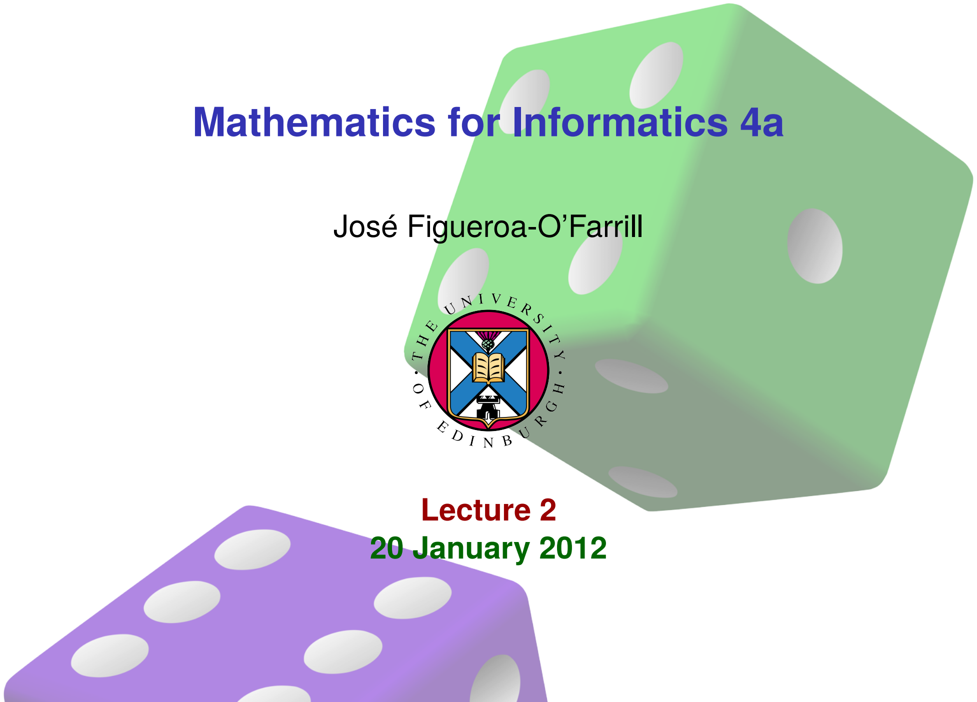


Mathematics for Informatics 4a

José Figueroa-O'Farrill

Lecture 2
20 January 2012

The story of the film so far...

Our first goal in this course is to formalise assertions such as

"the chance of A is p",

where

- the event A is a subset of the set of outcomes of some experiment, and
- p is some measure of the likelihood of event A occuring, by which we mean that the outcome of the experiment belongs to A.

The set of outcomes is denoted Ω and the possible events define a σ-field $\mathcal{F}$ of subsets of Ω; that is, a family of subsets which contains $\varnothing$ and Ω and is closed under complementation, countable union and countable intersection.

Probability as relative frequency

Imagine repeating an experiment N times.
We let $N(A)$ denote the number of times that the event A occurs. Clearly, $0 \leqslant N(A) \leqslant N$. If the following limit exists

$$\lim_{N\to\infty} \frac{N(A)}{N} = \mathbb{P}(A) ,$$

the number $\mathbb{P}(A)$ obeys $0 \leqslant \mathbb{P}(A) \leqslant 1$ and is called the **probability** of A occuring.
Since $N(\Omega) = N$, we have $\mathbb{P}(\Omega) = 1$.
If $A \cap B = \varnothing$, $N(A \cup B) = N(A) + N(B)$, whence $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$. By induction, if A_i is a finite family of pairwise disjoint events,

$$\mathbb{P}(A_1 \cup A_2 \cup \cdots \cup A_n) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \cdots + \mathbb{P}(A_n) .$$

As usual, it is (theoretically) convenient to extend this to countable families.

Probability measures

Definition

A **probability measure** on $(\Omega, \mathcal{F})$ is a function $\mathbb{P} : \mathcal{F} \to [0, 1]$ satisfying

- $\mathbb{P}(\Omega) = 1$
- if $A_i \in \mathcal{F}$, $i = 1, 2, \ldots$ are such that $A_i \cap A_j = \varnothing$ for all $i \neq j$, then
$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i) .$$

The triple $(\Omega, \mathcal{F}, \mathbb{P})$ is called a **probability space**.

Remark

Since $\Omega = A \cup A^c$ is a disjoint union, $\mathbb{P}(A) + \mathbb{P}(A^c) = 1$. In particular, $\mathbb{P}(\varnothing) = 0$.

