PROBABILITY & STATISTICS(MA20205, LTP- 3-0-0,CRD- 3)

BUDDHANANDA BANERJEE

Contents

Date: Updated on November 1, 2021,
Weekly classes:MON(11:00-11:55) , TUE(08:00-08:55) , TUE(09:00-09:55) .

1. Syllabus

(1) Probability: Classical, relative frequency and axiomatic definitions of probability, addition rule and conditional probability, multiplication rule, total probability, Bayes' Theorem and independence.
(2) Random Variables and its distribution: Discrete, continuous and mixed random variables, probability mass, probability density and cumulative distribution functions, mathematical, expectation, moments, moment generating function, Chebyshev's inequality. Joint, marginal and conditional distributions, product moments, correlation, independence of random variables.
(3) Special Distributions: Discrete uniform, Binomial, Geometric, Poisson, Hypergeometric, Negative binomial, Continuous uniform, Exponential, Gamma, Normal distributions. Bivariate normal distribution. Functions of Random Variable(s).
(4) Sampling Distributions: Distributions of the sample mean and the sample variance for a normal population, Chi-Square, t and F distributions.
(5) Law of large numbers: The Central Limit Theorem, Weak law of large numbers
(6) Estimation: The method of moments and the method of maximum likelihood estimation, confidence intervals for the mean(s) and variance(s) of normal populations.
(7) Testing of Hypotheses: Null and alternative hypotheses, the critical and acceptance regions, two types of error, power of the test, the most powerful test and Neyman-Pearson Fundamental Lemma, tests for one sample problem for normal populations

2. Books

(1) 1. An Introduction to Probability and Statistics by V.K. Rohatgi & A.K. Md. E. Saleh
(2) Mathematical Statistics and Data Analysis by John A. Rice
(3) Probability and Statistical Inference by Hogg, R. V., Tanis, E. A. & Zimmerman D. L.
(4) Introduction to Probability Theory by Paul G. Hoel, Sidney C. Port and Charles J. Stone
(5) Introduction to Probability and Statistics for Engineers and Scientists by S.M. Ross
(6) Introduction to Probability and Statistics by J.S. Milton & J.C. Arnold.
(7) Introduction to Probability Theory and Statistical Inference by H.J. Larson
(8) Probability and Statistics for Engineers and Scientists by R.E. Walpole, R.H. Myers, S.L. Myers, Keying Ye
(9) Modern Mathematical Statistics by E.J. Dudewicz & S.N. Mishra
(10) Introduction to the Theory of Statistics by A.M. Mood, F.A. Graybill and D.C. Boes

3. Evaluation

- Continuous evaluation: 6th September, 4th October, 15th November (Slot B3, 11am-12Noon)

4. Preliminary

Definition 1. A set Ω is said to be finite if there exists an $n \in \mathbb{N}$ and a bijection from Ω onto $\{1, 2, \ldots, n\}$. An infinite set Ω is said to be countable if there is a bijection from $\mathbb{N}$ onto Ω.

If Ω is an infinite countable set, then using any bijection $f : \Omega \to \mathbb{N}$, we can list the elements of Ω as a sequence $f(1), f(2), f(3), ...$ so that each element of Ω occurs exactly once in the sequence. Conversely, if you can write the elements of Ω as a sequence, it defines an injective function from natural numbers onto Ω (send 1 to the first element of the sequence, 2 to the second element, etc.).

Example 2. The set of integers $\mathbb{Z}$ is countable. Define $f : \mathbb{N} \to \mathbb{Z}$ by

$$f(n) = \begin{cases} n/2, & \text{if n is even} \\ -(n-1)/2, & \text{if n is odd} \end{cases}$$

It is clear that f maps $\mathbb{N}$ into $\mathbb{Z}$. Thus, we have found a bijection from $\mathbb{N}$ onto $\mathbb{Z}$.which shows that $\mathbb{Z}$. is countable. This function is a formal way of saying the we can list the elements of $\mathbb{Z}$ as

$$0, +1, -1, +2, -2, +3, -3, ...$$

Check that f is one-one and onto.

Example 3. The set $\mathbb{N} \times \mathbb{N}$ is countable. Rather than give a formula, we list the elements of $\mathbb{N} \times \mathbb{N}$ as follows;

$$(1, 1), (1, 2), (2, 1), (1, 3), (2, 2), (3, 1), (1, 4), (2, 3), (3, 2), (4, 1), ...$$

Exercise 4. Define a bijection from $\mathbb{N}$ onto $\mathbb{N} \times \mathbb{N}$, and hence show that $\mathbb{N} \times \mathbb{N}$ is countable.

Example 5. $\mathbb{Z} \times \mathbb{Z}$ is countable.

Example 6. The set of rational numbers $\mathbb{Q}$ is countable.

Theorem 7. *The set of real numbers $\mathbb{R}$ is not countable.*

Proof. The extraordinary proof of this fact is due to Cantor, and the core idea, called the 'diagonalization argument' is one that can be used in many other contexts. □

Consider any function $f : \mathbb{N} \to [0, 1]$. We show that it is not onto, and hence not a bijection. Indeed, use the decimal expansion to write a number $x \in [0, 1]$ as $0.x_1x_2x_3 \ldots$ where $x_i \in \{0, 1, \ldots, 9\}$. Write the decimal expansion for each of the numbers $f(1), f(2), f(3), \ldots.$ as follows:

$$\begin{aligned} f(1) &= 0.X_{1,1}X_{1,2}X_{1,3} \ldots \\ f(2) &= 0.X_{2,1}X_{2,2}X_{2,3} \ldots \\ f(3) &= 0.X_{3,1}X_{3,2}X_{3,3} \ldots \\ &\ldots\ldots \ldots \ldots \end{aligned}$$

Let $Y_1, Y_2, Y_3, \ldots$ be any numbers in $\{0, 1, \ldots, 9\}$ with the only condition that $Y_i \neq X_{i,i}$. Clearly, it is possible to choose Y_i like this. Now, consider the number

$y = 0.Y_1Y_2Y_3\ldots$ which is a number in $[0,1]$. However, it does not occur in the above list. Indeed, y disagrees with $f(1)$ in the first decimal place, disagrees with $f(2)$ in the second decimal place, etc. Thus, $y \neq f(i)$ for any $i \in \mathbb{N}$ which means that f is not onto $[0,1]$.

Theorem 8. *Thus, no function $f : \mathbb{N} \to [0,1]$ is onto, and hence there is no bijection from $\mathbb{N}$ onto $[0,1]$ and hence $[0,1]$ is not countable. Obviously, if there is no onto function onto $[0,1]$, there cannot be an onto function onto $\mathbb{R}$. Thus, $\mathbb{R}$ is also uncountable.*

Example 9. Let $A_1, A_2, ...$ be subsets of a set Ω. Suppose each A_i is countable (finite is allowed). Then, $\cup_i A_i$ is also countable.

5. Probability : Definition & laws

Definition 10. Random experiment: A random experiment is a physical phenomena which satisfies the followings.

(1) It has more than one outcomes.
(2) The outcome of a particular trial is not known in advance.
(3) It can be repeated countably many times in in *identical* condition.

Example 11. (a) Tossing a coin, (b) Rolling a die and (c) Arranging 52 cards etc.

Definition 12. Sample space: A set which is collection of all possible outcomes of a random experiment is known as sample space for the experiment and it is denoted by Ω or S.

Example 13. For the above examples the sample spaces are
(a) $\{H, T\}$, (b) $\{1, 2, 3, 4, 5, 6\}$, (c) $\{\pi | \pi$ is any permutation of 52 cards$\}$ respectively

Definition 14. Classical definition of probability: If the sample space (Ω) of a random experiment is a $finite$ set and $A \subseteq \Omega$ the probability of A is defined as

$$P(A) = \frac{|A|}{|\Omega|}$$

under the assumption that all outcomes are equally likely. Here $|\cdot|$ denotes the cardinality of a set.

Exercise 15. Consider the equation $a_1 + a_2 + \cdots + a_r = n$ where $r < n$. Suppose a computer provides an integral solution of it at random such that each $a_i \in \mathbb{N} \cup \{0\}$ for any solution. Find the probability that each $a_i \in \mathbb{N}$ only, for a solution.

Definition 16. Frequency definition of probability: If the sample space (Ω)of a random experiment is a *countable* set and $A \subseteq \Omega$ the probability of A is defined as

$$P(A) = \lim_{n \uparrow \infty} \frac{|A_n|}{|\Omega_n|}$$

where $\lim_{n\uparrow\infty} A_n = A$ and $\lim_{n\uparrow\infty} \Omega_n = \Omega$ and $A_n \subseteq \Omega_n$.

Exercise 17. What is the probability that a randomly chosen number from $\mathbb{N}$ will be a even number?

Exercise 18. What is the probability that a randomly chosen number from $\mathbb{N}$ will be a $k - digit$ number? Can you conclude an answer with frequency definition of probability?

Definition 19. Algebra: A collection $\mathcal{A}$ of the subsets of Ω is called an algebra if

(1) $\Omega \in \mathcal{A}$
(2) any $A \subseteq \Omega$ and $A \in \mathcal{A}$ then $A^c \in \mathcal{A}$ [closed under complementation]
(3) any $A, B \subseteq \Omega$ and $A, B \in \mathcal{A}$ implies $A \cup B \in \mathcal{A}$ [closed under finite union]

Definition 20. σ-algebra or σ-field: An algebra $\mathcal{A}$ of the subsets of Ω is called an σ-algebra/ field if $\{A_i\} \subseteq \Omega$ and $\{A_i\} \in \mathcal{A}$ implies $\cup_{i=1}^{\infty} A_i \in \mathcal{A}$ [closed countable union].

Example 21. (a) $\mathcal{A} = \{\emptyset, \Omega\}$, (b) $\mathcal{A} = \{\emptyset, \Omega, A, A^c\}$, (c) $\mathcal{A} = 2^{\Omega}$ [power set]

Definition 22. Axiomatic definition of probability (Kolmogorov): If $\mathcal{A}$ is σ-algebra of the subsets of a non-empty set Ω then the probability (P) is defined to be a function $P : \mathcal{A} \mapsto [0, 1]$ which satisfies,

(1) $P(\Omega) = 1$,
(2) $P(A) \geq 0$ for any $A \in \mathcal{A}$,
(3) $\{A_i\} \in \mathcal{A}$ implies $P\left(\cup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$ if $A_i \cap A_j = \emptyset \quad \forall i \neq j$.

Example 23. What is the probability that a randomly chosen number from $\Omega = [0, 1]$ will be

(1) a rational number?
(2) less than 0.4? Can you conclude an answer with classical / frequency definition of probability?

Definition 24. Probability space: $(\Omega, \mathcal{A}, P)$ is known as a probability space.

Definition 25. Event: For a given Probability space $(\Omega, \mathcal{A}, P)$ if $A \subseteq \Omega$ and $A \in \mathcal{A}$ then A is called an event.

Definition 26. Conditional Probability: For a given Probability space $(\Omega, \mathcal{A}, P)$ if A and$B$ are two events such that $P(B) > 0$ then the conditional probability of A given B is defined as

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

Definition 27. Independent events: For a given probability space $(\Omega, \mathcal{A}, P)$ the two events A and B are called *independent* if $P(A|B) = P(A)$, which implies

$$P(A \cap B) = P(A)P(B).$$

Remark 28. If the probability function P is changed to some P_1 on the same $(\Omega, \mathcal{A})$ then events A and B may not be independent any more.

Definition 29. Pairwise independence: For a given probability space $(\Omega, \mathcal{A}, P)$ consider a sequence of events $\{A_i\}$. This sequence of events are called pairwise independent if

$$P(A_i \cap A_j) = P(A_i)P(A_j) \quad \forall i \neq j.$$

Definition 30. Mutual independence: For a given probability space $(\Omega, \mathcal{A}, P)$ consider a sequence of events $\{A_i\}$. This sequence of events are called mutually independent if

$$P(\cap_{i_1, i_2 \cdots i_k} A_i) = \Pi_{i_1, i_2, \cdots i_k} P(A_i) \quad \forall i_1 \neq i_2, \neq \cdots \neq i_k.\text{for any} \quad k \in \mathbb{N}$$

Exercise 31. Give an example to show that pairwise independence does not imply mutual independence.

Definition 32. Mutually exclusive events: For a given probability space $(\Omega, \mathcal{A}, P)$ a sequence of events $\{A_i\}$ are called mutually exclusive if $A_i \cap A_j = \emptyset \quad \forall i \neq j$

Definition 33. Mutually exhaustive events: For a given probability space $(\Omega, \mathcal{A}, P)$ a sequence of events $\{A_i\}$ are called mutually exhaustive if $\cup_{i=1}^{\infty} A_i = \Omega$

Remark 34. Mutually exhaustive or exclusive events **does not depend the probability function**.

Definition 35. Partition: For a given probability space $(\Omega, \mathcal{A}, P)$ a sequence of events $\{A_i\}$ are called a partition of Ω if $\{A_i\}$ are **mutually exclusive and exhaustive.**

Exercise 36. Prove the following properties:

(1) $P(A^c) = 1 - P(A)$
(2) $P(\emptyset) = 0$
(3) If $A \subseteq B$ then $P(A) \leq P(B)$
(4) $1 - P(\cup_{i=1}^{\infty} A_i) = P(\cap_{i=1}^{\infty} A_i^c)$
(5) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
(6) $P(\cup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} P(A_i)$
(7) $P\left(\cup_{i=1}^{n} A_i\right) = \sum_{k=1}^{n} (-1)^{k-1} S_k$ where

$$S_k = \sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq n} P(A_{i_1} \cap \cdots \cap A_{i_k})$$

Exercise 37. Suppose n letters are put in n envelops distinct by addresses. What is the probability that no letter will reach to the correct address. What is the limiting probability as $n \uparrow \infty$?

Theorem 38. ***Bayes Theorem:*** *Let $A_1, A_2, \cdots A_k$ is a partition of Ω and $(\Omega, \mathcal{A}, P)$ be a probability space with $P(A_i) > 0 \ \ \forall i$ and $P(B) > 0$ for some $B \subseteq \Omega$. Then*

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{i=1}^{k} P(B|A_i)P(A_i)}.$$

Exercise 39. There are three drawers in a table. The first drawer contains two gold coins. The second drawer contains a gold and a silver coin. The third one contains two silver coins. Now a drawer is chosen at random and a coin is also chose randomly. It is found that the a gold coin has been selected then what is the probability that the second drawer was chosen?

Exercise 40. Consider the quadratic equation $u^2 - \sqrt{Y}u + X = 0$, where (X, Y) is a random point chosen uniformly from a unit square. What is the probability that the equation will have a real root?

Exercise 41. Give a randomized algorithm to approximate value of π.

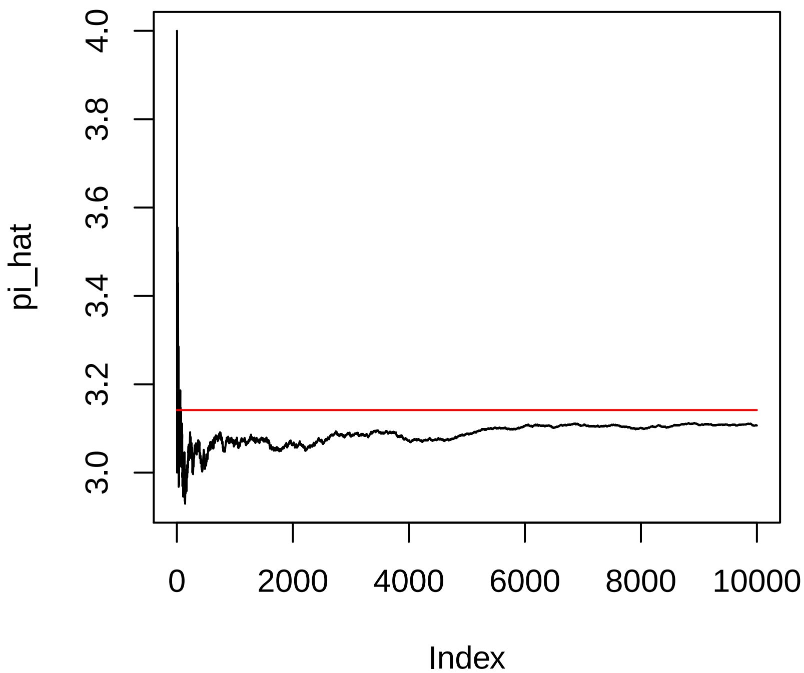
Exercise 42. Give a randomized algorithm to approximate value of e.

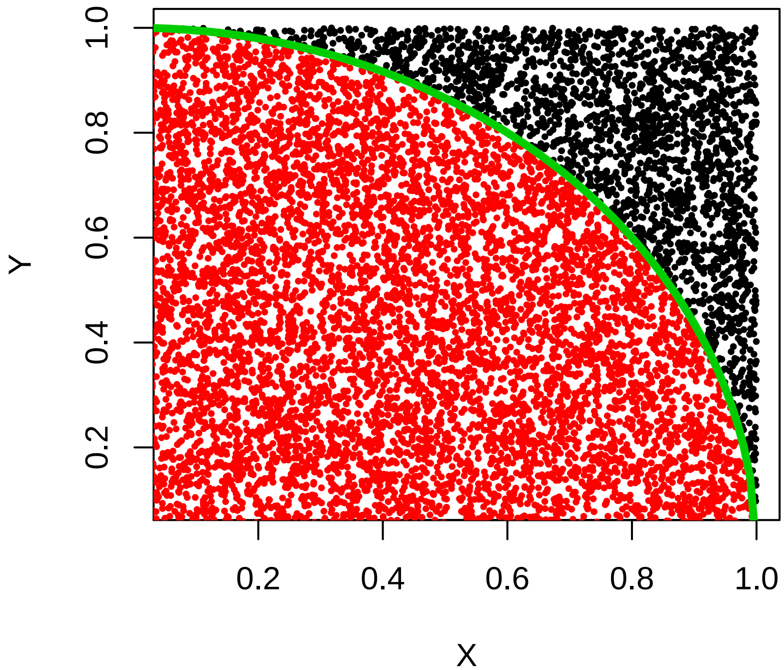
Exercise 43. Let a and b are randomly chosen from $\mathbb{N}$. Find the value

$$P(gcd(a, b) = 1)$$

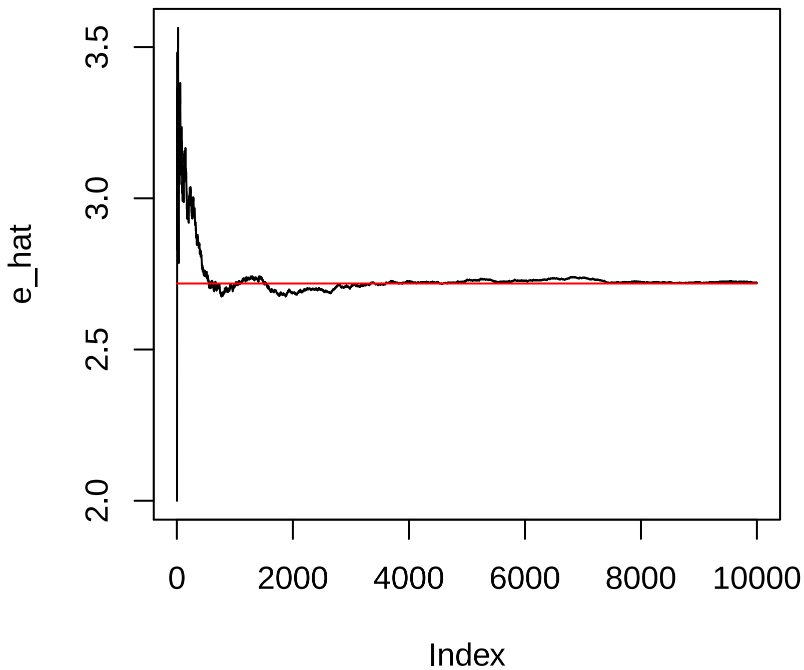
. Find its connection with Euler's Pi Prime Product and Riemann's Zeta Function. [FUN EXERCISE]

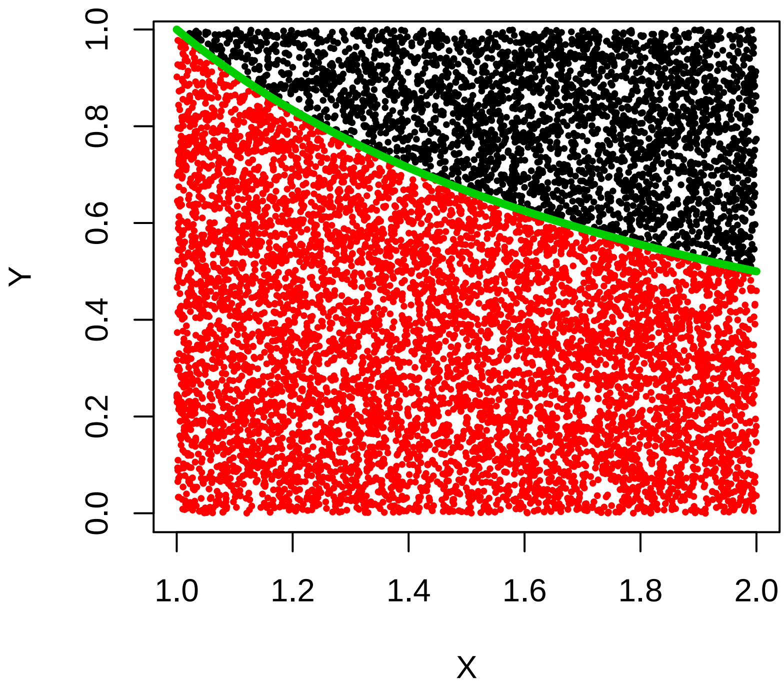
```r
# Date:26 July 2019
## Pi estimation
## Range of X
a1<-0
b1<-1
## Range of Y
a2<-0
b2<-1
itrn<- 10000 # iteration number
x<-runif(itrn,a1,b1) # generate X
y<-runif(itrn,a2,b2) # generate Y
pi_true<-rep(pi,itrn) # True value of pi
pi_hat<-array(0,dim=c(itrn))
count<-array(0,dim=c(itrn))
xx<-seq(a1,b1,by=0.01) # Sequence on [0,1]

for(i in 1 : itrn){
  count [i]<- (x[i]^2+y[i]^2<1) # (X,Y) in circle or not
  pi_hat[i]<-4*(sum(count)/i) # How many in cirecle amonge 'i' trials
}
# Plot
plot(pi_hat, type = 'l')
lines(pi_true, col=2)

s0<- which (count==0) # points out of circle
s1<- which (count==1) # points in circle
plot(y[s0]~x[s0], pch = 20, cex = 0.5, xlab="X", ylab = "Y")
lines (y[s1]~x[s1], col=2, type='p', pch = 20, cex = 0.5)
lines(sqrt(1-(xx)^2)~xx , col=3, lwd=4) # equation of circle
```


pi_hat
3.0
3.2
3.4
3.6
3.8
4.0
0
2000
4000
6000
8000
10000
Index


1.0
0.8
0.6
0.4
0.2
Y
0.2
0.4
0.6
0.8
1.0
X

```
# e estimation
# # Date:26 July 2019
a1<-1
b1<-2
a2<-0
b2<-1
itrn<- 10000
x<-runif(itrn,a1,b1)
y<-runif(itrn,a2,b2)
e_true<-rep(exp(1),itrn)
e_hat<-array(0,dim=c(itrn))
count<-array(0,dim=c(itrn))
xx<-seq(a1,b1,by=0.01)
for(i in 1 : itrn){
  count [i]<- (x[i]*y[i]<1)
  area<-(sum(count)/i)
  e_hat[i]<- 2^(1/area)

}
plot(e_hat, type = 'l')
lines(e_true, col=2)
s0<- which (count==0)
s1<- which (count==1)
plot(y[s1]~x[s1], col=2, pch = 20, cex = 0.5, xlab="X", ylab = "Y")
lines (y[s0]~x[s0], type='p', pch = 20, cex = 0.5)
lines((1/(xx))~xx , col=3, lwd=4)
```


e_hat
3.5
3.0
2.5
2.0
0
2000
4000
6000
8000
10000
Index



Y
X
0.0
0.2
0.4
0.6
0.8
1.0
1.0
1.2
1.4
1.6
1.8
2.0

6. Random Variable and it's moments

Definition 44. Random variable: Let $(\Omega, \mathcal{A}, P)$ be a probability space. Then a function $X : \Omega \to \mathbb{R}$ is called a random variable if

$$X^{-1}((-\infty, x]) \equiv \{\omega | X(\omega) \leq x\} \in \mathcal{A} \;\; \forall x \in \mathbb{R}$$

Remark 45. **Random variable is a deterministic function which has nothing random in it.**

Remark 46. Consider a function $g : \mathbb{R} \to \mathbb{R}$ and X is a random variable on $(\Omega, \mathcal{A}, P)$. Then $Y = g(X)$ is also a random variable. It implies that $X^{-1}(g^{-1}(-\infty, x]) \in \mathcal{A}$ for any $x \in \mathbb{R}$. It means $Q((-\infty, x]) = P(Y \in (-\infty, x]) = P(X \in g^{-1}((-\infty, x]))$, which is known as **push forward** of probability.

