Probability - Session 1

Introduction: Definitions and axioms

Elizabeth Williamson
with thanks to Jennifer Rogers

Foundations of Medical Statistics

Important information

- Lectures and practicals:

  Session 1 Thu 26 Sept, 2-5pm, LG9
  Session 2 Mon 30 Sept, 2-5pm, LG81
  Session 3 Tues 1 Oct, 2-5pm, LG81
  Session 4 Mon 7 Oct, 9.30-12.30, JS/SC-3A
  Session 5 Mon 14 Oct, 9.30-12.30, LG7
- Practicals will be in the form of written questions
- Practical facilitator:
  - Tess Poole
- Assignment
  - Hand-out date - Mon 7th October (Probability 4)
  - Hand-in date - Mon 21st October
  - Will be a written assignment, same form as practicals

Overall objectives

By the end of the 5 sessions you should be able to:

- explain basic concepts of probability theory
- draw a probability tree and obtain probabilities from it
- apply Bayes' Theorem to clinical examples
- state the probability distributions for the Normal, Poisson and Binomial distributions
- calculate the expectation and variance for these (and other) distributions

Session objectives

By the end of this session you should be able to:

- explain the basic idea of probability
- state key definitions and axioms in probability theory
- define conditional probability
- apply probability trees to examples
- define independence between events
- apply the theorem of total probability

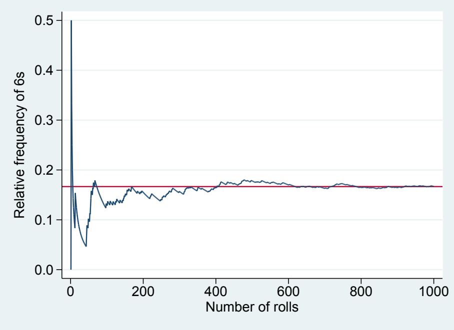
Outline

What is probability?

- We all have some intuitive sense of what probability is.
- For example:
  - When tossing a coin the probability of obtaining a head is 0.5.
  - The probability that it will rain tomorrow in London is 0.8.
- How can we define probability more formally?

Probability as relative frequency

- One way of defining probability is as the relative frequency of an event occurring when the process is repeated many times.
- For example:
  - Suppose we roll a standard die repeatedly, counting the number of 6s



Defining probability

- We can more formally define the probability by:

$$P(\text{roll a six}) = \lim_{n \to \infty} \frac{\text{number of 6s in } n \text{ rolls}}{n}$$

- This definition requires us to imagine an infinite repetition of the same process of experiment.
- How does this work for the example of the probability of rain tomorrow?
- We could say something like 'if you take all the days when I forecast an 80% of rain, the proportion of days when it actually rains will be close to 80%'.

Probability in medical statistics

- Probability is crucial to medical statistics.
- For example:
  - **Predicting events**
    'What is the probability that a particular patient will suffer from heart disease in the next 10 years?'
  - **Assessing whether two characteristics are related**
    'Is LVEF related to systolic blood pressure?'
  - **Quantifying uncertainty around estimates**
    'We estimate that this new drug decreases 10-year mortality by 5%. Can we provide a range of values which captures the uncertainty around this estimate?'

Outline

Example: Asthma and smoking

Research question: What is the prevalence of asthma in smokers and non-smokers?

Study design: Randomly select a number of individuals from the population and record whether they have asthma (A) or not ($\bar{A}$), whether or not they smoke (S or $\bar{S}$), and perhaps other basic information (e.g. age, sex).

(Unknown) true prevalences: Suppose prevalence of smoking is 20% among adults in general and that 9% of smokers suffer asthma, whereas 7% of non-smokers have asthma.