Bernoulli trials

Definition

A **Bernoulli trial** is any trial with only two outcomes.

Example

Tossing a coin is a Bernoulli trial: $\Omega = \{H, T\}$.

Let us call the two outcomes generically "success" (S) and "failure" (F), so that $\Omega = \{S, F\}$. Then we have

$$\mathbb{P}(\{S\}) = p \qquad \text{and} \qquad \mathbb{P}(\{F\}) = q\ .$$

Since $\Omega = \{S\} \cup \{F\}$ is a disjoint union, it follows that $q = 1 - p$.

Notation

We will often drop the $\{\ \ \}$ when talking about events consisting of a single outcome and will write $\mathbb{P}(S) = p$ and $\mathbb{P}(F) = 1 - p$.

Fair coins and fair dice

Tossing a coin has $\Omega = \{H, T\}$. Let $\mathbb{P}(H) = p$ and $\mathbb{P}(T) = 1 - p$. The coin is **fair** if $p = \frac{1}{2}$, so that both H and T are equally probable.

Similarly, a **fair** die is one where every outcome has the same probability. Since there are six outcomes, each one has probability $\frac{1}{6}$:

$$\mathbb{P}(⚀) = \mathbb{P}(⚁) = \mathbb{P}(⚂) = \mathbb{P}(⚃) = \mathbb{P}(⚄) = \mathbb{P}(⚅) = \frac{1}{6}\ .$$

Fair coins and fair dice are examples of "uniform probability spaces": those whose outcomes are all equally likely.

Uniform probability measures

Suppose that Ω is a finite set of cardinality $|\Omega|$, and suppose that every outcome is equally likely: $\mathbb{P}(\omega) = p$ for all $\omega \in \Omega$.
Since $\Omega = \bigcup_{\omega \in \Omega} \omega$ is a disjoint union, we have

$$1 = \mathbb{P}\left(\bigcup_{\omega \in \Omega} \omega\right) = \sum_{\omega \in \Omega} \mathbb{P}(\omega) = \sum_{\omega \in \Omega} p = p|\Omega|\ ,$$

whence $p = 1/|\Omega|$.
Now let $A \subseteq \Omega$ be an event:

$$\mathbb{P}(A) = \mathbb{P}\left(\bigcup_{\omega \in A} \omega\right) = \sum_{\omega \in A} \mathbb{P}(\omega) = \sum_{\omega \in A} \frac{1}{|\Omega|} = \frac{|A|}{|\Omega|}$$

Example

You draw a number at random from $\{1, 2, \ldots, 30\}$. *What is the probability of the following events:*

1. $A =$ *the number drawn is even*
2. $B =$ *the number drawn is divisible by 3*
3. $C =$ *the number drawn is less than 12*

There are 30 possible outcomes, all equally likely ("at random").

1. $A = \{2, 4, 6, \ldots, 30\}$, so $|A| = 15$ and hence $\mathbb{P}(A) = \frac{15}{30} = \frac{1}{2}$.
2. $B = \{3, 6, 9, \ldots, 30\}$, so $|B| = 10$ and hence $\mathbb{P}(B) = \frac{10}{30} = \frac{1}{3}$.
3. $C = \{1, 2, 3, \ldots, 11\}$, so $|C| = 11$ and hence $\mathbb{P}(C) = \frac{11}{30}$.

Example

Three fair dice are rolled and their scores added.
Which is more likely: a 9 or a 10?

There are 6^3 possible outcomes, all equally likely.

- $6 \times$ ⚀⚁⚅
- $6 \times$ ⚀⚂⚄
- $3 \times$ ⚀⚃⚃
- $3 \times$ ⚁⚁⚄
- $6 \times$ ⚁⚂⚃
- $1 \times$ ⚂⚂⚂

$\mathbb{P}(9) = 25/6^3$

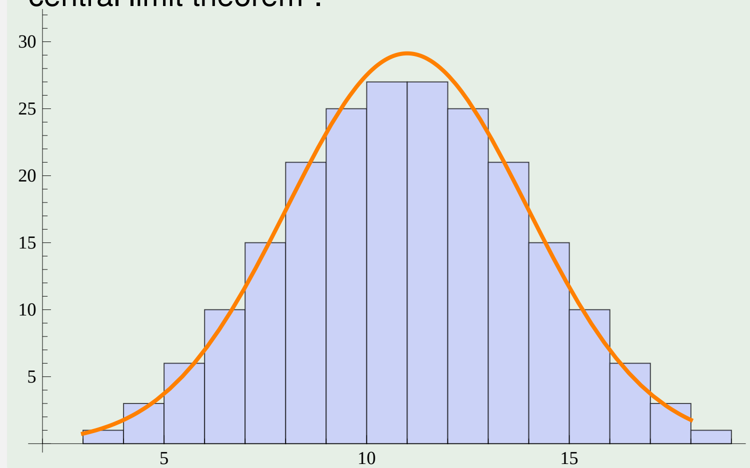
- $6 \times$ ⚀⚂⚅
- $6 \times$ ⚀⚃⚄
- $3 \times$ ⚁⚁⚅
- $6 \times$ ⚁⚂⚄
- $3 \times$ ⚁⚃⚃
- $3 \times$ ⚂⚂⚃

