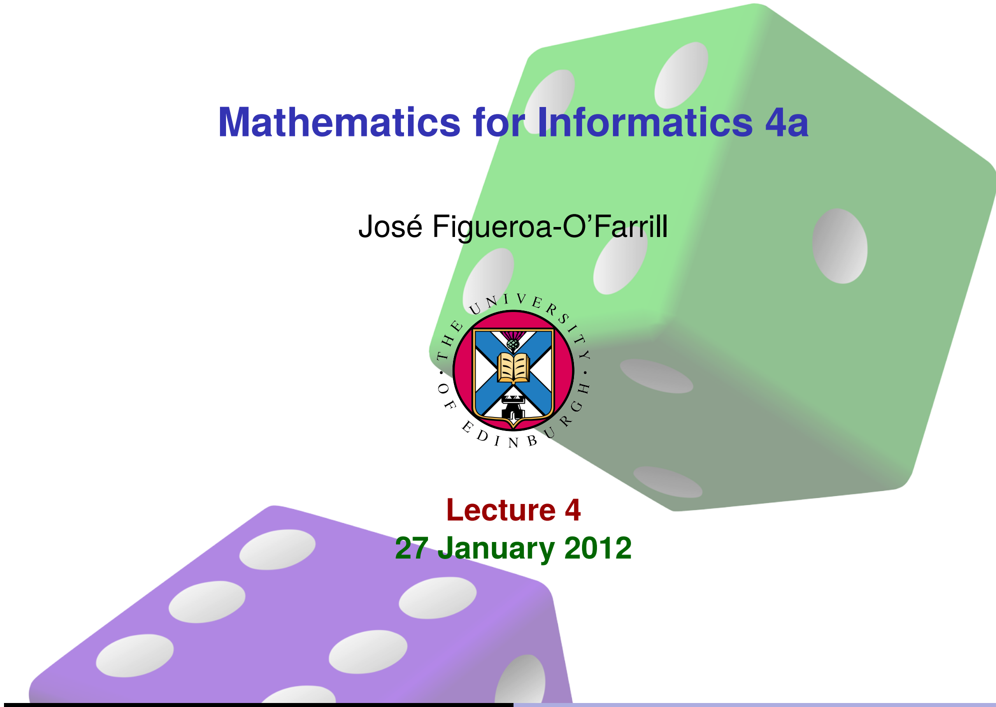


Mathematics for Informatics 4a

José Figueroa-O'Farrill

THE UNIVERSITY OF EDINBURGH

Lecture 4

27 January 2012

The story of the film so far...

- The probability of A occurring given that B has occurred is the conditional probability:

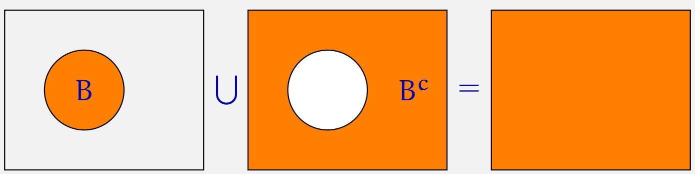
$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

- Don't confuse $\mathbb{P}(A|B)$ and $\mathbb{P}(B|A)$... or you can end up in jail!
- Events A and B are independent if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$
- Product rule: $\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B)$, so if A and B are independent, $\mathbb{P}(A|B) = \mathbb{P}(A)$.
- The method of hurdles:

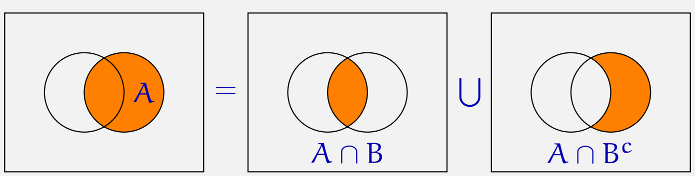
$$\mathbb{P}(A_1 \cap \cdots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_2 \cap A_1)\cdots$$

Partition theorems

For any event B, we have that $B \cup B^c = \Omega$:



So if A is any other event, $A = (A \cap B) \cup (A \cap B^c)$:



Formally,

$$A = A \cap \Omega = A \cap (B \cup B^c) = (A \cap B) \cup (A \cap B^c)$$

Because $A \cap B$ and $A \cap B^c$ are disjoint, their probabilities add:

$$\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c)\ .$$

Together with the multiplication rules

$$\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B) \qquad \text{and} \qquad \mathbb{P}(A \cap B^c) = \mathbb{P}(A|B^c)\mathbb{P}(B^c)\ ,$$

we arrive at

Theorem (The partition rule)

For any two events A, B, we have

$$\mathbb{P}(A) = \mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^c)\mathbb{P}(B^c)\ .$$

Remark

This is also called the *rule of total probability* or the *rule of alternatives*.