Definitions & notation

Definition	Example
Experiment: A process that produces one outcome from some set of alternatives.	Randomly select an individual from a population. Record asthma and smoking status (A or $\bar{A}$ and S or $\bar{S}$).
Sample space (Ω): The set of points representing all the possible outcomes of an experiment.	$\Omega = \{AS, A\bar{S}, \bar{A}S, \bar{A}\bar{S}\}$
Event: A subset of the sample space.	The event that the selected individual is a smoker: $\{AS, \bar{A}S\}$

Definitions & notation — Exercise

Experiment:

Randomly select an MSc Med Stats student. Ask student how many days a week (on average) they do vigorous exercise.

Sample space:

$\Omega =$

Event:

Event 1:

Event 2:

Definitions & notation — Exercise

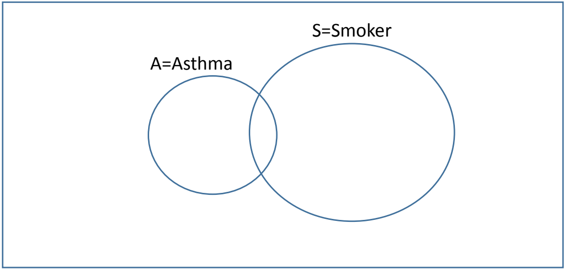
Experiment:
Randomly select an MSc Med Stats student. Ask student how many days a week (on average) they do vigorous exercise.

Sample space:
$\Omega = \{0, 1, 2, 3, 4, 5, 6, 7\}$

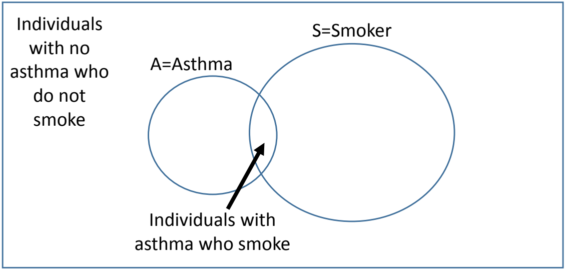
Event:
Event 1: Student does no exercise, $E_1 = \{0\}$
Event 2: Student exercises 3-4 days a week, $E_2 = \{3, 4\}$

Venn diagram for the smoking and asthma example

Venn diagrams are sometimes used to represent probabilities in the whole sample space graphically.



Venn diagram for the smoking and asthma example



Set notation for events

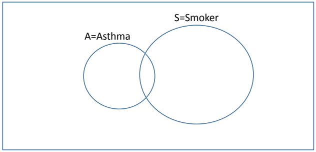
Union:
$A \cup S$ is the event that a randomly selected individual has asthma, *or* is a smoker *or* has asthma and is a smoker.

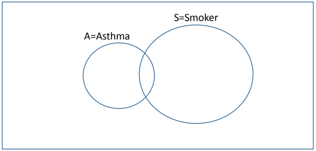
Intersection:
$A \cap S$ is the event that a randomly selected individual has asthma *and* is a smoker.

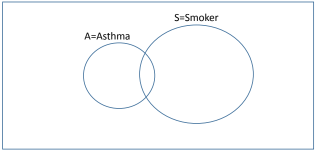
Complement:
$\bar{A}$ is the event that a randomly selected individual does *not* have asthma.

Set union, intersection and complement

Shade in $A \cup S$, $A \cap S$, and $\bar{A}$:

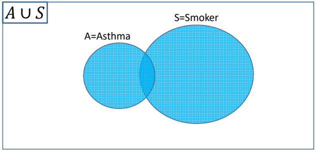


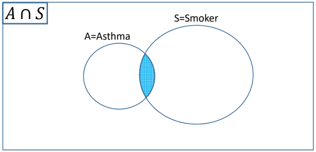


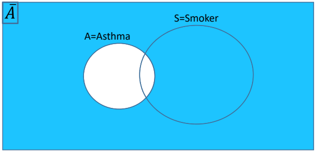


Set union, intersection and complement

Shade in $A \cup S$, $A \cap S$, and $\bar{A}$:







The axioms of probability

The axioms of probability are statements which probabilities must satisfy:

1. $0 \leq P(A) \leq 1$ for every event A.

