ENSAE

STATISTICS 1

SUBJECT 2

The beta distribution with two unknown parameters

using Newton-Raphson method

Author:
XU Kai
DANG ngoc hao

Supervisor:
Xavier
Guillet

October 23, 2017

1 Introduction

Point estimation seeks to get the knowledge of the parameters which yield the knowledge of the entire population. Since the parameters always have a meaningful physical meaning(ex:population mean,population variance etc),good point estimates are in interest.There are several ways to obtain the estimators of the parameters of the distributions,like the moments estimators and the maximum likelihood estimator(MLE). In this article,we focus on the estimation of the two parameters of the beta distribution α and β. With the help of the simulation study, we will compare two methods,one is the method of moments estimators, and the other is the method of maximum likelihood estimator. For the MLE, we will use the Newton-Raphson method to approximate the value of MLE, furthermore, a little improvement will be done to tackle the sensibility of the result to the original iteration point.

2 Content

2.1 The method of moments estimators

The moments estimators of the beta distribution is found by equating the sample moments to the corresponding population moments,and solving the resulting system of simultaneous equations. (George Casella and Roger L Berger 2nd) The moment generating function(mgf) can be used to generate moments as its name suggests.So here we shall use the mgf to explore the moments of the beta distribution.In fact,the n^{th} moment is equal to the n^{th} derivatives of mgf evaluated at t=0.

$$mgf(t) = Ee^{tX}$$

$$\frac{d^2 mgf(t)}{dt^2}|_{t=0} = EX^n e^{tX}|_{t=0} = EX^n$$

So we can deduct the moment of the order n by using the moment generating function as followed:

$$EX^n = \int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)+\Gamma(\beta)} x^{(\alpha-1)}(1-x)^{(\beta-1)} X^n dx$$

clearly we have $\int_0^1 x^{(\alpha+n-1)}(1-x)^{(\beta-1)}dx = \frac{\Gamma(\alpha+n)\Gamma(\beta)}{\Gamma(\alpha+n+\beta)}$ here,so

$$EX^n = \frac{\Gamma(\alpha+n)\Gamma(\alpha+\beta)}{\Gamma(\alpha+n+\beta)\Gamma(\alpha)}$$

Combined with the moments of the first order and second order,we can get

$$E(X) = \frac{\alpha}{\alpha+\beta}$$
$$E(X^2) = \frac{(\alpha+1)\alpha}{(\alpha+\beta+1)(\alpha+\beta)}$$

and we have

$$\begin{aligned} Var(x) &= E(X^2) - E^2(X) \\ &= \frac{(\alpha+1)\alpha}{(\alpha+\beta+1)(\alpha+\beta)} - \frac{\alpha^2}{(\alpha+\beta)^2} \\ &= \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \end{aligned}$$

then use the sample mean, $\overline{X} = \frac{1}{n}\sum X_i$ and sample variance,$S^2 = \frac{1}{n-1}\sum(X_i - \overline{X})^2$, we can obtain estimators of α and β by resolving the set of equations :

$$\begin{cases} \overline{X} &= \frac{\alpha}{\alpha+\beta} \\ S^2 &= \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \end{cases}$$

First,we solve for β

$$\begin{aligned} &\overline{X}(\alpha+\beta) = \alpha \\ \Leftrightarrow &\beta\bar{X} = \alpha(1-\bar{X}) \\ \Leftrightarrow &\beta = \alpha(\frac{1}{\bar{X}} - 1) \end{aligned}$$

Next,we solve for α

$$\begin{aligned} &\alpha\beta = S^2[(\alpha+\beta)^2 + (\alpha+\beta+1)] \\ \Leftrightarrow &(\frac{1}{\bar{X}} - 1)\frac{\bar{X}^2}{S^2} = \frac{\alpha}{\bar{X}} - 1 \\ \Leftrightarrow &\frac{\bar{X}(1-\bar{X})}{S^2} = \frac{\alpha}{\bar{X}} - 1 \\ \Leftrightarrow &\alpha = \bar{X}[\frac{\bar{X}(1-\bar{X})}{S^2} - 1] \end{aligned}$$

