MAT 2377
Probability and Statistics for Engineers

Chapter 1
Introduction to Probabilities

P. Boily (uOttawa)

Winter 2021

Contents

1.1 – Sample Spaces and Events

We will deal with **random experiments** (e.g. measurements of speed/ weight, number and duration of phone calls, etc.).

For any "experiment," the **sample space** is defined as the set of all possible outcomes. This is often denoted by the symbol $\mathcal{S}$.

A sample space can be **discrete** or **continuous**.

An **event** is a collection of outcomes from the sample space $\mathcal{S}$. Events will be denoted by A, B, E_1, E_2, etc.

Examples:

- Toss a fair coin. The (discrete) sample space is $\mathcal{S} = \{\text{Head, Tail}\}$.

- Roll a die: The (discrete) sample space is $\mathcal{S} = \{1, 2, 3, 4, 5, 6\}$. Various events:
  - Roll an even number: represent this as $\{2, 4, 6\}$.
  - Roll a prime number: $\{2, 3, 5\}$.

- Suppose we measure the weight (in grams) of a chemical sample. The (continuous) sample space can be represented by $(0, \infty)$, the positive half line. Events:
  - sample is less than 1.5 grams: $(0, 1.5)$;
  - sample exceeds 5 grams: $(5, \infty)$;

For any events A and B in $\mathcal{S}$:

- The **union** $A \cup B$ are all outcomes from $\mathcal{S}$ in either A or B;
- The **intersection** $A \cap B$ are all outcomes in both of A and B;
- The **complement** A^c of A (sometimes denoted $\overline{A}$ or $-A$) is the set of all outcomes in $\mathcal{S}$ that are **not** in A;
- If A and B have no outcomes in common, they are **mutually exclusive**; which is denoted by $A \cap B = \varnothing$ (the empty set). In particular, A and A^c are mutually exclusive.
- Graphical representation of events – Venn diagrams $\Rightarrow$ blackboard.

Examples:

- Roll a die. Let $A = \{2, 3, 5\}$ (a prime number) and $B = \{3, 6\}$ (multiples of 3). Then $A \cup B = \{2, 3, 5, 6\}$, $A \cap B = \{3\}$ and $A^c = \{1, 4, 6\}$.

- 100 plastic samples are analyzed for scratch and shock resistance.

		shock resistance	
		high	low
scratch resistance	high	40	4
	low	1	55

If A is the event that a sample has high shock resistance and B is the event that a sample has high scratch residence, then $A \cap B$ consists of 40 samples.

1.2 – Counting Techniques

Some basic combinatorial results make some probabilities easier to calculate.

A **two-stage procedure** can be modeled as having k bags, with m_1 items in the first bag, . . . , m_k items in k-th bag.

The first stage consists of picking a bag, and the second stage consists of drawing an item out of that bag.

This is equivalent to picking one of the $m_1 + m_2 + \cdots + m_k$ total items.

If all the bags have the same number of items $m_1 = \cdots = m_k = n$ then there are kn items in total, and this is the total number of ways the two-stage procedure can occur.

Examples:

- How many ways are there to first roll a die and then draw a card from a (shuffled) $52-$card pack?
  - There are 6 ways the first step can turn out, and for each of these (the stages are independent in fact) there are 52 ways to draw the card. Thus there are $6 \times 52 = 312$ ways this can turn out.

- How many ways are there to draw two tickets numbered 1 to 100 from a bag, the first with the right hand and the second with the left hand?
  - There are 100 ways to pick the first number; for *each of these* there are 99 ways to pick the second number. Thus $100 \times 99 = 9900$ ways.

Multi-Stage Procedures

A k-stage process is a process for which:

- there are n_1 possibilities at stage 1;
- regardless of the 1st outcome there are n_2 possibilities at stage 2,
- etc.,
- regardless of the previous outcomes, there are n_k choices at stage k.

Then there are $n_1 n_2 \cdots n_k$ total ways the process can turn out.

1.3 – Ordered Samples

Suppose we have a bag of n billiard balls numbered $1, 2, \ldots, n$. We draw an **ordered sample** of size r by picking balls from the bag:

- **with replacement**, or
- **without replacement**.

With how many different collection of r balls can we end up in each of those cases (each is an r-stage procedure)?

Key Notion: all the object (balls) can be differentiated (using numbers, colours, etc.)

Sampling With Replacement (order important)

If we replace each ball into the bag after it is picked, then every draw is the same (there are n ways it can turn out).

According to our earlier result, there are

$$\underbrace{n \times n \times \cdots \times n}_{r \text{ stages}} = n^r$$

ways to select an ordered sample of size r with replacement from a set with n objects $\{1, 2, \ldots, n\}$.

Sampling Without Replacement (order important)

If we **do not** replace each ball into the bag after it is drawn, then the choices for the second draw depend on the result of the first draw, and there are only $(n-1)$ possible outcomes.

Whatever the first two draws were, there are $(n-2)$ ways to draw the third ball, and so on.

Thus there are

$$\underbrace{n(n-1)\cdots(n-r+1)}_{r\text{ stages}} = {}_nP_r \quad \text{(common calculator symbol)}$$

ways to select an ordered sample of size $r \leq n$ without replacement from a set of n objects $\{1, 2, \ldots, n\}$.