The axioms of probability

The axioms of probability are statements which probabilities must satisfy:

1. $0 \leq P(A) \leq 1$ for every event A.
2. $P(\Omega) = 1$ where Ω is the total sample space.

The axioms of probability

The axioms of probability are statements which probabilities must satisfy:

1. $0 \leq P(A) \leq 1$ for every event A.
2. $P(\Omega) = 1$ where Ω is the total sample space.
3. For disjoint (mutually exclusive) events $A_1, .., A_n$:

   $$P(A_1 \cup A_2 \cup \ldots \cup A_n) = P(A_1) + P(A_2) + \ldots + P(A_n).$$

   The events $A_1, .., A_n$ are disjoint if there are no intersections between any of the events.

The axioms of probability

The axioms of probability are statements which probabilities must satisfy:

1. $0 \leq P(A) \leq 1$ for every event A.
2. $P(\Omega) = 1$ where Ω is the total sample space.
3. For disjoint (mutually exclusive) events $A_1, .., A_n$:

   $$P(A_1 \cup A_2 \cup \ldots \cup A_n) = P(A_1) + P(A_2) + \ldots + P(A_n).$$

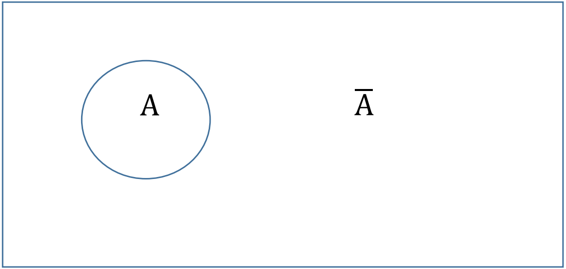
   The events $A_1, .., A_n$ are disjoint if there are no intersections between any of the events.

A number of results follow from the axioms. For example:

- A and $\bar{A}$ are exhaustive (one of them will certainly occur)
- $P(A) = 1 - P(\bar{A})$

Proving $P(A) = 1 - P(\bar{A})$

Venn diagram for proving $P(A) = 1 - P(\bar{A})$:



What is $A \cup \bar{A}$?

____

Axiom 1/2/3 may be useful

Are A and $\bar{A}$ mutually exclusive?

Yes/No

Axiom 1/2/3 may be useful

Proving $P(A) = 1 - P(\bar{A})$

1. Applying Axiom 2
   - Firstly, note we can write $\Omega = A \cup \bar{A}$.
   - Axiom 2 says $P(\Omega) = 1$

   $\implies$ $P(A \cup \bar{A}) = 1$

Proving $P(A) = 1 - P(\bar{A})$

1. Applying Axiom 2
   - Firstly, note we can write $\Omega = A \cup \bar{A}$.
   - Axiom 2 says $P(\Omega) = 1$

   $\implies$ $P(A \cup \bar{A}) = 1$

2. Applying Axiom 3
   - A and $\bar{A}$ are disjoint, so Axiom 3 can be used

   $\implies$ $P(A \cup \bar{A}) = P(A) + P(\bar{A})$

Proving $P(A) = 1 - P(\bar{A})$

1. Applying Axiom 2
   - Firstly, note we can write $\Omega = A \cup \bar{A}$.
   - Axiom 2 says $P(\Omega) = 1$

   $\implies$ $P(A \cup \bar{A}) = 1$

2. Applying Axiom 3
   - A and $\bar{A}$ are disjoint, so Axiom 3 can be used

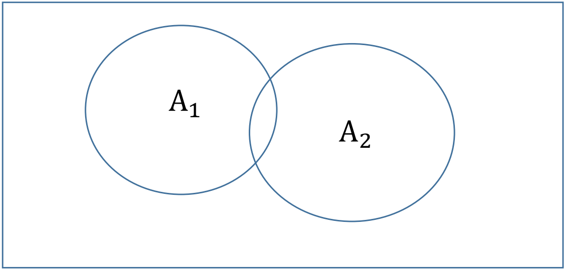
   $\implies$ $P(A \cup \bar{A}) = P(A) + P(\bar{A})$

Therefore $1 = P(A) + P(\bar{A})$

Re-arranging, $P(A) = 1 - P(\bar{A})$.

