

Elementary probability theory

The concept of probability is fundamental in statistical analysis. Theory of probability underpins most of the methods used in statistics.

1.1 Experiments, outcomes and sample spaces

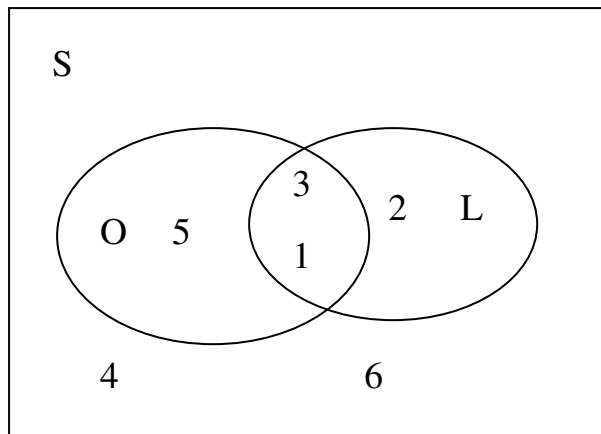
Tossing a coin or rolling a die are examples of experiments – that is, any action or observation whose outcome is uncertain. The first experiment has two basic outcomes head and tail. The second has six basic outcomes: 1,2,3,4,5 and 6. Events are defined in terms of the basic outcomes of an experiment, e.g. we could define event O as an odd score on a die, consisting of the three basic outcomes 1,3 and 5, while event L could be defined as the three lowest scores, 1, 2 and 3. An event which includes all outcomes *except* those belonging to an event A is denoted $\neg A$.

The Union of two events A and B is usually written $A \cup B$, and includes all basic outcomes which are included in A or in B or in both; e.g. $O \cup L$ includes 1,2,3 and 5.

The intersection of two events A and B, $A \cap B$ includes those basic outcomes that are in both A and B. E.g. $O \cap L$ includes 1 and 3.

The Sample Space, written as S, includes all basic outcomes from the experiment.

Venn diagrams are often used to represent events. Venn diagrams for events O and L in the die experiment are shown below:



The Empty Set (or Impossible event) \emptyset contains no basic outcomes

Mutually exclusive events have no common outcomes, i.e. A and B are mutually exclusive if $A \cap B = \emptyset$.

Exhaustive events include all the basic outcomes. If A and B are exhaustive, then $A \cup B = S$.

1.2. Definition of probability

a) Classical definition. If an experiment has n equally likely outcomes, and n_a of them are favourable to (included in) event A, then the probability of A is defined as

$$P(A) = n_a/n, \text{ where } 0 \leq P(A) \leq 1.$$

For example in the case of the die, if O=odd score, then $P(A) = \text{favourable outcomes/possible outcomes} = 3/6 = 0.5$. This of course assumes that the die is unbiased.

The classical definition is circular, in that it involves “equally likely” basic outcomes, which presupposes some notion of probability.

(b) Empirical definition. Here the probability of event A, $P(A)$, is defined as a limit of the ratio n_a/n as n tends to ∞ , that is as n gets larger.

Difficulties: (i) The estimate of $P(A)$ changes with each experiment. When to stop? (ii) The definition cannot be used when experimentation is impossible, e.g. what is the probability that inflation will be less than 3% in a year's time?

(c) Subjective probability. Here probability is estimated subjectively, as the experimenter's subjective view of the likelihood of the event. This is not entirely satisfactory, as it lacks any objective base.

In short, there is no universally accepted or unproblematic definition of probability. However, we can proceed on the basis of an intuitive understanding of what probability 'means', and using the mathematical laws of probability.

1.3 Laws of probability

Above we have discussed how to determine probabilities of basic outcomes like a tail when a coin is tossed or a six when a die is rolled. We now establish rules for finding probabilities of events when probabilities of basic outcomes are known or to find probabilities of more complex events using probabilities of simpler events.

(a) The addition law

If events A and B are mutually exclusive, i.e. $A \cap B = \emptyset$, then probability of the union of A and B is

$$P(A \cup B) = (n_a + n_b)/n = n_a/n + n_b/n = P(A) + P(B)$$

If A and B are not mutually exclusive, we have

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

Since when we add the basic events from A and B, we have counted those that occur in both A and B twice. (Check this using a Venn diagram)

(b) The multiplication law

We first define conditional probability, e.g. the probability of A happening given that B has happened:

$$P(A|B) = n_{ab}/n_b = [n_{ab}/n]/[n_b/n] = P(A \cap B)/P(B)$$

Where n_{ab} is the number of basic outcomes in $A \cap B$. Note that n_{ab}/n_b is the definition of the probability of A if the sample space is reduced from S to B – in other words, the probability of A|B is the proportion of the basic outcomes in B where A also happens.

$P(A|B)$ is called the conditional probability of A given B. By contrast, $P(A)$ or $P(B)$ can be called the unconditional probability of event A or B.

E.g. if events O and L are as defined above in the die roll experiment, then $P(O|L) = P(O \cap L)/P(L) = (2/6)/(3/6) = 2/3$.

