ISEG - Lisbon School of Economics and Management

Statistics I

(Date of this version: 6/03/2018)

Handout 1 – PROBABILITY

Issues covered:

1.1 Introduction

Why is statistics needed for?

It gives support and recommendations to decision making process under uncertainty.

Managers often base their decisions on an analysis of uncertainties such as the following:

1. What are the chances that sales will decrease if we increase prices?

2. What is the likelihood a new assembly method will increase productivity?

3. How likely is it that the project will be finished on time?

4. What is the chance that a new investment will be profitable?

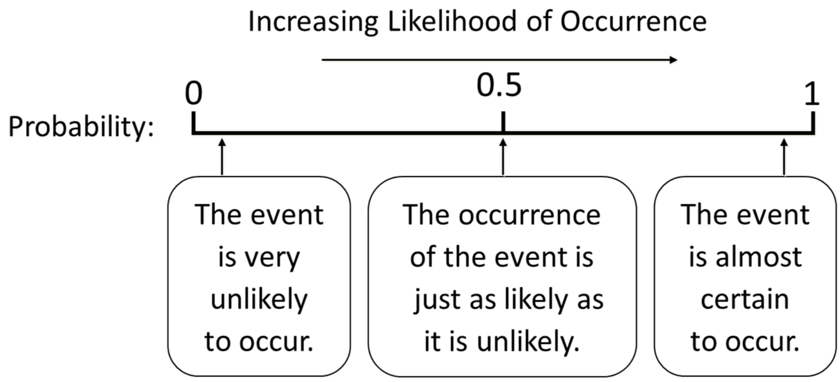
Statistics helps providing answers to this type of questions.

The beginnings of the research field of **statistics** may be found in XVIII century studies in **probability** motivated by interest in games of chance. The theory developed for "heads or tails" or "red or black" soon found applications in economics, management, physics, biology, psychology, etc. Researchers realised that the most rigorous way to take into account uncertainty in their studies was to resort to concept of **probability**

Probability is a numerical measure of the likelihood that an event will occur.

Thus, probability can be used as measures of the degree of uncertainty associated with an event

Probability values are always assigned on a scale from 0 to 1



1.2 Sample Spaces and Events

Since all probabilities pertain to the occurrence or nonoccurrence of events, let us explain first what we mean here by *event* and by the related terms *experiment, outcome*, and *sample space*.

An experiment is to any procedure that can be infinitely repeated and has a well-defined set of possible results which can be observed and measured.

Examples of experiments:

a) Flip of a coin and observation of head or tails.

b) Toss of a dice and observation of number of spots.

c) Observation of the number of meteors greater than 1 meter diameter that strike Earth in a year.

d) Observation of the daily change (Δ) in an index of stock market prices.

e) Observation of the time it takes before your next telephone call

Definition: Outcomes, Sample Space

Outcomes: The results one obtains from an experiment

Sample Space (S): Set of all possible outcomes from an experiment.

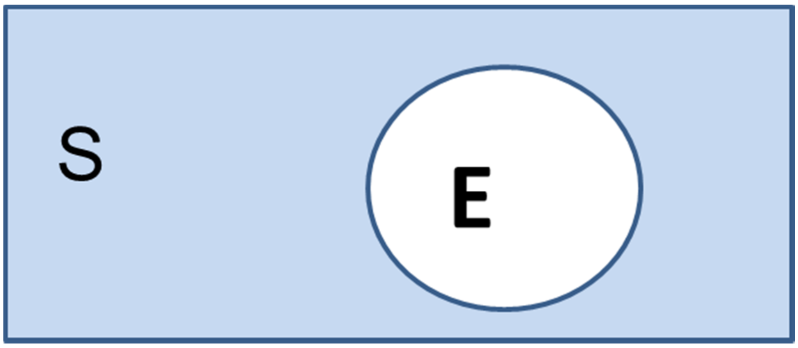
Element or sample point (*s*): Each of the possible outcomes of an experiment.