$\mathbb{P}(10) = 27/6^3$

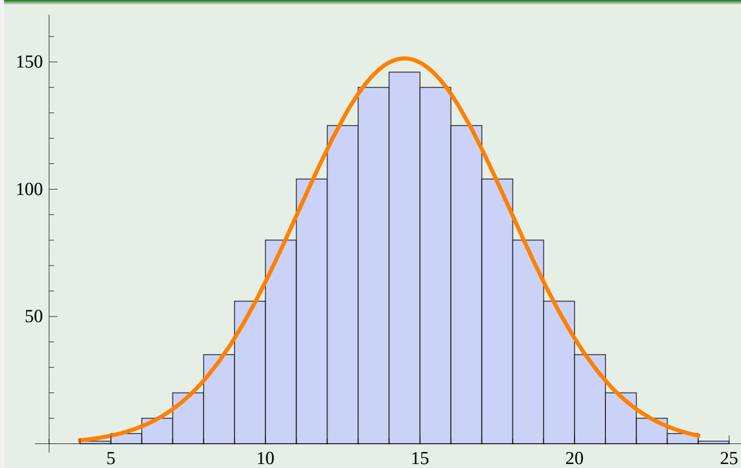
Therefore $\mathbb{P}(10) > \mathbb{P}(9)$.

Example (Continued)

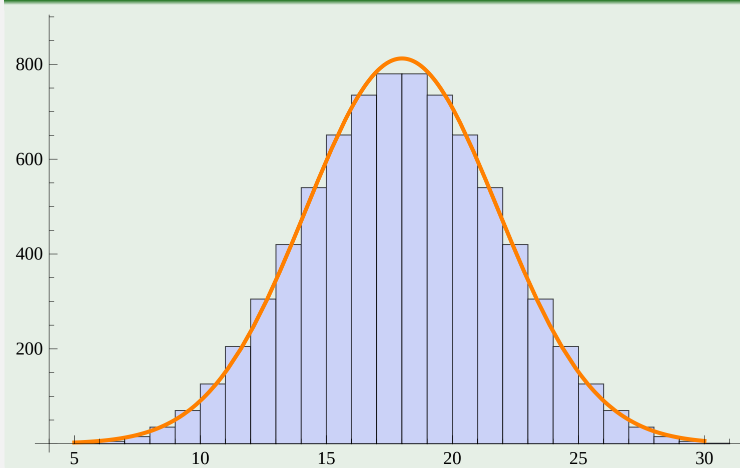
We could have answered this without enumeration. The average score of rolling three dice is $10\frac{1}{2}$. Since 10 is closer than 9 to the average, $\mathbb{P}(10) \geqslant \mathbb{P}(9)$ as a consequence of the "central limit theorem".



Example (Continued)



Example (Continued)



Example (Alice and Bob's game)

Alice and Bob toss a fair coin in turn and the winner is the first one to get H. Suppose that Alice goes first and consider the three events:

- $A =$ Alice wins
- $B =$ Bob wins
- $C =$ nobody wins

Let ω_i be the outcome $\underbrace{TT\cdots T}_{i-1}H$. Then $A = \{\omega_1, \omega_3, \omega_5, \dots\}$ and $B = \{\omega_2, \omega_4, \omega_6, \dots\}$. There is a further possible outcome ω_∞, corresponding to the unending game $TTT\cdots$ in which nobody wins. Hence $C = \{\omega_\infty\}$. The (countably infinite) sample space is $\Omega = \{\omega_1, \omega_2, \dots, \omega_\infty\}$, which is the disjoint union $\Omega = A \cup B \cup C$. Therefore $\mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) = 1$.

