

“Everyone is entitled to his own opinion, but not his own facts.”

–Patrick Moynihan (U.S. Senator)

PROBABILITY 01

N. SRINIVASAN

ABSTRACT. This is a brief introduction to Probability Theory. This paper is broken into several chapters. This is the first of several chapters. Other chapters are to follow. The emphasis is not on rigor, but on understanding. In these pages, what is interesting is not new, and what is new is not interesting.

1. INTRODUCTION

“There is probably no mathematical subject which shares with probability the features that on the one hand many of its most elementary theorems are based on rather deep mathematics, and that on the other hand many of its most advanced theorems are known and understood by many (statisticians and others) without the mathematical background to understand its proofs.”

–J. L. Doob [1]

Probability theory is a branch of Mathematics. Historically, it originated in the games of chance, also called gambling. There are numerous contributors to probability theory– too numerous in fact to list them here.

Why understand probability theory at all?

The author wishes to thank Prof. Nozer Singpurwalla (GWU), Prof. John Kulesza (GMU), Prof. Ron Levy (GMU), Prof. Ed Wegman (GMU), Arthi Srinivasan and Dr. Glenn Jones (CalTech). This work is partly funded by the U.S. Government.

Probability theory provides a means to quantify uncertainty. To cite a few examples, statistical inference, regardless of whether it is based on the frequentists' notion or the Bayesian view, plays a significant role in clinical trials, predicting the reliability and availability of equipment or systems, opinion polls, stock trading, and numerous other areas.

For example, when the weather forecast says that the probability of rain tomorrow is 0.9, what does it mean? If the probability of rain tomorrow is 0, we understand that it is not going to rain tomorrow. It may be cloudy, however. If the probability of rain tomorrow is stated as 1, it implies that it definitely is going to rain tomorrow.

In digital communication theory, such as satellite communications, the information is transmitted as 1's and 0's. However, the transmitted signals are often contaminated by noise. Due to the contamination by noise, the receiver may interpret the received signals erroneously. For example, a transmitted 1 may be interpreted as a 0, and a transmitted 0 may be interpreted as a 1. The noise is unpredictable, and can only be characterized only statistically (probabilistically). This probabilistic characterization is used to decode the signals correctly at the receiver.

What do we mean by the probability of transmitting a 1 is 0?

What do we mean by the probability of transmitting a 1 is 1?

What do we mean by the probability that a 1 is transmitted is 0.5?

We do understand what is meant by probability of 1 and probability of 0. However, what is meant by the received symbol has a probability of 0.5? It means that we cannot say anything about the received signal!

This example illustrates that the information content is not linearly related to the corresponding probabilistic statement.

There are some key results in probability theory that are profound.

Among them:

1. The Central Limit Theorem.

2. The Law of Large Numbers (both the Weak Law and The Strong Law).
3. Markov Chains.
4. Martingale and other limit theorems.
5. Renewal Theory.
6. Time Series Analysis and Forecasting.

The statement of the Central Limit Theorem (CLT) is fairly simple. One can empirically verify the truth of it fairly quickly using a spreadsheet such as the Open Office or even Excel. The OpenOffice is a free download— yes, it is free.

All one has to do to verify the CLT is to create a set of uniformly distributed random numbers in a spreadsheet: 20 Rows by 20 Columns, for example, or use your own number of columns and rows at least greater than or equal to 20. You will understand the reason for this if you play with it for different number of columns and rows. This is simply done using the random number generator `rand`. Add up the rows and columns. Create a histogram of the sums you calculated. you will notice that the histogram resembles the famous “Normal” or the Bell Curve. If you are so inclined, you can divide the column sum (and the row sum) by 20 and compute the “mean” or the average and plot the histogram.

Here is a simple R program used to produce the following figure to illustrate the CLT. 10000 samples of 20 random numbers are drawn from the uniform distribution $[0,1]$. The mean of each sample is calculated and plotted.

```
#Empirical Demonstration of the Central Limit Theorem
# This is a comment character.
#
#The function runif generates uniformly distributed random numbers
#in the interval [0,1]. The number of random numbers
#desired is specified as an argument to this function.
#The function hist creates a histogram.
```

```
library(grDevices)
postscript(file="uniform.ps")
hist(runif(10000), main="")
dev.off
#The function numeric creates a vector of specified length,
# in this case, 10000.
#xmeans is now a vector of length 10000, with each component set to 0.
#
xmeans = numeric(10000)
#The following command computes 10000 means from
#a random sample of 20,drwan from the uniform distribution [0,1].
#The computed means (averages) are stored in xmeans.
for (i in 1:10000) {xmeans[i] = mean(runif(20))}
postscript(file="clt.ps")
hist(xmeans, ylim=c(0,4000))
#The following command creates the mean of the 10000 means.
mean(xmeans)
#In this particular instance, the value turns out to be 0.500547.
#If you ran the program yourself, you may get a different value,
#but close to this one.
#[1] 0.500547
#The folowing sd command produces the standard deviation of the xmeans.
sd(xmeans)
#[1] 0.06435477
#We are going to plot the normal distribution curve along with
#the histogram we already generated. To do this, we need
#a set of X values and a set of Y values.
#The Y values are generated using the standard R function dnorm,
#which produces a set of normally distributed random variables
#from the normal density function,
#for each specified X value. We generate 1000 X values
#so that the curve looks smooth.
xv =seq(0,10,0.01)
#Multiplying by 80 is done for scaling.
yv = dnorm(xv, mean=0.500547, sd = 0.06435477)*500
#The following command overlays the normal pdf over the histogram
#for comparison.
lines(xv,yv)
dev.off()
#You should see a histogram and a normal pdf in the same graph.
```

