Statistics for Engineers
Lecture 2

antony.lewis@sussex.ac.uk

http://cosmologist.info/teaching/STAT/

Summary from last time

Complements Rule: $P(A^c) = 1 - P(A)$

Multiplication Rule: $P(A \cap B) = P(A)P(B|A) = P(B)P(A|B)$

Special case: if *independent* then $P(A \cap B) = P(A)P(B)$

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

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

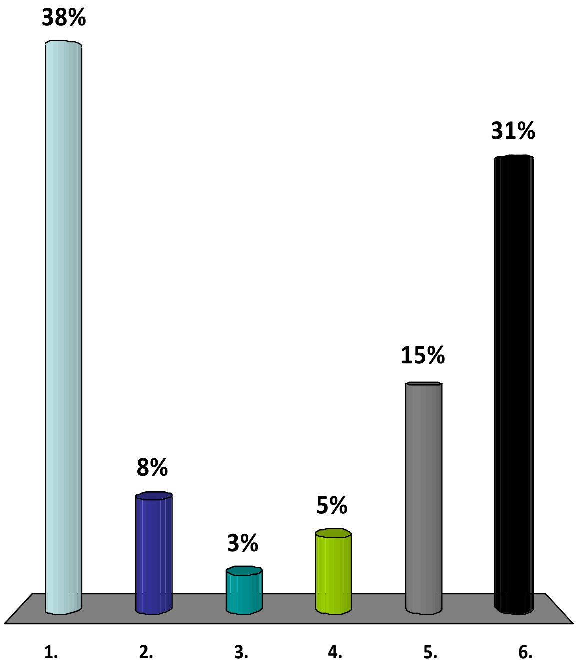
Special case: if mutually exclusive $P(A \cup B) = P(A) + P(B)$

Failing a drugs test

A drugs test for athletes is 99% reliable: applied to a drug taker it gives a positive result 99% of the time, given to a non-taker it gives a negative result 99% of the time. It is estimated that 1% of athletes take drugs.

A random athlete has failed the test. What is the probability the athlete takes drugs?

1. 0.01
2. 0.3
3. 0.5
4. 0.7
5. 0.98
6. 0.99



Similar example: TV screens produced by a manufacturer have defects 10% of the time.

An automated mid-production test is found to be 80% reliable at detecting faults (if the TV has a fault, the test indicates this 80% of the time, if the TV is fault-free there is a false positive only 20% of the time).

If a TV fails the test, what is the probability that it has a defect?

Split question into two parts

1. What is the probability that a random TV fails the test?

2. Given that a random TV has failed the test, what is the probability it is because it has a defect?

Example: TV screens produced by a manufacturer have defects 10% of the time.

An automated mid-production test is found to be 80% reliable at detecting faults (if the TV has a fault, the test indicates this 80% of the time, if the TV is fault-free there is a false positive only 20% of the time).

What is the probability of a random TV failing the mid-production test?

Answer:

Let D="TV has a defect"
Let F="TV fails test"

The question tells us: $P(D) = 0.1$ $P(F|D) = 0.8$ $P(F|D^c) = 0.2$

Two independent ways to fail the test:

TV has a defect and test shows this, -*OR*- TV is OK but get a false positive

$$P(F) = P(F \cap D) + P(F \cap D^c) = P(F|D)P(D) + P(F|D^c)P(D^c)$$

$$= 0.8 \times 0.1 + 0.2 \times (1 - 0.1) = 0.26$$

$$P(F) = P(F \cap D) + P(F \cap D^c) = P(F|D)P(D) + P(F|D^c)P(D^c)$$

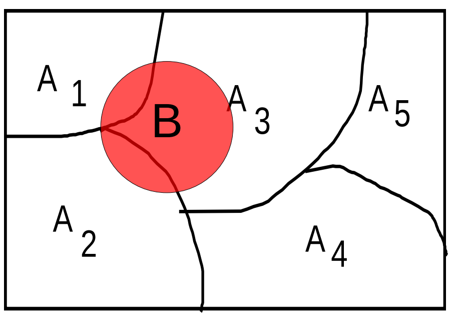
Is an example of the

Total Probability Rule

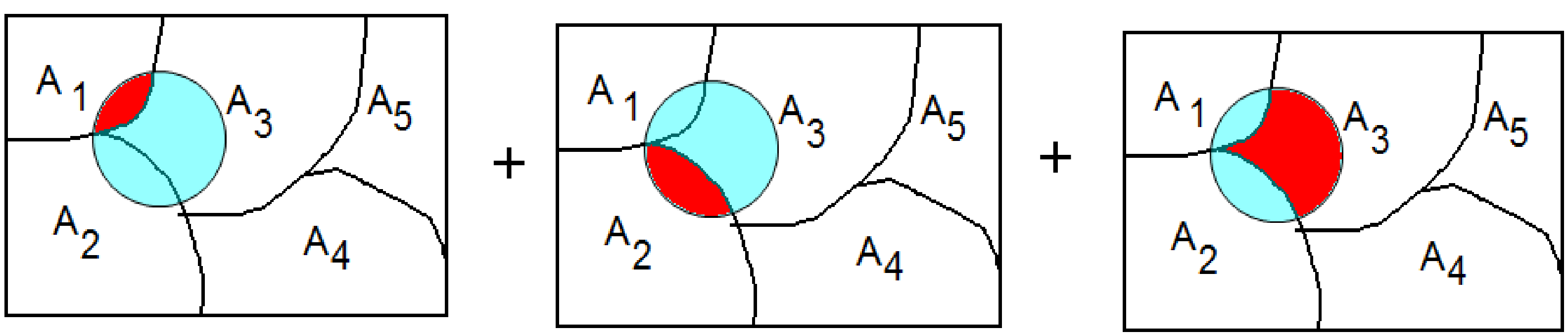
If $A_1, A_2 \ldots, A_k$ form a partition (a mutually exclusive list of all possible outcomes) and *B* is any event then

