NATIONAL INSTITUTE OF TECHNOLOGY SILCHAR

CACHAR, ASSAM

DEC, 2020

B.TECH. III[RD] SEMESTER

Subject: MA201

PROJECT ON: PROBABILITY DISTRIBUTION AND ITS APPLICATIONS

SUBMITTED BY:

Name: Subhojit Ghimire

Branch: CSE - B

Scholar Id.: 1912160

UNDER THE GUIDANCE OF:

Mr. Bijan Nath

Associate Professor

Department of Mathematics

NIT Silchar

Subhojit Ghimire
19/12/160
Ph. 9864391349

PREFACE:

The probability distributions are used to explain many of variations observed in real life phenomenon and show how these can be described in simple numerical terms. The distribution shows the probability of an event in an experiment; the event could be testing of human memory, survival of patient of cancer of a protest, time before the next telephone call, the distance between mutations on DNA stands, the number of cars passing through a certain point on road, the number of viruses that can affect a cell in cell culture and so on. The collection of more important facts about common distribution in statistical theory and practice would be useful to variety of scientific works and workers. The distribution can be discrete or continuous depending upon the random variable taken into consideration. For example, the number of objects one is able to remember in a memory test is discrete random variable whereas, the amount of rainfall measured over certain period is a continuous variable.

A discrete probability distribution is applicable to scenarios where the set of possible outcomes is discrete and the probabilities are encoded by a discrete list of probabilities of outcomes, known as probability mass function. On the other hand, continuous probability distribution is applicable to scenarios where the set of possible outcomes can take on values in continuous range. In this case, probabilities are typically described by a probability density function.

Subhojit Ghimire
19/12/160
Ph. 9864391349

PROBABILITY DISTRIBUTION:

A probability distribution can be described in various forms, such as by a probability mass function or a cumulative distribution function. One of the most general descriptions, which applies for continuous and discrete variables, is by means of a probability function $P: A \to \mathbb{R}$, whose input space A is related to the sample space, and gives a probability as its output.

The concept of probability function is made more rigorous by defining it as the element of a probability space (X, A, P),
where,

- X is the set of possible outcomes
- A is the set of all subsets $E \subset X$ whose probability can be measured.
- and, P is the probability function, or probability measure, that assigns a probability to each of these measurable subsets $E \in A$

PROBABILITY MASS FUNCTION:

Probability mass function is the probability distribution of a discrete random variable, and provides the possible values and their associated probabilities. It is the function $p: \mathbb{R} \to [0,1]$ defined by

$$p_X(x_i) = P(X = x_i) \quad ; \text{ for } -\infty < x < \infty$$

Subhojit Ghimire
19/12/160
Ph. 9864391349

where, P is a probability measure
$p_X(x)$ can also be simplified as $p(x)$.

The probabilities associated with each possible values must be positive and sum up to 1. for all other values, the probability needs to be 0.

$$\sum p_X(x_i) = 1$$
$$p(x_i) > 0$$
$$p(x) = 0 \text{ , for all other } x.$$

Thinking of probability as mass helps to avoid mistakes since the physical mass is conserved as is the total probability for all hypothetical outcomes x.

EXAMPLE. 1 :
Roll a die an infinite number of times and look at the proportion of 1, the proportion of 2 and so on. Call x the random variable that corresponds to the outcome of the dice roll. Thus, the random variable x can only take the following discrete values: 1, 2, 3, 4, 5 or 6. It is thus a discrete random variable.

The aim of the probability mass function is to describe the probability of each possible value. In this example, it describes the probability to get a 1, the probability to get a 2 and so on. In the case of rolling the dice, the possibility to get each value is the same, i.e.,