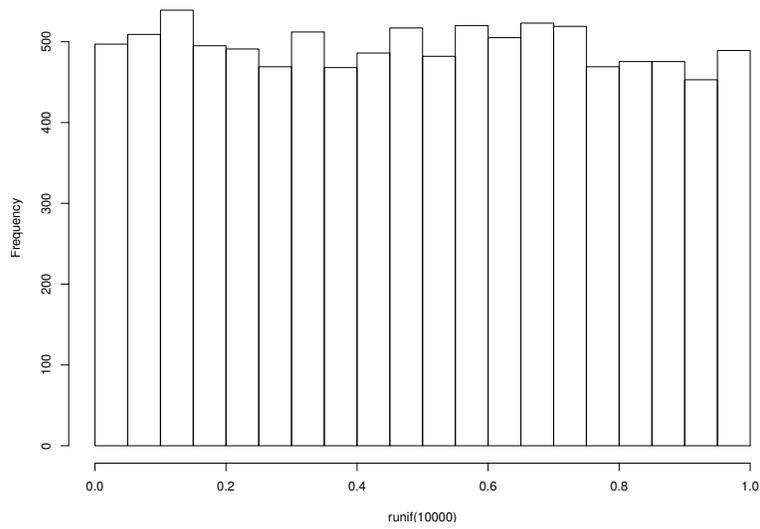


FIGURE 1. The Uniform Distribution

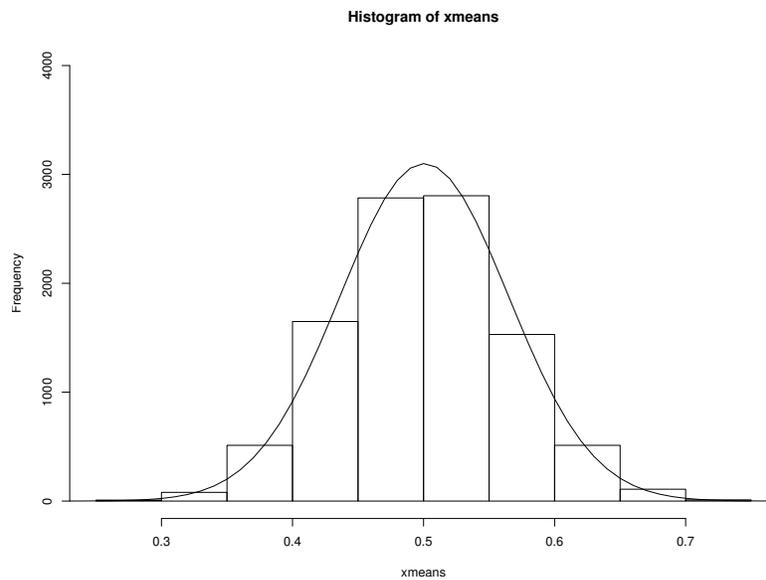


FIGURE 2. The Histogram of the means and its corresponding Normal Density Plot

The CLT states this formally as follows:

Theorem 1. *The distribution of the mean of the random numbers from any distribution tends to the normal distribution as the number in the sample which was used to compute the mean tends to be large.*

This statement is at once simple and startling.

Its simplicity is obvious. It is startling because the CLT says that this happens regardless of the distribution from which the random numbers come from. Yes, you can empirically verify this too by changing the random number generator from rand to any other you choose. Even more, the CLT says that something concrete and deterministic can be said about the set of random numbers. You can get something concrete or deterministic even though the data from which we obtain this concrete information is completely random.

Although you can verify the validity of the CLT in a simple way, mathematically proving this result is rather subtle.

2. A HINT OF PROBABILITY THEORY

Take a simple coin toss— assuming of course the coin is fair in the sense that it is not “loaded” or that both sides are marked alike. We call one side the “Head” and the other side the “Tail”.

One would expect that in repeated tosses the heads and tails will appear equally often in the sense that if you toss this fair coin 100 times, you would expect the heads to show up approximately 50 times and the tails to appear approximately 50 times. This is expressed symbolically:

$$P(H) = P(T) = \frac{1}{2}$$

It is read as the Probability of a Head and the Probability of a Tail is equal to $\frac{1}{2}$.

This experiment may not produce exactly $\frac{1}{2}$ but something close to it. As the number of repetitions increases, the ratio will approach $\frac{1}{2}$ even more closely.

One can perform a similar experiment with a die that has six faces, each marked with numbers from 1 to 6. The probability of any one of these numbers appearing in any given toss is approximately $\frac{1}{6}$, assuming the die is fair.

Symbolically:

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

The *set* of *all* possible outcomes in the experiment of tossing a single fair coin is $\{H, T\}$

The totality of *all* possible outcomes in the second experiment of tossing a single fair die is $\{1, 2, 3, 4, 5, 6\}$

The *set* of *all* possible outcomes in the performance of an experiment is called the Sample Space. The sample space is denoted by the Greek symbol Ω . In other words, the performance of the experiment should result in one of the outcomes that constitutes the sample space. The performance of the experiment should not result in any outcome not listed in the Sample Space

The sum of the probabilities of all the outcomes in the sample space should equal 1.

That is, $P(H) + P(T) = 1$. This is regardless of whether the coin is fair or not.

For example, if the coin is unfair, it may turn out that the $P(H) = 0.6$ and the $P(T) = 0.4$.

In such a case, the $P(H) + P(T) = 0.6 + 0.4 = 1.0$.

In defining the Sample Space, we have tacitly introduced the concept of a *set*.

The above examples illustrated a discrete sample space. There are other instances where the sample space is continuous. Some examples of a continuous sample space are taking the weights of patients, measuring the height of buildings, measuring the voltage across the terminals of an electrical outlet and so on.

Another Example : Brownian Motion

The following is an example of Brownian Motion, also called the Random Walk or The Drunkards Walk.

A particle suspended in a fluid undergoes random movements, or if you like, a drunk drunkard (well, a drunkard can be occasionally

sober) starts walking, and takes random steps— sometimes to the right and sometimes to the left. Also, this example can also be considered a simulation of tossing a fair coin a number of times, recording it as 1 when heads show up, and recording it as 0 when tails show up. add the recorded numbers and plot them. The program written in R is shown below.