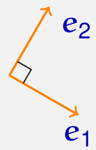
A geometric analogy

Consider $\mathbb{R}^2$ with the standard dot product:

$$\boldsymbol{x} \cdot \boldsymbol{y} = (x_1, x_2) \cdot (y_1, y_2) = x_1 y_1 + x_2 y_2 .$$

Let $(\boldsymbol{e}_1, \boldsymbol{e}_2)$ be an orthonormal basis for $\mathbb{R}^2$:

$$\boldsymbol{e}_i \cdot \boldsymbol{e}_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$



Any vector $\boldsymbol{x} \in \mathbb{R}^2$ can be decomposed in an unique way as

$$\boldsymbol{x} = (\boldsymbol{x} \cdot \boldsymbol{e}_1)\boldsymbol{e}_1 + (\boldsymbol{x} \cdot \boldsymbol{e}_2)\boldsymbol{e}_2$$

cf. $\quad \mathbb{P}(A) = \mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^c)\mathbb{P}(B^c)$

The analogue of orthogonality is now $\mathbb{P}(B|B^c) = 0$.

A general partition rule

Definition

By a (finite) **partition** of Ω we mean events $\{B_1, B_2, \dots, B_n\}$ such that $B_i \cap B_j = \varnothing$ for $i \neq j$ and $\bigcup_{i=1}^n B_i = \Omega$.

Theorem (General partition rule)

Let $\{B_1, \dots, B_n\}$ be a partition of Ω. Then for any event A,

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

Proof.

This is proved in exactly the same way as in the case of the partition $\{B, B^c\}$. $\square$

Example (More coins)

A box contains 3 double-headed coins, 2 double-tailed coins and 5 conventional coins. You pick a coin at random and flip it. What is the probability that you get a head?

Let H be the event that you get a head and let A, B, C be the events that the coin you picked was double-headed, double-tailed or conventional, respectively.

Then by the (general) partition rule

$$\begin{aligned} \mathbb{P}(H) &= \mathbb{P}(H|A)\mathbb{P}(A) + \mathbb{P}(H|B)\mathbb{P}(B) + \mathbb{P}(H|C)\mathbb{P}(C) \\ &= (1 \times \tfrac{3}{10}) + (0 \times \tfrac{2}{10}) + (\tfrac{1}{2} \times \tfrac{5}{10}) \\ &= \tfrac{3}{10} + \tfrac{1}{4} = \tfrac{11}{20} \end{aligned}$$

Example (Medical tests)

A virus infects a proportion p of individuals in a given population. A test is devised to indicate whether a given individual is infected. The probability that the test is positive for an infected individual is 95%, but there is a 10% probability of a false positive. Testing an individual at random, what is the chance of a positive result?

Let P denote the event that the result of the test is positive and V the event that the individual is infected. Then

$$\begin{aligned} \mathbb{P}(P) &= \mathbb{P}(P|V)\mathbb{P}(V) + \mathbb{P}(P|V^c)\mathbb{P}(V^c) \\ &= 0.95p + 0.1(1-p) \\ &= 0.85p + 0.1 \end{aligned}$$

(Not a very good test: if p is very small, most positive results are false positives.)

Example (Noisy channels)

Alice and Bob communicate across a noisy channel using a bit stream. Let S_0 (resp. S_1) denote the event that a 0(resp. 1) was sent, and let R_0 (resp. R_1) denote the event that a 0 (resp. 1) was received. Suppose that $\mathbb{P}(S_0) = \frac{4}{7}$ and that due to the noise $\mathbb{P}(R_1|S_0) = \frac{1}{8}$ and $\mathbb{P}(R_0|S_1) = \frac{1}{6}$. What is $\mathbb{P}(S_0|R_0)$?

$$\mathbb{P}(S_0|R_0) = \frac{\mathbb{P}(S_0 \cap R_0)}{\mathbb{P}(R_0)} = \frac{\mathbb{P}(S_0 \cap R_0)}{\mathbb{P}(S_0 \cap R_0) + \mathbb{P}(S_1 \cap R_0)}$$

$$\mathbb{P}(S_1 \cap R_0) = \mathbb{P}(R_0|S_1)\mathbb{P}(S_1) = \mathbb{P}(R_0|S_1)(1 - \mathbb{P}(S_0)) = \tfrac{1}{6} \times \tfrac{3}{7} = \tfrac{1}{14}$$
$$\mathbb{P}(S_0 \cap R_0) = \mathbb{P}(R_0|S_0)\mathbb{P}(S_0) = (1 - \mathbb{P}(R_1|S_0))\mathbb{P}(S_0) = \tfrac{7}{8} \times \tfrac{4}{7} = \tfrac{1}{2}$$

$$\therefore \quad \mathbb{P}(S_0|R_0) = \tfrac{1}{2}\Big/(\tfrac{1}{2} + \tfrac{1}{14}) = \tfrac{1}{2}\Big/\tfrac{4}{7} = \tfrac{7}{8}$$