Factorial Notation

For a positive integer n, write $n! = n(n-1)(n-2)\cdots 1$. We have

- when $r = n$, ${}_nP_r = n!$, and the ordered selection is called a **permutation**;
- when $r < n$, we can write

$$ {}_nP_r = \frac{n(n-1)\cdots(n-r+1)}{}\frac{(n-r)\cdots 1}{(n-r)\cdots 1} = \frac{n!}{(n-r)!} = n\cdots(n-r+1). $$

By convention, we take $0! = 1$, so that ${}_nP_r = \frac{n!}{(n-r)!}$ for all $r \leq n$.

Examples:

1. How many different ways can 6 balls be drawn *in order* without replacement from a bag of balls numbered 1 to 49?

   Answer: ${}_{49}P_6 = 49 \times 48 \times 47 \times 46 \times 45 \times 44 = 10,068,347,520$. This is the number of ways the actual drawing of the balls can occur for Lotto 6/49 in real-time (balls drawn one by one).

2. How many 6-digits PIN codes can you create from the set of digits $\{0, 1, \ldots, 9\}$?

   - If digits may be repeated: $10 \times 10 \times 10 \times 10 \times 10 \times 10 = 10^6 = 1,000,000$;
   - If digits may not be repeated: ${}_{10}P_6 = 10 \times 9 \times 8 \times 7 \times 6 \times 5 = 151,200$.

1.4 – Unordered Samples

Suppose now that we cannot distinguish between different ordered samples; when we look up the Lotto 6/49 results in the newspaper, for instance, we have no way of knowing the order in which the balls were drawn.

$1-2-3-4-5-6$ could mean that the first drawn ball was ball $\# \ 1$, the second drawn ball was ball $\# \ 2$, etc., but it could also mean that the first drawn ball was ball $\# \ 4$, the second one was ball $\# \ 3$, etc., or any other combinations of the first 6 balls.

Denote the (as yet unknown) number of unordered samples of size r from a set of size n by ${}_nC_r$.

We can derive the expression for ${}_nC_r$ by noting that the following two processes are equivalent:

- Take an ordered sample of size r (there are ${}_nP_r$ ways to do this);
- Take an unordered sample of size r (there are ${}_nC_r$ ways to do this) **and then** rearrange (permute) the objects in the sample (there are $r!$ ways to do this).

Thus

$${}_nP_r = {}_nC_r \times r! \quad \Rightarrow \quad {}_nC_r = \frac{{}_nP_r}{r!} = \frac{n!}{(n-r)!\, r!} = \binom{n}{r}.$$

This last notation is called a **binomial coefficient** (read as "n-choose-r") and is commonly used in textbooks.

Example:

In how many ways can the "Lotto 6/49 draw" be reported in the newspaper (where they are always reported in increasing order)?

Answer: this number is the same as the number of *unordered samples* of size 6 (different reorderings of same 6 numbers are indistinguishable), so

$$\begin{aligned} {}_{49}C_6 = \binom{49}{6} = \frac{49 \times 48 \times 47 \times 46 \times 45 \times 44}{6 \times 5 \times 4 \times 3 \times 2 \times 1} &= \frac{10,068,347,520}{720} \\ &= 13,983,816\,. \end{aligned}$$

Binomial Coefficient Identities $\Rightarrow$ blackboard

1.5 – Probability of an Event

For situations where we have a random experiment which has exactly N possible **mutually exclusive**, **equally likely** outcomes, we can assign a probability to an event A by counting the number of outcomes that correspond to A. If the count is a then

$$P(A) = \frac{a}{N}.$$

The probability of each individual outcome is thus $1/N$.

Examples:

1. Toss a fair coin. The sample space is $\mathcal{S} = \{\text{Head, Tail}\}$, i.e. $N = 2$. The probability of observing a Head is $\frac{1}{2}$.

2. Throw a fair six sided die. There are $N = 6$ possible outcomes. The sample space is
$$\mathcal{S} = \{1, 2, 3, 4, 5, 6\}.$$
If A corresponds to observing a multiple of 3, then $A = \{3, 6\}$ and $a = 2$, so that
$$\text{Prob(number is a multiple of } 3) = P(A) = \frac{2}{6} = \frac{1}{3}.$$

3. The probabilities of seeing an even/odd number are:
   - Prob$\{$even no.$\} = P(\{2,4,6\}) = \frac{3}{6} = \frac{1}{2}$.
   - Prob$\{$prime no.$\} = P(\{2,3,5\}) = 1 - P(\{1,4,6\}) = \frac{1}{2}$.

4. In a group of 1000 people it is known that 545 have high blood pressure. 1 person is selected randomly. What is the probability that this person has high blood pressure?

   Solution: the relative frequency of people with high blood pressure is 0.545. In the classical definition, this is the probability we are seeking.

Frequentist Interpretation of Probabilities $\Rightarrow$ blackboard

Axioms of Probability

1. For any event A, $1 \geq P(A) \geq 0$.

2. For the complete sample space $\mathcal{S}$, $P(\mathcal{S}) = 1$.

3. For the empty event $\varnothing$, $P(\varnothing) = 0$.

4. For two **mutually exclusive** events A and B, the probability that A or B occurs is $P(A \cup B) = P(A) + P(B)$.

Since $\mathcal{S} = A \cup A^c$, and since A and A^c are mutually exclusive, then

$$1 \overset{\textbf{A2}}{=} P(\mathcal{S}) = P(A \cup A^c) \overset{\textbf{A4}}{=} P(A) + P(A^c) \Rightarrow P(A^c) = 1 - P(A).$$

Examples:

1. Throw a single six sided die and record the number that is shown. Let A and B be the events that the number is a multiple of or smaller than 3, respectively. Then $A = \{3, 6\}$, $B = \{1, 2\}$ and A and B are mutually exclusive since $A \cap B = \varnothing$. Then

$$P(A \text{ or } B \text{ occurs}) = P(A \cup B) = P(A) + P(B) = \frac{2}{6} + \frac{2}{6} = \frac{2}{3}.$$

2. An urn contains 4 white balls, 3 red balls and 1 black ball. Draw one ball, and note the events $W = \{\text{the ball is white}\}$, $R = \{\text{the ball is red}\}$ and $B = \{\text{the ball is black}\}$. Then

$$P(W) = 1/2, \quad P(R) = 3/8, \quad P(B) = 1/8, \quad P(W \text{ or } R) = 7/8.$$

General Addition Rule

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Example: An electronic gadget consists of two components A and B. We know from experience that $P(A \text{ fails}) = 0.2$, $P(B \text{ fails}) = 0.3$ and $P(\text{both } A \text{ and } B \text{ fail}) = 0.15$. Find $P(\text{at least one of } A \text{ and } B \text{ fails})$ and $P(\text{neither } A \text{ nor } B \text{ fails})$.

Answer: Write A for "A fails" and similarly for B. Then we want

$$P(\text{at least one fails}) = P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.35\,;$$

$$P(\text{neither fail}) = 1 - P(\text{at least one fails}) = 0.65\,.$$

Market Basket Example $\Rightarrow$ blackboard

When A and B are mutually exclusive, $P(A \cap B) = P(\varnothing) = 0$ and

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

If there are more than two events, the rule expands as follows:

$$\begin{aligned} P(A \cup B \cup C) =& P(A) + P(B) + P(C) \\ & - P(A \cap B) - P(A \cap C) - P(B \cap C) \\ & + P(A \cap B \cap C). \end{aligned}$$

Set Theory Arithmetic $\Rightarrow$ blackboard

Patient: "Will I survive this risky operation?"
Surgeon: "Yes, I'm absolutely sure that you will survive the operation."
Patient: "How can you be so sure?"
Surgeon: "9 out of 10 patients die in this operation, and yesterday my ninth patient died."

One out of every four people is suffering from some form of mental illness. Check three friends. If they're OK, then it's you.

1.6 – Conditional Probability and Independent Events

Any two events A and B satisfying

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

are said to be **independent**; this is a purely mathematical definition, but it agrees with the intuitive notion of independence in simple examples.

When events are not independent, we say that they are **dependent** or **conditional**.

Mutually Exclusive vs Independent Events $\Rightarrow$ blackboard

Examples:

1. Flip a fair coin twice: the 4 possible outcomes are all equally likely: $\mathcal{S} = \{HH, HT, TH, TT\}$. Let $A = \{HH\} \cup \{HT\}$ denote "head on first flip", $B = \{HH\} \cup \{TH\}$ "head on second flip". Note that $A \cup B \neq \mathcal{S}$ and $A \cap B = \{HH\}$.

   By the General Addition Rule,

$$\begin{aligned} P(A) &= P(\{HH\}) + P(\{HT\}) - P(\{HH\} \cap \{HT\}) \\ &= \frac{1}{4} + \frac{1}{4} - P(\varnothing) = \frac{1}{2} - 0 = \frac{1}{2}. \end{aligned}$$

   Similarly, $P(B) = P(\{HH\}) + P(\{TH\}) = \frac{1}{2}$, and so $P(A)P(B) = \frac{1}{4}$. But $P(A \cap B) = P(\{HH\})$ is also $\frac{1}{4}$, so A and B are independent.

2. A card is drawn from a regular well-shuffled North American card deck. Let A be the event that it is an ace and D be the event that it is a diamond.

   These two events are independent: there are 4 aces ($P(A) = \frac{4}{52} = \frac{1}{13}$) and 13 diamonds $P(D) = \frac{13}{52} = \frac{1}{4}$) in such a deck, so that

$$P(A)P(D) = \frac{1}{13} \times \frac{1}{4} = \frac{1}{52},$$

   and exactly 1 ace of diamonds in the deck, so that $P(A \cap D)$ is also $\frac{1}{52}$.

3. A six-sided die numbered $1--6$ is loaded in such a way that the prob of getting each value is *proportional* to that value. Find $P(3)$.

   Solution: Let $\mathcal{S} = \{1,2,3,4,5,6\}$ be the result of a single toss; for some proportional constant v, we have $P(k) = kv$, for $k \in \mathcal{S}$. By Axiom **A2**, $P(\mathcal{S}) = P(1) + \cdots + P(6) = 1$, so that

$$1 = \sum_{k=1}^{6} P(k) = \sum_{k=1}^{6} kv = v \sum_{k=1}^{6} k = v\frac{(6+1)(6)}{2} = 21v\,.$$

   Hence $v = 1/21$ and $P(3) = 3v = 3/21 = 1/7$.

   Sigma Notation $\Rightarrow$ blackboard

4. Now the die is rolled twice, the second toss *independent* of the first. Find $P(3_1, 3_2)$.

   Solution: the experiment is such that $P(3_1) = 1/7$ and $P(3_2) = 1/7$, as seen in the previous example. Since the die tosses are independent, then

$$P\left(3_1 \cap 3_2\right) = P(3_1)P(3_2) = 1/49\,.$$

   Independent Tosses $\Rightarrow$ blackboard

5. Which plane is more likely to crash: a 2-engine one or a 3-engine one?

   Solution: this question is easier to answer if we assume that **engines fail independently** (convenient: yes; realistic: ???). Let p be the probability that an engine fails.