The mean value of the total number of steps is also plotted. This mean can be regarded as an empirical demonstration of the Law of Large Numbers.

However, what is depicted here is only one realization of the Brownian Motion. If you run the program the second time, you will see an entirely different path (or as it is commonly called, a realization). Brownian Motion is a stochastic process that evolves over time.

If you want to see the same process over and over again, that can be done by including a seed in the random number generator.

The random number generator is not a true random number generator. All implementations of random number generators are pseudo-random number generators in the sense that after a certain number of random numbers— depending on the one used— the numbers will repeat themselves, some sooner, some later.

```
# Simulation of the Brownian Motion using R.
#This simulation consists of random steps either
#to the right or to the left by a fixed amount.
#The step size = 0.001
# The random number generator used is runif (Built-in
#uniform random number generator in R that generates
#random numbers in (0,1).
#Read n, where n is the number of steps.
#
#print("Input the number of steps to be taken")
#print("After you input the numerical value,
#hit the carriage return twice")
#Clear the work space
rm(list =ls())
#
n = 10000
```

```
steps = 1:n
y = runif(n)
#
#n = scan()
#
step_size = 0.001
s= step_size
#
deviation = 1:n
deviation [1:n] = 0.0
MeanDeviation = 1:n
MeanDeviation [1:n] = 0.0
#
for (i in 1:(n-1)){
m = ifelse (y [i] < 0.5, -s, s)
deviation [i+1] = deviation [i] + m
MeanDeviation [i+1] = (deviation [i+1])/(i+1)}
#
#print(deviation) #This will create a lot of output!
#print(MeanDeviation) # This too!
print(MeanDeviation [10])
print(MeanDeviation [n/5])
print(MeanDeviation [n/4])
print(MeanDeviation [n/3])
print(MeanDeviation [n])
#
#Juggle a lot to create labels for the graph
z = abs(max(deviation))
zz = abs(min(deviation))
zzz = ifelse (z < zz, min(deviation), max(deviation))
#Labels finished
# Plot the Results and output to a file
postscript(file ="b1.ps")
plot(steps, deviation, type="l", xlab ="Number of Steps",
ylab ="Cumulative Deviation", col = "green")
lines(steps, MeanDeviation)
text(n/4, 0.002, "Running Mean of the Deviation")
text(0.8*n, zzz/2, "Cumulative Deviation")
title("Brownian Motion or The Drunkards Walk")
dev.off()
#
X11()
```

```
postscript(file = "mean.ps")  
plot(steps, MeanDeviation, type = "l", col = "red")  
dev.off()  
#Finished!
```

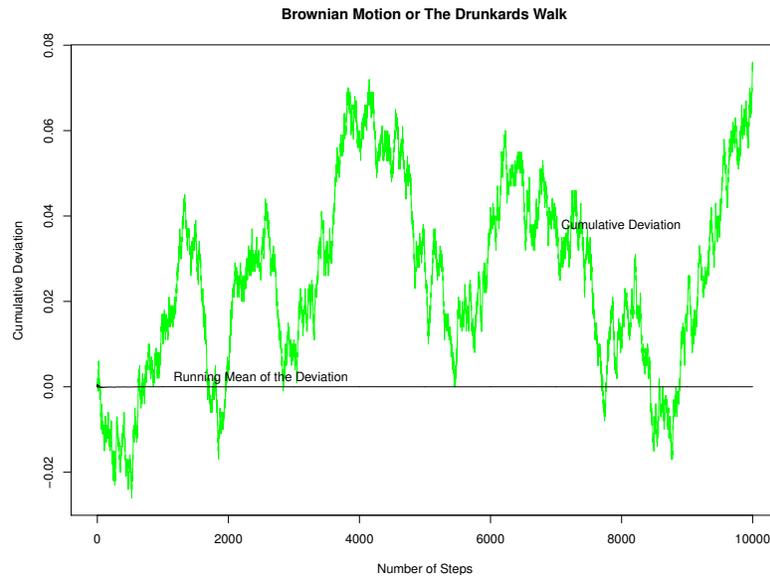


FIGURE 3. Brownian Motion

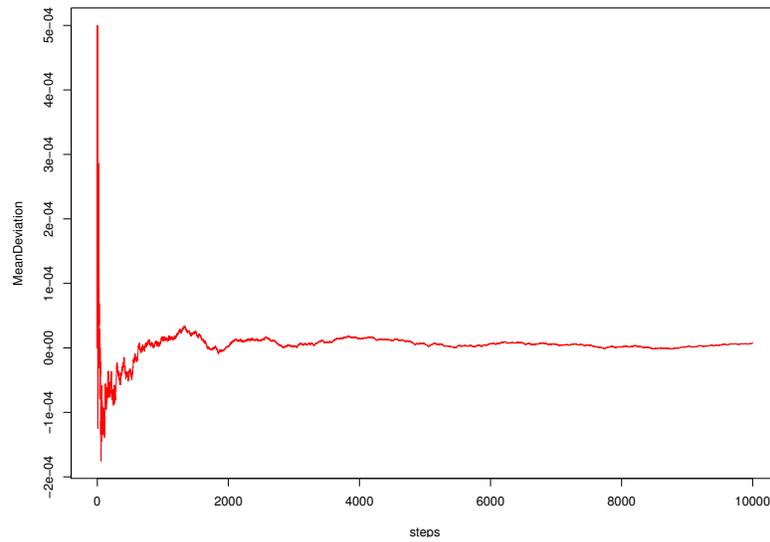


FIGURE 4. The Moving Average or The Mean of the Steps Taken

3. BARELY BASIC SET THEORY

One must start near to go far. Modern Mathematics is founded on set theory.

The concept of set and the element of a set will be taken as self evident primitive notions. Any attempt to define these will lead to a circular reasoning. For example, a set may be defined as an aggregate of elements. Then, one has to define what is meant by an aggregate. So, we take it for granted that everyone sort of knows what a set is and what it means for an element to belong to a set.

