Module 1
Elementary Probability Theory

1.1. Introduction

Example 1.1.1: The production manager of a bulb manufacturing company wishes to study the effect of a new manufacturing process on the lifetimes of bulbs produced through it.

Here the population under study is

$\mathcal{P}$: Collection of lifetimes of all electric bulbs produced using new manufacturing process

In most practical situations $\mathcal{P}$ is generally large (e.g. collection of lifetimes of all electric bulbs that would be produced using new manufacturing process) and it is not (due to time/cost constraints) to get complete information about $\mathcal{P}$. Thus a representative sample (a sample that in certain sense is a true representative of the population) is taken from $\mathcal{P}$ and using this representative sample inferences regarding various population characteristics of $\mathcal{P}$ (such as population mean, population variance etc.) are made. Note that the sample contains only partial information about $\mathcal{P}$ and the goal is to make inferences about various population characteristics based on partial information in the sample drawn from $\mathcal{P}$.

X: lifetime of a typical electric bulb manufactured using new manufacturing process (a typical element of $\mathcal{P}$)

X is random (called random variable) and its value varies across $\mathcal{P}$ according to some law

<u>Probability Theory</u>: A mathematical tool for modelling uncertainty (e.g. to describe the law according to which values of X vary across $\mathcal{P}$)

Statistics Concerns with procedures for analyzing data (sample) and drawing inferences about various characteristics of the population P

For understanding of statistics, one must have a sound background in probability theory.

The only way to elicit information about any random phenomenon is to perform experiments (e.g. selecting a set of bulbs manufactured by the new manufacturers process and putting them on test for measuring their lifetimes) Each experiment terminates in an outcome which can not be predicted in advance, prior to the performance of experiment (e.g. lifetimes of the bulbs put on test can not be predicted before they are put on test.

Definition 1.1.1 (Random Experiment): A random experiment is an experiment in which

(a) all possible outcomes of the experiment are known in advance;

(b) outcome of a particular performance (trial) of the experiment can not be predicted in advance;

(c) the experiment can be repeated under identical conditions

We will generally denote a random experiment by $\mathcal{E}$.

Definion 1.1.2. (Sample Space) The collection of all possible outcomes of a random experiment is called its sample space. A sample space will usually be denoted by Ω.

Example 1.1.2 (a) $\mathcal{E}$: Tossing a coin once

Sample space $\Omega = \{H, T\}$,

Where H: Heads

T: Tails

(b) $\mathcal{E}$: Casting one red die and one white die

Sample space:

$\Omega = \{(r, w):$ number of spots up on the red die and w is the number of spots up on the white die$\}$

$= \{(1, 1), (1, 2), \ldots, (1, 6), (2, 1), \ldots, (2, 6), \ldots, (6, 1), \ldots (6, 6)\}$

$= \{(i, j): i, j \in \{1, 2, \ldots, 6\}\}$

$= \{1, 2, \ldots, 6\} \times \{1, 2, \ldots, 6\}$

$\longrightarrow$ has 36 elements

(c) $\mathcal{E}$: Putting two electric bulbs produced by ~~a~~ new manufacturing process into test and measuring their lifetimes

Sample space:

$\Omega = \{(x_1, x_2): x_1 \geq 0, x_2 \geq 0\}$

$= [0, \infty)^2$

$\boxed{\text{Definition 1.1.3. (Event)}}$ If the outcome of a random experiment is a member of the set $E \subseteq \Omega$, we say that event E has occurred.

Generally we are interested in specific subsets of Ω, called events. Thus event ~~set~~ space (events under consideration) $\mathcal{F}$ is a subset of power set of Ω. In many situations $\mathcal{F} \equiv \mathcal{P}(\Omega)$, the power set of Ω. For all practical purposes $\mathcal{F} = \mathcal{P}(\Omega)$.

$\boxed{\text{Example 1.1.3.}}$ In Example 1.1.2 (b)

$A = \{(1, 5), (6, 2), (2, 2)\}$

is an event

In Example 1.1.2 (c)

$A = \{(x_1, x_2): x_1 \leq 6, x_2 \geq 8\}$

$= [0, 6] \times [8, \infty)$

may be an event.

The algebra of set theory is applicable in probability theory:

Probability is a measure of uncertainty.

We are interested in quantifying uncertainties associated with various outcomes of a random experiment by assigning probabilities to these outcomes.

Here we will not discuss how probabilities are assigned (which is a part of probability modelling). rather we will discuss properties of probability as a measure.

1.2. Probability Function (or Probability Measure)

$\mathcal{E}$: a random experiment

Ω: Sample space of $\mathcal{E}$

$\mathcal{F}$: Event space

For all practical purposes one may take $\mathcal{F} \equiv P(\Omega)$.

A set function is a function whose domain is a collection of sets (called class of sets)

Definition 1.2.1. (Probability Function or Probability measure)

A probability function (or probability measure) is a real-valued set function, defined on event space $\mathcal{F}$, satisfying the following axioms:

(a) $P(\Omega) = 1$;

(b) $P(E) \geq 0 \quad \forall E \in \mathcal{F}$

(c) If $E_1, E_2 \in \mathcal{F}$ are mutually exclusive / disjoint (i.e. $E_1 \cap E_2 = \phi$ the empty set), then

$$P(E_1 \cup E_2) = P(E_1) + P(E_2).$$

More generally (see property P.2 in the sequel) if $\{E_n\}_{n\geq 1}$ is a sequence of mutually exclusive / disjoint sets in $\mathcal{F}$ ($E_i \cap E_j = \phi$, $\forall (i \neq j)$) then

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i). \quad \text{(Countable additivity)}$$

The triplet $(\Omega, \mathcal{F}, P)$ is called probability space.

Remark 1.2.1. Axioms (b) and (c) are desirable for any measure (such as area, volume, probability etc.). Since the sample space Ω consists of all possible outcomes its occurrence is certain (100% chance of occurrence) and therefore Axiom (a) ($P(\Omega)=1$) is also reasonable.

Elementary Properties of Probability Function/Measure

Let $(\Omega, \mathcal{F}, P)$ be a probability space.

P.1 $P(\phi) = 0$.