Sample Spaces can be:

- **A) Discrete:**
  - **countable** and **finite: examples: a) S** = {H,T}. $\# S = 2^1$ =2 elements. **b) S** = {1, 2, 3, 4, 5, 6}. $\# S = 6$
  - **infinite though countable number of elements: examples: c) S** = {0, 1, 2, 3, 4, …}.
- ***b*) continuous – uncountable. Examples: d) S** =$\mathbb{R}$, **e)** $\boldsymbol{S} =$ {c: c $\geq$ 0}=[0,+∞)

Definition: EVENT: Any subset of the sample space **S**.

$\boldsymbol{E} = \{\boldsymbol{s_1}, \boldsymbol{s_2}, \cdots, \boldsymbol{s_p}\} \subset \boldsymbol{S}$



Remarks:

- If E = $\{s\}$ *it is an element.*
- S *is an event too.*

An event $\boldsymbol{E} \subset$S occurs if the experiment results is all of its elements

$\boldsymbol{s_i} \in \boldsymbol{E}, \boldsymbol{i} = \mathbf{1},\ \mathbf{2},\ \cdots,\ \boldsymbol{p}$

Example: Experiment - A coin is flipped twice.

The Sample space is: S = {HH, HT, TH, TT},

{HH},{HT},{TH},{TT} are elements.

Example of events:

E_1 = {HT,TH} «occurrence of exactly one head/tail»;

E_2 = {HH, HT} «occurrence of a head in first toss»;

E_3 = {HT,TH,HH} =«occurrence of at least one head».

Example

experiment: cast of two dices and observation of number of spots in first (i), and second (j) dices.

Sample space: S = {(i, j): i, j = 1,2,3,4,5,6}. # S = 6^2 = 36 elements.

Example of Events:

E_1 = {(4,4),(4,5),(5,4),(5,5)} - «occurrence of just 4 or 5 spots in both dices»;

E_2 = {(1,1),(1,2),(1,3),(2,1),(2,2),(3,1)} - «total number of spots less than 5».

Example

Experiment: observation of a light bulb life span (hrs).

Sample space: S = $\{x : x \geq 0\} \subset$ R

Example of Events:

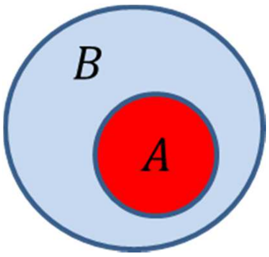
$E_1 = \{x : 17000 < x < 43000\}$ - «life span greater than 17000 and lower than 43000 hours »;

$E_2 = \{x : x \leq 50000\}$ - «life span less than or equal to 50000 hrs»;

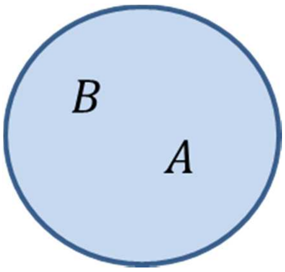
$E_3 = \{x : x > 60000\}$ «life span greater than 60000 hrs».

Let A, B be events in sample space S. In the study of probability, sets and events are interchangeable. Hence

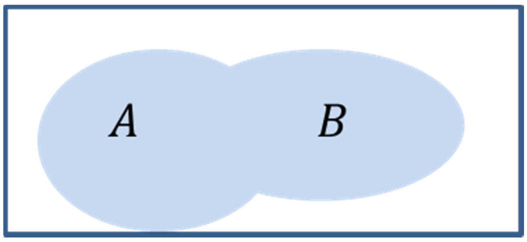
- Event A occurs if and only if B occurs: A $\subset$ B.