The notation for a set is $\{ \}$. The elements that belong to the set are placed within the braces, separated by commas.

The set of Natural Numbers $N : \{1, 2, 3, \dots\}$

The set of Integers $Z : \{ \dots -4, -3, -2, -1, 0, 1, 2, 3, 4, 5, \dots \}$

The set of Rational Numbers $Q: \{ \frac{m}{n} : m, n \in Z, \text{ and } n \neq 0 \}$

The set of Real Numbers R

The set of Irrational Numbers $: \{R \setminus Q : \text{The set that remains after the Rational Numbers are removed from the set of Real Numbers } R\}$

Some authors use the symbol J to denote irrational numbers. The irrational numbers are more numerous than the rational numbers.

The rational numbers are countable. "Countable" means exactly that : You can put the rational numbers into one-to-one correspondence with the set of natural numbers 1, 2, 3, 4, and so on. Countability is also referred to as denumerable. The set that is not countable is called uncountable or non-denumerable.

Probability theory has been axiomatized using set theory. One may be familiar with functions such as $y = f(x)$. In this case, this function assigns a value for y for each value of x . As will be seen below, probability theory is dependent on functions defined over sets called "Events". Kolomogorov axiomatized probability theory and provided the foundational framework for the modern probability theory as practiced today. . In general, proofs in this paper are given without recourse to diagrams and figures. Figures and diagrams are useful in developing an intuitive feel for the topic. However, any figure represents only a particular instance of the case under consideration.

Set theory is introduced only to the extent necessary. Set theory is a vast subject in its own right.

In mathematics, there is always a frame of reference called the “Universe of Discourse”. This “Universe of Discourse” will vary depending on the context. This context is sometimes explicitly stated or it is understandable from the context under discussion. For example, in plane geometry, the universe of discourse is the set of all points in the plane, for example as found in a graph paper. Points in a plane are represented by (x, y) . In the geometry of three dimensions, we need three points, such as (x, y, z) . Usually it is clear what the Universal Set is. Otherwise, one may have to infer this from the context.

For our purposes, the universe of discourse will be the set of complex numbers, the real numbers being a subset of the complex numbers, by setting the imaginary part to zero. There is nothing “complex” about the complex numbers. Similarly, there is nothing “imaginary” about the so-called imaginary numbers. The ancients did not know what to do when they encountered the square root of negative numbers. Hence, they coined the term “imaginary numbers” to the square root of negative numbers.

In all our discussions, we will consider only non-empty universal sets.

The universal set is denoted by U .

Here, we will use i to denote $\sqrt{-1}$. sometimes, we may switch to j to denote the $\sqrt{-1}$. The context will make this clear. One reason for this is that in electrical engineering, i is the symbol used to denote the electrical current in a circuit. Thus the need to resort to the symbol j . Here, the symbols i and j will be used interchangeably and synonymously. No confusion should arise as the context will make it clear.

Braces are used to name sets. For example, as we have seen already, the set of Natural Numbers are represented by $\{1, 2, 3, 4, \dots\}$.

Some examples of other sets:

$$\{1, 3, 5, 7, \dots\}$$

$$\{0.1, 0.7, -1.1, 2.8 \dots\}$$

Sets may be finite or infinite.

The set of all natural numbers are infinite. So are the set of all real numbers \mathbb{R} .

Some Examples of finite sets are:

$$\{2, 5, 8, 1, \}$$

$$\{9, 8, 7, 5.2, 3.14, 2\}$$

There are other ways of specifying sets, such as by using the set builder notation, as follows:

$$\{x \mid P(x)\}$$

The above statement is read as the set of *all* objects x that make the statement $P(x)$ true.

The word *all* in the above statement is important. Otherwise, one may end up with several sets that are subsets of the intended set comprised of elements that do satisfy the property $P(x)$.

Some examples using the set builder notation:

$$\{x \mid 1 \leq x \leq 4, x \text{ is an integer}\} \text{ specifies the set } \{1, 2, 3, 4\}.$$

It is read as the set of *all* integers equal to or greater than 1 and less than or equal to 4.

From now on, the words “object”, “element”, “and ”member“ will be used interchangeably and will mean the same thing.

Membership in a set is denoted by the symbol \in .

For example, a is a member of the set X will be denoted by $a \in X$.

a is not an element of the set X is denoted by $a \notin X$.

Examples:

$1 \in \mathbb{N}$.

$0.12 \notin \mathbb{N}$.

Here, as noted before, \mathbb{N} stands for the set of all natural numbers.

Definition 1. *Subset* : A is a subset of B if every element of the set A is also a member of the set B .

The phrase, A is included in B is also used to convey the same concept.

This is represented by $A \subset B$.

Some authors also use $A \subseteq B$. This implies that A is a subset of B , and A is also equal to B .

$A \subset B$ is also sometimes represented by $B \supset A$.

A is a subset of B means that B contains A .

A set is always a subset of itself by the definition (i.e. $A \subset A$. Every member of A is a member of A .)

Examples:

$$\{1, 2, 3\} \subset \{1, 2, 3, 4, 5\}$$

$$\{1, 2\} \subset \{1, 2\}$$

Definition 2. *Equality of Sets:*

Two sets are equal if and only if they have the same elements.

That is, if $A = B$, A and B have the same elements.

In other words, every element of A is also an element of B , and every element of B is also an element of A .

Another way of stating the equality of two sets is:

$$A \subset B \text{ and } B \subset A$$

All the above statements of equality of two sets directly flow from the definition.

Examples that demonstrate the concept of equality of two sets:

$$\{1, 2, 3, 4, \} = \{1, 2, 3, 4, \} = \{4, 3, 2, 1 \} = \{4, 4, 4, 4, 3, 3, 3, 2, 2, 1 \}$$

In listing the elements of a set, the order is irrelevant, and also the same element may be repeated any number of times. These two attributes follow from the definition of a set and the set equality.

