COMP9020

Foundations of Computer Science

Lecture 15: Probability

Lecturers: Katie Clinch (LIC)
Paul Hunter
Simon Mackenzie

Course admin: Nicholas Tandiono

Course email: `cs9020@cse.unsw.edu.au`

Announcements

- **myExperience surveys** - please fill these in
- **Final exam** - common questions
  - It will be in Inspera
  - You can complete it remotely on your own computer.
  - It is open book, and you can use any web-browser.
  - I'll go into further details in the **revision lecture**.
  - Vote for topics you'd like to cover in the revision lecture here.
- **Upcoming deadlines**:
  - Quiz 8 (deadline Wednesday 17th April)
  - Assignment 4 (deadline Thursday 18th April)

Outline

Outline

Elementary Probability

Definition

Sample space:

$$\Omega = \{\omega_1, \ldots, \omega_n\}$$

Each point represents an outcome.

Event: a collection of outcomes = subset of Ω

Probability distribution: A function $P : \mathrm{Pow}(\Omega) \to \mathbb{R}$ such that:

- $P(\Omega) = 1$
- E and F disjoint events then $P(E \cup F) = P(E) + P(F)$.

Fact

$$P(\emptyset) = 0, \quad P(E^c) = 1 - P(E)$$

Examples

Examples

Tossing a coin: $\Omega = \{H, T\}$

$$P(H) = P(T) = 0.5$$

Rolling a die: $\Omega = \{1, 2, 3, 4, 5, 6\}$

$$P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = \frac{1}{6}$$

Uniform distribution

Each outcome ω_i equally likely:

$$P(\omega_1) = P(\omega_2) = \ldots = P(\omega_n) = \frac{1}{n}$$

This a called a **uniform probability distribution** over Ω

Examples

Tossing a coin: $\Omega = \{H, T\}$

$$P(H) = P(T) = 0.5$$

Rolling a die: $\Omega = \{1, 2, 3, 4, 5, 6\}$

$$P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = \frac{1}{6}$$

Computing Probabilities by Counting

Computing probabilities with respect to a *uniform* distribution comes down to counting the size of the event.
If $E = \{e_1, \ldots, e_k\}$ then

$$P(E) = \sum_{i=1}^{k} P(e_i) = \sum_{i=1}^{k} \frac{1}{|\Omega|} = \frac{|E|}{|\Omega|}$$

Most of the counting rules carry over to probabilities wrt. a uniform distribution.

Important!

The expression "selected at random", when not further qualified, means:
"subject to / according to / … a *uniform* distribution."

Combining events

We can create complex events by combining simpler ones. Common constructions:

- A and B: $A \cap B$
- A or B: $A \cup B$
- Not A: $\Omega \setminus A$
- A followed by B

The first three involve events from the same set of outcomes. The last may involve events from different sets of outcomes (e.g. roll die and flip coin).

Inclusion-exclusion rule

Fact

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

Exercises

Exercises

$\boxed{\text{RW: 5.2.7}}$ Suppose an experiment leads to events A, B with probabilities $P(A) = 0.5, P(B) = 0.8, P(A \cap B) = 0.4$.
Find

- $P(B^c)$
- $P(A \cup B)$
- $P(A^c \cup B^c)$

$\boxed{\text{RW: 5.2.8}}$ Given $P(A) = 0.6,\ P(B) = 0.7$, show $P(A \cap B) \geq 0.3$

Exercises

$\boxed{\text{RW: 5.2.7}}$ Suppose an experiment leads to events A, B with probabilities $P(A) = 0.5, P(B) = 0.8, P(A \cap B) = 0.4$.
Find

- $P(B^c)$ $\quad 1 - P(B) = 0.2$
- $P(A \cup B)$ $\quad P(A) + P(B) - P(A \cap B) = 0.9$
- $P(A^c \cup B^c)$ $\quad 1 - P((A^c \cup B^c)^c) = 1 - P(A \cap B) = 0.6$

$\boxed{\text{RW: 5.2.8}}$ Given $P(A) = 0.6,\ P(B) = 0.7$, show $P(A \cap B) \geq 0.3$

Exercises

Exercises

$\boxed{\text{RW: 5.2.7}}$ Suppose an experiment leads to events A, B with probabilities $P(A) = 0.5, P(B) = 0.8, P(A \cap B) = 0.4$.
Find

- $P(B^c)$ $\quad 1 - P(B) = 0.2$
- $P(A \cup B)$ $\quad P(A) + P(B) - P(A \cap B) = 0.9$
- $P(A^c \cup B^c)$ $\quad 1 - P((A^c \cup B^c)^c) = 1 - P(A \cap B) = 0.6$

$\boxed{\text{RW: 5.2.8}}$ Given $P(A) = 0.6,\ P(B) = 0.7$, show $P(A \cap B) \geq 0.3$

$$\begin{aligned} P(A \cap B) &= P(A) + P(B) - P(A \cup B) \\ &= 0.6 + 0.7 - P(A \cup B) \\ &\geq 0.6 + 0.7 - 1 = 0.3 \end{aligned}$$

Examples

Example

Question. A four-digit number n is selected at random (i.e. randomly from $[1000, \ldots, 9999]$). Find the probability p that n has all of 0, 1, 2 among its digits.