Definition 47. Vector Valued Random variable: $(X_1(\omega), X_2(\omega), \cdots, X_k(\omega))$ is a vector valued random variable where $\omega \in \Omega$.

Remark 48. Let X and Y be random variables. Then,

- $aX + bY$ is a random variable for all$a, b \in \mathbb{R}$.
- $\max\{X, Y\}$ and $\min\{X, Y\}$ are random variables.
- XY is a random variable.
- Provided that $P(Y(\omega) = 0) = 0$ for each $\omega \in \Omega$, then X/Y is a random variable.

Definition 49. Cumulative distribution function (c.d.f.): Cumulative distribution function of a random variable X is a function $F : \mathbb{R} \to [0, 1]$ defined as

$$\begin{aligned} F(x) &= P(X \leq x) \\ &= P(X^{-1}(-\infty, x]) \\ &= P(\{\omega | X(\omega) \in (-\infty, x]\}) \;\; \forall x \in \mathbb{R}. \end{aligned}$$

Remark 50. Cumulative distribution function uniquely identifies a random variable.

Exercise 51. Prove the properties of a c.d.f. :

(1) $F(-\infty) = \lim\limits_{x \downarrow -\infty} F(x) = 0$
(2) $F(\infty) = \lim\limits_{x \uparrow \infty} F(x) = 1$
(3) $F(a) \leq F(b) \quad \forall a \leq b \in \mathbb{R}$ [non-decreasing]
(4) $F(a) = \lim\limits_{x \downarrow a} F(x) \quad \forall a \in \mathbb{R}$ [right-continuous]

Definition 52. Discrete valued random variable: For a given probability space $(\Omega, \mathcal{A}, P)$ a random variable X is said to be a discrete valued random variable if $S = \{X(\omega) | \omega \in \Omega\}$ is a **finite or countably infinite** set and $X^{-1}(s_i) \in \mathcal{A}$ for all $s_i \in S$.

Remark 53. There can be finitely or countably many jump discontinuities in a c.d.f. of a random variable. The sum of the magnitude of jumps is one, which is the total probability.

Definition 54. Probability mass function(p.m.f.): If X is a discrete valued random variable in a given probability space $(\Omega, \mathcal{A}, P)$ then a non-negative function

$f(x) := P(X = x)$ on $\mathbb{R}$ is called a probability mass function or discrete density function of the random variable X. Probability mass function has the following properties

- $f(x) \geq 0 \quad \forall x \in \mathbb{R}$.
- $S = \{x | f(x) > 0\}$ is finite or a countably infinite set.
- $\displaystyle\sum_{s \in S} f(s) = 1$

Definition 55. Continuous valued random variable: For a given probability space $(\Omega, \mathcal{A}, P)$ a random variable X is said to be a continuous valued random variable $P(X = x) = 0 \forall x \in \mathbb{R}$.

Definition 56. Probability density function(p.d.f.): If X is a continuous valued random variable in a given probability space $(\Omega, \mathcal{A}, P)$ with c.d.f $F(\cdot)$ then a non-negative function $f : \mathbb{R} \to [0, \infty)$ is called **a probability density function** of X if

$$P(X \in A) = \int_x f(x) \mathbf{1}_{\{x \in A\}} dx$$

Remark 57. In particular for $A = (-\infty, x]$ for any $x \in \mathbb{R}$ then

$$f(x) = \frac{d}{dx} F(x) = \int_{-\infty}^{x} f(t) dt.$$

Definition 58. Expectation: The expectation of a random variable X with c.d.f $F_X(\cdot)$ is defined as $E(X) = \int x dF_X(x)$ where,

$$\int x dF_X(x) = \begin{cases} \sum_x x f(x), & \text{if } \sum_x |x| f(x) < \infty \text{ for discrete } X, \\ \int_x x f(x), & \text{if } \int_x |x| f(x) < \infty \text{ for continuous } X. \end{cases}$$

Exercise 59. Find the expectation of the random variables with the following densities

(a) $f(x) = \frac{1}{\pi(1+x^2)}$ when $x \in \mathbb{R}$

(b) $f(x) = \frac{1}{|x|(1+|x|)}$ when $x \in S = \{(-1)^n n | n \in \mathbb{N}\}$

Definition 60. Moment generating function: The moment generating function (**m.g.f.**) of a random variable X is defined as

$$M_X(t) = E(e^{tX}) \text{ if } E(e^{tX}) < \infty \ \forall t \in (-\epsilon, \epsilon) \text{ for some } \epsilon > 0$$

- Cumulative distribution function (c.d.f) and uniquely identify the probability distribution of a random variable.
- Moment generating function (m.g.f.) if exists then uniquely identifies the probability distribution of a random variable.
- Probability density function identifies the probability distribution of a random variable up to some length or volume zero set. So it is not unique in general.

Remark 61. Probability mass function can be considered as discrete density function with respect to count measure. If X is a discrete valued random variable with $P(X \in S) = 1$, where S is a countable set, then a non-negative function f is called a probability mass function or discrete density function of the random variable X if $P(X \leq x) = \sum_{s \in S} f(s) \mathbf{1}_{\{s \leq x\}} \quad \forall x \in \mathbb{R}$.

6.1. **Moments of a random variable.**

Definition 62. Raw moment: Let X be a discrete valued random variable with p.m.f $f(\cdot)$ such that $\sum_x |x^r| f(x) < \infty$. Then the r^{th} order raw moment of X is defined as

$$\mu_r^{'} = E(X^r) = \sum_x x^r f(x)$$

Definition 63. Central moment: Let X be a discrete valued random variable with p.m.f $f(\cdot)$ such that $\sum_x |(x-\mu_x)^r| f(x) < \infty$. Then the r^{th} order raw moment of X is defined as

$$\mu_r = E(X - \mu_x)^r = \sum_x (x - \mu_x)^r f(x).$$

- Mean of random variable X is $\mu_1^{'} = E(X) = \sum_x x f(x) = \mu_x$.
- Variance of random variable X is $\mu_2 = E(X - \mu_x)^2 = \sum_x (x-\mu_x)^2 f(x) = Var(X)$.

Exercise 64. Prove that:

$$\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{1}{2n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} (x_i - x_j)^2$$

Exercise 65. Show that $g(a) = \frac{1}{n} \sum_{i=1}^{n} (x_i - a)^2$ is minimum if $a = \bar{x}$.

Theorem 66. *If X is a non-negative integer-valued random variable with finite expectation then*

$$E(X) = \sum_{k=1}^{\infty} P(X \geq k)$$

Moments from Moment generating function: Let X be a discrete valued random with moment generating function

$$M_X(t) = E(e^{tx}) = \sum_{k=0}^{\infty} \frac{t^k E(X^k)}{k!}$$

Then one can obtain the kth order raw moment from m.g.f. by

$$\frac{\partial^k}{\partial t^k} M_X(t)|_{t=0} = \mu_k^{'}$$

Exercise 67. Prove the following inequalities:

- **Markov's Inequity:** If X is a non-negative valued random variable then
$$P(X > t) \leq \frac{E(X)}{t} \quad \forall t > 0$$
- **Chebyshev's Inequity:** $P(|X - \mu_x| > \epsilon) \leq \frac{E(X-\mu_x)^2}{\epsilon^2}$

Definition 68. For a discrete probability distribution, a median is by definition any real number m that satisfies the inequalities

$$\mathrm{P}(X \leq m) \geq \frac{1}{2} \text{ and } \mathrm{P}(X \geq m) \geq \frac{1}{2}$$

and for a continuous probability distribution,

$$\mathrm{P}(X \leq m) \geq \frac{1}{2} = \mathrm{P}(X \geq m) = \frac{1}{2}\,.$$

Exercise 69. Graphically show that $g(a) = \frac{1}{n}\sum_{i=1}^{n} |x_i - a|$ is minimum if $a =$ *median* of $\{x_1, \cdots, x_n\}$.

7. Modeling with Random Variables

Definition 70. Independent Random Variables: Two random variables X and Y are said to be independently distributed if

$$P((X,Y) \in A \times B)) = P(X \in A)P(Y \in B) \text{ for any } A, B \in \mathcal{B}(\mathbb{R})$$

This implies
(a)$P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y)$ for any $x, y \in \mathbb{R}$
(b)$f(x,y) = f_x(x)f_y(y)$

Definition 71. Identically distributed random variables: Two random variables X and Y are said to be identically distributed if

$$P(X \in (-\infty, a]) = P(Y \in (-\infty, a]) \text{ for any } a \in \mathbb{R}.$$

Remark 72. Two random variables X and Y are **independently and identically distribute (i.i.d.)** random variables if the above two definitions hold.

Definition 73. Uniform Distribution[0,1]: A random variable X is said to have uniform distribution over $[0,1]$ if

$$P(X \in A) = \frac{length(A)}{length([0,1])} = length(A)$$

for any interval $A \subseteq [0,1]$. If $X \sim U[0,1]$ then the p.d.f. is

$$f(x) = \begin{cases} 1, & \text{if } x \in [0,1] \\ 0, & \text{otherwise} \end{cases}$$

and the c.d.f. is

$$F(x) = \begin{cases} 0, & \text{if } x < 0 \\ x, & \text{if } x \in [0,1] \\ 1, & \text{if } x > 1 \end{cases}$$

7.1. Examples of discrete random variables:

- **Discrete uniform:** [random sampling with replacement from a finite population]

$$P(X = s) = \begin{cases} \frac{1}{k}, & \text{if } s \in \mathcal{S} = \{s_1, s_2, \cdots s_k\} \\ 0, & \text{otherwise} \end{cases}$$

- **Bernoulli (p):** [binary $\{0,1\}$ random variable]

$$P(X = x) = \begin{cases} p^x(1-p)^{1-x}, & \text{if } x \in \mathcal{S} = \{0,1\} \\ 0, & \text{otherwise} \end{cases}$$

- **Binomial (n,p):** [sum of n i.i.d. Bernoulli(p)]

$$P(X = x) = \begin{cases} \binom{n}{x} p^x(1-p)^{n-x}, & \text{if } x \in \mathcal{S} = \{0, 1 \cdots, n\} \\ 0, & \text{otherwise} \end{cases}$$

- **Geometric (p):** [number of failures preceding to the first success]

$$P(X = x) = \begin{cases} p(1-p)^x, & \text{if } x \in \mathcal{S} = \{0\} \cup \mathbb{N} \\ 0, & \text{otherwise} \end{cases}$$

- **Negative Binomial (r,p)** [sum of r i.i.d. geometric(p)]

$$P(X=x)=\begin{cases}\binom{x+r-1}{r-1}p^r(1-p)^x, & \text{if } x\in\mathcal{S}=\{0,1,2\cdots\}\\ 0, & \text{otherwise}\end{cases}$$

- **Poisson (λ):** [limiting distribution of $bin(n,p_n)$ when $n\uparrow\infty$, $p_n\downarrow 0$ but $np_n\to\lambda>0$]

$$P(X=x)=\begin{cases}\frac{e^{-\lambda}\lambda^x}{x!}, & \text{if } x\in\mathcal{S}=\{0\}\cup\mathbb{N}\\ 0, & \text{otherwise}\end{cases}$$

- **Hyper-geometric:** [random sampling without replacement from a finite population divided into two categories]

$$P(X=x)=\begin{cases}\frac{\binom{n_1}{r_1}\binom{n_2}{r_2}}{\binom{n_1+n_2}{r}}, & \text{if } r_1=0,1,\cdots\min\{n_1,r\}; r_2=0,1,\cdots\min\{n_2,r\}; r=r_1+r_2,\\ 0, & \text{otherwise}\end{cases}$$

Theorem 74. *Let $\{X_n\}$ be a sequence of random variables with corresponding m.g.f.s as$M_{X_n}(t)$ such that $\lim_{n\uparrow\infty}M_{X_n}(t)=M_Y(t)$ for some random variable Y. Then*

$$\lim_{n\uparrow\infty}F_{X_n}(a)=F_Y(a)$$

for all such $a\in\mathbb{R}$, where $F_Y(a)$ is continuous. We say X_n converges in distribution to Y. [Proof is not included in the syllabus]

7.2. **Examples of continuous random variables:**

Exercise 75. Let $X\sim U[0,1]$. Find the c.d.f and p.d.f

(a)$Y_1=cX$ where $c>0$. [this is known as $U[0,c]$]
(b)$Y_2=a+(b-a)X$ where $b>a$.[this is known as $U[a,b]$]
(c) Show that $E(Y_2)=\frac{b+a}{2}$ and $Var(Y_2)=\frac{(b-a)^2}{12}$.

Exercise 76. Let $X\sim U[0,1]$. Find the c.d.f and p.d.f.
(a)$Z=\frac{-1}{\lambda}\log(1-X)$ where $\lambda>0$[it is known as $exponential(\lambda)$ distribution]
(b)Show that $E(Z)=\frac{1}{\lambda}$ and $Var(Z)=\frac{1}{\lambda^2}$.

- A positive valued random Y variable is said to follow *Gamma distribution* with shape parameter $\alpha(>0)$ and scale parameter $\lambda(>0)$ if it has p.d.f.

$$f(y)=\frac{\lambda^\alpha e^{-\lambda y}y^{\alpha-1}}{\Gamma(\alpha)}\ \mathbf{1}_{\{y>0\}}$$

Remark 77. Let $X_1\sim Gamma(a,\lambda)$ and $X_2\sim Gamma(b,\lambda)$ are independently distributed then $Y=X_1+X_2\sim Gamma(a+b,\lambda)$. Use MGF.

Remark 78. If $X\sim Gamma(a,\lambda)$, then $E(X)=\alpha/\lambda$ and $Var(X)=\alpha/\lambda^2$

- Let $X_1\sim G(a,\lambda)$ and $X_2\sim G(b,\lambda)$ are independently distributed then $Y=\frac{X_1}{X_1+X_2}$ is said to follow$Beta(a,b)$ and the p.d.f. of $Y$is given by

$$f(y)=\begin{cases}\frac{y^{a-1}(1-y)^{b-1}}{B(a,b)}, & \text{if } y\in[0,1], a>0, b>0\\ 0, & \text{otherwise}\end{cases}$$

Remark. Let $Y\sim B(a,b)$ then, $E(Y)=\frac{a}{a+b}$ and $Var(Y)=\frac{ab}{(a+b)^2(a+b+1)}$

- A random variable X is is said to follow $Cauchy(\mu, \sigma)$ if it has the p.d.f.

$$f(x) = \frac{\sigma}{\pi(\sigma^2 + (y-\mu)^2)}$$

and has c.d.f.

$$F(x) = \frac{1}{2} + \frac{1}{\pi}\tan^{-1}\left(\frac{y-\mu}{\sigma}\right)$$

- A random variable X is is said to follow $Normal(\mu, \sigma^2)$ if it has the p.d.f.

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

  where $x, \mu \in \mathbb{R}$ and $\sigma > 0$.

NOTE:

(1) $N(0,1)$ is also known as standard normal distribution.
(2) p.d.f of standard normal distribution is denoted by $\phi(x) = \frac{e^{-\frac{1}{2}x^2}}{\sqrt{2\pi}}$
(3) c.d.f of standard normal distribution is denoted by $\Phi(x) = \int_{-\infty}^{x} \frac{e^{-\frac{1}{2}t^2}}{\sqrt{2\pi}} dt$
(4) If $Z \sim N(0,1)$ then$Y = \mu + \sigma Z \sim N(\mu, \sigma^2)$ with $E(Y) = \mu$ & $Var(Y) = \sigma^2$
(5) The p.d.f of Y can be written as $f(y) = \frac{1}{\sigma}\phi(\frac{y-\mu}{\sigma})$
(6) $\phi(\cdot)$ is a symmetric function around zero i.e. $\phi(z) = \phi(-z)$
(7) $\Phi(-z) = 1 - \Phi(z)$
(8) $P(Z < 1.64) = 0.95$ and $P(Z < 1.96) = 0.975$.

- **Chi-squared Distribution:** If $Z \sim N(0,1)$ then random variable$Y = Z^2$ is said to follow χ_1^2 i.e. chi-squared distribution with one degree of freedom.

Remark 79. Y following χ_1^2 has same p.d.f of $Gamma(1/2, 1/2)$ distribution.

Remark 80. If Z_i be i.i.d.$N(0,1)$ then random variable $Y = \sum_{i=1}^{n} Z_i^2$ follows χ_n^2 i.e. chi-squared distribution with$n$ degree of freedom which is equivalent to $Gamma(n/2, 1/2)$ distribution.

Remark 81. Let $Y \sim \chi_n^2$ then show that $E(Y) = n$ and $Var(Y) = 2n$

- $\mathbf{t-distribution}$: If $Z \sim N(0,1)$ and $Y \sim \chi_k^2$ are independently distributed random variables the

$$X = \frac{Z}{\sqrt{Y/k}} \sim t_k, \text{ i.e. t-distribution with k degrees freedom.}$$

- $\mathbf{F-distribution}$:If $Y_1 \sim \chi_{k_1}^2$ and $Y_2 \sim \chi_{k_2}^2$ are independently distributed random variables the

$$X = \frac{Y_1/k_1}{Y_2/k_2} \sim F_{k_1,k_2}, \text{ i.e. F-distribution with } k_1, k_2 \text{ degrees freedom.}$$

Theorem 82. *Let ϕ be a strictly monotone function on $I = (a,b)$ with the range $\phi(I)$ and differentiable inverse function $\phi^{-1}(\cdot)$ on $\phi(I)$. Also assume that X be a continuous valued random variable with p.d.f $f_X(x) = 0$ if $x \notin I$. Then $Y = \phi(X)$ has density $g(\cdot)$ on $\phi(I)$ as*

$$g(y) = f(\phi^{-1}(y)) \left| \frac{d}{dy}\phi^{-1}(y) \right|.$$

Exercise 83. Let $X \sim bin(n,p)$. Show that $M_X(t) = (pe^t + 1 - p)^n$

Exercise 84. Let $Y \sim pois(\lambda)$. Show that $M_Y(t) = e^{-\lambda(1-e^t)}$

Exercise 85. $X \sim bin(n, p_n)$ such that $n \uparrow \infty$, $p_n \downarrow 0$ and $np_n \to \lambda > 0$. Show that X_n converges in distribution to Y, where $Y \sim pois(\lambda)$

Exercise 86. Let $Y \sim pois(\lambda)$. Show that $E(Y) = Var(Y) = E(Y-\lambda)^3 = \lambda$

Exercise 87. Let $X \sim N(\mu, \sigma^2)$ then find the density function of $Y = e^X$. [Y is said to follow $lognormal(\mu, \sigma^2)$]

Exercise 88. Let $Y \sim lognormal(\mu, \sigma^2)$ find $E(X)$ and $Var(X)$.

Exercise 89. Let $X \sim N(0,1)$ then find the density function of$Y = X^2$.

Exercise 90. Let $X \sim N(\mu, \sigma^2)$ then them MGF of X is $M_X(t) = e^{\mu t + \sigma^2 t^2/2}$

Exercise 91. Let $X \sim exp(\lambda)$. Show that$Y = [X]$ has geometric distribution.

Exercise 92. Let $X \sim geo(p)$. Show that X has **memory less property** i.e. $P(X > m+n | X > m) = P(X > n) = q^n$ where $m, n \in \mathbb{N}$

Exercise 93. Let $X \sim exp(\lambda)$. Show that X has **memory less** property i.e.$P(X > t+s | X > t) = P(X > s)$ where $s, t \in \mathbb{R}$

Exercise 94. Let X be a continuous values random variable. Find the distribution of $Y = F(X)$.

Exercise 95. Let $(\frac{X}{\lambda})^k \sim exp(1)$ for $\lambda > 0$,$k > 0$, find the p.d.f. X. [X is said to follow Weibull distribution]

Exercise 96. Let $1 - (\frac{\lambda}{X})^k \sim U(0,1)$ for $X > \lambda$,$k > 0$, find the p.d.f. X. [X is said to follow Pareto distribution]

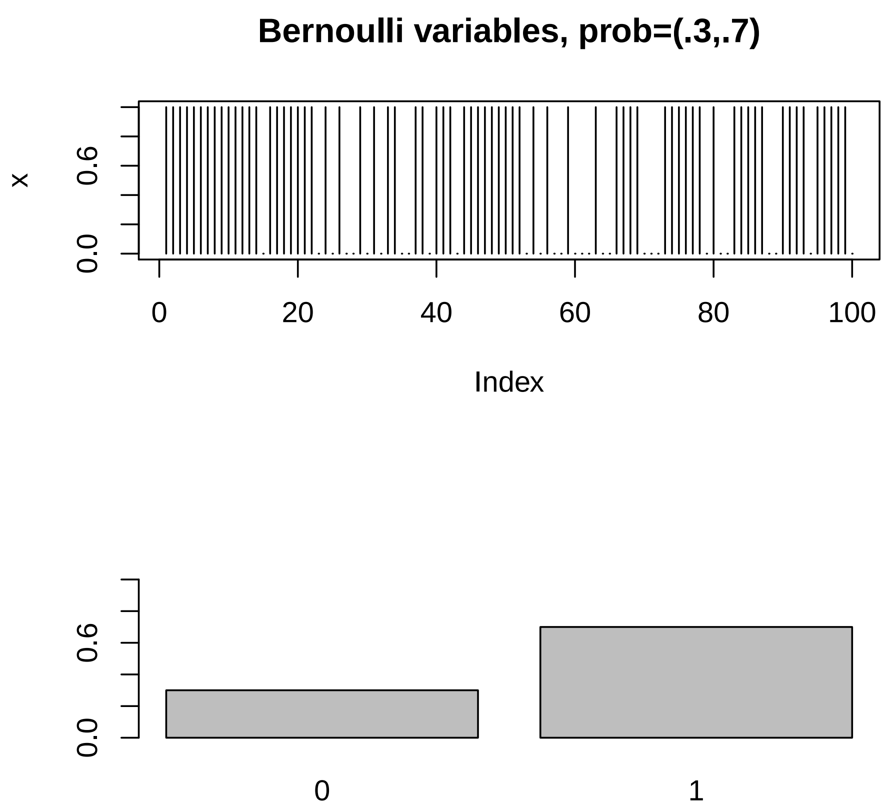
Exercise 97. Let $X \sim bin(n_1, p)$ and $Y \sim bin(n_2, p)$ independently . Use MGF to show that $Z = X + Y \sim bin(n_1 + n_2, p)$

Exercise 98. Let $X \sim pois(\lambda_1)$ and $Y \sim pois(\lambda_2)$ independently . Use MGF to show that $Z = X + Y \sim pois(\lambda_1 + \lambda_2)$

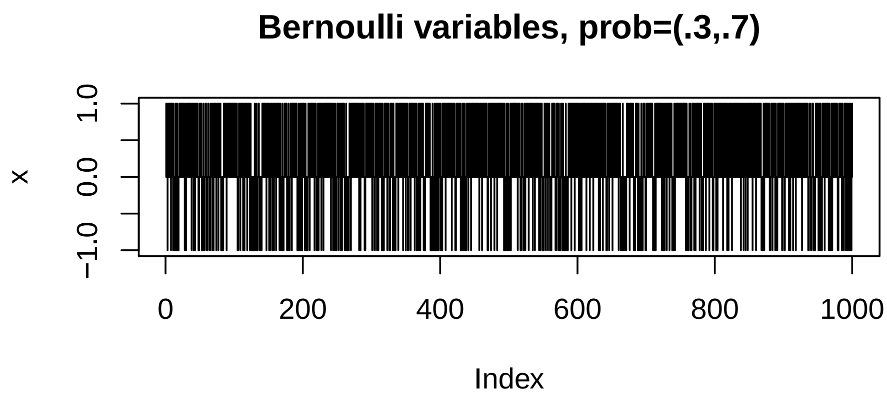
Exercise 99. Let $X \sim gamma(\alpha_1, \lambda)$ and $Y \sim bin(\alpha_2, \lambda)$ independently . Use MGF to show that $Z = X + Y \sim bin(\alpha_1 + \alpha_2, \lambda)$

