on 1,2, and 4. Rules 2 and 4 imply the complement rule for example.

How do we use these rules to get the probability of any event?

Need to start with the probability of each outcome in the sample space.

Example: Roll a fair 4-sided die twice and record the sum.

Outcome	Probability
2	$\frac{1}{16}$
3	$\frac{2}{16}$
4	$\frac{3}{16}$
5	$\frac{4}{16}$
6	$\frac{3}{16}$
7	$\frac{2}{16}$
8	$\frac{1}{16}$
Total	1

Note: The sum of the probabilities of the individuals outcomes must sum to 1 (Rule 2).

Any other set of 7 numbers that are between 0 and 1 and sum to 1 would be another possible set of probabilities, though not an accurate description of this probability experiment.

Then the probability of any event is the sum of the probabilities of each outcome making up the event (by rule 4).

- $A =$ sum is odd $= \{3, 5, 7\}$

$$\begin{aligned} P[A] &= P[\text{sum } = 3] + P[\text{sum } = 5] + P[\text{sum } = 7] \\ &= \frac{2}{16} + \frac{4}{16} + \frac{2}{16} = \frac{1}{2} \end{aligned}$$

- $B =$ sum is even $= \{2, 4, 6, 8\}$

$$\begin{aligned} P[B] &= P[\text{sum } = 2] + P[\text{sum } = 4] + P[\text{sum } = 6] + P[\text{sum } = 8] \\ &= \frac{1}{16} + \frac{3}{16} + \frac{3}{16} + \frac{1}{16} = \frac{1}{2} \end{aligned}$$

Note that $B = A^c$, so we could have used rule 3 instead to get

$$P[B] = 1 - P[A] = 1 - \frac{1}{2} = \frac{1}{2}$$

How to get the probabilities for each outcome?

We can't use the above rules unless we have the probabilities.

Often need long term observations of the phenomenon to estimate them.

For example, three different experiments looked at the probability of getting a head when flipping a coin.

- The French naturalist Count Buffon: 4040 tosses, 2048 heads ($\hat{p} = 0.5069$).

- While imprisoned during WWII, the South African statistician John Kerrich: 10000 tosses, 5067 heads ($\hat{p} = 0.5067$)

- Statistician Karl Pearson: 24000 tosses, 12012 heads ($\hat{p} = 0.5005$)

Sometimes you can appeal to the structure of the situation or intuition $\Rightarrow$ Probabilities models.

One possible model - equally likely outcomes (Discrete Uniform Distribution)

- Fair coin - 2 poss
- Fair 4-sided die - 4 possibilities

$$P[\text{Outcome}] = \frac{1}{\#\ \text{Possibilities}}$$

- Fair coin - $P[\text{Head}] = P[\text{Tail}] = \frac{1}{2}$
- Fair 4-sided die - $P[1] = P[2] = P[3] = P[4] = \frac{1}{4}$

If there are equally likely outcomes, then for any event A,

$$P[A] = \frac{\#\text{outcomes in } A}{\#\text{outcomes in } S}$$

This was the basis of the probabilities for the sum of two rolls earlier. The 4 possibilities on the first roll times the 4 possibilities on the second roll gives a total of 16 different pairs of rolls.

	1	2	3	4
1	2	3	4	5
2	3	4	5	6
3	4	5	6	7
4	5	6	7	8

Lets consider the following 3 events

- $A =$ Sum is even $= \{2, 4, 6, 8\}$
- $B =$ Sum is 5 $= \{5\}$
- $C =$ Sum $>$ 6 $= \{7, 8\}$

