MATH 230 - Probability and Statistics for Engineers

Introduction

Life runs on randomness and uncertainty - the realm of luck, risk, doubt, coincidence, fortune, and chance. We use probability and statistics to understand and measure uncertainty in a principled way.

2. PROBABILITY

Food for thought: What does Einstein mean when he says "God does not play dice with the universe." ?

2.1. Sample spaces and events

(*) Experiment is any action or process with an uncertain outcome.

Definition. The sample space of an experiment, often denoted by Ω (or S), is the set of all possible outcomes of that experiment.

Example. Consider the experiment of tossing a coin once. Here, the possible outcomes of the sample space, Ω, are either Heads (H) or Tails (T). That is, $\Omega = \{H, T\}$.

Example. If the experiment consists of examining a single fuse to determine whether it is defective, then the possible outcomes can be either Not defective (N) or Defective (D). Thus, the sample space is $\Omega = \{D, N\}$.

Example. If the experiment consists of examining three fuses in sequence to determine whether each is defective, then the possible outcomes can be either Not defective (N) and Defective (D). Then, $\Omega = \{NNN, DNN, NDN, NND, NDD, DND, DDN, DDD\}$.

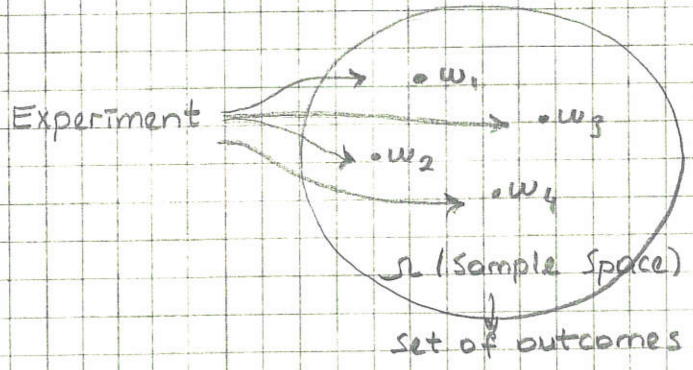
<u>Remark</u>. Note that when we say <u>in sequence</u> in the previous example, we mean "a sequence of N's and D's of length 3", where the order matters. Namely, we first check Fuse 1, then Fuse 2, and then Fuse 3.

<u>Example</u>. If the experiment consists of measuring (in hours) the lifetime of a transistor, then the sample space consists of all nonnegative real numbers; that is

$$\Omega = \{ x : 0 \leq x < \infty \}.$$

<u>Definition</u>. The outcomes of an experiment are sometimes called a <u>point</u>, <u>sample point</u>, or <u>elementary outcome</u> of the sample space, and are often denoted by ω.

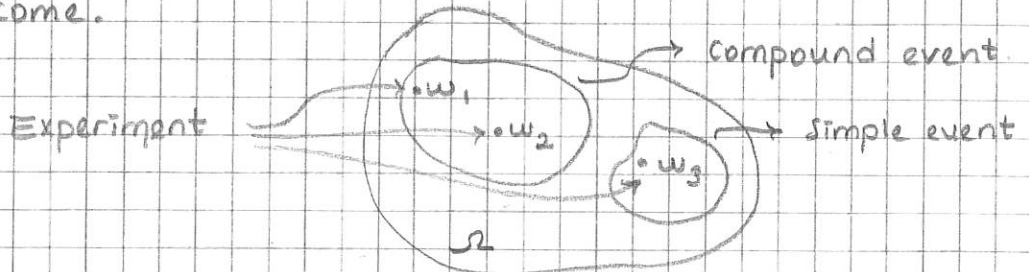
<u>Remark</u>. The sample space can be finite, e.g., $\Omega = \{\omega_1, \omega_2, \dots \omega_n\}$; countable, e.g., $\Omega = \mathbb{N}$; or uncountable, e.g., $\Omega = \mathbb{R}$ or $\Omega = \{0,1\}^{\infty}$.



<u>Example</u>. If the experiment consists of an infinite number of consecutive rolls of an ordinary die, the sample space $\Omega = \{1, 2, \dots, 6\}^{\infty}$. The element $\omega = (3, 1, 1, 5, \dots)$, which is an infinite sequence, is an example of a possible outcome.