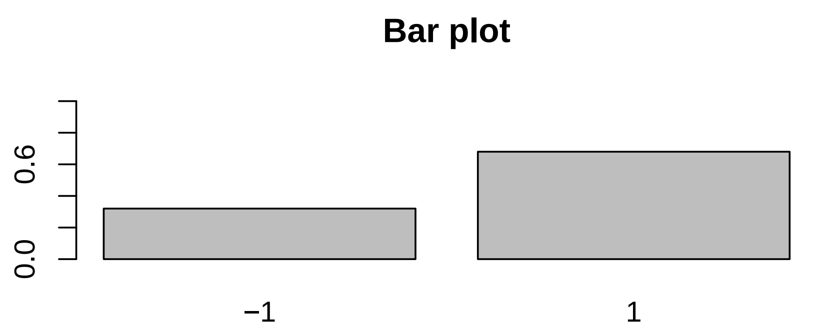
Exercise 100. Let $X \sim N(\mu_1, \sigma_1^2)$ and $Y \sim N(\mu_2, \sigma_2^2)$ independently . Use MGF to show that $Z = X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

```
# Bernoulli distribution of parameter p=0.7
n <- 100
x <- sample(c(0,1), n, replace=T, prob=c(.3,.7))
par(mfrow=c(2,1))
plot(x, type='h',main="Bernoulli variables, prob=(.3,.7)")
barplot(table(x)/n, ylim = c(0,1))
```

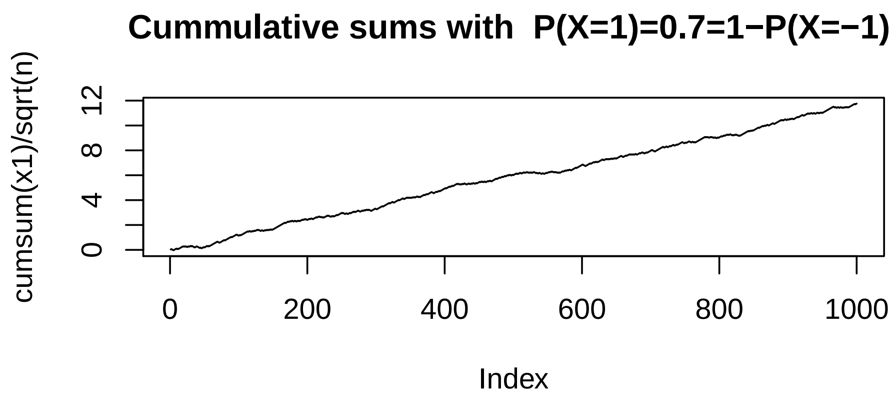


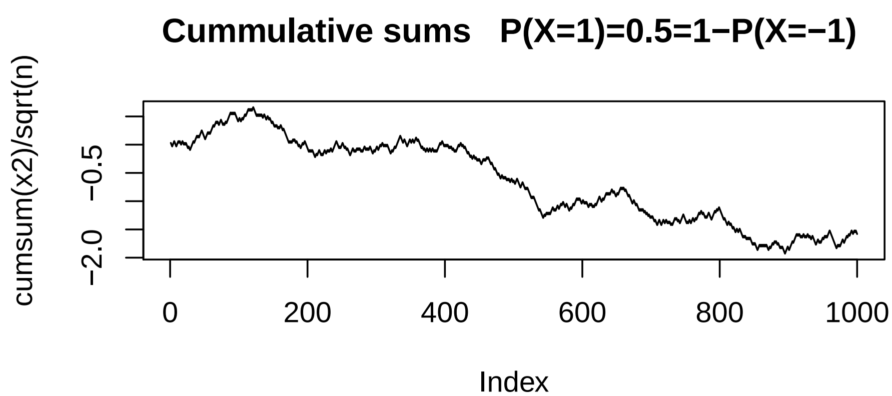

```
# Bernoulli distribution of parameter p=0.8 and X =-1 and +1
n <- 1000
x <- sample(c(-1,1), n, replace=T, prob=c(.3,.7))
par(mfrow=c(2,1))
plot(x, type='h', main="Bernoulli variables, prob=(.3,.7)")
barplot(table(x)/n, ylim = c(0,1), main = "Bar plot")
```



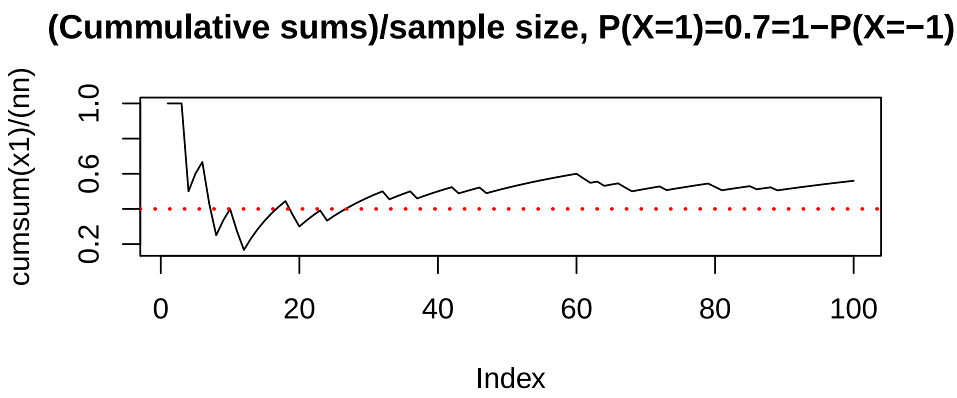


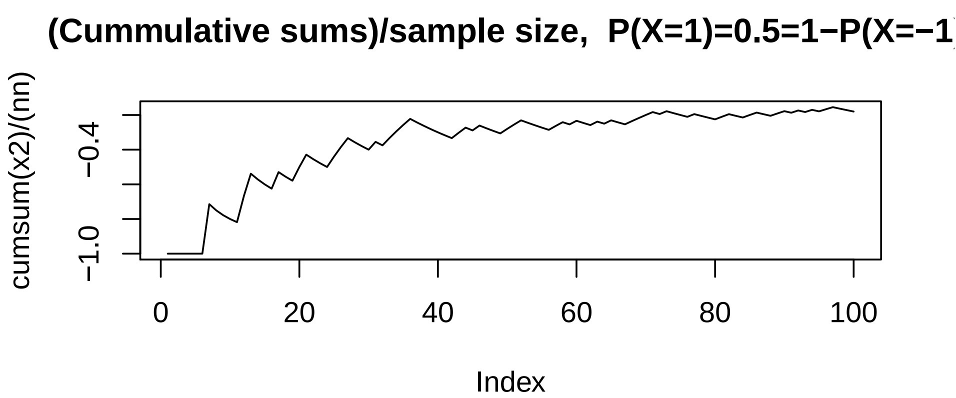

```r
# Cummulative sums  X =-1 and +1 with scaling 1/sqrt(n)
n <- 1000
x1 <- sample(c(-1,1), n, replace=T, prob=c(.3,.7))
x2 <- sample(c(-1,1), n, replace=T, prob=c(.5,.5))
par(mfrow=c(2,1))
plot(cumsum(x1)/sqrt(n), type='l',main="Cummulative sums with  P(X=1)=0.7=1-P(X=-1) ")
plot(cumsum(x2)/sqrt(n), type='l',main="Cummulative sums   P(X=1)=0.5=1-P(X=-1) ")
```



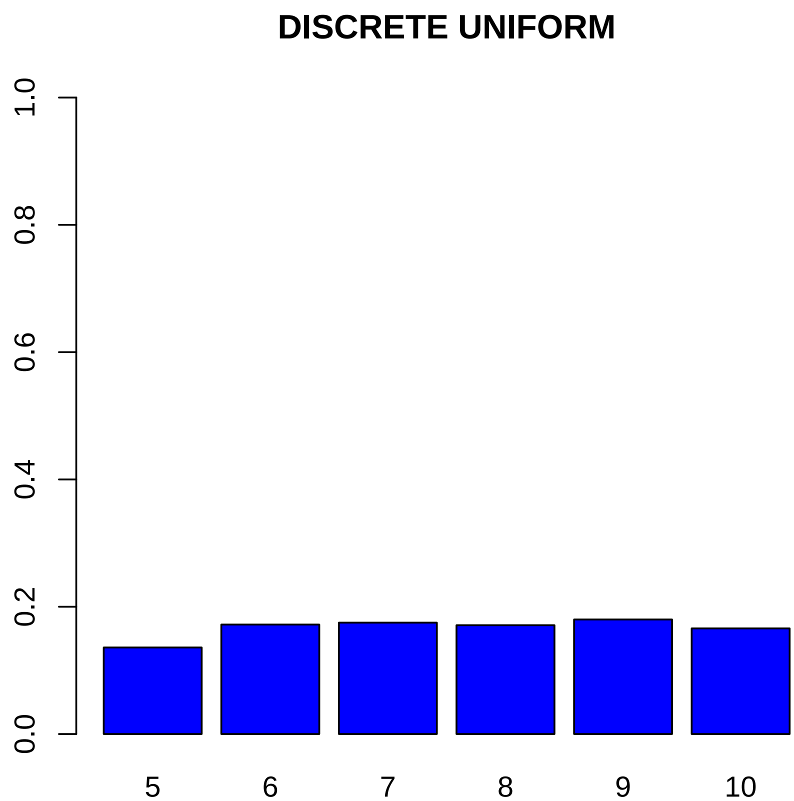



```
# Cummulative sums  X =-1 and +1 with scaling 1/(n)
n <- 100
nn<-((1:n))
x1 <- sample(c(-1,1), n, replace=T, prob=c(.3,.7))
x2 <- sample(c(-1,1), n, replace=T, prob=c(.5,.5))
par(mfrow=c(2,1))
plot(cumsum(x1)/(nn), type='l',
     main="(Cummulative sums)/sample size, P(X=1)=0.7=1-P(X=-1)  ")
abline(h=0.4, col=2,lwd=2, lty=3)
plot(cumsum(x2)/(nn), type='l',
     main="(Cummulative sums)/sample size,  P(X=1)=0.5=1-P(X=-1) ")
abline(h=0.0, col=2,lwd=2, lty=3)
```

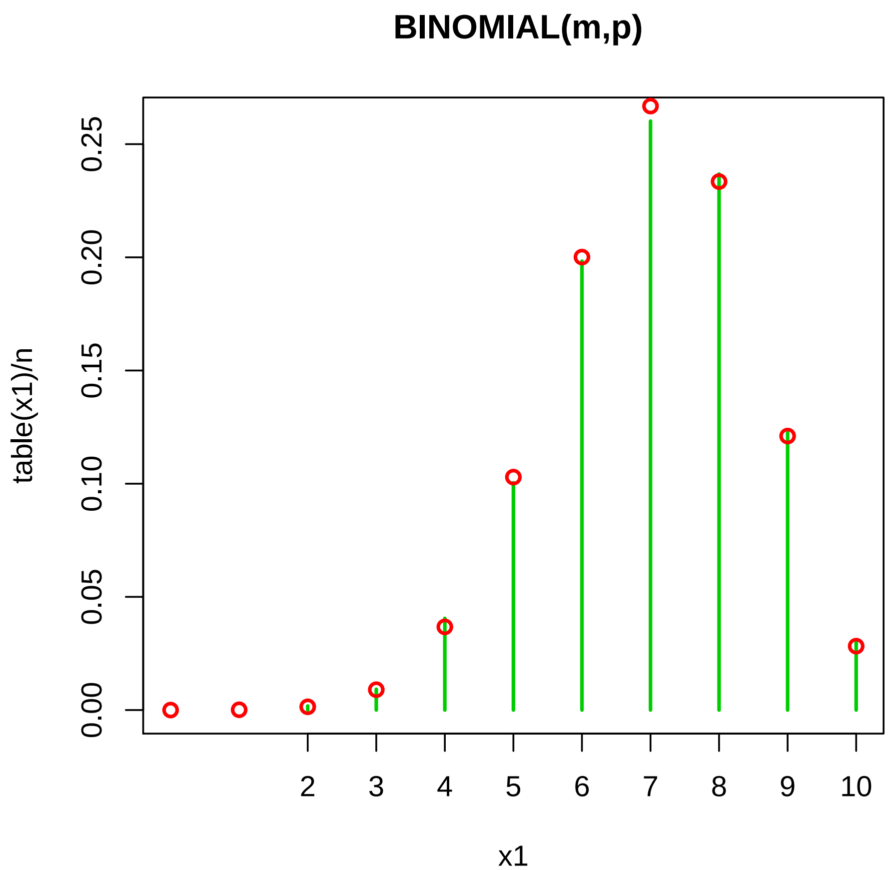




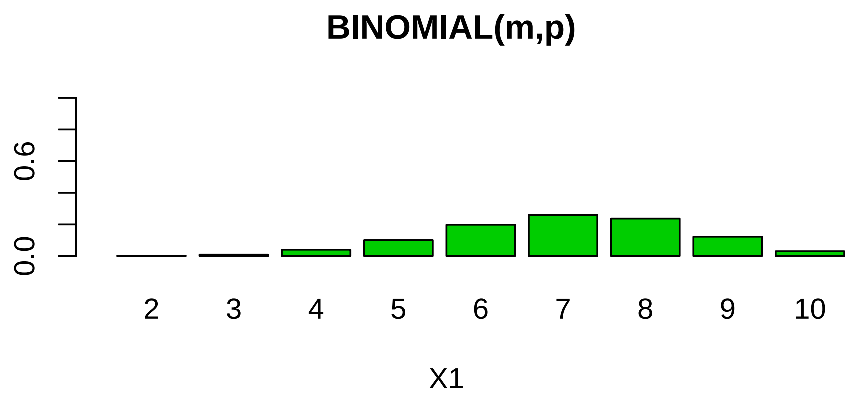

```
####################################################
#   DISCRETE UNIFORM
####################################################
n<- 1000 # sample size
sp<- 5:10 # sample space
x<-sample(sp, n, replace=T)   # data
par(mfrow=c(1,1))
barplot(table(x)/n, col=4, ylim=c(0,1), main = " DISCRETE UNIFORM ")
```

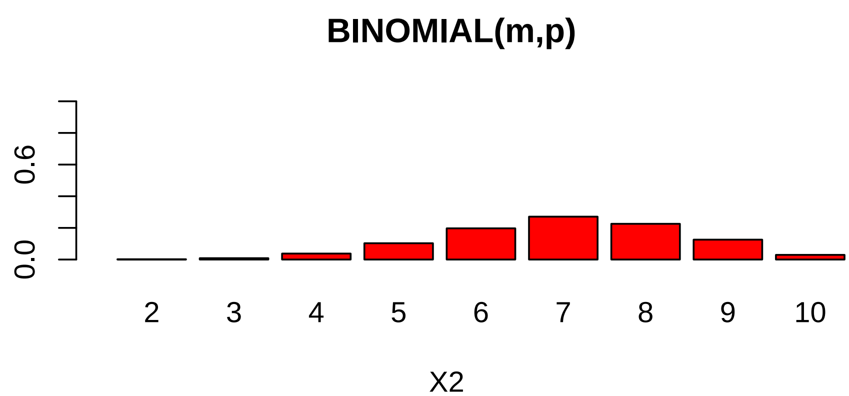



```
#####################################################
#   BINOMIAL(m,p)
#It is the total number of  successrs  in a series of  m Bernoulli events.
#####################################################
n<- 10000 # sample size
m<-10
p<-0.7
x1<-rbinom(n,m,p)   # data
par(mfrow=c(1,1))
plot(table(x1)/n, col=3, xlim=c(0,m), main = " BINOMIAL(m,p)")
lines(dbinom(0:m, m,p)~c(0:m), type='p', col=2, lwd=2)
```

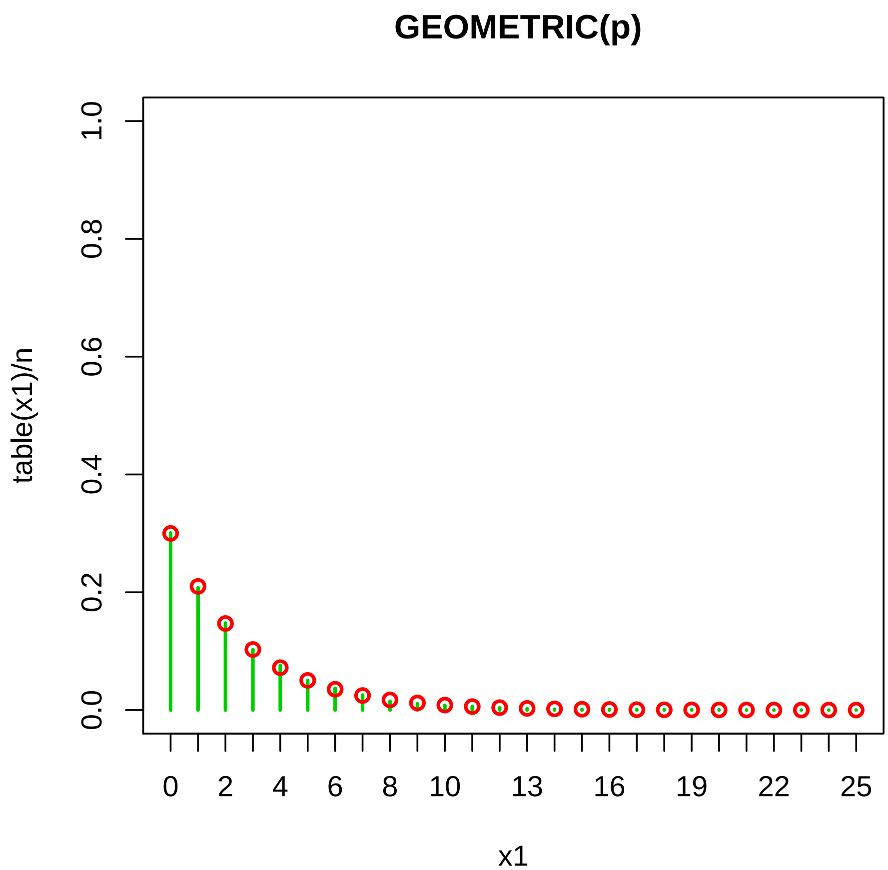



```
# Binomoal from Uniform(0,1)
x2 <- array(0,dim=c(n))
for (i in 1:n) {
  x2[i] <- sum(runif(m)<p)
}
par(mfrow=c(2,1))
barplot(table(x1)/n, col=3, ylim=c(0,1), xlim=c(0,m), xlab='X1', main = " BINOMIAL(m,p)")
barplot(table(x2)/n, col=2, ylim=c(0,1), xlim=c(0,m),xlab='X2', main = " BINOMIAL(m,p)")
```

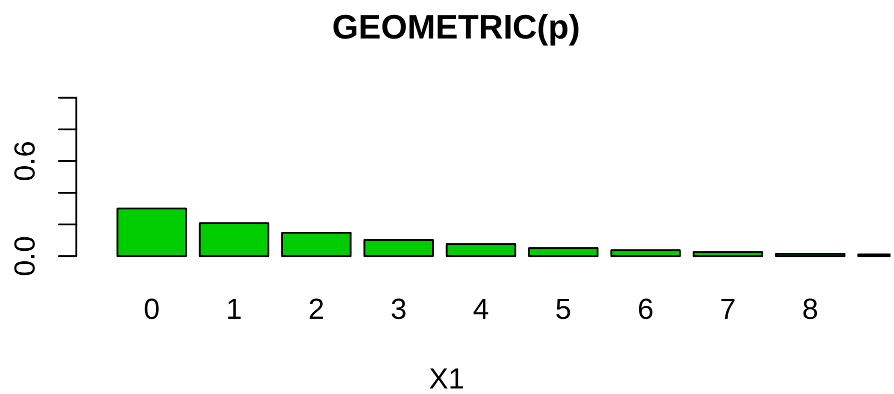


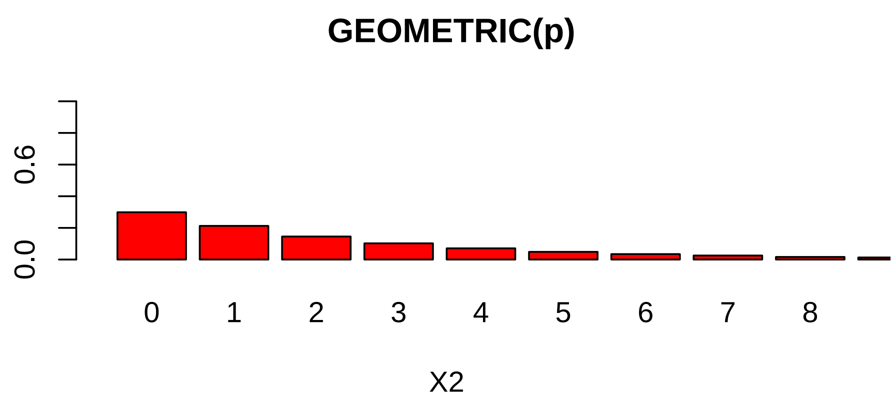



```
####################################################
#   GEOMETRIC(p)
#It is the number of failures  before a success in a series of Bernoulli event.
####################################################
n<- 10000 # sample size
p<-0.3
x1<-rgeom(n,p)   # data
par(mfrow=c(1,1))
plot(table(x1)/n, col=3, ylim=c(0,1), main = " GEOMETRIC(p)")
lines(dgeom(0:max(x1),p)~c(0:max(x1)), type='p', col=2, lwd=2)
```

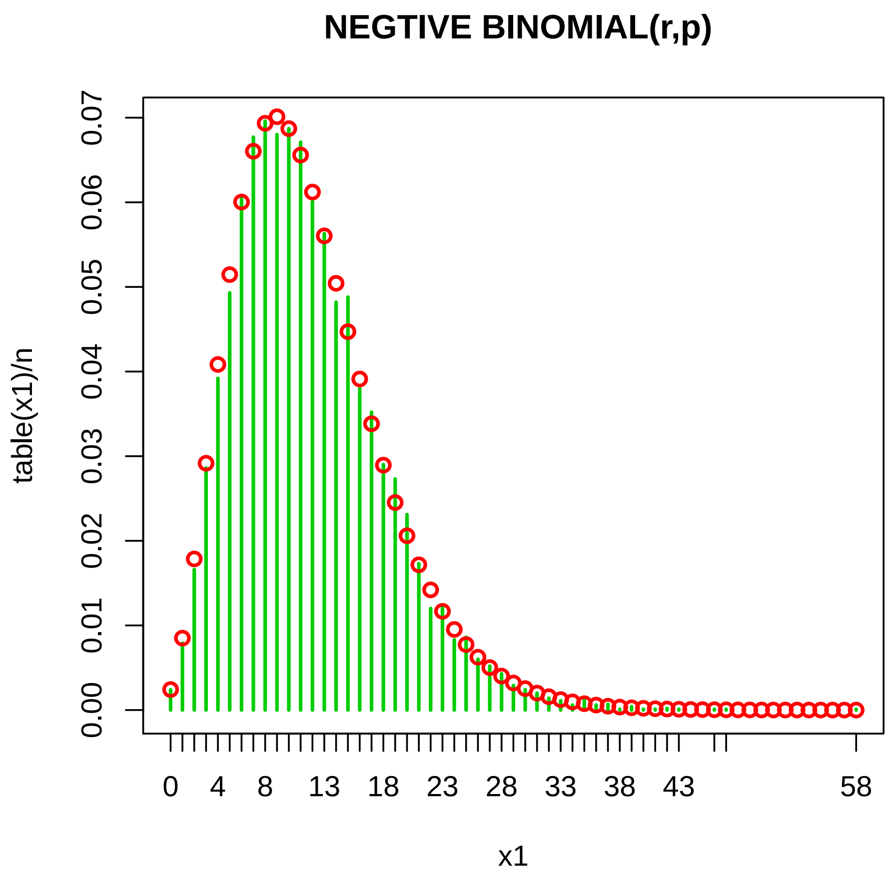



```
# Geometric from Uniform(0,1)
x2 <- array(0,dim=c(n))
for(i in 1 :n){
count<-0
s<-0
while (s==0) {

  count=count+1
  s<-(runif(1)<p)
}
x2[i]<-count-1
}
par(mfrow=c(2,1))
barplot(table(x1)/n, col=3, ylim=c(0,1), xlim=c(0,m), xlab='X1', main = "  GEOMETRIC(p)")
barplot(table(x2)/n, col=2, ylim=c(0,1), xlim=c(0,m), xlab='X2',main = " GEOMETRIC(p)")
```

