Statistics and Probability

free and open source book written for educational purposes

Joseph Mehdiyev

A book written for the author's educational purposes

Preface

About the Book

This book is sort of a big notebook to make (or force) the author to self study and understand the field of Probability and Statistics. I have to note that this book is an educational and fun project for the author himself. Through the book, author tries to explain the topics to *himself* .Be careful using the book as the main learning material, since the writer himself is not an expert in the field, there may be mathematical errors in the book.

I like to explain the mathematical concepts in more "traditional" way. I don't like long and complex theorems, lemmas with comically big proofs that reader must pray to understand. Through the book, I try to explain the concepts in everyday language. Of course, rigorous proofs are also provided as they are still an important part of mathematics.

To learn the field and write this book, I used various books from known authors and wikipedia (duh-duh) articles. These are some of the books I used majority of time:

- Larry Wasserman - All of Statistics - A Concise Course in Statistical Inference.
- Dimitri Bertsekas And John N Tsitsiklis - Introduction To Probability
- Mathematical Statistics with Applications by Dennis Wackerly, William Mendenhall, and Richard L. Scheaffer
- Joseph K. Blitzstein - Introduction to Probability
- Ross, Sheldon - First Course in Probability

I want to note that I did not, by any means, plagiarize any contents, diagrams or other things. I simply wrote whatever I learnt through the the brainstorm.

Book's source

Maybe you may already know this, this book is fully open source with its pictures and tex file shared in author's github. You may use the source code for whatever purposes you want to use it for. If you want to contribute, please send a pull request from the github.Currently the book is in development.

How to use the Book

Without the unnecessary historical chapters and big exercises, you may use the book as a revisit or a secondary material. The book shortly and simply explains the concepts and ideas. Important concepts' proofs are provided. However, other proofs explaned in sentences rather than other classic rigorous proofs.

Contents

Part I

Probability

Chapter 1

Introduction to Probability

The concept "probability" is used very often in everyday language to describe the chance of something happening. Mathematically, Probability is a language to quantify uncertainty. This chapter will introduce necessary and basic concepts and namely, **Probability Theory**. We will start the chapter with the elementary *Set Theory*.

1.1 Set Theory

Set Theory is a branch of mathematics that studies *sets*, which we will define shortly. This branch is, like other parts of mathematics, very deep and complex. We will learn only the most important concepts , which is in high-school level, needed to understand later sections and chapters.

We will quickly introduce the concepts and briefly explain them. The reader may skip this section if they already know about sets and their basic properties.

Sets

A **Set** is a collection of different objects, which are called *elements* of the set. The sets are notated as capital letters such as S. If x is an element of a set S, we write $x \in S$. Otherwise we write $x \notin S$. A set with no elements is called **empty set** and is notated as $\emptyset$.

If $x_1, x_2, ..., x_n$ are the elements of the set S, we write:

$$S \in \{x_1, x_2, ..., x_n\}$$

If S is set of all even numbers smaller than 12, we can draw the diagram as:

We can specify our set as a selection from a larger set. If we want to write the set of all even integers, we can write (Here the set of integers is the universal set):

$$S = \{n \in \mathbb{Z} : \frac{n}{2} \text{ is an integer}\}$$

If a set A's elements are also the elements of B, we say that A is a **subset** of B. We can notate it as:

$$A \subseteq B$$

If a set A is subset of B, but is not equal to B, we say that A is **proper subset** of B. We can notate it as:

$$A \subsetneq B$$

Set operations

Union of sets A,B is a set that contains the elements of A and B:

$$A \cup B = \{n : n \in A \vee n \in B\}$$

We can visualize the sets in 2D with circles and their intersections.

Intersection of sets A,B is a set that contains both the elements of A and B:

$$A \cap B = \{n : n \in A \wedge n \in B\}$$

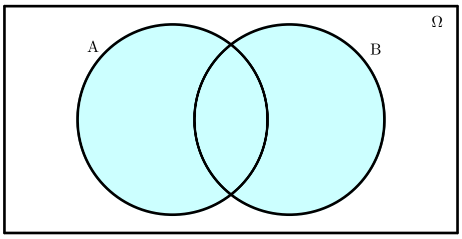

Figure 1.1: $A \cup B$

Sample Space and Events

The Sample Space, usually denoted as S or Ω, is the *set* of all possible outcomes of an experiment. It is also called **universal set**. Subsets of Ω are called **events**. A sample element of Ω is denoted as ω.