Since if we know that the score on a die is 1,2 or 3, then there is a 2/3 probability that the score is odd.

The multiplication law follows immediately from the definition of conditional probability:

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

This can be extended to probabilities involving three or more events, e.g.

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

1.4 Independence and dependence

Events A and B are said to be independent if $P(A|B) = P(A)$, or equivalently if $P(B|A) = P(B)$. (*Exercise: why are these two equivalent?*) It further follows from this that if A and B are independent events, then we have a particularly simple form of the multiplication law, namely:

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

Similarly, for mutually independent events A, B and C,

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

Example: Mortality records in a hospital show that the probability of dying from lung cancer is 0.1 and the probability of being a smoker is 0.5. If smoking and dying of lung cancer were independent events, then the probability of being a smoker and dying of lung cancer would be 0.05. If the probability of this intersection is greater than 0.05, then the two events are dependant, and the conditional probability of dying of lung

cancer given that you are a smoker is higher than the unconditional probability of dying of lung cancer.

Frequently, we will use the *assumption* of independence, where it is intuitively reasonable, to enable us to simplify our probability calculations. For example, any two individuals getting lung cancer can probably be treated as independent events for most purposes. (Though if two people knew each other as children, for example, the events of each of them starting smoking might not be independent due to peer pressure – so we must often be careful with assumptions of independence, and be prepared to question and revise such assumptions where appropriate.)

1.5 Bayes' Theorem

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

$$\text{Since } P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B|A)}{P(B)}$$

This is saying that if we know B has occurred, we can work out the probability that A has also occurred if we know the original probabilities of A and B, and the probability of B given A. This can be seen as revising our estimate of the probability of A in the light of B happening.

We frequently break down P(B) on the bottom of the fraction in the equation as $P(A)P(B|A) + P(\neg A)P(B|\neg A)$ – that is, by considering separately the probability of B happening if A has happened, and the probability of B happening if A has not happened.

Example:

You have a headache. You are worried this might be an early sign of meningitis.

Let H be the event that you have a headache, and M the event that you have meningitis. What we are interested in is P(M|H).

Suppose you know that 0.01% of the population has meningitis. (so $P(M)=0.0001$.) Suppose you also know that 90% of people with meningitis will get a headache as an early sign ($P(H|M)=0.9$), but that 5%

of people without meningitis will also have a headache on any given day. (So $P(H|\neg M)=0.05$).

Therefore, $P(H) = P(M)P(H|M)+P(\neg M)P(H|\neg M) = 0.0001*0.9 + 0.9999*.05 = 0.050085$.

Therefore, by Bayes' Theorem, $P(M|H) = P(H|M)P(M)/P(H) = 0.9*.0001/.050085 = .00180$ to 3 s.f., or less than one in 500. (Please note, these figures are hypothetical).

Suppose, however, that your headache gets worse the next day. You know that if you didn't have meningitis, the probability of this would have been 10%, while if you did, the probability would be 90%. Let W be the event that you have a headache that gets worse the second day. Combining the above information with the original conditional probabilities of H, we get $P(W|M) = 0.81$, and $P(W|\neg M)=0.005$. Thus, $P(W)=.005*.9999+.81*.0001=.0050805$

Then $P(M|W) = 0.81*.0001/.0050805 = 0.0159$, or 1.59%. No need to panic yet, but you might want to see your doctor just in case!

We see that, as new information arrives, we can successively use Bayes' theorem to revise our estimates of the probabilities of unknown events.

1.6 Probability calculus

The following results help to evaluate probabilities by finding the total number of outcomes and number of favourable outcomes of events.

- 1) Multiplication rule: if one experiment has n outcomes, and another has m outcomes, then there are $n*m$ joint outcomes of the combined experiment.
- 2) Factorials: if there are n different objects, they can be arranged in $n! = n(n-1)(n-2)*...*2*1$ different ways. E.g. for $n=3$, $n!=3*2*1=6$, i.e. ABC, ACB, BAC, BCA, CAB, CBA.
- 3) Permutations: Suppose we have n objects, of which we select k, in order. The number of permutations is given by:

$${}^n P_k = n * (n - 1) * (n - 2) * \dots * (n - k + 1) = \frac{n * (n - 1) * \dots * 2 * 1}{(n - k) * (n - k - 1) * \dots * 2 * 1} = \frac{n!}{(n - k)!}$$

Since we have n possibilities for the first choice, then $n-1$ for the second, etc.

E.g. ${}^4P_2 = 4 * 3 = 12$, i.e. AB, AC, AD, BA, BC, BD, CA, CB, CD, DA, DB, DC.

- 4) Combinations: suppose we must again select k objects from n , but ignoring the order they are drawn in (i.e. AB is the same as BA).
Then:

$${}^nC_k = \frac{n!}{k!(n-k)!}$$

Since we must divide the number of permutations by the number of internal orderings of the k objects chosen, that is, by $k!$.

E.g. ${}^4C_2 = 4!/(2!)(4-2)! = 24/2*2=6$, namely AB, AC, AD, BC, BD, CD.
(Note that we have lost the copies of each pair in the opposite order.)