**JSS Mahavidyapeetha**

**JSS Science and Technology University,**

**JSS Technical Institutions Campus,**

**Mysore – 570006**

****

DEPARTMENT OF INFORMATION SCIENCE AND ENGINEERING

Event report as a part of CIE event for the course

**Engineering Mathematics - IV (MA410)**

Report On,

**Probability Theory**

**(To find the Mean, Variance and Standard Deviation**

**of the given Data Set)**

**Name of the Candidates :**

|  |  |
| --- | --- |
| Name | USN |
| Dhanush Raj N | 01JST17IS016 |
| Kirthan M | 01JST17IS024 |
| Rajesh Chand N C | 01JST17IS038 |

**Under the Guidance of :**

Dr. T D Roopamala,

HOD of Mathematics Department

**PROBLEM STATEMENT**

* Probability distribution, Mean, Variance and Standard Deviation of the

Given Data Set

**Data Set :** Uber trips in New York City

**CONTENTS**

* **Abstract…………………………………………………………………………….…………...04**
* **Introduction…………………………………………...........................................05**
* **Uber Trips Data………………………………………………………………………………..07**
* **Probability Distribution Table………………………….................................08**
* **Mean, Variance and Standard Deviation……………..............................09**
* **Program………………………………………………………………………………..………..10**
* **Output…………………………………………………………………………………………….13**
* **Applications…………………………………………………………………………………….14**
* **Conclusion…………………………………………………......................................15**
* **References………………………………………………........................................16**

**ABSTRACT**

**Probability** is the measure of the likelihood that an event will occur. Probability quantifies as a number between 0 and 1, where, loosely speaking, 0 indicates impossibility and 1 indicates certainty. The higher the probability of an event, the more likely it is that the event will occur.

These concepts have been given an axiomatic mathematical formalization in probability theory, which is used widely in such areas of study as mathematics, statistics, finance, gambling, science (in particular physics), artificial intelligence/machine learning, computer science, game theory, and philosophy to, for example, draw inferences about the expected frequency of events. Probability theory is also used to describe the underlying mechanics and regularities of complex systems.

**INTRODUCTION**

In probability theory and statistics, a **probability distribution** is a mathematical function that provides the probabilities of occurrence of different possible outcomes in an experiment. In more technical terms, the probability distribution is a description of a random phenomenon in terms of the probabilities of events. For instance, if the random variable X is used to denote the outcome of a coin toss ("the experiment"), then the probability distribution of X would take the value 0.5 for X = heads, and 0.5 for X = tails (assuming the coin is fair). Examples of random phenomena can include the results of an experiment or survey.

A probability distribution is specified in terms of an underlying sample space, which is the set of all possible outcomes of the random phenomenon being observed. The sample space may be the set of real numbers or a set of vectors, or it may be a list of non-numerical values; for example, the sample space of a coin flip would be {heads, tails}

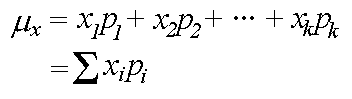
Probability distributions are generally divided into two classes. A **discrete probability distribution** (applicable to the scenarios where the set of possible outcomes is discrete, such as a coin toss or a roll of dice) can be encoded by a discrete list of the probabilities of the outcomes, known as a probability mass function. On the other hand, a **continuous probability distribution** (applicable to the scenarios where the set of possible outcomes can take on values in a continuous range (e.g. real numbers), such as the temperature on a given day) is typically described by probability density functions (with the probability of any individual outcome actually being 0). The normal distribution is a commonly encountered continuous probability distribution. More complex experiments, such as those involving stochastic processes defined in continuous time, may demand the use of more general probability measures.

In probability, a **random variable** is defined as a function that maps the outcomes of an unpredictable process to numerical quantities, typically real numbers. It is a variable (specifically a dependent variable), in the sense that it depends on the outcome of an underlying process providing the input to this function, and it is random in the sense that the underlying process is assumed to be random.

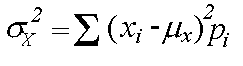
A random variable's possible values might represent the possible outcomes of a yet-to-be-performed experiment, or the possible outcomes of a past experiment whose already-existing value is uncertain (for example, due to imprecise measurements or quantum uncertainty). They may also conceptually represent either the results of an "objectively" random process (such as rolling a die) or the "subjective" randomness that results from incomplete knowledge of a quantity. The meaning of the

probabilities assigned to the potential values of a random variable is not part of probability theory itself but is instead related to philosophical arguments over the interpretation of probability. The mathematics works the same regardless of the particular interpretation in use.

