 Sardar Patel Institute of Technology,Mumbai

Department of Electronics and Telecommunication Engineering

T.E. Sem-V (2018-2019)

ETL54-Statistical Computational Laboratory

**Lab-2: Probability Distributions**

**Name:Yash Hemant Patil Roll No.2016120036**

**Objective:To compute probability density function (pdf) and cumulative distribution function (cdf)**

**Outcomes:**

To list and describe the well-known probability distributions with their characteristics.

To compute the probability distributions which are frequently occurs in Statistical Study

**System Requirements:** Ubuntu OS with R and RStudio installed

**Introduction to Probability distribution:**

A probability distribution describes how the values of a random variable is distributed. There are two types of probability distributions: Discrete and Continuous

Well-known probability distributions which are frequently occurred in statistical study:

[Binomial Distribution](http://www.r-tutor.com/elementary-statistics/probability-distributions/binomial-distribution)

[Poisson Distribution](http://www.r-tutor.com/elementary-statistics/probability-distributions/poisson-distribution)

[Continuous Uniform Distribution](http://www.r-tutor.com/elementary-statistics/probability-distributions/continuous-uniform-distribution)

[Exponential Distribution](http://www.r-tutor.com/elementary-statistics/probability-distributions/exponential-distribution)

[Normal Distribution](http://www.r-tutor.com/elementary-statistics/probability-distributions/normal-distribution)

[Chi-squared Distribution](http://www.r-tutor.com/elementary-statistics/probability-distributions/chi-squared-distribution)

[Student t Distribution](http://www.r-tutor.com/elementary-statistics/probability-distributions/student-t-distribution)

[F Distribution](http://www.r-tutor.com/elementary-statistics/probability-distributions/f-distribution)

**R Functions for Probability Distributions:**

Every distribution that R handles has four functions. There is a root name, for example, the root name for the normal distribution is norm. This root is prefixed by one of the letters

p for "probability", the cumulative distribution function (c. d. f.)

q for "quantile", the inverse c. d. f.

d for "density", the density function (p. f. or p. d. f.)

r for "random", a random variable having the specified distribution

For the normal distribution, these functions are pnorm, qnorm, dnorm, and rnorm. For the binomial distribution, these functions are pbinom, qbinom, dbinom, and rbinom. And so forth.

For a continuous distribution (like the normal), the most useful functions for doing problems involving probability calculations are the "p" and "q" functions (c. d. f. and inverse c. d. f.), because the the density (p. d. f.) calculated by the "d" function can only be used to calculate probabilities via integrals and R doesn't do integrals.

For a discrete distribution (like the binomial), the "d" function calculates the density (p. f.), which in this case is a probability

f(x) = P(X = x)

and hence is useful in calculating probabilities.

R has functions to handle many probability distributions. The table below gives the names of the functions for each distribution.

Table-1:Probability Distributions

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Distribution | Functions | | | |
| Binomial | pbinom | qbinom | dbinom | rbinom |
| Cauchy | pcauchy | qcauchy | dcauchy | rcauchy |
| Chi-Square | pchisq | qchisq | dchisq | rchisq |
| Exponential | pexp | qexp | dexp | rexp |
| F | pf | qf | df | rf |
| Gamma | pgamma | qgamma | dgamma | rgamma |
| Geometric | pgeom | qgeom | dgeom | rgeom |
| Hypergeometric | phyper | qhyper | dhyper | rhyper |
| Logistic | plogis | qlogis | dlogis | rlogis |
| Log Normal | plnorm | qlnorm | dlnorm | rlnorm |
| Normal | pnorm | qnorm | dnorm | rnorm |
| Poisson | ppois | qpois | dpois | rpois |
| Student t | pt | qt | dt | rt |
| Uniform | punif | qunif | dunif | runif |
| Weibull | pweibull | qweibull | dweibull | rweibull |

**Procedure:**

Open RStudio

Go to RConsole (>)

Probability distribution in R

>help(rnorm) #The normal Distribution

>help(dbinom) # The Binomial Distribution

**Probability Distributions in R:**

In R, probability functions take the form

***[dpqr]distribution\_abbreviation ()***

where the first letter refers to the aspect of the distribution returned:

d = density

p = distribution function

q = quantile function

r = random generation (random deviates)

# **1. Binomial Distribution**

The binomial distribution is a discrete probability distribution. It describes the outcome of n independent trials in an experiment. Each trial is assumed to have only two outcomes, either success or failure. If the probability of a successful trial is p, then the probability of having x successful outcomes in an experiment of n independent trials is as follows.