In probability, we are interested not only in the individual outcomes but in any collection of outcomes from Ω.

<u>Definition.</u> An <u>event</u> is any collection (subset) of outcomes contained in the sample space Ω. An event is said to be <u>simple</u> if it consists of exactly one outcome and <u>compound</u> if it consists of more than one outcome.



→ When an experiment is performed, a particular event A is said to <u>occur</u> if the resulting experimental outcome is contained in A. Namely,

$$A \text{ occurs if } \omega \in A \subseteq \Omega \quad ; \ \omega: \text{outcome}, \ \Omega: \text{sample space}.$$

<u>Example.</u> Consider an experiment in which each of three vehicles taking a particular freeway exit turns left (L) or right (R) at the end of the exit ramp. The sample space is

$$\Omega = \{LLL, RLL, LRL, LLR, LRR, RLR, RRL, RRR\}.$$

Here, there are eight simple events

$$A_1 = \{LLL\}, A_2 = \{RLL\}, \dots, A_8 = \{RRR\},$$

and some compounds events are

$$B_1 = \{RLL, LRL, LLR\},$$

$$B_2 = \{LLL, RRR\},$$

$$B_3 = \{LLL, RLL, LRL, LLR\}.$$

☞ <u>NOTE.</u> Data scientist rarely work with individual outcomes. Instead, they usually consider collection of outcomes (events).

Subject: Date :..../..../....

Example ⊛ A type-D battery is a failure (F) if its voltage is outside the prescribed limits, and a success (S) if it is within limits. Consider the experiment of testing each batteries one by one until the first success occurs.

a) How many simple events does this experiment have?

b) Find the event that at most three batteries are examined.

c) Find the event that an even number of batteries are examined.

Solution. a) Since the problem does not specify a fixed number of trials, we consider n trials for all possible values of $n = 1, 2, 3, \ldots$. Thus,

$$\Omega = \{ S, FS, FFS, FFFS, \ldots \},$$

which contains an infinite number of possible outcomes. Hence, there are infinitely many simple events.

b) $A = \{ S, FS, FFS \}$

one battery is examined — two batteries are examined — three batteries are examined.

c) $B = \{ FS, FFFS, FFFFFS, \ldots \}$

Some relations from set theory

Definition. For any events $A, B \in \Omega$,

i) The union of two events A and B, denoted by $A \cup B$ and read "A or B", is the event consisting of all outcomes that are either in A or in B or in both events.

ii) The intersection of two events A and B, denoted by either $A \cap B$ or AB and read "A and B", is the event consisting of all outcomes that belong to both A and B.

iii) The complement of an event A, denoted by $A^c = \Omega / A$, or A', or $\bar{A}$ is the set of all outcomes in Ω that are not contained in A.

Example. In the battery experiment (Example ⊛), define A, B, and C by

$$A = \{S, FS, FFS\},$$

$$B = \{S, FFS, FFFFS\},$$

$$C = \{FS, FFFS, FFFFFS, \dots\}.$$

Then,

$$A \cup B = \{S, FS, FFS, FFFFS\}$$

$$A \cap B = \{S, FFS\}$$

$$A^c = \Omega / A = \{FFFS, FFFFS, \dots\}$$

$$C^c = \Omega / C = \{S, FFS, FFFFS, \dots\} = \{\text{an odd number of batteries are examined}\}$$

even number of batteries are examined

Definition. If the events A and B have no outcomes in common, we say that A and B are disjoint or mutually exclusive, and it is denoted by $A \cap B = \phi$, where ϕ denotes the event consisting of no outcomes whatsoever (the "null" or "empty" event). A set of events $\{A_1, A_2, \dots\}$ is called mutually exclusive if the joint occurrence of any two of them is impossible; that is, if $\forall i \neq j$, $A_i \cap A_j = \phi$. Thus, $\{A_1, A_2, \dots\}$ is mutually exclusive if and only if every pair of them is mutually exclusive.

Mutually exclusive: No two event can occur simultaneously.

Example. At a busy international airport, arriving planes land on a first-come, first-served bases. Let

A = there are at least five planes waiting to land,

B = there are at most three planes waiting to land,

C = there are exactly two planes waiting to land.

