

# Probability Theory and Statistics

## *A Very Concise Approach*

Eduardo M. B. de A. Tenório

embatbr@gmail.com

The Men Who Stare at Codes

themenwhostareatcodes.wordpress.com

October 21, 2014

### Introduction

According to the *Encyclopaedia Britannica*, **probability theory** is “a branch of mathematics concerned with the analysis of random phenomena” [1]. This science is used to measure the likeliness that an event will occur. **Statistics** is defined by *The Oxford Dictionary of Statistical Terms* as “the study of the collection, analysis, interpretation, presentation and organization of data” [2].

A good example of probability and statistics as a tool is in election polls: researchers use the statistics of the voting population (schooling, region and etc.) to know the probability of victory for each candidate, by taking a sample, generalizing it and calculating the similarity to reality. Machine learning algorithms and insurance policies are also good examples of probability and statistics used to extract valuable information.

This document is a working in progress, where I intend to present the basic topics of Probability Theory and Statistics, so it will change with time. The main reference will be the book **A First Course in Probability - 8th ed.** [3], but probably other materials will be used. If the reader desires to go deeper, further readings are recommended.

### 1 Combinatorial Analysis

A communication system consisting of a transmitter, four antennas (signal repeaters) and a receiver is organized in a straight line. The system is called *functional* as long as two consecutive antennas are not defective. If we have exactly  $m$  of  $n$  defective antennas, what is the probability that the resulting system will be functional? For the particular case where we have  $n = 4$  and  $m = 2$ , there are 6 possible configurations, namely,

0 1 1 0  
0 1 0 1  
1 0 1 0  
0 0 1 1  
1 0 0 1  
1 1 0 0

where 1 means *working* and 0, *defective*. As the resulting system is functional in the first 3 arrangements and not

functional in the remaining 3, it is clear that the probability of having a functional system is  $\frac{3}{6} = \frac{1}{2}$ . So, if we generalize it to undefined  $m$  and  $n$  values, our probability problem reduces to a counting problem. The mathematical theory of counting is formally known as **combinatorial analysis**.

#### 1.1 The Generalized Basic Principle of Counting

If  $r$  experiments that are to be performed are such that the first one may result in any of  $n_1$  possible outcomes; and if, for each of these  $n_1$  possible outcomes, there are  $n_2$  possible outcomes of the second experiment; and if, for each of the possible outcomes of the first two experiments, there are  $n_3$  possible outcomes of the third experiment; and if  $\dots$ , then there is a total of  $n_1 n_2 \cdots n_r = \prod_{i=1}^r n_i$  possible outcomes of the  $r$  experiments [3].

**Example.** How many different 7-place license plates are possible if the first 3 places are to be occupied by letters and the final 4 by numbers?

**Solution.** By the generalized basic principle of counting, the answer is  $26 \cdot 26 \cdot 26 \cdot 10 \cdot 10 \cdot 10 \cdot 10 = 175,760,000$ .

#### 1.2 Permutations

In mathematics, **permutation** means “rearranging the elements of a set in a sequence”, and the number of possible sequences is a function of the set’s size. Given a set of  $n$  elements, we can *permute* its elements according to the formula:

$$P_n = n(n-1)(n-2) \cdots 3 \cdot 2 \cdot 1 = n! \quad (1)$$

As we are arranging the  $n$  elements in  $n$  positions, the first place can be chosen from  $n$  possibilities, the second from  $(n-1)$  and so on, until the last with just one remaining option. We use the  $!$  after a arbitrary number  $n$  to denote the **factorial** of  $n$ .

**Example.** In a group of 6 men and 4 women, how many possible arrangements can we have if (a) put all in the same group, and (b) divide by sex?

**Solution.** (a) We have 10 people to rearrange into 10 positions, so we just do  $P_{10} = 10! = 3,628,800$ .

(b) Using the **basic principle** we have two different experiments (rearranging men and rearranging women). So, our result is just  $P_6 P_4 = 6!4! = 720 \cdot 24 = 17,280$  possibilities.