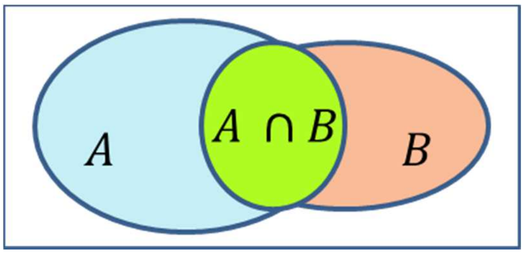
- Event A=Event B if and only if $A \subset B$ *and* $B \subset A$



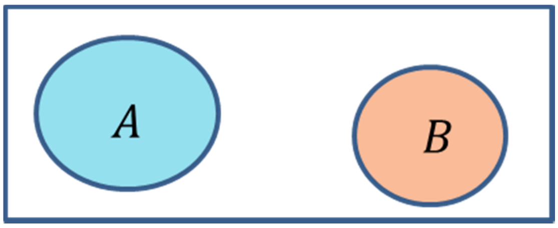
- **Union of events $\boldsymbol{A \cup B}$: $\boldsymbol{A \cup B}$** occurs if and only if either A or B or both occur.



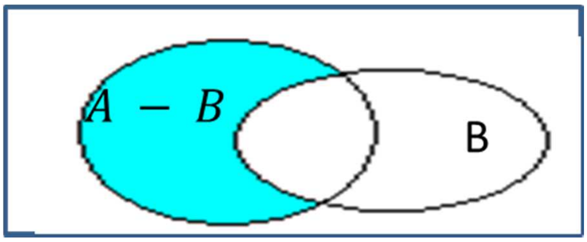
- **Null event** (Ø): event with no elements.
- **Intersection of events** $A \cap B$: is the set of all elements in S that belong to both A and B



- If events A and B have no common elements they are called **mutually exclusive events**. $A \cap B = \emptyset$



- **Difference of events** $A - B$: is the event that occurs if and only if A occurs but not B.



- **Complement** $A' = \mathrm{S} - A$: the set of elements belonging to S but not to A. It is obvious that $A \cap A' = \emptyset$ and $A \cup A' = \mathrm{S}$

Remark: Note that $A - B = A \cap B'$

Some properties of operations over events:

1. Associativity: $A \cup (B \cup C) = (A \cup B) \cup C$; $A \cap (B \cap C) = (A \cap B) \cap C$
2. Commutativity: $A \cup B = B \cup A$; $A \cap B = B \cap A$
3. Distributivity: $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$;

   $$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$
4. De Morgan Laws: $(A \cup B)' = A' \cap B'$; $(A \cap B)' = A' \cup B'$

Notation:

1. $\text{Union of } k \text{ events } A_{1,} A_{2,} \cdots, A_k = \cup_{i=1}^{k} A_i$
2. $\text{Intersection of } k \text{ events } A_{1,} A_{2,} \cdots, A_k \cdots = \cap_{i=1}^{k} A_i$

1.3 Interpretations of the concept of probability

Classical probability concept

Historically, the oldest way of defining probabilities is the **classical probability concept.** It applies when *all possible outcomes are equally likely*, as is the case in most games of chance.

We can then say that *if there are* N *equally likely possibilities, of which one must occur and* n *are regarded as favourable, or as a "success," then the probability of a "success" is given by the ratio n/N*

Example: What is the probability of drawing an ace from an ordinary deck of 52 playing cards? Since there are *n* = 4 aces among the *N* = 52 cards, the probability of drawing an ace is 4/52=1/13