Then, find A^c, B^c, $A \cap B^c$, $B \cap C$, $A \cap B$, $A \cap C$, and $B \cap C^c$.

<u>Solution.</u> $A = \{\text{at least five planes waiting}\} = \{5, 6, 7, \ldots\}$

$B = \{\text{at most three planes waiting}\} = \{0, 1, 2, 3\}$

$C = \{\text{exactly two planes waiting}\} = \{2\}$

Then, we have

$A^c = \{\text{at most four planes are waiting to land}\} = \{0, 1, 2, 3, 4\}$

$B^c = \{\text{at least four planes are waiting to land}\} = \{4, 5, 6, 7, \ldots\}$

$A \cap B^c = A$ since $A \subseteq B^c$.

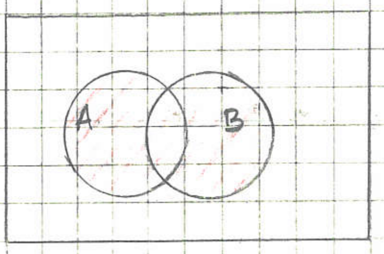
$B \cap C = C$ since $C \subseteq B$.

$A \cap B = \phi$ since A and B are mutually exclusive

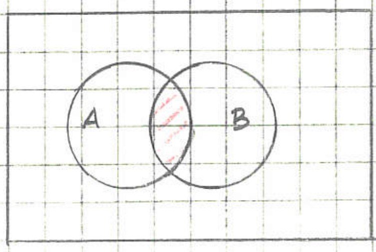
$A \cap C = \phi$ since A and C are mutually exclusive.

$B \cap C^c = \{0, 1, 3\}$ since $C^c = \{0, 1, 3, 4, \ldots\}$

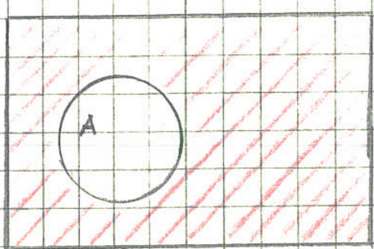
Venn diagrams are useful to visualize events and help improve understanding of their underlying logic.



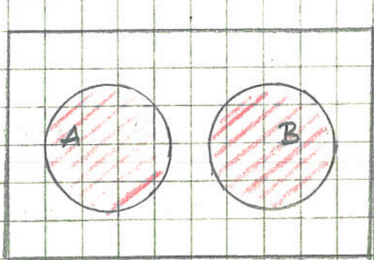
$A \cup B$



$A \cap B$



A^c



Mutually Exclusive

2.2. Axioms, Interpretations, and Properties of Probability

For an experiment with sample space Ω, the goal of probability is to assign each event A a number $P(A)$ that precisely measures the chance that A will occur.

<u>Definition.</u> Let Ω be the sample space of a random phenomenon. Suppose that to each event $A \subseteq \Omega$, a number denoted by $P(A)$ is associated with A. If the function P satisfies the following axioms, then it is called a <u>probability</u>, and the number $P(A)$ is said to be the <u>probability of A</u>.

1. <u>Non-negativity</u>: $P(A) \geq 0$, for every event A.
2. <u>Normalization</u>: $P(\Omega) = 1$.
3. <u>Countable additivity</u>: If $\{A_1, A_2, \dots\}$ is a sequence of mutually exclusive events, i.e.,

$$A_i \cap A_j = \emptyset, \text{ for all } i \neq j,$$

then

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

We add up all the individual chances to find the chance that one disjoint event happens.

That is,

$$P(A_1 \cup A_2 \cup \dots) = P(A_1) + P(A_2) + \cdots$$

Here,
- Axiom 1 states that the probability of an event, $P(A)$, is a non-negative number.
- Axiom 2 states that the probability of the outcome being in the sample space, Ω, is 1. Alternatively, it is also common to define $P(\Omega) = 100$ and interpret results as percentages.
- Axiom 3 states that the probability of one of the disjoint events occurring (left side), since they cannot occur at the same time, is given by the sum of their individual probabilities (right side).

Subject: Date : ..../...../.....

From Axiom 1 and Axiom 2, we can write that probability of an event A, $P(A)$, is a number between 0 and 1; that is, $0 \leq P(A) \leq 1$.

