Basics of Probability

Very short introduction

These notes are a quick and dirty introduction to probability theory! *Tanvir!*

Contents

1 | Probability Space

The object of probability theory is to describe and investigate mathematical models of random phenomena, primarily from a theoretical point of view. Closely related to probability theory is statistics, which is concerned with creating principles, methods, and criteria in order to treat data pertaining to such (random) phenomena or data from experiments and other observations of the real world, by using, for example, the theories and knowledge available from the theory of probability.

Probability theory thus aim at describing *random experiments*. We will shortly explain what do we mean by random experiments. But before that you should know the basis of probability theory is the *probability space* which is written $(\Omega, \mathcal{F}, P)$. You really need to understand each of the three things that we have written here. First let us give the names of these three elements. Ω is what we call *sample space*, $\mathcal{F}$ is what we call *event space* (another imprtant name σ-algebra) and P is our dearest *probability function* (another important name probability measure). Each of the three plays significant role in understanding probability. You might be wondering, why this complication? why not simply calculate probabilities from counting? Well welcome to mathematics, because in mathematics we hate inconsistencies. So we will *define* probability in a way so that we get meaningful probabilities. OK Lets start. Since everything starts from experiment. First we explain what is an experiment.

❖ **Definition 1.1** (Experiment, Random Experiment and Events)**.** An *experiment* is any process, real or hypothetical, in which *all possible outcomes* can be identified *ahead of time*. We call an experiment *random experiment* when we do not know which outcomes will appear. An *event* is a well-defined *set* of possible outcomes of the experiment.

<div align="right">—––❖</div>

♣ **Example 1.2** (Toissing a coin)**.** If we toss a coin we know the outcomes can be head or tail. So we know all possible outcomes, but we do not beforehand which will come when we toss a particular coin. So tossing a coin is the random experiment.

<div align="right">*(...end of example!)*</div>

♣ **Example 1.3** (Throwing a dice)**.** Similarly when we trow a dice we know that any number between 1 to 6 might appear, but before throwing it we do not know which one will exactly come. So throwing a dice is the random experiment.

<div align="right">*(...end of example!)*</div>

♣ **Example 1.4** (Tossing multiple coins)**.** A slightly complex example is when we trow toss more than one coins together. The experiment in this case is tossing more than one coins. For example, think about tossing three coins, we know all the outcomes, that is we know that our *realization* could be

$$(H, H, H) \text{ or } (H, H, T), \text{ or } (H, T, H), \text{ or }$$
$$(H, T, T), \text{ or } (T, H, H), \text{ or } (T, H, T), \text{ or }$$
$$(T, T, H), \text{ or } (T, T, T), \text{ or } (H, H, H)$$

<div align="right">*(...end of example!)*</div>

1.1 Sample space

Now that we have defined random experiment (again random means, beforehand we do not know which outocme will appear), we can define *Sample Space*. So sample space is a *set* (you should know some set theory from the basic math course!). It is the set of all possible outcomes. We will denote the sample space with Ω.

♣ **Example 1.5** (Continuation of Example 1.2)**.** When we toss a coin, the sample is the set $\Omega = \{H, T\}$

(...end of example!)

♣ **Example 1.6** (Continuation of Example 1.3)**.** When we throw a dice the sample space is $\Omega = \{1, 2, 3, 4, 5, 6\}$

(...end of example!)

ex:
tossmulticoin2 ♣ **Example 1.7** (Continuation of)**.** When we throw three coins together then we have a big sample space of 8 possible outcomes,

$$\Omega = \{(HHH), (THH)(HTH)(HHT) \\ (HTT), (THT), (TTH), (TTT)\}$$

We will use these example a lot, so let us write this example with easy notation, let us denote

$$\omega_1 : \text{HHH}, \quad \omega_2 : \text{THH}, \quad \omega_3 : \text{HTH} \quad \omega_4 : \text{HHT}$$
$$\omega_5 : \text{HTT}, \quad \omega_6 : \text{THT} \quad \omega_7 : \text{TTH}, \quad \omega_8 : \text{TTT}$$

Now we can write the same sample space as

$$\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6, \omega_7, \omega_8\},$$

which looks much better.

(...end of example!)