Example (continued)

There are 2^n possible outcomes of tossing the coin n times, all equally likely. Hence $\mathbb{P}(\omega_n) = 1/2^n$. Therefore since $A = \bigcup_{n=0}^\infty \{\omega_{2n+1}\}$ is a disjoint union,

$$\mathbb{P}(A) = \sum_{n=0}^\infty \mathbb{P}(\omega_{2n+1}) = \sum_{n=0}^\infty 2^{-2n-1} = \frac{1}{2}\sum_{n=0}^\infty \frac{1}{4^n} = \frac{\frac{1}{2}}{1-\frac{1}{4}} = \frac{2}{3} .$$

Similarly, since $B = \bigcup_{n=1}^\infty \{\omega_{2n}\}$ is also a disjoint union,

$$\mathbb{P}(B) = \sum_{n=1}^\infty \mathbb{P}(\omega_{2n}) = \sum_{n=1}^\infty 2^{-2n} = \frac{1}{4}\sum_{n=0}^\infty \frac{1}{4^n} = \frac{\frac{1}{4}}{1-\frac{1}{4}} = \frac{1}{3} .$$

Finally, since $\mathbb{P}(A) + \mathbb{P}(B) = 1$, we see that $\mathbb{P}(C) = 0$.

Warning

Although $\mathbb{P}(C) = 0$, the event C is **not** impossible.

Basic properties of probability measures

Theorem

1. $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$
2. *if* $B \supseteq A$ *then* $\mathbb{P}(B) \geqslant \mathbb{P}(A)$

Proof.

1. $\Omega = A \cup A^c$ is a disjoint union, whence
$$1 = \mathbb{P}(\Omega) = \mathbb{P}(A) + \mathbb{P}(A^c) .$$
2. Write B as the disjoint union $B = (B \setminus A) \cup A$, whence
$$\mathbb{P}(B) = \mathbb{P}(B \setminus A) + \mathbb{P}(A) \geqslant \mathbb{P}(A) .$$

□

Example (The Birthday problem)

What is the probability that among n people chosen at random, there are at least 2 people sharing the same birthday?

Let A_n be the event where at least two people in n share the same birthday. Then A_n^c is the event that no two people in n share the same birthday and $\mathbb{P}(A_n) = 1 - \mathbb{P}(A_n^c)$.
There are 365^n possible outcomes to the birthdays of n people and $365 \times 364 \times \cdots \times (365 - n + 1)$ possible outcomes consisting of n different birthdays, hence

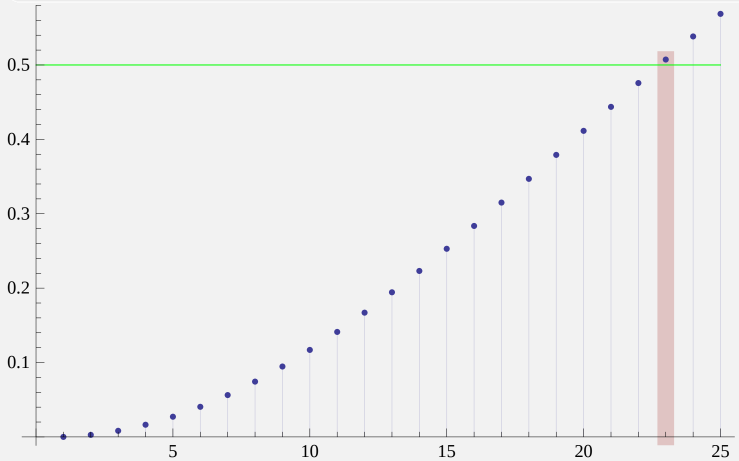
$$\mathbb{P}(A_n^c) = \frac{365 \times 364 \times \cdots \times (365 - n + 1)}{365^n} = \prod_{i=1}^{n-1} \left(1 - \tfrac{i}{365}\right)$$

and

$$\mathbb{P}(A_n) = 1 - \prod_{i=1}^{n-1} \left(1 - \tfrac{i}{365}\right) .$$

Example (continued)

For which value of n is the chance of two people sharing the same birthday better than evens?



Answer: **23** (!)

Inclusion-exclusion rule

Theorem

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

Proof.

- $A = (A \setminus B) \cup (A \cap B)$ whence $\mathbb{P}(A) = \mathbb{P}(A \setminus B) + \mathbb{P}(A \cap B)$
- $B = (B \setminus A) \cup (A \cap B)$ whence $\mathbb{P}(B) = \mathbb{P}(B \setminus A) + \mathbb{P}(A \cap B)$
- $A \cup B = (A \triangle B) \cup (A \cap B)$ whence

$$\begin{aligned}\mathbb{P}(A \cup B) &= \mathbb{P}(A \triangle B) + \mathbb{P}(A \cap B) \\ &= \mathbb{P}(A \setminus B) + \mathbb{P}(B \setminus A) + \mathbb{P}(A \cap B) \\ &= \mathbb{P}(A) - \mathbb{P}(A \cap B) + \mathbb{P}(B) - \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B) \\ &= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)\end{aligned}$$

□

Example

Historical meteorological records for a certain seaside location show that on New Year's day there is a 30% chance of rain, 40% chance of being windy and 20% chance of both rain and wind. What is the chance of it being dry? dry and windy? wet or windy?