The **mean** of a discrete random variable X is a weighted average of the possible values that the random variable can take. Unlike the sample mean of a group of observations, which gives each observation equal weight, the mean of a random variable weights each outcome xi according to its probability, pi. The common symbol for the mean (also known as the expected value of X) is, formally defined by,



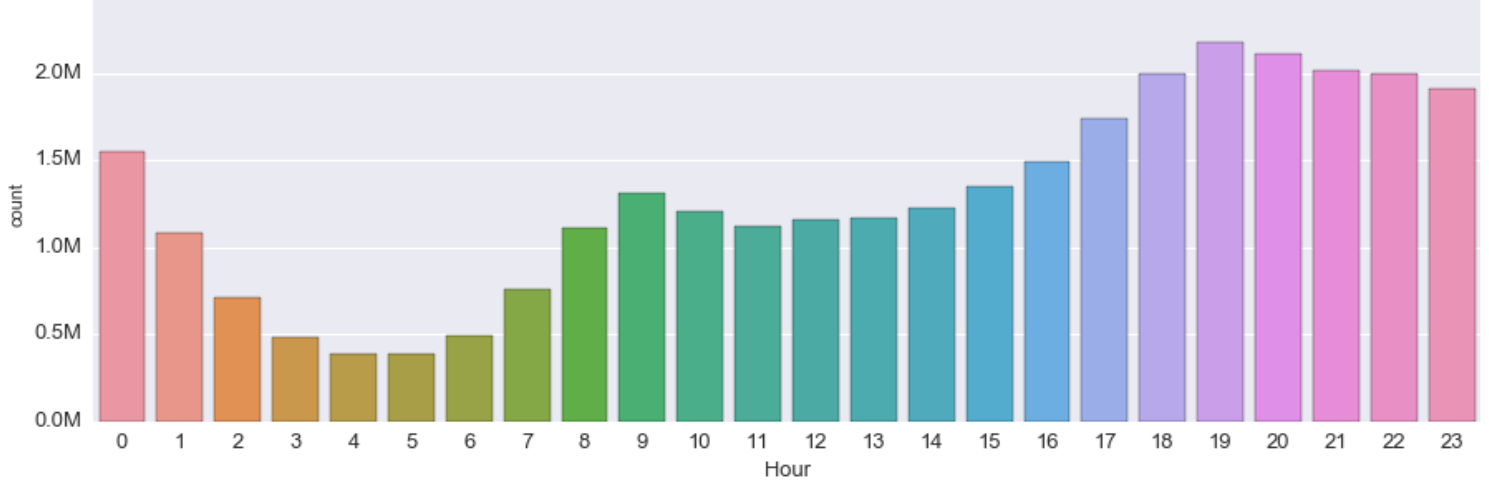
The **variance** of a discrete random variable *X* measures the spread, or variability, of the distribution, and is defined by



The **standard deviation** is the square root of the variance which is given by,

σ = σx2

**UBER TRIPS DATA**

***Graph* :** **1**

|  |  |
| --- | --- |
| nth Hour | Number of Trips  (in millions) |
| 0 | 1.60 |
| 1 | 1.10 |
| 2 | 0.80 |
| 3 | 0.50 |
| 4 | 0.40 |
| 5 | 0.40 |
| 6 | 0.50 |
| 7 | 0.80 |
| 8 | 1.20 |
| 9 | 1.35 |
| 10 | 1.25 |
| 11 | 1.20 |
| 12 | 1.30 |
| 13 | 1.30 |
| 14 | 1.35 |
| 15 | 1.40 |
| 16 | 1.50 |
| 17 | 1.75 |
| 18 | 2.00 |
| 19 | 2.20 |
| 20 | 2.10 |
| 21 | 2.00 |
| 22 | 2.00 |
| 23 | 1.90 |