Proof. Let $E_1 = \Omega$ and $E_i = \phi$, $i = 2, \ldots$. Then $E_i \cap E_j = \phi$, $\forall i \neq j$, and $\Omega = \bigcup_{i=1}^{\infty} E_i$. Therefore

$$P(\Omega) = P\left(\bigcup_{i=1}^{\infty} E_i\right)$$

$$\Rightarrow \quad 1 = \sum_{i=1}^{\infty} P(E_i) \quad \text{(Axioms (a) and (c))}$$

$$\Rightarrow \quad 1 = \lim_{n\to\infty} \sum_{i=1}^{n} P(E_i)$$

$$\Rightarrow \quad 1 = \lim_{n\to\infty} [P(\Omega) + (n-1)P(\phi)]$$

$$\Rightarrow \quad 1 = 1 + \lim_{n\to\infty} [(n-1)P(\phi)]$$

$$\Rightarrow \quad P(\phi) = 0.$$

P.2. For some natural number n, let $E_1, E_2, \ldots, E_n \in \mathcal{F}$ be mutually exclusive. Then

$$P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} P(E_i).$$

Proof. Let $E_{n+1} = E_{n+2} = \cdots = \phi$. Then $E_i \cap E_j = \phi$ $\forall i \neq j$ and

$$\bigcup_{i=1}^{n} E_i = \bigcup_{i=1}^{\infty} E_i$$

$$\Rightarrow \quad P\left(\bigcup_{i=1}^{n} E_i\right) = P\left(\bigcup_{i=1}^{\infty} E_i\right)$$

$$= \sum_{i=1}^{\infty} P(E_i) \quad \text{(Axiom (c))}$$

$$= \sum_{i=1}^{n} P(E_i) \quad (P(E_i) = P(\phi) = 0 \;\; \forall i = n+1, n+2, \ldots)$$

P.3 $0 \le P(E) \le 1$, $\forall E \in \mathcal{F}$ and
$P(E^c) = 1 - P(E)$, $\forall E \in \mathcal{F}$.

Proof. $\Omega = E \cup E^c$ and $E \cap E^c = \phi$. Therefore

$$P(\Omega) = P(E \cup E^c)$$

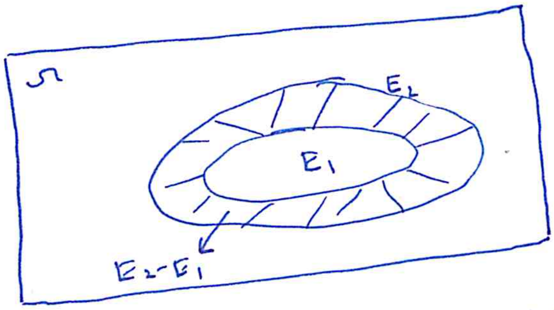
$$\Rightarrow \quad 1 = P(E) + P(E^c) \ge P(E) \quad \text{(using Axioms (a), (b) and P.2)}$$

$$\Rightarrow 0 \le P(E) \le 1 \text{ and}$$

$$\Rightarrow \quad P(E^c) = 1 - P(E).$$

P.4 Let $E_1, E_2 \in \mathcal{F}$ be such that $E_1 \subseteq E_2$. Then
$P(E_2 - E_1) = P(E_2) - P(E_1)$.

Proof.



$$E_2 = E_1 \cup (E_2 - E_1) \quad \text{and} \quad E_1 \cap (E_2 - E_1) = \phi$$

$$\Rightarrow \quad P(E_2) = P(E_1 \cup (E_2 - E_1))$$

$$= P(E_1) + P(E_2 - E_1) \quad \text{(using P.2)}$$

$$\Rightarrow \quad P(E_2 - E_1) = P(E_2) - P(E_1).$$

P.5 Let $E_1, E_2 \in \mathcal{F}$ be such that $E_1 \subseteq E_2$. Then $P(E_1) \le P(E_2)$ (i.e., $P(\cdot)$ is monotone).

Proof. By P.4

$$P(E_2 - E_1) = P(E_2) - P(E_1),$$ which implies that

By Axiom (b) we have $P(E_2 - E_1) \ge 0$,

$$P(E_2) - P(E_1) \ge 0$$

$$\Rightarrow \quad P(E_1) \le P(E_2).$$

~~**P.6.** $0 \le P(E) \le 1, \ \forall E \in \mathcal{F}$.~~

~~**Proof.** Since we have $\phi \subseteq E \subseteq \Omega, \ \forall E \in \mathcal{F}$, using P.5 we get~~

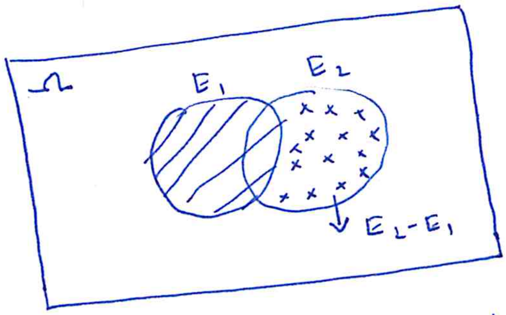
~~$$P(\phi) \le P(E) \le P(\Omega), \ \forall E \in \mathcal{F}$$~~

~~$$\Rightarrow 0 \le P(E) \le 1, \ \forall E \in \mathcal{F}.$$~~

P. 6 Let $E_1, E_2 \in \mathcal{F}$. Then

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2). \quad \text{(Inclusion-Exclusion principle for two events)}$$

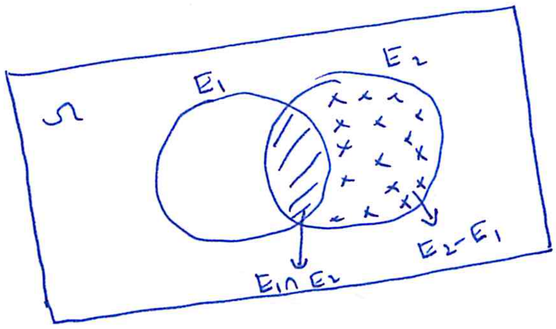
Proof.



$$E_1 \cup E_2 = E_1 \cup (E_2 - E_1) \quad \text{and} \quad E_1 \cap (E_2 - E_1) = \phi$$

$$\Rightarrow \quad P(E_1 \cup E_2) = P(E_1 \cup (E_2 - E_1))$$

$$= P(E_1) + P(E_2 - E_1) \qquad \text{(using P.2)} \ldots \boxed{(1.2.1)}$$



Also

$$E_2 = (E_1 \cap E_2) + (E_2 - E_1) \quad \text{and} \quad (E_1 \cap E_2) \cap (E_2 - E_1) = \phi.$$

Therefore

$$P(E_2) = P((E_1 \cap E_2) + (E_2 - E_1))$$

$$= P(E_1 \cap E_2) + P(E_2 - E_1)$$

$$\Rightarrow \quad P(E_2 - E_1) = P(E_2) - P(E_1 \cap E_2)$$

Using this in equation (1.2.1) we get

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$$

Remark 1.2.1 (a) If $P(A)=0$ and $B \subseteq A$, then $P(B)=0$ (using P.5 and Axiom (b)).