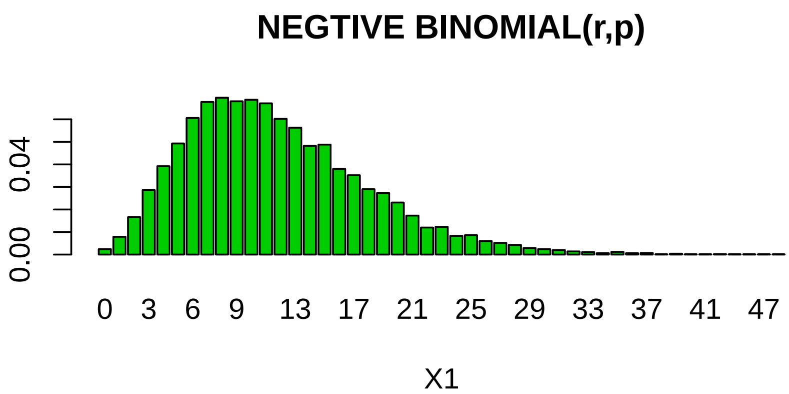


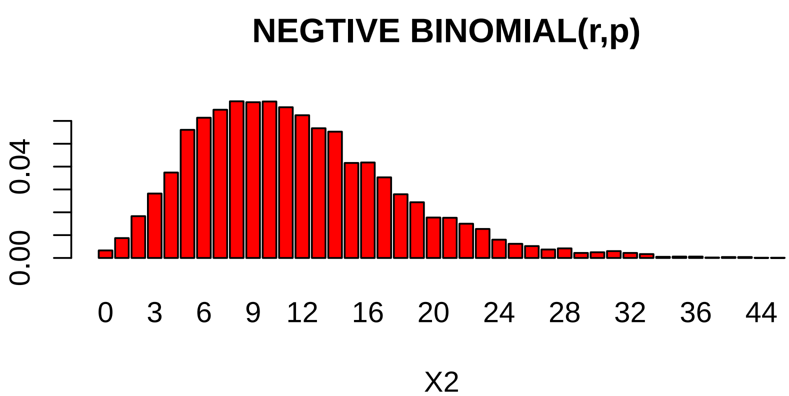



```
####################################################
#NEGTIVE BINOMIAL(r,p)
#It is the number of  failures  before a success in a series of Bernoulli events.
####################################################
n<- 10000 # sample size
r<-5
p<-0.3
x1<-rnbinom(n,r,p)   # data
par(mfrow=c(1,1))
plot(table(x1)/n, col=3, main = " NEGTIVE BINOMIAL(r,p)")
lines(dnbinom(0:max(x1),r,p)~c(0:max(x1)), type='p', col=2, lwd=2)
```

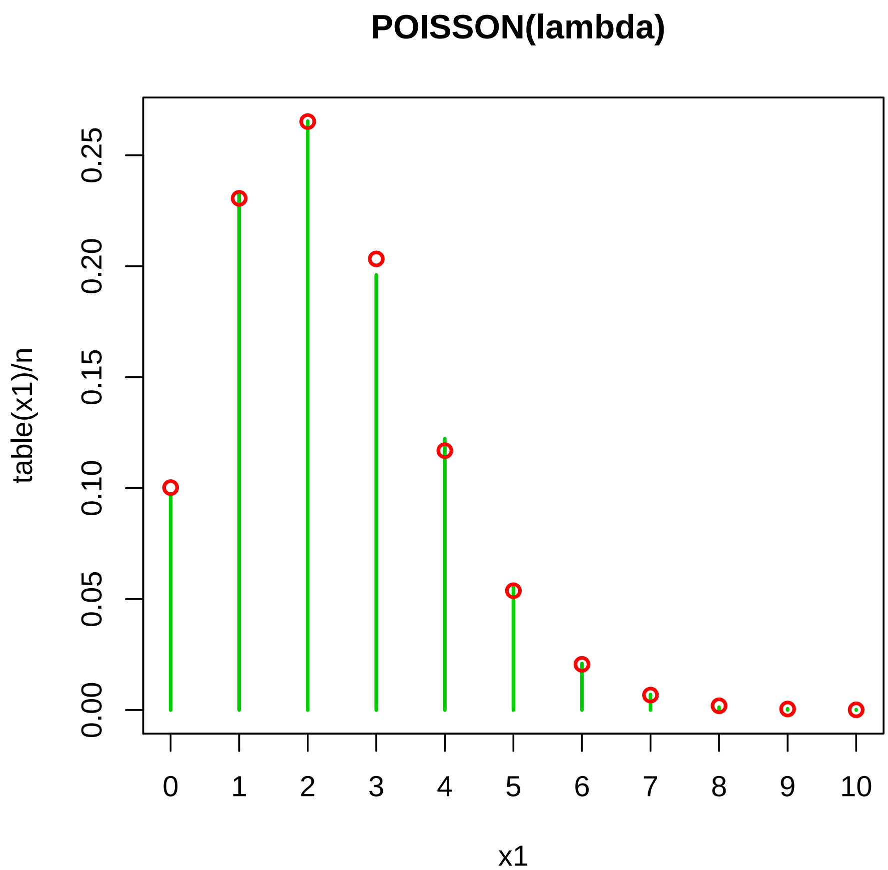



```
# NEGTIVE BINOMIAL(r,p) as a sum of r independent  GEOMETRIC(p)

x2 <- array(0,dim=c(n))
for(i in 1 : n){
  x2[i]<-sum(rgeom(r,p))
}

par(mfrow=c(2,1))
barplot(table(x1)/n, col=3, xlab='X1', main = "  NEGTIVE BINOMIAL(r,p)")
barplot(table(x2)/n, col=2, xlab='X2',  main = " NEGTIVE BINOMIAL(r,p)")
```

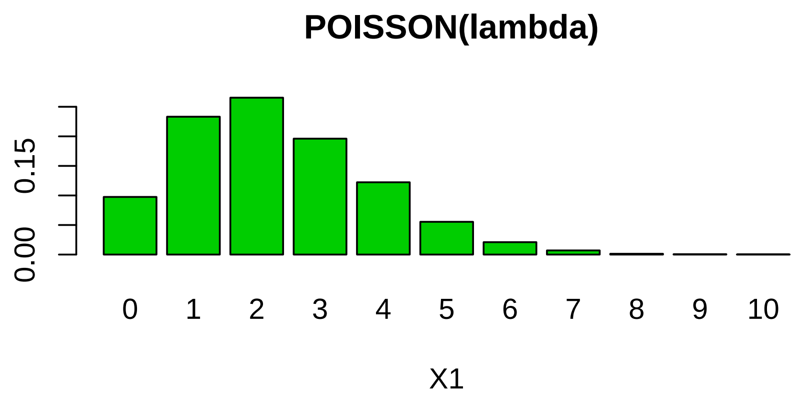


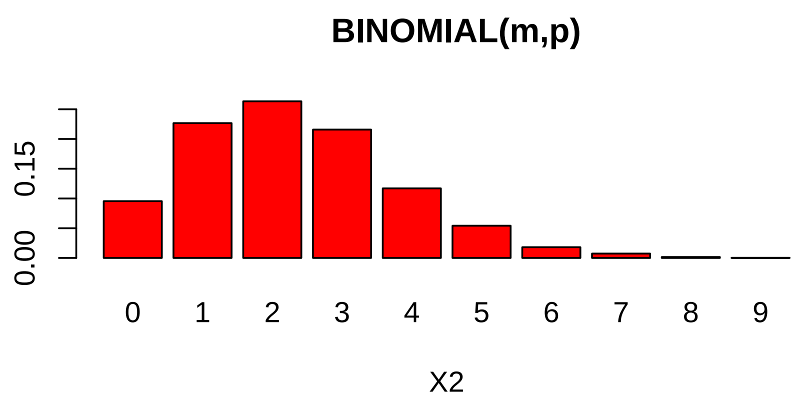



```
#####################################################
#POISSON(lambda)
# Binomal(m,p), m is large and p is small sych thet p*m converges to lambda
#####################################################
n<- 10000 # sample size
lambda<-2.3
x1<-rpois(n, 2.3)   # data
par(mfrow=c(1,1))
plot(table(x1)/n, col=3, main = " POISSON(lambda)")
lines(dpois(0:max(x1),lambda)~c(0:max(x1)), type='p', col=2, lwd=2)
```

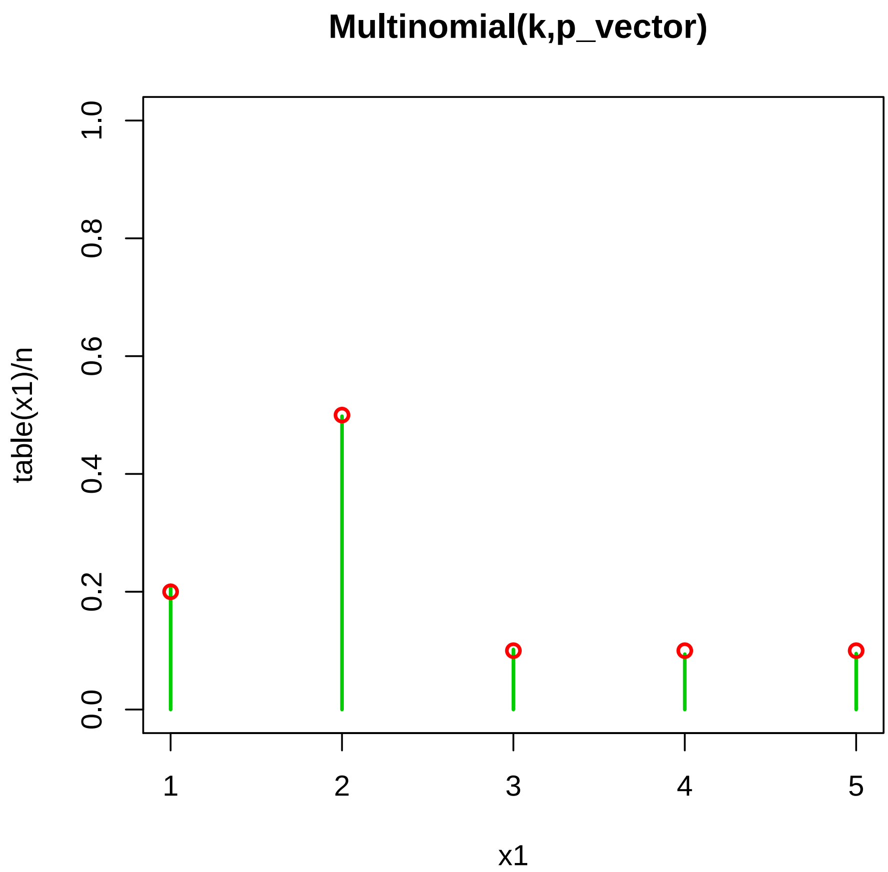



```r
# Poisson as  a limit of binomial
n<- 10000 # sample size
m<-100
p<-.023
x2<-rbinom(n,m,p)   # data
par(mfrow=c(2,1))
barplot(table(x1)/n, col=3, xlab='X1', main = " POISSON(lambda)")
barplot(table(x2)/n, col=2, xlab='X2',  main = "  BINOMIAL(m,p)")
```

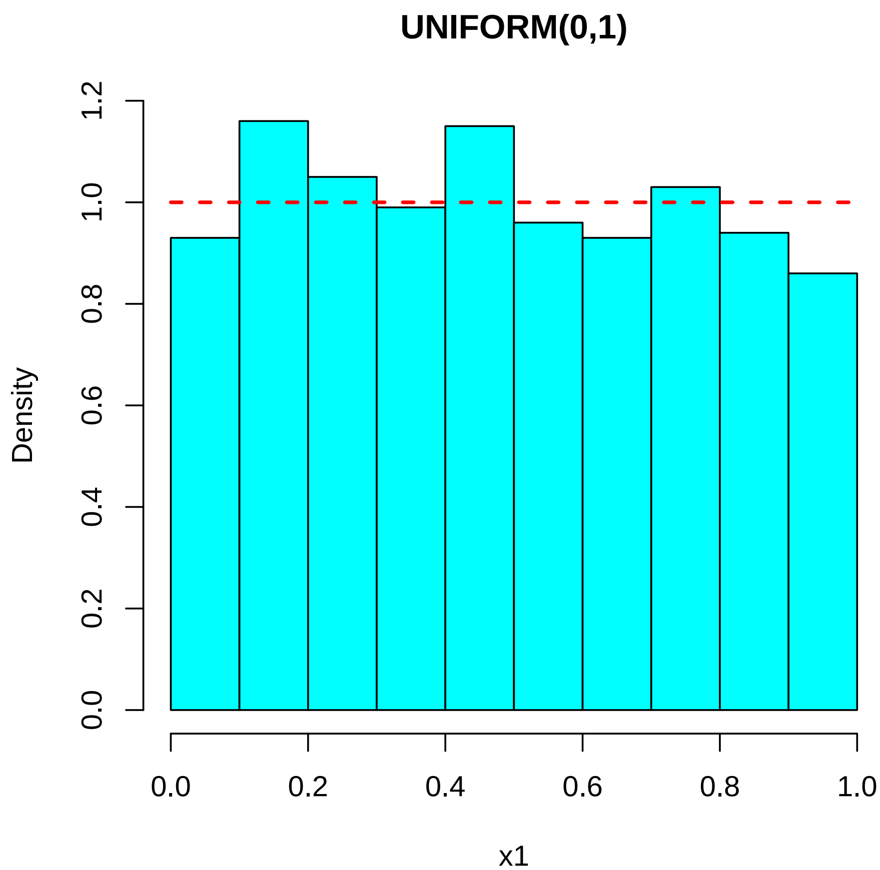




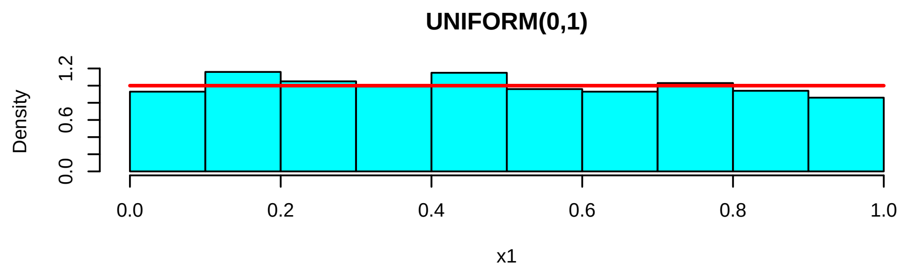

```
cat("\newpage")
##
## ewpage
##############################
#Multinomial(k,p_vector)
##############################
k <- 5  # categories
n <- 1000 # sample size
p <- c(.2,.5,.1,.1,.1) # probbility vector
x1 <- sample(1:k, n, replace=T, prob=p) # data
print(table(x1)/n )
## x1
##     1     2     3     4     5
## 0.211 0.498 0.102 0.094 0.095
par(mfrow=c(1,1))
plot(table(x1)/n,ylim=c(0,1), col=3, main = " Multinomial(k,p_vector)")
lines(p, type='p', col=2, lwd=2)
```

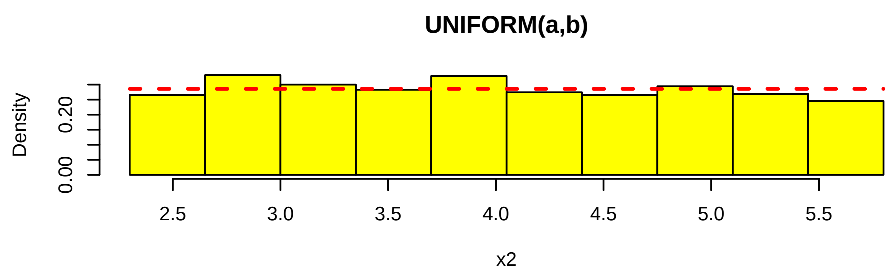


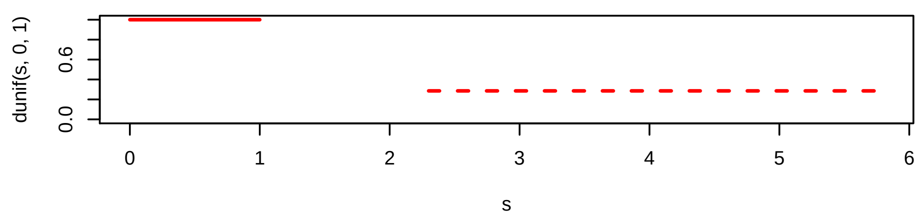

```
#########################################################
#COUNTINUOUS DISTRIBUTIONS
#########################################################

##############################
#UNIFORM(a,b)
##############################
a<-0
b<-1
n<- 1000 #Sample size
x1<-runif(n, a, b)
s<-seq(a,b,length=11)
par(mfrow=c(1,1))
hist(x1, col=5,probability = T, breaks =s, main='UNIFORM(0,1)')
lines(dunif(s,a,b)~s, col=2, lwd=2, lty=2)
```



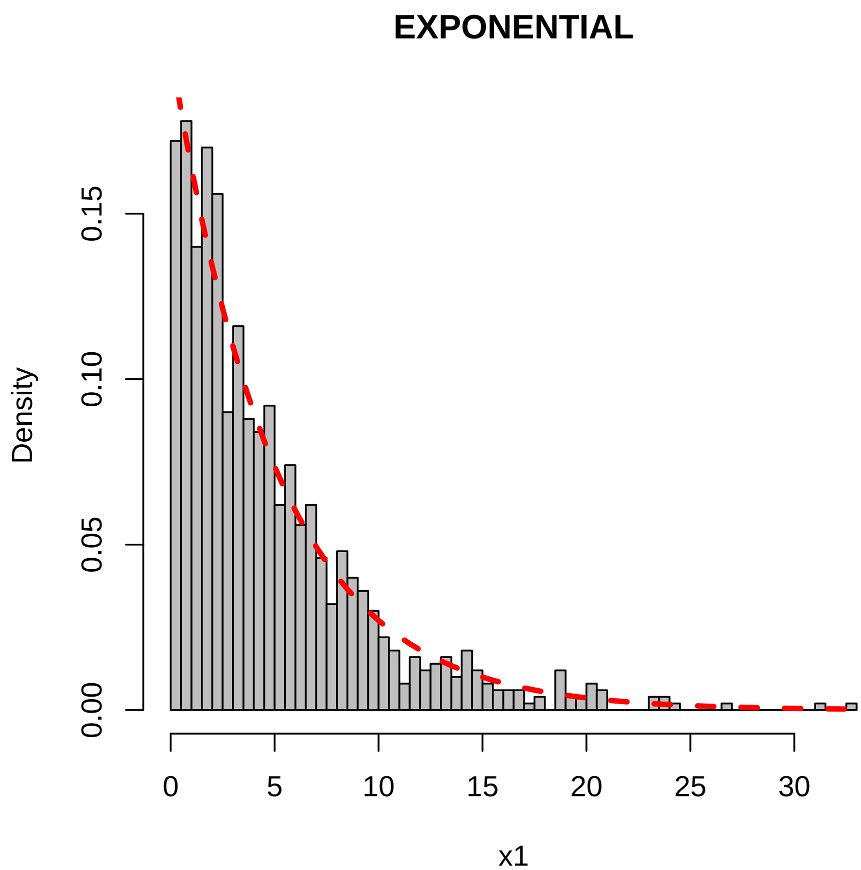

```
a<-2.3
b<-5.8

x2<-a+(b-a)*x1
ss<-a+s*(b-a)


par(mfrow=c(3,1))
hist(x1, col=5,probability = T, breaks =s, main='UNIFORM(0,1)')
lines(dunif(s,0,1)~s, col=2, lwd=2, lty=1)

hist(x2, col=7,probability = T, breaks =ss, main='UNIFORM(a,b)')
lines(dunif(ss,a,b)~ss, col=2, lwd=2, lty=2)

plot(dunif(s,0,1)~s, xlim=c(0,b),ylim=c(0,1), col=2, lwd=2, type='l')
lines(dunif(ss,a,b)~ss, col=2, lwd=2, lty=2)
```



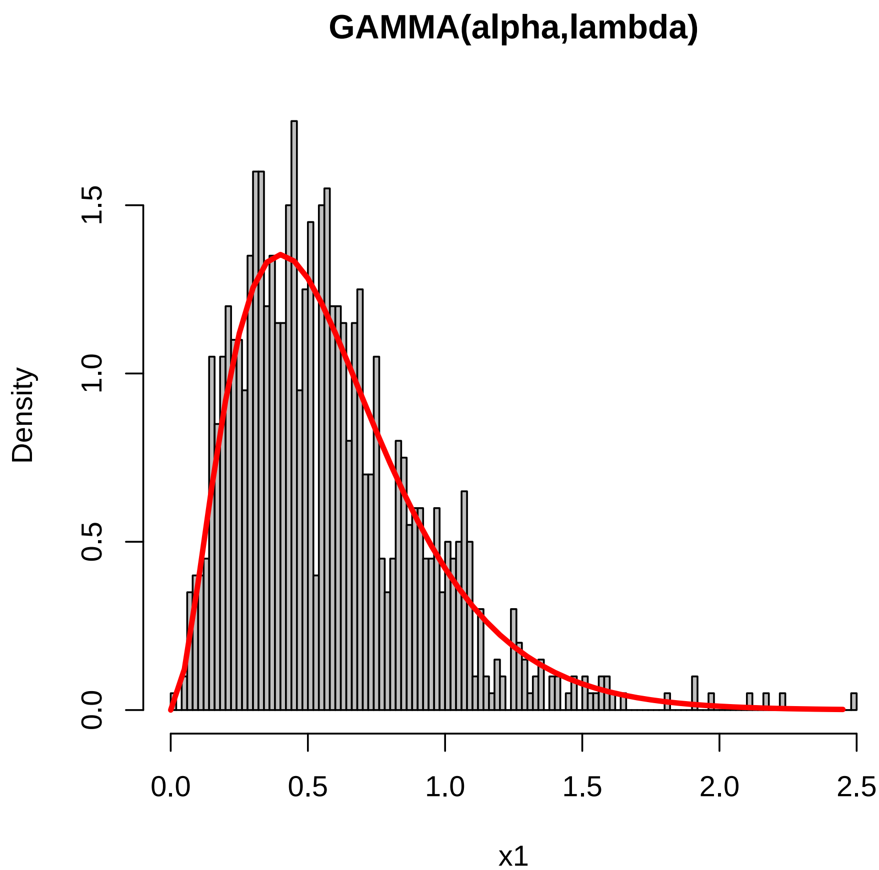




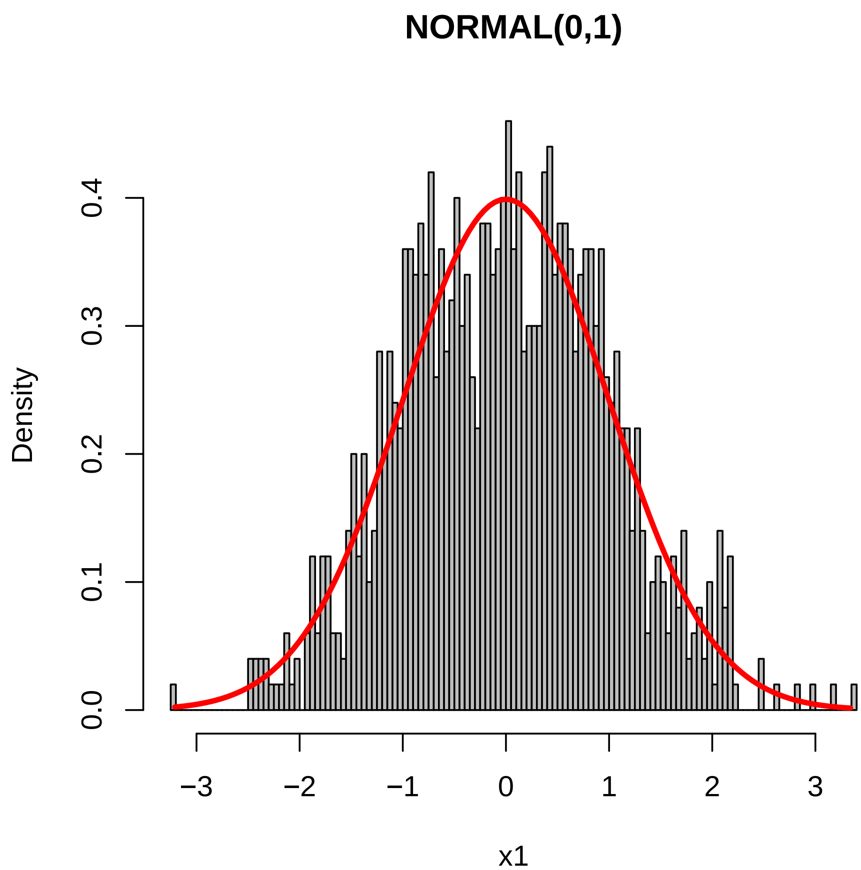

```
############################
#EXPONENTIAL (LAMBDA)
############################
lambda<-.2 #rate
mu=1
n<-1000 #sample size
x1 <- rexp(n, lambda)
s<-seq(0,max(x1),by=0.5)
par(mfrow=c(1,1))
hist(x1,probability = T,breaks = 100, col=8, main='EXPONENTIAL')
lines(dexp(s,lambda)~s, col=2, lwd=3, lty=2)
```