Answer (approach). Let $q = 1 - p$ be the complementary probability and define

$$A_i = \{n : \text{no digit } i\}, A_{ij} = \{n : \text{no digits } i, j\}, A_{ijk} = \{n : \text{no } i, j, k\}$$

Then define
$T = A_0 \cup A_1 \cup A_2 = \{n : \text{ missing at least one of } 0, 1, 2\}$
$S = (A_0 \cup A_1 \cup A_2)^c = \{n : \text{ containing all of } 0, 1, 2\}$

Example (cont'd)

Once we find the cardinality of T, the solution is

$$q = \frac{|T|}{9000}, \; p = 1 - q$$

To find $|A_i|, |A_{ij}|, |A_{ijk}|$ we reflect on how many choices are available for the first digit, for the second etc. A special case is the leading digit, which must be $1, \ldots, 9$

Example (cont'd)

Answer (arithmetic).

$$
\begin{aligned}
|A_0| &= 9^4, \quad |A_1| = |A_2| = 8 \cdot 9^3 \\
|A_{01}| &= |A_{02}| = 8^4, \quad |A_{12}| = 7 \cdot 8^3 \\
|A_{012}| &= 7^4
\end{aligned}
$$

$$
\begin{aligned}
|T| &= |A_0 \cup A_1 \cup A_2| \\
&= |A_0| + |A_1| + |A_2| - |A_0 \cap A_1| - |A_0 \cap A_2| - |A_1 \cap A_2| \\
&\qquad + |A_0 \cap A_1 \cap A_2| \\
&= 9^4 + 2 \cdot 8 \cdot 9^3 - 2 \cdot 8^4 - 7 \cdot 8^3 + 7^4 \\
&= 25 \cdot 9^3 - 23 \cdot 8^3 + 7^4 = 8850
\end{aligned}
$$

$$
q = \frac{8850}{9000}, \quad p = 1 - q \quad \approx 0.01667
$$

Example

Previous example generalised: Probability of an r-digit number having all of 0,1,2,3 among its digits.
We use the previous notation: A_i — set of numbers n *missing* digit i, and similarly for all $A_{ij\dots}$
We aim to find the size of $T = A_0 \cup A_1 \cup A_2 \cup A_3$, and then to compute $|S| = 9 \cdot 10^{r-1} - |T|$.

$$\begin{aligned} |A_0 \cup A_1 \cup A_2 \cup A_3| &= \text{sum of } |A_i| \\ &- \text{sum of } |A_i \cap A_j| \\ &+ \text{sum of } |A_i \cap A_j \cap A_k| \\ &- \text{sum of } |A_i \cap A_j \cap A_k \cap A_l| \end{aligned}$$

Exercises

$\boxed{\text{RW: 5.6.38}}$ (Supp) Of 100 problems, 75 are 'easy' and 40 'important'.

(b) n problems chosen randomly. What is the probability that all n are important?

Exercises

$\boxed{\text{RW: 5.6.38}}$ (Supp) Of 100 problems, 75 are 'easy' and 40 'important'.

(b) n problems chosen randomly. What is the probability that all n are important?

$$p = \frac{\binom{40}{n}}{\binom{100}{n}} = \frac{40 \cdot 39 \cdots (41 - n)}{100 \cdot 99 \cdots (101 - n)}$$

Exercises

Exercises

$\boxed{\text{RW: 5.2.3}}$ A 4-letter word is selected at random from Σ^4, where $\Sigma = \{a, b, c, d, e\}$. What is the probability that

(a) the letters in the word are all distinct?

(b) there are no vowels ("a", "e") in the word?

(c) the word begins with a vowel?

Exercises

$\boxed{\text{RW: 5.2.3}}$ A 4-letter word is selected at random from Σ^4, where $\Sigma = \{a, b, c, d, e\}$. What is the probability that

(a) the letters in the word are all distinct?

$$|E| = (5)_4, \quad P(E) = \frac{5 \cdot 4 \cdot 3 \cdot 2}{5^4} = \frac{120}{625} \approx 19\%$$

(b) there are no vowels (“a”, “e”) in the word?

$$|E| = 3^4, \quad P(E) = \frac{3^4}{5^4} = \frac{81}{625} \approx 13\%$$

(c) the word begins with a vowel?

$$|E| = 2 \cdot 5^3, \quad P(E) = \frac{2 \cdot 5^3}{5^4} = \frac{2}{5}$$

Outline

Unifying sets of outcomes

To combine events from different sets of outcomes we unify the sample space using the **product space**: $\Omega_1 \times \Omega_2 \times \ldots \times \Omega_n$.

Example

Flipping a coin and rolling a die:

$$\Omega_1 = \{\text{heads}, \text{tails}\} \qquad \Omega_2 = \{1, 2, 3, 4, 5, 6\}$$

$$\Omega = \Omega_1 \times \Omega_2 = \{(\text{heads}, 1), (\text{heads}, 2), \ldots\}$$

NB

This approach can also be used to model sequences of outcomes.