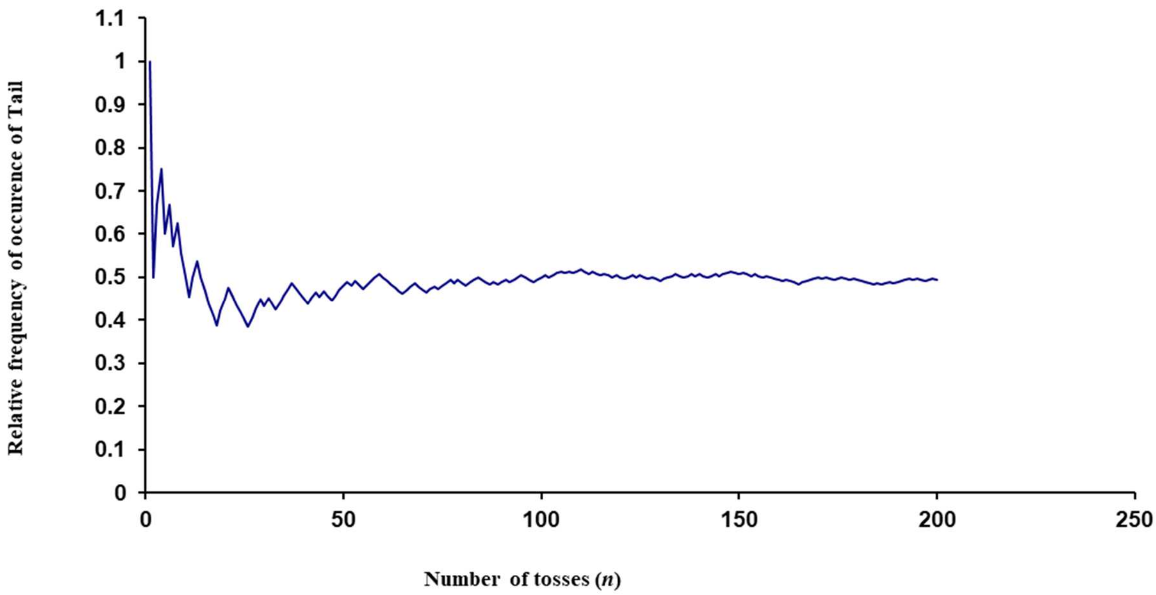
Frequency Interpretation

A major shortcoming of the classical probability concept is that there are many situations in which the possibilities that arise cannot all be regarded as equally likely. This would be the case, for instance, if we are concerned with the question whether it will rain on a given day.

Among the various probability concepts, most widely held is the **frequency interpretation**, according to which *the probability of an event* (*outcome or happening*) is the proportion of the time that events of the same kind will occur in the long run. Being (slightly) more precise we:

- Consider repeating a experiment $\boldsymbol{n}$ times.
- Count the number of times that event $\boldsymbol{A}$ occurred throughout these $\boldsymbol{n}$ repetitions. This number n_A is called the frequency of event $\boldsymbol{A}$.
- The ratio $\frac{n_A}{n}$ is called relative frequency of event $\boldsymbol{A}$.
- The probability of event $\boldsymbol{A}$ is the limit of the relative frequency of event $\boldsymbol{A}$ in a large number of trials: $P(A) = \lim_{n\to\infty} n_A/n$.

Example – A regular coin is tossed 200 times. The number of times that a Tail occurred throughout these $\boldsymbol{n}$ repetitions was counted. The relative frequency of occurence of a Tail calculated for $\boldsymbol{n} = 1,2,\cdots,200$.



1.4 Kolmogorov postulates and properties

The approach to probability that we shall use in this module is the **axiomatic approach**, in which probabilities are defined as "mathematical objects" that behave according to certain well-defined rules.

Any one of the preceding probability concepts, or interpretations, can be used in applications as long as it is consistent with these rules.

These postulates are due to Andrei Kolmogorov (1903-1987) and are known as **Kolmogorov's Probability postulates or axioms .**

Definition: Measure of probability

The probability of an event is a real valued set function *P that assigns to each event A in the sample space* S *($A \subset$ S) a real number $P(A)$, called probability of event A, such that the following postulates are satisfied*:

P1 – $P(A) \geq 0.$

P2 – $P(\mathrm{S}) = 1.$

P3 – If $A_1, A_2, \cdots, A_k$ are events satisfying $A_i \cap A_j = \emptyset$, $i \neq j$, then

$P(A_1 \cup A_2 \cup A_3 \cup \cdots \cup A_k) = P(A_1) + P(A_2) + \cdots + P(A_k).$

for any finite number of events *k.*

In some cases P3 should be replaced by:

P3*: If there is *an infinite but countable number of events* $A_{1,} A_{2,} \cdots, A_{k,} \cdots \Rightarrow \cup_{i=1}^{\infty} A_i$ *that are mutually exclusive events,* $A_i \cap A_j = \emptyset \, (i \neq j)$

then $P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$

Properties of the measure of probability in a sample space S:

1- $P(\emptyset) = 0.$

2-P(A')=1-P(A).