Similary if $P(C)=1$ and $C \subseteq D$, then $P(D)=1$ (using P.5 and P.6).

(b) If $P(D)=1$, then

$P(A) = P(A \cap D)$, $\forall A \in \mathcal{F}$.

Similarly if $P(D)=0$, then

$P(A) = P(A \cap D^c)$, $\forall A \in \mathcal{F}$.

Exercise.

(c) Let $E_1, E_2 \in \mathcal{F}$. Then, using P.7 and Axiom (b)

$P(E_1 \cup E_2) \leq P(E_1) + P(E_2)$ (Boole's inequality for two events)

(d) Let $E_1, E_2 \in \mathcal{F}$. Then using P.7 and P.6 we have

$P(E_1 \cap E_2) \geq \max\{0, P(E_1)+P(E_2)-1\}$

(Bonferroni's inequality for two events)

Theorem 1.2.1. (Inclusion-Exclusion Principle) For $E_1, E_2, \ldots, E_k \in \mathcal{F}$, ($k \geq 2$ is an integer), let

$p_{1,k} = P(E_1) + \ldots + P(E_k) = \sum_{i=1}^{k} P(E_i)$

$p_{2,k} = P(E_1 \cap E_2) + P(E_1 \cap E_3) + \ldots + P(E_1 \cap E_k) + P(E_2 \cap E_3) + \ldots + P(E_2 \cap E_k) + \ldots + P(E_{k-1} \cap E_k)$

$= \sum\sum_{1 \leq i < j \leq k} P(E_i \cap E_j)$ (sum of probabilities of all possible intersections involving 2 events out of the k events $E_1, \ldots, E_k$)

$\vdots$

$p_{i,k} = \sum\sum\cdots\sum_{1 \leq j_1 < j_2 < \cdots < j_i \leq k} P(E_{j_1} \cap E_{j_2} \cap \ldots \cap E_{j_i})$

(sum of probabilities of all possible intersections involving i events out of k events $E_1, \ldots, E_k$, $i=1, \ldots, k$)

Then

$P\left(\bigcup_{i=1}^{k} E_i\right) = p_{1,k} - p_{2,k} + p_{3,k} - p_{4,k} + \ldots + (-1)^{k-1} p_{k,k}$.

$\boxed{\text{Proof.}}$ Note that, for $k=2$,

$$p_{1\,2} = P(E_1) + P(E_2)$$
$$p_{2\,2} = P(E_1 \cap E_2)$$

and

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$$
$$= p_{1\,2} - p_{2\,2}.$$

Thus the result is true for $k=2$.

Now suppose that the result is true for $k = 2, 3, \ldots m$ i.e.

$$P\left(\bigcup_{i=1}^{k} E_i\right) = p_{1\,k} - p_{2,k} + p_{3\,k} + \cdots + (-1)^{k-1} p_{k,k}, \quad \forall\; k = 2, 3 \ldots m$$

Then

$$P\left(\bigcup_{i=1}^{m+1} E_i\right) = P\left(\left(\bigcup_{i=1}^{m} E_i\right) \cup E_{m+1}\right)$$

$$= P\left(\bigcup_{i=1}^{m} E_i\right) + P(E_{m+1}) - P\left(\left(\bigcup_{i=1}^{m} E_i\right) \cap E_{m+1}\right)$$

(using result for $k=2$)

$$= \sum_{j=1}^{m} (-1)^{j-1} p_{j\,m} + P(E_{m+1}) - P\left(\bigcup_{i=1}^{m} (E_i \cap E_{m+1})\right)$$

(using the result for $k=m$ on $\bigcup_{i=1}^{m} E_i$)

$$= \sum_{j=1}^{m} (-1)^{j-1} p_{j\,m} + P(E_{m+1}) - \sum_{j=1}^{m} (-1)^{j} t_{j\,m},$$

(using the result for $k=m$ on $\bigcup_{i=1}^{m} (E_i \cap E_{m+1})$)

where

$$t_{1\,m} = \sum_{i=1}^{m} P(E_i \cap E_{m+1})$$

$$t_{2\,m} = \sum\sum_{1 \le i < j \le m} P(E_i \cap E_j \cap E_{m+1})$$

$$\vdots$$

$$t_{j\,m} = \sum\sum\cdots\sum_{1 \le i_1 < i_2 < \cdots < i_j \le m} P(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_j} \cap E_{m+1}), \quad j = 1, 2, \ldots m$$

$$t_{m\,m} = P(E_1 \cap E_2 \cap \cdots \cap E_m \cap E_{m+1})$$

Therefore

$$P\left(\bigcup_{i=1}^{m+1} E_i\right) = (p_{1,m} + P(E_{m+1})) - (p_{2,m} + t_{1,m}) + (p_{3,m} + t_{2,m})$$
$$+ \cdots + (-1)^{m-1}(p_{m,m} + t_{m-1,m}) + (-1)^m t_{m,m},$$

$$= p_{1,m+1} - p_{2,m+1} + p_{3,m+1} + \cdots + (-1)^{m-1} p_{m,m+1} + (-1)^m p_{m+1,m+1},$$

as

$$p_{1,m} + P(E_{m+1}) = \sum_{j=1}^{m} P(E_j) + P(E_{m+1}) = p_{1,m+1}$$

$$p_{2,m} + t_{1,m} = \sum\sum_{1 \le i < j \le m} P(E_i \cap E_j) + \sum_{i=1}^{m} P(E_i \cap E_{m+1})$$

$$= \sum\sum_{1 \le i < j \le m+1} P(E_i \cap E_j) = p_{2,m+1}$$

$$\vdots$$

$$p_{m,m} + t_{m-1,m} = P(E_1 \cap E_2 \cap \cdots \cap E_m) + \sum\sum\cdots\sum_{1 \le i_1 < i_2 < \cdots < i_{m-1} \le m} P(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_{m-1}} \cap E_{m+1})$$

$$= p_{m,m+1}$$

and

$$t_{m,m} = P(E_1 \cap E_2 \cap \cdots \cap E_m \cap E_{m+1}) = p_{m+1,m+1}$$

The result now follows by induction.