Events in the product space

Events are lifted into the product space by restricting the appropriate co-ordinate. E.g. $A \subseteq \Omega_1$ translates to $A' = A \times \Omega_2 \times \ldots \times \Omega_n$.

Example

Coin shows heads and die shows an even number:

$$\Omega_1 = \{\text{heads}, \text{tails}\} \qquad A = \{\text{heads}\}$$
$$\Omega_2 = \{1, 2, 3, 4, 5, 6\} \qquad B = \{2, 4, 6\}$$

$$\Omega = \Omega_1 \times \Omega_2 = \{(\text{heads}, 1), (\text{heads}, 2), \ldots\}$$
$$A' = A \times \Omega_2 \qquad B' = \Omega_1 \times B$$

"A and B" or "A followed by B" corresponds to:

$$A' \cap B' = (A \times \Omega_2) \cap (\Omega_1 \times B) = A \times B$$

Probability in the product space

NB

Cannot assume that $P(A \times B) = P(A)P(B)$

Example

Toss two coins.

- A: First coin shows heads
- B: Both coins show tails

$$\begin{array}{ll}
\Omega_1 = \{H, T\} & \Omega_2 = \{HH, HT, TH, TT\} \\
A = \{H\} & A' = \{(H, HH), (H, HT), (H, TH), (H, TT)\} \\
B = \{TT\} & B' = \{(H, TT), (T, TT)\} \\
A' \cap B' = A \times B = \{(H, TT)\} &
\end{array}$$

$$P(A) = \frac{1}{2} \quad P(B) = \frac{1}{4} \quad P(A' \cap B') = 0$$

Product distribution

Given probability distributions on the component spaces, there is a natural probability distribution on the product space:

$$P(E_1 \times E_2 \times \ldots \times E_n) = P_1(E_1) \cdot P_2(E_2) \cdots P_n(E_n)$$

Intuitively, the probability of an event in one dimension is not affected by the outcomes in the other dimensions.

Product distribution

Given probability distributions on the component spaces, there is a natural probability distribution on the product space:

$$P(E_1 \times E_2 \times \ldots \times E_n) = P_1(E_1) \cdot P_2(E_2) \cdots P_n(E_n)$$

Intuitively, the probability of an event in one dimension is not affected by the outcomes in the other dimensions.

Fact

If the P_i are uniform distributions then so is the product distribution.

Independence

Informally, events are *independent* if the outcomes in one do not affect the outcomes in the other.

Independence

Informally, events are *independent* if the outcomes in one do not affect the outcomes in the other.
More generally, we define independence on events of the **same** sample space.

Definition

A and B are **(stochastically) independent** (notation: $A \perp B$) if $P(A \cap B) = P(A) \cdot P(B)$

NB

Informal notion of independence corresponds to the stochastic independence of the "lifted" events A' and B'

Important!

Unless specified otherwise, we assume independence when unifying events (where appropriate).

Independence of multiple events

Independence of $A_1, \ldots, A_n$ ($A_1 \perp A_2 \perp \cdots \perp A_n$)

$$P(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}) = P(A_{i_1}) \cdot P(A_{i_2}) \cdots P(A_{i_k})$$

This is often called (for emphasis) a *full* independence

Independence of multiple events

Independence of $A_1, \dots, A_n$ ($A_1 \perp A_2 \perp \cdots \perp A_n$)

$$P(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}) = P(A_{i_1}) \cdot P(A_{i_2}) \cdots P(A_{i_k})$$

This is often called (for emphasis) a *full* independence

Pairwise independence is a *weaker* concept.

Example

Toss of two coins

$$\left.\begin{array}{l} A = \langle \text{first coin } H \rangle \\ B = \langle \text{second coin } H \rangle \\ C = \langle \text{exactly one } H \rangle \end{array}\right\} \begin{array}{l} P(A) = P(B) = P(C) = \frac{1}{2} \\ P(A \cap B) = P(A \cap C) = P(B \cap C) = \frac{1}{4} \\ \text{However: } P(A \cap B \cap C) = 0 \end{array}$$

One can similarly construct a set of n events where any k of them are independent, while any $k+1$ are dependent (for $k < n$).

Example: Dependent events

Example

Basic non-independent sets of events (assuming non-trivial probabilities)

- $A \subseteq B$
- $A \cap B = \emptyset$
- Any pair of one-point events $A = \{x\}$, $B = \{y\}$:
  either $x = y$ and $A \subseteq B$
  or $x \neq y$ and $A \cap B = \emptyset$

Exercise

$\boxed{\text{RW: 9.1.25}}$ Does $A \perp B \perp C$ imply $(A \cap B) \perp (A \cap C)$?

Exercise

$\boxed{\text{RW: 9.1.25}}$ Does $A \perp B \perp C$ imply $(A \cap B) \perp (A \cap C)$?

No; this is almost never the case.
If somehow $(A \cap B) \perp (A \cap C)$ then it would give

$$P(A \cap B \cap C) = P(A \cap B \cap A \cap C) = P(A \cap B) \cdot P(A \cap C)$$

As A is independent of B and of C it would suggest

$$P(A \cap B \cap C) \stackrel{?}{=} P(A) \cdot P(B) \cdot P(A) \cdot P(C)$$

instead of the correct

$$P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C)$$

Example: Sequences of independent events

Example

Team A has probability $p = 0.5$ of winning a game against B. What is the probability P_p of A winning a best-of-seven match if

a. A already won the first game?
b. A already won the first two games?
c. A already won two out of the first three games?