***Table* :** **1**

**PROBABILTY DISTRIBUTION TABLE**

**Let ‘X’ be a random variable which denotes the hour at which uber trips are available.**

|  |  |
| --- | --- |
| X | P(X) |
| 0 | 0.0502 |
| 1 | 0.0345 |
| 2 | 0.0251 |
| 3 | 0.0157 |
| 4 | 0.0125 |
| 5 | 0.0125 |
| 6 | 0.0157 |
| 7 | 0.0251 |
| 8 | 0.0376 |
| 9 | 0.0423 |
| 10 | 0.0392 |
| 11 | 0.0376 |
| 12 | 0.0408 |
| 13 | 0.0408 |
| 14 | 0.0423 |
| 15 | 0.0439 |
| 16 | 0.0470 |
| 17 | 0.0549 |
| 18 | 0.0627 |
| 19 | 0.0690 |
| 20 | 0.0658 |
| 21 | 0.0627 |
| 22 | 0.0627 |
| 23 | 0.0596 |

**To find P(X):**

**Sample space**,

**S** = 1.6+1.1+0.8+0.5+0.4+0.4+0.5+0.8+1.2+1.35+1.25+1.2

+1.3+1.3+1.35+1.4+1.5+1.75+2+2.2+2.1+2+2+1.9

**S = 31.9**

**P(X=0)** = 1.6 / 31.9 = 0.0502 **P(X=12)** = 1.3 / 31.9 = 0.0408

**P(X=1)** = 1.1 / 31.9 = 0.0345 **P(X=13)** = 1.3 / 31.9 = 0.0408

**P(X=2)** = 0.8 / 31.9 = 0.0251 **P(X=14)** = 1.35 / 31.9 = 0.0423

**P(X=3)** = 0.5 / 31.9 = 0.0157 **P(X=15)** = 1.4 / 31.9 = 0.0439

**P(X=4)** = 0.4 / 31.9 = 0.0125 **P(X=16)** = 1.5 / 31.9 = 0.0470

**P(X=5)** = 0.4 / 31.9 = 0.0125 **P(X=17)** = 1.75 / 31.9 = 0.0549

**P(X=6)** = 0.5 / 31.9 = 0.0157 **P(X=18)** = 2.0 / 31.9 = 0.0627

**P(X=7)** = 0.8 / 31.9 = 0.0251 **P(X=19)** = 2.2 / 31.9 = 0.0690

**P(X=8)** = 1.2 / 31.9 = 0.0376 **P(X=20)** = 2.1 / 31.9 = 0.0658

**P(X=9)** = 1.35 / 31.9 = 0.0423 **P(X=21)** = 2.0 / 31.9 = 0.0627

**P(X=10)** = 1.25 / 31.9 = 0.0392 **P(X=22)** = 2.0 /31.9 = 0.0627

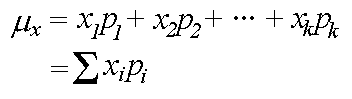
**P(X=11)** = 1.2 / 31.9 = 0.0376 **P(X=23)** = 1.9 / 31.9 = 0.0596

***Table* :** **2**

**MEAN, VARIANCE AND**

**STANDARD DEVIATION**

**Mean,**



= (0 \* 0.0502) + (1 \* 0.0345) + (2 \* 0.0251) + (3 \* 0.0157) + (4 \* 0.0125) +

(5 \* 0.0125) + (6 \* 0.0157) + (7 \* 0.0251) + (8 \* 0.0376) + (9 \* 0.0423) +

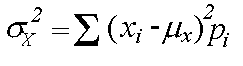
(10 \* 0.0392) + (11 \* 0.0376) + (12 \* 0.0408) + (13 \* 0.0408) + (14 \* 0.0423) +

(15 \* 0.0439) + (16 \* 0.0470) + (17 \* 0.0549) + (18 \* 0.0627) + (19 \* 0.0690) +

(20 \* 0.0658) + (21 \* 0.0627) + (22 \* 0.0627) + (23 \* 0.0596)

**µx = 13.7774**

**Variance,**



= ((0 - 13.7774)2 \* 0.0502) + ((1 - 13.7774)2 \* 0.0345) + ((2 - 13.7774)2 \* 0.0251) +

((3 - 13.7774)2 \* 0.0157) + ((4 - 13.7774)2 \* 0.0125) + ((5 - 13.7774)2 \* 0.0125) +