   How can a plane crash? (another set of assumptions)

   - A 2-engine plane will crash if both engines fail – the probability is p^2.
   - A 3-engine plane will crash if any pair of engines fail, or if all 3 fail.
     - **Pair:** the probability that exactly 1 pair of engines will fail independently (i.e. two engines fail and one does not) is

$$p \times p \times (1 - p).$$

The order in which the engines fail does not matter: there are ${}_3C_2 = \frac{3!}{2!1!} = 3$ ways in which a pair of engines can fail: for 3 engines A, B, C, these are AB, AC, BC.

– **All 3:** the probability of all three engines failing independently is p^3.

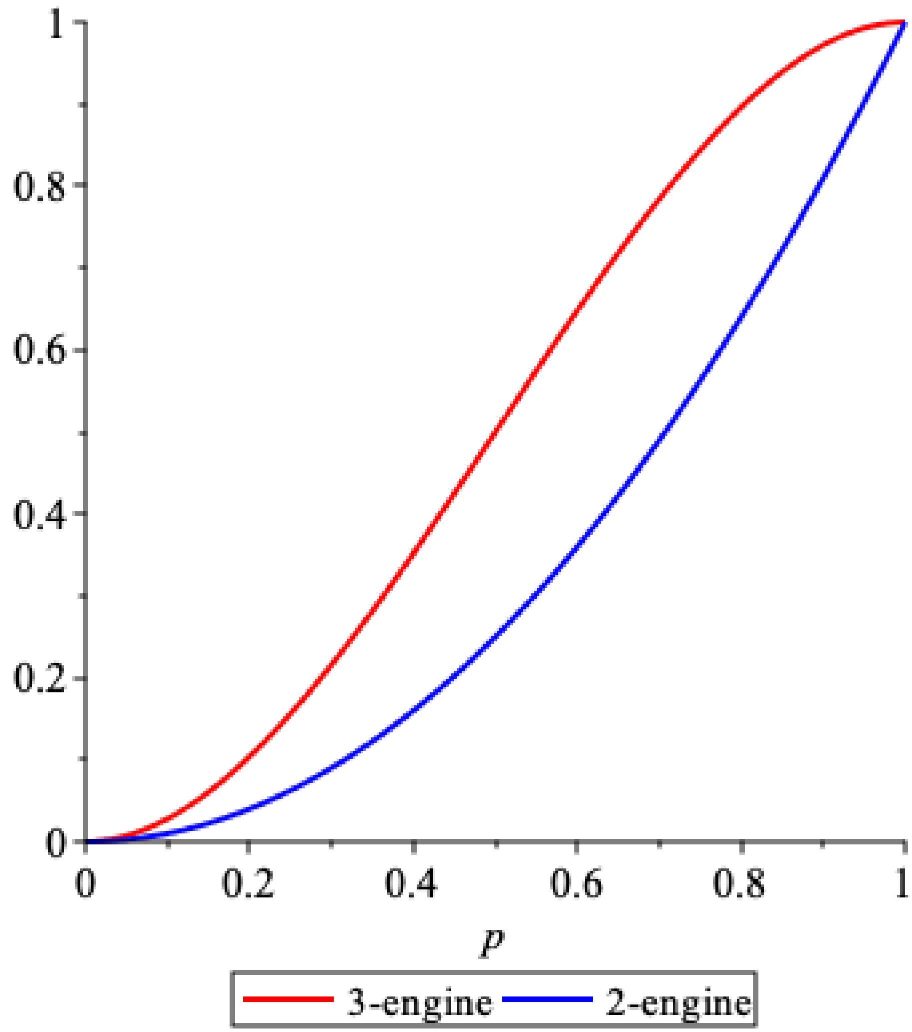
The probability of at least 2 engines failing is:

$$P(\text{at least } 2 \text{ engines fail}) = 3p^2(1-p) + p^3 = 3p^2 - 2p^3.$$

Basically it's safer to use a 2-engine plane than a 3-engine plane: the 3-engine plane will crash more often, assuming it needs 2 engines to fly.

This "makes sense": the 2-engine plane needs 50% of its engines working, while the 3-engine plane needs 66%.

What do you think a realistic value of p could be?


