ESTIMATION

BUDDHANANDA BANERJEE

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

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

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.f f_θ or F_θ through the estimation of θ itself. Parametric estimation has two aspects, namely,

- ▷ **Point estimation**
  - (a) Definition of an estimator
  - (b) Good properties of an estimator
  - (c) Methods of estimation (MME and MLE)
- ▷ **Interval estimation**
  - (a) Definition of confidence interval
  - (b) Construction of confidence interval

Date: Last updated April 23, 2024.

Definition 1. 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 2. 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.

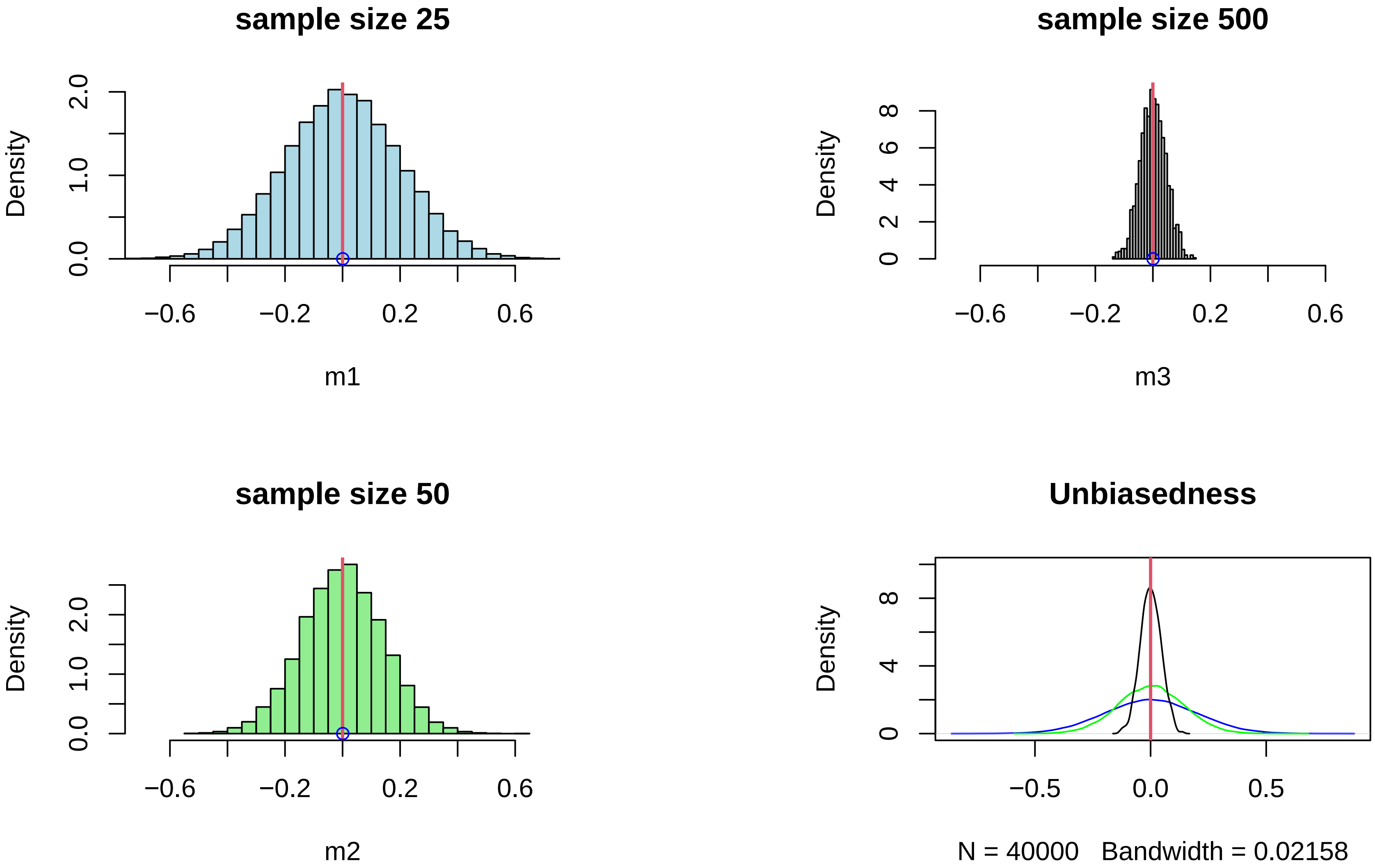
1. Properties

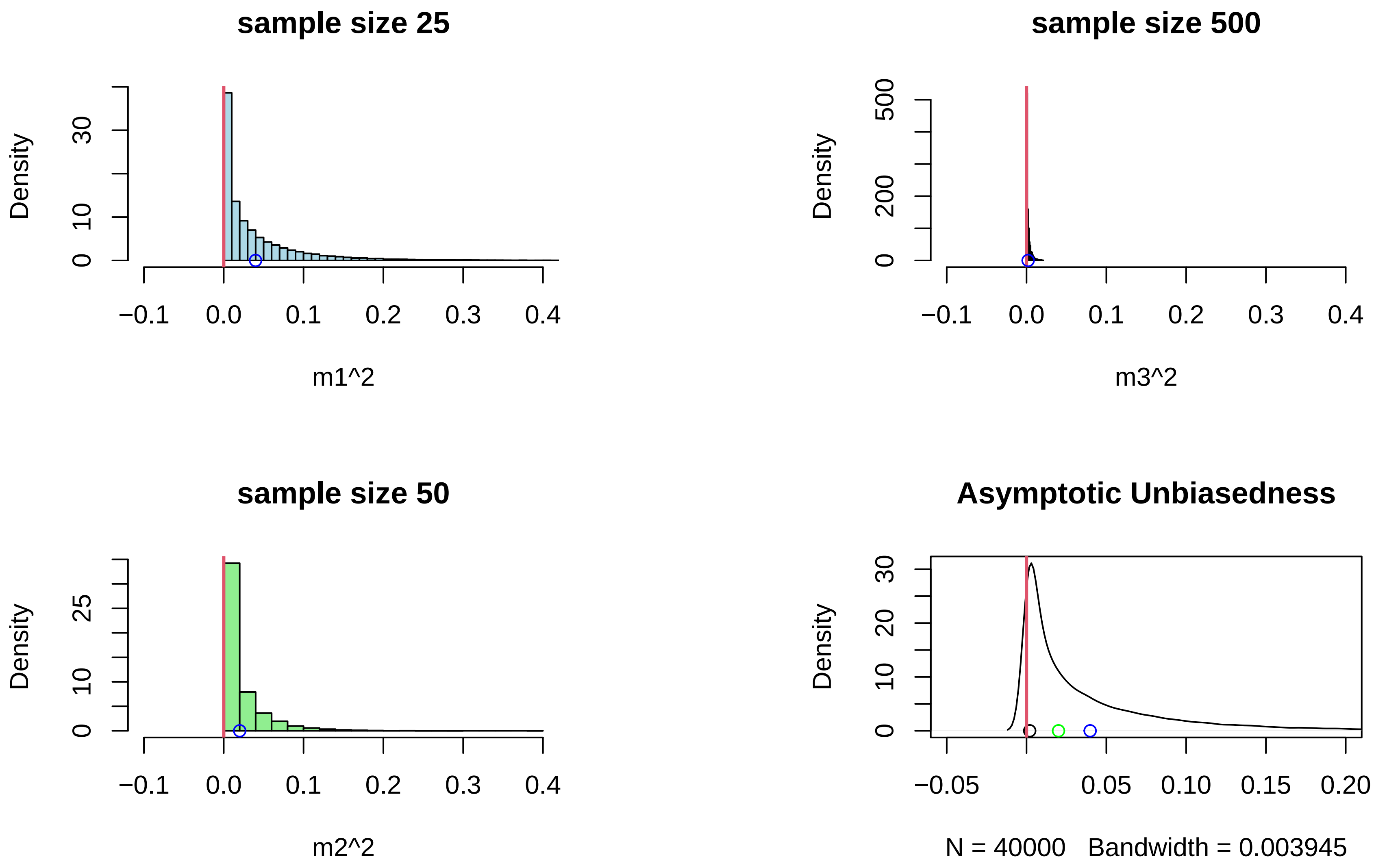
Definition 3. Unbiased estimator: An estimator $T(\mathbf{X})$ is said to be an unbiased estimator of a parametric function $g(\theta)$ if $E(T(\mathbf{X}) - g(\theta)) = 0 \ \forall \ \theta \in \Theta$.