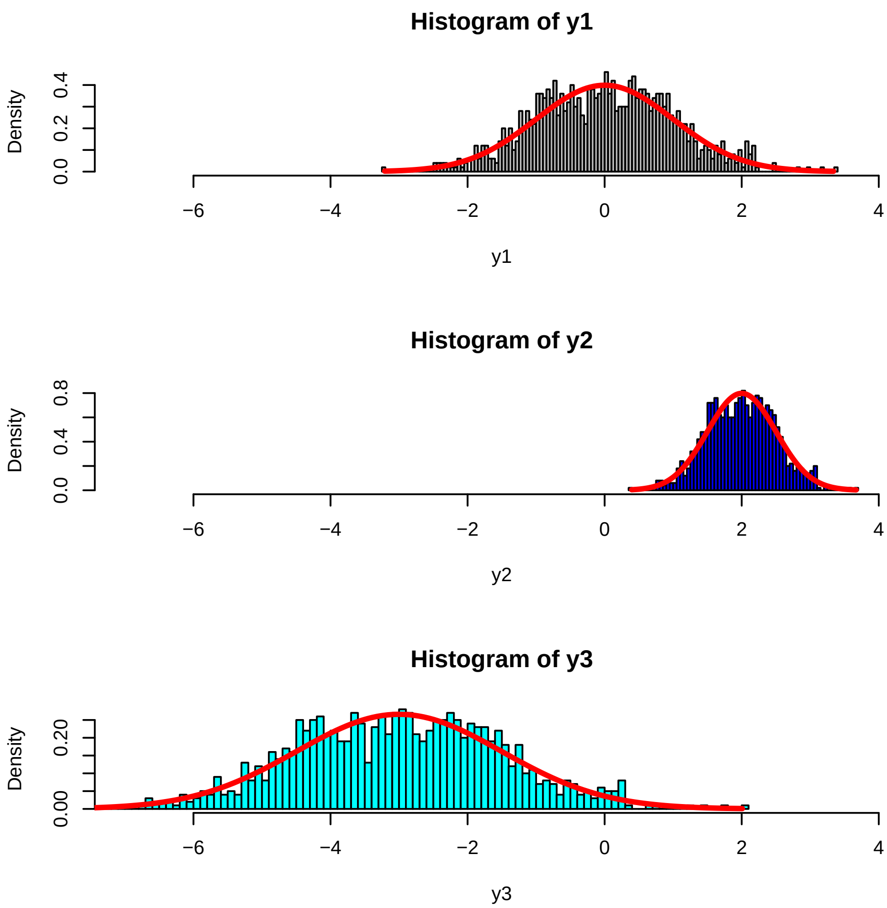

```
###############################
#GAMMA(alpha,lambda)
################################
alpha<-3 # shape
lambda<-.2 # rate

n<-1000 #sample size
x1 <- rgamma(n, shape=alpha, scale=lambda)
s<-seq(0,max(x1),by=0.05)
par(mfrow=c(1,1))
hist(x1,probability = T,breaks = 100, col=8, main='GAMMA(alpha,lambda)')
lines(dgamma(s, shape=alpha, scale=lambda)~s, col=2, lwd=3)
```



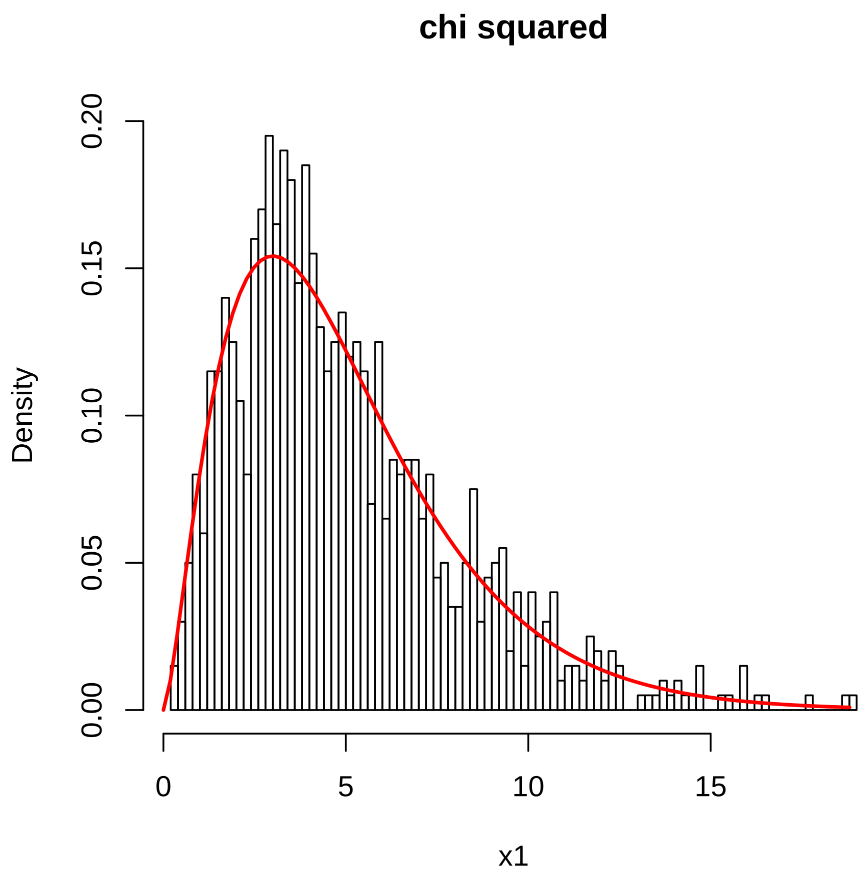

```
################################
# NORMAL(mu, sigma)
################################

mu=0  # location
sigma=1 # scale

n<-1000
x1<-rnorm(n, mu,sigma)
s<-seq(min(x1),max(x1),by=0.05)

par(mfrow=c(1,1))
hist(x1,probability = T,breaks = 100, col=8, main='NORMAL(0,1)')
lines(dnorm(s, mean=0, sd=1)~s, col=2, lwd=3)
```



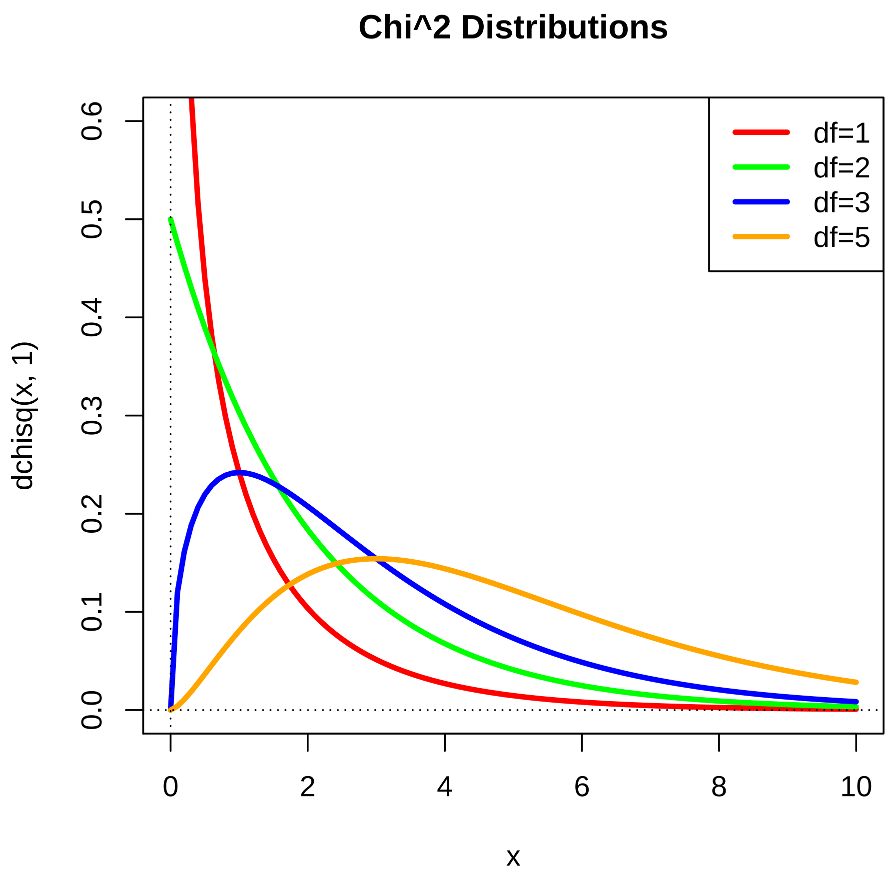

```
y1<-x1
s1<-s
y2<-2+0.5*x1
s2<-2+0.5*s
y3<- -3+1.5*x1
s3<- -3+1.5*s

par(mfrow=c(3,1))
hist(y1,probability = T,breaks = 100, col=8, xlim=c(-7,4))
lines(dnorm(s1, mean=0, sd=1)~s1, col=2, lwd=3)

hist(y2,probability = T,breaks = 100, col=4, xlim=c(-7,4))
lines(dnorm(s2, mean=2, sd=0.5)~s2, col=2, lwd=3)
hist(y3,probability = T,breaks = 100, col=5, xlim=c(-7,4))
lines(dnorm(s3, mean= -3, sd=1.5)~s3, col=2, lwd=3)
```



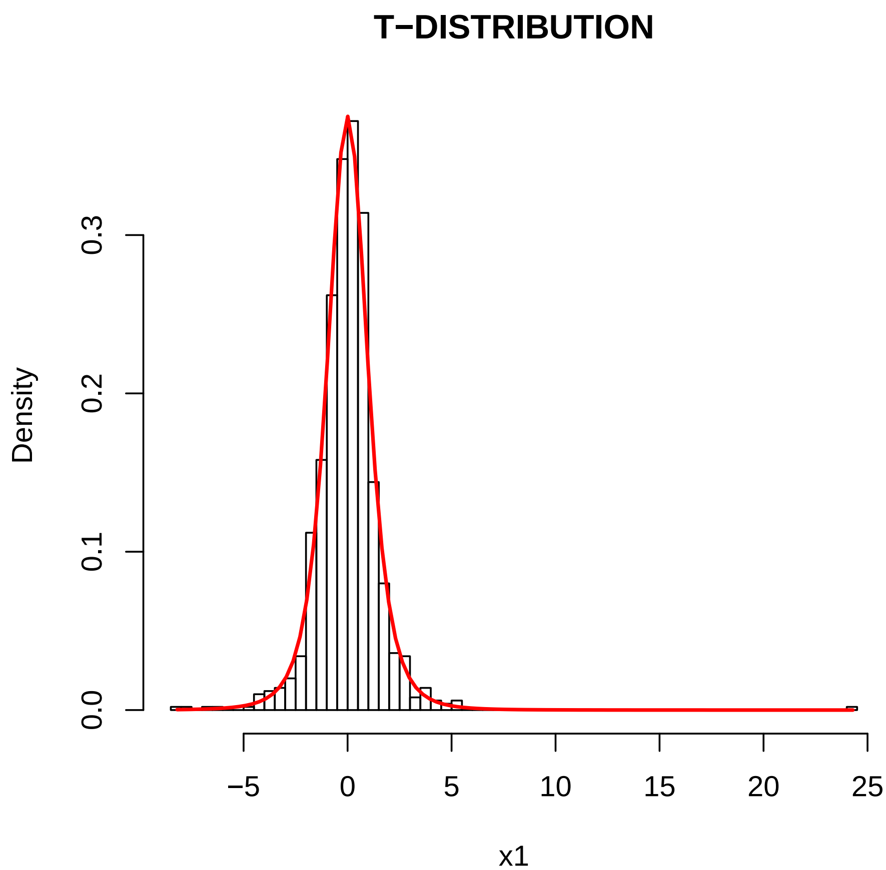

```
cat("mean(Y1)=", mean(y1), "sd(Y1)=",sd(y1),"\n")
## mean(Y1)= 0.005077315 sd(Y1)= 0.9932519
cat("mean(Y2)=", mean(y2), "sd(Y2)=",sd(y2),"\n")
## mean(Y2)= 2.002539 sd(Y2)= 0.4966259
cat("mean(Y3)=", mean(y3), "sd(Y2)=",sd(y3),"\n")
## mean(Y3)= -2.992384 sd(Y2)= 1.489878
############################
#Chi^2- distribution
############################
df<-5
x1<-rchisq(n = 1000,df=df)
par(mfrow=c(1,1))
s<-seq(0,max(x1),length=100)
hist(x1, probability = T, breaks=100,  ylim=c(0,.2), main="chi squared")
lines( dchisq(s,df)~s, col='red', lwd=2 )
```



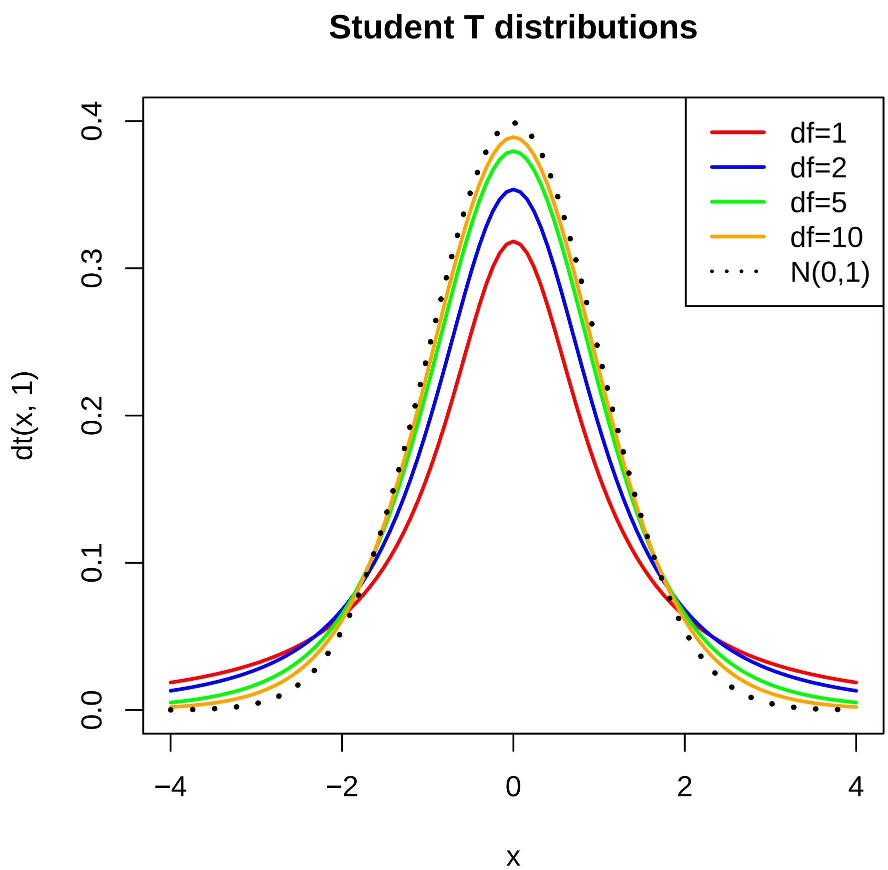

```
par(mfrow=c(1,1))
curve(dchisq(x,1), xlim=c(0,10), ylim=c(0,.6), col='red', lwd=3)
curve(dchisq(x,2), add=T, col='green', lwd=3)
curve(dchisq(x,3), add=T, col='blue', lwd=3)
curve(dchisq(x,5), add=T, col='orange', lwd=3)
abline(h=0,lty=3)
abline(v=0,lty=3)
legend(par('usr')[2], par('usr')[4], xjust=1,
       c('df=1', 'df=2', 'df=3', 'df=5'),
       lwd=3,
       lty=1,
       col=c('red', 'green', 'blue', 'orange')
)
title(main='Chi^2 Distributions')
```



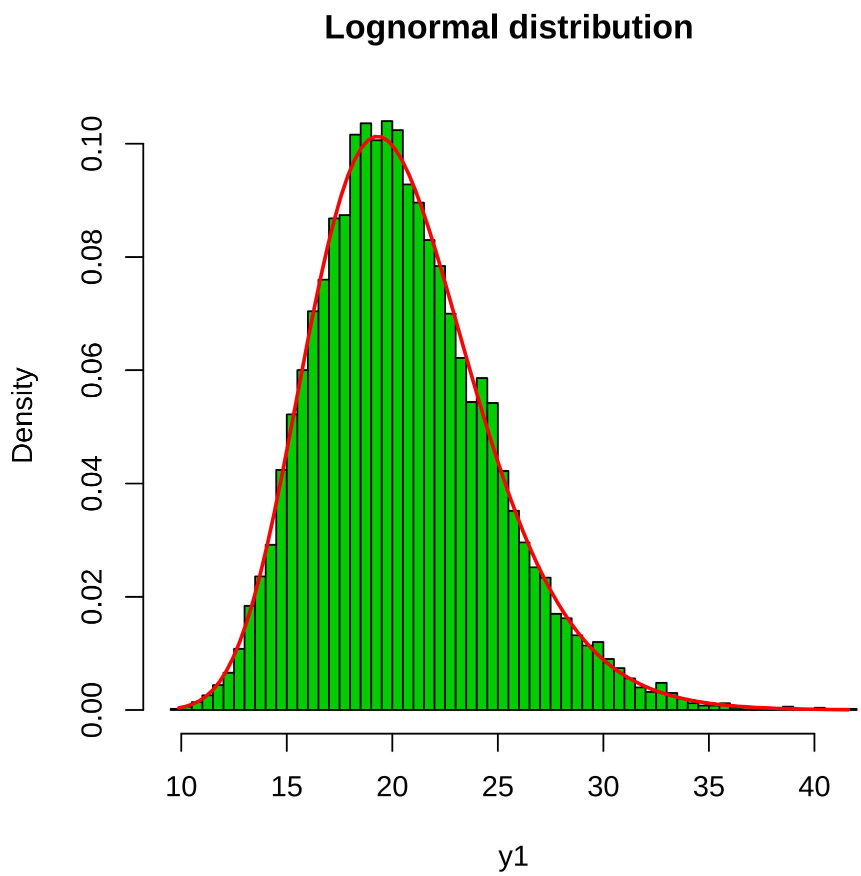

```
###################################
#T-DISTRIBUTION
###################################
df<-4
x1<-rt(n = 1000,df=df)
par(mfrow=c(1,1))
s<-seq(min(x1),max(x1),length=100)
hist(x1, probability = T, breaks=100, main="T-DISTRIBUTION" )
lines( dt(s,df )~s, col='red', lwd=2 )
```



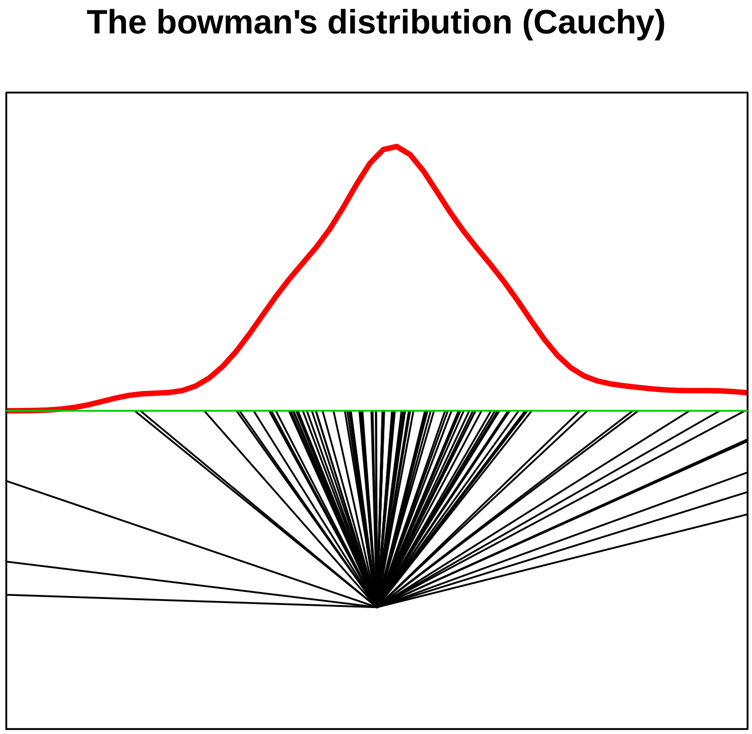

```
par(mfrow=c(1,1))
curve( dt(x,1), xlim=c(-4,4), ylim=c(0,.4), col='red', lwd=2 )
curve( dt(x,2), add=T, col='blue', lwd=2 )
curve( dt(x,5), add=T, col='green', lwd=2 )
curve( dt(x,10), add=T, col='orange', lwd=2 )
curve( dnorm(x), add=T, lwd=3, lty=3 )
title(main="Student T distributions")
legend(par('usr')[2], par('usr')[4], xjust=1,
       c('df=1', 'df=2', 'df=5', 'df=10', 'N(0,1)'),
       lwd=c(2,2,2,2,2),
       lty=c(1,1,1,1,3),
       col=c('red', 'blue', 'green', 'orange', par("fg")))
```




```
######################################
# Lognormal distribution
#######################################


mu=3
sigma=.2

n<-10000

x1<-rnorm(n, mean=mu,sd=sigma)
y1<-exp(x1) # rlnorm(n,mu, sigma)
s<-seq(min(y1),max(y1), length=100)

hist(y1,breaks = 100, probability = T, col=3, main = "Lognormal distribution ")
lines(dlnorm(s,mu,sigma)~s, col=2, lwd=2)
```




```
######################################
#Cauchy distribution
#######################################

n <- 100                              # sample size
alpha <- runif(n, -pi/2, pi/2)    # Direction of the arrow
x <- tan(alpha)                     # Arrow impact
plot.new()
plot.window(xlim=c(-5, 5), ylim=c(-1.5, 1.5))
segments( 0, -1,      # Position of the Bowman
          x, 0   )    # Impact
d <- density(x)
lines(d$x, 5*d$y, col="red", lwd=3 )
box()
abline(h=0, col=3)
title(main="The bowman's distribution (Cauchy)")
```


The bowman's distribution (Cauchy)

8. Joint and conditional distributions

Definition 101. Vector Valued Random variable: $(X_1(\omega), X_2(\omega), \cdots, X_k(\omega))$ is a vector valued random variable where $\omega \in \Omega$.