Now, since we have defined the sample space, events, and the axioms of probability, we are ready to describe a probabilistic model for the experiment.

There are many properties of a probability law that are not stated in the previous definition, but can be derived from these axioms.

Proposition. $P(\emptyset) = 0$, where $\emptyset$ is the null event.

Proof. Consider the infinite collection

$$A_1 = \emptyset,\ A_2 = \emptyset,\ A_3 = \emptyset, \dots \qquad \dots (1)$$

Since $A_i \cap A_j = \emptyset$, for $i \neq j$, the family (1) is pairwise disjoint. That is, $\bigcup_{i=1}^{\infty} A_i = \emptyset$. By Axiom 3, we write

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

$$\Rightarrow P(\emptyset) = \sum_{i=1}^{\infty} P(\emptyset).$$

Assume $1 \geq P(\emptyset) = p \geq 0$ by Axiom 1 and Axiom 2. Then, we would have $p = p + p + p + \dots$, which only happens when $p = 0$; that is $P(\emptyset) = 0$.

Proposition. Axiom 3 also holds in the finite case.

Proof. If we consider the pairwise disjoint events $A_1, A_2, \dots, A_k$, and infinite collection of empty events $A_{k+1} = \emptyset$, $A_{k+2} = \emptyset$, $A_{k+3} = \emptyset, \dots$, we obtain

$$P\left(\bigcup_{i=1}^{k} A_i\right) = P\left(\bigcup_{i=1}^{\infty} A_i\right)$$ → adding the infinite collection: $\emptyset$ doesn't change the union

$$= \sum_{i=1}^{\infty} P(A_i)$$ by Axiom 3

$$= \sum_{i=1}^{k} P(A_i)$$ by removing the infinite collection: $\emptyset$ doesn't change the sum.

Example. Consider tossing a thumbtack in the air. The sample space consists of two possible outcomes: up (U) or down (D) when it falls on the ground. That is, $\Omega = \{U, D\}$. From Axiom 2, we have $P(\Omega) = 1$, and from the previous proposition, we write

$$P(\Omega) = 1 = P(U) + P(D).$$

$$\Rightarrow P(U) = 1 - P(D)$$

For any number p between 0 and 1 representing $P(D)$, we have $P(U) = 1-p$. For instance, if $p = .7$, then $P(U) = 1-(.7) = .3$.

Example. In the battery experiment (Example ⊛), suppose the probability of any particular battery being satisfactory is .99. We know that

$$\Omega = \{S, FS, FFS, FFFS, \dots\}.$$

Here, the simple events are

$$A_1 = \{S\},\ A_2 = \{FS\},\ A_3 = \{FFS\},\ \dots$$

Then,

$$P(A_1) = .99,\ P(A_2) = (.01)(.99),\ P(A_3) = (.01)^2(.99), \dots .$$

These probabilities satisfy the axioms. Therefore,

$$\begin{aligned} P(\Omega) &= P(A_1) + P(A_2) + P(A_3) + \dots \\ &= .99 + (.01)(.99) + (.01)^2(.99) + \dots \\ &= .99\left(1 + .01 + (.01)^2 + (.01)^3 + \dots\right) \\ &= .99\left(\frac{1}{1-.01}\right) \\ &= .99\left(\frac{1}{.99}\right) \\ &= 1 \end{aligned}$$

by geometric series formula

$$\boxed{a + ar + ar^2 + \dots = \frac{a}{1-r}}$$

Interpreting probability

The axioms serve only to rule out assignments inconsistent with our intuitive notions of probability. For the appropriate or correct assignment of experiments, we need to introduce some additional notions:

- Suppose an experiment is conducted n times, and let A be an event associated with this experiment. Let $n(A)$ denote the number of replications in which A occurs. Then the ratio $\frac{n(A)}{n}$ is called the relative frequency of occurrence of the event A in the sequence of n trials.

- The relative frequency $\frac{n(A)}{n}$ tends to stabilize as n becomes large. More precisely, as $n \to \infty$, the relative frequency approaches a limiting value, called the limiting relative frequency of the event A.

▼ The objective interpretation of the limiting relative frequency of an event A is identified with $P(A)$. It is called objective because it does not depend on individual preferences and relies solely on the properties of an experiment.