Remark 1. It does not require$T(\mathbf{x}) = g(\theta)$ to be hold or it may hold with probability zero.

Definition 4. 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

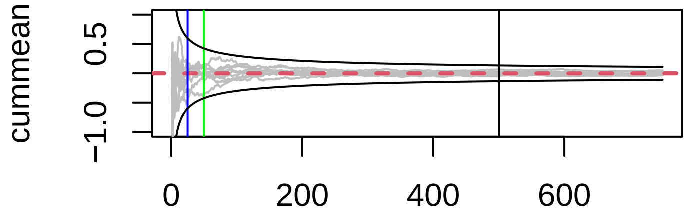
$$\lim_{n\to\infty} B_{g(\theta)}(T_n) = \lim_{n\to\infty} E(T_n - g(\theta)) = 0$$

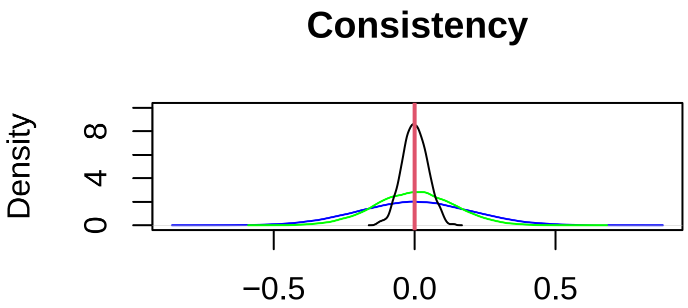

sample size 25
Density
0.0
1.0
2.0
−0.6
−0.2
0.2
0.6
m1
sample size 500
Density
0
2
4
6
8
−0.6
−0.2
0.2
0.6
m3
sample size 50
Density
0.0
1.0
2.0
−0.6
−0.2
0.2
0.6
m2
Unbiasedness
Density
0
4
8
−0.5
0.0
0.5
N = 40000 Bandwidth = 0.02158


sample size 25
Density
0
10
30
−0.1
0.0
0.1
0.2
0.3
0.4
m1^2
sample size 500
Density
0
200
500
−0.1
0.0
0.1
0.2
0.3
0.4
m3^2
sample size 50
Density
0
10
25
−0.1
0.0
0.1
0.2
0.3
0.4
m2^2
Asymptotic Unbiasedness
Density
0
10
20
30
−0.05
0.05
0.10
0.15
0.20
N = 40000 Bandwidth = 0.003945

Definition 5. Consistent estimator: An estimator T_n is said to be consistent estimator $g(\theta)$ if $T_n \xrightarrow{P} g(\theta)$ i.e.

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





2. Accuracy Measures

Definition 6. 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 7. Mean squared error (MSE): The MSE of an estimator $T(\mathbf{X})$ while estimating a parametric function $g(\theta)$ is

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

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

Remark 3. 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 4. 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 1. 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

- ▷ $T_n(\mathbf{X})$ is an unbiased estimator of μ.
- ▷ S_1^2 is an unbiased estimator of σ^2
- ▷ S_2^2 is an asymptotically unbiased estimator of σ^2.

Remark 5. Let $(X_1, X_2, \cdots, X_n) \overset{iid}{\sim} N(\mu, \sigma^2)$ then $MSE(S_2^2) < MSE(S_1^2)$. Unbiased estimator need not have minimum MSE.