If our permutation do not preserve beginning and end (called **circular permutation**), the formula is no longer  $n!$  (simple permutation), but  $(n-1)!$ . For example, if we permute  $abc$  circularly, the results  $abc$ ,  $cab$  and  $bca$  are equal. The same to  $acb$ ,  $bac$  and  $cba$ . For  $n$  elements, as the first  $n$  possibilities of the previous “first position” are the same, we do

$$P_{c_n} = \frac{P_n}{n} = \frac{n!}{n} = (n-1)! \quad (2)$$

**Remember:** the order is still importante, but now the line does not have a beginning and an end, but a circular organization.

### 1.3 $k$ -permutations of $n$

A permutation of  $n$  elements in  $k$  positions is also called an “arrangement of  $n$ ,  $k$  to  $k$ ” and can be symbolized by  $A_{n,k}$ . Given the simple permutation shown before ( $P_n = n!$ ), if we permute the  $n$  elements in  $k$  positions, this means that the first position has  $n$  possible values, the second has  $(n-1)$  possibilities and so on, until the last position having  $(n-k+1)$  possibilities.

**Example.** We have 7 athletes in a competition and want to rank them in the three best positions: gold, silver and bronze medals. How many possibilities there is?

**Solution.** The gold medal can be won by all 7 athletes. Ranked the first place, the second has 6 options and the third, 5. This means that  $A_{7,3} = 7 \cdot 6 \cdot 5 = 210$ .

So, we can generalize it, think a little and have the simple formula  $A_{n,k} = \frac{n!}{(n-k)!}$  for an arrangement.

### 1.4 Combinations

If we join the ideas of a circular permutation and arrangement, we develop a new interesting one. In the previous example (the competition), chosen the first three places, we can permute them in  $3! = 6$  ways. If we abolish the different medals and just give a “well done” medal to all three, the six different configurations involving these particular athletes are the same and we can count them as just one. The same to the other 209 configurations. So, now we have  $\frac{A_{7,3}}{6} = 35$ . This is called **combination**,

$$C_{n,k} = \binom{n}{k} = \frac{A_{n,k}}{k!} = \frac{n!}{k!(n-k)!} \quad (3)$$

and is read “combination of  $n$ ,  $k$  to  $k$ ”. Remember that  $r \leq n$  always (but I shouldn’t need to say that).

Now you have the tools to answer the question from the beginning of this section and go deeper into this subject, if you desire. After see the solution, try to read about **binomial theorem** and to generalize the combinatorial analysis to **multinomial coefficients**.

## 1.5 Answering The Problem

Let us go back to the antennas problem from the beginning of the section. Given a number of antennas  $n = 4$ , with  $m = 2$  defectives, we name the working ones as 1 and the defective one as 0. Let us consider two 0s and two 1s, being the set of possibilities  $S = \{0_1, 0_2, 1_1, 1_2\}$ , just to illustrative purposes. The configurations are:

$$\begin{aligned} 0_1 0_2 1_1 1_2 &= 0011 \\ 0_1 1_1 0_2 1_2 &= 0101 \\ 0_1 1_1 1_2 0_2 &= 0110 \\ 1_1 0_1 0_2 1_2 &= 1001 \\ 1_1 0_1 1_2 0_2 &= 1010 \\ 1_1 1_2 0_1 0_2 &= 1100 \end{aligned}$$

Take notice that as there is no difference between the 0s and no difference between the 1s, the index in each number is irrelevant. So, we are back to the configurations from the beginning of this section and realized we are talking about a combination.

Now, generalizing, we have  $n$  antennas, with  $m$  defectives. This means we can line up  $n-m$  functional antennas, with one or none defective antenna between two functional (and also between the transmitter and the receiver). Then, we have  $n-m+1$  possible positions to fill these spaces and must select  $m$  of them. This is  $C_{n-m+1,m} = \binom{n-m+1}{m}$ . Fill it with  $n = 4$  and  $m = 2$  and we find the same answer.

## 2 Axioms of Probability

This section is where we really start to talk about probability theory. The first subsection discusses sample spaces and events, basically *set theory* with a wider vision. So, although I will explain many basic concepts, it is interesting that you have a background in set theory.