IMG_256

#### **Problem**

Suppose there are twelve multiple choice questions in an English class quiz. Each question has five possible answers, and only one of them is correct. Find the probability of having four or less correct answers if a student attempts to answer every question at random.

**Example Solution:**

Since only one out of five possible answers is correct, the probability of answering a question correctly by random is 1/5=0.2. We can find the probability of having exactly 4 correct answers by random attempts as follows.

> dbinom(4, size=12, prob=0.2)   
[1] 0.1329

To find the probability of having four or less correct answers by random attempts, we apply the function dbinom with x = 0,…,4.

> dbinom(0, size=12, prob=0.2) +   
+ dbinom(1, size=12, prob=0.2) +   
+ dbinom(2, size=12, prob=0.2) +   
+ dbinom(3, size=12, prob=0.2) +   
+ dbinom(4, size=12, prob=0.2)   
[1] 0.9274

Alternatively, we can use the cumulative probability function for binomial distribution pbinom.

> pbinom(4, size=12, prob=0.2)   
[1] 0.92744

#### **Answer:**The probability of four or less questions answered correctly by random in a twelve question multiple choice quiz is 92.7%.

**2. Poisson Distribution**

The Poisson distribution is the probability distribution of independent event occurrences in an interval. If λ is the mean occurrence per interval, then the probability of having x occurrences within a given interval is:

IMG_256

**Problem**

If there are twelve cars crossing a bridge per minute on average, find the probability of having seventeen or more cars crossing the bridge in a particular minute.

Answer:

> lam=12

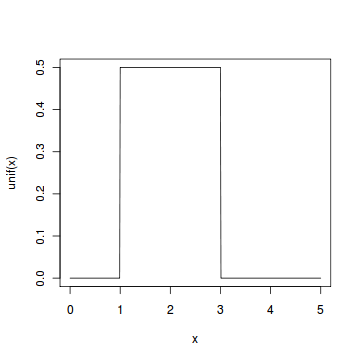
> ppois(16, lam, lower.tail = FALSE, log.p = FALSE)

[1] 0.101291

# 3. **Continuous Uniform Distribution**

The continuous uniform distribution is the probability distribution of random number selection from the continuous interval between a and b. Its density function is defined by the following.

IMG_256

Here is a graph of the continuous uniform distribution with a = 1, b = 3. 

#### **Problem**

Select ten random numbers between one and three.

**Answer:**

> runif(10, min = 1, max = 3)

[1] 2.961773 2.872257 1.965110 2.826935 2.422518 1.363877 2.768416 1.101335 2.920489

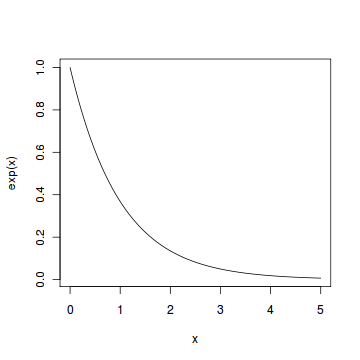
[10] 1.293350

# 4. **Exponential Distribution**

The exponential distribution describes the arrival time of a randomly recurring independent event sequence. If μ is the mean waiting time for the next event recurrence, its probability density function is:

IMG_256

Here is a graph of the exponential distribution with μ = 1.



#### **Problem**

Suppose the mean checkout time of a supermarket cashier is three minutes. Find the probability of a customer checkout being completed by the cashier in less than two minutes.