$$P(x=1) = P(x=2) = P(x=3) = P(x=4) = P(x=5) = P(x=6)$$

Subhojit Ghimire
19/12/160
Ph. 9864391349

Since there are 6 possible outcomes, they are equiprobable,

i.e., $P(x=1) = \frac{1}{6}$

$P(x=2) = \frac{1}{6}$

$P(x=3) = \frac{1}{6}$

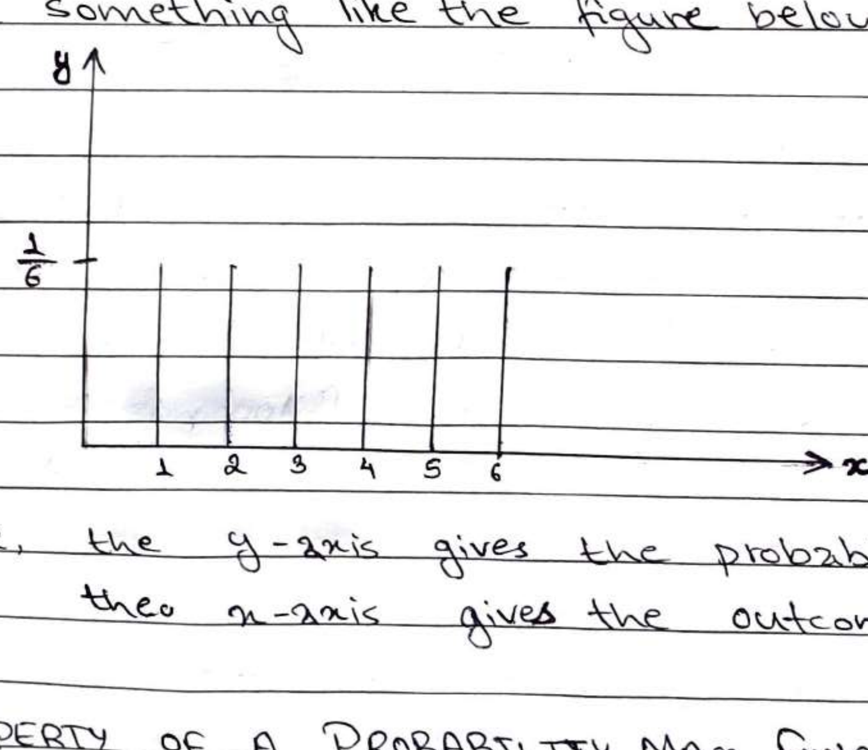
$P(x=4) = \frac{1}{6}$

$P(x=5) = \frac{1}{6}$

and, $P(x=6) = \frac{1}{6}$

This distribution shows the same probability for each value, hence it is called uniform distribution.

The probability mass function for this example would look something like the figure below:



Here, the y-axis gives the probability and the x-axis gives the outcome.

PROPERTY OF A PROBABILITY MASS FUNCTION:

A function is a probability mass function if,

$\forall x \in X \;;\; 0 \leq P(x) \leq 1$

i.e., for every possible value x in the range of x, the probability that the outcome corresponds

Subhojit Ghimire
19/12/160
Ph. 9864391349

to this value is between 0 and 1. A probability of 0 means that the event is impossible and the probability of 1 means that the event is a surety.

The sum of the probabilities associated with each possible value is equal to 1.

$$\sum_{i=1}^{n} P(x_i) = 1$$

PROBABILITY DENSITY FUNCTION:

A random variable X with values in a measurable space (X, A), usually $\mathbb{R}^n$ with the Borel sets as measurable subsets, has as probability distribution the measure X_*P on (X, A). The density of X with respect to a reference measure μ on (X, A) is the Radon-Nikodym derivative:

$$f = \frac{d\,X_*P}{d\mu}$$

where, f is any measurable function with the property that,

$$P_r[X \in A] = \int_{X^{-1}A} dP = \int_A f\,d\mu$$

for any measurable set $A \in \mathcal{A}$.

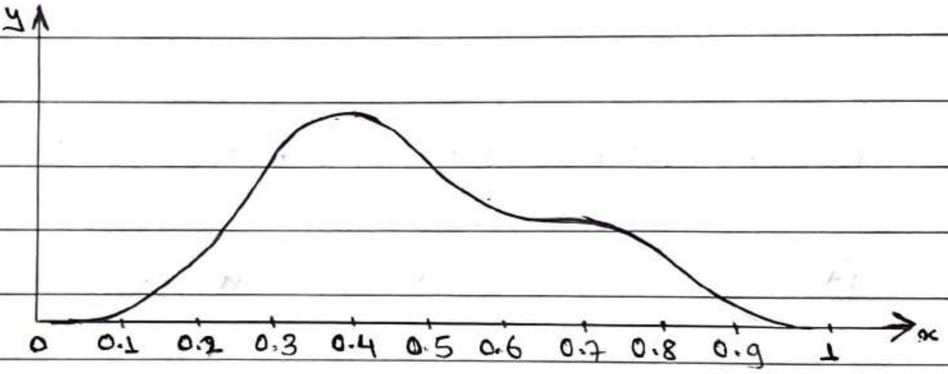
In the continuous univariate case, the reference measure is the Lebesgue measure. The probability mass function of a discrete random variable is the density with respect to the counting measure over the sample space.