### 2.1 Sample Space and Events

In probability theory we deal with problems of uncertainty, with experiments we cannot guarantee the outcome. However, we know the set of possible outcomes for an experiment, as *head* or *tail* for a coin tossing (not an “even number” or “male/female”). This set of possible outcomes for an experiment is called **sample space** and is denoted  $S$ . Any subset of  $S$  is known as an **event**. The following examples will clarify these concepts:

1. If the experiment consists of flipping two coins, then the sample space consists of the following four points:

$$S = \{(H, H), (H, T), (T, H), (T, T)\}$$

where  $H = \text{head}$  and  $T = \text{tail}$ . Valuable events are  $A = \{\text{all outcomes are heads}\}$  and  $B = \{\text{has at least one tail}\}$ , with

$$A = \{(H, H)\} \text{ and } B = \{(H, T), (T, H), (T, T)\}$$

2. If the experiment consists of tossing two dice, then the sample space consists of the 36 points

$$S = \{(i, j) : i, j = 1, 2, 3, 4, 5, 6\}$$

where the outcome  $(i, j)$  is said to occur if  $i$  appears on the leftmost die and  $j$  on the other die. Valuable events are  $A = \{i + j = 7\}$  and  $B = \{\text{all outcomes with } i = 3\}$ , with

$$A = \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\} \text{ and } B = \{(3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6)\}$$

Similar to what (I believe) you learned in set theory, events can be *unified*, *intersected*, *complemented* and etc. As this is piece of cake, any doubt can be answered with a quick consultation to a regular book about set theory.

## 2.2 Axioms of Probability

When we say an event  $E$  has a probability  $P(E) = 30\%$ , that doesn't mean we will have 30 occurrences of  $E$  for each 100 repetitions of the experiment, but that when the number is really big, the outcomes of  $E$  approach this percentage. The leap here is named *relative frequency*, that is the ratio

$$f(E) = \frac{n(E)}{n} \quad (4)$$

where  $n$  is the number of outcomes and  $n(E)$  is the number of occurrences of the event  $E$ . For a small  $n$ , let us say 10,  $n(E)$  hardly is 3. But when  $n \rightarrow \infty$ ,  $f(E) \rightarrow P(E)$ , so

$$P(E) = \lim_{n \rightarrow \infty} \frac{n(E)}{n} \quad (5)$$

Now, let us do it properly, showing the **axioms of probability**. They are few and they are simple. Just a glance and you will think “of course, this is so obvious”.

**Axiom 1**  $0 \leq P(E) \leq 1$

**Axiom 2**  $P(S) = 1$

**Axiom 3**  $P(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} P(E_i)$ , for  $E_i E_j = \emptyset$  when  $i \neq j$

Notice that from Axiom 1 an event cannot have a probability greater than the probability of  $S$  (Axiom 2). If  $P(E) = P(S) \rightarrow E = S$ . From Axiom 3 we can state that  $P(\emptyset) = 0$  (how would you prove that?).

**Example.** If we roll an honest die (all sides are equally probable) we have  $P(\{1\}) = P(\{2\}) = P(\{3\}) = P(\{4\}) = P(\{5\}) = P(\{6\}) = \frac{1}{6}$ . From Axiom 3  $P(\{1, 2, 3\}) = P(\{1\}) + P(\{2\}) + P(\{3\}) = \frac{1}{2}$ .

Hopefully, the reader will agree that the axioms are natural and in accordance with our intuitive concept of probability as related to chance and randomness. Furthermore, using these axioms we shall be able to prove that if an experiment is repeated over and over again, then, with probability 1, the proportion of time during which any specific event  $E$  occurs will equal  $P(E)$  [3].

## References

- [1] “Probability theory, Encyclopaedia Britannica”. Britannica.com. Retrieved 2014-10-11
- [2] Dodge, Y. (2006) *The Oxford Dictionary of Statistical Terms*, OUP. ISBN 0-19-920613-9
- [3] Ross, Sheldon M., 2010, “A First Course in Probability”, 8th ed., Prentice Hall