Definition 3. *The empty set is denoted by \emptyset .*

The empty set has no elements. The empty set can also be represented by $\{ \}$

Also, the empty set is a subset of every set. That is, $\emptyset \subset A$, where A is any set.

A Note on Null Sets Vs. Empty Sets:

In Axiomatic Set Theory, it is now more common to refer to the Empty Set, rather than to the Null Set, to avoid the confusion with the concept of a Null Set as used in Measure Theory. Probability Theory is based on Measure Theory.

Null Sets may be uncountable. Or, stated in another way, uncountable Sets may be Null Sets. For more on this subject, refer to books on Set Theory, or books on Measure Theory.

The empty set is not the same thing as nothing. It is a set with nothing inside it, and a set is something in the sense it exists. This often causes difficulty among those who first encounter it. Think of a set as a package containing its elements. An empty package may be empty, but the package itself certainly exists.

The following are some mathematical properties of the empty set:

The empty set is NOT the number 0.

The empty set is a subset of any set A.

For any set A, the union of A with the empty set is A.

For any set A, the intersection of A with the empty set is the empty set.

For any set A, the Cartesian product of A and the empty set is empty.

The only subset of the empty set is the empty set itself.

An example: Consider the set $\{\emptyset\}$, whose only member is the set \emptyset . That is, a set within a set. Note that $\{\emptyset\} \neq \emptyset$.

The number of elements in the empty set is zero; That means, its cardinality is zero. In particular, the empty set is finite.

Some warnings:

1. $A \notin A$. That is, a set is not a member of itself.
2. $x \neq \{x\}$. The left hand side is an element, while the right hand side is a set.

For example, each of the following is false:

$$\{x, y\} \in \{x, y\}$$

$$\emptyset \in \emptyset$$

$$\emptyset = \{\emptyset\}$$

$$5 = \{5\}$$

Definition 4. *Union of two sets:*

The union of two sets A and B is the set of all elements that are in A or B or both.

The union of two sets is represented thus: $A \cup B$

If $C = A \cup B$, then, every element of C belongs to at least one of A or B or both.

Examples to demonstrate the union of two sets:

$$1. \{1, 2, \} \cup \{3, 4, 5\} = \{1, 2, 3, 4, 5\}$$

$$2. \{x \mid x > 0\} \cup \{x \mid x > 2\} = \{x > 0\}$$

Note: The Real Line that represents all the real numbers is the open set: $\{-\infty, +\infty\}$.

Also, note that ∞ is not a number, but merely a symbol.

$$3. \{x \mid x > 0\} \cup \{x \mid x > -\infty\} = \{x \mid x > -\infty\}$$

The definition of the union of two sets can be easily extended to multiple sets:

Let $A_1, A_2, A_3, A_4, \dots, A_n$ be arbitrary sets, each being a subset of the universal set U .

Let A represent the union of the above sets.

$$\text{Then, } A = A_1 \cup A_2 \cup A_3 \cup A_4 \dots \dots \cup A_n$$

A consists of all the elements that belong to either A_1 , or A_2 , or A_3 , \dots or A_n or all of them.

To rephrase it, every element of A_i belongs to A .

$$\text{This is concisely represented as } A = \bigcup_{i=1}^n A_i$$

Definition 5. *Intersection of two sets.*

The intersection of two sets A and B is the set of all elements that are common to both A and B .

If A and B share no elements in common, their intersection is the null set \emptyset

Symbolically, the intersection of A and B is represented as $A \cap B$.

Examples to demonstrate the intersection of two sets:

$$1. \{1, 2, \} \cap \{3, 4, 5\} = \emptyset$$

$$2. \{x \mid x > 0\} \cap \{x \mid x > 2\} = \{x > 2\}$$

Note: The Real Line that represents all the real numbers is the open set: $\{-\infty, +\infty\}$.

$$3. \{x \mid x > 0\} \cap \{x \mid x > -\infty\} = \{x \mid x > 0\}$$

The definition of the intersection of two sets can be easily extended to multiple sets:

Let $A_1, A_2, A_3, A_4, \dots, A_n$ be arbitrary sets, each being a subset of the universal set U .

Let A represent the intersection of the above sets.

$$\text{Then, } A = A_1 \cap A_2 \cap A_3 \cap A_4 \dots \dots \cap A_n$$

A consists of all the elements that are common to A_1 , and A_2 , and A_3, \dots and A_n .

$$\text{Symbolically, } A = \bigcap_{i=1}^n A_i$$

Definition 6. *Definition: Disjoint Sets*

For any two arbitrary, nonempty sets A and B if $A \cap B = \emptyset$, then A and B are called disjoint sets.

If any one of the sets A or B is an empty set, then the above statement is trivially true.

Let $A_1, A_2, A_3, \dots, A_n$ be any arbitrary collection of sets.

If $A_i \cap A_j = \emptyset$, for $i \neq j$, for all i and j , then the collection of sets $A_i, i = 1, 2, 3, \dots, n$ is said to be pairwise disjoint.

4. A FEW EXAMPLES

1. A set, as mentioned before, is a collection of objects or things. Numbers happen to be objects too. To generalize this notion, here are some examples:

A collection of books one has is a *set* of books. In a park, you may find a *set* of trees. In a rock garden, you may find a *set* of rocks.

1, 2, 3, 4 is a *set* of numbers represented thus: $\{1, 2, 3, 4\}$.

2. Sets may have a few elements. These kinds of sets are called finite sets.

Sets may also have an indefinite number of elements such as 1, 2, 3, 4, In this case, we call them infinite sets.

There are two kinds of infinite sets: 1. Countable, and 2. Uncountable.

Countable means that you can actually count the elements of the set as one, two, three, four, and so on.

The finite sets are countable in the sense that the counting process comes to an end, or the set has no elements.

The set of fingers in both your hands is an example of a finite set.

Uncountable sets have the property that you cannot count the elements of the set. The Real Numbers are an example of an uncountable set. Even more, the elements of the set contained in the interval $[0, 1]$ are also uncountable. These are said to be uncountably infinite sets.