1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
p
3-engine
2-engine

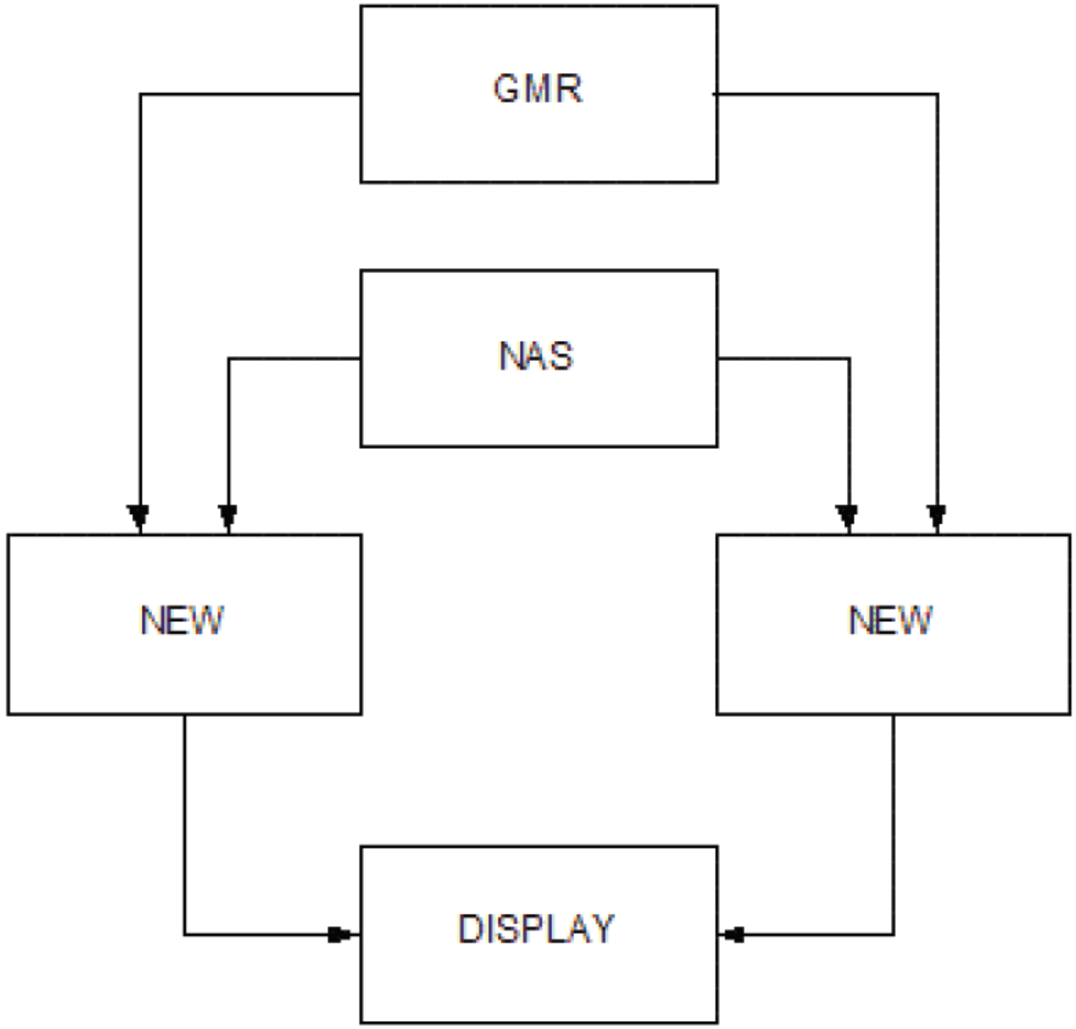
Scenario: Air Traffic Control

Air traffic control is a safety-related activity.

Each piece of equipment is designed to the highest safety standards and in many cases duplicate equipment is provided so that if one item fails another takes over.

A new system is to be provided passing information from Heathrow Airport to Terminal Control at West Drayton. As part of the system design a decision has to be made as to whether it is necessary to provide duplication.

The new system takes data from the *Ground Movements Radar* (GMR) at Heathrow, combines this with data from the *National Airspace System NAS*, and sends the output to a display at *Terminal Control*.


GMR
NAS
NEW
NEW
DISPLAY

For all existing systems, records of failure are kept and an experimental probability of failure is calculated annually using the previous 4 years.

The **reliability** of a system is defined as $R = 1 - P$, where $P = P(\text{failure})$.

Given: $R_{\text{GMR}} = R_{\text{NAS}} = 0.9999$ (i.e. 1 failure in $10,000$ hours).

Assumption: the components' failure probabilities are independent.

For the system above, if a single NEW module is introduced the reliability of the system (STD – single thread design) is

$$R_{\text{STD}} = R_{\text{GMR}} \times R_{\text{NEW}} \times R_{\text{NAS}}.$$

If the NEW module is duplicated, the reliability of the dual thread design is

$$R_{\text{DTD}} = R_{\text{GMR}} \times (1 - (1 - R_{\text{NEW}})^2) \times R_{\text{NAS}}.$$

Duplicating the NEW module causes an improvement in reliability of

$$\rho = \frac{R_{\text{DTD}}}{R_{\text{STD}}} = \frac{(1-(1-R_{\text{NEW}})^2)}{R_{\text{NEW}}} \times 100\%\,.$$

For the NEW module, no historical data is available. Instead, we work out the improvement achieved by using the dual thread design for various values of R_{NEW}.

R_{NEW}	0.1	0.2	0.5	0.75	0.99	0.999	0.9999	0.99999
ρ (%)	190	180	150	125	101	100.1	100.01	100.001

If the NEW module is very unreliable (i.e. R_{NEW} is small), then there is a significant benefit in using the dual thread design (ρ is large).

But why would we install a module which we know to be unreliable?

If the new module is as reliable as NAS and GMR, that is, if

$$R_{\text{GMR}} = R_{\text{NEW}} = R_{\text{NAS}} = 0.9999,$$

then the single thread design has a combined reliability of 0.9997 (i.e. 3 failures in $10,000$ hours), whereas the dual thread design has a combined reliability of 0.9998 (i.e. 2 failures in $10,000$ hours).

If the probability of failure is independent for each component, we could conclude from this that the reliability gain from a dual thread design probably does not justify the extra cost.

Conditional Probability

We can better understand when independence applies by defining the **conditional probability of an event B given that another event A has occurred** as

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

Note that this only makes sense when "A can happen" i.e. $P(A) > 0$.

If $P(A)P(B) > 0$, then

$$P(A \cap B) = P(A) \times P(B|A) = P(B) \times P(A|B) = P(B \cap A);$$

A and B are thus independent if $P(B|A) = P(B)$ and $P(A|B) = P(A)$.

