Introduction to probability - Part II

Axioms and a few probability rules

Remarks:

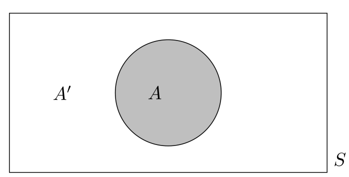
- The modern theory of probability is based on three basic principles, called the axioms of probability. Both the Classical Approach and the Relative Frequency Approach will satisfy the axioms of probability. Thus this approach is more general and any consequences of the axioms will hold true for the other two models.
- All approaches, that is the Subjective, the Classical Approach and the Relative Frequency Approach, can be used. The approach that we are using will never be explicitly stated, but it should be obvious from the context which approach is being used or is more appropriate.
- Before defining the axioms, we will need operations on events.

Operations on Events:

Consider a random experiment with sample space S. We will assume that all the events below are subsets of S.

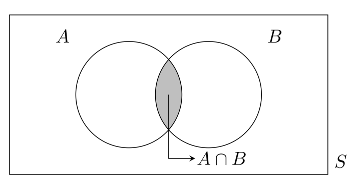
Complement: A' occurs means that A does **not** occur.

Here is a visual representation of the event A and its complement A'. The rectangle represents the sample space S. Here S is divided into two disjoint sets: A and its complement $'A$.



Intersection:

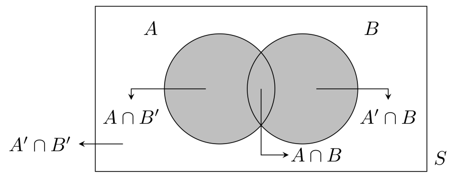
$A \cap B$ occurs means that A occurs **and** B occurs.



Generalizing: $A_1 \cap A_2 \cap \cdots \cap A_n$ occurs means that **all** of the events $A_1, A_2, \ldots, A_n$ occur.

Union:

$A \cup B$ occurs means that A occurs, or B occurs, or both occur.



$A \cup B$ is the shaded region

Remark: If $A \cup B$ occurs, it means that only one of the following events events has occurred:

- $A \cap B'$ occurs, i.e. A occurs but B does not occur;
- $B \cap A'$ occurs, i.e. B occurs but A does not occur;
- $A \cap B$ occurs, i.e. A occurs and B occurs.

Generalizing: $A_1 \cup A_2 \cup \cdots \cup A_n$ occurs means that **at least one** the events $A_1, A_2, \ldots, A_n$ occurs.

DeMorgan Laws:

a) $(A_1 \cup A_2 \cup \cdots \cup A_n)'$ occurs $\equiv$ "none of the events $A_1, A_2, \ldots, A_n$ occur"
$\equiv$ $A_1' \cap A_2' \cap \ldots \cap A_n'$ occurs

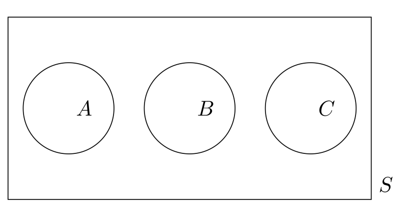
b) $(A_1 \cap A_2 \cap \cdots \cap A_n)'$ occurs $\equiv$ "at least one of the events $A_1, A_2, \ldots, A_n$ does not occur"
$\equiv$ $A_1' \cup A_2' \cup \ldots \cup A_n'$ occurs

Mutually Exclusive Events:

Definition: We say that the events $A_1, A_2, \ldots, A_n$ are **mutually exclusive** events if

$$A_i \cap A_j = \emptyset \quad \text{for } i \neq j.$$

Here is an illustration of 3 mutually exclusive events.



"The theory of probability as a mathematical discipline can and should be developed from axioms in exactly the same way as geometry and algebra." - Kolmogorov, Andrey

Axioms of Probability

Let S be a sample space. A function P that assigns real numbers to events is called a *probability* function if it satisfies the following three axioms:

1. **[Positivity]** $P(A) \geq 0$, for all events A

2. **[Certainty]** $P(S) = 1$

3. **[Additivity]** Let $A_1, A_2, \ldots$ be a sequence of mutually exclusive events, then

$$P\left(A_1 \cup A_2 \cup \cdots\right) = P(A_1) + P(A_2) + \cdots$$

Direct consequences of the axioms (probability rules):

Here are some direct consequences of the axioms of probability. Let S be a sample space with some probability function P, i.e. P satisfies the axioms of probability, then the following hold:

