Information Theory

EE4334

First Course

By Dr. Naser Al-Falahy

UPON THE COMPLETION OF THIS COURSE THE STUDENT SHOULD LEARN:

- ✓ Understand the meaning of information in communication system.
- ✓ How to measure the entropy and data rate of the source.
- ✓ The channel capacity and its limitations
- ✓ Potential algorithms for source coding.
- ✓ Error detection and correction through channel coding techniques.

Recommended Textbook(s):

- S. Haykin; “Communication Systems”, 4th ed.
- Glavieux, “Channel Coding in Communication Networks”, ISTE, 2007.
- Viterbi, and Omura, “Principles of Digital Communication and Coding”, 1979.

WEEK 1

INTRODUCTION TO PROBABILITY

Consider an experiment that can produce a number of outcomes. The collection of all results is called the sample space of the experiment. The power set of the sample space is formed by considering all different collections of possible results. For example, rolling a die produces one of six possible results. One collection of possible results corresponds to getting an odd number. Thus, the subset {1,3,5} is an element of the power set of the sample space of die rolls. These collections are called events. In this case, {1,3,5} is the event that the die falls on some odd number. If the results that actually occur fall in a given event, that event is said to have occurred.

Probability is a way of assigning every "event" a value between zero and one, with the requirement that the event made up of all possible results (in our example, the event {1,2,3,4,5,6}) be assigned a value of one. To qualify as a probability distribution, the assignment of values must satisfy the requirement that if you look at a collection of mutually exclusive events (events that contain no common results, e.g., the events {1,6}, {3}, and {2,4} are all mutually exclusive), the probability that at least one of the events will occur is given by the sum of the probabilities of all the individual events.
The probability that any one of the events {1,6}, {3}, or {2,4} will occur is 5/6. This is the same as saying that the probability of event {1,2,3,4,6} is 5/6. This event encompasses the possibility of any number except five being rolled. The mutually exclusive event {5} has a probability of 1/6, and the event {1,2,3,4,5,6} has a probability of 1 - absolute certainty. For convenience's sake, we ignore the possibility that the die, once rolled, will be obliterated before it can hit the table.

Definition 1 (Probability Based on Logic). This view presupposes we are dealing with a finite set of possibilities which are drawn at random. For example, in drawing from a deck of cards there are a finite set of possible hands in a deck of cards. Such problems are easily solved if we can recognize all possible outcomes and count the number of ways that a particular event may occur. Then, the probability of the event is the number of times it can occur divided by the total number of possibilities. To use a familiar example, we all recognize that there are 52 cards in a deck and 4 of these cards are Kings. Thus, the probability of drawing a King = 4 / 52 = .0769.

Definition 2 (Probability Based on Experience). This second view of probability assumes that *if* a process is repeated a large enough number of times n, and if event A occurs x of these times, then the probability of event A will *converge* on x / n as n becomes large. If we flip a coin many times we expect to see half of the flips turn up heads.
Such estimates will become increasingly reliable as the number of replications (n) increases. For example, if a coin is flipped 10 times, there is no guarantee that exactly 5 heads will be observed -- the proportion of heads will range from 0 to 1, although in most cases we would expect it to be closer to .50 than to 0 or 1. However, if the coin is flipped 100 times, chances are better that the proportions of heads will be close to .50. With 1000 flips, the proportion of heads will be an even better reflection of the true probability.

Definition 3 (Subjective Probability). In this view, probability is treated as a quantifiable level of belief ranging from 0 (complete disbelief) to 1 (complete belief). For instance, an experienced physician may say "this patient has a 50% chance of recovery." Presumably, this is based on an understanding of the relative frequency of might occur in similar cases. Although this view of probability is subjective, it permits a constructive way for dealing with uncertainty.

All probabilities are a type of relative frequency—the number of times something can occur divided by the total number of possibilities or occurrences. Thus,

$$\text{The probability of event A } = \frac{\text{no. of times event A can occur}}{\text{total no. of occurrences}}$$

Several Properties of Probabilities

At this point, many statistical texts would cover the laws and axioms of probability. We will take a less formal approach by introducing only selected properties of probabilities:

1- ***The range of possible probabilities.*** This may seem obvious, but keep in mind that probabilities can be no less than zero and no more than one. A statement that the probability is 110%, of course, is ridiculous.