$$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \cdots + P(B|A_k)P(A_k)$$

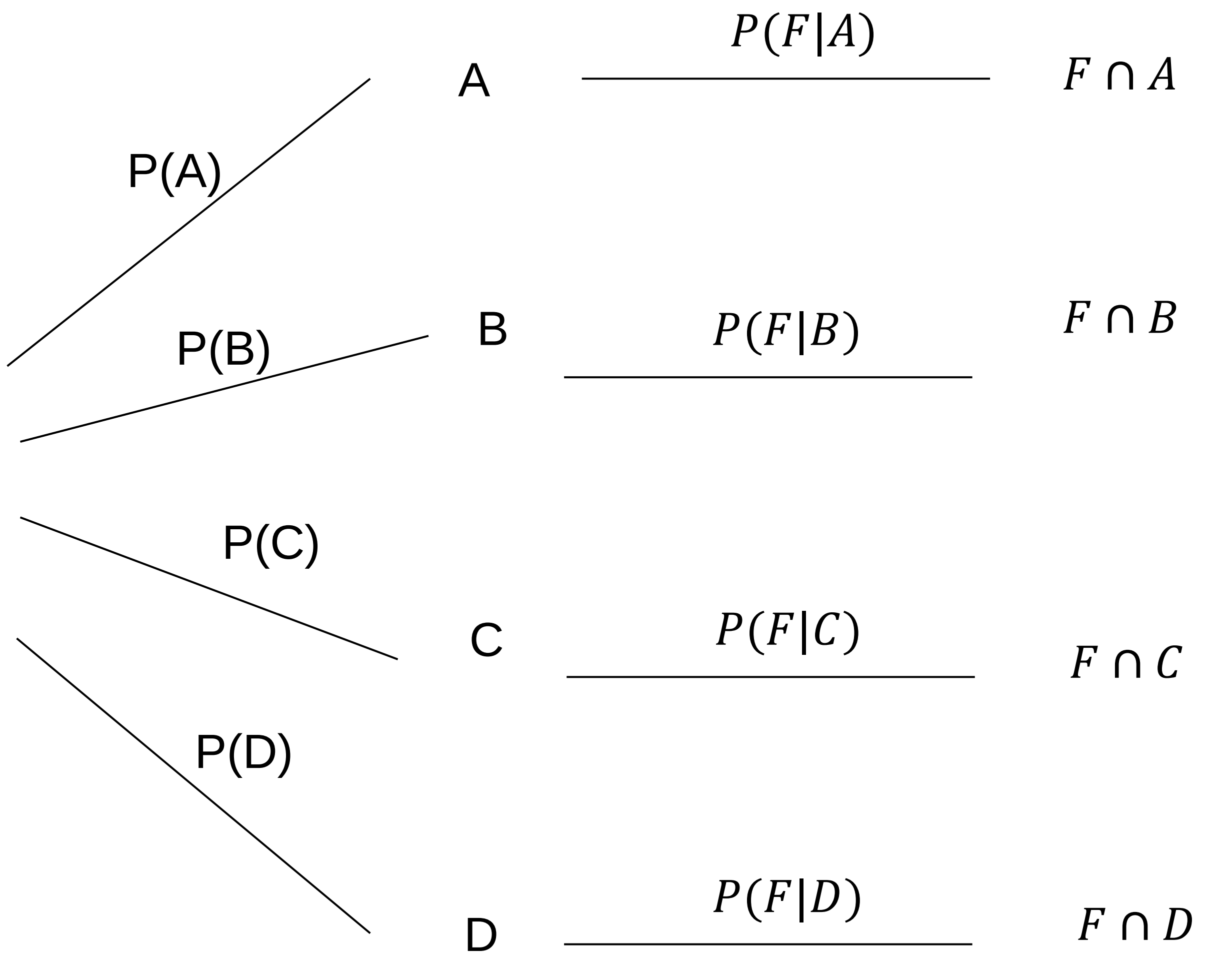
$$= \sum_k P(B|A_k)P(A_k)$$



=



$P(A_1 \cap B) = P(B|A_1)P(A_1)$ $P(A_2 \cap B) = P(B|A_2)P(A_2)$ $P(A_3 \cap B) = P(B|A_3)P(A_3)$



$$P(F) = P(A)P(F|A) + P(B)P(F|B) + P(C)P(F|C) + P(D)P(F|D)$$

Example: TV screens produced by a manufacturer have defects 10% of the time.

An automated mid-production test is found to be 80% reliable at detecting faults (if the TV has a fault, the test indicates this 80% of the time, if the TV is fault-free there is a false positive only 20% of the time).

If a TV fails the test, what is the probability that it has a defect?

Answer:

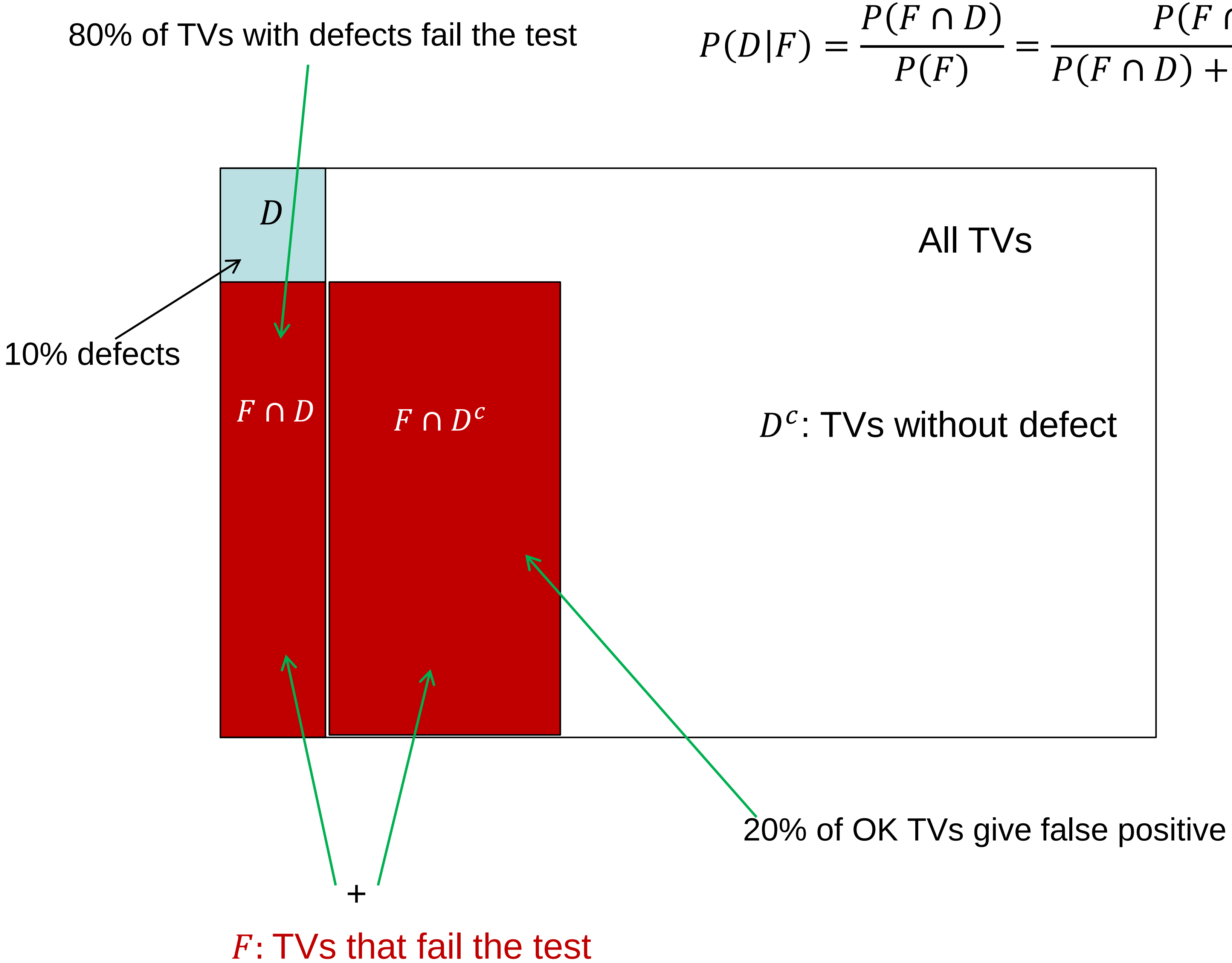
Let D="TV has a defect"
Let F="TV fails test"