Examples:

1. From a group of 100 people, 1 is selected. What is the probability that this person has high blood pressure (HBP)?

   Solution: if we know nothing else about the population, this is an **(unconditional) probability**, namely

$$P(\text{HBP}) = \frac{\#\text{individuals with HBP in the population}}{100}.$$

2. If instead we first filter out all people with low cholesterol level, and then select 1 person. What is the probability that this person has HBP? This is a **conditional probability** $P(\text{HBP}|\text{high cholesterol})$; the probability of selecting a person with HBP, given high cholesterol levels, presumably different from $P(\text{HBP}|\text{low cholesterol})$.

3. A sample of 249 individuals is taken and each person is classified by blood type and tuberculosis (TB) status.

	O	**A**	**B**	**AB**	Total
TB	34	37	31	11	113
no TB	55	50	24	7	136
Total	89	87	55	18	249

The (unconditional) probability that a random individual has TB is $P(\text{TB}) = \frac{\#\text{TB}}{249} = \frac{113}{249} = 0.454$. Among those individuals with type **B** blood, the (conditional) probability of having TB is

$$P(\text{TB}|\text{type }\mathbf{B}) = \frac{P(\text{TB} \cap \text{type }\mathbf{B})}{P(\text{type }\mathbf{B})} = \frac{31}{55} = \frac{31/249}{55/249} = 0.564.$$

4. A family has two children (not twins). What is the probability that the youngest child is a girl given that at least one of the children is a girl? Assume that boys and girls are equally likely to be born.

   Solution: Let A and B be the events that the youngest child is a girl and that at least one child is a girl, respectively:

   $$A = \{\text{GG}, \text{BG}\} \quad \text{and} \quad B = \{\text{GG}, \text{BG}, \text{GB}\}, \quad \text{so that} \quad A \cap B = A.$$

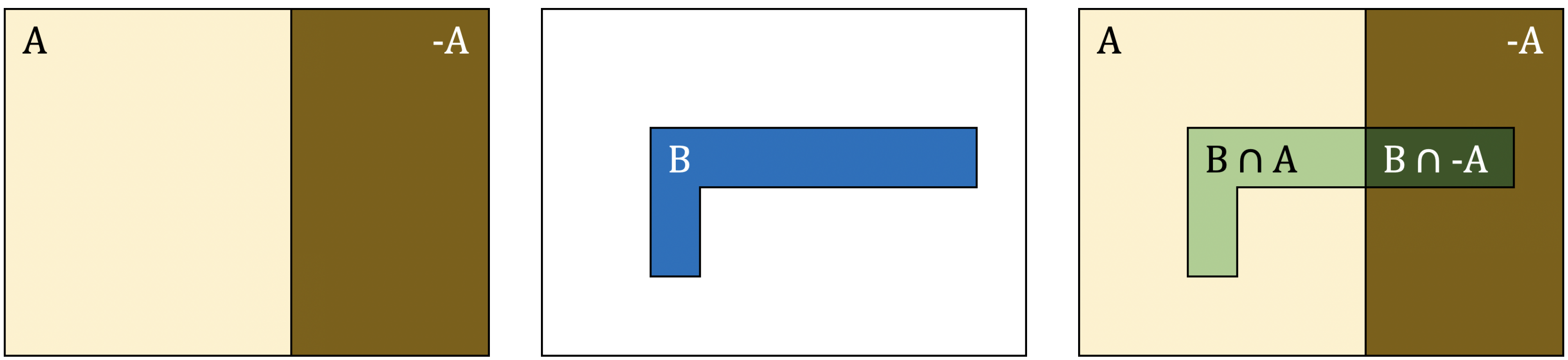
   Then $P(A|B) = \frac{P(A\cap B)}{P(B)} = \frac{P(A)}{P(B)} = \frac{2/4}{3/4} = \frac{2}{3}$ (and not $\frac{1}{2}$).

   Incidentally, $P(A \cap B) = P(A) \neq P(A) \times P(B)$ which means that A and B are dependent events.

Law of Total Probability

Let A and B be two events. From set theory, we have

$$B = (A \cap B) \cup (\overline{A} \cap B).$$



Note that $A \cap B$ and $\overline{A} \cap B$ are mutually exclusive.

According to Axiom **A4**,

$$P(B) = P(A \cap B) + P(\overline{A} \cap B).$$

Now, assuming that $\varnothing \neq A \neq \mathcal{S}$,

$$P(A \cap B) = P(B|A)P(A) \quad \text{and} \quad P(\overline{A} \cap B) = P(B|\overline{A})P(\overline{A}),$$

so that $P(B) = P(B|A)P(A) + P(B|\overline{A})P(\overline{A})$.

This generalizes as follows: if $A_1, ...A_k$ are **mutually exclusive** and **exhaustive** (i.e. $A_i \cap A_j = \varnothing$ for all $i \neq j$ and $A_1 \cup .... \cup A_k = \mathcal{S}$), then for any event B

$$P(B) = \sum_{j=1}^{k} P(B|A_j)P(A_j) = P(B|A_1)P(A_1) + ... + P(B|A_k)P(A_k).$$

Example: Use the Law of Total Probability to compute $P(\text{TB})$ using the data from the previous example.

