# WEEK 2

## 2. Basic Set Theory and Probability

The goal of statistics (or any other science) is to understand the (real) world. Here, understanding means that we want to get some information about the unknown of the population which we call parameter. The parameter/parameters) are unknown non-measurable real values and we will never know the actual values. Therefore, we perform experiments to get some information.

As it is mentioned before, statistics helps us to solve three problems in the nature.

- collection of data
- organization (or analysis)
- Interpretation

We do not have a chance to reach all the individuals of the population. Therefore, we take a sample (by experiment) and based on this sample, we try to get some information about the population parameter/parameters (estimation, hypothesis testing, etc.). The parameters usually mean ($\mu$) and variance ($\sigma^2$).

**Definition:** A set of all possible outcomes of an experiment is called the ***sample space*** and denoted by either $\Omega$ or $S$. Any subset of the sample space will be called an ***EVENT***.

Since, we do not have a chance to reach all the individuals of the populations, we work on a subset of the sample space. Sample values are observed by an experiment. At the same conditions, if you repeat the experiment, it is possible to observe different sample values.

| | |
|---|---|
| | Here $\Omega$ is the sample space which is a set of all possible outcomes of an experiment. Any subset of the sample space is an event. |
| Figure 2.1 | |

In the following, we will summarize basic set theory. Here, $\Omega$ is the sample space (the whole set).

**Subset:** $A \subset \Omega$

| | |
|---|---|
| | $A \subset \Omega$ : $A$ is a subset of $\Omega$ means that any element in $A$ is also an element of $\Omega$. Mathematically,<br><br>$A \subset \Omega$ if $\forall x \in A \Rightarrow x \in \Omega$ |
| Figure 2.2 | |

**Complement:** $A^c$

| | |
|---|---|
| | Let $A \subset \Omega$, the complement of $A$ is a set of all points that they are not in $A$. Mathematically,<br><br>$A^c = \{x \in \Omega : x \notin A\}$ |
| Figure 2.3 | |

**Union:** $A \cup B$

| | |
|---|---|
| | Let $A, B \subset \Omega$ . The union of $A$ and $B$ is a set of elements in $\Omega$ such that the elements are either $A$ or $B$. Mathematically,<br><br>$A \cup B = \{x \in \Omega : x \in A \text{ or } x \in B\}$ |
| Figure 2.4 | |

**Intersection**: $A \cap B$

| | |
|---|---|
| | Let $A, B \subset \Omega$ . The intersection of $A$ and $B$ is a set of elements in $\Omega$ such that the elements are in both $A$ and $B$. Mathematically, $A \cap B = \{x \in \Omega : x \in A \text{ and } x \in B\}$ . |
| Figure 2.5 | |

**Difference:** $A \setminus B$

| | |
|---|---|
| | Let $A, B \subset \Omega$ . The difference of $A$ and $B$ is a set of points which are in $A$ and not in $B$ and denoted by $A \setminus B$ . Mathematically, $A \setminus B = \{x \in \Omega : x \in A \text{ and } x \notin B\} = A \cap B^c$ . |
| Figure 2.6 | |

**Symmetric difference:** $A \Delta B$

| | |
|---|---|
| | Let $A, B \subset \Omega$ . The symmetric difference of $A$ and $B$ denoted by $A \nabla B$ where $A \Delta B = (A \setminus B) \cup (B \setminus A) = (A \cap B^c) \cup (B \cap A^c)$ |
| Figure 2.7 | |

Note that two sets are said to be disjoint if their intersection is empty. That is, if $A \cap B = \varnothing$ then the sets $A$ and $B$ are disjoint.

Secondly, let $A, B$ and $C$ be any three subsets of $\Omega$ . Then we have the following equalities:

- $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
- $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
- $(A \cup B)^c = A^c \cap B^c$ and $(A \cap B)^c = A^c \cup B^c$

**Example:** Let $\Omega = \{a, b, c, d, e, f\}$ and $A = \{a, b, c\}$, $B = \{b, d, e\}$ and $C = \{a, c\}$. Then find the following sets.

$A \cup B = \{a, b, c, d, e\}$, $A \cap B = \{b\}$,

Since, $C \subset A$ then $A \cap C = C = \{a, c\}$ and $A \cup C = A = \{a, b, c\}$

$B \cup C = \{a, b, c, d, e\}$ and $B \cap C = \varnothing$, that is the subsets $B$ and $C$ are disjoint.