1.2 Event space

Now let us talk about event space, the idea about event space is this is where we would like to assign probabilities. This will become clear soon, but let us define first what is an event.

❖ *Definition* **1.8** (Event)**.** An event is a subset of the sample space. And often we denote this with the letters A. B, C, etc.

–––❖

So if the sample space is the big set Ω, an event is just a subset. Note it can be any subset as long as it is a subset of Ω.

♣ **Example 1.9.** If we want to define an event in Example 1.7, we can simply take any subset $A \subset \Omega$, where $A = \{\omega_1, \omega_8\}$

(...end of example!)

The last example just took a subset which may not habe any meaning. This is a valid event. But in probability theory we often define event to be a meaningful set. So let us define some events.

♣ **Example 1.10** (Continuation of Example 1.7)**.** If A is an event that *at least* one head is obtained in the 3 tosses; then

$$A = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6, \omega_7\}$$

If B is an event that *one head is obtained on the second toss*;

$$B = \{\omega_1, \omega_2, \omega_4, \omega_6\}$$

let C be the event that *one tail is obtained on the third toss*, so

$$C = \{\omega_4, \omega_5, \omega_6, \omega_8\}$$

finally let D be the event that *no heads are obtained*. Accordingly,

$$D = \{\omega_8\}$$

(...end of example!)

So event is simply a subset. What about *event space*. Event space is the set of *all (well defined) events*. So simply we should take all "nice" events and put them in a set, so that we get a *set of events* and that set is what we call the event space. Note, the events itself are sets, so when we again make another set with events we have a *set of sets*. For example, power set is the set of all possible subsets. The problem is how do we define "nice" or "well behaved" that we have written in the definition. We will not into the details of this concept. You will learn it in detail in the Statistical Theory course. But we get the nice set of subsets. When we have a finite sample space, for all of our example, 1.2 to 1.4 we have finite Ω, the nice sets or the event space is just simply the power set. Let us look at some example

♣ **Example 1.11** (Power set as Event space)**.** For example we can take $\Omega = \{\omega_1, \omega_2, \omega_3\}$. So this is our sample. Now for this set Ω, what is the power set of this set? It is the set all possible outcomes

$$
\begin{aligned}
\mathscr{P}(\Omega) = \{ &\varnothing, \\
&\{\omega_1\}, \{\omega_2\}, \{\omega_3\}, \\
&\{\omega_1, \omega_2\}, \{\omega_2, \omega_3\}, \{\omega_1, \omega_3\} \\
&\{\omega_1.\omega_2, \omega_3\}\}
\end{aligned}
$$

In this case the event space which is the *set of nice subsets* of Ω is the power set, which is the set of all outcomes, so the event space is same

$$
\begin{aligned}
\mathcal{F} = \{ &\varnothing, \\
&\{\omega_1\}, \{\omega_2\}, \{\omega_3\}, \\
&\{\omega_1, \omega_2\}, \{\omega_2, \omega_3\}, \{\omega_1, \omega_3\} \\
&\{\omega_1.\omega_2, \omega_3\}\}
\end{aligned}
$$

(...end of example!)

Again we emphasize it, event space is a particular set of subsets of Ω. We have not given you the definition of event space, you do not need it for now. Just remember its a special set of subsets. It may or may note be the power set. In the case of finite sample space, when Ω is finite, power set satisfies the definition of the event space, so that is why power set is the event space. For infinite cases, meaning Ω is infinite. for example, when $\Omega = \mathbb{R}$, power set IS NOT an event space, we need a special set of sets then.

You can ask why do we need to define event space, the short answer is we would like to calculate probabilities on events of the sample space. For example if $A = \{\omega_1, \omega_2, \omega_3\}$ is an

event of Ω, we would like to calculate the probability of happening Ω, or in other words $P(A)$. But what is probability P? This is what we will define in the next section.