1. $$P(\emptyset) = 0,$$
   where $\emptyset$ is the empty set, i.e. an event with no outcomes.
2. Let A_1 and A_2 be events such that $A_1 \subseteq A_2$, that is if A_1 occurs then A_2 occurs, then
   $$P(A_1) \leq P(A_2).$$
3. $$0 \leq P(A) \leq 1, \quad \text{for all events } A \subseteq S.$$
4. $$P(A') = 1 - P(A), \quad \text{for all events } A \subseteq S.$$

Addition Rules

a)

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

b)

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

Example 7: Suppose that the probability that a fruit fly will have a wing mutation is 5%, that it will have an eye mutation is 10% and that it will have both is 2%.

(a) What is the probability that a fruit fly will have at least one of the mutations?

(b) What is the probability that it will have the wing mutation but not the eye mutation?

(c) What is the probability that the fly will have neither of the two mutations?

Conditional Probabilities

Remark: Sometimes we wish to restrict our focus on the outcomes of a particular event. Then, relative to the outcomes in this event what is the probability that another event will occur?

Definition: Let S be a sample space with a probability function P. Assume that $P(B) > 0$, then the *conditional probability* that A occurs given that B occurs is defined as

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

Remarks:

1. We can show that for a fixed event B the function $P(\cdot|B)$ satisfies the axioms of probability. Thus, the consequences of the axioms also apply to conditional probabilities, for example
   (a) $P(\emptyset|B) = 0$.
   (b) If $A \subseteq C$, then $P(A|B) \leq P(C|B)$.
   (c) $0 \leq P(A|B) \leq 1$, for all $A \subseteq S$.
   (d) $P(A'|B) = 1 - P(A|B)$, for all $A \subseteq S$.

2. For the Classical Approach, the definition of conditional probability is equivalent to

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

   In the Classical Approach case, we can interpret $P(A|B)$ as the proportion of outcomes in B that also satisfy the criteria to be in A.

Example 7: Refer to Example 6. What is the probability that the fly will have a wing mutation given that it has an eye mutation?

Multiplication Rule

$$P(A \cap B) = P(A|B)\,P(B) \qquad \text{and} \qquad P(A \cap B) = P(B|A)\,P(A)$$

✓

Example 8: In a given fruit fly population, 25% have a certain wing mutation. Of the group with the wing mutation, 40% have an eye mutation. What is the probability that a fly selected at random will have both mutations?

Example 9: Suppose that the probability that a certain type of fish will die if exposed to high levels of radiation is 85% and that the probability of exposure to high levels of radiation is 0.01%. What is the probability that a fish is exposed to high levels of radiation and it died.

The Law of Total Probability

Let $A_1, A_2, \ldots, A_n$ be a collection of events that are **exhaustive** and **mutually exclusive**, then

$$P(B) = P(B|A_1)\,P(A_1) + P(B|A_2)\,P(A_2) + \cdots + P(B|A_n)\,P(A_n).$$

✓

Bayes' Theorem

Let A and B be two probable events (i.e. $P(A) > 0$ and $P(B) > 0$). Then

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

Remark: Bayes' Theorem is useful when we want $P(A|B)$, but we know $P(B|A)$. It allows us to do a role reversal.

✓

Example 10: Samples taken from a river near an industrial plant are tested for toxic levels of lead and mercury: 32% contain toxic levels of mercury, 16% contain toxic levels of lead, 38% contain toxic levels of at least one of the two substances.

(a) A sample is chosen at random. It contains toxic levels of mercury. What is the probability that is also contains toxic levels of lead?

(b) It is known that 45% of the samples containing toxic levels of mercury also contain toxic levels of arsenic. What is the probability that a sample contains toxic levels of both mercury and arsenic?

(c) Of the samples which do not contain toxic levels of mercury, only 2% contain toxic levels of arsonic. What is the probability that a sample contains toxic levels of arsenic?

Example 11: Select a person at random from a population of 3 ethnic groups A, B, and C. Suppose that group A represents 50% of the population while B and C represent 30% and 20% of the population, respectively. It is known that 1% of A, 0.5% of B and 2% of C will test positive for Tuberculosis.

(a) What is the probability that the person will test positive for tuberculosis?

(b) It is known that the person tested positive for tuberculosis, what is the probability that he or she was from group C?