For these events, the 3 pairs satisfy

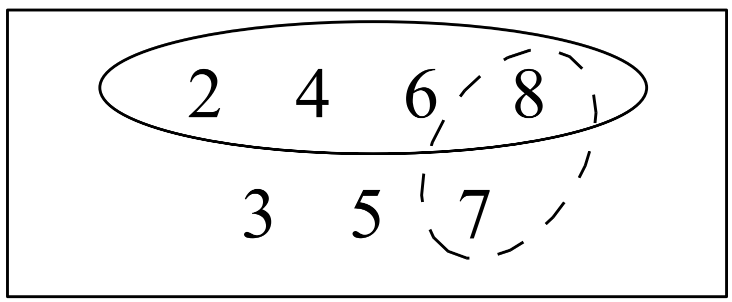
- $A \& B$: Disjoint
- $A \& C$: Not disjoint (8 in both)
- $B \& C$: Disjoint

So we can use the addition rule to get $P[A \text{ or } B]$ and $P[B \text{ or } C]$, but not $P[A \text{ or } C]$

What happens if we try to use the addition rule to determine $P[A \text{ or } C]$

$$
\begin{aligned}
P[A \text{ or } C] &= \frac{10}{16} \\
P[A] + P[C] &= \frac{8}{16} + \frac{3}{16} = \frac{11}{16} > P[A \text{ or } C]
\end{aligned}
$$

In fact the amount we are over is $\frac{1}{16} = P[8]$



Using the addition rule when events are not disjoint will give a probability that's too big. We'll come back to a correction in section 4.5.

Independence and the Multiplication Rule

In many situations, events that we are interested are a combination of two or more random phenomena.

- What is the probability that two dice both come up 5?
- What is the probability that a coin flipped twice comes up a head on the first flip and a tail on the second?
- In a quality control check, 10 switches are examined to see if they work correctly. What is the probability that all 10 are ok?
- In a similar problem, there are switches from two companies available. If I choose a switch at random, what is the probability that I select a switch from company 1 and it works properly? What about company 2?

In some situations, like the dice rolling and coin flipping examples, knowing what occurs on the first flip or roll shouldn't affect your belief of what happens on the second flip or roll.

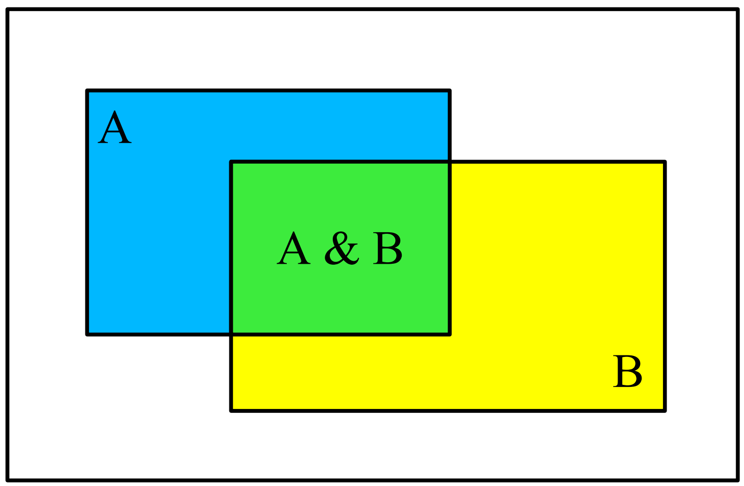
However for the situation with the two companies switches, if I know which company the switch comes from, I probably want to change my belief about whether the switch works or not.

For the first switch example, it isn't clear you would want to change your belief about the 10th switch based on the earlier 9 or not. More likely it should be closer to the coin flipping situation.

Rule 5. Multiplication Rule for Independent Events

Two event A and B are **independent** if knowing that one occurs does not change the probability that the other occurs. If A and B are independent, then

$$P[A \text{ and } B] = P[A]P[B]$$



As the book notes, this definition of independence is informal and we will make it precise later.

This rule, will give us for example a probability of

$$\frac{1}{36} = \frac{1}{6} \times \frac{1}{6}$$

for the chance of getting two 5's when two fair 6-sided dice are rolled.

Note that independence is an assumption that may or may not be valid for a particular problem. (Later we will discuss how to test this assumption in real problems.)