1.3 Probability

The mathematical definition of probability will satisfy three *axioms*. You can think about axioms as some kind of grand rules, that we just take it as some rules that we will assume will always hold. Its actually very much like definition. In probability theory mathematicians came up with three axioms, or rules of probability and then they said for P to be a probability on any set A, following three rules must hold

❖*Axioms* **1.12** (Axioms of Probability)**.**

a1. For every event $A \subset \Omega$ $P(A) \geq 0$.

a2. $P(\Omega) = 1$

a3. For every infinite sequence of disjoint events $A_1, A_2, \ldots,$

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

Now that we have the axioms, we can define *probability*,

❖ *Definition* **1.13** (Probability)**.** *A probability measure*, or simply a *probability*, on a sample space Ω is a set function, such that $P(A)$ for all events A will satisfy Axioms 1.12.

$$---❖$$

So any set function as long as it satisfies the axioms will be a probability function. Let us quickly mention some of the rules of probability. You should try to prove this, if not now at least remember this

❖ *Theorem* **1.14.** For any set $A \subset \Omega$, and $B \subset \Omega$
- $P(\varnothing) = 0$.
- For every finite sequence of n disjoint events $A_1 \ldots \ldots, A_n$
$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$$

- $P(A^c) = 1 - P(A)$
- If $A \subset B$, then $P(A) \leq P(B)$
- $0 \leq P(A) \leq 1$
- For A and B
$$P(A \cap B^c) = P(A) - P(A \cap B)$$

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

$$---❖$$

So we have already explained the three elements of the probability space that $\Omega, \mathcal{F}, P$, Ω is the sample space which is the set of all possible outcomes, $\mathcal{F}$ is the *set of nice subsets of* Ω, when Ω is a finite set, then $\mathcal{F}$ is a power set, but if Ω is any infinite set. e.g., $\Omega = \mathbb{R}$. It tuns out we cannot take power set as a event space, because we cannot consistently assign probabilities. So we have to choose in general. You do not need more details on this now. Finally P is any *set* function that will satisfy certain axioms, in particular Axioms 1.12. Let

us see one complete example,

♣ **Example 1.15** (Throwing a dice example)**.** Recall the dice example, our sample space was

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

You can think about the event space, which is the power set in this case. Now we need to come up with an idea how we can assign probabilities for each subset A of $\Omega = \{1, 2, 3, 4, 5, 6\}$. Let us do the following, let $P(A)$ be the number of elements of A divided by 6. Then it is easy to see that this satisfies the first two axioms. There are only finitely many distinct collections of nonempty disjoint events. It is not difficult to see that the third axiom is also satisfied by this example.

There are also ways to assign probabilities. For example, if we believe that the die is loaded, we might believe that some sides have different probabilities of turning up. To be specific, suppose that we belicve that 6 is twice as likely to come up as each of the other five sides. We could set $p_j = 1/7$ for $i = 1, 2, 3, 4, 5$ and $p_6 = 2/7$. Then, for each event A, define $P(A)$ to be the sum of all p_i such that $i \in A$. For example, if $A = \{1, 3, 5\}$, then $P(A) = p_1 + p_3 + p_5 = 3/7$ It is not difficult to check that this also satisfies all three axioms.
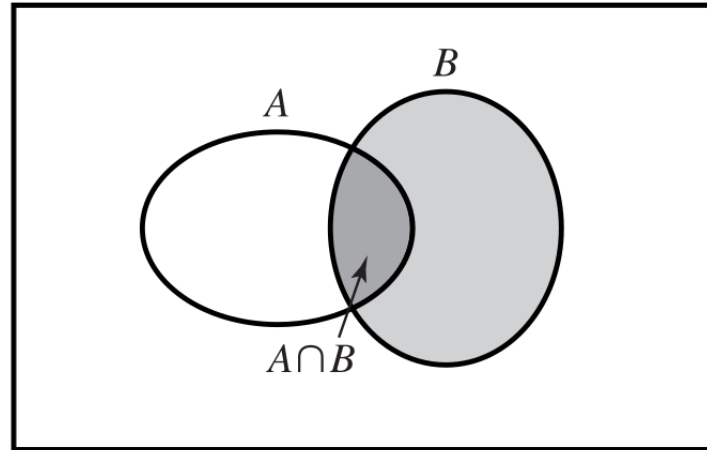
So the idea is, if I design/know how will I calculate the probabilities, then if you give me any subset A of Ω (that is part of the nice set of subsets), I can calculate the probabilities of that set.

(...end of example!)