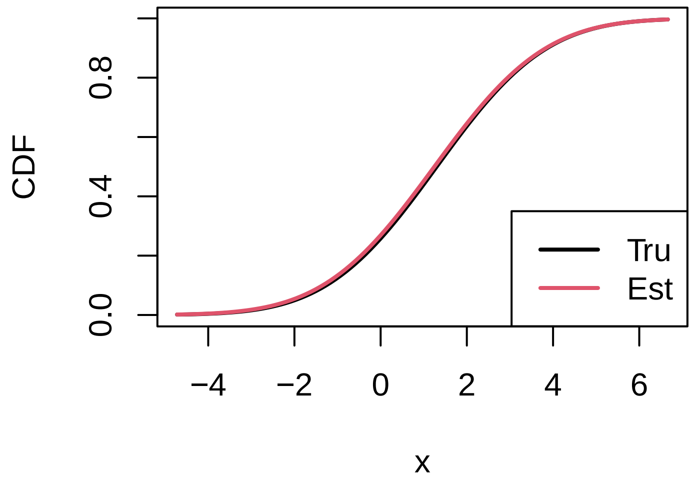
Method of Moment for Estimation (MME): Consider $\mathbf{x} = (x_1, x_2, \cdots, x_n)$ be the observed/ realized values of a set of i.i.d. random variables $\mathbf{X} = (X_1, X_2, \cdots, X_n)$ where $X_i \overset{iid}{\sim} f_\theta$ for some $\theta \in \Theta$. Then

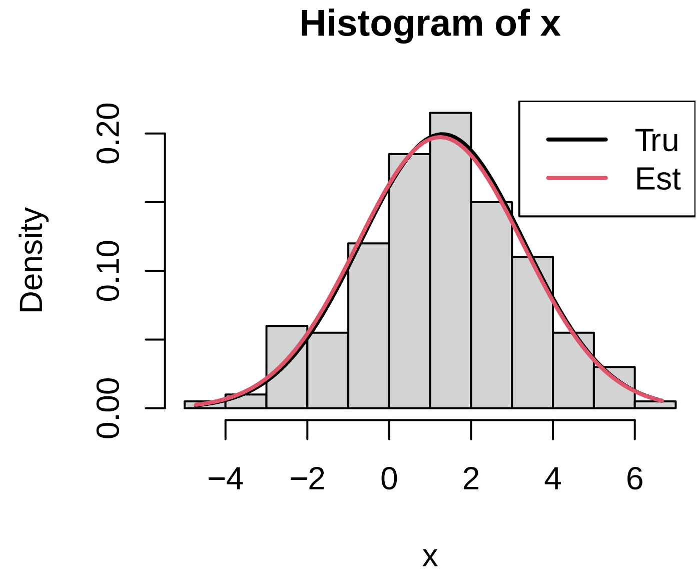
Step 1: Computer theoretical moments from the p.d.f.

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.






```
## True mean= 1.3 estimated mean= 1.246376
## True sigma= 2 estimated sigma= 2.021477
```

Remark 6. We can not use MME to estimate the parameters of $C(\mu, \sigma)$, because the moments does not exists for Cauchy distribution.

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.

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

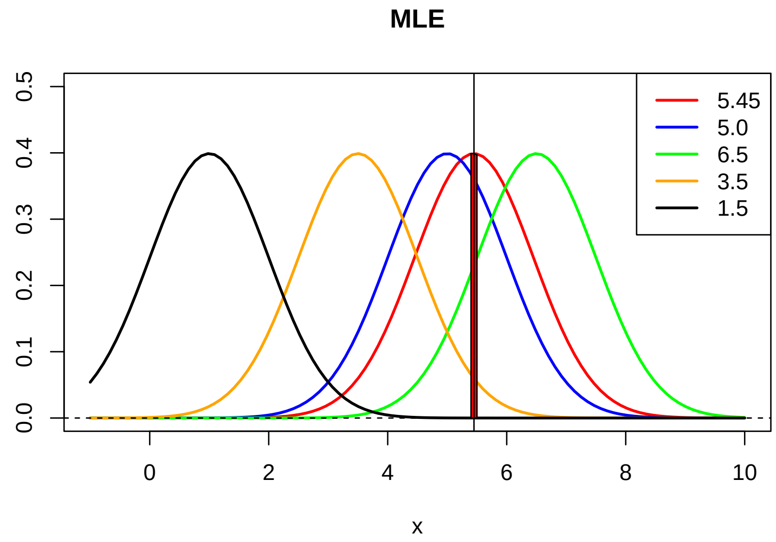
and the likelihood of a function of parameter for a given set of data $\mathbf{X} = \mathbf{x}$ i.e.