Example 1.1.1. If we toss a six sided dice once, then $\Omega = \{1, 2, 3, 4, 5, 6\}$, the even that the side is even is $A = \{2, 4, 6\}$ while $\omega \in \{1, 2, 3, 4, 5, 6\}$

Example 1.1.2. If we toss a two sided coin twice, then

$$\Omega = \{(HH), (TT), (HT), (TH)\} \ \wedge \ \omega \in \{(HH), (TT), (HT), (TH)\}$$

Example 1.1.3. If we toss a 2 sided coin forever, then

$$\Omega = \{\omega = (\omega_1, \omega_2, ...) : \ \omega_i \in \{H, T\}\}$$

Example 1.1.4. Let E be the event that only even numbers appear in the six sided dice toss. Then,

$$E = \{2, 4, 6\}$$

With the new definition, we can make more set operation: **complement** of the event A is a set of elements Ω that do not belong to A.

$$A^c = \{n : n \in \Omega \wedge n \notin A\}$$

difference of the set A from B is a set of elements of A that do not also belong to B

$$A \setminus B = A \cap B^c$$

we say that $E_1, E_2, ..., E_N$ are **disjoint** if

$$A_i \cap A_j = \emptyset$$

A partition of Ω is a sequence of disjoint events such that

$$\bigcup^{\infty} E_i = \Omega$$

Similiar to **monotone functions**, we define **monotone increasing** sequence of sets $A_1, A_2, ...$ as the sequence of sets such that $A_1 \subset A_2 \subset ...$ and $\lim_{n\to\infty} A_n = \bigcup A_i$

Moreover, we can define certain rules similar to the rules of algebra:

Commutative laws	$A \cup B = B \cup A$
Associative laws	$(A \cup B) \cup C = A \cup (B \cup C)$
Distributive laws	$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

And lastly, **DeMorgan's laws** states that

$$\left(\bigcup_{i=1}^{n} A_i\right)^c = \bigcap_{i=1}^{n} A_i^c$$

$$\left(\bigcap_{i=1}^{n} A_i\right)^c = \bigcup_{i=1}^{n} A_i^c$$

Which is, in my opinion, very intuitive and can be easily understood with sketching venn diagrams. These are all of the terminology and notations we will be using for learning the probability.

1.2 Probability Law

To show the probability of a event A, we assign a real number $P(A)$ or $\mathbb{P}(A)$ in some textbooks, called **probability of** A. In other words, $P()$ is a unique function with unique properties that inputs an event A, and outputs its probability.

To qualify as probability, P must satisfy 3 axioms:

Axiom 1 $P(A) \geq 0$ for every A

Axiom 2 $P(\Omega) = 1$

Axiom 3 If $A_1, A_2, ...$ are disjoing:

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

Let's explain the axioms. The first axiom is very simple, a probability can't be negative, since the meaning of the word probability. Second axiom is also very simple, the probability of any possible outcomes happening is 1, since there must be a outcome at the end of the experiment. Third axiom, assume we have 2 disjoint sets. Then

$$P(A \cup B) = P(A) + P(B)$$

This is true simply because sets are disjoint. Similiarly, we can use induction to prove the above property for n sets. Proving for infinite sets are out of scope of this section, therefore we will skip it.

We can derive many properties from these axioms. These are the most simple and intuitive ones:

$$\begin{array}{rcl} P(\emptyset) & = & 0 \\ A \subset B & \implies & P(A) \leq P(B) \\ 0 \leq & P(A) & \leq 1 \\ P(A^c) & = & 1 - P(A) \end{array}$$

And a less obvious property:

Lemma 1.2.1. For events A and B,

$$P\left(A \cup B\right) = P(A) + P(B) - P(A \cap B)$$

Proof. We can rewrite $A \cup B$ as union of $A \setminus B$, $B \setminus A$, and $A \cap B$, since these are the slices of the thing we want to begin with. Moreover, these slices are disjoint, therefore we can apply our third axiom (P is additive):

$$\begin{aligned} P\left(A \cup B\right) &= P\big((A \setminus B) \cup (B \setminus A) \cup (A \cap B)\big) \\ &= P(A \setminus B) + P(B \setminus A) + P(A \cap B) \\ &= P(A \setminus B) + P(A \cap B) + P(B \setminus A) + P(A \cap B) - P(A \cap B) \\ &= P(A) + P(B) - P(A \cap B) \end{aligned}$$

□

1.3 Uniform Probability Distribution