We previously showed using the total probability rule that

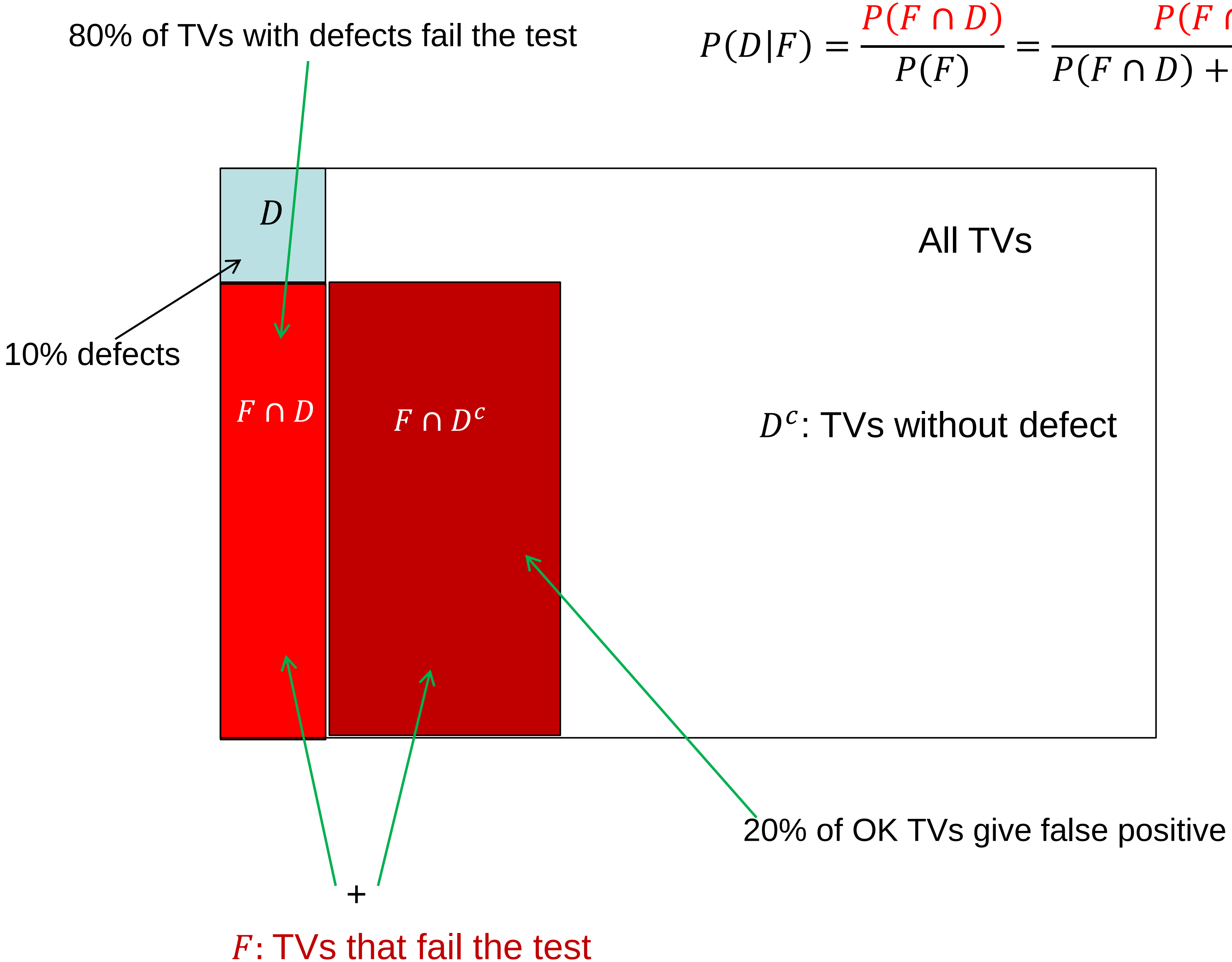
$$P(F) = P(F|D)P(D) + P(F|D^c)P(D^c) = 0.8 \times 0.1 + 0.2 \times (1 - 0.1) = 0.26$$

When we get a test fail, what fraction of the time is it because the TV has a defect?