Conditional partition rule

Theorem

Let $\{B_1, \dots, B_n\}$ be a partition of Ω and let C be an event with $\mathbb{P}(C) > 0$. Then for any event A,

$$\mathbb{P}(A|C) = \sum_{i=1}^{n} \mathbb{P}(A|B_i \cap C)\mathbb{P}(B_i|C) .$$

Proof

The partition rule holds in *any* probability space, so in particular it holds for the conditional probability $\mathbb{P}'(A \cap C) = \mathbb{P}(A|C)$. Since $\{B_1 \cap C, \dots, B_n \cap C\}$ is a partition of C,

$$\mathbb{P}'(A \cap C) = \sum_{i=1}^{n} \mathbb{P}'(A \cap C|B_i \cap C)\mathbb{P}'(B_i \cap C)$$

Proof (continued)

We rewrite $\mathbb{P}'(A \cap C) = \sum_{i=1}^{n} \mathbb{P}'(A \cap C|B_i \cap C)\mathbb{P}'(B_i \cap C)$ as

$$\mathbb{P}(A|C) = \sum_{i=1}^{n} \mathbb{P}'(A \cap C|B_i \cap C)\mathbb{P}(B_i|C) .$$

We finish the proof by rewriting $\mathbb{P}'(A \cap C|B_i \cap C)$ as follows

$$\begin{aligned}
\mathbb{P}'(A \cap C|B_i \cap C) &= \frac{\mathbb{P}'(A \cap B_i \cap C)}{\mathbb{P}'(B_i \cap C)} = \frac{\mathbb{P}(A \cap B_i|C)}{\mathbb{P}(B_i|C)} \\
&= \frac{\mathbb{P}(A \cap B_i \cap C)}{\mathbb{P}(C)}\Big/\frac{\mathbb{P}(B_i \cap C)}{\mathbb{P}(C)} \\
&= \frac{\mathbb{P}(A \cap B_i \cap C)}{\mathbb{P}(B_i \cap C)} \\
&= \mathbb{P}(A|B_i \cap C) .
\end{aligned}$$

□

Example

There are number of different drugs to treat a disease and each drug may give rise to side effects. A certain drug C has a 99% success rate in the absence of side effects and side effects only arise in 5% of cases. If they do arise, however, then C has only a 30% success rate. If C is used, what is the probability of the event A that a cure is effected?

Let B be the event that **no** side effects occur. We are given that

$$\mathbb{P}(A|B \cap C) = 0.99 \qquad \mathbb{P}(B|C) = 0.95 \qquad \mathbb{P}(A|B^c \cap C) = 0.3 ,$$

whence $\mathbb{P}(B^c|C) = 0.05$. By the conditional partition rule corresponding to the partition $\{B, B^c\}$ and condition C,

$$\begin{aligned}
\mathbb{P}(A|C) &= \mathbb{P}(A|B \cap C)\mathbb{P}(B|C) + \mathbb{P}(A|B^c \cap C)\mathbb{P}(B^c|C) \\
&= (0.99 \times 0.95) + (0.3 \times 0.05) = 0.9555 \simeq 96\%
\end{aligned}$$

Bayes's rule

Recall the product rule

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

which immediately gives

Theorem (Bayes's rule)

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

Using the partition rule $\mathbb{P}(B) = \mathbb{P}(B|A)\mathbb{P}(A) + \mathbb{P}(B|A^c)\mathbb{P}(A^c)$ we get a modified version of Bayes's rule:

Theorem (Bayes's rule too)

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B|A)\mathbb{P}(A) + \mathbb{P}(B|A^c)\mathbb{P}(A^c)}$$

Example (False positives)

You get tested for the virus in the earlier example *and it shows positive. What is the probability that you are actually infected?*
In the notation of the earlier example, we want to compute $\mathbb{P}(V|P)$. By Bayes's rule

$$\mathbb{P}(V|P) = \frac{\mathbb{P}(P|V)\mathbb{P}(V)}{\mathbb{P}(P)} = \frac{0.95p}{0.85p + 0.1}$$

So that if half the population is infected ($p = 0.5$), then $\mathbb{P}(V|P) \simeq 90\%$ and the test looks good, but if the virus affects only one person in every thousand ($p = 10^{-3}$), then $\mathbb{P}(V|P) \simeq 1\%$, so not very conclusive at all!