2 | Conditional Probability

2.1 Conditional Probability: Definition and exmaples

❖ **Definition** 2.1 (Conditional Probability). Suppose that we learn that an event *B* has occurred and that we wish to compute the probability of another event *A* taking into account that we know that *B* has occurred. The new probability of *A* is called the conditional probability of the event A given that the event *B* has occurred and is denoted P(*A*|*B*) If P(*B*) > 0, we compute this probability as

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{2.1}$$

{eq: conp}

The conditional probability P(*A*|*B*) is not defined if P(*B*) = 0

–––❖

For convenience, the notation $P(..|..)$ in Eq. (2.1) is read simply as the conditional probability of *A* given *B*, indicates that P(*A*|*B*) is computed as the proportion of the total probability P(*B*) that is represented by P(*A* ∩ *B*), intuitively the proportion of *B* that is also part of *A*. Lets see an example

♣ **Example 2.2** (A Clinical Trial). It is very common for patients with episodes of depression to have a recurrence within two to three years. Prien et al. (1984) studied three treatments for depression: imipramine, lithium carbonate, and a combination. As is traditional in such studies (called clinical trials), there was also a group of patients who received a placebo. (A placebo is a treatment that is supposed to be neither helpful nor harmful. Some patients are given a placebo so that they will not know that they did not receive one of the other treatments. None of the other patients knew which treatment or placebo they received, either.)

In this example, we shall consider 150 patients who entered the study after an episode of depression that was classified as "unipolar" (meaning that there was no manic disorder). They were divided into the *four* groups (three treatments plus placebo) and followed to see how many had recurrences of depression. Following table summarizes the results.

If a patient were selected at random from this study and it were found that the patient received the placebo treatment, what is the conditional probability that the patient had a relapse? SO we want to calculate the conditional probability of relapse, conditional on

placebo.

	Treatment group				
Response	Imipramine	Lithium	Combination	Placebo	Total
Relapse	18	13	22	24	77
No relapse	22	25	16	10	73
Total	40	38	38	34	150

Let B be the event that the patient received the placebo, and let A be the event that the patient had a relapse. So we want to calculate $P(A|B)$.

We can calculate $P(B) = 34/150$ and $P(A \cap B) = 24/150$ directly from the table. Then $P(A|B) = 24/34 = 0.706$ Note there is a rescaling. If you would calculate the probability of relapse out of whole sample then you would calculate 24/150. But now since you know placebo happened (we are conditioning on a placebo set), we need to adjust the probability of relapse, based on this new information.

(...end of example!)

❖ **Theorem** 2.3 (Theorem Multiplication Rule for Conditional Probabilities). Let A and B be events. If $P(B) > 0$ Eq. (2.1) then
$$P(A \cap B) = P(B) \, P(A|B)$$
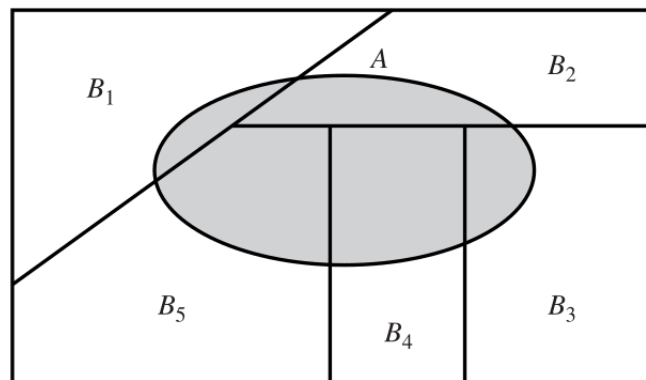
Also if $P(A) > 0$ then,
$$P(A \cap B) = P(A) \, P(B|A)$$

– – –❖

The baove theoem is a direct consequence of the definition,

2.2   Conditional Probability and Partitions

Think about a situation sample space Ω can be partitioned into other sets B_1, B_2, B_3, B_4 and B_5. So here is the picture of the partition.

It turns out that in this case we can write the probability of A with the conditional probability of $P(A|B_j)$ and $P(B_j)$. Following theorem shows this

❖ **Theorem** 2.4 (Law of total probability). Suppose that the events $B_1, \ldots, B_k$ form a partition of the sample space Ω and $P(B_j) > 0$ for $j = 1, \ldots, k$. Then, for every event A in Ω
$$P(A) = \sum_{j=1}^{k} P(B_j) \, P(A|B_j)$$

2.3 Independent Events

The conditional probability of the event A given that the event B has occurred is the revised probability of A after we learn that B has occurred. It might be the case, however, that no revision is necessary to the probability of A even after we learn that B occurs. In this case, we say that A and B are independent events.