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

Hence the maximum likelihood estimator of θ is

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

Remark 7. Finding the maxima through differentiation is possible **only if** 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.**



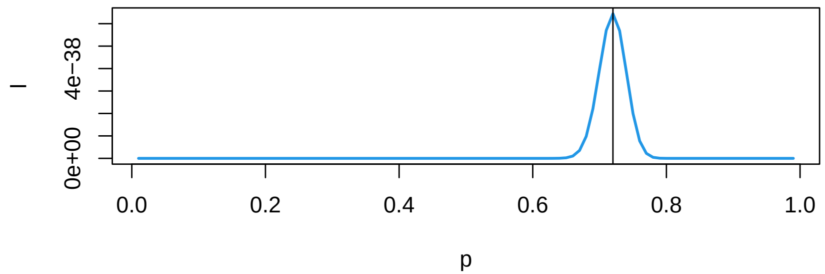
Propertied of MLE:

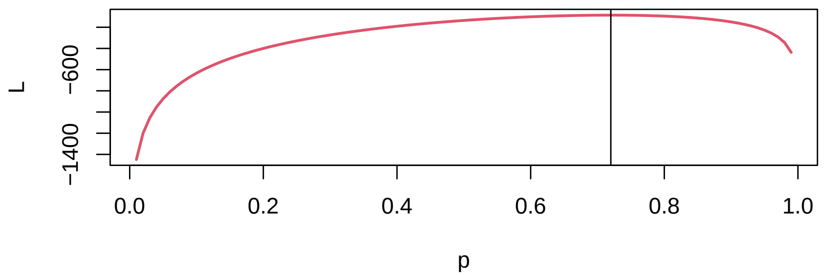
- ▷ MLE need not be unique.
- ▷ MLE need not be an unbiased estimator.
- ▷ MLE is always a consistent estimator.
- ▷ MLE is asymptotically normally distributed up to some location and scale when some **{regularity condition}** satisfied like

  (1) Range of the random variable is free from parameter.

  (2) Likelihood is smoothly differentiable for up to 3rd order and corresponding expectations exists.

```
##       [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
## [1,]    4    6    6    6    7    7    8    8    8     9
## [2,]    5    6    6    7    7    7    8    8    8     9
## [3,]    5    6    6    7    7    7    8    8    9     9
## [4,]    5    6    6    7    7    8    8    8    9    10
## [5,]    5    6    6    7    7    8    8    8    9    10
```






```
## MLE1= 0.72 MLE2= 0.72
```

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

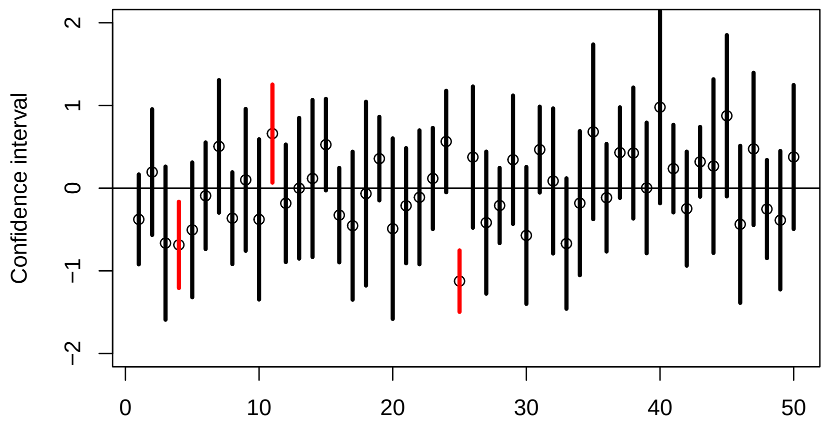
$$P_\theta(\theta \in [L(\mathbf{X}), U(\mathbf{X})]) = 1 - \alpha$$