Example (Multiple choice exam)

A student is taking a multiple choice exam, each question having c available choices. The student either knows the answer to the question with probability p or else guesses at random with probability $1 - p$. Given that the answer selected is correct, what is the probability that the student knew the answer?
Let A denote the event that the answer is correct and let K denote the event that the student knew the answer. We are after $\mathbb{P}(K|A)$. Bayes's rule says

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

so we need to compute $\mathbb{P}(A)$. We will use the partition rule,

$$\mathbb{P}(A) = \mathbb{P}(A|K)\mathbb{P}(K) + \mathbb{P}(A|K^c)\mathbb{P}(K^c) .$$

Example (Multiple choice exam – continued)

We notice that $\mathbb{P}(A|K) = 1$ and $\mathbb{P}(A|K^c) = 1/c$, whence

$$\begin{aligned}\mathbb{P}(A) &= \mathbb{P}(A|K)\mathbb{P}(K) + \mathbb{P}(A|K^c)\mathbb{P}(K^c) \\ &= (1 \times p) + (\tfrac{1}{c} \times (1 - p)) \\ &= p + \tfrac{1-p}{c} .\end{aligned}$$

Finally,

$$\mathbb{P}(K|A) = \frac{p}{p + (1 - p)/c} = \frac{cp}{1 + (c - 1)p} .$$

Notice that the larger the number c, the more likely that the student knew the answer.

Conditional probability in Mendelian genetics I

Basic question

How does your family's medical history affect your carrying a certain gene?

First some basic genetics:

- **Genes** are nucleotide sequences forming part of a **chromosome**.
- Humans have 23 pairs of chromosomes, one of which are the sex chromosomes which come in two varieties X and Y. Females have two X chromosomes, whereas males have one X and one Y chromosome.
- Certain genetic traits are passed on through genes in the X chromosomes: the so-called **X-linked genetic traits**.
- A given gene can come in two (or more) mutated forms called **alleles**.

Conditional probability in Mendelian genetics II

- Let us consider a gene contained inside the X chromosome and having two alleles: A and a.
- We will assume that a male with allele A in his one X chromosome does **not** present the genetic trait, whereas one with a does.
- We will assume that a female will present the trait if and only if both her X chromosomes contain the allele a.
- One says that allele A is **dominant** and a is **recessive**.
- We will assume the following laws of inheritance:
  - a son gets one of his mother's two X chromosomes at random
  - a daughter gets her father's X chromosome and one of her mother's at random
- Males can therefore be A or a, whereas females can be AA, Aa and aa. (We don't distinguish between Aa and aA.)

Conditional probability in Mendelian genetics III

Example

Suppose that a male with genotype A and a female with genotype Aa have a daughter. She can have genotype AA or Aa, both with probability $\frac{1}{2}$. Now suppose that she herself has a son with genotype A. What is the (conditional) probability that she has genotype AA?

Let G_{AA} (resp. G_{Aa}) denote the event that the daughter has genotype AA (resp. Aa) and let S_A denote the event that the daughter's son has genotype A. We want $\mathbb{P}(G_{AA}|S_A)$. Notice that $\mathbb{P}(S_A|G_{AA}) = 1$ and $\mathbb{P}(S_A|G_{Aa}) = \frac{1}{2}$. By the partition rule

$$\begin{aligned}\mathbb{P}(S_A) &= \mathbb{P}(S_A|G_{AA})\mathbb{P}(G_{AA}) + \mathbb{P}(S_A|G_{Aa})\mathbb{P}(G_{Aa}) \\ &= (1 \times \tfrac{1}{2}) + (\tfrac{1}{2} \times \tfrac{1}{2}) = \tfrac{1}{2} + \tfrac{1}{4} = \tfrac{3}{4}\,.\end{aligned}$$

Bayes's: $\mathbb{P}(G_{AA}|S_A) = \mathbb{P}(S_A|G_{AA})\mathbb{P}(G_{AA})/\mathbb{P}(S_A) = \frac{1}{2}\Big/\frac{3}{4} = \frac{2}{3}$.

Summary

- The partition rule: $\mathbb{P}(A) = \mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^c)\mathbb{P}(B^c)$
- This generalises to a partition $\{B_i\}$ of the sample space:

$$\mathbb{P}(A) = \sum_i \mathbb{P}(A|B_i)\mathbb{P}(B_i)$$

- It also applies to conditional probability:

$$\mathbb{P}(A|C) = \sum_i \mathbb{P}(A|B_i \cap C)\mathbb{P}(B_i|C)$$

- Bayes's rule allows us to compute $\mathbb{P}(A|B)$ from a knowledge of $\mathbb{P}(B|A)$ via

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B|A)\mathbb{P}(A) + \mathbb{P}(B|A^c)\mathbb{P}(A^c)}$$