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.

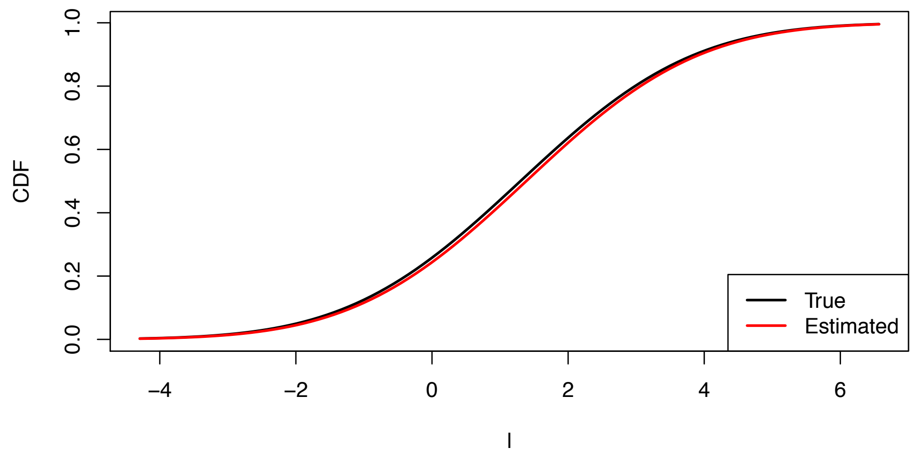
Exercise 140. Let $(X_1, X_2, \cdots, X_n)$ be i.i.d random variables with $E(X) = \mu$ and $Var(X) = \sigma^2$. and define $T_n(\mathbf{X}) = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, $S_1^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ and $S_2^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$. Show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

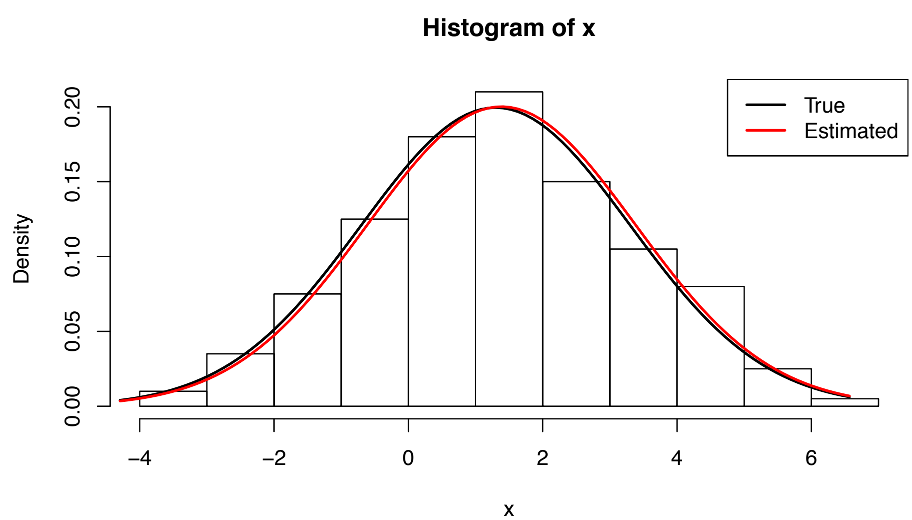
Definition 142. Method of Moment for Estimation (MME): Consider $\mathbf{x} = (x_1, x_2, \cdots, x_n)$ be the observed/ realized values of a set of i.i.d. random variables $\mathbf{X} = (X_1, X_2, \cdots, X_n)$ where $X_i \overset{iid}{\sim} f_\theta$ for some $\theta \in \Theta$. Then
Step 1: Computer 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.