$A \cup B \cup C = \{a, b, c, d, e\}$,

$B^c = \{a, c, e, f\}$ , $A^c = \{d, e, f\}$ $\Rightarrow$ $A \setminus B = A \cap B^c = \{a, c\}$ and $B \setminus A = B \cap A^c = \{d, e\}$

$C^c = \{b, d, e, f\}$ and therefore, $A \setminus C = A \cap C^c = \{b\}$ and $B \setminus C = B \cap C^c = \{b, d, e\}$

$A \cup (B \setminus C) = \{a, b, c, d, e\}$ and $A \cap (B \setminus C) = \{b\}$.

## Probability

Let $\Omega$ be the sample space which is a set of all possible outcomes of an experiment. Let $\mathcal{U}$ denote the class of all possible subsets of the sample space. As we mentioned before, any subset of the sample space is an event. Consider a set function $P$,

$$P : \mathcal{U} \to [0,1]$$
$$A \to P(A).$$

If this set function satisfies:

**a)** $P(\Omega) = 1$ **b)** $P(A) \geq 1$ for all $A \in \mathcal{U}$ **c)** $A_i \in \mathcal{U}$ and $A_i$'s are disjoint $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$

then the set function $P$ is called a probability measure and the number $P(A)$ is called the probability of the event $A$. If $P$ is a probability measure defined on $\mathcal{U}$, the triple $(\Omega, \mathcal{U}, P)$ is called a probability measure.

**Example**: Consider an experiment of tossing a coin twice. Than the sample space is

$$\Omega = \{HH, HT, TH, TT\} = \{a, b, c, d\}, say$$

The class of all possible subsets of $\Omega$ is

$\mathcal{U} = \{\varnothing, \Omega, \{a\}, \{b\}, \{c\}, \{d\}, \{a,b\}, \{a,c\}, \{a,d\}, \{b,c\}, \{b,d\}, \{c,d\}, \{a,b,c\}, \{a,b,d\}, \{a,c,d\}, \{b,c,d\}\}$.

Define a set function $P$ as $P(A) = n(A)/4$ for all $A \in \mathcal{U}$ where $n(A)$ is the number of elements in $A$ then $P$ is a probability measure. Now, we can calculate the following probabilities:

$A = \{no\, tail\} = \{HH\} \Rightarrow P(A) = 1/4$

$B = \{one\, tail\} = \{HT, TH\} \Rightarrow P(B) = 2/4$

$C = \{two\, tails\} = \{TT\} \Rightarrow P(C) = 1/4$

$D = \{at\, least\, one\, tail\} = \{HT, TH, TT\} \Rightarrow P(D) = 3/4$.

## ***Some Rules of a Probability Measure***

In the following, we are going to list a few properties of a probability measure without proofs.

- $P(A^c) = 1 - P(A)$
- $P(\varnothing) = 0$
- If $A \subset B$ then $P(A) \le P(A)$ and $BP(B \setminus A) = P(B) - P(A)$

  this implies that since $\varnothing \subset A \subset \Omega$ we have $P(\varnothing) \le P(A) \le P(\Omega)$ that is , $0 \le P(A) \le 1$ and therefore, probability of an event is always between $0$ and $1$
- If $A, B \in \mathcal{U}$ then $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ . If $A$ and $B$ are disjoint events then $P(A \cap B) = P(\varnothing) = 0$ and therefore $P(A \cup B) = P(A) + P(B)$
- If $A_n, n \in \mathbb{N}$ be any sequence of events then $P\left(\bigcup_{n=1}^{\infty} A_n\right) \le \sum_{n=1}^{\infty} P(A_n)$

***Conditional probability***:

Let $\Omega$ be the sample space which is a set of all possible outcomes of an experiment and $\mathcal{U}$ be the class of all possible subsets of $\Omega$. And, let $P$ be a probability measure defined on $\mathcal{U}$. Simply, let $(\Omega, \mathcal{U}, P)$ be a probability space. Take an event $B$ in $\mathcal{U}$ such that $P(B) > 0$ and define a set function $P_B$ as in terms of the probability measure $P$ as follows:

$$P_B : \mathcal{U} \to [0,1]$$
$$A \to P_B(A) = \frac{P(A \cap B)}{P(B)}.$$

Then the set function $P_B$ is a probability measure. In fact,

- $P_B(\Omega) = \dfrac{P(\Omega \cap B)}{P(B)} = \dfrac{P(B)}{P(B)} = 1$

- Since $P$ is a probability measure we have $P(A \cap B)$ and therefore,

$$P_B(A) = \frac{P(A \cap B)}{P(B)} \geq 0 \text{ for any } A \in \mathcal{U}$$

- Let $A_n$ be any disjoint events in $\mathcal{U}$ ( $A_n \in \mathcal{U}$ and $A_i \cap A_j = \varnothing$ for $i \neq j$ and since $A_n$ 's are disjoint then the new sequence of events $C_n = A_n \cap B$ is also disjoint) and therefore,

$$P_B\left(\bigcup_{i=1}^{\infty} A_i\right) = \frac{P\left(B \cap \left(\bigcup_{i=1}^{\infty} A_i\right)\right)}{P(B)} = \frac{P\left(\bigcup_{i=1}^{\infty} (A_i \cap B)\right)}{P(B)} = \frac{\sum_{i=1}^{\infty} P(A_i \cap B)}{P(B)} = \sum_{i=1}^{\infty} \frac{P(A_i \cap B)}{P(B)} = \sum_{i=1}^{\infty} P_B(A_i).$$

Since, the new set function $P_B$ satisfies three conditions, it is a probability measure. This probability measure is called the conditional probability measure with respect to $B$ and the number $P_B(A)$ is called the conditional probability of $A$ given $B$ and usually denoted by $P(A|B)$. Notice that,

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

which implies that $P(A \cap B) = P(B)P(A|B)$.

**<u>Example</u>**: Consider two boxes, one contains 5 red and 5 blue balls and other contains 10 red and 10 blue balls.

| | |
|---|---|
| | We randomly pick a ball from Box I and put into Box II. Next, we pick a ball randomly from Box II. We observe that the second ball is RED. What is the probability that the first ball is RED? |
| Figure 2.8 | |

**<u>Solution</u>**: Define the following events:

$A_1 = \{ \text{first ball is RED} \}$ ,

$A_2 = \{ \text{first ball is BLUE} \}$ ,

$B = \{ \text{second ball is RED} \}$ .

We want to calculate $P(A_1 | B)$.

Note that $A_1 \cap A_2 = \varnothing$ and $A_1 \cup A_2 = \Omega$ and therefore for any subset $B$ of $\Omega$ can be written as

$$B = B \cap \Omega = B \cap (A_1 \cup A_2) = (B \cap A_1) \cup (B \cap A_2)$$

here the events (subsets ) $(B \cap A_1)$ and $(B \cap A_2)$ are disjoint. And therefore,

$$P(B) = P(B \cap A_1) + P(B \cap A_2) = P(A_1)P(B | A_1) + P(A_2)P(B | A_2) = \left(\frac{1}{2}\right)\left(\frac{11}{21}\right) + \left(\frac{1}{2}\right)\left(\frac{10}{21}\right) = \frac{1}{2}$$

using this probability we can calculate the conditional probability as we want:

$$P(A_1 | B) = \frac{P(A_1 \cap B)}{P(B)} = \frac{P(A_1)P(B | A_1)}{P(B)} = \frac{(1/2)(11/21)}{(1/2)} = \frac{11}{21}.$$

**Bayes Formula:**

Let $A_1, A_2, ..., A_k$ be any subsets of the sample space $\Omega$ such that $A_i \cap A_j = \varnothing$, for $i \neq j$ and $\Omega = (A_1 \cup A_2 \cup ... \cup A_k)$. A set of subsets satisfies the above conditions is called a partitioned of the sample space. Then we can calculate conditional probabilities in general as

$$P(A_s | B) = \frac{P(A_s \cap B)}{P(B)} = \frac{P(A_s)P(B | A_s)}{\sum_{i=1}^{k} P(A_i)P(B | A_i)}, i = 1, 2, ..., k \quad .$$

The formula

$$\boxed{P(A_s | B) = \frac{P(A_s)P(B | A_s)}{\sum_{i=1}^{k} P(A_i)P(B | A_i)}, i = 1, 2, ..., k}$$

is known to be the BAYES formula.

**Independent Events:**

Let $A$ and $B$ be two events. If $P(A \cap B) = P(A)P(B)$ then the events $A$ and $B$ are said to be independent.