Remark 1.2.3. Let $E_1, E_2, E_3 \in \mathcal{F}$. Then

$$P(E_1 \cup E_2 \cup E_3) = p_{1,3} - p_{2,3} + p_{3,3}$$
$$= (P(E_1) + P(E_2) + P(E_3))$$
$$- (P(E_1 \cap E_2) + P(E_1 \cap E_3) + P(E_2 \cap E_3))$$
$$+ P(E_1 \cap E_2 \cap E_3)$$

Theorem 1.2.2. For some positive integer $k \geq 2$ let $E_1, E_2, \dots, E_k \in \mathcal{F}$. Then

$$p_{1,k} - p_{2,k} \leq P\left(\bigcup_{i=1}^{k} E_i\right) \leq p_{1,k}$$

Proof. Note that for $k=2$

$$p_{1,2} = P(E_1) + P(E_2)$$

$$p_{2,2} = P(E_1 \cap E_2)$$

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2) \leq P(E_1) + P(E_2)$$

$$\Rightarrow \quad p_{1,2} - p_{2,2} = P(E_1 \cup E_2) \leq P(E_1) + P(E_2)$$

Thus the result is true for $k=2$.

Now suppose that for some positive integer $m (\geq 2)$

$$p_{1,k} - p_{2,k} \leq P\left(\bigcup_{i=1}^{k} E_i\right) \leq p_{1,k}, \quad \forall k = 1, 2, \dots, m$$

Then

$$P\left(\bigcup_{i=1}^{m+1} E_i\right) = P\left(\left(\bigcup_{i=1}^{m} E_i\right) \cup E_{m+1}\right)$$

$$\leq P\left(\bigcup_{i=1}^{m} E_i\right) + P(E_{m+1})$$

(using result for $k=2$, $A = \bigcup_{i=1}^{m} E_i$ and $B = E_{m+1}$, $P(A \cup B) \leq P(A) + P(B)$)

$$\leq p_{1,m} + P(E_{m+1})$$

$$= p_{1,m+1} \quad \dots$$

(1.2.2)

Also using the result for $k=m$ we get

$$P\left(\bigcup_{i=1}^{m} E_i\right) \geq p_{1,m} - p_{2,m}$$

and

$$P\left(\bigcup_{i=1}^{m} (E_i \cap E_{m+1})\right) \leq \sum_{i=1}^{m} P(E_i \cap E_{m+1})$$

Thus

$$P\left(\bigcup_{i=1}^{m+1} E_i\right) = P\left(\left(\bigcup_{i=1}^{m} E_i\right) \cup E_{m+1}\right)$$

$$= P\left(\bigcup_{i=1}^{m} E_i\right) + P(E_{m+1}) - P\left(\bigcup_{i=1}^{m} (E_i \cap E_{m+1})\right)$$

$$\geq p_{1,m} - p_{2,m} + P(E_{m+1}) - \sum_{i=1}^{m} P(E_i \cap E_{m+1})$$

$$= \left(p_{1,m} + P(E_{m+1})\right) - \left(p_{2,m} + \sum_{i=1}^{m} P(E_i \cap E_{m+1})\right)$$

$$= p_{1,m+1} - p_{2,m+1} \qquad \cdots \boxed{(1.2.3)}$$

Using (1.2.2) and (1.2.3) we get

$$p_{1,m+1} - p_{2,m+1} \leq P\left(\bigcup_{i=1}^{m+1} E_i\right) \leq p_{1,m+1}$$

and the result follows using principle of mathematical induction.

Remark 1.2.4 One can also show that

$$p_{1,k} - p_{2,k} + p_{3,k} - p_{4,k} \leq P\left(\bigcup_{i=1}^{k} E_i\right) \leq p_{1,k} - p_{2,k} + p_{3,k}$$

$$\vdots$$

$$p_{1,k} - p_{2,k} + p_{3,k} - \cdots + p_{2m,k} \leq P\left(\bigcup_{i=1}^{k} E_i\right) \leq p_{1,k} - p_{2,k} + p_{3,k} - \cdots + p_{2m-1,k},$$

$$m = 1, 2, \ldots, \left[\frac{k}{2}\right]$$

Theorem 1.2.3. (**Bonferroni's Inequality**) Let $E_1, E_2, \ldots, E_k \in \mathcal{F}$.

Then

$$P\left(\bigcap_{i=1}^{k} E_i\right) \geq \max\left\{\sum_{i=1}^{k} P(E_i) - (k-1), 0\right\}$$

Proof. We have

$$P\left(\bigcap_{i=1}^{k} E_i\right) = P\left(\left(\bigcup_{i=1}^{k} E_i^c\right)^c\right) \qquad \text{(De-Morgan's law)}$$

$$= 1 - P\left(\bigcup_{i=1}^{k} E_i^c\right)$$

$$\geq 1 - \sum_{i=1}^{k} P(E_i^c) \qquad \text{(Boole's inequality)}$$

$$= 1 - \sum_{i=1}^{k} [1 - P(E_i)]$$

$$= \sum_{i=1}^{k} P(E_i) - (k-1) \qquad \cdots \boxed{(1.2.4)}$$

Also

$$P\left(\bigcap_{i=1}^{k} E_i\right) \geq 0 \qquad \cdots \boxed{(1.2.5)}$$

Combining (1.2.4) and (1.2.5) we get

$$P\left(\bigcap_{i=1}^{k} E_i\right) \geq \max\left\{\sum_{i=1}^{k} P(E_i) - (k-1), 0\right\}$$

$\boxed{\text{Example 1.2.1}}$ Random experiment: Casting a red and a white die

Sample Space: $\Omega = \{(i, j): i \in \{1, 2, \cdots, 6\},\ j \in \{1, 2, \cdots, 6\}\}$

For $(i, j) \in \Omega$

i: number of spots up on the red die;

j: number of spots up on the white die

Event space $\mathcal{F} \equiv$ power set of Ω.

For $E \in \mathcal{F}$, define $Q: \mathcal{F} \to \mathbb{R}$ as

$$Q(E) = \frac{|E|}{36}, \quad \text{where } |E| = \#\text{ of elements in } E.$$

Then

(a) $Q(\Omega) = \frac{|\Omega|}{36} = \frac{36}{36} = 1$

(b) $Q(E) = \frac{|E|}{36} \geq 0, \quad \forall E \in \mathcal{F}$

(c) for mutually exclusive events $E_1, E_2, \ldots$

$$Q\left(\bigcup_{i=1}^{\infty} E_i\right) = \frac{\left|\bigcup_{i=1}^{\infty} E_i\right|}{36}$$

$$= \frac{\sum_{i=1}^{\infty} |E_i|}{36} \qquad \left(E_i\text{s are disjoint} \Rightarrow \left|\bigcup_{i=1}^{\infty} E_i\right| = \sum_{i=1}^{\infty} |E_i|\right)$$

$$= \sum_{i=1}^{\infty} \frac{|E_i|}{36}$$

$$= \sum_{i=1}^{\infty} Q(E_i)$$

Thus $(\Omega, Q, \mathcal{F})$ is a probability space

Equally Like Probability Models for Finite Sample Space

Suppose that the sample space

$$\Omega = \{\omega_1, \omega_2, \dots, \omega_k\}$$

is finite (has k elements). Here singletons $\{\omega_i\}$ are called elementary events and

$$\Omega = \bigcup_{i=1}^{k} \{\omega_i\}.$$

Suppose that

$$P(\{\omega_i\}) = \frac{1}{k}, \quad i = 1, \dots, k \quad \left(\text{each elementary event is equally likely}\right)$$