Example 12: We are testing for tuberculosis (TB) but tests are not perfect. Suppose that the false negative rate is 5%, that is 5% of people with the disease will test negative. Suppose that the false positive rate is 1%, that is 1% of people without the disease will test positive. Suppose that the prevalence of the disease is 10%, that is 10% of the population has TB. A patient has tested positive for TB, what is the probability that the patient has TB?

Terminology for diagnostic tests

Consider a diagnostic test for a certain disease D. The test is not perfect. Let $+$ be the event that the result is positive. (i.e. the test indicates the presence of the disease).

For a good diagnostic test:

(i) the chances of observing a positive result, if the condition is present, should be high:

$$P(+|D) = \text{sensitivity of the test.}$$

(ii) the chances of observing a negative result, if the condition is absent, should be high:

$$P(-|D') = \text{specificity of the test.}$$

Parameters concerning the predictive values of the test:

(i)

$$P(D|+) = \text{PPV} = \text{positive predictive value.}$$

(ii)

$$P(D'|-) = \text{NPV} = \text{negative predictive value.}$$

Independence of Events

Definition: Let A and B be two events. We say that A and B are independent if

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

Motivation: Why is this multiplication rule used as a definition of independence? The motivation comes from the following result.

Theorem: Let A and B be two events such that $P(A) > 0$ and $P(B) > 0$, then the following three statements are equivalent (in the sense when one is true then all are true):

1. $P(A|B) = P(A)$
2. $P(B|A) = P(B)$
3. $P(A \cap B) = P(A)\, P(B)$

Remarks:

- Intuitively we would want independence of two events to mean the following. If A and B are independent, then the probability that A will occur should not depend on the fact that B has occurred or not. This is conveyed by the statements (1) and (2).
- As long as the conditional probabilities $P(A|B)$ and $P(B|A)$ are defined, that is $P(B) > 0$ and $P(A) > 0$, respectively, then the three above statements are equivalent.
- Since the quantities in statement (3) are always defined, then it is more general than the other two statements. We will use this concept to generalize the concept of independence to more than two events.

Exemple 13: Consider the crossing of Drosophila with the following genotype $RrWw$, where R is red eyes (r is vermilion eyes) and W is long wings (w is rudimentary wings). Here is a contingency table that can be used to describe 458 offspring in terms of the colour of their eyes and the shape of their wings.

	Eyes		
Wings	Red	Vermilion	Total
long	105	316	421
rudimentary	33	4	37
	138	320	458

Are the events vermilion eyes and rudimentary wings independent?

Definition: Let $A_1, A_2, \ldots, A_n$ be a collection of events. We say that the events are independent if for any sub-collection $A_{i1}, A_{i2}, \ldots, A_{ik}$ the following holds

$$P(A_{i1} \cap A_{i2} \cap \ldots \cap A_{ik}) = P(A_{i1})\, P(A_{i2}) \cdots P(A_{ik}).$$

Remark: Let $A_1, A_2, \ldots, A_n$ be a collection of independent events. If we replace some of these events by their complement, for example replace A_i with A_i', then the new collection is still a collection of independent events.

For example, if A and B are independent events, then

- A and B' are independent events;
- A' and B are independent events;
- A' and B' are independent events.

Example 14: In the last 30 years there are been about 105 snake bites in Ontario. Among these people only 2 have died. If three individuals were bitten in the last year. What is the probability that at least one has survived? *Assume that the individuals will survive independently of each other.*

Example 15 [Sampling without replacement]: Suppose that we have a group of 12 students: 5 biology students and 7 chemistry students. We select two students at random among the 12 students without replacement. Define the events B_i="ith selected student is a biology student", for $i = 1, 2$. Are B_1 and B_2 independent?
Hint: It is reasonable to take $P(B_2|B_1) = 4/11 \approx 0.3636$. Use the total probability rule, to show that $P(B_2) = 5/12 \approx 0.41667 \neq P(B_2|B_1)$.

Remark: Because the population is small in size, as remove a few elements from the population, this causes a change in the composition of the population. However, as the population becomes large, removing a few elements will have little effect on the composition of the population. The trials are still dependent, but this size of this dependence is very small. When sampling from a large population, it is reasonable to assume that the trials are independent.

Remark: Do **not** confuse the concept of independent events with the concept of mutually exclusive events. The two concepts imply different things.

- If A and B are **independent events**, then

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

  and

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

- If A and B are **mutually exclusive events**, then

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

  and

  $$P(A \cup B) = P(A) + P(B).$$