▼ If the interpretations varies from individual to individual, then it is called subjective interpretation.

More probability properties

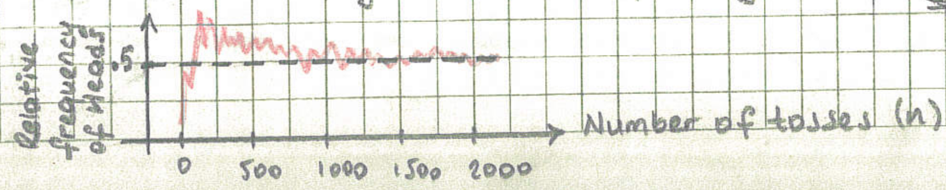
Proposition. For any event A, $P(A) = 1 - P(A^c)$.

Proof. We know that $A + A^c = \Omega$. Considering Axiom 2, we have

$$1 = P(\Omega) = P(A \cup A^c) = P(A) + P(A^c).$$

Thus, $P(A) = 1 - P(A^c)$.

Example for limiting relative frequency:



$$\Rightarrow P(H) = \lim_{n \to \infty} \frac{n(H)}{n} = .5$$

Example. Consider a system of five identical components connected in series. For each component, let S denote success and F denote failure. The system fails if at least one component fails and it works only if all five succeed.

a) How many outcomes correspond to the system working and to the system failing?

b) Each component works independently with $.9$. What is the probability that the system works?

c) What is the probability that the system fails?

Solution. System of five components in series: 

a) Each component can either work (S) or failure (F).

Then, there are $2^5 = 32$ possible outcomes.

There is one outcome, where the system works: $SSSSS$.

Thus, there are $32-1 = 31$ outcomes, where the system fails.

b) We are interested in the case of $A = \{SSSSS\}$. Then,

$P(A) = (.9)(.9)(.9)(.9)(.9) = (.9)^5 = .59$

c) Here, we wish to find $P(A^c)$, which can be found by using the previous proposition: $P(A) = 1 - P(A^c)$. Hence,

$P(A^c) = 1 - P(A) = 1 - (.59) = .41$.

Proposition. For any event A, $P(A) \leq 1$.

Proof. Considering the previous proposition, we write

$$\begin{aligned} 1 &= P(A) + P(A^c) \\ &\geq P(A), \end{aligned}$$

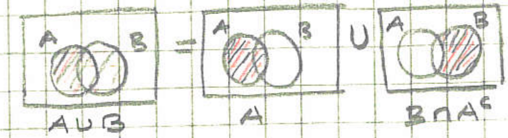
because $P(A^c) \geq 0$. ■

When A and B are disjoint, we have $P(A \cup B) = P(A) + P(B)$. Now, we will give this probability for disjoint events.

Subject:

Proposition. For any events A and B, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

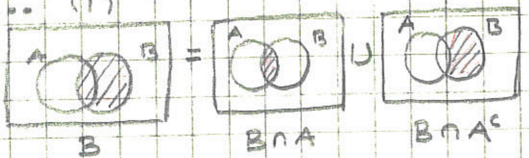
Proof. First, note that $A \cup B = A \cup (B \cap A^c)$.



Here, A and $(B \cap A^c)$ are two disjoint events. Then, we can write

$$P(A \cup B) = P(A) + P(B \cap A^c). \quad \cdots (1)$$

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



Similarly, since $(B \cap A)$ and $(B \cap A^c)$ are disjoint, we can write

$$P(B) = P(B \cap A) + P(B \cap A^c).$$

$$\Rightarrow P(B \cap A^c) = P(B) - P(B \cap A). \quad \cdots (2)$$

From (1) and (2), we obtain

$$\begin{aligned} P(A \cup B) &= P(A) + P(B \cap A^c) \quad \text{by (2)} \\ &= P(A) + \big(P(B) - P(B \cap A)\big) \\ &= P(A) + P(B) - P(B \cap A). \ \blacksquare \end{aligned}$$

Example. In a suburb,
- ⊛ 60% of households get internet from the cable company,
- ⊛ 80% get TV from the company
- ⊛ 50% get both internet and TV.

If one household is chosen at random,

a) What is the probability it gets at least one of the two services?

b) What is the probability it gets exactly one of the two services?