For any event $E \subseteq \Omega$, we have

$$E = \{\omega_{i_1}, \omega_{i_2}, \dots, \omega_{i_r}\}$$

for some $i_1, \dots, i_r \in \{1, 2, \dots, k\}$ $1 \le r \le k$. Then $E = \bigcup_{j=1}^{r} \{\omega_{i_j}\}$

$$\begin{aligned}
P(E) &= P\left(\bigcup_{j=1}^{r} \{\omega_{i_j}\}\right) \\
&= \sum_{j=1}^{r} P(\{\omega_{i_j}\}) \\
&= \sum_{j=1}^{r} \frac{1}{k} \\
&= \frac{r}{k} \\
&= \frac{\#\text{ of ways favorable to event } E}{\text{total number of ways in which the random experiment can terminate}}
\end{aligned}$$

Here the assumption of equally likely ($P(\{\omega_i\}) = \frac{1}{k}$, $i = 1, \dots, k$) is a part of probability modelling.

"At random" In a random experiment with finite sample space S whenever we say that the experiment has been performed at random it means that all the outcomes in the sample space are equally likely.

Example 1.2.2. (Birthday Problem) Suppose that a college has n students, including you. Each of them were born on non-leap years.

(a) Find the probability that at least two of them have the same birthday. For what values of n this probability is more than 0.5, 0.8, 0.95?

(b) For what value of n the probability that you will find someone who shares your birthday is $\frac{1}{2}$?

Solution

(a) Required probability = 1 − P(all of them have different birthdays)

$$= 1 - \frac{365 \times 364 \times \cdots \times (365-n+1)}{365^n}$$

(b) Required probability = 1 − P(no one shares the same birthday as mine)

$$= 1 - \frac{364^n}{365^n}$$

For $1 - \frac{364^{n-1}}{365^{n-1}} \approx 0.5$

$n \approx 253$

Example 1.2.3. Five cards are drawn at random and without replacement from a deck of 52 cards. Find the probability that:

(i) each card is spade (event E_1);

(ii) at least one card is spade (event E_2)

(iii) exactly tree cards are king and two cards are queen (event E_3)

(iv) exactly two kings, two queens and one jack are drawn.

Solution (i) $P(E_1) = \dfrac{\binom{13}{5}}{\binom{52}{5}}$

(ii) $P(E_2) = 1 - P(E_2^c)$

$= 1 - P(\text{no card is spade})$

$= 1 - \dfrac{\binom{39}{5}}{\binom{52}{5}}$

(iii) $P(E_3) = \dfrac{\binom{4}{3}\binom{4}{2}}{\binom{52}{5}}$

(iv) $P(E_4) = \dfrac{\binom{4}{2}\binom{4}{2}\binom{4}{1}}{\binom{52}{5}}$.

Example 1.2.4. (Capture/Recapture Method for estimating population size)

In a wildlife population suppose that the population size n is unknown. To estimate the population size n, 20 animals are captured, tagged and then released back. Thereafter 40 animals are captured at random and it is found that 8 of them are tagged. Find an estimate of the population size n based on the given data.

Solution We have

$n =$ total # of animals

$20 =$ # of tagged animals in the population

$n-20 =$ # of untagged animals in the population

Data: Sample of 40 animals yield

# of tagged animals $= 8$

# of untagged animals $= 32$

The probability of obtaining this data is

$$l(n) = \frac{\binom{20}{8}\binom{n-20}{32}}{\binom{n}{40}}, \quad 0 \le n-20, \; 0 \le n, \quad n-20 \ge 32, \; n \ge 40$$

$$= \frac{\binom{20}{8}\binom{n-20}{32}}{\binom{n}{40}}, \quad n \ge 52$$

$$l(n+1) > l(n) \iff \frac{\binom{n-19}{32}}{\binom{n+1}{40}} > \frac{\binom{n-20}{32}}{\binom{n}{40}}$$

$$\iff \frac{\lfloor n-19}{\lfloor 32 \; \lfloor n-51} \times \frac{\lfloor 40 \; \lfloor n-39}{\lfloor n+1} > \frac{\lfloor n-20}{\lfloor 32 \; \lfloor n-52} \times \frac{\lfloor 40 \; \lfloor n-40}{\lfloor n}$$

$$\iff \frac{(n-19)}{(n-51)(n+1)} > 1$$

$$\iff n < 99$$

$\iff$

Similarly

$$l(n+1) < l(n) \iff n > 99$$

Thus l is maximized at $n=99$, i.e. for $n=99$ the observed data (among 40 captured animals 8 are tagged and 32 are untagged) is most probable.

Thus an estimate of n is

$$\hat{n} = 99 \quad \text{(Maximum likelihood estimator)}$$

1.3. Conditional Probability

Consider a probability space $(\Omega, \mathcal{F}, P)$, where $\Omega = \{\omega_1, \dots, \omega_n\}$ is finite and

$$P(\{\omega_i\}) = \frac{1}{n}, \quad i = 1, \dots, n \quad \text{(Equally likely probability model)}$$

Then, for any $C \in \mathcal{F}$

$$P(C) = \frac{\text{\# of cases favorable to } C}{\text{total \# of cases}}$$

$$= \frac{|C|}{|\Omega|} = \frac{|C|}{n}$$

Now suppose it is known apriori that event A has occurred (i.e. outcome of the experiment is an element of A), where $|A| \geq 1$. (so that $P(A) = \frac{|A|}{n} > 0$). Given this prior information (that the event A has occurred) we want to define probability function, say $P(B|A)$ on event space $\mathcal{F}$. A natural way to define $P(B|A)$ is

$$P(B|A) = \frac{|A \cap B|}{|A|} = \frac{|A \cap B|/n}{|A|/n} = \frac{P(A \cap B)}{P(A)}, \quad B \in \mathcal{F}.$$

Definition 1.3.1 Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $A \in \mathcal{F}$ be such that $P(A) > 0$. Then

$$P(B|A) = \frac{P(A \cap B)}{P(A)}, \quad B \in \mathcal{F}.$$

is called the conditional probability of event B given the event.