- $\mathbb{P}(\text{dry}) = 1 - \mathbb{P}(\text{wet}) = 1 - \frac{3}{10} = \frac{7}{10}$
- $\mathbb{P}(\text{dry and windy}) = \mathbb{P}(\text{windy but not wet}) = \mathbb{P}(\text{windy}) - \mathbb{P}(\text{wet and windy}) = \frac{4}{10} - \frac{2}{10} = \frac{2}{10}$
- $\mathbb{P}(\text{wet or windy}) = \mathbb{P}(\text{wet}) + \mathbb{P}(\text{windy}) - \mathbb{P}(\text{wet and windy}) = \frac{3}{10} + \frac{4}{10} - \frac{2}{10} = \frac{1}{2}$

Boole's inequality

Theorem

$\mathbb{P}(A_1 \cup A_2 \cup \cdots \cup A_n) \leqslant \mathbb{P}(A_1) + \mathbb{P}(A_2) + \cdots + \mathbb{P}(A_n)$

Proof.

$A_1 \cup A_2 \cup \cdots \cup A_n = (A_1 \cup A_2 \cup \cdots \cup A_{n-1}) \cup A_n$, and by the inclusion-exclusion rule,

$$\mathbb{P}(A_1 \cup A_2 \cup \cdots \cup A_n) \leqslant \mathbb{P}(A_1 \cup A_2 \cup \cdots \cup A_{n-1}) + \mathbb{P}(A_n)$$

But now $A_1 \cup A_2 \cup \cdots \cup A_{n-1} = (A_1 \cup A_2 \cup \cdots \cup A_{n-2}) \cup A_{n-1}$, so that

$$\mathbb{P}(A_1 \cup A_2 \cup \cdots \cup A_n) \leqslant \mathbb{P}(A_1 \cup A_2 \cup \cdots \cup A_{n-2}) + \mathbb{P}(A_{n-1}) + \mathbb{P}(A_n)$$

et cetera. □

General inclusion-exclusion rule

Theorem

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathbb{P}(A_i) - \sum_{1\leqslant i<j\geqslant n} \mathbb{P}(A_i \cap A_j)$$

$$+ \sum_{1\leqslant i<j<k\leqslant n} \mathbb{P}(A_i \cap A_j \cap A_k) - \cdots + (-1)^n \mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right)$$

Proof.

Use induction from the simple inclusion-exclusion rule. □

Continuity

Theorem

1. *Let* $A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots$ *and let* $A = \bigcup_{i=1}^{\infty} A_i = \lim_{i\to\infty} A_i$. *Then* $P(A) = \lim_{i\to\infty} \mathbb{P}(A_i)$.
2. *Let* $B_1 \supseteq B_2 \supseteq B_3 \supseteq \cdots$ *and let* $B = \bigcap_{i=1}^{\infty} B_i = \lim_{i\to\infty} B_i$. *Then* $P(B) = \lim_{i\to\infty} \mathbb{P}(B_i)$.

Proof.

1. $A = A_1 \cup (A_2 \setminus A_1) \cup (A_3 \setminus A_2) \cup \cdots$ is a disjoint union:

$$\begin{aligned}\mathbb{P}(A) &= \mathbb{P}(A_1) + \mathbb{P}(A_2 \setminus A_1) + \mathbb{P}(A_3 \setminus A_2) + \cdots \\ &= \mathbb{P}(A_1) + (\mathbb{P}(A_2) - \mathbb{P}(A_1)) + (\mathbb{P}(A_3) - \mathbb{P}(A_2)) + \cdots \\ &= \lim_{n\to\infty} \mathbb{P}(A_n).\end{aligned}$$

2. Take complements of the previous proof.

Summary

Every experiment has an associated **probability space** $(\Omega, \mathcal{F}, \mathbb{P})$, where

- Ω is the sample space (set of all outcomes),
- $\mathcal{F}$ is the σ-field of events, and
- $\mathbb{P} : \mathcal{F} \to [0, 1]$ is a probability measure:
  - normalised so that $\mathbb{P}(\Omega) = 1$
  - countably additive over disjoint unions

Probability spaces with Ω finite and $\mathbb{P}$ uniformly distributed ("all outcomes equally likely") are particularly amenable to counting techniques from combinatorial analysis.