3- If $A \subset B \Rightarrow P(A) \leq P(B).$

4- $0 \leq P(A) \leq 1.$

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

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

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

Remark: The fact that $P(A) = 0$ *does not imply that* $A = \emptyset$.

Exercise: In a large metropolitan area, the probabilities are 0.86, 0.35, and 0.29 respectively that a family (randomly chosen for a sample survey) owns a colour television set, a HDTV set, or both kinds of sets. What is the probability that a family owns either or both kinds of sets?

Exercise: A corporation takes delivery of some new machinery that must be installed and checked before it becomes available. The corporation is sure that it will take no more than 7 days for the installation and check-up to take place.

Let: $A = \{It\ will\ be\ more\ than\ 4\ days\ before\ the\ machinery\ becomes\ available\}$

$B = \{It\ will\ be\ less\ than\ 6\ days\ before\ the\ machinery\ becomes\ available\}$

(a) Describe the event that is the complement of A.

(b) Are events A and B mutually exclusive?

Exercise: If $P(A) = 0.4; P(B) = 0.5; P(A \cup B) = 0.7$

Find $P(A \cap B)$, $P(A - B)$, $P(A')$, $P(A' \cup B')$, $P(A' \cap B')$

Exercise: Let x equal a number randomly selected from the closed interval $[0,1]$. Use your intuition to assign values to:

(a) $P\left(x: 0 \le x \le {}^{1}/_{3}\right)$, (b) $P\left(x: {}^{1}/_{3} \le x \le 1\right)$, (c) $P\left(x: x = {}^{1}/_{3}\right)$

(d) $P\left(x: {}^{1}/_{2} < x < 5\right)$

1.5 Methods of Enumeration

Often the computation of probabilities require that we count outcomes in the sample space. Counting all the outcomes could be very time consuming if we first had to identify every possible outcome. Fortunately, there are tools that make this task easier.

Theorem: Multiplication Principle

If an operation consists of two steps, of which the 1st can be done in n_1 ways and for each of these the 2nd one can be done in n_2 ways, then the whole operation can be done in $n_1 \times n_2$ ways.

Extending to a sequence of more than 2 experiments suppose that a experiment E_i has $m_i, i = 1,2,\cdots,k$ possible elements

Theorem If an operation consists of k steps of which the 1st can be done in n_1 ways, for each of these the 2nd step can be done in n_2 ways, and so forth, then the whole operation can be done in $n_1 \times n_2 \times \cdots \times n_k$ ways.

A helpful graphical representation of a multiple-step experiment is a tree diagram

Example: John has invested in two stocks, Marrley Oil and Cullins Mining. John has determined that the possible outcomes of these investments three months from now are as follows.

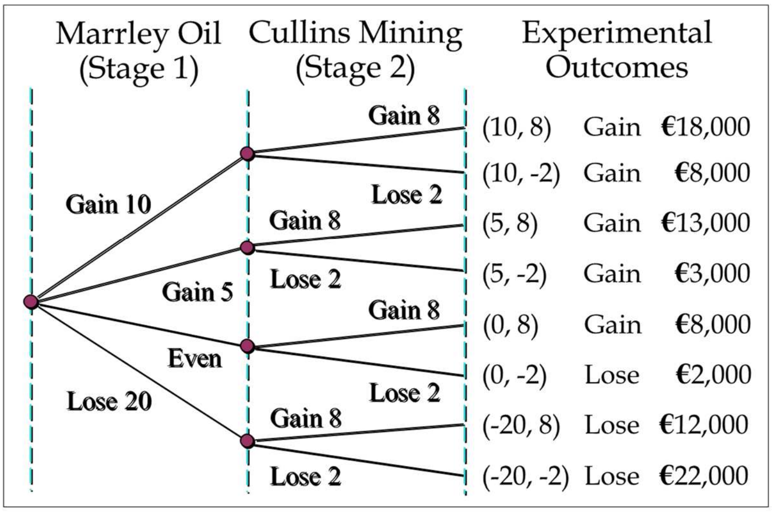
Investment Gain or Loss in 3 Months (in €000)

Marrley Oil	Cullins Mining
10	8
5	–2
0	
–20	

John's Investments can be viewed as a two-step experiment. It involves two stocks, each with a set of experimental outcomes.