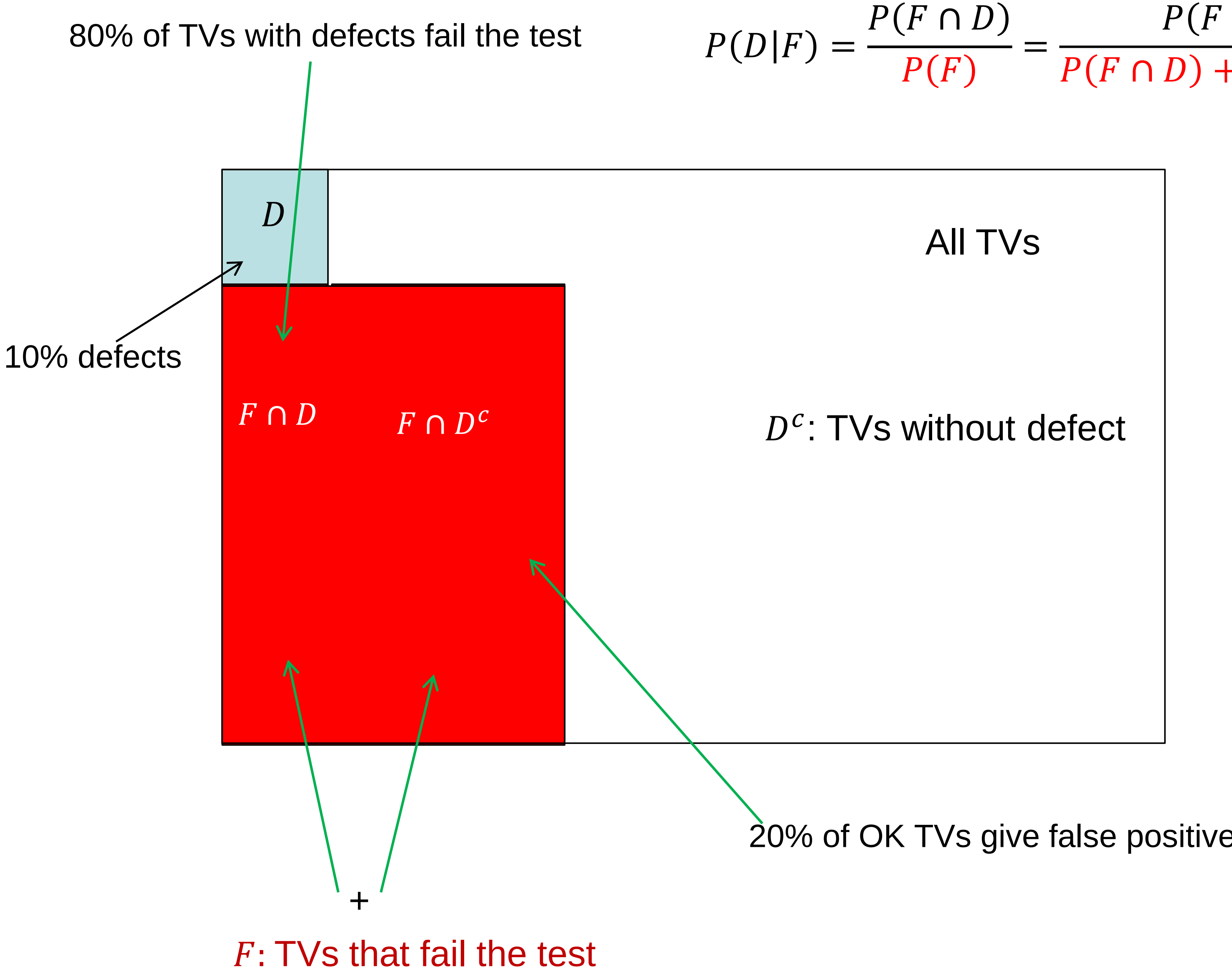
$$P(D|F) = \frac{P(F \cap D)}{P(F)} = \frac{P(F \cap D)}{P(F \cap D) + P(F \cap D^c)}$$



$$P(D|F) = \frac{P(F \cap D)}{P(F)} = \frac{P(F \cap D)}{P(F \cap D) + P(F \cap D^c)}$$



$$P(D|F) = \frac{P(F \cap D)}{P(F)} = \frac{P(F \cap D)}{P(F \cap D) + P(F \cap D^c)}$$



Example: TV screens produced by a manufacturer have defects 10% of the time.

An automated mid-production test is found to be 80% reliable at detecting faults (if the TV has a fault, the test indicates this 80% of the time, if the TV is fault-free there is a false positive only 20% of the time).

If a TV fails the test, what is the probability that it has a defect?

Answer:

Let D="TV has a defect"
Let F="TV fails test"

We previously showed using the total probability rule that

$$P(F) = P(F|D)P(D) + P(F|D^c)P(D^c) = 0.8 \times 0.1 + 0.2 \times (1 - 0.1) = 0.26$$

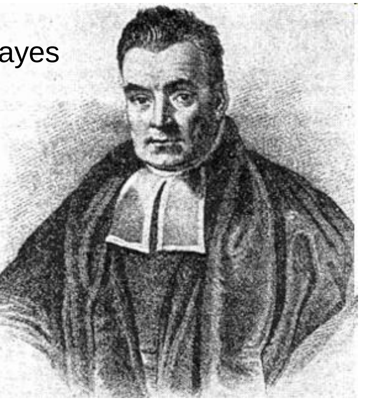
When we get a test fail, what fraction of the time is it because the TV has a defect?

$$P(D|F) = \frac{P(D \cap F)}{P(F)}$$

Know $P(F|D) = 0.8, P(D) = 0.1$:

$$P(D|F) = \frac{P(F|D)P(D)}{P(F)} = \frac{0.8 \times 0.1}{0.26} \approx 0.3077$$

The Rev Thomas Bayes
(1702-1761)



Bayes' Theorem

The multiplication rule gives

Bayes' Theorem

$$P(A \text{ and } B) = P(A|B)P(B) = P(B|A)P(A)$$

Note: as in the example, the Total Probability rule is often used to evaluate P(B):

$$P(A|B) = \frac{P(B|A)P(A)}{\sum_k P(B|A_k)P(A_k)}$$
$$= \frac{P(A \text{ and } B)}{P(A \text{ and } B) + P(A_2 \text{ and } B) + P(A_3 \text{ and } B) + \cdots}$$

If you have a model that tells you how likely B is given A, Bayes' theorem allows you to calculate the probability of A if you observe B. This is the key to learning about your model from statistical data.

Example: **Evidence in court**

The cars in a city are 90% black and 10% grey.

A witness to a bank robbery briefly sees the escape car, and says it is grey. Testing the witness under similar conditions shows the witness correctly identifies the colour 80% of the time (in either direction).

What is the probability that the escape car was actually grey?

Answer: Let G = car is grey, B=car is black, W = Witness says car is grey.

Know $P(W|G)$ want $P(G|W)$

Bayes' Theorem $$P(G|W) = \frac{P(W \cap G)}{P(W)} = \frac{P(W|G)P(G)}{P(W)}.$$

Use total probability rule to write

$$P(W) = P(W|G)P(G) + P(W|B)P(B) \quad = \ 0.8 \times 0.1 + 0.2 \times 0.9 \ = 0.26$$

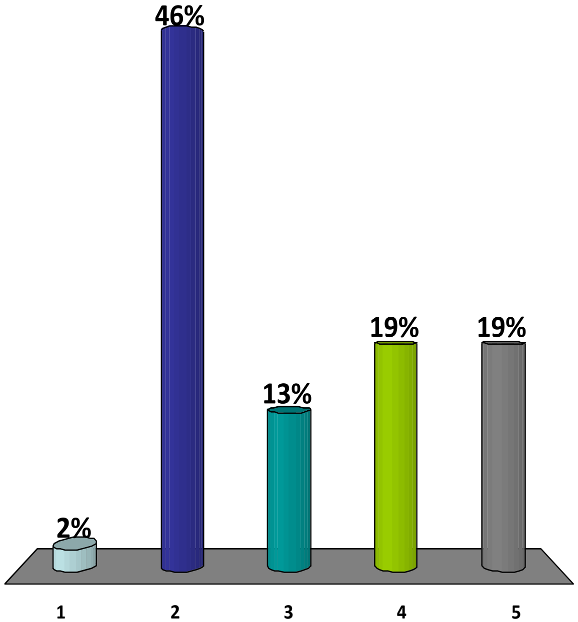
Hence: $$P(G|W) = \frac{P(W|G)P(G)}{P(W)} = \frac{0.8 \times 0.1}{0.26} \approx 0.31$$

Failing a drugs test

A drugs test for athletes is 99% reliable: applied to a drug taker it gives a positive result 99% of the time, given to a non-taker it gives a negative result 99% of the time. It is estimated that 1% of athletes take drugs.