Example: Sequences of independent events

Example

Team A has probability $p = 0.5$ of winning a game against B.
What is the probability P_p of A winning a best-of-seven match if

a. A already won the first game?
b. A already won the first two games?
c. A already won two out of the first three games?

(a) Sample space S — 6-sequences, formed from wins (W) and losses (L)

$$|S| = 2^6 = 64$$

Favourable sequences F — those with three to six W

$$|F| = \binom{6}{3} + \binom{6}{4} + \binom{6}{5} + \binom{6}{6} = 20 + 15 + 6 + 1 = 42$$

Therefore $P_{0.5} = \frac{42}{64} \approx 66\%$

Example: Sequences of independent events

Example

Team A has probability $p = 0.5$ of winning a game against B.
What is the probability P_p of A winning a best-of-seven match if

a. A already won the first game?
b. A already won the first two games?
c. A already won two out of the first three games?

(b) Sample space S — 5-sequences of W and L

$$|S| = 2^5 = 32$$

Favourable sequences F — those with two to five W

$$|F| = \binom{5}{2} + \binom{5}{3} + \binom{5}{4} + \binom{5}{5} = 10 + 10 + 5 + 1 = 26$$

Therefore $P_{0.5} = \frac{26}{32} \approx 81\%$

Example: Sequences of independent events

Example

Team A has probability $p = 0.5$ of winning a game against B.
What is the probability P_p of A winning a best-of-seven match if

- **a** A already won the first game?
- **b** A already won the first two games?
- **c** A already won two out of the first three games?

(c)

$$|S| = 2^4 = 16$$

$$|F| = \binom{4}{2} + \binom{4}{3} + \binom{4}{4} = 6 + 4 + 1 = 11$$

Therefore $P_{0.5} = \frac{11}{16} \approx 69\%$

Binomial distribution

A useful corollary:

Fact

In a sequence of n independent trials, each with a probability of p of success:

$$P(exactly\ k\ successes) = \binom{n}{k} p^k q^{n-k}$$

where $q = (1 - p)$.

NB

This leads to a probability distribution on sequences of outcomes, known as the **binomial distribution**.

Exercises

Exercise

RW: 5.2.11 Two dice, a red die and a black die, are rolled.
What is the probability that
(a) the sum of the values is even?

(b) the number on the red die is bigger than on the black die?

(c) the number on the black die is twice the one on the red die?

Exercises

Exercise

$\boxed{\text{RW: 5.2.11}}$ Two dice, a red die and a black die, are rolled.
What is the probability that
(a) the sum of the values is even?

$$P\big(R + B \in \{2, 4, \dots, 12\}\big) = \frac{18}{36} = \frac{1}{2}$$

(b) the number on the red die is bigger than on the black die?

(c) the number on the black die is twice the one on the red die?

Exercises

Exercise

$\boxed{\text{RW: 5.2.11}}$ Two dice, a red die and a black die, are rolled.
What is the probability that
(a) the sum of the values is even?

$$P\big(R + B \in \{2, 4, \ldots, 12\}\big) = \frac{18}{36} = \frac{1}{2}$$

(b) the number on the red die is bigger than on the black die?
$P(R > B) = P(R < B)$; also $P(R = B) = \frac{1}{6}$
Therefore $P(R < B) = \frac{1}{2}(1 - P(R = B)) = \frac{5}{12}$

(c) the number on the black die is twice the one on the red die?

Exercises

Exercise

$\boxed{\text{RW: 5.2.11}}$ Two dice, a red die and a black die, are rolled.
What is the probability that
(a) the sum of the values is even?

$$P\big(R + B \in \{2, 4, \ldots, 12\}\big) = \frac{18}{36} = \frac{1}{2}$$

(b) the number on the red die is bigger than on the black die?
$P(R > B) = P(R < B)$; also $P(R = B) = \frac{1}{6}$
Therefore $P(R < B) = \frac{1}{2}(1 - P(R = B)) = \frac{5}{12}$

(c) the number on the black die is twice the one on the red die?

$$P(R = 2 \cdot B) = P(\{(2,1), (4,2), (6,3)\}) = \frac{3}{36} = \frac{1}{12}$$

Exercises

Exercise

$\boxed{\text{RW: 5.2.12}}$ Two dice, a red die and a black die, are rolled.
What is the probability that

(a) the maximum of the numbers is 4?

(b) their minimum is 4?

Exercises

Exercise

RW: 5.2.12 Two dice, a red die and a black die, are rolled. What is the probability that

(a) the maximum of the numbers is 4?

$$P(E_1) = \frac{7}{36}$$

(b) their minimum is 4?