So the expression of β in terms of $\bar{X}$ and S^2 is:

$$\beta = (1-\bar{X})(\frac{\bar{X}(1-\bar{X})}{S^2} - 1)$$

Thus, the methode of moments estimates of α and β are:

$$\hat{\alpha}_{MOM} = \bar{X}[\frac{\bar{X}(1-\bar{X})}{S^2} - 1]$$
$$\hat{\beta}_{MOM} = (1-\bar{X})(\frac{\bar{X}(1-\bar{X})}{S^2} - 1)$$

2.2 Maximum likelihood estimators and its application with Newton-Raphson method

The likelihood function of the β function is:

$$\begin{aligned} L(\alpha,\beta|X) &= \prod_{i=1}^{n} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x_i^{\alpha-1}(1-x_i)^{\beta-1} \\ &= [\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}]^n \prod_{i=1}^{n} x_i^{\alpha-1} \prod_{i=1}^{n} (1-x_i)^{\beta-1} \end{aligned}$$

and the log likelihood is:

$$\begin{aligned} logL(\alpha,\beta|X) &= nlog(\Gamma(\alpha+\beta)) - nlog(\Gamma(\alpha)) - nlog(\Gamma(\beta)) \\ &+ (\alpha-1)\sum_{i=1}^{n} log(x_i) + (\beta-1)\sum_{i=1}^{n} log(1-x_i) \end{aligned}$$

To solve MLEs of α and β, we take the derivative of the log likelihood with respect to each parameter and set them equal to zero:

$$\frac{\partial}{\partial\alpha} logL(\alpha,\beta|X) = \frac{n\Gamma'(\alpha+\beta)}{\Gamma(\alpha+\beta)} - \frac{n\Gamma'(\alpha)}{\Gamma(\alpha)} + \sum_{i=1}^{n} log(x_i) = 0$$
$$\frac{\partial}{\partial\beta} logL(\alpha,\beta|X) = \frac{n\Gamma'(\alpha+\beta)}{\Gamma(\alpha+\beta)} - \frac{n\Gamma'(\beta)}{\Gamma(\beta)} + \sum_{i=1}^{n} log(1-x_i) = 0$$

We want to estimate parameters α and β but this system of equations depend on the fraction $\frac{\Gamma'(.)}{\Gamma(.)}$ of α, β and $\alpha+\beta$,so we can't get the close-form solution. Since we can not find a explicit solution of the maximum likelihood estimators,we have to get the estimation of the MLEs with the numerical method.We introduce the Newton-Raphson method to solve the problem.The procedure is as followed: Assuming $X \sim f(x,\theta)$,$L(\theta) = logf(x,\theta)$,the MLEs $\hat{\theta}$ with p dimensions satisfy the following conditions:

$$\dot{L}(\hat{\theta}) = 0, \dot{L}(\theta) = (\frac{\partial L}{\partial \theta_1}, ..., \frac{\partial L}{\partial \theta_p})^T$$

we extend at some initial point θ^0,

$$\dot{L}(\hat{\theta}) = \dot{L}(\theta^0) + \ddot{L}(\theta^0)(\hat{\theta} - \theta^0) + \circ(|\hat{\theta} - \theta^0|)$$

$$\hat{\theta} = \theta^0 - [\ddot{L}(\theta^0)]^{-1}\dot{L}(\theta^0) + remainder$$

and we iterate the value of $\hat{\theta}$ as below:

$$\theta^1 = \theta^0 - [\ddot{L}(\theta^0)]^{-1}\dot{L}(\theta^0)$$

$$.....$$

$$\theta^{i+1} = \theta^i - [\ddot{L}(\theta^i)]^{-1}\dot{L}(\theta^i)$$

until$|\theta^{i+1} - \theta^i| < \epsilon$,in reality,$\epsilon$ is some positive number which is sufficiently small, and we take θ^{i+1} as the approximation of the $\hat{\theta}$. Since the shape of the beta distribution changes largely as the parameters change, we shall introduce various combinations of the parameters to calculate and to compare the estimators.