However if it is reasonable, we can answer questions based on the first switch example

1. What is the probability that the first two switches checked are ok?

2. What is the probability that one of the first two is ok and the other is faulty.

3. What is the probability that all 10 are ok?

4. What is the probability that at least one of the 10 are faulty?

Lets assume that for each switch $P[\text{OK}] = 0.95$ and $P[\text{Faulty}] = 0.05$

Solutions:

1. What is the probability that the first two switches checked are ok?

$$
\begin{aligned}
P[\text{Both OK}] &= P[1^{st} \text{ OK and } 2^{nd} \text{ OK}] \\
&= P[1^{st} \text{ OK}] \times P[2^{nd} \text{ OK}] \\
&= 0.95 \times 0.95 = 0.9025
\end{aligned}
$$

2. What is the probability that one of the first two is ok and the other is faulty.

$$
\begin{aligned}
&P[\text{One OK and one Faulty}] \\
&\quad = P[1^{st} \text{ OK and } 2^{nd} \text{ Faulty}] + P[1^{st} \text{ Faulty and } 2^{nd} \text{ OK}] \\
&\quad = P[1^{st} \text{ OK}] \times P[2^{nd} \text{ Faulty}] + P[1^{st} \text{ Faulty}] \times P[2^{nd} \text{ OK}] \\
&\quad = 0.95 \times 0.05 + 0.05 \times 0.95 = 0.095
\end{aligned}
$$

3. What is the probability that all 10 are ok?

$$
\begin{aligned}
P[\text{All 10 OK}] &= P[1^{st} \text{ OK and } 2^{nd} \text{ OK and } \ldots \text{ and } 10^{th} \text{ OK}] \\
&= P[1^{st} \text{ OK}] \times P[2^{nd} \text{ OK}] \times \ldots P[10^{th} \text{ OK}] \\
&= 0.95 \times 0.95 \times \ldots \times 0.95 = 0.95^{10} = 0.5987
\end{aligned}
$$

4. What is the probability that at least one of the 10 are faulty?

Approach 1:

$$P[\text{At least one Faulty}] = \\ P[\text{Exactly 1 Faulty}] + P[\text{Exactly 2 Faulty}] \\ +P[\text{Exactly 3 Faulty}] + \ldots + P[\text{Exactly 10 Faulty}]$$

We would need to figure out exactly how you could have exactly one faulty switch, two faulty switches, etc, and then figure out the probability of each case with a method analogous to finding $P[1^{st} \text{ OK and } 2^{nd} \text{ Faulty}]$.

This could take a long time.

For example there are 252 different ways to have exactly 5 faulty switches out of the 10.

There are 1024 $(= 2^{10})$ different ways the switches could come out, with 1023 having at least one faulty switch.

Approach 2:

At least one faulty switch is the complement to the event all switches are ok

$$
\begin{aligned}
P[\text{At least one Faulty}] &= 1 - P[\text{All 10 OK}] \\
&= 1 - 0.5987 = 0.4013
\end{aligned}
$$

As mentioned before, independence is an assumption that may or not be true.

Lets look at the Company/Switch example where the probabilities of the different combinations of company and switch status ($P[\text{Company } = x \text{ and Status } = y]$) are

Company	OK	Faulty
1	0.4	0.2
2	0.32	0.08

We can get the probabilities associated with which company a switch was made by and its status by adding across the rows and down the columns of the table, giving what is sometimes referred to as the marginal distribution.

Company	OK	Faulty	$P[\text{Company}]$
1	0.4	0.2	0.6
2	0.32	0.08	0.4
$P[\text{Status}]$	0.72	0.28	1

In this example, company and switch status are not independent since

$$\begin{aligned} P[\text{Company 1 and OK}] &= 0.4 \\ &\neq P[\text{Company 1}]P[\text{OK}] = 0.6 \times 0.72 = 0.432 \end{aligned}$$

As we will see later, if I know that I have a switch from company 1, $P[\text{OK}] = \frac{2}{3}$. If the switch is known to be from company 2, $P[\text{OK}] = \frac{4}{5}$