Exercises

Exercise

$\boxed{\text{RW: 5.2.12}}$ Two dice, a red die and a black die, are rolled. What is the probability that

(a) the maximum of the numbers is 4?

$$P(E_1) = \frac{7}{36}$$

(b) their minimum is 4?

$$P(E_2) = \frac{5}{36}$$

Check:

$$P(E_1 \cup E_2) = \frac{7}{36} + \frac{5}{36} - P(E_1 \cap E_2) = \frac{7+5-1}{36} = \frac{11}{36}$$

Exercise

Exercises

$\boxed{\text{RW: 5.2.5}}$ An urn contains 3 red and 4 black balls. 3 balls are removed without replacement. What are the probabilities that
(a) all 3 are red
(b) all 3 are black
(c) one is red, two are black

Exercise

Exercises

$\boxed{\text{RW: 5.2.5}}$ An urn contains 3 red and 4 black balls. 3 balls are removed without replacement. What are the probabilities that
(a) all 3 are red
(b) all 3 are black
(c) one is red, two are black

All probabilities are computed using the same sample space: all possible ways to draw three balls without replacement.

The size of the sample space is $\dfrac{7 \cdot 6 \cdot 5}{3!} = 35$
(a) $E =$ All balls are red: 1 combination
(b) $E =$ All balls are black: $\binom{4}{3} = 4$ combinations
(c) $E =$ One red and two black: $\binom{3}{1} \cdot \binom{4}{2} = 18$ combinations

Outline

Infinite sample spaces

Probability distributions generalize to infinite sample spaces **with some provisos**.

- In continuous spaces (e.g. $\mathbb{R}$):
  - Probability distributions are *measures*;
  - Sums are *integrals*;
  - Non-zero probabilities apply to *ranges*;
  - Probability of a single event is 0.
    Note: Probability 0 is not the same as impossible.
- In discrete spaces (e.g. $\mathbb{N}$):
  - Probability 0 is the same as impossible.
  - No uniform distribution!
  - Non-uniform distributions exist, e.g. $P(0) = 1$, $P(n) = 0$ for $n > 0$; or $P(0) = 0$, $P(n) = \frac{1}{2^n}$ for $n > 0$.
  - May consider limiting probabilities if that makes sense.

Asymptotic Estimate of Relative Probabilities

Example

Event $A \stackrel{\text{def}}{=}$ one die rolled n times and you obtain two 6's
Event $B \stackrel{\text{def}}{=}$ n dice rolled simultaneously and you obtain one 6

$$P(A) = \frac{\binom{n}{2} \cdot 5^{n-2}}{6^n} \qquad P(B) = \frac{\binom{n}{1} \cdot 5^{n-1}}{6^n}$$

Therefore $\frac{P(A)}{P(B)} = \frac{\binom{n}{2}}{\binom{n}{1}} \cdot \frac{1}{5} = \frac{n(n-1)}{2} \cdot \frac{1}{5n} = \frac{n-1}{10} \in \Theta(n)$

n	1	2	3	4	...	11	...	20	...
$P(A)$	0	$\frac{1}{36}$	$\frac{5}{72}$	$\frac{25}{216}$	...	0.296	...	0.198	...
$P(B)$	$\frac{1}{6}$	$\frac{10}{36}$	$\frac{25}{72}$	$\frac{125}{324}$	...	0.296	...	0.104	...

Outline

Use of Recursion in Probability Computations

Question

Given n tosses of a coin, what is the probability of two HEADS *in a row?*

Use of Recursion in Probability Computations

Question

Given n tosses of a coin, what is the probability of two HEADS *in a row?*

Answer

Recall $N(n)$: the number of sequences without HH.

$N(n) = N(n-1) + N(n-2)$: $N(n) = \text{FIB}(n+1)$

$N(n) \approx \frac{1}{\sqrt{5}} \left(\frac{\sqrt{5}+1}{2}\right)^{n+1} \approx 0.72 \cdot (1.6)^n$

$p_n = 1 - \frac{\text{FIB}(n+1)}{2^n} \approx 1 - 0.72 \cdot (0.8)^n$

Example

Question

Given n tosses, what is the probability q_n of at least one HHH?

$q_0 = q_1 = q_2 = 0; q_3 = \frac{1}{8}$

Then recursive computation:

$$
\begin{aligned}
q_n = & \frac{1}{2} q_{n-1} && \text{(initial: T)} \\
& + \frac{1}{4} q_{n-2} && \text{(initial: HT)} \\
& + \frac{1}{8} q_{n-3} && \text{(initial: HHT)} \\
& + \frac{1}{8} && \text{(start with: HHH)}
\end{aligned}
$$

Example

Question

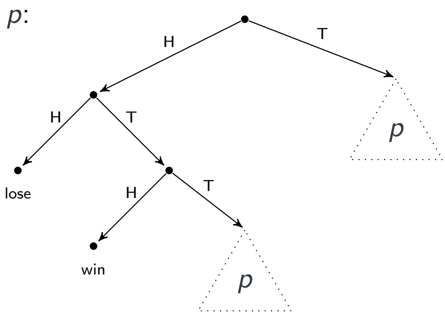
A coin is tossed 'indefinitely'. Which pattern is more likely (and by how much) to appear first, HTH or HHT?