Remark 1.3.1. (a) In the above definition the event A (with $P(A) > 0$) is fixed and for this fixed $A \in \mathcal{F}$, $P(\cdot|A)$ is a set function defined on $\mathcal{F}$.

(b) $P(A \cap B) = P(A)\,P(B|A) = P(B)\,P(A|B)$, $A, B \in \mathcal{F}$.

Theorem 1.3.1 Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $A \in \mathcal{F}$, with $P(A) > 0$, be fixed. Then $P(\cdot|A) : \mathcal{F} \to \mathbb{R}$ is a probability function (called the conditional probability function) on $\mathcal{F}$ (so that $(\Omega, \mathcal{F}, P(\cdot|A))$ is a prob. space).

Proof Clearly

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \geq 0, \quad \forall B \in \mathcal{F}$$

and

$$P(\Omega|A) = \frac{P(A \cap \Omega)}{P(A)} = 1.$$

Let $\{B_n\}_{n \geq 1}$ be a sequence of disjoint events in $\mathcal{F}$. Then

$$P\left(\bigcup_{n\geq 1}^{\infty} B_n \Big| A\right) = \frac{P\left(\left(\bigcup_{n\geq 1}^{\infty} B_n\right) \cap A\right)}{P(A)}$$

$$= \frac{P\left(\bigcup_{n\geq 1}^{\infty} (B_n \cap A)\right)}{P(A)}.$$

Since $\{B_n\}_{n\geq 1}$ are disjoint their subsets $\{B_n \cap A\}_{n\geq 1}$ are also disjoint. Since $P(\cdot)$ is a prob. measure, we get

$$P\left(\bigcup_{n\geq 1}^{\infty} B_n \Big| A\right) = \frac{\sum_{n\geq 1}^{\infty} P(B_n \cap A)}{P(A)} = \sum_{n\geq 1}^{\infty} \frac{P(B_n \cap A)}{P(A)}$$

$$= \sum_{n\geq 1}^{\infty} P(D_n|A)$$

It follows that $P(\cdot|A)$ is a prob. function on $\mathcal{F}$, for any fixed $A\in\mathcal{F}$ with $P(A)>0$).

Example 1.3.1. Five cards are dealt at random (without replacement) from a deck of 52 cards. Define events

B: all spade hand
A: atleast 4 spades in hand

Find $P(B|A)$.

Solution We have

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

$$= \frac{P(B)}{P(A)} \quad (\text{since } B\subseteq A)$$

$$= \frac{\binom{13}{5}\Big/\binom{52}{5}}{\left[\binom{13}{4}\binom{39}{1}+\binom{13}{5}\right]\Big/\binom{52}{5}} = 0.441.$$

Remark 1.3.2. (Multiplication Law)

(i) $P(A\cap B) = P(A)\, P(B|A)$, if $P(A)>0$

(ii) $P(A\cap B\cap C) = P(A\cap B)\, P(C|A\cap B)$
$= P(A)\, P(B|A)\, P(C|A\cap B)$,
provided $P(A\cap B)>0$ (which ensures that $P(A)>0$)

(iii) Using principle of mathematical induction, we have

$$P\left(\bigcap_{i=1}^{n} C_i\right) = P(C_1\cap C_2\cap\cdots\cap C_n)$$
$$= P(C_1\cap C_2\cap\cdots\cap C_{n-1})\, P(C_n|C_1\cap C_2\cap\cdots\cap C_{n-1})$$
$$= P(C_1\cap C_2\cap\cdots\cap C_{n-2})\, P(C_{n-1}|C_1\cap C_2\cap\cdots\cap C_{n-2})\, P(C_n|C_1\cap C_2\cap\cdots\cap C_{n-1})$$
$$\vdots$$
$$= P(C_1)\, P(C_2|C_1)\, P(C_3|C_1\cap C_2)\cdots P(C_n|C_1\cap C_2\cap\cdots\cap C_{n-1})$$

provided $P(C_1\cap C_2\cap\cdots\cap C_{n-1})>0$ (which also ensures that $P(C_1\cap C_2\cap\cdots\cap C_i)>0$, $i=1,\ldots,n-2$)

Due to symmetry, if $(\alpha_1, \dots, \alpha_n)$ is a permutation of $(1, \dots, n)$ then

$$P\left(\bigcap_{i=1}^{n} C_i\right) = P(C_{\alpha_1} \cap C_{\alpha_2} \cap \dots \cap C_{\alpha_n})$$

$$= P(C_{\alpha_1})\, P(C_{\alpha_2} | C_{\alpha_1})\, P(C_{\alpha_3} | C_{\alpha_1} \cap C_{\alpha_2}) \dots P(C_{\alpha_n} | C_{\alpha_1} \cap C_{\alpha_2} \cap \dots \cap C_{\alpha_{n-1}})$$

provided $P(C_{\alpha_1} \cap C_{\alpha_2} \cap \dots \cap C_{\alpha_{n-1}}) > 0$ (which also ensures that $P(C_{\alpha_1} \cap C_{\alpha_2} \cap \dots \cap C_{\alpha_i}) > 0$, $i = 1, \dots, n-2$).

Example 1.3.2. A bowl contains 3 red and 5 blue chips. All chips that are of the same colour are identical. Two chips are drawn successively at random and without replacement. Define events

A: first draw resulted in a red chip

B: second draw resulted in a blue chip.

Find $P(A \cap B)$, $P(A)$ and $P(B)$.

Solution

$P(A) = \frac{3}{8}$, $P(B|A) = \frac{5}{7}$, $P(B) = P(A \cap B) + P(A^c \cap B)$

$= P(B|A)\, P(A) + P(B|A^c)\, P(A^c)$

$= 5/7 \times 3/8 + 4/7 \times 5/8 = 35/56$

Note that here the outcome of second draw is dependent on outcome of first draw ($P(B|A) \neq P(B)$).