Another useful rule

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$$



We will prove this rule in the practical session.

Outline

Conditional probability

Consider two events A (e.g. asthma) and S (smoking).

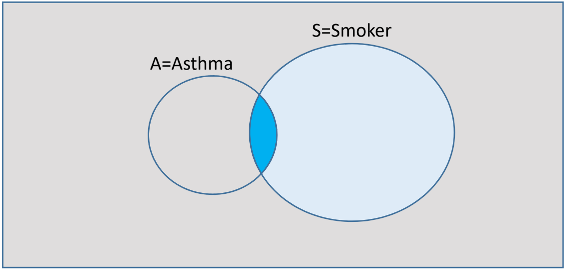
- Suppose we know that an individual is a smoker. What, then, is the probability that they suffer from asthma?
- We call this the conditional probability of A 'given' (or 'conditional on') S
- We write $P(A|S)$ for the probability that A occurs, given or conditional on S occurring.

The idea of conditional probability is of fundamental importance in medical statistics. For example:

- What is the probability that a patient will have a myocardial infarction within 5 years given she has hypertension?

Conditional probability

- Once we condition on S occurring, the sample space is reduced to S.
- The probability that A now occurs is the probability of the intersection relative to the reduced sample space



Defining conditional probability

- The probability that A now occurs is the probability of the intersection relative to the reduced sample space:

$$P(A|S) = \frac{P(A \cap S)}{P(S)}.$$

- Multiplying through by $P(S)$, we see this implies that

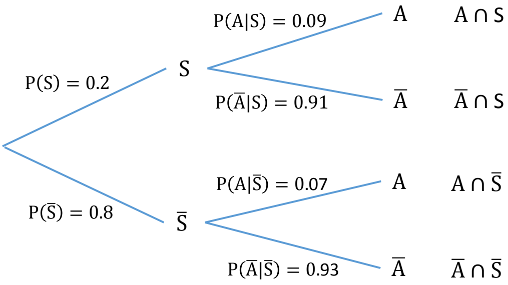
$$P(A \cap S) = P(S)P(A|S).$$

Outline

Probability trees

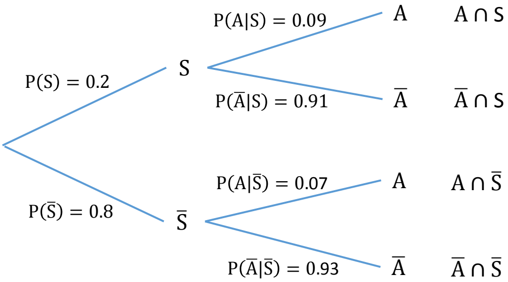
A good way of displaying and calculating simple conditional probabilities is through the use of a *probability tree*.

Remember: prevalence of smoking is 20% among adults in general; 9% of smokers and 7% of non-smokers have asthma.



Asthma and smoking

What is the probability that someone has asthma and is a smoker?



$$P(A \cap S) = P(S) \times P(A|S) = 0.2 \times 0.09 = 0.018$$

Outline

Independence

The idea of independence is another concept of fundamental importance in medical statistics.

Two events are independent if knowing information about one of the events *does not* give us information about the other.

For example, knowing that Sam has a beard tells us that Sam is more likely to be male, thus facial hair and sex are not independent.

Are the following pairs of events independent?