Marrley Oil:	$n_1 = 4$
Cullins Mining:	$n_2 = 2$
Total Number of Experimental Outcomes:	$n_1 n_2 = (4)(2) = 8$

Tree Diagram



Definition: Permutations.

A permutation is a distinct arrangement of n different elements of a set

Theorem: The number of permutations of n distinct objects is $n(n-1)(n-2)\cdots(2)(1) = n!$

Remark: We define $0! = 1$

Example: Consider the set {1,2,3}. The possible permutations of this set

Are:

{1,2,3}

{1,3,2}

{2,1,3}

{2,3,1}

{3,1,2}

{3,2,1}

Hence there are 3!=6 permutations.

Permutation of $\boldsymbol{n}$ objects -**ordered sampling is without replacement**: This is n!

Example: The number of permutations of the four-letters a,b,c,d =$4! = 24$

Permutation of $\boldsymbol{n}$ objects (**ordered sampling is with replacement**) : The number of possible arrangements are $\boldsymbol{n^n}$

Example: The number of possible four-letters code words using a,b,c,d =$4^4 = 256$

Ordered sampling without replacement

Theorem: The number of permutations of *n* distinct objects taken *r* at a time is

$$ {}_nP_r = n(n-1)(n-2)\cdots(n-r+1) = \frac{n!}{(n-r)!} $$

Example: {1,2,3,4}, hence we can have the following permutations of 4 distinct objects taken 2 at a time: {1,2}, {1,3}, {1,4}, {2,1}, {2,3}, {2,4}, {3,1}, {3,2}, {3,4}, {4,1}, {4,2}, {4,3}, hence we have ${}_4P_2 = 12$

Exercise: With pieces of cloth of 4 different colours how many distinct three band vertical coloured- flags can one make if the colours can't be repeated?

Theorem The number of combinations of $\boldsymbol{n}$ objects taken $\boldsymbol{k}$ at a time without repetition is

$$ C_k^n = \binom{n}{k} = \frac{n!}{k!\,(n-k)!} $$

C_k^n is also known as Binomial Coefficient

Example: {1,2,3,4}, hence we can have the following combinations of 4 distinct objects taken 2 at a time: {1,2}, {1,3}, {1,4}, {2,3}, {2,4}, {3,4}, hence we have $C_2^4 = 6$.

Exercise: In a restaurant there are 10 different meals. How many subsets of 5 different meals can we make from the 10 meals available?

Theorem: Distinguishable Permutation

Suppose a set of $\boldsymbol{n}$ objects of $\boldsymbol{r}$ distinguishable types. From these $\boldsymbol{n}$ objects, k_1 are similar, k_2 are similar, $\cdots$, k_r are similar, so that $k_1 + k_2 + \cdots + k_r = n$,

The number of distinguishable permutations of these $\boldsymbol{n}$ objects is:

$$ \frac{n!}{k_1! \times k_2! \times \cdots \times k_r!} $$

which is known a **Multinomial Coefficient**

When r =2 , we have the **Binomial Coefficient** $=C_k^n = \frac{n!}{k_1!(n-k_1)!}$.

Exercise: With 9 balls of 3 different colours, 3 black, 4 green and 2 yellow, how many distinguishable groups can one make?

1.6 Conditional probabilities

Example: Two dice are cast, one red and one green. We are interested in the sum of spots of both dice. Let event $\boldsymbol{A}$ = « get a sum of 5 spots»

$$ P(A) = P(Sum\ of\ spots = 5) = 4/36 = 1/9 $$

What is the probability that we get the sum equal to 5 given that we know that the number of spots in green die is 4?