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

Example 1. 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

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

True mean need not be in the confidence interval always



Example 2. 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

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

R Code

```
################
#####Data#######
################
nn<-1000000
mu<-0
sd=1
x<-rnorm(nn,0,1)
n1<-25
n2<-50
n3<-500
x1<-matrix(x,nrow = n1)
x2<-matrix(x,nrow = n2)
x3<-matrix(x,nrow = n3)

#####Estimation of MEAN#######
##### Unbiased estimator #####
m1<-apply(x1, MARGIN = 2, mean)
m2<-apply(x2, MARGIN = 2, mean)
m3<-apply(x3, MARGIN = 2, mean)
par(mfrow=c(1,1))
tt<-paste("sample size",n1)
hist(m1,probability = T, col="lightblue", main = tt, breaks = 50, xlim=c(-0.7,0.7))
abline(v=mu, col=2, lwd=2)
points(mean(m1),0, col="blue")
tt<-paste("sample size",n2)
```

```r
hist(m2,probability = T, col="lightgreen", main = tt,breaks = 20, xlim=c(-0.7,0.7))
abline(v=mu, col=2, lwd=2)
points(mean(m2),0, col="blue")
tt<-paste("sample size",n3)

hist(m3,probability = T, col="lightgray", main = tt,breaks = 20, xlim=c(-0.7,0.7))
abline(v=mu, col=2, lwd=2)
points(mean(m1),0, col="blue")
plot(density(m1), col="blue", ylim=c(0,10), main='Unbiasedness')
lines(density(m2), col="green")
lines(density(m3), col="black")
abline(v=mu, col=2, lwd=2)

##### Estimation of square of MEAN#########
##### Asymptotically Unbiased estimator ###

par(mfrow=c(1,1))
tt<-paste("sample size",n1)
hist(m1^2,probability = T, col="lightblue", main = tt, breaks = 50, xlim=c(-0.1,0.4))
abline(v=mu, col=2, lwd=2)
points(mean(m1^2),0, col="blue")
tt<-paste("sample size",n2)
hist(m2^2,probability = T, col="lightgreen", main = tt,breaks = 20, xlim=c(-0.1,0.4))
abline(v=mu, col=2, lwd=2)
points(mean(m2^2),0, col="blue")
tt<-paste("sample size",n3)
```

```r
hist(m3^2,probability = T, col="lightgray", main = tt,breaks = 20, xlim=c(-0.1,0.4))
abline(v=mu, col=2, lwd=2)
points(mean(m3^2),0, col="blue")


plot(density(m1^2), xlim=c(-0.05,0.2), main="Asymptotic Unbiasedness")
points(mean(m1^2),0, col="blue", main='consistency')
points(mean(m2^2),0, col="green", main='consistency')
points(mean(m3^2),0, col="black", main='consistency')
abline(v=mu, col=2, lwd=2)


##### Estimation of MEAN#########
##### Consistency ###

n<-750
xsample<-sample(x,n,replace = F)
cummean<-cumsum(xsample)/(1:n)
plot(cummean,type='l',col="gray",ylim=c(-1,1) )
abline(h=mu,col=2,lwd=2,lty=2)
for( i in 2 :10){
xsample<-sample(x,n,replace = F)
cummean<-cumsum(xsample)/(1:n)
lines(cummean,type='l',col="gray")
}
abline(h=mu,col=2,lwd=2,lty=2)
abline(v=25, col="blue")
```

```
abline(v=50, col="green")
abline(v=500, col="black")
lines(-3/sqrt((1:n)))
lines(3/sqrt((1:n)))


par(mfrow=c(1,1))
plot(density(m1), col="blue", ylim=c(0,10), main='Consistency')
lines(density(m2), col="green")
lines(density(m3), col="black")
abline(v=mu, col=2, lwd=2)

#########################
# Method of Moments
# # 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")
```

```r
cat("True sigma=", s, "estimated sigma=", sh,"\n")
###############################

plot(pnorm(q = l,mean = mu,sd = s)~l, type = 'l', col=1, lwd=2, ylab = "CDF", xlab
     ='x')
lines(pnorm(q = l,mean = muh,sd = sh)~l, type = 'l', col=2, lwd=2)
#lines(ecdf(x),col=3, lty=2)
legend("bottomright",legend =  c("True", "Estimated"), col = c(1,2), lwd = c(2,2))

hist(x,probability = T, xlab
     ='x')
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))



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

# MLE finding
p<-seq(0.01,0.99,by = 0.01)
l<-array(0,dim=c(length(p)))
```

```r
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))]

abline(v=mle1)
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))]
abline(v=mle2)

cat("MLE1=",mle1,"MLE2=",mle2,"\n")
```


Department of Mathematics, IIT Kgaragpur
URL: https://sites.google.com/site/buddhanandastat/
E-mail address: bbanerjee@maths.iitkgp.ac.in