COMP9020 22T1
Week 9
Conditional Probability

NB

- Public holiday on Friday
- Last regular lecture on Wednesday Week 10
- Course review, Fun quiz on Friday Week 10

- Textbook (R & W) - Ch. 9, Sec. 9.1-9.4
- Problem set week 9 + last quiz

Assessment

Assessment isn't a "one-way street" ...

- I get to assess you in the final exam
- you get to assess me in UNSW's MyExperience Evaluation
    - go to `https://myexperience.unsw.edu.au/`
    - log on using zID@ad.unsw.edu.au and your zPass
    - or follow the direct link when on Moodle

Please fill it out ...

- give me some feedback on how you'd envision the course contents and delivery – lectures, problem sets, quizzes – in the future
- even if that is "Exactly the same. I liked how it was run."

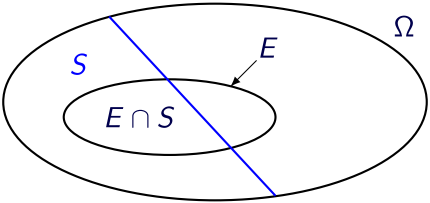
Conditional Probability

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 $P(S) \neq 0$



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 $P(A \cap B) = 0$.

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

$\boxed{9.1.6}$ 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) $\frac{8}{16}$ (b) $\frac{8}{11}$

Exercise

$\boxed{9.1.12}$ What is the probability of a flush ($\neq$ straight/royal 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, not of sequential rank

$$P(\text{flush} \mid \text{all five cards are Red}) = \frac{2 \cdot \left(\binom{13}{5} - 10 \right)}{\binom{26}{5}} \approx 4\%$$

Exercise

$\boxed{9.1.12}$ What is the probability of a flush ($\neq$ straight/royal 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, not of sequential rank

$$P(\text{flush} \,|\, \text{all five cards are Red}) = \frac{2 \cdot \left(\binom{13}{5} - 10\right)}{\binom{26}{5}} \approx 4\%$$

Some General Rules

Fact

- $A \subseteq B \Rightarrow P(A|B) \geq P(A)$
- $A \subseteq B \Rightarrow P(B|A) = 1$
- $P(A \cap B|B) = P(A|B)$
- $P(\emptyset|A) = 0 \text{ for } A \neq \emptyset$
- $P(A|\Omega) = P(A)$
- $P(A^c|B) = 1 - P(A|B)$

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

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 $|S| = 26$

NB

Bayes' Formula: $P(S|B) \cdot P(B) = P(B|S) \cdot P(S)$

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.

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

Exercise

$\boxed{9.1.9}$ 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

Using conditional probabilities, 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:

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

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

(b) both Black:

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

Exercise

$\boxed{9.1.9}$ 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

Using conditional probabilities, 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:

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

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

(b) both Black:

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

Exercise

(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

(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

$\boxed{9.1.22}$ Prove the following:

If $P(A|B) > P(A)$ ("positive correlation") then $P(B|A) > P(B)$

$P(A|B) > P(A)$

$\Rightarrow P(A \cap B) > P(A) \cdot P(B)$

$\Rightarrow \frac{P(A \cap B)}{P(A)} > P(B)$

$\Rightarrow P(B|A) > P(B)$

Exercise

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

$P(A|B) > P(A)$

$\Rightarrow P(A \cap B) > P(A) \cdot P(B)$

$\Rightarrow \frac{P(A \cap B)}{P(A)} > P(B)$

$\Rightarrow P(B|A) > P(B)$

Stochastic Independence

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*:

- $P(A \cap B) = P(A)P(B)$
- $P(A|B) = P(A)$ (i.e. B does not affect the probability of A)
- $P(B|A) = P(B)$ (i.e. A does not affect the probability of 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 means that

$$A \perp B \Leftrightarrow A^c \perp B \Leftrightarrow A \perp B^c \Leftrightarrow A^c \perp B^c$$

Basic non-independent sets of events (if $P(A), P(B) > 0$)

- $A \subseteq B$
- $A \cap B = \emptyset$
- Any pair of one-point events $\{x\}, \{y\}$:
  either $x = y$ and $P(x|y) = 1$
  or $x \neq y$ and $P(x|y) = 0$

Independence of $A_1, \ldots, A_n$

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

for all possible collections $A_{i_1}, A_{i_2}, \ldots, A_{i_k}$.
This is often called (for emphasis) a *full* independence

Basic non-independent sets of events (if $P(A), P(B) > 0$)

- $A \subseteq B$
- $A \cap B = \emptyset$
- Any pair of one-point events $\{x\}, \{y\}$:
  either $x = y$ and $P(x|y) = 1$
  or $x \neq y$ and $P(x|y) = 0$

Independence of $A_1, \ldots, A_n$

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

for all possible collections $A_{i_1}, A_{i_2}, \ldots, A_{i_k}$.
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 $k < n$).

NB

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.

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

Exercise

$\boxed{9.1.7}$ 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

$\boxed{9.1.7}$ 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

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

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

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

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

Exercise

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

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

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

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

Exercise

$\boxed{\text{9.5.5 (supp)}}$ 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) $P(A) = P(A|B)P(B) + P(A|B^c)P(B^c)$ — true

NB

Total probability: $P(A) = P(A|B)P(B) + P(A|B^c)P(B^c)$

Exercise

$\boxed{\text{9.5.5 (supp)}}$ 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) $P(A) = P(A|B)P(B) + P(A|B^c)P(B^c)$ — true

NB

Total probability: $P(A) = P(A|B)P(B) + P(A|B^c)P(B^c)$

Summary

- Conditional probability $P(A|B)$
- Independence $A \perp B$
- Bayes' formula, total probability

Coming up ...

- Expectations
- Course review