Definition 102. Let (X, Y) be a pair of random variable with joint distribution function F on same probability space$(\Omega, \mathcal{A}, P)$ then

$$\begin{aligned} F(x,y) &= P(X \le x, Y \le y) \\ &= P(\{\omega | \omega \in \Omega, X(\omega) \in (-\infty, x], Y(\omega) \in (-\infty, y]\}) \end{aligned} \tag{8.1}$$

Properties:
(a) $\lim_{x \downarrow -\infty, y \downarrow -\infty} F(x,y) = 0$
(b) $\lim_{x \uparrow \infty, y \uparrow \infty} F(x,y) = 1$
(c) $\lim_{y \uparrow \infty} F(x,y) = F_X(x)$ [Marginal distribution of X]
(d) $\lim_{x \uparrow \infty} F(x,y) = F_Y(y)$ [Marginal distribution of Y]
(e) $F(a < X \le b, c < Y \le d) = F(b,d) - F(a,d) - F(b,c) + F(a,c) \ge 0$ [Non decreasing]
(f) $F(a < X \le b, c < Y \le d) = (F_X(b) - F_X(a))(F_Y(d) - F_Y(c))$ iff X and Y are independent.
(g) Joint p.d.f of (X, Y) is $f(x,y) = \frac{\partial^2}{\partial x \partial y} F(x,y)$ such that $\int_x \int_y f(u,v) du dv = 1$
(h) Marginal densities: $\int_x f(u,y) du = f_Y(y)$ and $\int_y f(x,v) dv = f_X(x)$.
(i) $f(x,y) = f_X(x) f_Y(y)$ iff X and Y are independent.
(j) Conditional density of $Y|X = x$ is $f_{Y|x}(y|x) = \frac{f(x,y)}{f_X(x)}$ if $f_X(x) > 0$
(k) Conditional exception (regression) of $g(Y)$ given $X = x$ i.e. $E(g(Y)|X = x) = \int_{y|x} g(y) f_{Y|x}(y|x) dy$ is a function of x.

Definition 103. A non negative function $P(X = x, Y = y) = f(x,y)$ is said to be the the joint p.m.f of (X, Y) if they have the following properties
(a) $\sum_x \sum_y f(x,y) = 1$
(b) $\sum_y f(x,y) = f_X(x)$ i.e. marginal p.m.f. of X.
(c) $\sum_x f(x,y) = f_Y(y)$ i.e. marginal p.m.f. of Y.

8.1. **Laws of expectation.** Assume that (X, Y) has joint p.m.f. $f(x,y)$ with $\sum_x |x|^2 f_X(x) < \infty$ and $\sum_y |y|^2 f_Y(y) < \infty$ then
(1) $E(\alpha X) = \alpha E(X)$
(2) $E(X + Y) = E(X) + E(Y)$ NOTE: Independence NOT required.
(3) $E(XY) = E(X)E(Y)$ if$X$ and Y are independent but the reverse is not true in general.
(4) **Product moment** $Cov(X,Y) = E(X - \mu_x)(Y - \mu_y) = E(XY) - \mu_x \mu_y$.
(5)$Cov(X,X) = Var(X) = \sigma_x^2$
(6)$Var(aX) = a^2 Var(X)$
(7)$Var(aX + bY) = a^2 \ Var(X) + b^2 \ Var(Y) + 2ab \ \ Cov(X,Y)$
(8)$E(Y) = E_X E_{Y|X}(Y|X = x)$
(9)$Var(Y) = E_X V_{Y|X}(Y|X = x) + V_X E_{Y|X}(Y|X = x)$
(10) If $P(X \ge Y) = 1$ then $E(X) \ge E(Y)$
(11) If X has MGF $M_X(t)$ then $Y = a + bX$ has MGF $M_Y(t) = e^{at} M_X(bt)$
(12) If two independent random variables X has MGF $M_X(t)$ and Y has MGF $M_Y(t)$ then $Z = X + Y$ has MGF $M_Z(t) = M_X(t) M_Y(t)$

Definition 104. $Corr(X,Y) = \rho(X,Y) = \frac{E(XY)-E(X)E(Y)}{\sqrt{(Var(X)Var(Y)}}$

Remark 105. Correlation can measure **linear dependency** between two random variables.

Exercise 106. Show that $|\rho(X,Y)| \leq 1$.

Exercise 107. Let $X_1 \sim bin(n_1, p)$ and $X_2 \sim bin(n_2, p)$ be independently distributed. Find the conditional distribution of X_1 when it is given $X_1 + X_2 = k$.

Exercise 108. Let $X_1 \sim pois(\lambda_1)$ and $X_2 \sim pois(\lambda_2)$ be independently distributed. Find the conditional distribution of X_1 when it is given $X_1 + X_2 = k$.

Exercise 109. Let $P \sim U[0,1]$ and $Y|P = p \sim bin(n,p)$. Find the marginal distribution of Y.

Exercise 110. Let $\Lambda \sim exponential(1)$ and $Y|\Lambda = \lambda \sim pois(\lambda)$. Find the marginal distribution of Y.

Exercise 111. Let $P \sim B(a,b)$ and $Y|P = p \sim bin(n,p)$. Find the marginal distribution of Y.

Definition 112. If (X,Y) has joint density function $f(x,y)$ as

$$f(x,y) = \frac{e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2+\left(\frac{y-\mu_y}{\sigma_y}\right)^2-2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right)\right]}}{2\pi\sigma_x,\sigma_y\sqrt{1-\rho^2}}$$

then (X,Y) is said to follow Bivariate normal$(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$

Exercise 113. If (X,Y) follow Bivariate normal$(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$ the show that $Y|X = x$ follows $N(\mu_y + \rho\frac{\sigma_y}{\sigma_x}(x-\mu_x), (1-\rho^2)\sigma_y^2)$

Theorem 114. *Let (X,Y) be continuous random variables with density function $f(x,y)$ then the density function of $(U,V) = (U(X,Y), V(X,Y))$ is*

$$g(u,v) = f(x(u,v), y(u,v)) \left\|\frac{\partial(x,y)}{\partial(u,v)}\right\|$$

where $\left\|\frac{\partial(x,y)}{\partial(u,v)}\right\|$ *stands for the absolute value of the determinant of the Jacobian matrix*$\frac{\partial(x,y)}{\partial(u,v)}$.

Exercise 115. Let X and Y be i.i.d. $N(0,1)$ random variables. Find the joint density of $U = X+Y$ and $V = X-Y$. Are (U,V) independently distributed?

Exercise 116. Let X and Y be i.i.d. $N(0,1)$ random variables. Find the p.d.f. of (r,θ) where $X = r\cos\theta$ and $Y = r\sin\theta$.

Exercise 117. Give an algorithm to generate X,Y which are i.i.d. $N(0,1)$ from U,V which are i.i.d. $U(0,1)$.

Exercise 118. Let X and Y be i.i.d. $N(0,1)$ random variables. Find the p.d.f. of $U = X/Y$.

Exercise 119. Show that t_1 and $C(0,1)$ are the same distribution. [Alternative statement: Let X and Y be i.i.d. $N(0,1)$ random variables. Then $U = X/|Y|$ has $C(0,1)$ distribution.]

Exercise 120. Let $X_1, X_2, \ldots, X_n$ be a random sample from a $\mathrm{N}(\mu, \sigma^2)$ population. Then show that

(1) $\overline{X}$ has a $\mathrm{N}(\mu, \sigma^2/n)$ distribution,
(2) $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ has a $\mathrm{N}(0,1)$ distribution,
(3) $\frac{(n-1)S^2}{\sigma^2}$ has a χ^2-distribution with $n-1$ degrees of freedom
(4) $\overline{X}$ and $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ are independent random variables,
(5) $\frac{\overline{X}-\mu}{S/\sqrt{n}}$ has a t-distribution with $n-1$ degrees of freedom

FIGURE 8.1. Bivariate normal joint p.d.f.

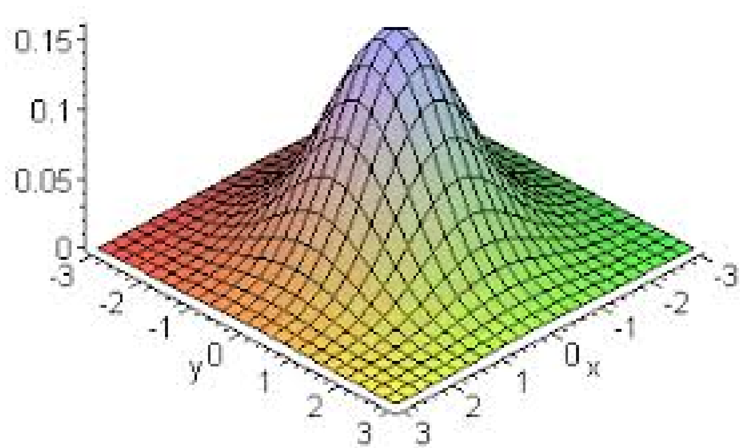
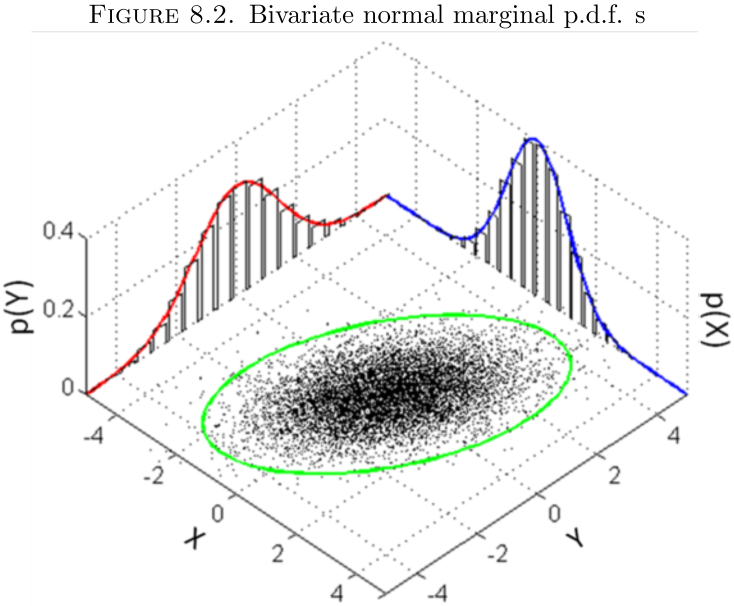
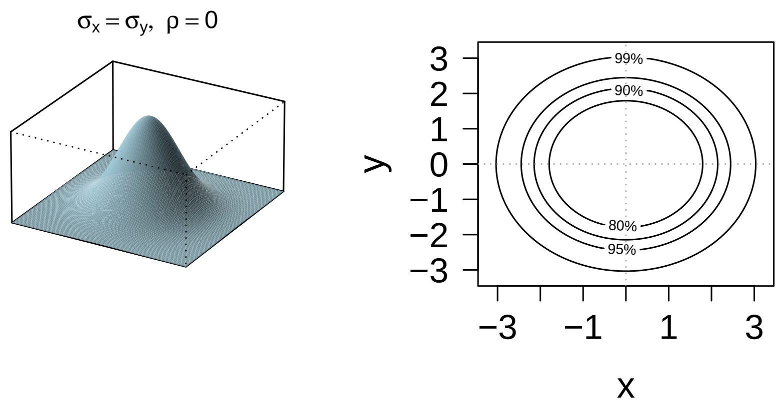
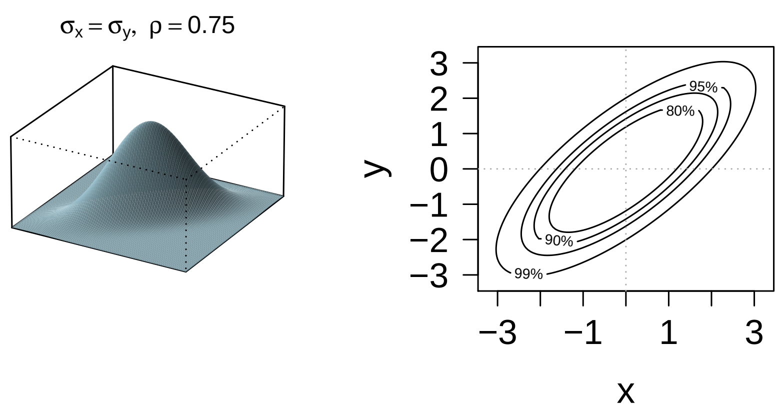
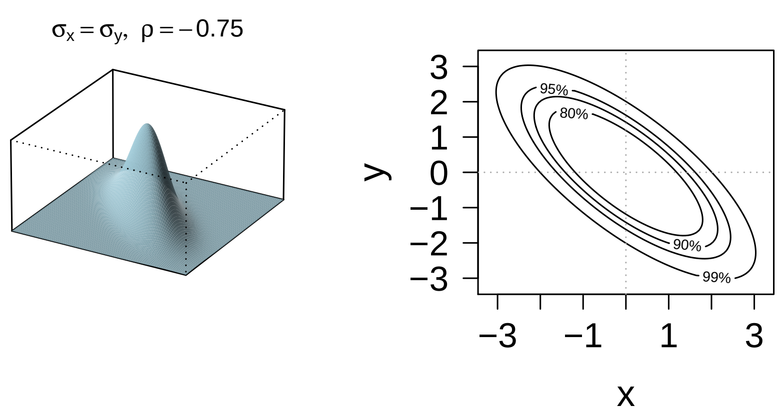


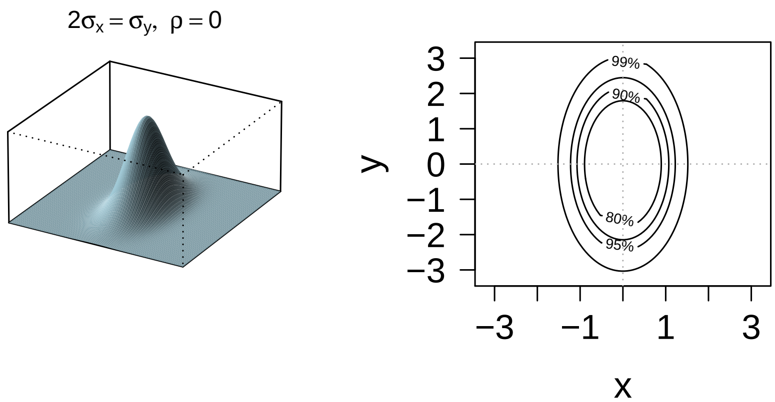
FIGURE 8.2. Bivariate normal marginal p.d.f. s

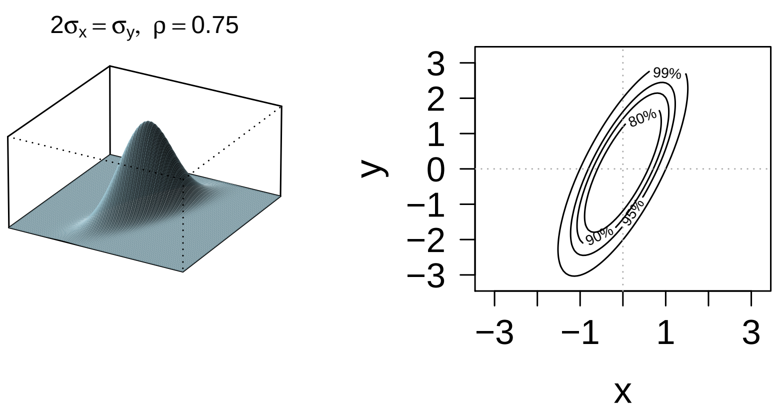



$\sigma_x = \sigma_y,\ \rho = 0$
99%
90%
80%
95%
y
x


$\sigma_x = \sigma_y,\ \rho = 0.75$
95%
80%
90%
99%
y
x


$\sigma_x = \sigma_y,\ \rho = -0.75$
95%
80%
90%
99%
y
x


$2\sigma_x = \sigma_y,\ \rho = 0$
99%
90%
80%
95%
y
x


$2\sigma_x = \sigma_y,\ \rho = 0.75$
99%
80%
90%
95%
y
x

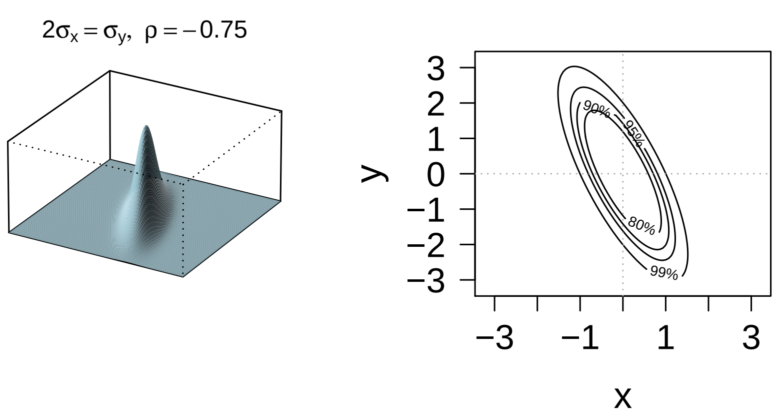

$2\sigma_x = \sigma_y,\ \rho = -0.75$
90%
95%
80%
99%
y
x

FIGURE 8.3. Conditional distribution

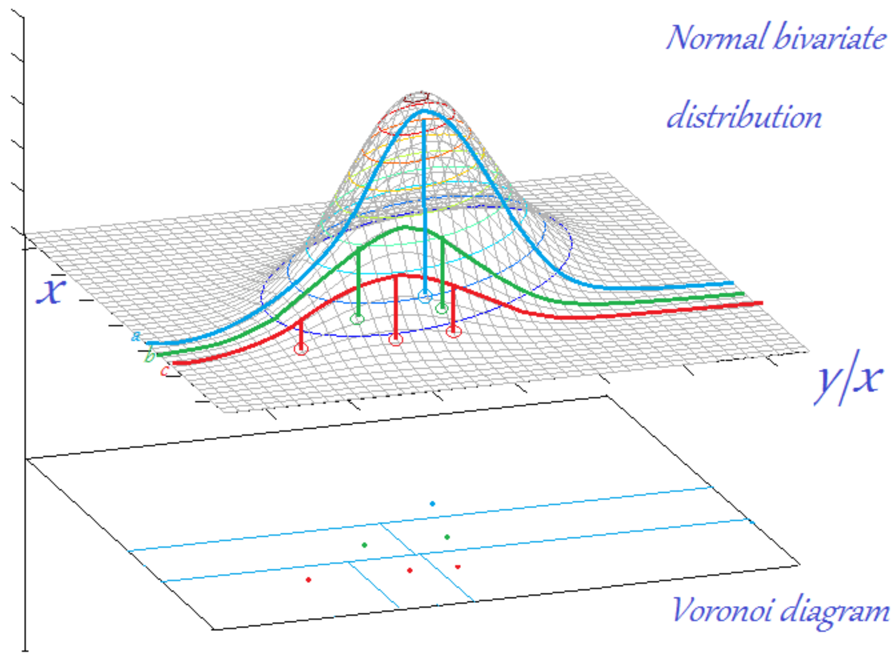
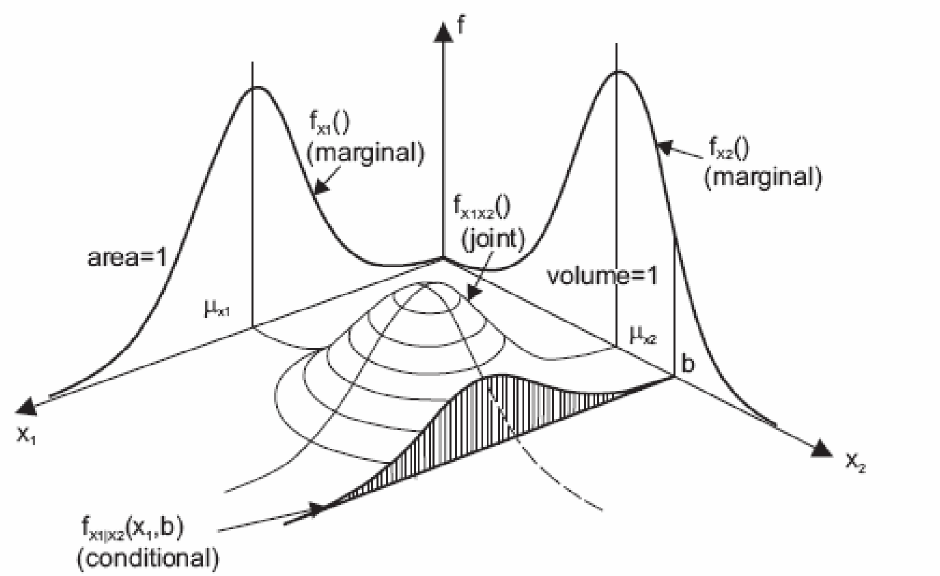
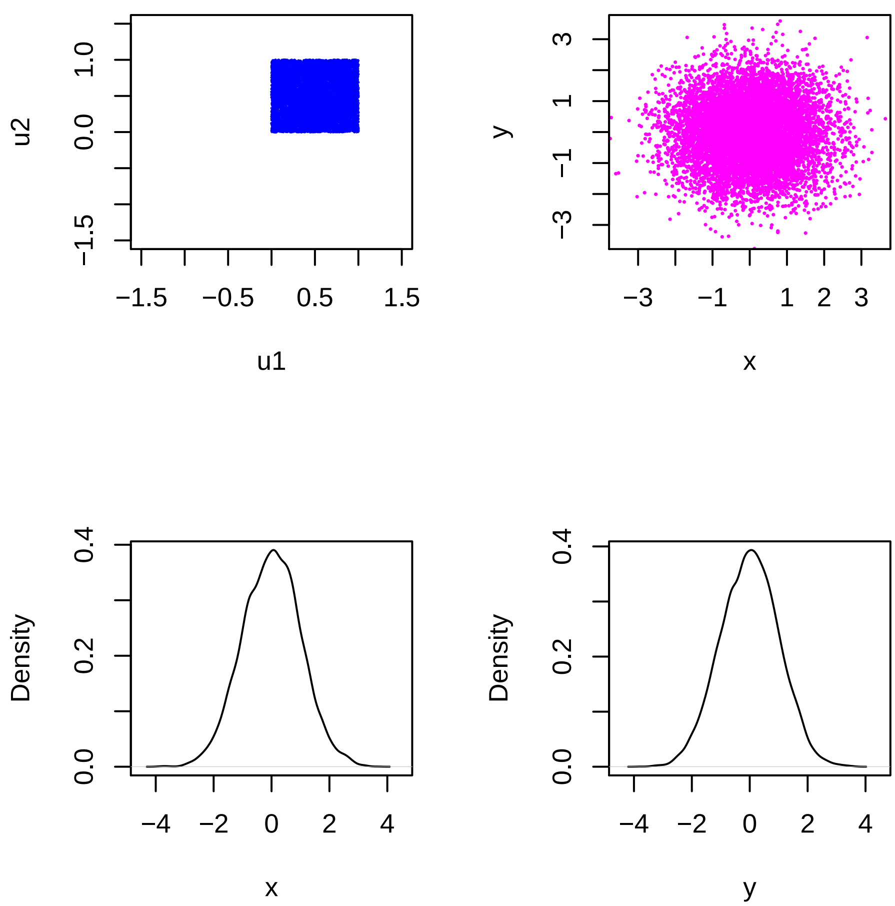


FIGURE 8.4. Regression



```
###################################
#Box-Muller method
###################################
set.seed(123)
n<-9000
u1<- runif(n,min = 0,max = 1)
u2<- runif(n,min = 0,max = 1)
x<-sqrt(-2*log(u1))*cos(2*pi*u2)
y<-sqrt(-2*log(u1))*sin(2*pi*u2)
par(mfrow=c(2,2))
plot(u2~u1,xlim=c(-1.5,1.5), ylim=c(-1.5,1.5), type="p", pch = 20, cex = 0.1, col=4)
plot(y~x, xlim=c(-3.5,3.5), ylim=c(-3.5,3.5), type="p",  pch = 20, cex = 0.2, col=6)
plot(density(x), xlim=c(-4.5,4.5),main = "", xlab = "x")
plot(density(y), xlim=c(-4.5,4.5),main = "", xlab= "y")
```



9. Law of Large Numbers

Definition 121. Convergence in probability: A sequence of random variables X_n $\forall n \in N$ on probability space $(\Omega, \mathcal{F}, P)$ is said to converge in probability to X on $(\Omega, \mathcal{F}, P)$ if

$$\lim_{n\uparrow\infty} P(\{\omega||X_n(\omega) - X(\omega)| < \epsilon\}) = 1 \forall\ \epsilon > 0$$

Theorem 122. ***Weak law of large number (WLLN):*** *Let $\{X_n\}$ be a sequence of i.i.d random variables with finite variance then*

$$\lim_{n\uparrow\infty} P(|\bar{X} - \mu_x| \leq \epsilon) = 1$$

where $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $\epsilon > 0$.