♣ **Example 2.5** (Toss a coin and roll a dice togther)**.** If we toss a coin and then roll a die, we could let A be the event that the die shows 3 and let B be the event that the coin lands with heads up. If the tossing of the coin is done in isolation of the rolling of the die, we might be quite comfortable assigning P($A|B$) = P(A) = 1/6. In this case, we say that A and B are independent events.

(...end of example!)

In general, if $\Pr(B) > 0$, the equation P($A|B$) = P(A) can be rewritten as P($A \cap B$)/P(B) = P(A). If we multiply both sides of this last equation by P(B), we obtain the equation P($A \cap B$) = P(A) P(B). In order to avoid the condition P(B) > 0, the mathematical definition of the independence of two events is stated as follows:

❖ *Definition* **2.6** (Independent Events)**.** Two events A and B are independent if
$$\Pr(A \cap B) = \Pr(A) \Pr(B)$$

Lets look at an example,

♣ **Example 2.7.** Suppose that a fair coin is tossed twice. The experiment has four outcomes, HH, HT, TH, and TT, that tell us how the coin landed on each of the two tosses. We can assume that this sample space is simple so that each outcome has probability 1/4. Suppose that we are interested in the second toss. In particular, we want to calculate the probability of the event $A = \{H$ on second toss $\}$. We see that $A = \{HH, TH\}$, so that P(A) = 2/4 = 1/2. If we learn that the first coin landed T, we might wish to compute the conditional probability P($A|B$) where $B = \{$ T on first toss $\}$ Using the definition of conditional probability, we easily compute

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1/4}{1/2} = \frac{1}{2}$$

because $A \cap B = \{TH\}$ has probability 1/4. We see that P($A|B$) = P(A); hence, we don't change the probability of A even after we learn that B has occurred.

(...end of example!)

2.4 Indepdendence of several events

The definition of independent events can be extended to any number of events, $A_1 \ldots \ldots, A_k$. Intuitively, if learning that some of these events do or do not occur does not change our probabilitics for any events that depend only on the remaining events, we would say that all k events are independent.

A | Appendix: Set Theory

Sets are constantly encountered in mathematics. One speaks of sets of points, collections of real numbers, and families of functions. A set is conceived simply as a collection of definable and unique objects. The words set, collection, and family are all synonymous. The notation $x \in A$ means that x is an element of the set A; the notation $x \notin A$ means that x is not an element of the set A. A set can be described by listing its elements, usually within braces For example,

$$A = \{-1, 2, 5, 4\}$$

describes the set consisting of the numbers $-1, 2, 4$, and 5. More generally, a set A may be defined as the collection of all elements x in some larger collection satisfying a given property. Thus the notation

$$A = \{x : P(x)\}$$

defines A to be the set of all objects x having the property $P(x)$. This is usually read as "A equals the set of all elements x such that $P(x)$." For example, if x ranges over all real numbers, the set A defined by

$$A = \{x : 1 < x < 5\}$$

is the set of all real numbers that lie between 1 and 5. For this example, $3.75 \in A$ whereas $5 \notin A$. We will also use the notation $A = \{x \in X : P(x)\}$ to indicate that only those x that are elements of X are being considered. Some basic sets that we will encounter throughout the text are the following:

$\mathbb{N}$ = the set of natural numbers or positive integers = $\{1, 2, 3, \ldots, \}$
$\mathbb{Z}$ = the set of all integers = $\{\ldots, -2, -1, 0, 1, 2, \ldots, \}$
Q = the set of rational numbers = $\{p/q : p, q \in \mathbb{Z}, q \neq 0\}$,
and $\mathbb{R}$ = the set of real numbers.

§. Using Venn diagrams for sets

We will often use Venn diagrams to represent sets. Normally in a Venn diagram we draw sets using some objects like circles or squares and then show different relations between sets.