- Smoking and being left-handed
- Having asthma and liking spicy food
- Smoking and having asthma

Definition of independence

Formal definition:

- A and B are said to be **independent** if and only if

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

Connection to intuitive definition:

- Since we know that $P(A \cap B) = P(A|B)P(B)$, we see that the equality above will only be true when

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

  i.e. when knowing about B tells us nothing about A.

Example of non-independence

Are smoking and asthma independent?

- We know that $P(A|S) = 0.09$ and $P(A|\bar{S}) = 0.07$.
- If smoking and asthma were independent, we would have:
  - $P(A|S) = P(A)$
  - $P(A|\bar{S}) = P(A)$
  - *but* this implies $P(A|S) = P(A|\bar{S})$, which is not true
- So $P(A \cap S) = P(A|S)P(S) \neq P(A)P(S)$
- So smoking and asthma are *not* independent.

Outline

Theorem of total probability

Suppose we wish to know the overall prevalence of asthma i.e. $P(A)$, but we only have information on the prevalence of asthma by age group.

Age-group (years)		Prevalence of asthma	Fraction of population
$0-5$	(G_1)	4%	7%
$6-17$	(G_2)	9%	13%
$18-40$	(G_3)	7%	26%
$41-60$	(G_4)	8%	37%
$61-100$	(G_5)	7%	17%

The theorem of total probability allows us to obtain the overall probability from the conditional probabilities of asthma given each age group.

Partition of a sample space

A set of events $G_1, \cdots, G_n$ partition the sample space if

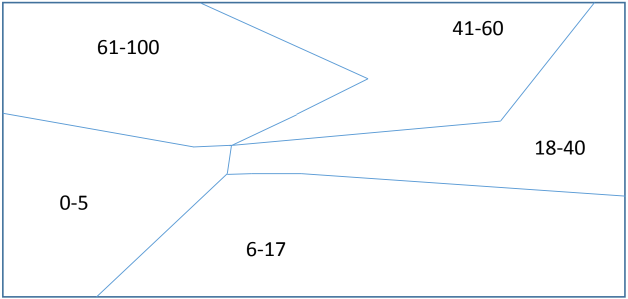
- all events are possible
- at least one event must occur, but
- no two events can occur simultaneously

For example, for the sample space of ages in the general population (of people at most 100 years old).

- The age-groups: $\{0-5, 6-17, 18-40, 41-60, 61-100\}$ partition the sample space

Partition of a sample space

A diagram showing a partition of the sample space of ages in the general population (of people at most 100 years old).



Partition of a sample space - formal definition

The events $G_1, G_2, ...G_n$ partition the sample space Ω if:

- **all events are possible**
  - $P(G_i) > 0$ for all i
  - Each event in the partition has a non-zero probability of occurring

Partition of a sample space - formal definition

The events $G_1, G_2, ...G_n$ partition the sample space Ω if:

- **all events are possible**
  - $P(G_i) > 0$ for all i
  - Each event in the partition has a non-zero probability of occurring
- **at least one event must occur**
  - $\bigcup_{i=1}^{n} G_i = \Omega$
  - i.e. the union of the events = the sample space

Partition of a sample space - formal definition

The events $G_1, G_2, \ldots G_n$ partition the sample space Ω if:

- **all events are possible**
  - $P(G_i) > 0$ for all i
  - Each event in the partition has a non-zero probability of occurring
- **at least one event must occur**
  - $\bigcup_{i=1}^{n} G_i = \Omega$
  - i.e. the union of the events = the sample space
- **no two events can occur simultaneously**
  - $G_i \cap G_j = \emptyset$ (empty) for all $i \neq j$.
  - i.e. G_i and G_j are disjoint

Examples of partitions

If the sample space is ages of people under 20 years old, which of these are partitions?