The set of points on a line is an example of an uncountably infinite set or a non-denumerable set.

The set that has a countably infinite number of elements is called denumerable or in some books, enumerable.

3. We make a list enclosed in braces to depict a set. The order of the listing is not important. Thus, $\{a, b, c, d, \}$ = $\{c, b, d, a\}$

4. When we state how many elements are in a set, we do not count a member of the set more than once. Thus, $\{0, 1, 1, 1, 1, 0\}$ = $\{0, 1\}$. The number of elements in this set is just two

5. The concept of a point, line, plane, etc., is assumed to be self evident and is not defined, as mentioned before.

6. The statement that b is an element of a set A is represented by $b \in A$

It can also be stated as b belongs to the set A .

The symbol \in is the greek letter epsilon.

7. The statement that b is not an element of a set A is represented by $b \notin A$.

To illustrate, let $A = \{1, 4, 9, 16\}$.

$$1 \in A$$

$$2 \notin A$$

$$3 \notin A$$

$$4 \in A$$

$$5 \notin A$$

$$9 \in A$$

$$10 \notin A$$

$$16 \in A$$

8. A set that has only one element is called a singleton or a unit set.
Example: $\{7\}$. This set has only one element 7.

9. $A = 2, 3, 5, 7, 8$ is not a set. To denote a set, the elements should be enclosed in braces as follows:

$$A = \{2, 3, 5, 7, 8, \}$$

10. Definition: *Solution Set*

A *solution set* means the set of all those numbers which when substituted in the corresponding equation satisfy that equation.

Example: $x^2 = 4$. The solution set for this equation is $\{+2, -2\}$.

11. Notation for the null set is \emptyset .

$\emptyset \neq 0$. that is, the empty set is not equal to the number zero.

$\emptyset \neq \{0\}$. $\{0\}$ has one element 0.

12. Write the solution set for the following equations:

$$4 - 4, 4 \times 0, 0/4$$

$$\text{Answer: } \{0, 0, 0\} = \{0\}$$

13. The set of all integers, negative, positive, and zero:

$$\{\dots - 3, -2, -1, 0, 1, 2, 3, 4, \dots\}$$

14. The set of all nonnegative integers:

$$\{0, 1, 2, 3, 4, 5, \dots\}$$

15. The set of all positive integers:

$$\{1, 2, 3, 4, 5, 6, \dots\}$$

16. What is the set A of all integers between 1 and 2?

Answer: $A = \emptyset$

17. What is the set A of all integers between -1 and 1 ?

Answer: $A = \{0\}$

18. *Set-builder notation:*

An example: $A = \{x \mid x \text{ is an integer, and } x > 8\}$

The above statement is read as “The set of *all* integers x greater than eight.”

Another example: $A = \{x \mid x \text{ is an integer, and } 1 \leq x < 8\}$

The above statement is read as “The set of *all* integers x greater than or equal to 1, but less than eight.”

19. Definition: *Prime Number:*

A Prime Number is any positive integer greater than 1 which is divisible by no positive integers other than itself and 1.

The number 1 is not included in the set of Prime Numbers by definition.

Example: List the set A of all Prime Numbers less than or equal to 17.

$A = \{2, 3, 5, 7, 11, 13, 17\}$

5. INTERVALS

In this section, the open intervals, half open intervals, and the closed intervals will be defined along with their notation.

Intervals are sets. Intervals occur quite frequently in mathematics, and hence, for brevity, a special notation has been adopted to denote

the sets that represent intervals. These are indicated below for the corresponding intervals.

Let a , b , and x be any numbers from the set of real numbers.

Also, let $a < b$.

Closed interval:

$\{x \mid a \leq x \leq b\}$. This set is called the closed interval. The notation for this is $[a,b]$.

The concept behind the closed interval is that both the end points a and b are included in the interval.

Example:

$\{x \mid -1.1 \leq x \leq 2.3\}$ is denoted by $[-1.1,2.3]$

Open interval:

$\{x \mid a < x < b\}$ is called the open interval. The notation for this is (a,b) .

In the case of the open interval, the end points a and b are excluded in the interval.

Example:

$\{x \mid -1.1 < x < 2.3\}$ is denoted by $(-1.1,2.3)$

Half-open or semi-closed intervals:

$\{x \mid a \leq x < b\}$ is called the half-open interval. The notation for this is $[a,b)$.

Example:

$\{x \mid -1.1 \leq x < 2.3\}$ is denoted by $[-1.1,2.3)$

In this case, a is included in the interval, and b is not part of the interval (That is b is excluded).

$\{x \mid a < x \leq b\}$ is also called the half-open interval. The notation for this is $(a,b]$.

In this instance, a is excluded and b is included in the interval.

Example:

$\{x \mid -1.1 < x \leq 2.3\}$ is denoted by $(-1.1,2.3]$

$(a, a) = \emptyset$

Example:

Let $A_n = [0, \frac{1}{n}]$, where $n \in \mathbb{N}$, where \mathbb{N} is the set of natural numbers.

Then, $\lim_{n \rightarrow \infty} \bigcap_{i=1}^n A_i = \emptyset$

Because,

$A_1 = [0,1], A_2 = [0,\frac{1}{2}], A_3 = [0,\frac{1}{3}] \dots A_{100} = [0,\frac{1}{100}] \dots$

Example:

Let $X = [-2,3)$ and $Y = [3,8]$

$X \cap Y = \emptyset, X \cup Y = [-2,8]$

Note that X and Y are disjoint.

Length of the interval $X = 5$. Length of $Y = 5$.

Length of $X \cup Y = 10$.

example:

A point has zero dimensions. i.e., a point has zero length, zero width, and zero height. It is true to visualize a point, one has to make a mark on a paper, which obviously has dimensions— otherwise you cannot see it. But mathematically, a point has no dimensions— regardless of

whether it is in one dimension such as on a line, or in two dimensions such as a plane, or in a space of n dimensions.