2- ***Complements.*** We often speak of the complement of an event. The complement of an event is its "opposite," or event NOT happening. For example, if the event under consideration is being correct, the complement of the event is being incorrect. If we are denote an event with the symbol A, the complement may be denoted as the same symbol with a line overhead (A'). The sum of the probabilities of an event and its complement is always equal to one:

$$P(A) + P(A') = 1$$

Therefore, the probability of the complementary of an event is equal to 1 minus the probability of the event:

$$P(A') = 1 - P(A)$$

For example, if the probability of being correct is 0.95, the probability of being incorrect = 1 - 0.95 = 0.05. In contrast if the probability of being incorrect is 0.01, then the probability of being incorrect = 1 - 0.01 = 0.99.

Joint Probability

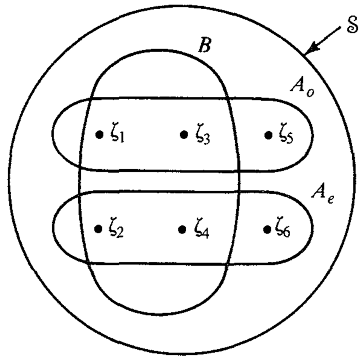
The intersection of events A and B, denoted by (A∩B)or simply A B, is the event that contains points common to A and B. This is the event "both A and B " also known as the joint event AB. Observe that

AB = BA

All these concepts can be demonstrated if the events A and B are such that

$$AB = \emptyset$$

then A and B are said to be disjoint, or mutually exclusive, events. This means events A and B cannot occur simultaneously. In Fig. below, events A_e and A_o are mutually exclusive, meaning that in any trial of the experiment if A_e occurs, A_o cannot occur at the same time, and vice versa. So it describe the probability of both event occurring, P(A∩B) or P(AB)



Conditional Probability and Independent Events

Conditional Probability: One often comes across a situation where the probability of one event is influenced by the outcome of another event. As an example, consider drawing two cards in succession from a deck. Let A denote the event that the first card drawn is an ace. We do not replace the card drawn in the first trial. Let B denote the event that the second card drawn is an ace. It is evident that the probability of drawing an ace in the second trial will be influenced by the outcome of the first draw. If the first draw does not result in an ace, then the probability of obtaining an ace in the second trial is $4/51$. The probability of event B thus depends on whether or not event A occurs. We now introduce the **conditional probability** $P(B \setminus A)$ to denote the probability of event B when it is known that event A has occurred. $P(B \setminus A)$ is read as *"probability of B given A"*

Example An experiment consists of rolling a die once. Let X be the outcome. Let F be the event $\{X = 6\}$, and let E be the event $\{X > 4\}$. We assign the distribution function $m(n) = 1/6$ for n = 1,2,….. 6. Thus, $P(F) = 1/6$. Now suppose that the die is rolled and we are told that the event E has occurred. This leaves only two possible outcomes: 5 and 6. In the absence of any other information, we would still regard these outcomes to be equally likely, so the probability of F becomes $1/2$, making $P(F/E) = 1/2$.

We call $P(F/E)$ the *conditional probability of F occurring given that E occurs,* and compute it using the formula**:**

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

Let us return to the example of rolling a die. Recall that F is the event $X = 6$, and E is the event $X > 4$. Note that $E \cap F$ is the event F. So, the above formula gives **:**

$$P(F|E) = \frac{P(F \cap E)}{P(E)} = \frac{1/6}{1/3} = \frac{1}{2}$$

It's agreed with the calculations performed earlier !!

Independent Events

It often happens that the knowledge that a certain event E has occurred has no effect on the probability that some other event F has occurred, that is, that $P(F/E) = P(F)$. One would expect that in this case, the equation $P(E/F) = P(E)$ would also be true. In fact each equation implies the other. If these equations are true, we might say the F is *independent* of E. For example, you would not expect the knowledge of the outcome of the first toss of a coin to change the probability that you would assign to the possible outcomes of the second toss, that is, you would not expect that the second toss depends on the first. This idea is formalized in the following definition of independent events.