- Age groups 0-12, 10-19 years
- Age groups 0-5, 6-9, 10-25 years
- Age groups 0-3, 4-13, 14-19 years
- Age groups 0-5, 6-9, 11-14, 15-19 years

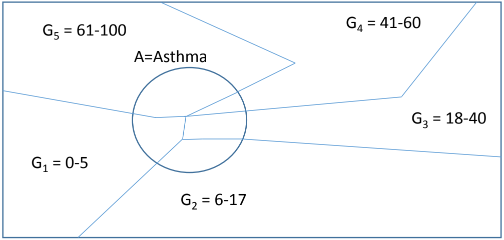
Theorem of total probability

We have:

- an event A (asthma)
- a partition $G_1, .., G_5$ (the five agegroups) of the sample space.

The theorem of total probability says that:

$$P(A) = P(A|G_1)P(G_1) + P(A|G_2)P(G_2) + \ldots + P(A|G_5)P(G_5).$$



Proof: Theorem of total probability

1. We can express A as:

$$A = (A \cap G_1) \cup (A \cap G_2) \cup \ldots \cup (A \cap G_5).$$

Proof: Theorem of total probability

1. We can express A as:

$$A = (A \cap G_1) \cup (A \cap G_2) \cup \ldots \cup (A \cap G_5).$$

2. The events $G_1, ... G_5$ are mutually exclusive (disjoint). Therefore, the sets $(A \cap G_1), (A \cap G_2), ..., (A \cap G_5)$ are also mutually exclusive. So Axiom 3 gives

$$P(A) = P(A \cap G_1) + P(A \cap G_2) + \ldots + P(A \cap G_5).$$

Proof: Theorem of total probability

1. We can express A as:

$$A = (A \cap G_1) \cup (A \cap G_2) \cup \ldots \cup (A \cap G_5).$$

2. The events $G_1, ... G_5$ are mutually exclusive (disjoint). Therefore, the sets $(A \cap G_1), (A \cap G_2), ..., (A \cap G_5)$ are also mutually exclusive. So Axiom 3 gives

$$P(A) = P(A \cap G_1) + P(A \cap G_2) + \ldots + P(A \cap G_5).$$

3. Finally, $P(A \cap G_i) = P(A|G_i)P(G_i)$. So

$$P(A) = P(A|G_1)P(G_1) + P(A|G_2)P(G_2) + \ldots + P(A|G_5)P(G_5).$$

Example: Theorem of total probability

$$P(A) = P(A|G_1)P(G_1) + P(A|G_2)P(G_2) + \ldots + P(A|G_5)P(G_5).$$

Age-group (years)		Prevalence of asthma $P(A\|G_i)$	Fraction of population $P(G_i)$	$P(A\|G_i)P(G_i)$
$0-5$	(G_1)	0.04	0.07	0.0028
$6-17$	(G_2)	0.09	0.13	0.0117
$18-40$	(G_3)	0.07	0.26	0.0182
$41-60$	(G_4)	0.08	0.37	0.0296
$61-100$	(G_5)	0.07	0.17	0.0119
Overall			$\sum_{i=1}^{5} P(A\|G_i)P(G_i) =$	0.074

So the overall prevalence of asthma is 7.4%.

Theorem of total probability: General statement

We have:

- an event A
- a partition $G_1, .., G_n$ of the sample space Ω.

The theorem of total probability says that:

$$P(A) = P(A|G_1)P(G_1) + P(A|G_2)P(G_2) + \ldots + P(A|G_n)P(G_n),$$

or

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

Outline

Summary

- Probability
  - Probability can be defined in terms of relative frequencies.
  - Probability theory is formally defined by specifying three *axioms*.
  - Venn diagrams are useful for proving probability theorems.
- Conditional probability
  - A conditional probability expresses the probability that one event occurs given that another event has occurred.
  - Probability trees are useful for expressing conditional probabilities.
- Independence
  - Two events are not independent if knowing whether one event occurred changes the probability that the second will occur.
- Theorem of total probability
  - Useful for calculating $P(A)$ based on partition $B_1, .., B_n$.