((6 - 13.7774)2 \* 0.0157) + ((7 - 13.7774)2 \* 0.0251) + ((8 - 13.7774)2 \* 0.0376) +

((9 - 13.7774)2 \* 0.0423) + ((10 - 13.7774)2 \* 0.0392) + ((11 - 13.7774)2 \* 0.0376) +

((12 - 13.7774)2 \* 0.0408) + ((13 - 13.7774)2 \* 0.0408) + ((14 - 13.7774)2 \* 0.0423) +

((15 - 13.7774)2 \* 0.0439) + ((16 - 13.7774)2 \* 0.0470) + ((17 - 13.7774)2 \* 0.0549) +

((18 - 13.7774)2 \* 0.0627) + ((19 - 13.7774)2 \* 0.0690) + ((20 - 13.7774)2 \* 0.0658) +

((21 - 13.7774)2 \* 0.0627) + ((22 - 13.7774)2 \* 0.0627) + ((23 - 13.7774)2 \* 0.0596)

**σx2 = 46.9317**

**Standard Deviation,**

**σ = σx2**

**σ** =

**σ = 6.8507**

**PROGRAM**

*//C program for Uber data analysis*

**#include <stdio.h>**

**#include <stdlib.h>**

**#include <math.h>**

**float** mean (**int** X [ ], **float** P\_X [ ], **float** S)

{

**int** i;

**float** x=0.0;

**for**(i=0; i<24; i++)

{

x = x+ (X[i] \* P\_X[i]);

}

**return** x;

}

**float** variance (**int** X [ ], **float** P\_X [ ], **float** S, **float** mean)

{

**int** i;

**float** var=0.0;

for (i=0; i<24; i++)

{

var = var + (pow ((X[i] - mean), **2**) \* P\_X[i]);

}

return var;

}

**int** main ()

{

**int** i;

**int** X [50];

**float** x, var, sd, S = 0.0, total = 0.0;

**float** P\_X [50];

**float** data [ ] = {1.6, 1.1, 0.8, 0.5, 0.4, 0.4, 0.5, 0.8, 1.2, 1.35, 1.25, 1.2,

1.3, 1.3, 1.35, 1.4, 1.5, 1.75, 2, 2.2, 2.1, 2, 2, 1.9};

**for** (i=0; i<24; i++)

{

X[i] = i;

}

**for** (i=0; i<24; i++)

{

S = S + data[i];

}

**for** (i=0; i<24; i++)

{

P\_X[i] = data[i] / S;

}

**for** (i=0; i<24; i++)

{

total = total + P\_X[i];

}

printf ("\nProbabilityDistributionTable:\n\n");

printf (**"** ===================\n**"**);

printf ("| X \t| P(X) |\n");

printf (**"** ===================\n**"**);

**for** (i=0; i<24; i++)

{

printf ("| %d \t| %.4f |\n",X[i],P\_X[i]);

}

printf (**"** -------------------\n**"**);

printf ("\nTotal:%f\n",total);

printf ("\nSample Space : %f\n",S);

x = mean (X, P\_X, S);

printf ("\nMean of the given probability distribution is:\n");

printf ("µ= %.4f\n",x);

var = variance (X, P\_X, S, x);

printf ("\nVariance of the given probability distribution is:\n");

printf ("σ^2= %.4f\n",var);

sd = sqrt(var);