Example

Question

A coin is tossed 'indefinitely'. Which pattern is more likely (and by how much) to appear first, HTH or HHT?

let $p = P(HTH \text{ first})$

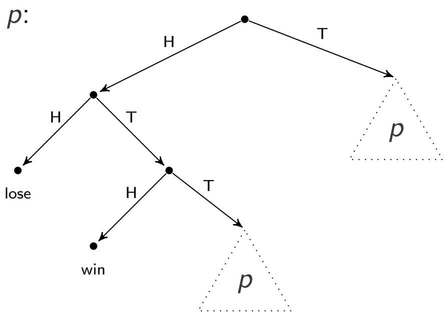


Example

Question

A coin is tossed 'indefinitely'. Which pattern is more likely (and by how much) to appear first, HTH or HHT?

let $p = P(HTH \text{ first})$



$$p = \frac{1}{8} + \frac{1}{8}p + \frac{1}{2}p \implies \frac{3}{8}p = \frac{1}{8} \implies p = \frac{1}{3}$$

Difficult probability calculations

NB

The majority of problems in probability and statistics do not have such elegant solutions. Hence the use of computers for either precise calculations or approximate simulations is mandatory. However, it is the use of recursion that simplifies such computing or, quite often, makes it possible in the first place.

Outline

Conditional probability

Definition

Conditional probability of E **given** S:

$$P(E|S) = \frac{P(E \cap S)}{P(S)}, \quad E, S \subseteq \Omega$$

It is defined only when $P(S) \neq 0$

NB

$P(A|B)$ and $P(B|A)$ are, in general, not related — one of these values predicts, by itself, essentially nothing about the other.
The only exception, applicable when $P(A), P(B) \neq 0$, is that $P(A|B) = 0$ iff $P(B|A) = 0$ iff $P(A \cap B) = 0$.

E.g. let A denote the event that a person is a student, and B denote the event that a person studies at UNSW. Then $P(A|B) \approx 1$ but $P(B|A) \approx 0$.

Conditional probability in uniform distributions

If P is the uniform distribution over a finite set Ω, then

$$P(E|S) = \frac{\frac{|E \cap S|}{|\Omega|}}{\frac{|S|}{|\Omega|}} = \frac{|E \cap S|}{|S|}$$

This observation can help in calculations...

Example

$\boxed{\text{RW: 9.1.6}}$ A coin is tossed four times. What is the probability of
(a) two consecutive HEADS
(b) two consecutive HEADS *given* that ≥ 2 tosses are HEADS

Conditional probability in uniform distributions

If P is the uniform distribution over a finite set Ω, then

$$P(E|S) = \frac{\frac{|E \cap S|}{|\Omega|}}{\frac{|S|}{|\Omega|}} = \frac{|E \cap S|}{|S|}$$

This observation can help in calculations...

Example

$\boxed{\text{RW: 9.1.6}}$ A coin is tossed four times. What is the probability of
(a) two consecutive HEADS
(b) two consecutive HEADS *given* that ≥ 2 tosses are HEADS

T	T	T	T	H	T	T	T
T	T	T	H	H	T	T	H
T	T	H	T	H	T	H	T
T	T	H	H	H	T	H	H
T	H	T	T	H	H	T	T
T	H	T	H	H	H	T	H
T	H	H	T	H	H	H	T
T	H	H	H	H	H	H	H

Conditional probability in uniform distributions

If P is the uniform distribution over a finite set Ω, then

$$P(E|S) = \frac{\frac{|E \cap S|}{|\Omega|}}{\frac{|S|}{|\Omega|}} = \frac{|E \cap S|}{|S|}$$

This observation can help in calculations...

Example

$\boxed{\text{RW: 9.1.6}}$ A coin is tossed four times. What is the probability of
(a) two consecutive HEADS
(b) two consecutive HEADS *given* that ≥ 2 tosses are HEADS

T	T	T	T	H	T	T	T
T	T	T	H	H	T	T	H
T	T	H	T	H	T	H	T
T	T	H	H	H	T	H	H
T	H	T	T	H	H	T	T
T	H	T	H	H	H	T	H
T	H	H	T	H	H	H	T
T	H	H	H	H	H	H	H

(a) $\frac{8}{16}$

Conditional probability in uniform distributions

If P is the uniform distribution over a finite set Ω, then

$$P(E|S) = \frac{\frac{|E \cap S|}{|\Omega|}}{\frac{|S|}{|\Omega|}} = \frac{|E \cap S|}{|S|}$$

This observation can help in calculations...