❖ *Definition* **A.1** (Containment). It is said that a set A is contained in another set B if every element of the set A also belongs to the set B. This relation between two set is expressed symbolically by the expression $A \subset B$, which is the set-theoretic expression for saying that A is a subset of B. Equivalently, if $A \subset B$, we may say that B contains A and may write $B \supset A$

–––❖

For containment, if $A \subset B$, we can draw this in a Venn diagram,
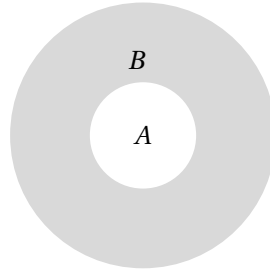
Now we can also probe can prove following theorem (try this!)

❖ **Theorem A.2.** Let A, B, and C be sets. If $A \subset B$ and $B \subset A$, then $A = B$. If $A \subset B$ and $B \subset C$, then $A \subset C$

––––❖

❖ **Definition A.3.** The set contains no elements is called the empry set, or null set, and it is denoted by the symbol $\varnothing$.

––––❖

Some sets contain only finitely many elements, while others have infinitely many elements. There are two sizes of infinite sets that we need to distinguish.

❖ **Definition A.4** (Countable/Uncountable.). An infinite set A is countable if there is a one-to-one corrspondence between the elements of A and the set of natural numbers $(1, 2, 3, \ldots)$. A set is uncountable if it is neither finite nor countable. If we say that a set has at most countably many elements, we mean that the set is either finite or countable.

––––❖

❖ **Definition A.5** (Empty set). A set which has no elements is called an empty set and we denote this with $\varnothing$

––––❖

Note if you take any set A, then $\varnothing \subset A$. This might look strange. But this is true because empty set has nothing if include all the elements of the empty set we include nothing. Also another important point any set $A \subset A$.

❖ **Definition A.6** (Power set). If A is a set, the set of all subsets of A is called the power set of A and we denote this with $\mathscr{P}(A)$.

––––❖

♣ **Example A.7** (example of a power set). If $A = \{1, 2\}$, then

$$\mathscr{P}(A) = \{\phi, \{1\}, \{2\}, \{1, 2\}\}$$

(...end of example!)

In the last example, the set A has 2 elements and $\mathscr{P}(A)$ has 4 or 2^2 elements, the elements in this instance being the subsets of A. If we take a set with 3 elements, then by listing the subsets of A it is easily seen that there are exactly 2^3 subsets of A (Exercise 8). On the basis of these two examples we are inclined to conjecture that if A contains n elements, then $\mathscr{P}(A)$ contains 2^n elements. Note in the example we also included $\varnothing$ and the whole set A as well.

A.1 Operations on sets

Now we come to certain operations on sets. Here we will need a special concept, that we call universal set. A universal set is just like a set. But the idea is whenever we have an universal set. All the other sets are subsets of the universal set. Often its denoted with U

❖ *Definition* **A.8** (Complement). The complement of a set A with respected to the universal set A is defined to be the set that contains all elements of U that do not belong to A. The notation for the complement of A is A^c.

<div align="right">———❖</div>

Note the concept complement needs a universal set, otherwise it does not make it sense. Here is the Venn diagram of A^c.
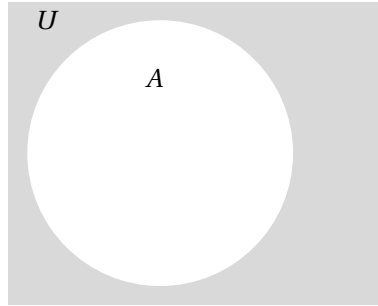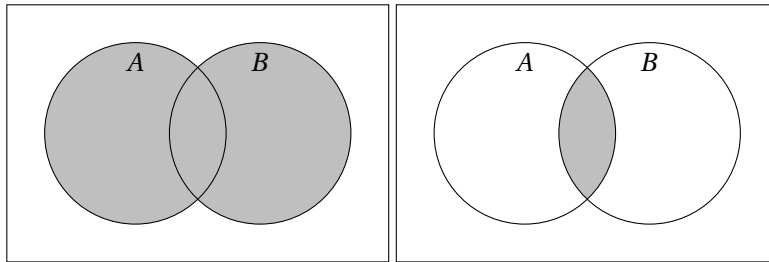
following are the events,

Figure A.1.: Left: $A \cup B$ and right $A \cap B$