$P(A \cap B) = P(A)P(B)$
$P(F/E) = P(F)$
$P(E/F) = P(E)$

Example It is often, but not always, intuitively clear when two events are independent, let A be the event "the first toss is a head" and B the event "the two outcomes are the same." Then

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

Note: Sine $P(B|A) = \frac{P(AB)}{P(A)}$ & $P(A|B) = \frac{P(AB)}{P(B)}$

It follows that :

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

The above two equations are called **Bayes' rule.** On conditional probability is expressed in terms of the reversed conditional probability

EXAMPLE 1

Two dice are thrown. Determine the probability that the sum on the dice is seven.

Sol:

For this experiment, the sample space contains 36 sample points because 36 possible outcomes exist. All the outcomes are equally likely. Hence, the probability of each outcome is 1/36.
A sum of seven can be obtained by the six combinations: (1,6), (2, 5), (3,4), (4,3) (5, 2), and (6, 1). Hence, the event "a seven is thrown" is the union of six outcomes, each with probability 1/36. Therefore,

$$P('a\ seven\ is\ thrown'') = \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} = \frac{1}{6}$$

EXAMPLE 2

A coin is tossed four times in succession. Determine the probability of obtaining exactly two heads.

A total of $2^4 = 16$ distinct outcomes are possible, all of which are equally likely because of the symmetry of the situation. Hence, the sample space consists of 16 points, each with probability 1/16. The 16 outcomes are as follows:

HHHH	TTTT
HHHT	TTTH
HHTH	TTHT
<u>HHTT</u>	<u>TTHH</u>
HTHH	THTT
<u>HTHT</u>	<u>THTH</u>
<u>HTTH</u>	<u>THHT</u>
HTTT	THHH

Six out of these 16 outcomes are favorable to the event "obtaining exactly two heads" (shown by underline). Because all of the six outcomes are disjoint (mutually exclusive),

P (obtaining exactly two heads) $=\frac{1}{16}+\frac{1}{16}+\frac{1}{16}+\frac{1}{16}+\frac{1}{16}+\frac{1}{16}=\frac{6}{16}=\frac{3}{8}$

Bernoulli Trials

In Bernoulli trials, if a certain event A occurs, we call it a "success." If $P(A) = p$, then the probability of success is p. If q is the probability of failure, then $q = 1 - p$. We shall find the probability of k successes in n (Bernoulli) trials. The outcome of each trial is independent of the outcomes of the other trials. It is clear that in n trials, if success occurs in k trials, failure occurs in $n - k$ trials. Since the outcomes of the trials are independent, the probability of this event is clearly $p^k(1-p)^{n-k}$, that is,

$$P(k \text{ successes in a specific order in } n \text{ trials}) = p^k(1-p)^{n-k}$$

But the event of "k successes in n trials" can occur in many different ways (different orders). It is well known from the combinatorial analysis that there are ways in which k things can be taken from n things (which is the same as the number of ways of achieving k successes in n trials), where

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

This means the probability of k successes in n trials is

$$P(k \text{ successes in } n \text{ trials}) = \binom{n}{k} p^k (1-p)^{n-k}$$
$$= \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

So, tossing a coin and observing the number of heads is a Bernoulli trial with p = 0.5. Hence, the probability of observing k =2 heads in n=4 tosses is

$$P(2heads\ in\ 4\ trials) = \frac{4!}{2!\,(4-2)!}(0.5)^2(0.5)^2 = 6 * \frac{1}{4} * \frac{1}{4} = \frac{3}{8}(the\ same\ answer)$$

Example. Suppose a treatment is successful 75% of the time (probability of success =0 .75). This treatment is used in 4 patients (n = 4). What is the probability of seeing 2 successes in these 4 patients?