The events $A_1$ and $A_2$ are independent if $P(A_1 \cap A_2) = P(A_1)P(A_2)$

The events $A_1$, $A_2$ and $A_3$ are independent if $P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3)$

The events $A_1$, $A_2, \ldots, A_k$ are independent if $P(A_1 \cap A_2 \cap ... \cap A_k) = P(A_1)P(A_2)...P(A_k)$.

Let $A$ and $B$ be two independent events. Then

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

**Example:** If $A$ and $B$ are two independent events, then

a) $A^c$ and $B$ b) $B^c$ and $A$ c) $A^c$ and $B^c$

are all independent.

**Solution**: a) If the events $A$ and $B$ are independent then $P(A\cap B)=P(A)P(B)$. In order to show that $A^c$ and $B$ are independent, we need to show that $P(A^c\cap B)=P(A^c)P(B)$. Notice that the events $A^c\cap B$ and $A\cap B$ are disjoint and

$$B=B\cap\Omega=B\cap(A\cup A^c)=(A\cap B)\cup(A^c\cap B)\,.$$

Therefore,

$$\begin{aligned} P(B)=P(A\cap B)+P(A^c\cap B) &\Rightarrow P(A^c\cap B)=P(B)-P(A\cap B)\\ &\Rightarrow P(A^c\cap B)=P(B)-P(A)P(B)=P(B)[1-P(A)]=P(B)P(A^c)\\ &\Rightarrow P(A^c\cap B)=P(B)P(A^c)\end{aligned}$$

which implies that the events $A^c$ and $B$ are independent.

b) Similarly,

$$\begin{aligned} P(A)=P(A\cap B)+P(B^c\cap A) &\Rightarrow P(B^c\cap A)=P(A)-P(A\cap B)\\ &\Rightarrow P(B^c\cap A)=P(B)-P(A)P(B)=P(A)[1-P(B)]=P(A)P(B^c)\\ &\Rightarrow P(B^c\cap A)=P(A)P(B^c)\end{aligned}$$

which implies that the events $A$ and $B^c$ are independent.

c) Note that

$$\begin{aligned} P(A^c\cap B^c)=P((A\cup B)^c)&=1-P(A\cup B)=1-[P(A)+P(B)-P(A\cap B)]\\ &=1-[P(A)+P(B)-P(A)P(B)]=1-P(A)-P(B)+P(A)P(B)\\ &=[1-P(A)][1-P(B)]=P(A^c)P(B^c)\end{aligned}$$

which implies that the events $A^c$ and $B^c$ are independent.

**Example**: Let $\Omega=\{a,b,c,d\}$ and $\mathcal{U}$ be the class of all subsets of $\Omega$ . Define a set function $P$ as

$P(A)=n(A)/4$ for any subset of $\Omega$ (or for any $A\in\mathcal{U}$ ) .

a) Consider the events $A=\{a,b\}$ and $B=\{b,c\}$. Notice that $P(A)=P(B)=1/2$ and $A\cap B=\{b\}$ therefore $P(A\cap B)=1/4$. Notice that

$$P(A\cap B)=\frac{1}{4}=\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)=P(A)\,P(B)\,.$$

Thus, the events $A$ and $B$ are independent. As you see, the events are not disjoint ( $A \cap B = \{b\} \neq \varnothing$) but they are independent.

On the other hand, consider the event $C = \{c, d\}$ and notice that the events $A$ and $C$ are disjoint ( $A \cap C = \varnothing$ ). Note that

$$P(A \cap C) = P(\varnothing = 0 \neq \left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = P(A)P(B)$$

and therefore the events $A$ and $C$ are not independent.

**Result**: Let $A$ be any event (any subset of the sample space). Then

a) The events $A$ and $\Omega$ are independent

b) The events $A$ and $\varnothing$ are independent.

**Proof**:

a) Since $P$ is a probability measure we have $P(\Omega) = 1$ and $A \subset \Omega$ we also have $A \cap \Omega = A$. Therefore, we can write $P(A \cap \Omega) = P(A) = P(A)1 = P(A)P(\Omega)$ which implies that the events $A$ and $\Omega$ are independent.

b) Similarly, since $P$ is a probability measure, we have $P(\varnothing) = 0$ and $\varnothing \cap A = \varnothing$. Therefore,

$$P(A \cap \varnothing) = P(\varnothing) = 0 = P(A)\ 0 = P(A)P(\varnothing)$$

which implies that the events $A$ and $\varnothing$ are independent.