Part 1. What fraction of randomly tested athletes fail the test?

1. 1%
2. 1.98% ✓
3. 0.99%
4. 2%
5. 0.01%



Failing a drugs test

A drugs test for athletes is 99% reliable: applied to a drug taker it gives a positive result 99% of the time, given to a non-taker it gives a negative result 99% of the time. It is estimated that 1% of athletes take drugs.

What fraction of randomly tested athletes fail the test?

Let F="fails test"
Let D="takes drugs"

Question tells us
$P(D) = 0.01, P(F|D) = 0.99$, $P(F|D^c) = 0.01$

From total probability rule:

$$P(F) = P(F|D)P(D) + P(F|D^c)P(D^c) = 0.99 \times 0.01 + 0.01 \times 0.99$$

$$=0.0198$$

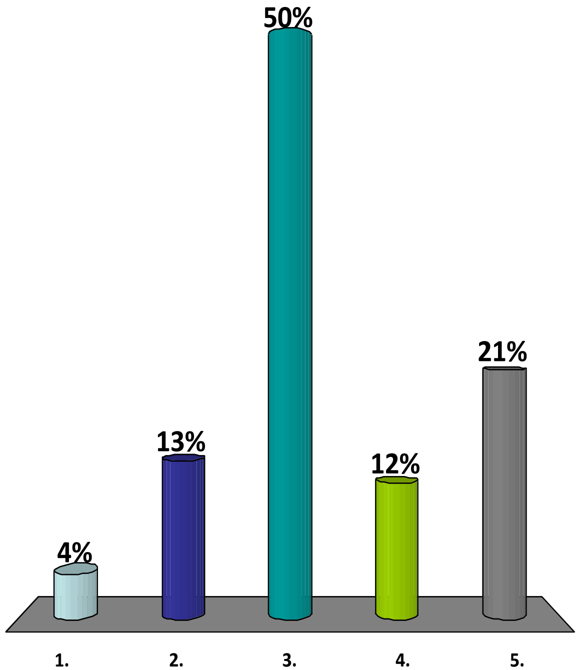
i.e. 1.98% of randomly tested athletes fail

Failing a drugs test

A drugs test for athletes is 99% reliable: applied to a drug taker it gives a positive result 99% of the time, given to a non-taker it gives a negative result 99% of the time. It is estimated that 1% of athletes take drugs.

A random athlete has failed the test. What is the probability the athlete takes drugs?

1. 0.01
2. 0.3
3. 0.5
4. 0.7
5. 0.99



Failing a drugs test

A drugs test for athletes is 99% reliable: applied to a drug taker it gives a positive result 99% of the time, given to a non-taker it gives a negative result 99% of the time. It is estimated that 1% of athletes take drugs.

A random athlete is tested and gives a positive result. What is the probability the athlete takes drugs?

Let F="fails test"
Let D="takes drugs"

Question tells us
$P(D) = 0.01, P(F|D) = 0.99,\ P(F|D^c) = 0.01$

Bayes' Theorem gives $P(D|F) = \dfrac{P(F|D)P(D)}{P(F)}$

We need $P(F) = P(F|D)P(D) + P(F|D^c)P(D^c) = 0.99 \times 0.01 + 0.01 \times 0.99$
= 0.0198

Hence: $P(D|F) = \dfrac{P(F|D)P(D)}{P(F)} \quad = \dfrac{0.99 \times 0.01}{0.0198} = \dfrac{0.0099}{0.0198} = \dfrac{1}{2}$

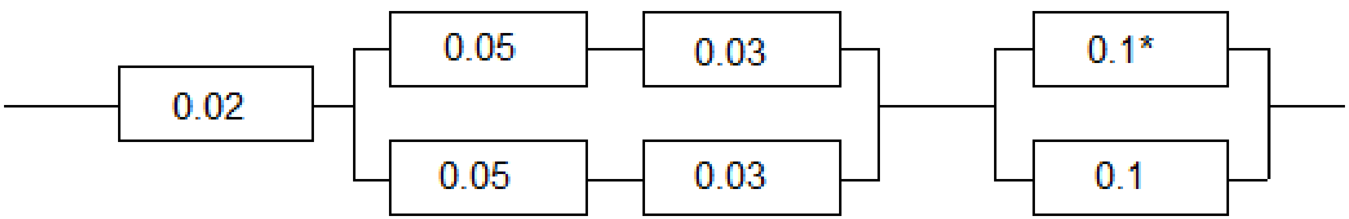
Reliability of a system

General approach: bottom-up analysis. Need to break down the system into subsystems just containing elements in series or just containing elements in parallel.

Find the reliability of each of these subsystems and then repeat the process at the next level up.

Example: reliability of a system

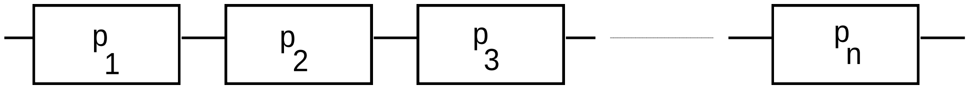
The reliability of a critical system has to be determined. An assessment has already been made of the reliability of components making up the system. The probabilities of failure of the various components in the next year are indicated in the diagram below. It can be assumed that components fail independently of one another.



(a) What is the probability that the system does not fail in the next year?

(b) Find the probability that within one year the system does not fail but component * does fail.

Series subsystem: in the diagram p_i= probability that element *i* fails, so $1-p_i$= probability that it does not fail.



The system only works if all *n* elements work. Failures of different elements are assumed to be independent (so the probability of Element 1 failing does alter after connection to the system).

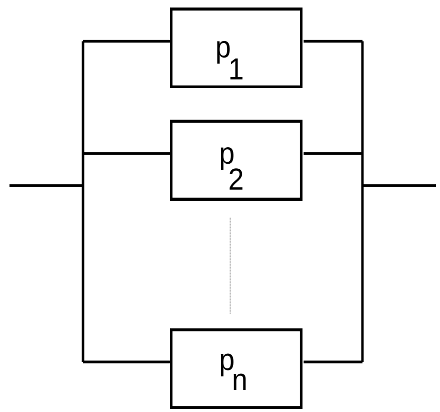
$$P(system\ does\ not\ fail) = P(1\ does\ not\ fail\ AND\ 2\ does\ not\ fail\ AND\ \ldots n\ does\ not\ fail)$$

$$= (1-p_1)(1-p_2)\ldots(1-p_n) = \prod_{i=1}^{n}(1-p_i)$$

Hence $P(system\ does\ fail) = 1 - P(system\ does\ not\ fail)$

$$= 1 - \prod_{i=1}^{n}(1-p_i)$$

Parallel subsystem: the subsystem only fails if all the elements fail.