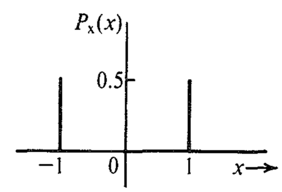
Sol: Let X represent the number of successful treatments, then

$$P(x=2) = C_2^4\,(0.75)^2(0.25)^2 = (6)(0.5625)(0.0625) = 0.2109$$

RANDOM VARIABLES

The outcome of a random experiment may be a real number (as in the case of rolling a die), or it may be non-numerical and describable by a phrase (such as "heads" or "tails" in tossing a coin). From a mathematical point of view, it is desirable to have numerical values for all outcomes. For this reason, we assign a real number to each sample point according to some rule. If there are m sample points $\zeta_1, \zeta_2 \ldots \zeta_m$ then using some convenient rule, we assign a real number $x(\zeta_i)$ to sample point ζ_i (i = 1, 2, …, m). In the case of tossing a coin, for example, we may assign the number 1 for the outcome heads and the number —1 for the outcome tails .

Thus, **x(.)** is a function that maps sample points $\zeta_1, \zeta_2 \ldots \zeta_m$ into real numbers $x_1, x_2, \ldots, x_n$, We now have a **random variable** x that takes on values $\boldsymbol{x_1, x_2, \ldots, x_n}$, We shall use Roman type (x) to denote a random variable (RV) and *italic* type (for example, $\boldsymbol{x_1, x_2, \ldots, x_n}$, etc.) to denote the value it takes. The probability of an RV x taking a value x_1 is $P_x(x_1)$.



DISCRETE RANDOM VARIABLES

A random variable is discrete if there exists a sequence of distinct numbers $\boldsymbol{x_i}$ such that:

$$\sum_i P_x(x_i) = 1$$

Thus, a discrete RV can assume only certain discrete values. An RV that can assume any value from a continuous interval is called a continuous random variable.

Example

Two dice are thrown. The sum of the points appearing on the two dice is an RV x. find the values taken by x, and the corresponding probabilities.

Sol: x can take on all integral values from 2 through 12. There are 36 sample points in all, each with probability 1/36. Dice outcomes for various values of x are shown in the below table. Note that although there are 36 sample points they all map into 11 values of x. This is because more than one sample point maps into the same value of x. For example, six sample points map into x = 7.

The student can verify that

$$\sum_{i=2}^{12} \mathrm{Px}(x_i) = 1$$

Value of x x_i	Dice Outcomes	$P_x(x_i)$
2	(1, 1)	1/36
3	(1, 2), (2, 1)	2/36 = 1/18
4	(1, 3), (2, 2), (3, 1)	3/36 = 1/12
5	(1, 4), (2, 3), (3, 2), (4, 1)	4/36 = 1/9
6	(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)	5/36
7	(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)	6/36 =1/6
8	(2, 6), (3, 5), (4, 4), (5, 3), (6, 2)	5/36
9	(3, 6), (4, 5), (5, 4), (6, 3)	4/36 = 1/9
10	(4, 6), (5, 5), (6, 4)	3/36 = 1/12
11	(5, 6), (6, 5)	2/36 = 1/18
12	(6, 6)	1/36

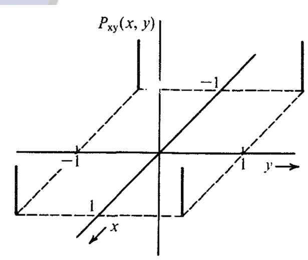
The preceding discussion can be extended to two RVs x and y. The joint probability Pxy(*xi* , *yj*) is the probability that x = *xi* and y = *yj*. Consider, for example, the case of a coin tossed twice in succession. If the outcomes of the first and second tosses are mapped into RVs x and y,then x and y take values 1 and -1. Because the outcomes of the two tosses are independent, x and y are independent, and

Pxy(*xi, yj*) = Px(*xi*) Py(*yj*)

and

Pxy(1, 1) = Pxy(l, -1) = Pxy(-1, 1) = Pxy(-1, -1) = 1/4

These probabilities are plotted in the figure on right,



For a general case where the variable x can take values $x_1, x_2, \ldots, x_n$ and the variable y can take values $y_1, y_2, \ldots, y_m$, we have

$$\sum_i \sum_j P_{xy}(x_i, y_j) = 1$$

This follows from the fact that the summation on the left is the probability of the union of all possible outcomes and must be unity (a certain event).