Example

$\boxed{\text{RW: 9.1.6}}$ A coin is tossed four times. What is the probability of
(a) two consecutive HEADS
(b) two consecutive HEADS *given* that ≥ 2 tosses are HEADS

				H	T	T	H
				H	T	H	T
T	T	H	H	H	T	H	H
				H	H	T	T
T	H	T	H	H	H	T	H
T	H	H	T	H	H	H	T
T	H	H	H	H	H	H	H

(a) $\frac{8}{16}$ (b) $\frac{8}{11}$

Some General Rules

Fact

- $A \subseteq B \to P(A|B) \geq P(A)$
- $A \subseteq B \to P(B|A) = 1$
- $P(A \cap B|B) = P(A|B)$
- $P(\emptyset|A) = 0$ *for* $A \neq \emptyset$
- $P(A|\Omega) = P(A)$
- $P(A^c|B) = 1 - P(A|B)$

NB

- $P(A|B)$ *and* $P(A|B^c)$ *are not related*
- $P(A|B), P(B|A), P(A^c|B^c), P(B^c|A^c)$ *are not related*

Example

Example

Two dice are rolled and the outcomes recorded as b for the black die, r for the red die and $s = b + r$ for their total.
Define the events $B = \{b \geq 3\},\ R = \{r \geq 3\},\ S = \{s \geq 6\}$.

$$P(S|B) = \frac{4+5+6+6}{24} = \frac{21}{24} = \frac{7}{8} = 87.5\%$$

$$P(B|S) = \frac{4+5+6+6}{26} = \frac{21}{26} = 80.8\%$$

The (common) numerator $4+5+6+6 = 21$ represents the size of the $B \cap S$ — the common part of B and S, that is, the number of rolls where $b \geq 3$ and $s \geq 6$. It is obtained by considering the different cases: $b = 3$ and $s \geq 6$, then $b = 4$ and $s \geq 6$ etc.

The denominators are $|B| = 24$ and $|S| = 26$

Example

Example (cont'd)

Recall: $B = \{b \geq 3\},\ R = \{r \geq 3\},\ S = \{s \geq 6\}$

$$P(B) = P(R) = 2/3 = 66.7\%$$

$$P(S) = \frac{5+6+5+4+3+2+1}{36} = \frac{26}{36} = 72.22\%$$

$$P(S|B \cup R) = \frac{2+3+4+5+6+6}{32} = \frac{26}{32} = 81.25\%$$

The set $B \cup R$ represents the event 'b or r'.
It comprises all the rolls except for those with *both* the red and the black die coming up either 1 or 2.

Example

Example (cont'd)

Recall: $B = \{b \geq 3\},\ R = \{r \geq 3\},\ S = \{s \geq 6\}$

$$P(B) = P(R) = 2/3 = 66.7\%$$

$$P(S) = \frac{5+6+5+4+3+2+1}{36} = \frac{26}{36} = 72.22\%$$

$$P(S|B \cup R) = \frac{2+3+4+5+6+6}{32} = \frac{26}{32} = 81.25\%$$

The set $B \cup R$ represents the event 'b or r'.
It comprises all the rolls except for those with *both* the red and the black die coming up either 1 or 2.

$$P(S|B \cap R) = 1 = 100\% \text{ — because } S \supseteq B \cap R$$

Exercise

$\boxed{\text{RW: 9.1.9}}$ Consider three red and eight black marbles; draw two without replacement. We write b_1 — Black on the first draw, b_2 — Black on the second draw, r_1 — Red on first draw, r_2 — Red on second draw

Find the probabilities

(a) both Red:

Exercise

Exercise

$\boxed{\text{RW: 9.1.9}}$ Consider three red and eight black marbles; draw two without replacement. We write b_1 — Black on the first draw, b_2 — Black on the second draw, r_1 — Red on first draw, r_2 — Red on second draw

Find the probabilities

(a) both Red:

$$P(r_1 \wedge r_2) = P(r_1)P(r_2|r_1) = \frac{3}{11} \cdot \frac{2}{10} = \frac{3}{55}$$

Equivalently:

$|\text{two-samples}| = \binom{11}{2} = 55$; $|\text{Red two-samples}| = \binom{3}{2} = 3$

$P(\cdot) = \frac{\binom{3}{2}}{\binom{11}{2}} = \frac{3}{55}$

Exercise

(b) both Black:

(c) one Red, one Black:

Exercise

(b) both Black:

$$P(b_1 \wedge b_2) = P(b_1)P(b_2|b_1) = \frac{8}{11} \cdot \frac{7}{10} = \frac{28}{55} = \frac{\binom{8}{2}}{\binom{11}{2}}$$

(c) one Red, one Black:

$$P(r_1 \wedge b_2) + P(b_1 \wedge r_2) = \frac{3 \cdot 8}{\binom{11}{2}} \quad \text{— why?}$$

By textbook (the 'hard way')

$$P(r_1 \wedge b_2) + P(b_1 \wedge r_2) = \frac{3}{11} \cdot \frac{8}{10} + \frac{8}{11} \cdot \frac{3}{10}$$

or

$$P(\cdot) = 1 - P(r_1 \wedge r_2) - P(b_1 \wedge b_2) = \frac{55 - 3 - 28}{55}$$

Exercise

Exercise

$\boxed{\text{RW: 9.1.12}}$ What is the probability of a flush given that all five cards in a Poker hand are red?

Exercise

Exercise

$\boxed{\text{RW: 9.1.12}}$ What is the probability of a flush given that all five cards in a Poker hand are red?