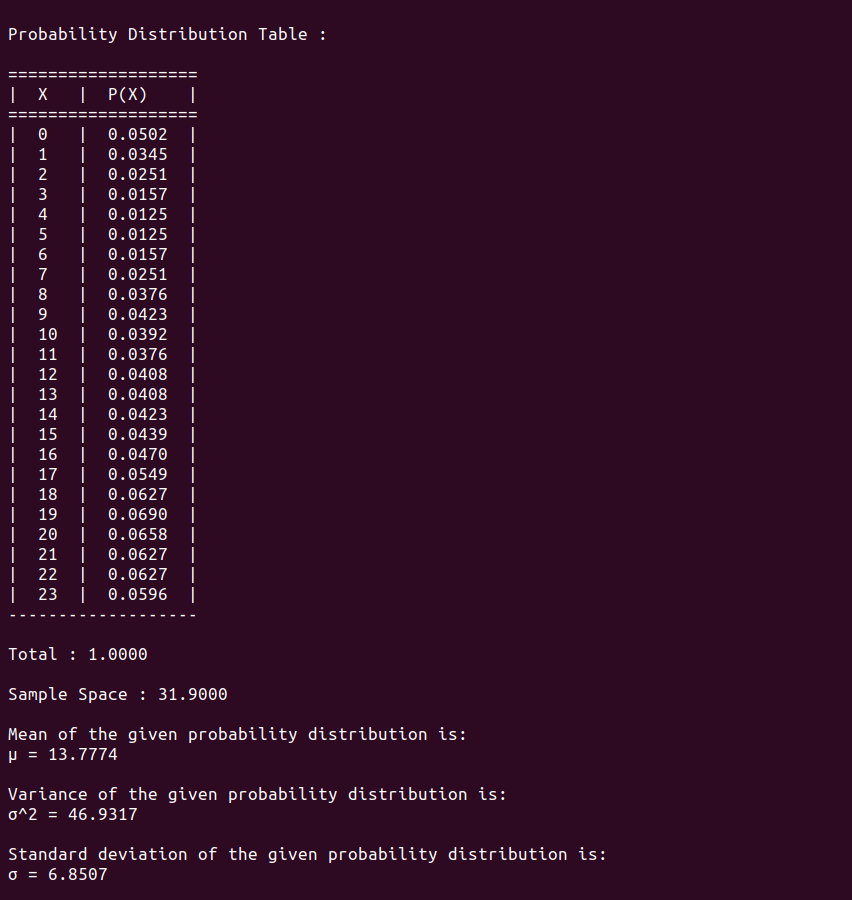
printf ("\nStandard deviation of the given probability distribution is:\n**"**);

printf ("σ= %.4f\n",sd);

**return** 0;

}

**OUTPUT**

****

**APPLICATIONS**

Probability theory is applied in everyday life in risk assessment and modelling. The insurance industry and markets use actuarial science to determine pricing and make trading decisions. Governments apply probabilistic methods in environmental regulation, entitlement analysis (Reliability theory of aging and longevity), and financial regulation.

A good example of the use of probability theory in equity trading is the effect of the perceived probability of any widespread Middle East conflict on oil prices, which have ripple effects in the economy as a whole. An assessment by a commodity trader that a war is more likely can send that commodity's prices up or down, and signals other traders of that opinion. Accordingly, the probabilities are neither assessed independently nor necessarily very rationally. The theory of behavioural finance emerged to describe the effect of such groupthink on pricing, on policy, and on peace and conflict.

In addition to financial assessment, probability can be used to analyse trends in biology (e.g. disease spread) as well as ecology (e.g. biological Punnett squares). As with finance, risk assessment can be used as a statistical tool to calculate the likelihood of undesirable events occurring and can assist with implementing protocols to avoid encountering such circumstances. Probability is used to design games of chance so that casinos can make a guaranteed profit, yet provide pay-outs to players that are frequent enough to encourage continued play.

The discovery of rigorous methods to assess and combine probability assessments has changed society.

Another significant application of probability theory in everyday life is reliability. Many consumer products, such as automobiles and consumer electronics, use reliability theory in product design to reduce the probability of failure. Failure probability may influence a manufacturer's decisions on a product's warranty.

The cache language model and other statistical language models that are used in natural language processing are also examples of applications of probability theory.

**CONCLUSION**

Statistics don't just look at the data and calculate the average, what statistics try to do is to find the original probability distribution from which the collected data originated. Pdf's are central to most statistics.  
  
 Since we can't have the full dataset (we just have a sample) probability distributions, help us calculate the probability of error in our estimators. As such, we can say that the mean (average) of the original distribution (not our sample) is between such and such values with a 95% probability. We can calculate the lifetime of a light bulb, or how many yogurts are still fresh two days after the expiration date (fun fact, most manufacturers calculate expiration date as the day before about 99.99% of all product is still good to eat, as such they only have to deal with about 1 in 10,000 products gone bad, they vary the percentage for products where freshness is vital).

**REFERENCES**

***For the concepts of Probability Theory :***

**Textbook :** Engineering Mathematics – IV , Dr . K.S.C

***For the Data Set :***

**Website :** https://github.com