Subhojit Ghimire
19/12/160
Ph. 9864391349

It is not possible to define a density with reference to an arbitrary measure. furthermore, when it does exit, the density is almost everywhere unique.

EXAMPLE. 2:

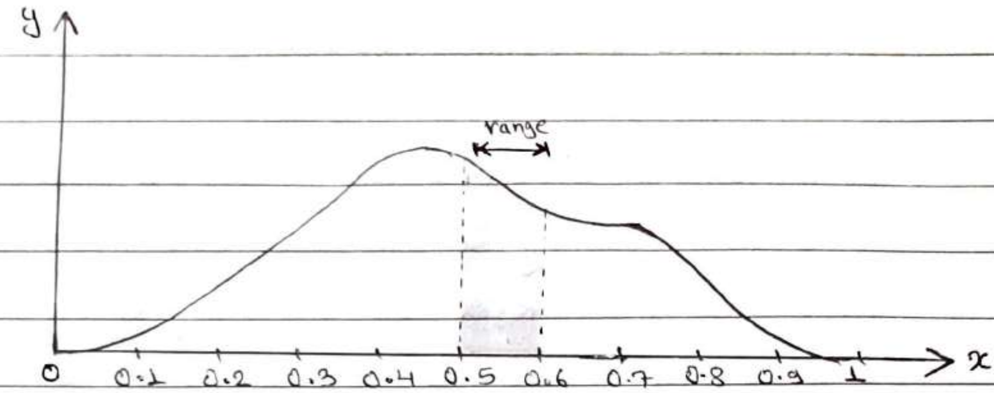
Let a random variable x can take values between 0 and 1, such that its probability density function is as shown in the figure below:



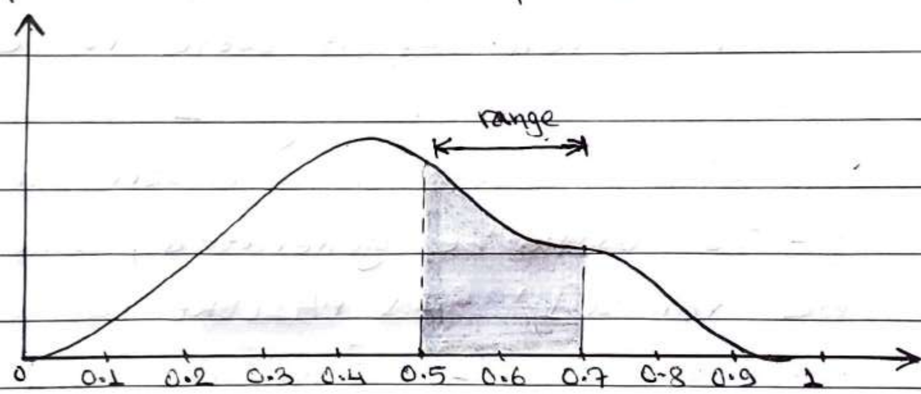
It can be seen that 0 seems not to be possible (i.e., probability around 0) and neither 1. The graph around 0.4 means that it will get a lot of outcomes around this value.

Finding probabilities from probability density function between a certain range of values can be done by calculating the area under the curve for this range. For example, the probability of drawing a value between 0.5 and 0.6 corresponds to the following area:

Subhojit Ghimire
19/12/160
Ph. 9864391349



It can be seen that with an increase in the range, the probability increases as well. for instance, the range of 0.5 to 0.7 is as follows:



The probability associated with a specific range can be obtained by calculating the area under the curve.

PROPERTY OF A PROBABILITY DENSITY FUNCTION:

A function is a probability density function if,

$$\forall x \in x \quad ; \quad p(x) \geq 0$$

In this case, $p(x)$ is not necessarily less than 1 because it doesn't correspond to the probability (the probability itself will still need to be between 0 and 1).

