DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
SUBJECTCODE: 21MT2103RA
PROBABILITY STATISTICS AND QUEUING THEORY

Tutorial 6:
Demonstrate Random Variables, Probability Distribution Function
Date of the Session: //______________________________Time of the Session:_____ to ______

Learning outcomes:

- Develop the Probability Distribution Function
- Understanding Higher-Order Moments, Variance, Standard Deviation

PRE-TUTORIAL

1. Describe probability distribution and its types?

Solution:

In Statistics, the probability distribution gives the possibility of each outcome of a random experiment or event. It provides the probabilities of different possible occurrences. Also read, events in probability, here.

To recall, the **probability is a measure of uncertainty of various phenomena**. Like, if you throw a dice, the possible outcomes of it, is defined by the probability. This distribution could be defined with any random experiments, whose outcome is not sure or could not be predicted. Let us discuss now its definition, function, formula and its types here, along with how to create a table of probability based on random variables.

Probability distribution yields the possible outcomes for any random event. It is also defined based on the underlying sample space as a set of possible outcomes of any random experiment. These settings could be a set of real numbers or a set of vectors or a set of any entities. It is a part of probability and statistics.

Random experiments are defined as the result of an experiment, whose outcome cannot be predicted. Suppose, if we toss a coin, we cannot predict, what outcome it will appear either it will come as Head or as Tail. The possible result of a random experiment is called an outcome. And the set of outcomes is called a sample point. With the help of these experiments or events, we can always create a probability pattern table in terms of variables and probabilities.

Sampling Distributions and Confidence Intervals

Inferential testing uses the sample mean ($\bar{x}$) to estimate the population mean (μ). Typically, we use the data from a single sample, but there are many possible samples of the same size that could be drawn from that population. As we saw in the previous chapter, the sample mean ($\bar{x}$) is a random variable with its own distribution.

- The distribution of the sample mean will have a mean equal to μ.
- It will have a standard deviation (standard error) equal to $\sigma/\sqrt{n}$

Sampling Distribution of the Sample Proportion

The population proportion (p) is a parameter that is as commonly estimated as the mean. It is just as important to understand the distribution of the sample proportion, as the mean. With proportions, the element either has the characteristic you are interested in or the element does not have the characteristic. The sample proportion ($\hat{p}$) is calculated by

$$\hat{p} = \frac{x}{n}$$

where x is the number of elements in your population with the characteristic and n is the sample size

2. The adjacent table (joint Probability Mass Function of two random variables X and Y) shows the probabilities for events X and Y happening at the same time. What is the probability of the two random variables (X and Y) taking the same value simultaneously? In other words, write P(X = k, Y = k)

Solution:

From the table, we see that

P(X = 1, Y = 1) = 0

P(X = 2, Y = 2) = 0

P(X = 3, Y = 3) = 0

So,

P(X = k, Y = k) = 0 for all values that X and Y can take.

<u>IN-TUTORIAL:</u>

1. Write a SAS code that generates 10 random samples of a poison distribution with lambda=4, and plot a bar graph of the probability mass function using sgplot procedure.

Solution:

```
%let lambda=4;

data Poisson_PMF;
  do k=0 to 10;
    PMF=pdf('Poisson', k, &lambda);
    output;
  end;
run;

title "Poisson Probability Mass Function.";
title2 "For (*ESC*){unicode lambda} = &lambda.";
proc sgplot data=Poisson_PMF noautolegend;
  vbar k / response=PMF barwidth=0.5 legendlabel="PMF";
  keylegend / location=inside position=NE across=1;
  yaxis display=(nolabel);
run;
title;
```

2. Assume that the pair of dice is thrown, and the random variable X is the sum of numbers that appears on two dice. Find the mean or the expectation of the random variable X using SAS

Solution:

The PDF function for the *T* distribution returns the probability density function of a *T* distribution, with degrees of freedom *df* and the noncentrality parameter *nc*. The PDF function is evaluated at the value *x*. This PDF function accepts noninteger degrees of freedom. If *nc* is omitted or equal to zero, the value returned is from the central *T* distribution. In this equation,

If two dice are thrown, then the total number of sample spaces obtained is 36.

Given that, the random variable X is the sum of numbers that appear on two dice, such as 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 or 12.

Therefore,

P(X= 2) = 1/36

P(X= 3) = 2/36

P(X= 4) = 3/36

P(X= 5) = 4/36

P(X= 6) = 5/36

P(X= 7) = 6/36

P(X= 8) = 5/36

P(X= 9) = 4/36

P(X= 10) = 3/36

P(X= 11) = 2/36

P(X= 12) = 1/36

Hence, the probability distribution of the random variable X is:

X	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
P(X)	**1/36**	**2/36**	**3/36**	**4/36**	**5/36**	**6/36**	**5/36**	**4/36**	**3/36**	**2/36**	**1/36**

Therefore,

The mean or the expectation of the random variable X is:

= 2(1/36) + 3(2/36) + 4(3/36) + 5(4/36) + 6(5/36) + 7(6/36) + 8(5/36) + 9(4/36) + 10(3/36) + 11(2/36) + 12(1/36)

= (2+6+12+20+30+42+40+36+30+22+12)/36

=7

Therefore, the mean of the random variable X is 7.

The following program illustrates the PDF T distribution function

```
proc ds2;
data _null_;
   dcl double y;
   method init();
      y=pdf('T',.9,5);
      put 'T dist:      ' y;
   end;
enddata;
run;
quit;
```

Out Put:

SAS writes the following output to the log:

T dist: 0.24194434361358

POST-TUTORIAL

1. How to use a SAS data step to obtain an **approximate-sized random sample without replacement**. Specifically, the program uses the **ranuni** function and a WHERE statement to tell SAS to randomly sample approximately 30% of the 50 observations from the permanent SAS data set mailing.

Solution:

```
DATA sample1A (where = (random le 0.30));
```

```
    LIBNAME phc6089 '/home/u61435383/aroragaurav1260/data';
    set phc6089.mailing;
    random = ranuni(43420);
RUN;

PROC PRINT data=sample1A NOOBS;
    title1 'Sample1A: Approximate-Sized Simple Random Sample';
    title2 'without Replacement';
RUN;
```

Note:
Upload data set "mailing.sas7bdat" in File data folder.

Sample output:

Sample1A: Approximate-Sized Simple Random Sample
without Replacement

Num	Name	Street	City	State	random
1	Jonathon Smothers	103 Oak Lane	Bellefonte	PA	0.07478
2	Jane Doe	845 Main Street	Bellefonte	PA	0.25203
4	Mark Adams	312 Oak Lane	Bellefonte	PA	0.08918
6	Delilah Fequa	2094 Acorn Street	Bellefonte	PA	0.02253
7	John Doe	812 Main Street	Bellefonte	PA	0.15570
8	Mamie Davison	102 Cherry Avenue	Bellefonte	PA	0.05460
9	Ernest Smith	492 Main Street	Bellefonte	PA	0.05662
14	William Edwards	79 Oak Lane	Bellefonte	PA	0.15432