Lecture 3:

Conditional probability

Thinking conditionally

Definition and intuition

Definition 2.2.1 (Conditional probability). If A and B are events with $P(B) > 0$, then the **conditional probability** of A **given** B:

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

$P(A)$ – **prior** probability of A, $P(A \mid B)$ – **posterior** probability of A.

(Posterior = updated based on evidence, prior = before this update)

For any event A, $P(A \mid A) = P(A \cap A)/P(A) = 1$ – if A occurred, our updated probability for A is 1.

Definition and intuition

Example 2.2.2 (Two cards) Two cards are picked from a well-shuffled standard deck. Let A = first card is a ♥ (#♥=13), and B = second card is red (#red=26). **Find:** $P(A\,|\,B)$ and $P(B\,|\,A)$.

Solution: $P(A \cap B) = \dfrac{13 \cdot 25}{52 \cdot 51} = \dfrac{25}{204}$. Also $P(A) = 1/4$ and $P(B) = \dfrac{26 \cdot 51}{52 \cdot 51} = \dfrac{1}{2}$ (second is red, first is **any** of the rest 51)

So, $P(A\,|\,B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{25/204}{1/2} = \dfrac{25}{102}$

and $P(B\,|\,A) = \dfrac{P(B \cap A)}{P(A)} = \dfrac{25/204}{1/4} = \dfrac{25}{51}$

Definition and intuition

Some observations from this example:

1. **Order** matters – in general, $P(A \mid B) \neq P(B \mid A)$. Confusing these two is called the ***prosecutor's fallacy*** (we'll discuss that later).

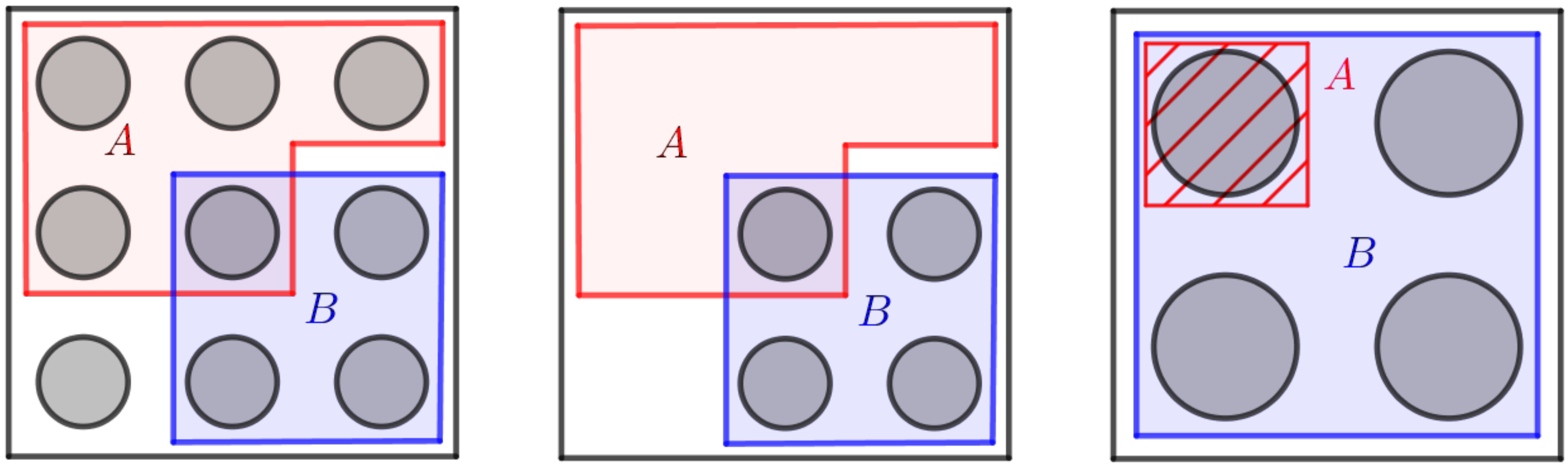
2. Both $P(A \mid B)$ and $P(B \mid A)$ make sense – **information** observing one provides about another, but **not if one causes another!**

3. Obtaining $P(B \mid A) = 25/51$ could be easier: if 1st is a ♥, there are 25 red + 26 black left, so $P(B \mid A) = 25/(25 + 26)$. It's harder to find $P(A \mid B)$ this way (if 1st is red, could be both a ♦ or a ♥).

But we'll figure out general rules for doing that!