Subhojit Ghimire
19/12/160
Ph. 9864391349

BINOMIAL DISTRIBUTION:

A binomial distribution can be thought of as simply the probability of a SUCCESS or FAILURE outcome in an experiment or survey that is repeated multiple times. The binomial is a type of distribution that has two possible outcomes. For example, a coin toss has only two possible outcomes: heads or tails.

For example, to find the probability of getting a 1 on a die roll. If it were rolled 20 times, the probability of rolling a 1 in any throw is $\frac{1}{6}$. In 20 rolls, a binomial distribution of $n=20$ and $p=1/6$ would be generated, where SUCCESS would be 'roll a 1' and FAILURE would be 'roll any number except 1'. Similarly, the probability of the die landing on an even number, when the die is rolled 20 times would be given by the binomial distribution having $n=20$ and $p=1/2$. because the probability of rolling an even number at each attempt is one half.

The binomial distribution formula is:

$$b(x; n, P) = {}^{n}C_{x} \cdot P^{x} \cdot (1-P)^{n-x}$$

where, b is binomial probability
n is total number of events or repetitions (trials)
P is probability of success on an individual trial
x is total number of successes
$(1-P) = Q$ or q is probability of failure on a trial.
${}^{n}C_{x}$ is Combination formula $= \frac{n!}{(n-x)!\,x!}$

Subhojit Ghimire
19/12/160
Ph. 9864391349

EXAMPLE. 3:

When a coin is tossed 10 times, what is the probability of getting exactly 6 heads?

Here,

Using the binomial distribution formula,

$$b(x; n, P) = {}^nC_x P^x (1-P)^{n-x}$$

The number of trials, n is 10

The odds of success (tossing heads) is, $p = 0.5$

So, $-P+1 = 1-0.5 = 0.5$

The total number of successes, $x = 6$

So,

$$P(x=6) = {}^{10}C_6 \cdot (0.5)^6 \cdot (0.5)^4$$
$$= 210 \times 0.015625 \times 0.0625$$
$$= 0.205078125$$

PROPERTIES OF A BINOMIAL DISTRIBUTION:

1. The number of observations or trials is fixed. That means, one can only figure out the probability of something happening if it is repeated a certain number of times.

2. Each observation or trial is independent. In other words, none of the previously done trials will have an effect on the probability on the next trial.

3. The probability of success is exactly the same from one trial to another.

Subhojit Ghimire
19/12/160
Ph. 9864391349.

REAL LIFE EXAMPLES OF BINOMIAL DISTRIBUTION:

Basically, anything that can be thought to be either a success or a failure can be represented by a binomial distribution.

If a new drug is introduced to cure a disease, it is either a success, meaning it cures the disease, or it doesn't cure the meaning, meaning it's a failure.

If a lottery ticket is purchased, the buyer either wins the prize or doesn't win the prize.

POISSON DISTRIBUTION:

A Poisson distribution is a tool that helps to predict the probability of certain events from happening when it is known how often the event has occurred. It gives the probability of a given number of events happening in a fixed interval of time.

The formula for Poisson probability mass function is

$$p(x;\lambda) = \frac{e^{-\lambda}\lambda^{x}}{x!}$$

where, $x = 0, 1, 2, \ldots$

λ is the expected number of occurrences. It is sometimes denoted by μ and is called the event rate or rate parameter.

Subhojit Ghimire
19/12/160
Ph. 9864391349

EXAMPLE. 4:

If the average number of major storms in a city is 2 per year, what is the probability that exactly 3 storms will hit that city the next year?

Here,

Average number of storms per year, $\mu = 2$

Number of storms to hit the next year, $x = 3$

Euler's number, $e = 2.71828$

So,

Placing these component values in the Poisson distribution formula,

$$P(3; 2) = \frac{(2.71828)^{-2} \cdot 2^3}{3!}$$

$$= \frac{0.13534 \times 8}{6}$$

$$= 0.180$$

i.e, the probability of 3 storms happening next year is 0.180, or 18%.

As it can probably be seen, Poisson distribution can be calculated manually, but it would taker an extraordinary amount of time unless a simple set of data is provided. The usual way to calculate a Poisson distribution in real life situations is with software like IBM SPSS.