Similarly, a line has zero width and zero height. However, it has a length.

The above two statements will have a significant meaning in probability theory through its relationship with measure theory, because probability is defined as a measure.

6. PARTITION

Let the non-empty set $A = B_1 \cup B_2 \cup B_3 \dots B_n$

If for any sets B_i, B_j for all $i, j = 1$ to n , either $B_i = B_j$, or $B_i \cap B_j = \emptyset$, for $i \neq j$, then the sets B_i are said to be a partition of A .

Examples:

Let $N = \{1, 2, 3, 4, \dots\}$,

$B_1 = \{1, 3, 5, 7, \dots\}$, and $B_2 = \{2, 4, 6, 8, 10, \dots\}$

The sets B_1 and B_2 are a partition of N .

In other words, a partition implies no overlap among the sets that constitute the partition.

Another example:

Consider a closed interval $[a, b]$. A partition of $[a, b]$ is defined as:

$a = x_0 < x_1 < x_2 < x_3 \dots < x_n = b$ and

$[a, b] = \bigcup [x_0, x_1), [x_1, x_2) \dots [x_{n-1}, x_n]$

7. THE INTEGRALS

Definition 7. *Subdivision*

A subdivision Δ of an interval (a,b) is a set of numbers or points such that

$$a = x_0 < x_1 < x_2 < \dots < x_n = b$$

A subdivision involving $(n + 1)$ points divides the interval into n adjoining subintervals $(x_0, x_1), (x_1, x_2) \dots, (x_n, x_{n-1})$

Definition 8. *Definition: The Norm of a Subdivision*

The norm $\|\Delta\|$ of a subdivision Δ is

$$\|\Delta\| = \max (x_1 - x_0, x_2 - x_1, \dots, x_n - x_{n-1})$$

$\|\Delta\|$ is the length of the largest of the subintervals.

Definition 9. *Definition: Riemann Integral*

The Riemann integral of $f(x)$ from a to b , with respect to x is defined as

$$\int_a^b f(x) dx = \lim_{\|\Delta\| \rightarrow 0} \sum_{k=1}^n f(\xi_k)[x_k - x_{k-1}]$$

To have $\|\Delta\| \rightarrow 0$, n must $\rightarrow \infty$

The Riemann sum is defined as:

$$\sum_{j=1}^n f(\xi_j)[x_j - x_{j-1}]$$

where ξ_j is any point in the interval $[x_j - x_{j-1}]$

The value of the Riemann integral is obtained by letting the largest of the partitions tend to 0. The result is given by:

$$S = \int_a^b f(x) dx$$

An Example of Riemann Integration: Numerical Computation: A Python Program:

```
# Calculate the area under a curve using Riemann Integration
#
```

```

# Example Function y = x^2
#
# This program integrates the function from x1 to x2
# x2 must be greater than x1, otherwise the program will print an error message
#
x1 = float(input('x1='))
x2 = float (input('x2='))

if x1 > x2:
    print('The calculated area will be negative')

# Compute delta_x for the integration interval
#
delta_x = ((x2-x1)/1000)
j = abs ((x2-x1)/delta_x)
i = int (j)
print('i =', i)
# initialize
n=0
A= 0.0
x = x1
# Begin Numerical Integration
while n < i:
    delta_A = x**2 * delta_x
    x = x + delta_x
    A = A + delta_A
    n = n+1
print('Area Under the Curve =', A)

```

The Darboux Integrals:

One of the difficulties in the Riemann's definition of the integral arises because ξ_j can take any value in the interval $(x_j - x_{j-1})$.

What is this difficulty?

If you ask 10 people to compute the Riemann integral using the above definition, each answer will be quite different depending on the value of ξ_j each one chose in computing the Riemann sum, however close their sums may be to each other depending on the granularity of

the subdivision.

Darboux proposed to remedy this by introducing the upper and lower Darboux sums.

Given a function f defined on an interval $[a,b]$ and a subdivision as defined above, we define:

$$M_j = \sup\{f(x) | x_{j-1} \leq x \leq x_j\}$$

$$m_j = \inf\{f(x) | x_{j-1} \leq x \leq x_j\}$$

The upper Darboux sum is:

$$\bar{S}(P; f) = \sum_{j=1}^n M_j [x_j - x_{j-1}]$$

The lower Darboux sum is:

$$\underline{S}(P; f) = \sum_{j=1}^n m_j [x_j - x_{j-1}]$$

The upper Darboux integral over $[a,b]$ is defined as:

$$\int_a^b f(x) dx = \inf \bar{S}(P; f) \text{ for all subdivisions } P$$

The lower Darboux integral over $[a,b]$ is defined as:

$$\int_a^b f(x) dx = \sup \underline{S}(P; f) \text{ for all partitions } P$$

(Note: sup is nothing but the maximum, and inf is the minimum value in that interval.)

However, the value of the sums will be the same regardless of how the intervals are specified because, as mentioned before, a point has no dimensions.

Instead of randomly picking some ξ in the interval, Darboux's approach fixed its value in the aforementioned manner.

Thus, given a subdivision, everyone who does the computation will arrive at the same value for the approximation of the integral *for that*

subdivision.

The Darboux integrability condition:

Let f be a bounded function on $[a,b]$. This function is Riemann integrable over this interval if and only if the upper and lower Darboux integrals are equal.

This method of integration was first proposed by Volterra. However, these integrals bear Darboux's name because they are defined in terms of the sums defined by him.

Darboux's definition is a restatement of the Riemann integral, with the exception of how the ξ is defined.

The Stieltjes Integral:

This is the integral that is often used in probability theory and Statistics, although quite a bit of probability theory can be handled quite nicely using the Riemann integral.