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

$$= \frac{3}{8} \times \frac{5}{7} = 0.2679$$

Theorem 1.3.2. (Theorem of Total Probability) For a countable set Λ (that is elements of Λ can either be put in 1-1 correspondence with $\{1, 2, \dots, n\}$ for some $n \in \mathbb{N}$), or with $\mathbb{N} = \{1, 2, \dots\}$

let $\{E_\alpha : \alpha \in \Lambda\}$ be a countable collection of mutually exclusive (i.e. $E_\alpha \cap E_\beta = \phi \ \forall \alpha \neq \beta$) and exhaustive (i.e. $P(\bigcup_{\alpha \in \Lambda} E_\alpha) = 1$) events. Then, for any $E \in \mathcal{F}$,

$$P(E) = \sum_{\alpha \in \Lambda} P(E \cap E_\alpha) = \sum_{\substack{\alpha \in \Lambda \\ P(E_\alpha) > 0}} P(E|E_\alpha)\, P(E_\alpha).$$

Proof. Since $P(\bigcup_{\alpha\in\Lambda} E_\alpha)=1$, we have

$$P(E) = P(E \cap (\bigcup_{\alpha\in\Lambda} E_\alpha))$$

$$= P(\bigcup_{\alpha\in\Lambda} (E\cap E_\alpha))$$

$$= \sum_{\alpha\in\Lambda} P(E\cap E_\alpha) \quad \left(\begin{array}{l} E_\alpha\text{'s are disjoint} \Rightarrow \text{their subsets} \\ (E\cap E_\alpha)\text{'s are disjoint} \end{array} \right)$$

$$= \sum_{\substack{\alpha\in\Lambda \\ P(E_\alpha)>0}} P(E\cap E_\alpha) \qquad \left(P(E_\alpha)=0 \Rightarrow P(E\cap E_\alpha)=0, \ \alpha\in\Lambda \right)$$

$$= \sum_{\substack{\alpha\in\Lambda \\ P(E_\alpha)>0}} P(E\,|\,E_\alpha)\,P(E_\alpha)$$

Example 1.3.3. A population comprises of 40% female and 60% female. Suppose that 15% of female and 30% of male in the population smoke. A person is selected at random from the population.

(a) Find the probability that he/she is a smoker

(b) Given that the selected person is smoker, find the probability that he is male.

Solution Define the events

M: selected person is a male

$F = M^c$: selected person is a female

S: Selected person is a smoker

$T = S^c$: Selected person is a non-smoker.

We have

$P(F) = 0.4$, $P(M) = 0.6$, $P(F\cup M) = P(F)+P(M) = 1.0$

$P(S|F) = 0.15$, $P(T|F) = 0.85$

$P(S|M) = 0.30$ $P(T|M) = 0.70$

(a) By Theorem of total probability

$$P(S) = P(S \cap F) + P(S \cap M)$$

$$= P(S|F)\,P(F) + P(S|M)\,P(M)$$

$$= 0.15 \times 0.4 + 0.30 \times 0.6$$

$$= 0.06 + .18 = 0.24$$

(b) $$P(M|S) = \frac{P(M \cap S)}{P(S)} = \frac{P(S|M)\,P(M)}{P(S)} = \frac{0.30 \times 0.60}{0.24}$$

$$= \frac{0.18}{0.24} = \frac{3}{4}.$$

Theorem 1.3.3. (Bayes' Theorem) Let $\{E_\alpha : \alpha \in \Lambda\}$ be a countable collection of mutually exclusive and exhaustive events and let E be any event with $P(E) > 0$. Then, for any $j \in \Lambda$ with $P(E_j) > 0$,

$$P(E_j|E) = \frac{P(E_j)\,P(E|E_j)}{\sum\limits_{\substack{\alpha \in \Lambda \\ P(E_\alpha) > 0}} P(E_\alpha)\,P(E|E_\alpha)}.$$

Proof. For $j \in \Lambda$

$$P(E_j|E) = \frac{P(E_j \cap E)}{P(E)}$$

$$= \frac{P(E_j)\,P(E|E_j)}{\sum\limits_{\substack{\alpha \in \Lambda \\ P(E_\alpha) > 0}} P(E_\alpha)\,P(E|E_\alpha)} \quad \text{(using Theorem of total probability)}$$

Remark 1.3.3. (a) Suppose that occurrence of any of the mutually exclusive and exhaustive events $\{E_\alpha : \alpha \in \Lambda\}$ (where Λ is a countable set) may cause the occurrence of an event E. Given that the event E has occurred (i.e. given the effect) the Bayes' Theorem provides the conditional probability that the event E (effect) is caused by occurrence of event E_j, $j \in \Lambda$.

(b) In Bayes' Theorem

$\{P(E_j) : j \in \Lambda\}$ are called prior probabilities

and

$\{P(E_j|E) : j \in \Lambda\}$ are called posterior probabilities.

Example 1.3.4. Bowl C_1 contains 3 red and 7 blue chips
Bowl C_2 contains 8 red and 2 blue chips
Bowl C_3 contains 5 red and 5 blue chips
All chips of the same colour are identical.

A die is cast and a bowl is selected as per the following scheme:

- Bowl C_1 is selected if 5 or 6 spots show on the upper side
- Bowl C_2 is selected if 2, 3 or 4 spots show on the upper side
- Bowl C_3 is selected if 6 spots show on the upper face.

The selected bowl is handed over to another person who drawn two chips at random from this bowl. Find the probability that:

(a) two red chips are drawn.

(b) Given that drawn chips are both red, find the probability that it came from bowl C_3.

Solution Define the events

A_i: selected bowl is C_i $i = 1, 2, 3$

R: the chips drawn from the selected bowl are both red.

Then

$P(A_1) = \frac{2}{6} = \frac{1}{3}$; $P(A_2) = \frac{3}{6} = \frac{1}{2}$, $P(A_3) = \frac{1}{6}$

$\{A_1, A_2, A_3\}$ are mutually exclusive and exhaustive.

(a) $P(R) = P(R|A_1)P(A_1) + P(R|A_2)P(A_2) + P(R|A_3)P(A_3)$

$$= \frac{\binom{3}{2}}{\binom{10}{2}} \frac{1}{3} + \frac{\binom{8}{2}}{\binom{10}{2}} * \frac{1}{2} + \frac{\binom{5}{2}}{\binom{10}{2}} \times \frac{1}{6}$$

$$= \frac{1}{6}\left[\frac{2\times 3}{45} + \frac{3\times 28}{45} + \frac{10}{45}\right]$$

$$= \frac{1}{6} \times \frac{100}{45} = \frac{10}{27}$$

(b) $P(A_3|R) = \dfrac{P(R|A_3)P(A_3)}{P(R)} = \dfrac{\binom{5}{2}/\binom{10}{2} \times \frac{1}{6}}{10/27}$

$$= \frac{10}{45} \times \frac{1}{6} \times \frac{27}{10} = \frac{1}{10}.$$

$\boxed{\text{Remark 1.3.4.}}$ In the above example

$$P(A_1|R) = \frac{P(R|A_1)P(A_1)}{P(R)} = \frac{\binom{3}{2}/\binom{10}{2} \times \frac{1}{3}}{10/27} = \frac{3}{45} \times \frac{1}{3} \times \frac{27}{10} = \frac{3}{50}$$

$$P(A_2|R) = \frac{P(R|A_2)P(A_2)}{P(R)} = \frac{\binom{8}{2}/\binom{10}{2} \times \frac{1}{2}}{\frac{10}{27}} = \frac{28}{45} \times \frac{1}{2} \times \frac{27}{10} = \frac{21}{25}$$