Definition 123. Convergence in distribution: A sequence of random variables X_n $\forall n \in N$ on probability space$(\Omega_n, \mathcal{F}_n, P_n)$ is said to converge in distribution to Y on $(\Omega, \mathcal{F}, P)$ if

$$\lim_{n\uparrow\infty} F_{X_n}(a) = F_Y(a)$$

for all such $a \in \mathbb{R}$, where $F_Y(a)$ is continuous.

Remark 124. $X_n \xrightarrow{p} X \implies X_n \xrightarrow{d} X$ **but the reverse implications are not true in general**. [Proof is not in syllabus]

Remark 125. If $X_n \xrightarrow{p} X$ then $h(X_n) \xrightarrow{p} h(X)$ for any continuous function h. [Proof is not in syllabus]

Theorem 126. ***Continuity theorem:*** *Let $\{X_n\}$ be a sequence of random variables with corresponding characteristic functions as $E(e^{itX_n}) = \phi_{X_n}(t)$ such that $\lim\limits_{n\uparrow\infty} \phi_{X_n}(t) = \phi_Y(t)$ for some random variable Y. Then*

$$\lim_{n\uparrow\infty} F_{X_n}(a) = F_Y(a)$$

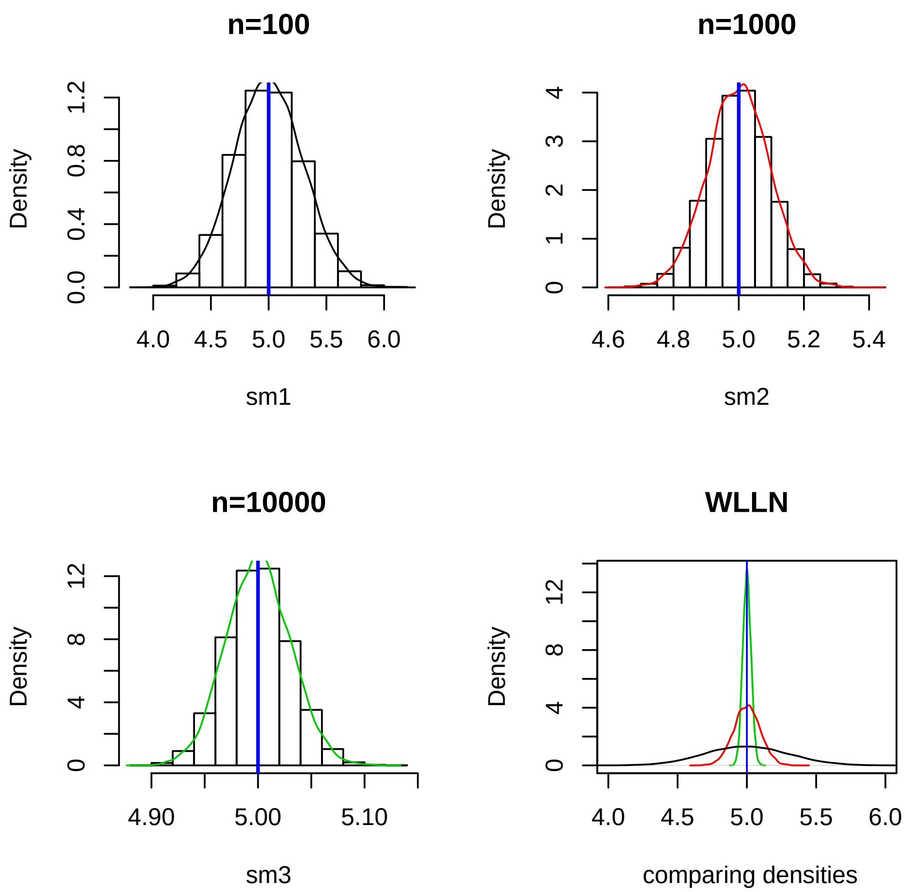
for all such $a \in \mathbb{R}$, where $F_Y(a)$ is continuous. We say X_n converges in distribution to Y. [We can have the above result with MGF, if it exists. Proof is not in syllabus]

Theorem 127. ***Central Limit Theorem(CLT)****: IfX_is be i.i.d. random variables with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2 < \infty$. Define $S_n = \sum_{i=1}^{n} X_i = n\bar{X}$ and $T_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}$. Then*

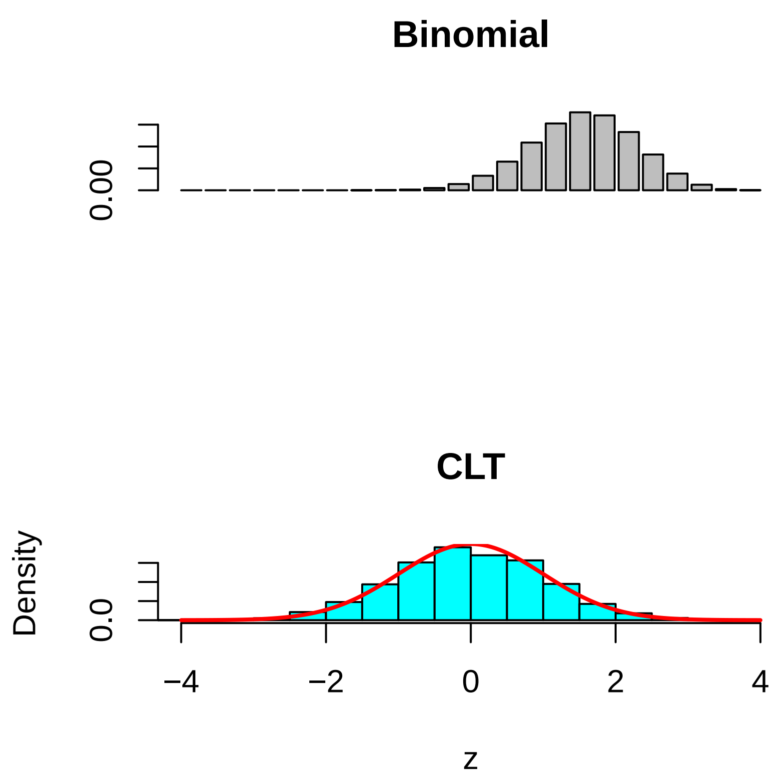
$$\lim_{n\uparrow\infty} P(T_n \leq t) = \Phi(t)$$

Exercise 128. Find the value of $\lim_{n\uparrow\infty} e^{-n}\sum_{k=0}^{n} \frac{n^k}{k!}$.

```
itrn<-10000
sm1<-array(0,dim = c(itrn))
sm2<-array(0,dim = c(itrn))
sm3<-array(0,dim = c(itrn))

for(i in 1 : itrn){
n<-100
x<-rnorm(n = n, mean = 5,sd = 3)
sm1[i]<-(mean(x))

n<-1000
x<-rnorm(n = n, mean = 5,sd = 3)
sm2[i]<-(mean(x))

n<-10000
x<-rnorm(n = n, mean = 5,sd = 3)
sm3[i]<-(mean(x))

}
par(mfrow=c(2,2))
hist(sm1, probability = T, main = "n=100")
lines(density(sm1),col=1)
abline(v=5, col=4, lwd=2)

hist(sm2, probability = T, main = "n=1000")
lines(density(sm2),col=2)
abline(v=5, col=4, lwd=2)

hist(sm3, probability = T, main = "n=10000")
lines(density(sm3),col=3)
abline(v=5, col=4, lwd=2)


plot(density(sm3), col=3,xlim=c(4,6),main = "WLLN", xlab = " comparing densities")
lines(density(sm2),col=2)
lines(density(sm1),col=1)
abline(v=5, col=4, lwd=1)
```

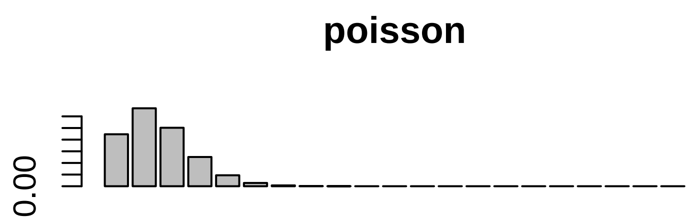

n=100
Density
0.0
0.4
0.8
1.2
4.0
4.5
5.0
5.5
6.0
sm1
n=1000
Density
0
1
2
3
4
4.6
4.8
5.0
5.2
5.4
sm2
n=10000
Density
0
4
8
12
4.90
5.00
5.10
sm3
WLLN
Density
0
4
8
12
4.0
4.5
5.0
5.5
6.0
comparing densities

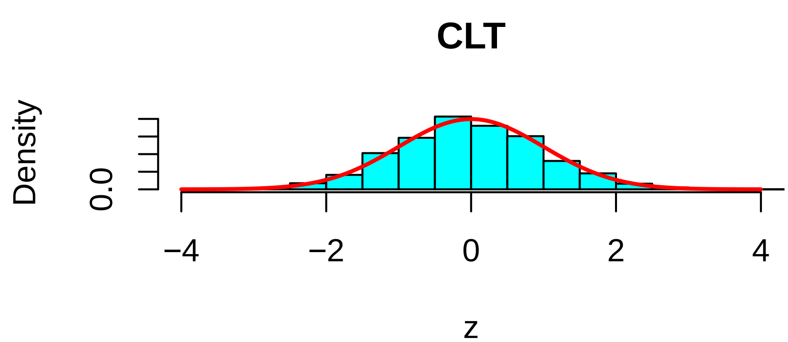
```
# Binomial CLT
bin_clt<-function(itrn,ss,n,p){
  z<-array(0,dim=c(itrn))
  for(i in 1 : itrn){
    x<-rbinom(ss,n,p)
    z[i]<-sqrt(ss)*(mean(x)-n*p)/sqrt(n*p*(1-p))
  }
  par(mfrow=c(2,1))
  barplot(dbinom(0:n,n,p), main = "Binomial")
  hist(z, probability = T, xlim = c(-4,4) ,col=5 ,main = "CLT" )
  s<-seq(-4,4,by=0.01)
  lines(dnorm(s,0,1)~s , col=2, lwd=2)
}
ss<-100 # sample size
itrn<-10000 #iteration
n=23
p=0.7
bin_clt(itrn,ss,n,p)
```




```
#Poisson CLT
poiss_clt<-function(itrn,ss,lambda){
```

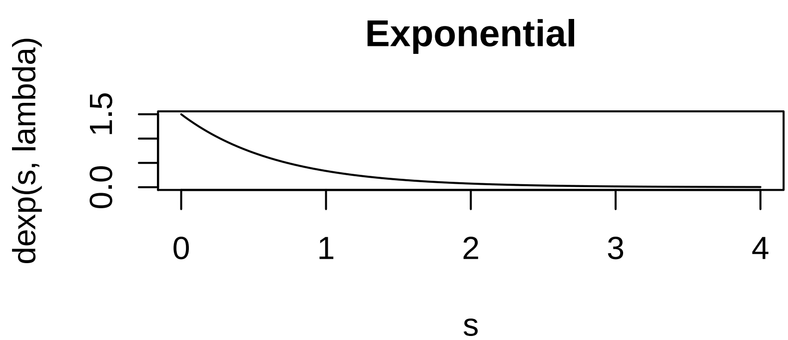
```
  z<-array(0,dim=c(itrn))
  for(i in 1 : itrn){
    x<-rpois(ss,lambda)
    z[i]<-sqrt(ss)*(mean(x)-lambda)/sqrt(lambda)

  }
  par(mfrow=c(2,1))
  barplot(dpois(0:20,lambda), main = "poisson")
  hist(z, probability = T, xlim = c(-4,4) ,col=5 ,main = "CLT" )
  s<-seq(-4,4,by=0.01)
  lines(dnorm(s,0,1)~s , col=2, lwd=2)
}
ss<-150 # sample size
itrn<-10000 #iteration
lambda= 1.5
poiss_clt(itrn,ss,lambda)
```

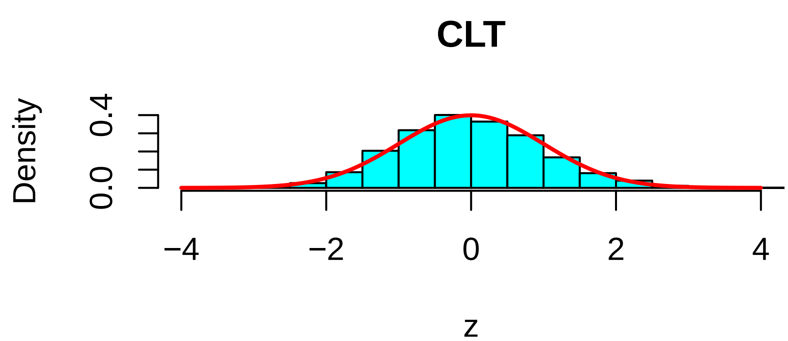





```
#exponential CLT
exp_clt<-function(itrn,ss,lambda){
  z<-array(0,dim=c(itrn))
  for(i in 1 : itrn){
```

```
    x<-rexp(ss,lambda)
    z[i]<-sqrt(ss)*(mean(x)-1/lambda)/sqrt(1/lambda^2)
  }
  par(mfrow=c(2,1))
  s<-seq(0,4, by=0.01)
  plot(dexp(s,lambda)~s, main = "Exponential", type="l")
  hist(z, probability = T, xlim = c(-4,4) ,col=5 ,main = "CLT" )
  s<-seq(-4,4,by=0.01)
  lines(dnorm(s,0,1)~s , col=2, lwd=2)
}
ss<-150 # sample size
itrn<-10000 #iteration
lambda= 1.5
exp_clt(itrn,ss,lambda)
```





10. Estimation

Let $\mathbf{x} = (x_1, x_2, \cdots, x_n)$ be the observed/ realized values of a set of i.i.d. random variables $\mathbf{X} = (X_1, X_2, \cdots, X_n)$ where$X_i \overset{iid}{\sim} f_\theta$ for some $\theta \in \Theta$. Here a family of distributions is denoted by

$$\mathcal{F} = \{f(x|\theta)|\theta \in \Theta\} \text{ or } \{F(x|\theta)|\theta \in \Theta\}$$

Parametric Estimation: In a parametric inference problem it is assumed that the family of the distribution is known but the particular value of the parameter is unknown. We estimate the value of the parameter θ as a function of the observations $\mathbf{x}$. The ultimate goal is to approximate the p.d.ff_θ or F_θ through the estimation of θ itself. Parametric estimation has two aspects, namely, (a) **Point estimation** and (b) **Interval estimation** .[We will learn it after Testing]

In point estimation we will learn
(a) Definition of an estimator
(b) Good properties of an estimator
(c) Methods of estimation (MME and MLE)

Definition 129. Statistic: A statistic is a function of random variables and it is free from any unknown parameter. Being a (measurable) function,$T(\mathbf{X})$ say , of random variables it is also a random variable.

Definition 130. Estimator: If the statistic $T(\mathbf{X})$ is used to estimate a parametric function $g(\theta)$ then $T(X)$ is said to be {an estimator of $g(\theta)$. And a realized value of it for $\mathbf{X} = \mathbf{x}$ i.e. $T(\mathbf{x})$ is know as **an estimate** of θ. We often abuse the notation as $g(\hat{\theta}) = T(\mathbf{x})$ and $g(\hat{\theta}) = T(\mathbf{X})$ which are understood from the context.

Definition 131. Unbiased estimator: An estimator $T(\mathbf{X})$ is said to be an unbiased estimator of a parametric function $g(\theta)$ if $E(T(\mathbf{X}) - g(\theta)) = 0 \ \forall\ \theta \in \Theta$.

Remark 132. It does not require$T(\mathbf{x}) = g(\theta)$ to be hold or it may hold with probability zero.

Definition 133. Bias: The bias of an estimator $T(\mathbf{X})$ while estimating a parametric function $g(\theta)$ is $B_{g(\theta)}(T(\mathbf{X})) = E(T(\mathbf{X}) - g(\theta)) \ \forall\ \theta \in \Theta$.

Definition 134. Asymptotically unbiased estimator: Denoting $T_n = T(X_1, X_2, \cdots, X_n)$ an estimator T_n is said to be asymptotically unbiased of $g(\theta)$ if

$$\lim_{n\to\infty} B_{g(\theta)}(T_n) = \lim_{n\to\infty} E(T_n - g(\theta)) = 0$$

Definition 135. Consistent estimator: An estimator T_n is said to be consistent estimator $g(\theta)$ if $T_n \xrightarrow{P} g(\theta)$ i.e.

$$\lim_{n\to\infty} P(|T_n - g(\theta)| < \epsilon) = 1 \ \forall\ \theta \in \Theta, \epsilon > 0$$

Definition 136. Mean squared error (MSE): The MSE of an estimator $T(\mathbf{X})$ while estimating a parametric function $g(\theta)$ is

$$MSE_{g(\theta)}(T(\mathbf{X})) = E[(T(\mathbf{X}) - g(\theta))^2] \ \forall\ \theta \in \Theta.$$

Exercise 137. Show that $MSE_{g(\theta)}(T(\mathbf{X})) = Var(T(\mathbf{X})) + B^2_{g(\theta)}(T(\mathbf{X}))$

Exercise 138. If $MSE_{g(\theta)}(T_n(\mathbf{X})) \downarrow 0$ as $n \uparrow \infty$ then show that $(T_n(\mathbf{X}))$ is a consistent estimator.

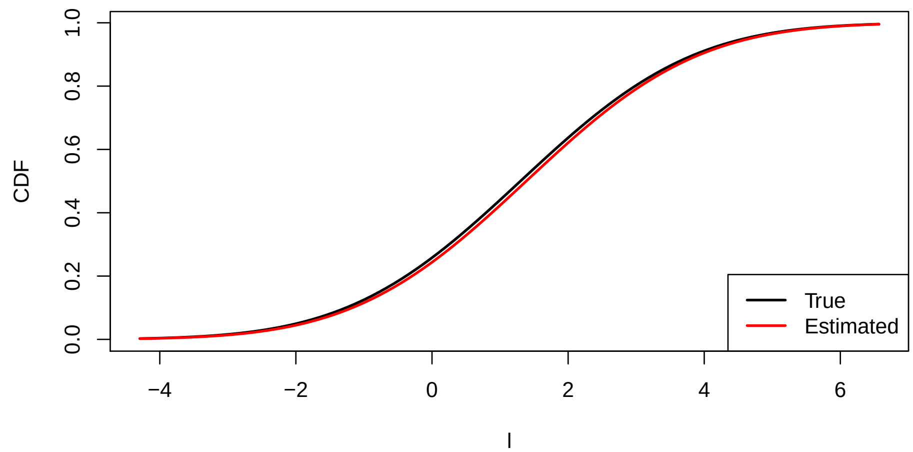
Remark 139. Asymptotic unbiasedness and consistency are large sample properties and both are based on L_1 norm. . MSE is defined based on L_2 norm.

Exercise 140. Let $(X_1, X_2, \cdots, X_n)$ be i.i.d random variables with $E(X) = \mu$ and $Var(X) = \sigma^2$. and define $T_n(\mathbf{X}) = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, $S_1^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ and $S_2^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$. Show that
(a) $T_n(\mathbf{X})$ is an unbiased estimator of μ.
(b) S_1^2 is an unbiased estimator of σ^2
(c) S_2^2 is an asymptotically unbiased estimator of σ^2

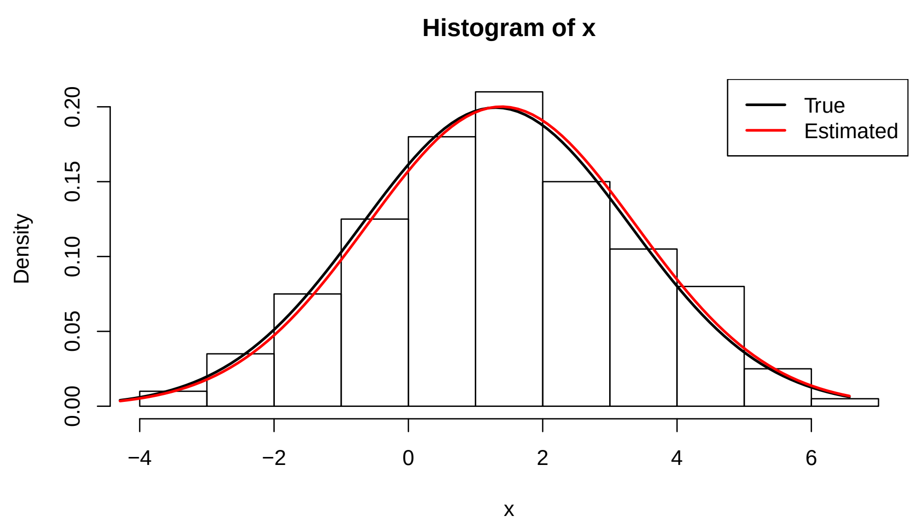
Exercise 141. Let $(X_1, X_2, \cdots, X_n) \overset{iid}{\sim} N(\mu, \sigma^2)$. Show that $MSE(S_2^2) < MSE(S_1^2)$.
Note: Unbiased estimator need not have minimum MSE.

Definition 142. Method of Moment for Estimation (MME): Consider $\mathbf{x} = (x_1, x_2, \cdots, x_n)$ be the observed/ realized values of a set of i.i.d. random variables $\mathbf{X} = (X_1, X_2, \cdots, X_n)$ where $X_i \overset{iid}{\sim} f_\theta$ for some $\theta \in \Theta$. Then
Step 1: Computer theoretical moments from the p.d.f.
Step 2: Computer empirical moments from the data.
Step 3: Construct k equations if you have k unknown parameters.
Step 4: Solve the equations for the parameters.

```
# Distribution : Normal
mu<-1.3 # mean
s<- 2   # sigma
n<- 200 # sample size
x<- rnorm(n,mean = mu,sd = s) # data
xmin<- min(x) # min of data
xmax<-max(x)  # max data
l<- seq(xmin-0.5, xmax+0.5, length=100)
######### Estimation ##########
muh<-mean(x)
sh<-sd(x)
###############################
cat("True mean=", mu, "estimated mean=", muh,"\n")
## True mean= 1.3 estimated mean= 1.385195
cat("True sigma=", s, "estimated sigma=", sh,"\n")
## True sigma= 2 estimated sigma= 1.993788
###############################

plot(pnorm(q = l,mean = mu,sd = s)~l, type = 'l', col=1, lwd=2, ylab = "CDF")
lines(pnorm(q = l,mean = muh,sd = sh)~l, type = 'l', col=2, lwd=2)
legend("bottomright",legend =  c("True", "Estimated"), col = c(1,2), lwd = c(2,2))
```




```
hist(x,probability = T)
lines(dnorm(x = l,mean = mu,sd = s)~l, type = 'l', col=1, lwd=2, ylab = "PDF")
lines(dnorm(x=l,mean = muh,sd = sh)~l, type = 'l', col=2, lwd=2)
legend("topright",legend =  c("True", "Estimated"), col = c(1,2), lwd = c(2,2))
```



Remark 143. We can not use MME to estimate the parameters of $C(\mu, \sigma)$, because the moments does not exists for Cauchy distribution.