The Stieltjes integral of $f(x)$ from a to b , with respect to $\alpha(x)$ is defined as

$$\int_a^b f(x) d\alpha(x) = \lim_{\|\Delta\| \rightarrow 0} \sum_{k=1}^n f(\xi_k) [\alpha(x_k) - \alpha(x_{k-1})]$$

Where $x_{k-1} \leq \xi_k \leq x_k$

$$\lim_{\|\Delta\| \rightarrow 0} \sum_{k=1}^n f(\xi_k) [\alpha(x_k) - \alpha(x_{k-1})]$$

The above limit, if it exists, must hold for *all* partitions of the interval $[a, b]$ for the integral to exist. This point is significant especially where the functions $f(x)$ or $\alpha(x)$ are discontinuous in the interval of integration.

To have $\|\Delta\| \rightarrow 0$, n must $\rightarrow \infty$

The need to perform the integration of discontinuous functions arises in probability theory. For example, the Cumulative Distribution Function (CDF) of a discrete probability distribution is discontinuous. To compute the Expectation of such a discrete random variable, one has

to perform the integral (or summation) of a discontinuous CDF.

The Stieltjes integral has the unique feature that it is defined for continuous and discontinuous functions.

Some examples of Stieltjes integral:

Example 1:

Let the interval $[a, b] = [0, 2]$

And let $f(x) = \alpha(x) = 0$ over the interval $0 \leq x \leq 1$

And let $f(x) = \alpha(x) = 0$ over the interval $1 < x \leq 2$

The limit defining the Stieltjes integral does not exist, and hence the Stieltjes integral does not exist, as can easily be verified as follows.

The sum of the differences $\alpha(x_k) - \alpha(x_{k-1})$ take two different values, either 1 or 0, depending on the partition, and hence do not converge towards a unique limit.

8. SOME RESULTS

Let $F(x)$ be the probability distribution function of a random variable x , where $-\infty < x < +\infty$.

$F(x)$ is also called the Cumulative Distribution Function (CDF).

$F(x)$ has the following properties:

$$\lim_{x \rightarrow -\infty} F(x) = 0$$

$$\lim_{x \rightarrow +\infty} F(x) = 1$$

$F(x)$ is right continuous, that is, for any $h > 0$ and any $a \in R$, $\lim_{h \rightarrow 0} F(a+h) = F(a)$

$$\begin{aligned}
 P(a, b] &= P(a < x \leq b) \\
 (1) \quad &= P(x \leq b) - P(x \leq a) \\
 &= F(b) - F(a)
 \end{aligned}$$

$$\begin{aligned}
 P(a, b) &= P(a < x < b) \\
 (2) \quad &= P(x < b) - P(x \leq a) \\
 &= F(b^-) - F(a)
 \end{aligned}$$

$$\begin{aligned}
 P[a, b) &= P(a \leq x < b) \\
 (3) \quad &= P(x < b) - P(x < a) \\
 &= F(b^-) - F(a^-)
 \end{aligned}$$

$$\begin{aligned}
 P[a, b] &= P(a \leq x \leq b) \\
 (4) \quad &= P(x \leq b) - P(x < a) \\
 &= F(b) - F(a^-)
 \end{aligned}$$

9. THE CENTRAL LIMIT THEOREM

If $X_1, X_2, X_3 \dots X_n$ are independent, identically distributed random variables each with mean μ , and each with a variance $= \sigma^2$, then

$$(5) \quad S_n = \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^n (X_i - \mu)$$

converges to $N(0, 1)$

10. SOME PRELIMINARIES

In this section, it is necessary to establish certain preliminary results—sort of a mathematical scaffolding—that will be required in the proof of the Central Limit Theorem.

In undergraduate texts and even in graduate-level texts in Mathematical Statistics use the Moment Generating function to prove the CLT. However, the drawback in using the MGF is that not all probability distribution functions have a MGF. Hence, we will use the Characteristic Functions. The advantage in using the Characteristic Functions is that these exist for all probability distribution functions,

and thus offers a lot more utility and generality in obtaining results.

Having mentioned the MGF, let us look at what a MGF is before proceeding to the Characteristic Function:

Definition 10.

The Moment Generating function, if it exists, of a real random variable X is defined as:

$M_X(t) = E[e^{tX}]$ for some real number t , $t \in R$, $-h < t < h$.

The mgf may not exist for all real numbers $t \in R$. It is defined only for values of t for which $E[e^{tx}]$ exists. E is the Expectation operator.

Recall that the expectation of a random variable X is defined as:

For continuous random variables:

$$(6) \quad E(X) = \int_{-\infty}^{+\infty} X f(X) dX$$

where $f(X)$ is the probability density function of the r.v. X .

For discrete random variables:

$$(7) \quad E(X) = \sum_{\forall X_i \in \Omega} X_i f(X_i)$$

Where Ω is the entire sample space for the discrete random variable X .

Hence, for discrete random variables, the mgf is:

$$(8) \quad M_X(t) = \sum_{\forall X_i \in \Omega} e^{tX_i} f(X_i)$$

For continuous random variables, the mgf is:

$$(9) \quad M_X(t) = \int_{-\infty}^{+\infty} e^{tX} f(X) dX$$

The above integral is the Riemann integral. The above Riemann integral presupposes there is a probability density function that is piecewise continuous over the domain of integration. If the probability density function does not exist, the cumulative distribution does exist both for discrete and continuous random variables. Hence, we switch over to the Riemann-Stieltje's integral and directly deal with probabilities as opposed to probability densities, which may exist or may not exist. After all, the expectation of a random variable is with respect to its probability distribution. The cumulative distribution function (CDF) is a probability, regardless of whether the random variable is discrete or continuous.

Also note:

$$f(X) dX = dF(X)$$

where $dF(X)$ is the probability of X in the interval dX .

REFERENCES

Doob1
Sing1

- [1] J. L. Doob, *Stochastic Processes*, John Wiley & Sons, NY 1953.
- [2] N. R. Mann, R. E. Schaefer, N. D. Singpurwalla, *Methods for the Statistical Analysis of Reliability and Life Data*, John Wiley & Sons, 1974.