Solution: the blood types $\{\mathbf{O},\mathbf{A},\mathbf{B},\mathbf{AB}\}$ form a mutually exclusive partition of the population, with

$$P(\mathbf{O})=\frac{89}{249},\ P(\mathbf{A})=\frac{87}{249},\ P(\mathbf{B})=\frac{55}{249}\text{ and }P(\mathbf{AB})=\frac{18}{249}.$$

It is easy to see that $P(\mathbf{O})+P(\mathbf{A})+P(\mathbf{B})+P(\mathbf{AB})=1.$ Furthermore,

$$P(\text{TB}|\mathbf{O})=\frac{P(\text{TB}\cap\mathbf{O})}{P(\mathbf{O})}=\frac{34}{89},\ P(\text{TB}|\mathbf{A})=\frac{P(\text{TB}\cap\mathbf{A})}{P(\mathbf{A})}=\frac{37}{87},$$

$$P(\text{TB}|\mathbf{B})=\frac{P(\text{TB}\cap\mathbf{B})}{P(\mathbf{B})}=\frac{31}{55},\ P(\text{TB}|\mathbf{AB})=\frac{P(\text{TB}\cap\mathbf{AB})}{P(\mathbf{AB})}=\frac{11}{18}.$$

According to the Law of Total Probability,

$$\begin{aligned} P(\mathsf{TB}) &= P(\mathsf{TB}|\mathbf{O})P(\mathbf{O}) + P(\mathsf{TB}|\mathbf{A})P(\mathbf{A}) \\ &\quad + P(\mathsf{TB}|\mathbf{B})P(\mathbf{B}) + P(\mathsf{TB}|\mathbf{AB})P(\mathbf{AB}), \end{aligned}$$

so that

$$\begin{aligned} P(\mathsf{TB}) &= \frac{34}{89}\cdot\frac{89}{249} + \frac{37}{87}\cdot\frac{87}{249} + \frac{31}{55}\cdot\frac{55}{249} + \frac{11}{18}\cdot\frac{18}{249} \\ &= \frac{34+37+31+11}{249} = \frac{113}{249} = 0.454, \end{aligned}$$

which matches with the result of the previous example.

1.7 – Bayes' Theorem

After an experiment generates an outcome, we are often interested in the probability that a certain condition was present given an outcome (or that a particular hypothesis was valid, say).

We have noted before that if $P(A)P(B) > 0$, then

$$P(A \cap B) = P(A) \times P(B|A) = P(B) \times P(A|B) = P(B \cap A);$$

this can be re-written as **Bayes' Theorem**:

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

BT is a simple corollary of the rules of probability.

Central Data Analysis Question

Given everything that was known prior to the experiment, does the collected/observed data support (or invalidate) the hypothesis/presence of a certain condition?

Problem: this is nearly always impossible to compute directly.

Solution: using Bayes' Theorem, we can re-write the CDAQ as

$$\begin{aligned} P(\text{hypothesis}|\text{data}) &= \frac{P(\text{data}|\text{hypothesis}) \times P(\text{hypothesis})}{P(\text{data})}, \\ &\propto P(\text{data}|\text{hypothesis}) \times P(\text{hypothesis}) \end{aligned}$$

in which the terms on the right might be easier to compute.

Bayesian Vernacular

The following terms are used in Bayesian analysis:

- $P(\text{hypothesis})$ is the probability of the hypothesis being true prior to the experiment (called the **prior**);
- $P(\text{hypothesis}|\text{data})$ is the probability of the hypothesis being true once the experimental data is taken into account (called the **posterior**);
- $P(\text{data}|\text{hypothesis})$ is the probability of the experimental data being observed assuming that the hypothesis is true (called the **likelihood**).

BT is often presented as posterior $\propto$ likelihood $\times$ prior, which is to say, **beliefs should be updated in the presence of new information**.

Formulations

If A and B are events for which $P(A)P(B) > 0$, then BT can be re-written (using the LTP) as

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\overline{A})P(\overline{A})}.$$

This generalizes as follows: if $A_1, ...A_k$ are **mutually exclusive** and **exhaustive** events, then for any event B and for each i,

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)} = \frac{P(B|A_i)P(A_i)}{P(B|A_1)P(A_1) + ... + P(B|A_k)P(A_k)}.$$

Examples:

1. In 1999, Nissan sold three car models in North America: the Sentra (S), the Maxima (M), and the Pathfinder (Pa). Of the vehicles sold, 50% were S, 30% were M and 20% were Pa. In the same year 12% of the S, 15% of the M, and 25% of the Pa had a particular defect D.

   (a) If you own a 1999 Nissan, what is the probability that it has the defect?

   Solution: In the language of conditional probability,

$$\begin{aligned}
P(\mathsf{S}) &= 0.5, \quad P(\mathsf{M}) = 0.3, \quad P(\mathsf{Pa}) = 0.2, \\
P(D|\mathsf{S}) &= 0.12, \quad P(D|\mathsf{M}) = 0.15, \quad P(D|\mathsf{Pa}) = 0.25, \text{ so that} \\
P(D) &= P(D|\mathsf{S})P(\mathsf{S}) + P(D|\mathsf{M})P(\mathsf{M}) + P(D|\mathsf{Pa})P(\mathsf{Pa}) \\
&= 0.12 \times 0.5 + 0.15 \times 0.3 + 0.25 \times 0.2 = 0.155 = 15.5\%
\end{aligned}$$

(b) My 1999 Nissan has defect D. What model am I likely to own?