Definition 144. Maximum Likelihood Estimator: Consider $\mathbf{x} = (x_1, x_2, \cdots, x_n)$ be the observed/ realized values of a set of i.i.d. random variables $\mathbf{X} = (X_1, X_2, \cdots, X_n)$ where $X_i \overset{iid}{\sim} f_\theta$ for some $\theta \in \Theta$. Then the joint p.d.f. of $\mathbf{X} = (X_1, X_2, \cdots, X_n)$ is a function of $\mathbf{x}$ when the parameter value is fixed i.e.

$$f(\mathbf{x}|\theta) = \prod_{i=1}^{n} f(x_i, \theta)$$

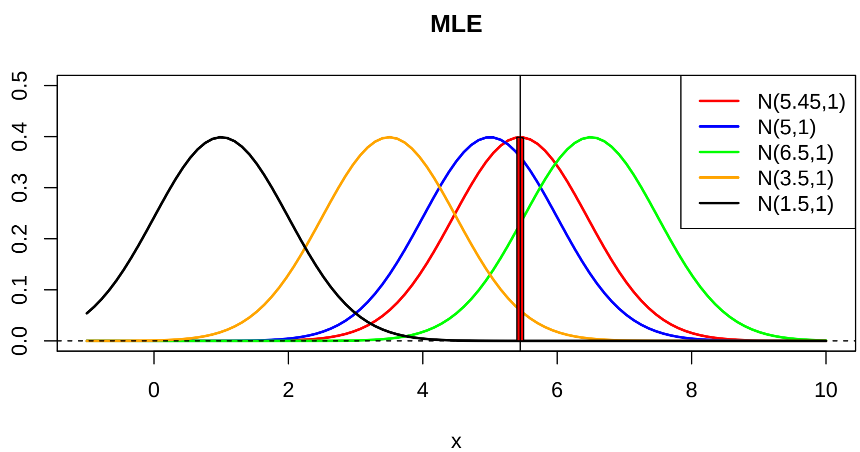
and the likelihood of a function of parameter for a given set of data $\mathbf{X} = \mathbf{x}$ i.e.

$$\ell(\theta|\mathbf{x}) = \prod_{i=1}^{n} f(x_i, \theta).$$

Hence the maximum likelihood estimator of θ is

$$\hat{\theta}_{mle} = \arg\max_{\theta\in\Theta} \ell(\theta|\mathbf{x}) = \arg\max_{\theta\in\Theta} \log \ell(\theta|\mathbf{x})$$

NOTE: Finding the maxima through differentiation is possible **only of** ℓ is a smoothly differentiable function w.r.t θ. Otherwise it has to be maximized by some other methods. **Differentiation is not the only way of finding maxima or minima.**



Exercise 145. $(X_1, X_2, \cdots, X_n) \overset{iid}{\sim} U(0, \theta)$. where $\theta \in \Theta = (0, \infty)$.
(a)Find the MLE of θ.
(b) Is it an unbiased estimator ?
(c) Find the MSE.

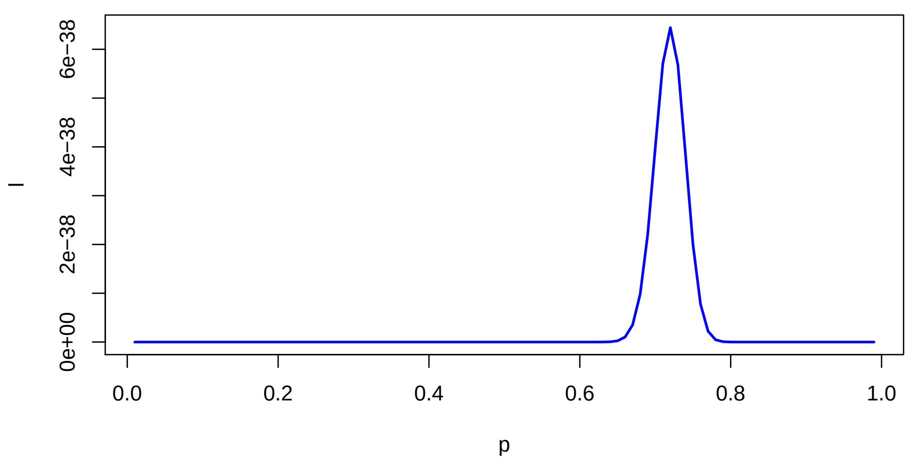
Exercise 146. Let $(X_1, X_2, \cdots, X_n) \overset{iid}{\sim} N(\mu, \sigma^2)$.
(a) Find the MME and MLE of μ and σ^2. Are they same ?
(b) Are they unbiased estimators?

Exercise 147. Let $(X_1, X_2, \cdots, X_n) \overset{iid}{\sim} Gamma(\alpha, \lambda)$.
(a) Find the MME of (α, λ)?
(b) Find MLE of (α, λ) by by an iterative method of solution.
NOTE: You may use the MME as an initial value of iteration to obtain the MLE.

```
# MLE of binomal parameter
set.seed(12)
n<-10  # size of binomial
x<- sort(rbinom (50, n, 0.7))  # sample  given
print(x)
```

```
##  [1]  4  5  5  5  5  6  6  6  6  6  6  6  6  6  6  6  7  7  7  7  7  7  7
## [24]  7  7  7  7  7  8  8  8  8  8  8  8  8  8  8  8  8  8  8  9  9  9  9
## [47]  9  9 10 10
```

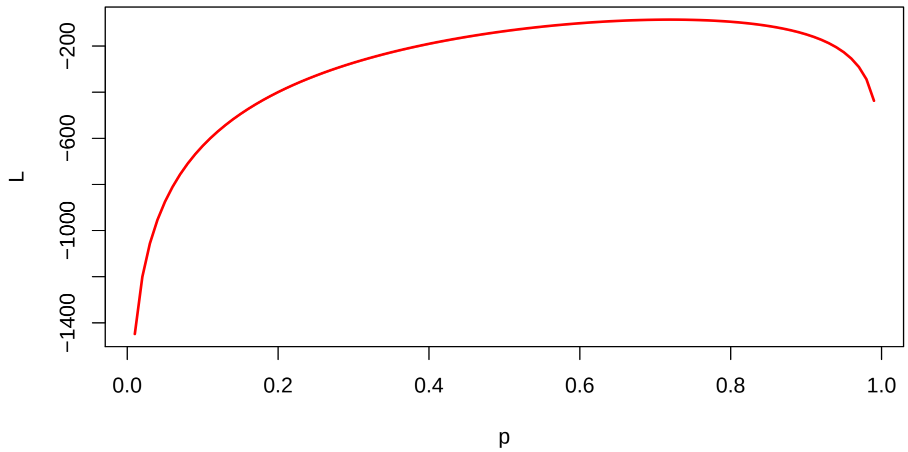
```
# MLE finding
p<-seq(0.01,0.99,by = 0.01)
l<-array(0,dim=c(length(p)))
for (i in 1 : length(p)){
  l[i]<-prod(dbinom(x,n,p[i]))  # product of likelihood
}

plot(l~p, type='l', col=4, lwd=2)
```




```
mle1<-p[which(l==max(l))]
print(mle1)
```

```
## [1] 0.72
```

```
L<-array(0,dim=c(length(p)))
for (i in 1 : length(p)){
  L[i]<-sum(log(dbinom(x,n,p[i])))  #sum of log likelihood
}

plot(L~p,type='l', col=2, lwd=2)
```




```
mle2<-p[which(L==max(L))]
print(mle2)
## [1] 0.72
```

Propertied of MLE:
(a) MLE need not be unique.
(b) MLE need not be an unbiased estimator.
(c) MLE is always a consistent estimator.
(d) MLE is asymptotically normally distributed up to some location and scale when some regularity condition satisfied like
(1) Range of the random variable is free from parameter.
(2) Likelihood is smoothly differentiable for up to 3rd order and corresponding
expectations exists.

Exercise 148. $(X_1, X_2, \cdots, X_n) \overset{iid}{\sim} U(\theta - 0.5, \theta + 0.5)$ where $\theta \in \Theta = (-\infty, \infty)$.
(a)Find the MLE of θ.
(b) Is it unique?
(c) Is it consistent? Find the MSE.

Definition 149. Interval Estimation: Consider a pair of statistic $(L(\mathbf{X}), U(\mathbf{X}))$ such that for a parameter θ ,

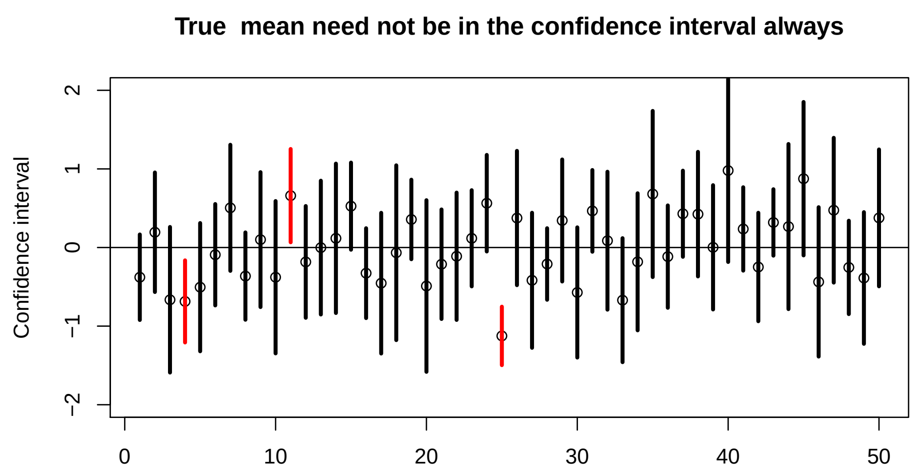
$$P_\theta(\theta \in [L(\mathbf{X}), U(\mathbf{X})]) = 1 - \alpha$$

Then a $100(1-\alpha)\%$ confidence interval of θ is considered to be $[L(\mathbf{X}), U(\mathbf{X})]$.

Example 150. If $X_1, X_2, \ldots, X_n$ are i.i.d random variables with $N(\mu, \sigma^2)$ distribution with known value of σ^2. Then a $100(1-\alpha)\%$ CI of μ is

$$\left[L(\mathbf{X}) = \overline{X} - \frac{\sigma}{\sqrt{n}} z_{\alpha/2}, U(\mathbf{X}) = \overline{X} + \frac{\sigma}{\sqrt{n}} z_{\alpha/2}\right]$$

```
set.seed(10)
N <- 50
n <- 8 # sample size
v <- matrix(c(0,0),nrow=2)
for (i in 1:N) {
  x <- rnorm(n)
  v <- cbind(v, t.test(x)$conf.int)
}
v <- v[,2:(N+1)]
plot(apply(v,2,mean), ylim=c(-2,2), ylab='Confidence interval', xlab='')
abline(0,0)
c <- apply(v,2,min)>0 | apply(v,2,max)<0
segments(1:N,v[1,],1:N,v[2,], col=c(par('fg'),'red')[c+1], lwd=3)
title(main="True  mean need not be in the confidence interval always")
```



Example 151. If $X_1, X_{2,} \ldots, X_n$ are i.i.d random variables with $N(\mu, \sigma^2)$ distribution . Then a $100(1-\alpha)\%$ CI of μ is

$$\left[L(\mathbf{X}) = \overline{X} - \frac{\hat{\sigma_u}}{\sqrt{n}}\tau_{\alpha/2,n-1}, U(\mathbf{X}) = \overline{X} + \frac{\hat{\sigma_u}}{\sqrt{n}}\tau_{\alpha/2,n-1}\right]$$

$\hat{\sigma}_u^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ is an unbiased estimator of unknown variance and a $100(1-\alpha)\%$ CI of σ^2 is

$$\left[L(\mathbf{X}) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\chi^2_{\alpha/2,(n-1)}}, U(\mathbf{X}) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\chi^2_{1-\alpha/2,(n-1)}}\right]$$

11. Testing of Hypothesis

Definition 152. Hypothesis: A hypothesis in parametric inference is a statement about the population parameter. It has two categories. A **null hypothesis** (H_0) specifies a subset Θ_0 in the parameter space Θ. If Θ_a is a singleton set then it called a **simple null**, otherwise a **composite null.** On the other hand an **alternative hypothesis** (H_1) specifies another subset $\Theta_a \subset \Theta$ which is disjoint to Θ_0.

Definition 153. Test Rule: A test rule is a statistical procedure, based on the distribution of the test statistic, which will reject the null hypothesis in favour of the alternative hypothesis.

Definition 154. Rejection Region or Critical region: A rejection Region or critical region is a subset $C \subset \mathbb{R}^n$ such that $\mathbf{X} \in C \Leftrightarrow T(\mathbf{X})$ will reject the null hypothesis.

Definition 155. Level-α test: For any $\alpha \in (0,1)$, a test is said to be level-α test if

$$\sup_{\theta \in \Theta_0} P_\theta(\mathbf{X} \in C) \leq \alpha.$$

Definition 156. Size-α test: For any $\alpha \in (0,1)$, a test is said to be size-α test if

$$\sup_{\theta \in \Theta_0} P_\theta(\mathbf{X} \in C) = \alpha.$$

Definition 157. Power-function: A power function is a function

$$P_\theta(\mathbf{X} \in C) : \Theta_a \to [0,1]$$

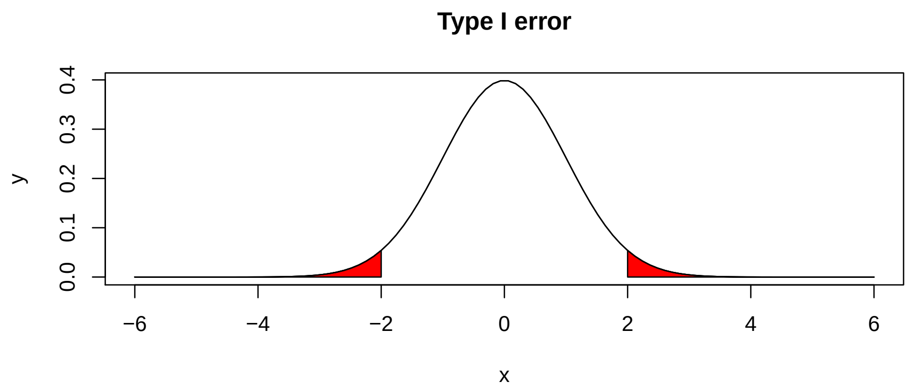
Remark 158. More than one tests with same level can be compared in terms of power functions. A test procedure with more power than the other with same level can be considered a better test.

Definition 159. Type-I error: The event $\mathbf{X} \in C$ when $\theta \in \Theta_0$ is known as Type-I error.

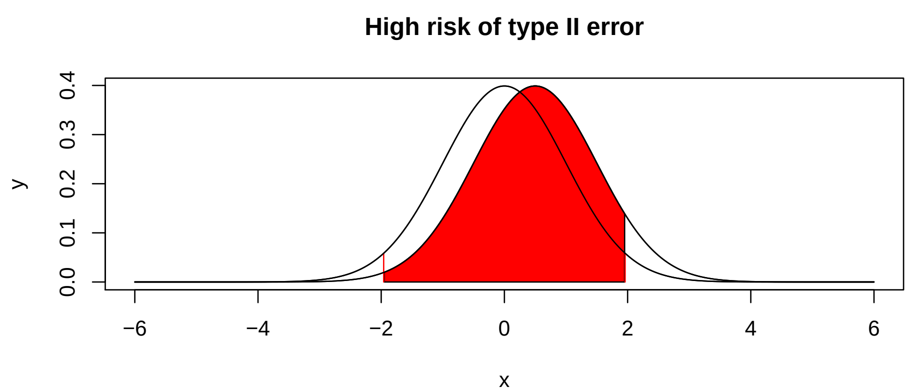
Definition 160. Type-II error: The event $\mathbf{X} \in C^c$ when $\theta \in \Theta_a$ is known as Type-II error. Power is 1-P(Type-II error).

```r
colorie <- function (x, y1, y2, N=1000, ...) {
  for (t in (0:N)/N) {
    lines(x, t*y1+(1-t)*y2, ...)
  }
}
# No, there is already a function to do this
colorie <- function (x, y1, y2, ...) {
  polygon( c(x, x[length(x):1]), c(y1, y2[length(y2):1]), ... )
}
x <- seq(-6,6, length=100)
y <- dnorm(x)
plot(y~x, type='l')
i = x<qnorm(.025)
colorie(x[i],y[i],rep(0,sum(i)) ,col='red')
i = x>qnorm(.975)
```

```
colorie(x[i],y[i],rep(0,sum(i)) ,col='red')
lines(y~x)
title(main="Type I error")
```

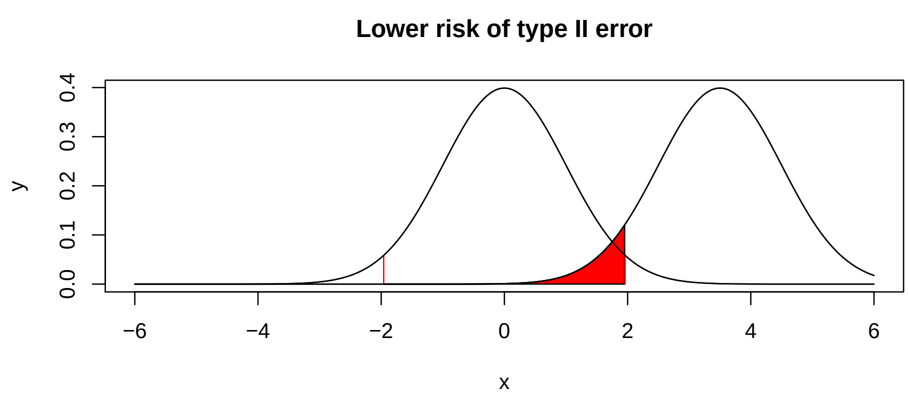



```
x <- seq(-6,6, length=1000)
y <- dnorm(x)
plot(y~x, type='l')
y2 <- dnorm(x-.5)
lines(y2~x)
i <- x>qnorm(.025) & x<qnorm(.975)
colorie(x[i],y2[i],rep(0,sum(i)), col='red')
segments( qnorm(.025),0,qnorm(.025),dnorm(qnorm(.025)), col='red' )
segments( qnorm(.975),0,qnorm(.975),dnorm(qnorm(.975)), col='red' )
lines(y~x)
lines(y2~x)
title(main="High risk of type II error")
```

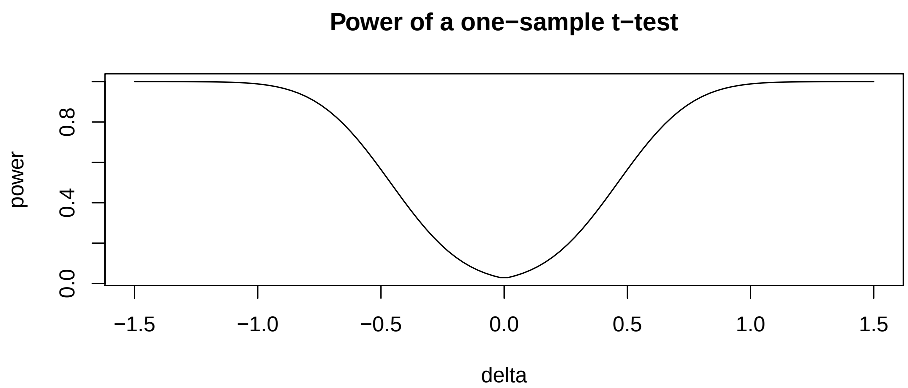



```
x <- seq(-6,6, length=1000)
y <- dnorm(x)
```

```
plot(y~x, type='l')
y2 <- dnorm(x-3.5)
lines(y2~x)
i <- x>qnorm(.025) & x<qnorm(.975)
colorie(x[i],y2[i],rep(0,sum(i)), col='red')
segments( qnorm(.025),0,qnorm(.025),dnorm(qnorm(.025)), col='red' )
segments( qnorm(.975),0,qnorm(.975),dnorm(qnorm(.975)), col='red' )
lines(y~x)
lines(y2~x)
title(main="Lower risk of type II error")
```




```
delta <- seq(-1.5, 1.5, length=100)
p <- NULL
for (d in delta) {
  p <- append(p,
              power.t.test(delta=d, sd=1, sig.level=0.05, n=20,
                           type='one.sample')$power)
}
plot(p~delta, type='l',
     ylab='power', main='Power of a one-sample t-test')
```



Lemma 161. ***Neyman-Pearson Lemma (1933):*** *To test $H_0 : \theta = \theta_0$ vs $H_1 : \theta = \theta_1$ reject H_0 in favour of H_1 at level/ size α if*

$$\Lambda(\mathbf{x}) = \frac{f(\mathbf{x}|\theta_0)}{f(\mathbf{x}|\theta_1)} \leq \xi \quad \textit{such that} \quad P_{\theta_0}(\Lambda(\mathbf{X}) \leq \xi) = \alpha$$

How to perform a test ??

Step1: Estimate the parameter for which the testing to be done.
Step2: Estimate the unknown parameters if any.
Step3: Construct the test statistic and obtain its value.
Step4: Obtain the exact or asymptotic distribution of the test statistic under the null hypothesis.
Step5: Depending on the alternative hypothesis (H_1) and level (α) decide the cut-off value or rejection condition.
Step6: Compare the observed value of test statistic (from Step 4) and the cut off value (from Step 5) to conclude the test. You may use **p-value** also.

Exercise 162. Let $X_1, ..., X_n \sim N(\mu_, \sigma^2)$ Perform a test at size 0.05 for
(a)$H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$. when σ^2 is known
(b)$H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$. when σ^2 is unknown
(a)$H_0 : \sigma^2 = \sigma_0^2$ vs $H_1 : \sigma^2 \neq \sigma_0^2$ when μ is unknown

```
library("TeachingDemos")
n<-10
mu_true<-10.5
sd_true<-1.2
x<-rnorm(10,mu_true,sd_true) # generate data
#########################
print(x)
##  [1] 10.404652 11.918102 13.123373 10.987410  9.613969  8.152216  8.159945
##  [8]  9.370803 11.937344  9.750913
cat("Unbiased estimate of mean =",mean(x), "\n")
## Unbiased estimate of mean = 10.34187
cat("Unbiased estimate of variance =",var(x), "\n")
```

```
## Unbiased estimate of variance = 2.729432
alpha<-0.05
## (a)H_0: mu = 10  vs H_1: mu not equal to 10  when sigma^2 = (1.2)^2 is known
za<-z.test(x,mu = 10,stdev =  sd_true ,alternative =c("two.sided"),conf.level = (1-alpha))
print(za)
##
##  One Sample z-test
##
## data:  x
## z = 0.90091, n = 10.00000, Std. Dev. = 1.20000, Std. Dev. of the
## sample mean = 0.37947, p-value = 0.3676
## alternative hypothesis: true mean is not equal to 10
## 95 percent confidence interval:
##   9.598119 11.085627
## sample estimates:
## mean of x
##  10.34187
##(b)H_0: mu = 10  vs H_1: mu not equal to 10  when sigma^2  is unknown
 ta<-t.test(x, mu = 10,alternative =c("two.sided"),conf.level = (1-alpha))
print(ta)
##
##  One Sample t-test
##
## data:  x
## t = 0.65438, df = 9, p-value = 0.5292
## alternative hypothesis: true mean is not equal to 10
## 95 percent confidence interval:
##   9.160032 11.523713
## sample estimates:
## mean of x
##  10.34187
##(c)H_0: sigma^2 = 1  vs H_0: sigma^2 neq 1  when mu is unknown
 va<-sigma.test(x, sigma = 1,alternative = "two.sided", conf.level = (1-alpha))
print(va)
##
##  One sample Chi-squared test for variance
##
## data:  x
## X-squared = 24.565, df = 9, p-value = 0.006984
## alternative hypothesis: true variance is not equal to 1
## 95 percent confidence interval:
##  1.291341 9.096795
## sample estimates:
## var of x
## 2.729432
```

Exercise 163. Let $X_1, ..., X_n \sim N(\mu_1, \sigma^2)$ (iid) and $Y_1, ..., Y_m \sim N(\mu_2, \sigma^2)$ (iid) are independent. Perform a test at size 0.05 for $H_0 : \mu_1 = \mu_2$ vs $H_1 : \mu_1 \neq \mu_2$.

Exercise 164. Let $(X_i, Y_i) \sim N(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho),\ i = 1, 2, ..., n$. Perform a test at size 0.05 for $H_0 : \mu_x = \mu_y$ vs $H_1 : \mu_x \neq \mu_y$. (this is known as paired-T test).

Exercise 165. Let $X_1, X_2 \ldots, X_n \sim Bernoulli(p)$ Perform a test for $H_0 : p = 0.5$ vs $H_1 : p = 0.5$ at size 0.05.

Exercise 166. Let $X_1, ..., X_n \sim N(\mu_1, \sigma_1^2)$ (iid) and $Y_1, ..., Y_m \sim N(\mu_2, \sigma_2^2)$ (iid) are independent. Perform a test at size 0.05 for $H_0 : \sigma_1^2 = \sigma_2^2$ vs $H_1 : \sigma_1^2 \neq \sigma_2^2$.

List of Test Statistic: http://en.wikipedia.org/wiki/Test_statistic

Department of Mathematics, IIT Kgaragpur
E-mail address: bbanerjee@maths.iitkgp.ac.in