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

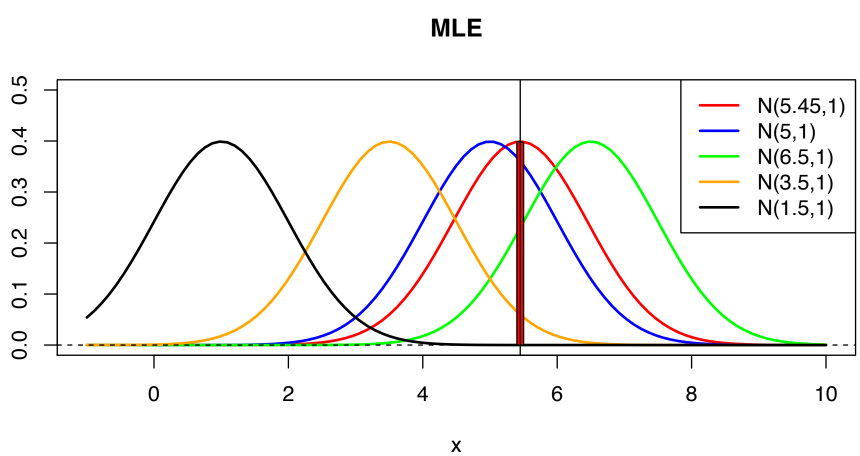
and the likelihood of a function of parameter for a given set of data $\mathbf{X} = \mathbf{x}$ i.e.

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

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?

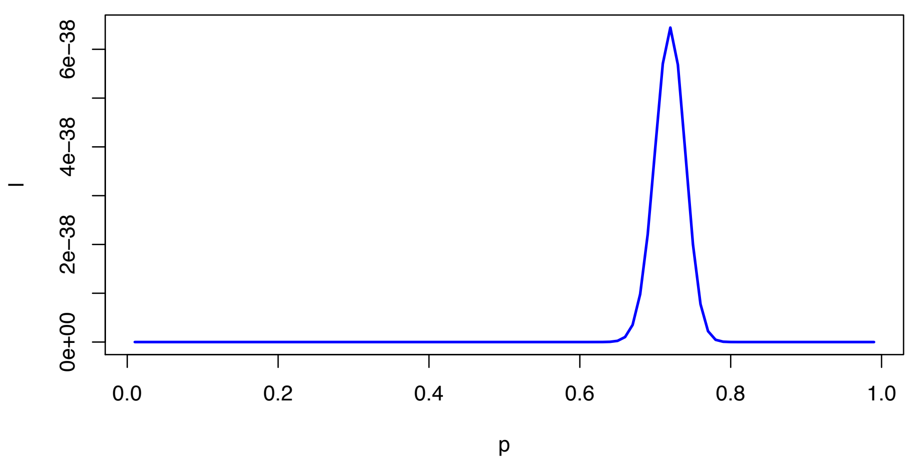
Exercise 147. Let $(X_1, X_2, \cdots, X_n) \overset{iid}{\sim} Gamma(\alpha, \lambda)$.
(a) Find the MME of (α, λ)?
(b) Find MLE of (α, λ) by by an iterative method of solution.
NOTE: You may use the MME as an initial value of iteration to obtain the MLE.

```
# 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)
```