Red cards $= \diamondsuit$'s $+ \heartsuit$'s
flush $=$ all cards of the same suit

$$P(\text{flush} \mid \text{all five cards are Red}) = \frac{2 \cdot \binom{13}{5}}{\binom{26}{5}} = \frac{9}{230} \approx 4\%$$

Exercise

Exercise

$\boxed{\text{RW: 9.1.22}}$ Prove the following:
If $P(A|B) > P(A)$ ("positive correlation") then $P(B|A) > P(B)$

Exercise

Exercise

$\boxed{\text{RW: 9.1.22}}$ Prove the following:
If $P(A|B) > P(A)$ ("positive correlation") then $P(B|A) > P(B)$

$$
\begin{aligned}
P(A|B) &> P(A) \\
\therefore\ P(A \cap B) &> P(A)P(B) \\
\therefore\ \frac{P(A \cap B)}{P(A)} &> P(B) \\
\therefore\ P(B|A) &> P(B)
\end{aligned}
$$

Outline

Stochastic Independence, again

Definition

A and B are **(stochastically) independent** (notation: $A \perp B$) if $P(A \cap B) = P(A) \cdot P(B)$

If $P(A) \neq 0$ and $P(B) \neq 0$, all of the following are *equivalent* definitions:

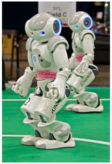
- $P(A \cap B) = P(A)P(B)$
- $P(A|B) = P(A)$
- $P(B|A) = P(B)$
- $P(A^c|B) = P(A^c)$ or $P(A|B^c) = P(A)$ or $P(A^c|B^c) = P(A^c)$

The last one claims that

$$A \perp B \leftrightarrow A^c \perp B \leftrightarrow A \perp B^c \leftrightarrow A^c \perp B^c$$

Using independence to simplify calculations

Independence of events, even just pairwise independence, can greatly simplify computations and reasoning in AI applications. It is common for many expert systems to make an approximating assumption of independence, even if it is not completely satisfied.



$$P(\mathsf{sense}_t \,|\, \mathsf{loc}_t, \mathsf{sense}_{t-1}, \mathsf{loc}_{t-1}, \ldots) = P(\mathsf{sense}_t \,|\, \mathsf{loc}_t)$$

Exercise

Exercise

$\boxed{\text{RW: 9.1.7}}$ Suppose that an experiment leads to events A, B and C with $P(A) = 0.3$, $P(B) = 0.4$ and $P(A \cap B) = 0.1$

(a) $P(A|B) =$

(b) $P(A^c) =$

(c) Is $A \perp B$?

(d) Is $A^c \perp B$?

Exercise

Exercise

$\boxed{\text{RW: 9.1.7}}$ Suppose that an experiment leads to events A, B and C with $P(A) = 0.3$, $P(B) = 0.4$ and $P(A \cap B) = 0.1$

(a) $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1}{4}$

(b) $P(A^c) = 1 - P(A) = 0.7$

(c) Is $A \perp B$? No. $P(A) \cdot P(B) = 0.12 \neq P(A \cap B)$

(d) Is $A^c \perp B$? No, as can be seen from (c).

Note: $P(A^c \cap B) = P(B) - P(A \cap B) = 0.4 - 0.1 = 0.3$
$P(A^c) \cdot P(B) = 0.7 \cdot 0.4 = 0.28$

Exercise

Exercise

$\boxed{\text{RW: 9.1.8}}$ Given $A \perp B,\ P(A) = 0.4,\ P(B) = 0.6$

$P(A|B) =$

$P(A \cup B) =$

$P(A^c \cap B) =$

Exercise

Exercise

$\boxed{\text{RW: 9.1.8}}$ Given $A \perp B,\ P(A) = 0.4,\ P(B) = 0.6$

$$P(A|B) = P(A) = 0.4$$

$$P(A \cup B) = P(A) + P(B) - P(A)P(B) = 0.76$$

$$P(A^c \cap B) = P(A^c)P(B) = 0.36$$

Supplementary Exercise

Exercise

$\boxed{\text{RW: 9.5.5}}$ (Supp) We are given two events with
$P(A) = \frac{1}{4},\ P(B) = \frac{1}{3}$.
True, false or could be either?

- **a** $P(A \cap B) = \frac{1}{12}$
- **b** $P(A \cup B) = \frac{7}{12}$
- **c** $P(B|A) = \frac{P(B)}{P(A)}$
- **d** $P(A|B) \geq P(A)$
- **e** $P(A^c) = \frac{3}{4}$
- **f** $P(A) = P(B)P(A|B) + P(B^c)P(A|B^c)$

Supplementary Exercise

Exercise

$\boxed{\text{RW: 9.5.5}}$ (Supp) We are given two events with
$P(A) = \frac{1}{4},\ P(B) = \frac{1}{3}$.
True, false or could be either?

- **a** $P(A \cap B) = \frac{1}{12}$ — possible; it holds when $A \perp B$
- **b** $P(A \cup B) = \frac{7}{12}$ — possible; it holds when A, B are disjoint
- **c** $P(B|A) = \frac{P(B)}{P(A)}$ — false; correct is: $P(B|A) = \frac{P(B \cap A)}{P(A)}$
- **d** $P(A|B) \geq P(A)$ — possible (it means that B "supports" A)
- **e** $P(A^c) = \frac{3}{4}$ — true, since $P(A^c) = 1 - P(A)$
- **f** $P(A) = P(B)P(A|B) + P(B^c)P(A|B^c)$
  — true (also known as *total probability*)