Let event $\boldsymbol{B}$ be the number of spots in the green die is 4. Now probability of event A should be reassessed because the **sample space narrowed** . The occurrence of event $\boldsymbol{A}$ will now happen only if the other die has one spot.

$$P(Sum\ of\ spots = 5 | The\ green\ die\ got\ 4) = 1/6$$

Green\Red	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Definition Conditional Probability

Let $\boldsymbol{A}$ and $\boldsymbol{B}$, be any two events in a sample space S,

$P(\boldsymbol{A}) > 0$. The **conditional probability** of event $\boldsymbol{B}$ given that event $\boldsymbol{A}$ has occurred is denoted by the symbol $P(A|B)$ and is defined by:

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

The **conditional probability is a measure of probability** and verifies the Kolmogorov postulates.

The **conditional probability** can be seen as a reassessment of the probability of an event when there is information about the occurrence of another event. Once one got the new information that event $\boldsymbol{B}$ has occurred, the sample space is no longer S but a subset o S, associated to the occurrence of event $\boldsymbol{A}$.

Exercise : Suppose that we are given 20 tulip bulbs very similar in appearance and told that 8 will bloom early and 13 will be red in accordance with the following combinations:

	Bloom		**Total**
Colour	Early	Late	
Red	5	8	13
Yellow	3	4	7
Total	8	12	20

If one bulb is selected at random, what is the probability that it will produce a red tulip?

Suppose that a close examination of the bulb reveals that it will bloom early. What is the probability that it will produce a red tulip, knowing that it will bloom early?

Theorem: Multiplication rule:

The probability that two events, A and B in sample space S, occur simultaneously is

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

with $P(B) \neq 0$ and/or $P(A) \neq 0$

This rule can easily be generalized to three or more events.

Theorem If A, B and C are any three events in a sample space S such that $P(A \cap B) \neq 0$, then

$$P(A \cap B \cap C) = P[(A \cap B) \cap C] \qquad = P(A \cap B) \times P(C|A \cap B) = P(A) \times P(B|A) \times P(C|A \cap B)$$

Example: Four cards are to be dealt successively at random and without replacement from a deck of playing cards. Compute the probability of receiving in order a spade (S), a heart (H), a diamond (D) and a club(C).

1.7- Independence

Definition: Independence

Let A and B be any two events. These events are said to be **statistically independent** if and only if

$$\boldsymbol{P(A \cap B) = P(A) \times P(B)}$$

From the multiplication rule it follows that

$$P(A|B) = P(A) \; if \; P(B) > 0$$

$$P(B|A) = P(B) if \; P(A) > 0$$

Exercise: Toss a coin twice and observe the sequence of head and tails.

$$\text{S} = \{HH, \; HT, \; TH, TT\}$$

Consider the events:

$$A = \{heads \; on \; the \; first \; toss\} = \{HH, \; HT\}$$

$$B = \{tails \; on \; the \; second \; toss\} = \{HT, \; TT\}$$

$$C = \{Tails \; on \; both \; tosses\} = \{TT\}$$

Show that

- $P(B \cap C) \neq P(B) \times P(C)$so the events B and C are not statistically independent.
- $P(A \cap C) \neq P(A) \times P(C)$ so the events A and C are not statistically independent.
- $P(A \cap B) = P(A) \times P(B)$, so events A and B are statistically independent.

Remarks:

1. Statistically independent events and mutually exclusive events mean different relations between events.
2. If $P(A) = 0$, then A is an event independent from any other event.
3. If $P(A) > 0$ and $P(B) > 0$, and events A and B, are mutually exclusive events then A and B are statistically dependent events, since $P(A \cap B) = 0 \neq P(A) \times P(B)$.

Theorem : If A and B are statistically independent events, A and B', $A'$and B, A' and $B'$are statistically independent events too.

Remark: When three events, A, B and C, are considered we can find situations such that:

1. Events are two by two independent but $P(A \cap B \cap C) \neq P(A) \times P(B) \times P(C)$.
2. $P(A \cap B \cap C) = P(A) \times P(B) \times P(C)$ but A and B or A and C or B and C are statistically dependent events .
3. Events are two by two independent and $P(A \cap B \cap C) = P(A) \times P(B) \times P(C)$.