Subhojit Ghimire
19/12/160
Ph. 9864391349

PRACTICAL USE OF POISSON DISTRIBUTION:

A textbook store rents an average of 200 books every Saturday night. Using this data, one can predict the probability that more books (perhaps 300 or 400) will sell on the following Saturday nights.

Another example is the number of diners in a certain restaurant every day. If the average number of diners for seven days is 500, it can be predicted that the probability of a certain day having more customers.

Because of this application, Poisson distributions are used by businessmen to make forecasts about the number of customers or sales on certain days or seasons of the year. In business, overstocking will sometimes mean losses if the goods are not sold. Likewise, having too few stock would still mean a lost business. By using poisson distribution, businessmen are able to estimate the time when demand is unusually higher, so they can purchase more stock. Hotels and restaurants could prepare for an influx of customers, hire extra temporary workers in advance, purchase more supplies etc.

With the Poisson distribution, companies can adjust supply to demand in order to keep their business earning good profit. In addition, waste of resource is prevented.

Subhojit Ghimire
19/12/160
Ph. 9864391349

POISSON DISTRIBUTION VS BINOMIAL DISTRIBUTION:

It can be challenging to figure out whether to use a Binomial distribution or a Poisson distribution in exam-related questions. The following two points are general guidelines to distinguish between the two:

1. If the question has an average probability of an event happening per unit (i.e., per unit time, cycle, event) and it has been asked to find probability of a certain number of events happening in a period of time (or number of events), then the Poisson Distribution is used.

2. If the question has provided an exact probability and it has been asked to find the probability of the event happening a certain number of times out of n (i.e., 10 times out of 100, or 99 times out of 9000), the Binomial Distribution formula is used.

NORMAL DISTRIBUTION:

Normal distribution, also known as Gaussian distribution, is a probability distribution that is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean. In graph form, normal distribution will appear as a bell curve.

Subhojit Ghimire
19/12/160
Ph. 9864391349

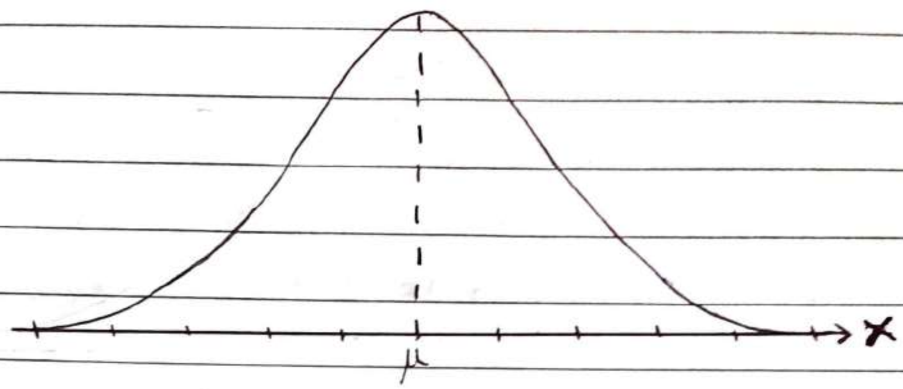


fig.: A normal (bell) curve

This random variable X is said to be normally distributed with mean μ and standard deviation σ, if its probability distribution is given by,

$$f(X) = \frac{1}{\sigma\sqrt{2\pi}} \times e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

A normal distribution is the proper term for a probability bell curve. In a normal distribution the mean is zero and the standard deviation is 1. It has zero skew and a kurtosis is 3. Normal distributions are symmetrical but not all symmetrical distributions are normal. In reality, most pricing distributions are not perfectly normal.

SKEWNESS AND KURTOSIS:

The skewness and kurtosis coefficients measure how different a given distribution is from a normal distribution. The skewness measures the symmetry

Subhojit Ghimire
19/12/160
Ph. 9864391349

of a distribution. The normal distribution is symmetric and has a skewness of zero. If the distribution of a data set has a skewness less than zero, or negative skewness, then the left tail of the distribution is longer than the right tail; positive skewness implies that the right tail of the distribution is longer than the left.

The kurtosis statistic measures the thickness of the tail ends of a distribution in relation to the tails of the normal distribution. Distributions with large kurtosis exhibit tail data exceeding the tails of the normal distribution. Distributions with low kurtosis exhibit tail data that is generally less extreme than the tails of the normal distribution.