Definition and intuition

Intuition 2.2.3 (Pebble world)



1) B occurred, so only take pebbles from B
2) Renormalise, so that total mass is still 1

So $P(A \mid B) = \frac{1}{4}$

Definition and intuition

Intuition 2.2.4 (Frequentist interpretation). We repeat an experiment n (large) times, n_A, n_B, n_{AB} = # of occurrences of $A, B, A \cap B$ resp. Frequentist interpretation is that

$$P(A) \approx \frac{n_A}{n},\ P(B) \approx \frac{n_B}{n},\ P(A \cap B) \approx \frac{n_{AB}}{n}$$

Then $P(A \mid B) = \dfrac{n_{AB}}{n_B}$, which equals $\dfrac{n_{AB}/n}{n_B/n} = \dfrac{P(A \cap B)}{P(B)}$

Definition and intuition

Example 2.2.5 (Two children) Martin Gardner once posed such a puzzle in his column in Scientific American:

> 1) Mr. Jones has two children. The older child is a girl. What's the probability that both children are girls?
>
> 2) Mr. Smith has two children. At least one of them is a boy. What's the probability that both children are boys?

Gardner gave the answers $1/2$ and $1/3$ respectively.

Assumptions are that 1) gender is binary, 2) $P(\text{boy}) = P(\text{girl})$ and 3) genders of two children are **independent**

Definition and intuition

Example 2.2.5

1) Mr. Jones has two children. The older child is a girl. What's the probability that both children are girls?

$$P(\text{both girls} \,|\, \text{elder is girl}) = \frac{P(\text{both girls, elder is girl})}{P(\text{elder is a girl})} = \frac{1/4}{1/2} = \frac{1}{2}$$

2) Mr. Smith has two children. At least one of them is a boy. What's the probability that both children are boys?

$$P(\text{both boys} \,|\, \text{at least one boy}) = \frac{P(\text{both boys, at least one boy})}{P(\text{at least one boy})} = \frac{1/4}{3/4} = \frac{1}{3}$$

Bayes' rule,

Law of total probability (LOTP)

Bayes' rule and LOTP

Theorem 2.3.1 (Probability of intersection). For any two events A and B with positive probabilities,

$$P(A \cap B) = P(B)\,P(A\,|\,B) = P(A)\,P(B\,|\,A)$$

Applying this repeatedly, we get (actually, $n!$ such statements):

Theorem 2.3.2 (Probability of intersection of n events). For any events $A_1, \ldots, A_n$ with $P(A_1, A_2, \ldots, A_{n-1}) > 0$,

$$P(A_1, A_2, \ldots, A_n) = P(A_1)\,P(A_2\,|\,A_1)\,P(A_3\,|\,A_1, A_2)\ldots P(A_n\,|\,A_1, \ldots, A_{n-1})$$

where , $\equiv \cap$, e.g. $P(A_3\,|\,A_1, A_2) \equiv P(A_3\,|\,A_1 \cap A_2)$.

Bayes' rule and LOTP

Theorem 2.3.3 (Bayes' rule)

$$P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B)}$$

A very important statement in probability & statistics.

It helps you find $P(A \mid B)$ from $P(B \mid A)$ (it is often the case that one is **easier** to find than the other!)

Bayes' rule and LOTP

Theorem 2.3.3 (Bayes' rule)

$$P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B)}$$

Definition 2.3.4 (Odds). The **odds** of event A are

odds$(A) = \frac{P(A)}{P(A^c)}$ – e.g. if $P(A) = 2/3$, we say the odds are 2 to 1.

Conversely, $P(A) =$ odds$(A)/(1 +$ odds$(A))$.

Bayes' rule and LOTP

Theorem 2.3.5 (Odds form of Bayes' rule). For any events A and B with positive probabilities, the odds of A after conditioning on B are:

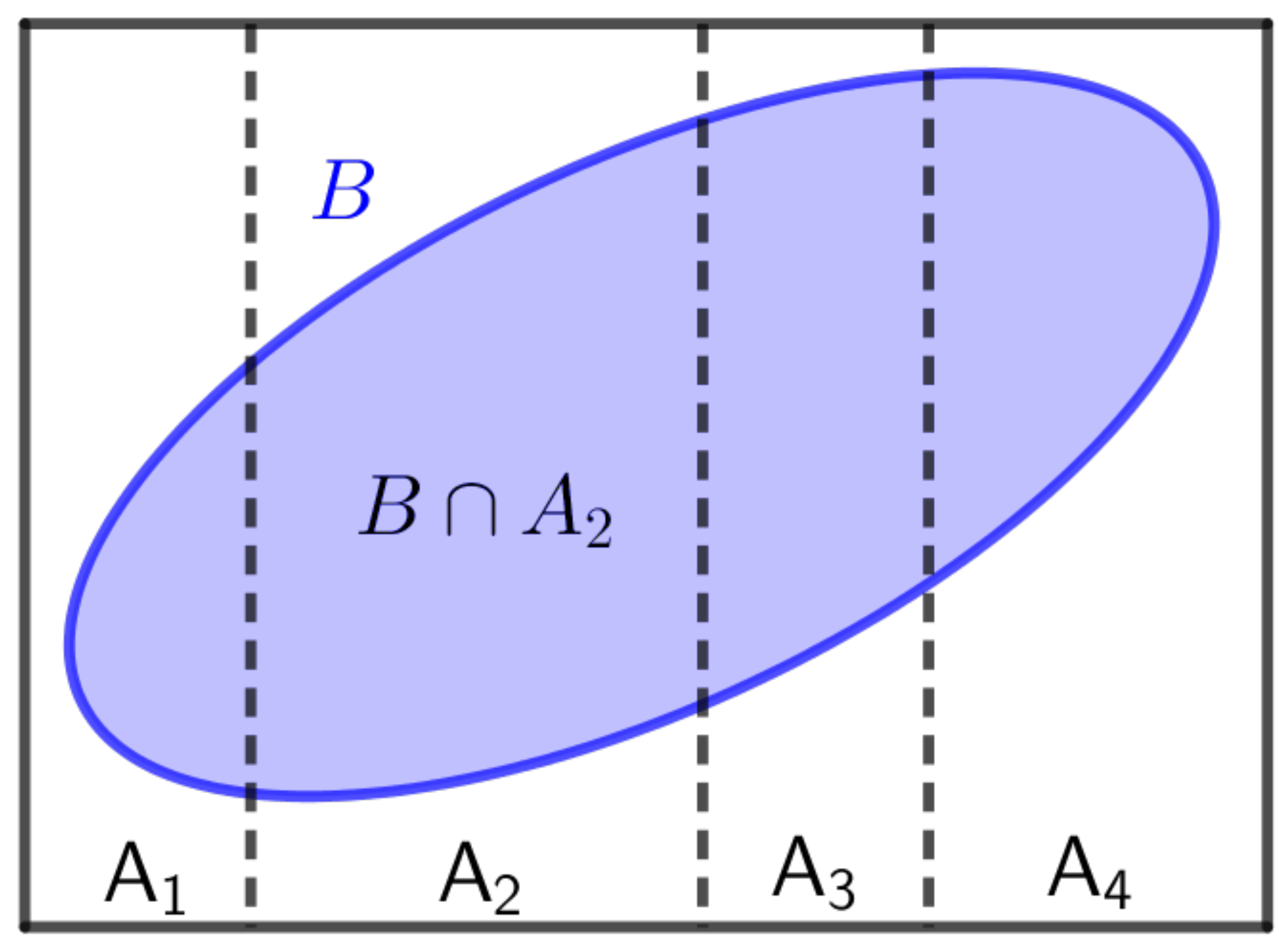
$$\frac{P(A|B)}{P(A^c|B)} = \frac{P(B|A)}{P(B|A^c)}\frac{P(A)}{P(A^c)}$$

That is: **posterior odds** $\frac{P(A|B)}{P(A^c|B)}$ equal to **prior odds** $\frac{P(A)}{P(A^c)}$ times the factor $\frac{P(B|A)}{P(B|A^c)}$ – called the **likelihood ratio**

Bayes' rule and LOTP

Theorem 2.3.6 (Law of total probability, LOTP). Let $A_1, \ldots, A_n$ be a partition of the sample space S – that is, 1) A_i are disjoint, 2) their union is S. Then, if $P(A_i) > 0$ for all i,

$$P(B) = \sum_{i=1}^{n} P(B \,|\, A_i)\, P(A_i)$$



Bayes' rule and LOTP

Theorem 2.3.6 (Law of total probability, LOTP). Let $A_1, \ldots, A_n$ be a partition of the sample space S – that is, 1) A_i are disjoint, 2) their union is S. Then, if $P(A_i) > 0$ for all i,

$$P(B) = \sum_{i=1}^{n} P(B \mid A_i)\, P(A_i)$$

Proof: 1) A_i-s – partition, so: $B = (B \cap A_1) \cup \ldots \cup (B \cap A_n)$

2) so by the second axiom, $P(B) = P(B \cap A_1) + \ldots + P(B \cap A_n)$

so, from $P(B \cap A_i) = P(B \mid A_i)\, P(A_i)$ – Q.E.D.