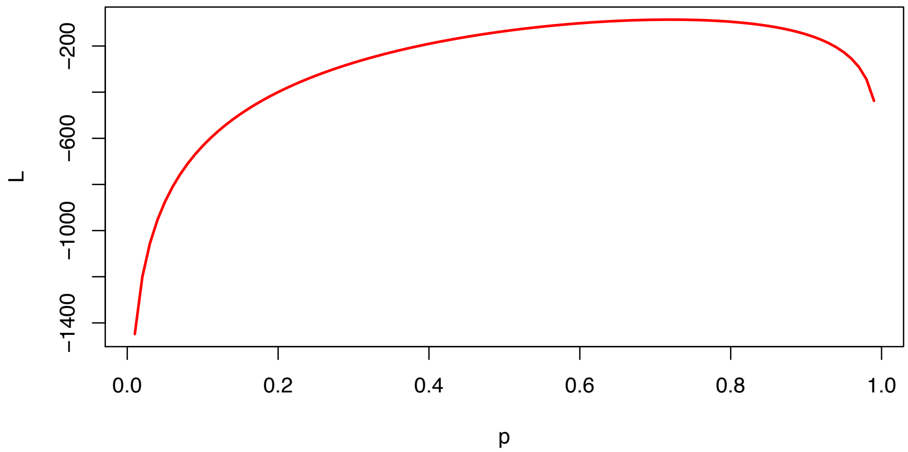



```
mle1<-p[which(l==max(l))]
print(mle1)

## [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 ,

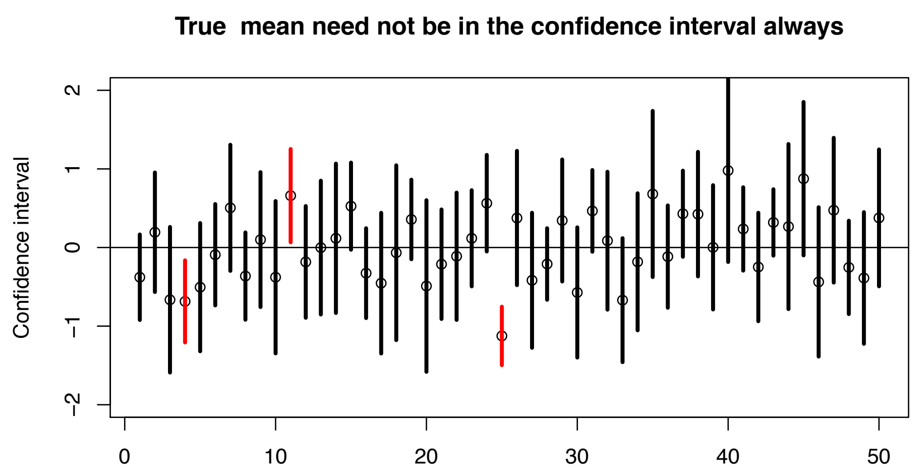
$$P_\theta(\theta \in [L(\mathbf{X}), U(\mathbf{X})]) = 1 - \alpha$$

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

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

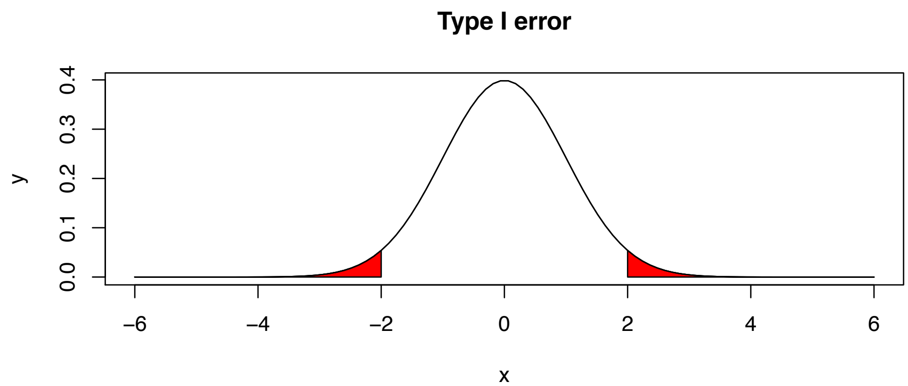
Remark 158. More than one tests with same level can be compared in terms of power functions. A test procedure with more power than the other with same level can be considered a better test.

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

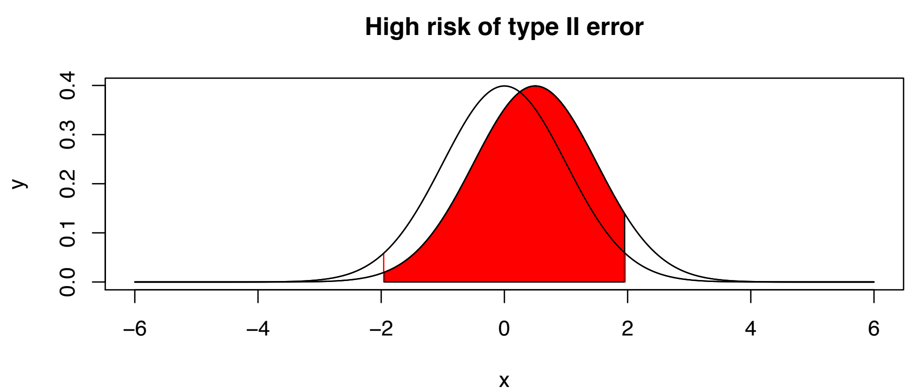
Definition 160. Type-II error: The event $\mathbf{X} \in C^c$ when $\theta \in \Theta_a$ is known as Type-II error. Power is 1-P(Type-II error).