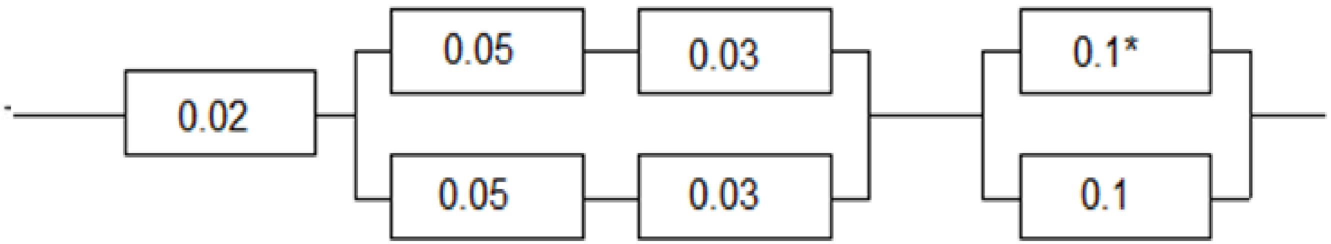
$$P(system\ fails) = P(1\ fails\ AND\ 2\ fails\ AND\ \dots n\ fails)$$

$$= P(1\ fails)P(2\ fails)\dots P(n\ fails)$$ [Special multiplication rule assuming failures independent]

$$= p_1 p_2 \dots p_n = \prod_{i=1}^{n} p_i$$

Example:

The reliability of a critical system has to be determined. An assessment has already been made of the reliability of components making up the system. The probabilities of failure of the various components in the next year are indicated in the diagram below. It can be assumed that components fail independently of one another.



→ *Subsystem 3:*
P(Subsystem 3 fails)
= 0.1 x 0.1 = 0.01

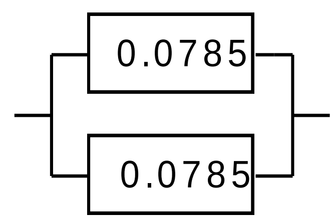
(a) What is the probability that the system does not fail in the next year?

Subsystem 1:

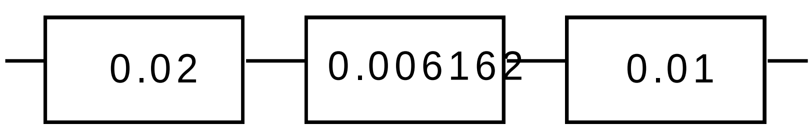
P(Subsystem 1 doesn't fail)
$= (1 - 0.05)(1 - 0.03) = 0.9215$

Hence P(Subsystem 1 fails)= 0.0785

Subsystem 2: (two units of subsystem 1)



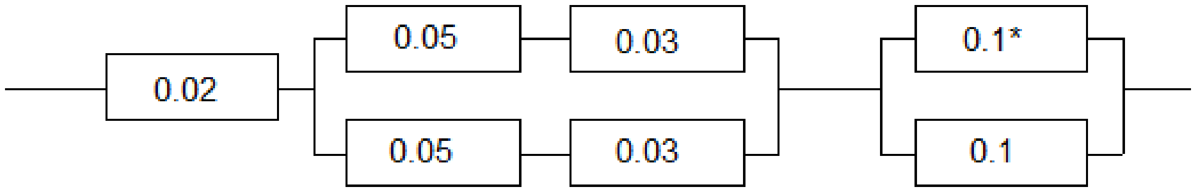
P(Subsystem 2 fails)
=
0.0785 x 0.0785 = 0.006162



Answer:
P(System doesn't fail) =
(1 - 0.02)(1 - 0.006162)(1 - 0.01)
= 0.964

Example: reliability of a system

The reliability of a critical system has to be determined. An assessment has already been made of the reliability of components making up the system. The probabilities of failure of the various components in the next year are indicated in the diagram below. It can be assumed that components fail independently of one another.



(a) What is the probability that the system does not fail in the next year?

(b) Find the probability that within one year the system does not fail but component * does fail.

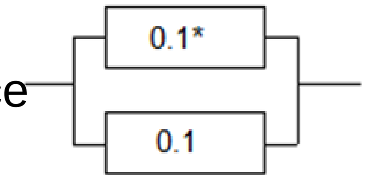
Answer to (b)

Let B = event that the system does not fail
Let C = event that component * does fail

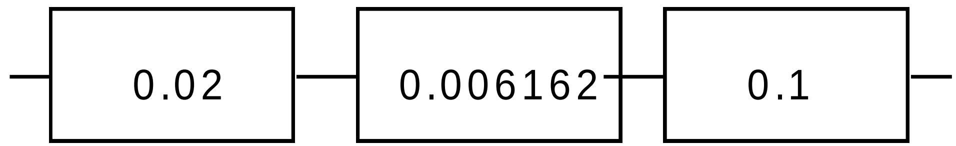
We need to find P(B and C).

Use $P(B \cap C) = P(B|C)P(C)$. We know P(C) = 0.1.

P(*B* | *C*) = P(system does not fail given component * has failed)

If * failed replace 

with 

Final diagram is then 

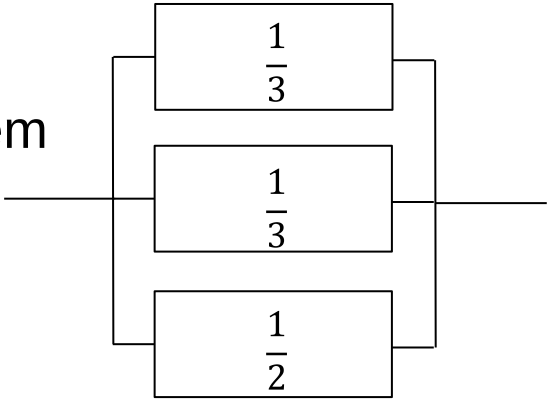
P(*B* | *C*) = (1 - 0.02)(1 – 0.006162)(1 - 0.1) = 0.8766

Hence since P(*C*) = 0.1

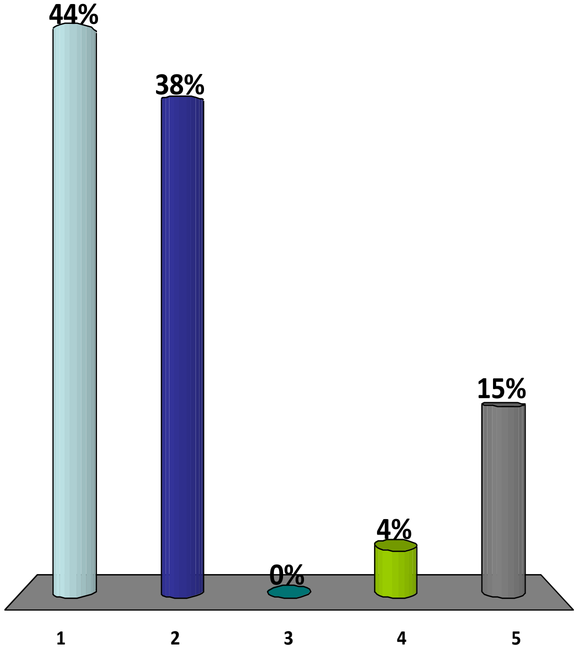
P(*B* and *C*) = P(*B* | *C*) P(*C*) = 0.8766 x 0.1 = 0.08766

Triple redundancy

What is probability that this system does not fail, given the failure probabilities of the components?