Bayes' rule and LOTP

Example 2.3.9 (Testing for a disease). A patient is tested for a disease that affects 1% of the population. Suppose the test is "95% accurate" – for now that means that: $P(T | D) = 0.95$ (**true positive** rate of the test), as well as $P(T^c | D^c) = 0.95$ (**true negative**).
Q: Given test is positive, find the probability of the patient being sick.

A: $P(D | T) = \dfrac{P(T | D)\, P(D)}{P(T)} = \dfrac{P(T | D)\, P(D)}{P(T | D)\, P(D) + P(T | D^c)\, P(D^c)} =$

$$= \frac{0.95 \cdot 0.01}{0.95 \cdot 0.01 + 0.05 \cdot 0.99} \approx 0.16$$

There are two factors at play: 1) evidence from the test and 2) **prior information** about the prevalence of the disease

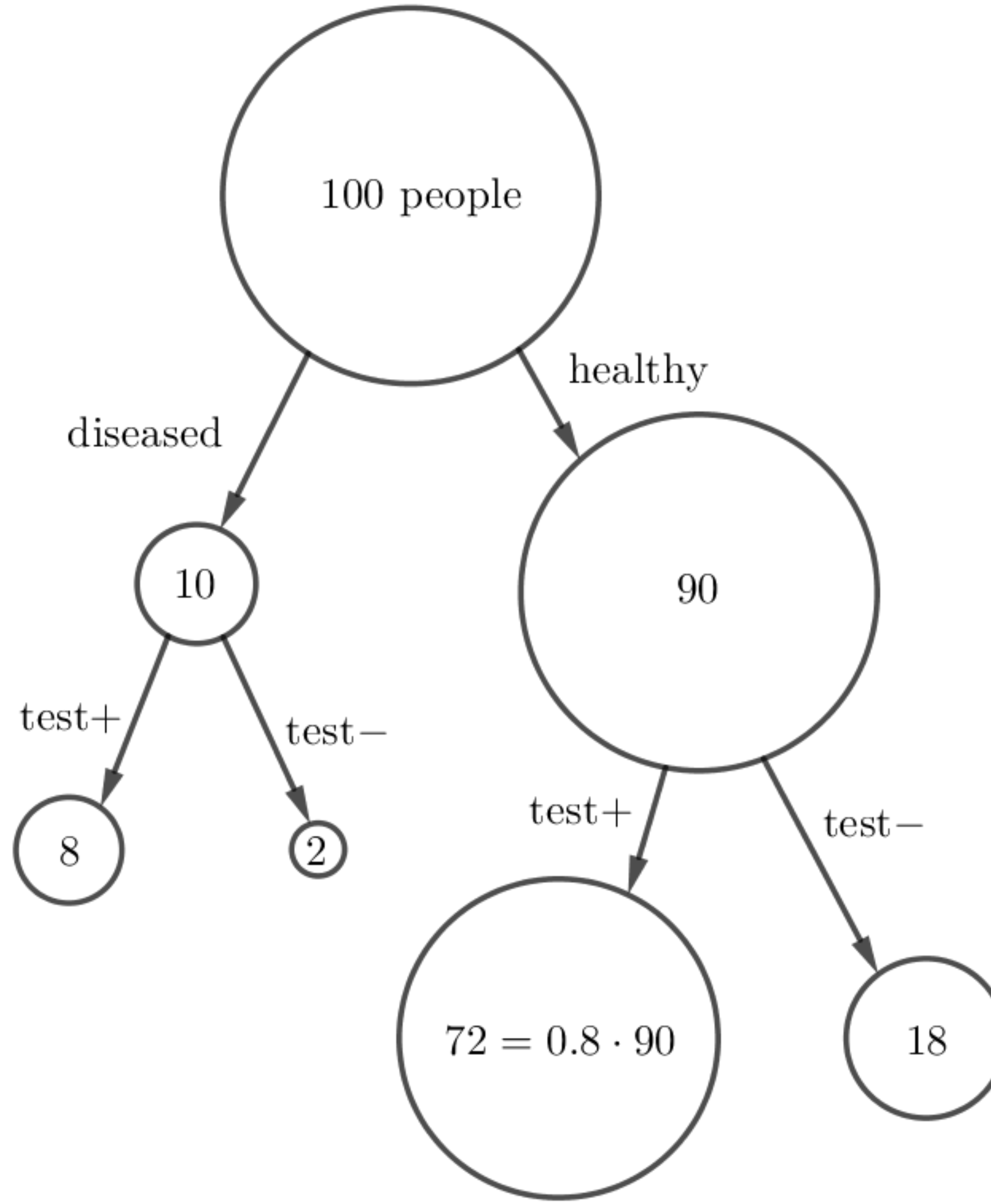
Bayes' rule and LOTP

Example 2.3.9

Here's an example with
$P(T \mid D) = P(T^c \mid D^c) = 0.8$

and $P(D) = 0.1$

In proper scale (area = # of people)