```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")
```

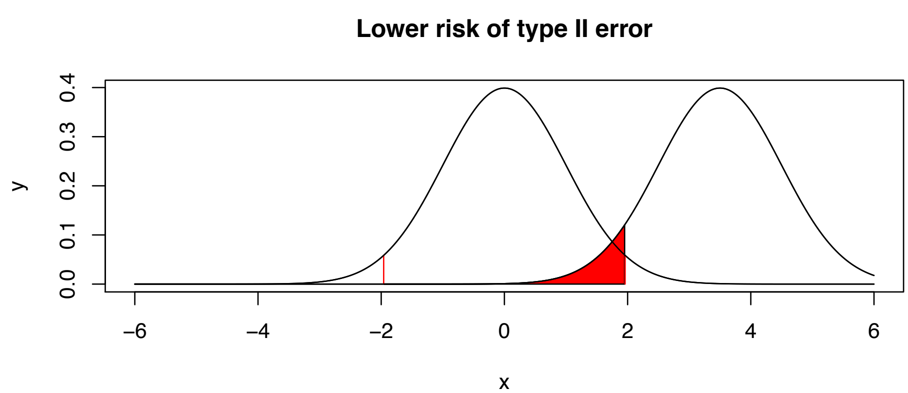



```
x <- seq(-6,6, length=1000)
y <- dnorm(x)
plot(y~x, type='l')
y2 <- dnorm(x-.5)
lines(y2~x)
i <- x>qnorm(.025) & x<qnorm(.975)
colorie(x[i],y2[i],rep(0,sum(i)), col='red')
segments( qnorm(.025),0,qnorm(.025),dnorm(qnorm(.025)), col='red' )
segments( qnorm(.975),0,qnorm(.975),dnorm(qnorm(.975)), col='red'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

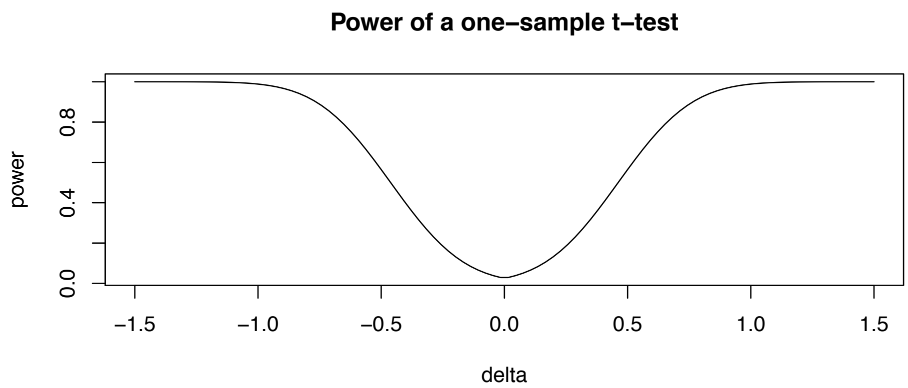



```
x <- seq(-6,6, length=1000)
y <- dnorm(x)
```

```
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

(1) Mathematical Statistics and Data Analysis by John A. Rice
(2) Probability and Statistical Inference by Hogg, R. V., Tanis, E. A. & Zimmerman D. L.

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