**Answer:**

> mew = 3

> pexp(2, 1/mew)

[1] 0.4865829

# **5.Normal Distribution**

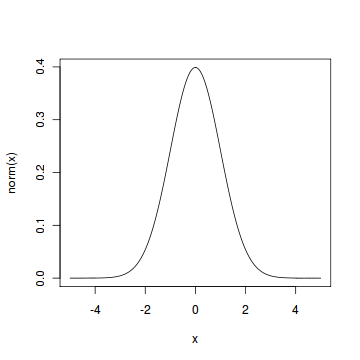
The normal distribution is defined by the following probability density function, where μ is the population mean and σ2 is the variance.

IMG_256

If a random variable X follows the normal distribution, then we write:

IMG_257

In particular, the normal distribution with μ = 0 and σ = 1 is called the standard normal distribution, and is denoted as N(0,1). It can be graphed as follows.



The normal distribution is important because of the Central Limit Theorem, which states that the population of all possible samples of size n from a population with mean μ and variance σ2 approaches a normal distribution with mean μ and σ2∕n when n approaches infinity.

#### **Problem**

Assume that the test scores of a college entrance exam fits a normal distribution. Furthermore, the mean test score is 72, and the standard deviation is 15.2. What is the percentage of students scoring 84 or more in the exam?

**Answer:**

> pnorm(83, mean = 72, sd = 15.2, lower.tail = FALSE)

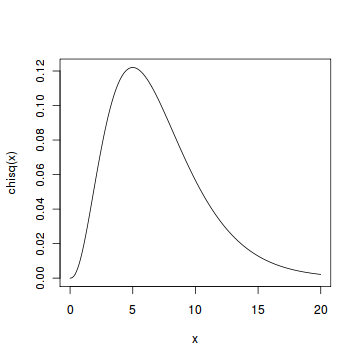
[1] 0.2346298

# **6.Chi-squared Distribution**

If X1,X2,…,Xm are m independent random variables having the standard normal distribution, then the following quantity follows a Chi-Squared distribution with m degrees of freedom. Its mean is m, and its variance is 2m.

IMG_256

Here is a graph of the Chi-Squared distribution 7 degrees of freedom.



#### **Problem**

Find the 95th percentile of the Chi-Squared distribution with 7 degrees of freedom.

Answer:

> qchisq(0.95, 7)

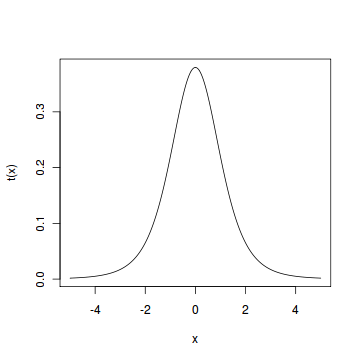
[1] 14.06714

# **8.Student t Distribution**

Assume that a random variable Z has the standard normal distribution, and another random variable V has the Chi-Squared distribution with m degrees of freedom. Assume further that Z and V are independent, then the following quantity follows a Student t distribution with m degrees of freedom.

IMG_256

Here is a graph of the Student t distribution with 5 degrees of freedom.



#### **Problem**

Find the 2.5th and 97.5th percentiles of the Student t distribution with 5 degrees of freedom.

**Answer:**

> qt(0.025, 5)

[1] -2.570582

> qt(0.975, 5)

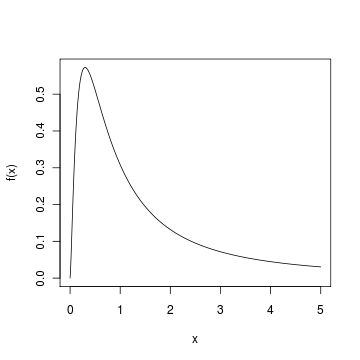
[1] 2.570582

# **8. F Distribution**

If V 1 and V 2 are two independent random variables having the Chi-Squared distribution with m1 and m2 degrees of freedom respectively, then the following quantity follows an F distribution with m1 numerator degrees of freedom and m2 denominator degrees of freedom, i.e., (m1,m2) degrees of freedom.