$P(A_1|R) = \frac{3}{50} < \frac{1}{3} = P(A_1) \Leftrightarrow P(A_1 \cap R) < P(A_1)P(R)$

$\Leftrightarrow$ R has negative information about A_1

$P(A_2|R) = \frac{21}{25} > \frac{1}{2} = P(A_2) \Leftrightarrow P(A_2 \cap R) > P(A_2)P(R)$

$\Leftrightarrow$ R has positive information about A_2

$P(A_3|R) = \frac{1}{10} < \frac{1}{6} = P(A_3) \Leftrightarrow P(A_3 \cap R) < P(A_3)P(R)$

$\Leftrightarrow$ R has negative information about A_3

Note that

proportion of red chips in C_2 > proportion of red chips in C_i, $i = 1, 3$.

Definition 1.3.2 Let $\{E_j : j \in \Delta\}$ be a collection of events.

(a) Events $\{E_j : j \in \Delta\}$ are said to be pairwise independent if for any pair of events E_α and E_β ($\alpha, \beta \in \Delta$, $\alpha \neq \beta$) in the collection $\{E_j : j \in \Delta\}$ we have

$$P(E_\alpha \cap E_\beta) = P(E_\alpha) P(E_\beta)$$

(b) Events $\{E_1, \dots, E_n\}$ are said to be independent if for any subcollection $\{E_{\alpha_1}, \dots, E_{\alpha_k}\}$ of $\{E_1, \dots, E_n\}$ ($k = 2, 3, \dots, n$)

$$P\left(\bigcap_{j=1}^{k} E_{\alpha_j}\right) = \prod_{j=1}^{k} P(E_{\alpha_j})$$

(c) Let $\Delta \subseteq \mathbb{R}$ be an arbitrary index set so that $\{E_\alpha : \alpha \in \Delta\}$ is an arbitrary collection of events. Events $\{E_\alpha : \alpha \in \Delta\}$ are said to be independent if any finite subcollection of events in $\{E_\alpha : \alpha \in \Delta\}$ forms a collection of independent events.

Theorem 1.3.4. Let $E_1, E_2, \dots$ be a collection of independent events. Then

$$P\left(\bigcap_{k=1}^{\infty} A_k\right) = \prod_{k=1}^{\infty} P(A_k)$$

Proof Let

$$B_n = \bigcap_{k=1}^{n} A_k, \quad n = 1, 2, \dots$$

Then $B_n \downarrow$. Using Problem 5 (b), Assignment-I,

$$P\left(\bigcap_{n=1}^{\infty} B_n\right) = \lim_{n\to\infty} P(B_n)$$

But $\bigcap_{n=1}^{\infty} B_n = \bigcap_{k=1}^{\infty} A_k$ and $P(B_n) = P\left(\bigcap_{k=1}^{n} A_k\right) = \prod_{k=1}^{n} P(A_k)$. Thus

$$P\left(\bigcap_{k=1}^{\infty} A_k\right) = \lim_{n\to\infty} \prod_{k=1}^{n} P(A_k) = \prod_{k=1}^{\infty} P(A_k).$$

Remark 1.3.5. (a) To verify that n events $E_1, \ldots, E_n$ are independent one must verify

$$\binom{n}{2} + \binom{n}{3} + \cdots + \binom{n}{n} = 2^n - n - 1$$

Conditions. For example to conclude that three events E_1 E_2 and E_3 are independent the following 4 (2^3-3-1) Conditions must be verified:

$P(E_1 \cap E_2) = P(E_1 \wedge E_2)$; $\qquad P(E_1 \cap E_3) = P(E_1)\,P(E_3)$

$P(E_2 \cap E_3) = P(E_2 \cap E_3)$; $\qquad P(E_1 \cap E_2 \cap E_3) = P(E_1)\,P(E_2)\,P(E_3)$

(b) Any subcollection of independent events is of independent events. In particular the independence of a collection of events implies their pairwise independence.

(c) If E_1 and E_2 are independent events ($P(E_1)>0$, $P(E_2)>0$), then

$$P(E_1 | E_2) = \frac{P(E_1 \cap E_2)}{P(E_2)} = \frac{P(E_1)\,P(E_2)}{P(E_2)} = P(E_2)$$

i.e. $P(E_1|E_2) = P(E_1)$ (Conditional probability of E_1 given E_2 is the same as unconditional probability of E_1)

Similarly if E_1 E_2 and E_3 are independent events then

$$P(E_1 | E_2 \cap E_3) = P(E_1)$$

Example 1.3.5 Consider the probability space $(\Omega, \mathcal{F}, P)$ with $\Omega = \{1, 2, 3, 4\}$ and $P(\{i\}) = \frac{1}{4}$, $i = 1, 2, 3, 4$. Let $A = \{1, 4\}$, $B = \{2, 4\}$ and $C = \{3, 4\}$. Then show that A, B and C are pairwise independent but not independent.

Solution We have

$$P(A) = P(B) = P(C) = \frac{1}{2}$$

$$P(A \cap B) = P(A \cap C) = P(B \cap C) = P(\{4\}) = \frac{1}{4}.$$

Thus

$P(A \cap B) = P(A)\,P(B)$; $\quad P(A \cap C) = P(A)\,P(C)$, & $P(B \cap C) = P(B)\,P(C)$, implying that A, B and C are pairwise independent.

However

$P(A \cap B \cap C) = P(\{4\}) = \frac{1}{4} \neq \frac{1}{8} = P(A)P(B)P(C)$

$\Rightarrow$ A, B and C are not independent although they are pairwise independent.

$\boxed{\text{Example 1.3.6}}$ Let $E_1, E_2, \dots, E_n$ be a collection of independent events. Show that,

(a) for any permutation $(\alpha_1, \dots, \alpha_n)$ of $\{1, \dots, n\}$, $E_{\alpha_1}, \dots, E_{\alpha_n}$ are independent;

(b) $E_1, \dots, E_k, E_{k+1}^c, \dots, E_n^c$ are independent for any $k \in \{0, 1, \dots, n\}$

(c) E_1^c and $E_2 \cup E_3^c \cup E_5$ are independent

(d) $E_1 \cup E_2^c$, E_3^c and $E_4 \cap E_5^c$ are independent.

$\boxed{\text{Remark 1.3.6}}$ When we say that the two random experiments are performed independently it means that the events associated with two random experiments are independent.