Exercise 1: A box has four balls (1, 2, 3, 4). Two balls are taken out (each with replacement) from the numbers observed.

Consider the events $\mathrm{A} = \{1,2\}$, $\mathrm{B} = \{1,3\}$, $\mathrm{C} = \{1,4\}$.

Show that:

1. $\mathrm{P(A \cap B) = P(A) \times P(B)}$,
2. $\mathrm{P(A \cap C) = P(A) \times P(C)}$,
3. $\mathrm{P(B \cap C) = P(B) \times P(C)}$ but

$$\mathrm{P(A \cap B \cap C) \neq P(A) \times P(B) \times P(C)}$$

Exercise 2: A regular coin is tossed 3 times. Let A={TTT,HTT,THH,HHH}, B={TTH,THT, HHT, HHH} and C={TTH, THT, HTH, HHH} be events from

$$\mathrm{S = \{TTT, TTH, THT, HTT, THH, HTH, HHT, HHH\}}$$

Show that $P(A \cap B \cap C) = P(A) \times P(B) \times P(C)$ but $\mathrm{P(B \cap C) \neq P(B) \times P(C)}$

Completely independent or mutually independent events

Events A, B and C belonging to the same sample space are **Completely independent events** if and only if:

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

and

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

Exercise 3: A regular coin is tossed 3 times. The sample space is

$$\mathrm{S = \{TTT, TTH, THT, HTT, THH, HTH, HHT, HHH\}}$$

Let A={TTT,HTT,THH,HHH}, B={TTH,THT,HTT,HHH} and C={TTT,HTT,TTH,HHT} be events of the sample space.

Show that

1. $P(A \cap B) = P(A) \times P(B)$;
2. $P(A \cap C) = P(A) \times P(C)$
3. $P(B \cap C) = P(B) \times P(C)$

$4 - P(A \cap B \cap C) = P(A) \times P(B) \times P(C)$

1.8 Total probability theorem and Bayes Theorem

Sample Space Partition

The class of events $\{B_1, B_2, \cdots, B_k, \cdots\}$ is said to be a sample space partition if and only if

$\cup_j B_j = \mathrm{S}$ and $B_i \cap B_j = \emptyset (i \neq j), i, j = 1,2, \cdots, k, \cdots$

then $P(\cup_j B_j) = \sum_j P(B_j) = 1, j = 1,2, \cdots, k, \cdots$

Theorem: Total probability theorem

If $\{B_1, B_2, \cdots, B_k \cdots\}$ is a partition of sample spaceS and $P(B_j) > 0$ $(j = 1,2, \cdots, k, \cdots)$, for any event A,

$$P(A) = \sum_j P(A \cap B_j) = \sum_j P(B_j) \times P(A|B_j)$$

Exercise: The members of a consulting firm rent cars from three rental agencies: 60 percent from agency 1, 30 percent from agency 2, and 10 percent from agency 3. If 9 percent of the cars from agency 1 need an oil change, 20 percent of the cars from agency 2 need an oil change, and 6 percent of the cars from agency 3 need an oil change, what is the probability that a rental car delivered to the firm will need an oil change?

Theorem: Bayes Theorem

If $\{B_1, B_2, \cdots, B_k, \cdots\}$ is a partition of S and if

$P(B_j) > 0\ (j = 1,2,\cdots,k,\cdots)$, for any event A

such that $P(A) > 0$,

$$P(B_j|A) = \frac{P(B_j) \times P(A|B_j)}{\sum_j P(B_j) \times P(A|B_j)}, (j = 1,2,\cdots,k,\cdots)$$

Exercise: A rare but serious disease, *D*, has been found in 0.01 percent of a certain population. A test has been developed that will be positive, *p*, for 98 percent of those who have the disease and be positive for only 3 percent of those who do not have the disease. Find the probability that a person tested as positive does not have the disease.