IMG_256

Here is a graph of the F distribution with (5, 2) degrees of freedom.



#### **Problem**

Find the 95th percentile of the F distribution with (5, 2) degrees of freedom.

**Answer:**

> qf(0.95,5,2)

[1] 19.29641

**Describe the following with respect to probability distributions:**

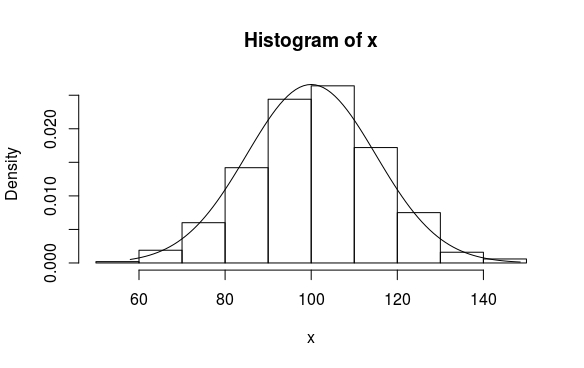
**1.**

x <- rnorm(1000, mean=100, sd=15)

hist(x, probability=TRUE)

xx <- seq(min(x), max(x), length=100)

lines(xx, dnorm(xx, mean=100, sd=15))

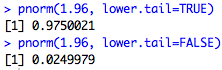


What is ***P*(*X* > 19)** when *X* has the **N(17.46, 375.67)** distribution?

> pnorm(19, mean=17.46, sd=sqrt(375.67), lower.tail=FALSE)

[1] 0.4683356

Interpret the following



In case 1, the output is the value of cumulative distribution less than or equal to the specified value (1.96 in this case).

In case 2, the output is the value of cumulative distribution strictly more than the specified value(1.96 in this case).

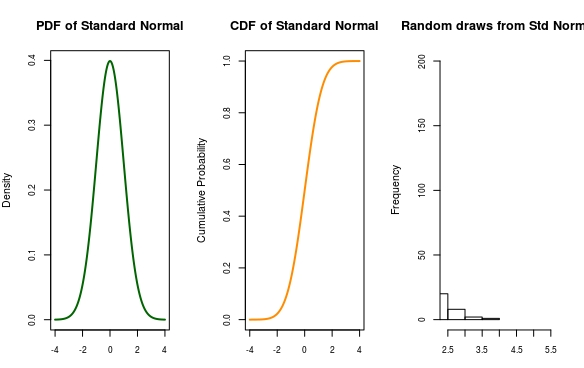
Run this in RStudio Script editor and explain it from plot

set.seed(3000)  
xseq<- seq(-4,4,.01)  
densities<- dnorm(xseq, 0,1)  
cumulative<- pnorm(xseq, 0, 1)  
randomdeviates<- rnorm(1000,0,1)  
 par(mfrow=[c](http://inside-r.org/r-doc/base/c)(1,3), mar=[c](http://inside-r.org/r-doc/base/c)(3,4,4,2))

plot(xseq, densities, col="darkgreen",xlab="", ylab="Density", type="l",lwd=2, cex=2, main="PDF of Standard Normal", cex.axis=.8)

plot(xseq, cumulative, col="darkorange", xlab="", ylab="Cumulative Probability",type="l",lwd=2, cex=2, main="CDF of Standard Normal", cex.axis=.8)

hist(randomdeviates, main="Random draws from Std Normal", cex.axis=.8, xlim=[c](http://inside-r.org/r-doc/base/c)(4,4))



**Conclusion:**

1. The different types of probability distributions were studied.
2. While plotting the histogram in exercise 1, different values were generated after each independent execution, which gave different looking graphs after each case.
3. In the last exercise, the first graph is a standard normal distribution. It means that the maximum probability is centred around the mean value (zero in this case). The second graph is the CDF, which is supposed to and is ending at the value 1.