EXAMPLE. 5:

If the bottom 30% of students failed an end of semester exam, where the mean for the test was 120 and the standard deviation was 17, what was the passing score?

Here,

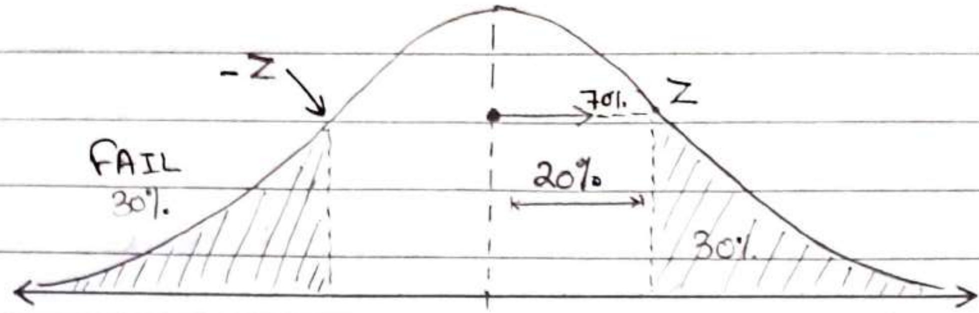
$$z = \frac{x - \mu}{\sigma}$$

$$\text{or, } x = z\sigma + \mu$$

$$= 17z + 120$$

To find the value of z, we will be looking at the normal distribution curve.

Subhojit Ghimire
19/12/160
Ph. 9864391349



Understanding the curve,
the curve is symmetrical, so the bottom 30% and the z-point are same on both the sides of the curve.

The area under the curve is 100% and hence the half area under the curve is 50%.
Looking up 20% in z-table, we can see the value is $z = 0.52$; $-z = -0.52$
So,

$$x = 17 \times (-0.52) + 120$$
$$= 120 - 8.84$$
$$= 111.16$$

So the passing score was 111.16, which is slightly below the mean 120.

PROPERTIES OF A NORMAL DISTRIBUTION:

1. The normal curve is symmetrical about the μ
2. The mean is at middle and divides area into halves
3. The total area under the curve is equal to 1.
4. It is completely determined by its mean and standard deviation σ.

Subhojit Ghimire
19/12/160
Ph. 9864391349

REAL LIFE APPLICATION OF NORMAL DISTRIBUTION:

1. Height: Height of the population is the example of normal distribution. Most of the people on a specific population are of average height. The number of people taller and shorter than the average height people is almost equal, and a very small number of people are either extremely tall or extremely short. Therefore, it follows the normal distribution.

2. Technical Stock Market: The falling and hiking in the price of the shares, the changes in the log values of Forex rates, price indices and stock prices return often from a bell shaped curve.

3. Income Distribution in Economy: The income of a country lies in the hands of enduring politics and government. It depends upon how they distribute the income among the rich and poor community. The middle-class population is higher than the rich and poor population, so, the wages of the middle-class population makes the mean in the normal distribution curve.

4. Birth Weight: The normal birth weight of a newborn ranges from 25 to 3.5 kg. The majority of newborns have normal birthweight whereas, only a few percentage have a weight higher or lower than the normal. Hence, birth weight follows the normal distribution curve.

Subhojit Ghimire
19/12/160
Ph. 9864391349

BIBLIOGRAPHY:

1. Some Probability Distributions and their Applications || S.B. Kalyankar, H.S. Acharya || 2009
2. Statistical Learning Theory || Vlasimir N. Vapnik || 1998
3. Deep Learning Book || I. Goodfellow, Y. Bengio, A. Courville || 2016
4. Probability Mass and Density Function || KDnuggets.com || Hadrien Jean || 2019
5. CRC Standard Mathematical Tables and Formulae || W.H. Beyer || 1987
6. Theory and Problems of Probability and Statistics || M.R. Spiegel || 1992
7. Schaum's Easy Outlines of Statistics || D. Lindstorm || 2010
8. Binomial Distribution || StatisticsHowTo.com || Stephanie Glen || 2013
9. Poisson Distribution/ Poisson Curve || StatisticsHowTo.com || Stephanie Glen || 2013
10. Normal Distribution || Investopedia.com || James Chen || 2020
11. 9 Real Life Examples of Normal Distribution || StudiousGuy.com