Solution: in the first part we computed the total probability $P(D)$; in this part, we compare the posterior probabilities $P(\text{M}|D)$, $P(\text{S}|D)$, and $P(\text{Pa}|D)$ (and not the priors!), computed using Bayes' Theorem:

$$P(\text{S}|D) = \frac{P(D|\text{S})P(\text{S})}{P(D)} = \frac{0.12 \times 0.5}{0.155} \approx 38.7\%$$

$$P(\text{M}|D) = \frac{P(D|\text{M})P(\text{M})}{P(D)} = \frac{0.15 \times 0.3}{0.155} \approx 29.0\%$$

$$P(\text{Pa}|D) = \frac{P(D|\text{Pa})P(\text{Pa})}{P(D)} = \frac{0.25 \times 0.2}{0.155} \approx 32.3\%$$

Even though Sentras are the least likely to have the defect D, their overall prevalence in the population carry them over the hump.

2. Suppose that a test for a particular disease has a very high success rate. If a patient

    - has the disease, the test reports a 'positive' with probability 0.99;
    - does not have the disease, the test reports a 'negative' with prob 0.95.

    Assume that only 0.1% of the population has the disease. What is the probability that a patient who tests positive does not have the disease?

    Solution: Let D be the event that the patient has the disease, and A be the event that the test is positive. The probability of a true positive is

$$\begin{aligned} P(D|A) &= \frac{P(A|D)P(D)}{P(A|D)P(D) + P(A|D^c)P(D^c)} \\ &= \frac{0.99 \times 0.001}{0.99 \times 0.001 + 0.05 \times 0.999} \approx 0.019. \end{aligned}$$

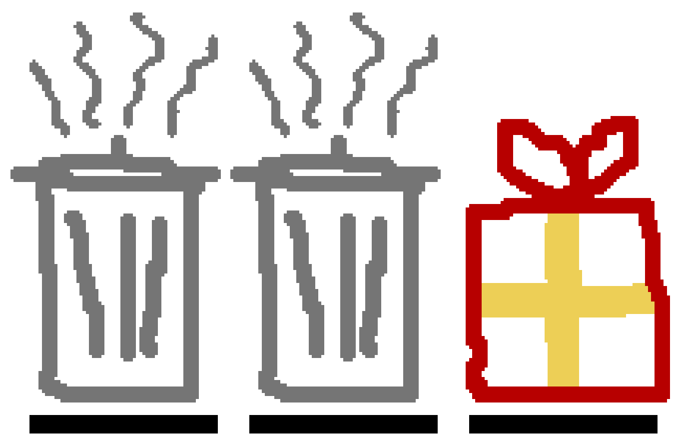
The probability of a false positive is thus $1 - 0.019 \approx 0.981$.

Despite the apparent high accuracy of the test, the incidence of the disease is so low (1 in a 1000) that the vast majority of patients who test positive (98 in 100) do not have the disease.

The 2 in 100 who are true positives still represent 20 times the proportion of positives found in the population (before the outcome of the test is known).

IMPORTANT TAKE-AWAY: when dealing with probabilities, be sure to take **both** the likelihood and the prevalence into account!

3. (**Monty Hall Problem**) Suppose you are on a game show, and you are given the choice of three doors. Behind one door is a prize; behind the others, dirty and smelly rubbish bins. You pick a door, say No. 1, and the host, who knows what is behind the doors, opens another door, say No. 3, behind which is a bin. She then says to you, "Do you want to switch from door No. 1 to No. 2?" Is it to your advantage to do so?



Solution: in what follows, let S and D be the events that switching to another door is a successful strategy and that the prize is behind the original door, respectively.

- Let's first assume that the host opens no door. What is the probability that switching to another door in this scenario would prove to be a successful strategy?

  If the prize is behind the original door, switching would succeed 0% of the time: $P(\mathsf{S}|\mathsf{D}) = 0$. Note that the prior is $P(\mathsf{D}) = 1/3$.

  If the prize is not behind the original door, switching would succeed 50% of the time: $P(\mathsf{S}|\mathsf{D}^c) = 1/2$. Note that the prior is $P(\mathsf{D}^c) = 2/3$.

  Thus, $P(\mathsf{S}) = P(\mathsf{S}|\mathsf{D})P(\mathsf{D}) + P(\mathsf{S}|\mathsf{D}^c)P(\mathsf{D}^c) = 0 \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{2}{3} \approx 33\%.$

- Now let's assume that the host opens one of the other two doors to show a rubbish bin. What is the probability that switching to another door in this scenario would prove to be a successful strategy?

  If the prize is behind the original door, switching would succeed 0% of the time: $P(\mathsf{S}|\mathsf{D}) = 0$. Note that the prior is $P(\mathsf{D}) = 1/3$.

  If the prize is not behind the original door, switching would succeed 100% of the time: $P(\mathsf{S}|\mathsf{D}^c) = 1$. Note that the prior is $P(\mathsf{D}^c) = 2/3$.

  Thus, $P(\mathsf{S}) = P(\mathsf{S}|\mathsf{D})P(\mathsf{D}) + P(\mathsf{S}|\mathsf{D}^c)P(\mathsf{D}^c) = 0 \cdot \frac{1}{3} + 1 \cdot \frac{2}{3} \approx 67\%.$

If no door is opened, switching is not a winning strategy, resulting in success only 33% of the time. If a door is opened, however, switching becomes the winning strategy, resulting in success 67% of the time.

Appendix – Probabilistic Fallacies

Be on the lookout for these easy mistakes to make:

- **gambler's fallacy**: frequentist vs Bayesian interpretations
- **base rate/false positive fallacy**: low incidence/high test accuracy
- **prosecutor's fallacy**: small probabilities and large populations
- **confusion of the inverse**: $P(A|B) \neq P(B|A)$, in general
- **Simpson's paradox**: patterns in sub-groups may not generalize

Let's go to the blackboard!