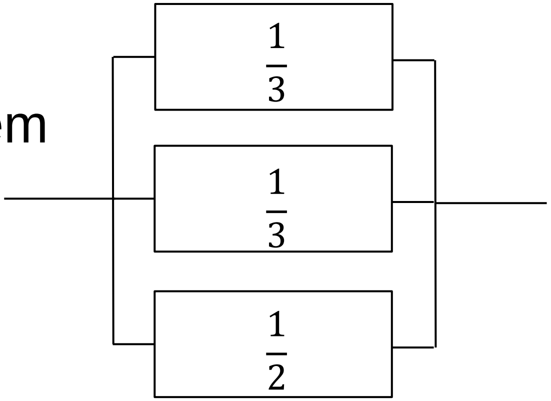


1. 17/18
2. 2/9
3. 1/9
4. 1/3
5. 1/18



Triple redundancy

What is probability that this system does not fail, given the failure probabilities of the components?



P(failing) = P(1 fails)P(2 fails)P(3 fails)= $\frac{1}{3} \times \frac{1}{3} \times \frac{1}{2} = \frac{1}{18}$

Hence: P(not failing) = 1 – P(failing) = 1 $- \frac{1}{18} = \frac{17}{18}$

Combinatorics

Permutations - ways of ordering k items: k!

Factorials: for a positive integer k, $k! = k(k-1)(k-2) \ldots 2.1$

e.g. $3! = 3 \times 2 \times 1 = 6$.
By definition, $0! = 1$.

The first item can be chosen in k ways, the second in k-1 ways, the third, in k-2 ways, etc., giving k! possible orders.

A	B	C	D	E	F
6 choices	5 choices	4 choices	3 choices	2 choices	1 choice

Total of $6 \times 5 \times 4 \times 3 \times 2 \times 1 = 6!$ possible orderings of the 6 items

e.g. ABC can be arranged as ABC, ACB, BAC, BCA, CAB and CBA, a total of 3! = 6 ways.

Ways of choosing *k* things from *n, irrespective of ordering:*

Binomial coefficient*:* for integers *n* and *k* where $n \geq k \geq 0$:

$$C_k^n = \binom{n}{k} = \frac{n!}{k!\,(n-k)!}$$

Sometimes this is also called "*n* choose *k*". Other notations include ${}_nC_k$ and variants.

Justification: Choosing *k* things from *n* there are *n* ways to choose the first item, *n-1* ways to choose the second…. and (*n-k+1*) ways to choose the last, so

$$n(n-1)(n-2)\ldots(n-k+1) = \frac{n!}{(n-k)!}$$

ways. This is the number of different orderings of *k* things drawn from *n*. But there *k*! orderings of *k* things, so only 1/k! of these is a distinct set, giving the C_k^n distinct sets.

Example: choosing 3 items from 6

A B C D E F

6 choices 5 choices 4 choices

Total of $6 \times 5 \times 4 = \frac{6\times5\times4\times3\times2\times1}{3\times2\times1} = \frac{6!}{3!}$ possible orderings of a choice of 3 items

But: the same three choices can be ordered in 3!=6 Different ways

C E B   C B E   E C B   E B C

B E C   B C E

Hence: $\frac{6!}{3!3!} = \frac{6\times5\times4}{3\times2\times1} = 20$ distinct sets of 3 things chosen from 6

ABC, ABD, ABE, ABF, ACD, ACE,ACF, ADE, ADF, AEF, BCD, BCE, BCF, BDE, BDF, BEF, CDE, CDF, CEF, DEF

Example: What is the probability of winning the National Lottery? (picking 6 numbers from a choice of 49)

Answer: the numbers of ways of choosing 6 numbers from 49 (1, 2, ... , 49) is:

$$C_6^{49} = \frac{n!}{k!\,(n-k)!} = \frac{49!}{6!\,43!} = \frac{49 \times 48 \times 47 \times 46 \times 45 \times 44}{6 \times 5 \times 4 \times 3 \times 2 \times 1} = 13{,}983{,}816$$

Each possible combination of 6 numbers is equally likely.
So the probability of winning with a given random ticket is about 1/(14 million).

Calculating factorials and C_k^n

Many calculators have a factorial button, but they become very large very quickly: $15! = 1{,}307{,}674{,}368{,}000 \approx 1.3 \times 10^{12}$, so be careful they do not overflow.

Some calculators have a button for calculating C_k^n or you can calculate it directly by cancelling factorials.

Beware that C_k^n can also become very large for large *n* and *k*, for example there are 100891344545564193334812497256 $\approx 10^{29}$ ways to choose 50 items from 100.

For computer users: In MatLab the function is called “nchoosek”, in other systems like Maple and Mathematica it is called “binomial”.

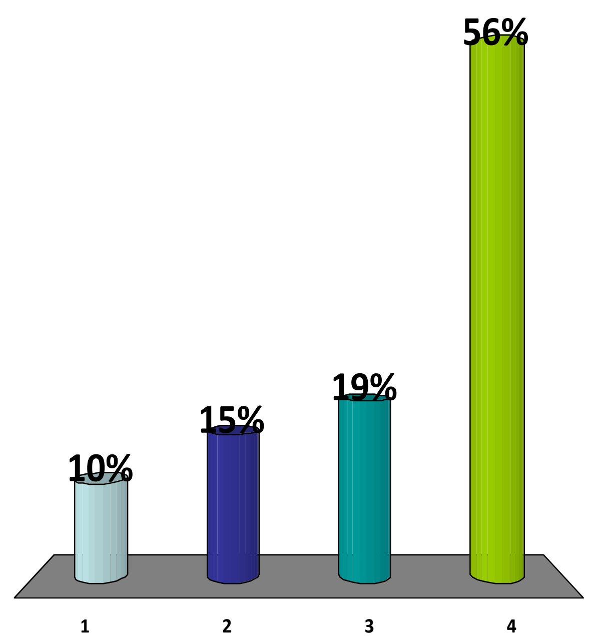
Coin tosses

A fair coin is tossed four times.