Conditional probabilities are probabilities

Conditional probabilities are probabilities

Conditional probability satisfies all the properties of probability:

- Conditional probabilities are between 0 and 1
- $P(S\,|\,E) = 1$ and $P(\varnothing\,|\,E) = 0$
- If $A_1, A_2, \ldots$ are disjoint, then $P(\cup_{j=1}^{\infty} A_j\,|\,E) = \Sigma_{j=1}^{\infty} P(A_j\,|\,E)$
- $P(A^c\,|\,E) = 1 - P(A\,|\,E)$
- Inclusion-exclusion:

$$P(A \cup B\,|\,E) = P(A\,|\,E) + P(B\,|\,E) - P(A \cap B\,|\,E)$$

Conditional probabilities are probabilities

Important **note**: $P(A \mid E)$ does not mean $A \mid E$ is an event!

Rather, $P(\,\cdot \mid E)$ is a probability function that gives the probability of an event given that E has occurred

$P(\,\cdot\,)$ is a **different** probability function, it gives probabilities without regard of whether E occurred or not

Conditional probabilities are probabilities

Theorem 2.4.2: (Bayes' rule with extra conditioning). Provided that $P(A \cap E) > 0$ and $P(B \cap E) > 0$,

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

Theorem 2.4.2: (LOTP with extra conditioning). Let $A_1, \ldots, A_n$ be a partition of S. Provided that $P(A_i \cap E) > 0$,

$$P(B \,|\, E) = \sum_{i=1}^{n} P(B \,|\, A_i, E)\, P(A_i \,|\, E)$$

Conditional probabilities are probabilities

To summarise,

> *Conditional probabilities are probabilities,*
>
> *and all probabilities are conditional.*

Independence of events

Independence of events

Definition 2.5.1 (Independence of two events). Events A and B are **independent** if: $P(A \cap B) = P(A)\, P(B)$

If $P(A) > 0$ and $P(B) > 0$, this is equivalent to $P(A \,|\, B) = P(A)$ and also to $P(B \,|\, A) = P(B)$.

Note that independence is a ***symmetric relation:*** if A is independent of B, then B is independent of A.

Note that **independence $\neq$ disjointness**! If A and B are disjoint, then $P(A \cap B) = 0$, so disjoint events can be independent only if $P(A) = 0$ or $P(B) = 0$: $P(A \cap B) = 0$ – knowing that A occurred tells us that B definitely didn't occur, so they are not independent (except if A or B have zero probability).

Independence of events

Proposition 2.5.3 If A and B are independent, then A and B^c are independent, A^c and B are independent, and A^c and B^c are too.

Let me leave the proof to you :)

Definition 2.5.4 (Independence of 3 events). Events A, B, C are independent if **all** these hold:

$$P(A \cap B) = P(A)\,P(B), \quad P(A \cap C) = P(A)\,P(C),$$

$$P(B \cap C) = P(B)\,P(C) \quad \text{and} \quad P(A \cap B \cap C) = P(A)\,P(B)\,P(C)$$

Independence of events

Definition 2.5.4 (Independence of 3 events). Events A, B, C are independent if **all** these hold:

$$P(A \cap B) = P(A)\,P(B), \quad P(A \cap C) = P(A)\,P(C),$$

$$P(B \cap C) = P(B)\,P(C) \quad \text{and} \quad P(A \cap B \cap C) = P(A)\,P(B)\,P(C)$$

If the first 3 hold, we say that A, B, C are **pairwise independent**. Pairwise independence does **not** imply independence!

Example 2.5.5: Two coin tosses, A = 1st is H, B = 2nd is H, C = both tosses have same result. Pairwise independent, but $P(A \cap B \cap C) = 1/4$ while $P(A)\,P(B)\,P(C) = 1/8$

Independence of events

Definition 2.5.6 (Independence of many events). For n events $A_1, \ldots, A_n$ to be independent, for all pairs $P(A_i \cap A_j) = P(A_i)\,P(A_j)$, and for all triplets $P(A_i \cap A_j \cap A_k) = P(A_i)\,P(A_j)\,P(A_k)$, and so on.

Definition 2.5.7 (Conditional independence). Events A and B are **conditionally independent** given E if

$$P(A \cap B \,|\, E) = P(A \,|\, E)\,P(B \,|\, E)$$

Conditional independence does **not** imply independence!

Nor does independence imply conditional independence!