Solution. Let be $A = \{\text{gets internet from the cable company}\}$, and $B = \{\text{gets TV from the company}\}$.

We are given $P(A) = .6$, $P(B) = .8$, and $P(A \cap B) = .5$.

a) $P(A \cup B) = P(A) + P(B) - P(A \cap B) = (.6) + (.8) - (.5) = 1.4 - (.5) = .9$.

b) $P(B) = P(A \cap B) \cup P(A^c \cap B)$
$\quad = P(A \cap B) + P(A^c \cap B)$
$\Rightarrow .8 = .5 + P(A^c \cap B)$
$\Rightarrow P(A^c \cap B) = .3$

and

$P(A) = P(A \cap B) \cup P(B^c \cap A)$
$\quad = P(A \cap B) + P(B^c \cap A)$
$\Rightarrow .6 = .5 + P(B^c \cap A)$
$\Rightarrow P(B^c \cap A) = .1$

Then, $P(\text{exactly one}) = (.3) + (.1) = .4$ //

Remark. For three events A, B, and C, the probability of the union is

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

We remove the three since they were counted twice.

We add back this term because we previously removed it totally.

Determining probabilities systematically

A large number of simple events leads to many compound events. A simple way to determine probabilities for these events is to first determine probabilities $P(A_i)$ for all simple events. Here,

$$P(A_i) \geq 0 \quad \text{and} \quad \sum_{\text{all } i} P(A_i) = 1.$$

Then, the probability of any compound event C is computed by adding together the $P(A_i)$'s for all A_i's in C. Namely,

$$P(C) = \sum_{\text{all } A_i\text{'s in } C} P(A_i).$$

Example. A passenger train has five cars. A passenger is twice as likely to select the middle car (#3) as either adjacent car (#2 or #4), and twice as likely to select an adjacent car as either end car (#1 or #5). What is the probability that one of the three middle cars (a compound event) is selected?

Solution. Let $p_i = P(\text{car } i \text{ is selected})$. We are given

(*) #3 is twice as likely as #2 or #4 $\Rightarrow p_3 = 2p_2 = 2p_4$

(*) #2 or #4 are twice as likely as #1 or #5 $\Rightarrow \begin{cases} p_2 = 2p_1 = 2p_5 \\ p_4 = 2p_1 = 2p_5 \end{cases}$

We wish to find $p_2+p_3+p_4$. First, we can use $\sum_{\text{all } i} P(A_i) = 1$. Namely,

$$\begin{aligned} 1 = \sum_{i=1}^{5} P(A_i) &= P(A_1)+P(A_2)+P(A_3)+P(A_4)+P(A_5) \\ &= p_1 + 2p_1 + 4p_1 + 2p_1 + p_1 \\ &= 10p_1 \end{aligned}$$

$\Rightarrow p_1 = .1 \Rightarrow p_2 = .2,\ p_3 = .4, \text{ and } p_4 = .2 \Rightarrow \boxed{p_2+p_3+p_4 = .8}$

Date : ..../..../....

Subject:

Equally likely outcomes

Definition. Let A and B be events in a sample space Ω. We say that A and B are equally likely if $P(A) = P(B)$.

Suppose an experiment has N equally likely simple events $A_1, A_2, \ldots, A_N$. Let $P(A_i)$ for each i. Then,

$$1 = \sum_{i=1}^{N} P(A_i) = \sum_{i=1}^{N} p = pN$$

$$\Rightarrow \quad p = P(A_i) = \frac{1}{N}.$$

Now, let $A \subset \Omega$ be an event, and let $N(A)$ denote the number of outcomes contained in A. Then,

$$P(A) = \sum_{A_i \in A} P(A_i) = \sum_{A_i \in A} \frac{1}{N} = \frac{N(A)}{N}.$$

Example. Consider a six-sided die, where each side is equally likely to appear. Then, the probabilities of each outcome are

$$P(\{1\}) = P(\{2\}) = P(\{3\}) = P(\{4\}) = P(\{5\}) = P(\{6\}) = \frac{1}{6}.$$

The probability of rolling an even number is then

$$P(\{2,4,6\}) = P(\{2\}) + P(\{4\}) + P(\{6\}) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2} \left(\frac{N(A)}{N} = \frac{3}{6} \right).$$