Let $P(A)$ be probability of 4 heads
Let $P(B)$ be probability of 2 heads and then 2 tails
Let $P(C)$ be probability of 1 head, then two tails, then 1 head

What is the relation between these probabilities?

1. $P(A)<P(B)<P(C)$
2. $P(A)<P(B)=P(C)$
3. $P(A)=P(B)<P(C)$
4. $P(A)=P(B)=P(C)$



Coin tosses

A fair coin is tossed four times.

Let $P(A)$ be probability of 4 heads
Let $P(B)$ be probability of 2 heads and then 2 tails
Let $P(C)$ be probability of 1 head, then two tails, then 1 head

What is the relation between these probabilities?

Answer:

For each coin toss $P(H) = P(T) = \frac{1}{2}$

For a *fair* coin, tosses are *independent:*

P(Head then tails)=P(first toss heads)P(second toss tails)=P(H)P(T), etc.

$$P(A) = P(H)P(H)P(H)P(H) \quad = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{16}$$

$$P(B) = P(H)P(H)P(T)P(T) \quad = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{16}$$

$$P(C) = P(H)P(T)P(T)P(H) \quad = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{16}$$

Any specific ordering of results has the same probability

Random variables

.

If the chance outcome of the experiment is the result of a random process, it is called a *random variable*.

Discrete random variable*:* the possible outcomes can be listed e.g. 0, 1, 2, ..., or yes/no, or A, B, C.. etc.

Continuous random variable*:* the possible outcomes are on a continuous scale e.g. weights, strengths, times or lengths.

Notation for random variables: capital letter near the end of the alphabet e.g. X, Y.

Discrete Random variables

P(X = k) denotes "The probability that X takes the value k".

Note: $0 \leq P(X = k) \leq 1$ for all k, and $\sum_k P(X = k) = 1$.

How to we quickly quantify the main properties of the distribution of the variable?

Mean (or expected value) of a variable

For a random variable *X* taking values 0, 1, 2, ... , the mean value of *X* is:

$$\mu = \sum_{k} k\, P(X = k) = 0 \times P(X = 0) + 1 \times P(X = 1) + 2 \times P(X = 2) + \cdots$$

The mean is also called: population mean
expected value of *X*
average of X
E(*X*)
$\langle X \rangle$

Intuitive idea: if *X* is observed in repeated independent experiments and

$$\bar{X}_n = \frac{1}{n} \sum_{i+1}^{n} X_i$$

is the sample mean after *n* observations then as *n* gets bigger, $\bar{X}_n$ tends to the mean μ.

Example: mean of a random whole number k between 1 and 10 is

$\mu = \sum_{k=1}^{10} k \times \frac{1}{10}$ $= (1 + 2 + 3 + 4 + 5 + 6 + 7 + 8 + 9 + 10)/10$ = 5.5

Mean (or expected value) of a function of a random variable

For a random variable X, the expected value of $f(X)$ is given by

$$E\big(f(X)\big) \equiv \langle f(X) \rangle = \sum_k f(k) P(X = k)$$

For example $f(X)$ might give the winnings on a bet on X. The expected winnings of a bet would be the sum of the winnings for each outcome multiplied by the probability of each outcome.

Note: "expected value" is a somewhat misleading technical term, equivalent in normal language to the mean or average of a population. The expected value is not necessarily a likely outcome, in fact often it is impossible. It is the average you would expect if the variable were sampled very many times.

E.g. you bet on a coin toss: Heads (H) wins W=50p, Tails (T) loses W=-£1. What is your expected winnings?

The expected "winnings" W is

$$E(W) = \langle W \rangle = £0.5 \times P(H) + (-£1) \times P(T) = £0.5 \times \frac{1}{2} - £1 \times \frac{1}{2} = -£0.25$$

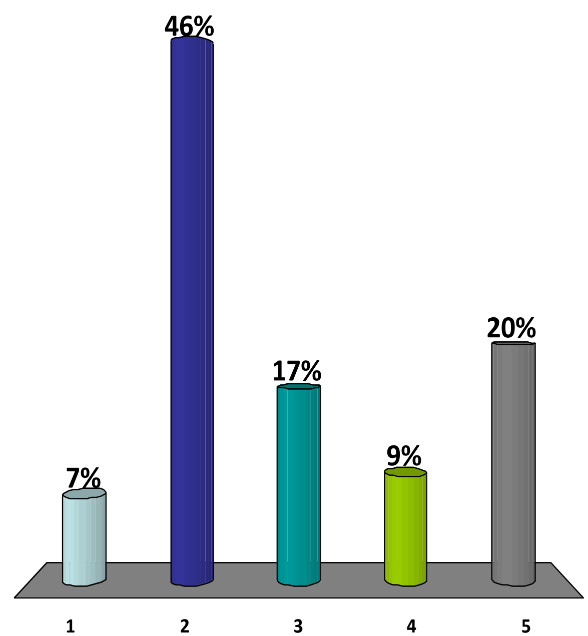
Roulette

A roulette wheel has 37 pockets.

£1 on a number returns £36 if it comes up (i.e. your £1 back + £35 winnings). Otherwise you lose your £1.

What is the expected winnings (in pounds) on a £1 number bet?

1. -1/36
2. -1/37 ✓
3. -2/37
4. -1/35
5. 1/36



Roulette

A roulette wheel has 37 pockets.

£1 on a number returns £36 if it comes up (i.e. your £1 back + £35 winnings). Otherwise you lose your £1.

What is the expected winnings (in pounds) on a £1 number bet?

Expected winnings is

£35 × P(right number) + £(-1) × P(wrong number)

$$= £35 \times \frac{1}{37} - £1 \times \frac{36}{37} = £\left(-\frac{1}{37}\right) \approx -£0.027$$

Sums of expected values

Means simply add, so e.g.

$$\langle f(X) + g(X)\rangle = \sum_k \big(f(k) + g(k)\big) P(X = k)$$

$$= \sum_k f(k) P(X = k) + \sum_k g(k) P(X = k)$$

$$= \langle f(X)\rangle + \langle g(X)\rangle$$

This also works for functions of two (or more) different random variables X and Y.

e.g. for constants a and b, and random variables X and Y:

$$\langle aX + bY\rangle = \langle aX\rangle + \langle bY\rangle = a\langle X\rangle + b\langle Y\rangle = a\mu_X + b\mu_Y$$

Note $\langle aX\rangle = \sum_k a\ k\ P(X = k) = a\ \sum_k k\ P(X = k) = a\langle X\ \rangle$

This also works for continuous random variables