	1	2	3	4	5	6
α	0.2	0.4	0.5	1	2	2
β	0.5	2	0.5	1	2	6

Table 1: The parameters combinations.

Considering that the Newton-Raphson method is a local method which is sensible to the initial point.In the iteration,we use the estimations obtained by the method of moments as the initial point.We use a simulated sample with the size from 50 to 300 ,after the iterations,we got the results of the moment estimators and the MLEs using Newton-Raphson method,

(α,β)	50	100	200	300
(0.2,0.5)	(0.234,0.469)	(0.205,0.454)	(0.191,0.458)	(0.20,0.457)
(0.4, 2)	(0.488,2.185)	(0.384,2.051)	(0.383,1.989)	(0.381,1.937)
(0.5,0.5)	(0.648,0.62)	(0.526,0.494)	(0.513,0.503)	(0.484,0.478)
(1,1)	(1.156,0.999)	(1.092,0.960)	(1.092,0.970)	(1.02,0.99)
(2,2)	(2.456,2.772)	(2.193,2.297)	(2.014,2.063)	(1.986,1.932)
(2,6)	(2.246,7.04)	(2,6)	(2.011,6.097)	(1.860,5.351)

Table 2: The MLEs estimations.

(α,β)	50	100	200	300
(0.2,0.5)	(0.171,0.363)	(0.177,0.421)	(0.172,0.429)	(0.19,0.443)
(0.4, 2)	(0.410,1.841)	(0.369,1.962)	(0.361,1.863)	(0.378,1.890)
(0.5,0.5)	(0.658,0.654)	(0.51,0.485)	(0.493,0.482)	(0.448,0.447)
(1,1)	(1.337, 1.153)	(1.126,1)	(1.104,0.976)	(0.967,0.962)
(2,2)	(2.467,2.773)	(2.121, 2.224)	(1.947,2.018)	(1.898,1.846)
(2,6)	(2.263,7.112)	(1.967, 5.900)	(1.904,5.800)	(1.813,5.225)

Table 3: The moment estimations.

From the tables,we can see the MLES seems to outperform the moment ones.The value obtained by the two methods are obviously different.
There is one point needed to be mentioned,the Newton-Raphson method relies on the initial point,and the iteration doesn't guarantee increment of the value of the log likelihood function.That means when we choose a 'bad' initial point,it won't be easy to get the approximation of the MLEs,even cause the divergence of result of the algorithm.So we make a little improvement here to ensure the increment of the value of the log likelihood function during the iteration.We just need add a factorλ_i to the original iteration to make sure that$L(\theta^{i+1}) = L(\theta^i - \lambda_i[\ddot{L}(\theta^i)]^{-1}\dot{L}(\theta^i)) \geq L(\theta^i)$,and luckily the$\lambda_i$ is near 0,so we don't need to make too much effort to search it.
We shall give a proof of the existence of the λ_i , Since $-\ddot{L}(\theta)$equals to the Fisher information matrix in the case of the beta distribution,so $-\ddot{L}(\theta)$is positive semi-definite.We can use this property to prove the existence of the λ_i.
If $-\ddot{L}(\theta)$positive definite,$-\dot{L}(\theta) \neq 0$,then there exist a$\lambda \geq 0$ to ensure $L(\tilde{\theta}) = L(\theta - \lambda[\ddot{L}(\theta)]^{-1}\dot{L}(\theta)) \geq L(\theta)$.

Proof: f(λ)=$L(\theta-\lambda[\ddot{L}(\theta)]^{-1}\dot{L}(\theta))$, according to the differential rule, $L(\tilde{\theta})-L(\theta)$=f($\lambda$)-f(0)=$\dot{f}(0)\lambda$+$\alpha\lambda$, here $\alpha\lambda$ represents the remainder, obviously when $\lambda\rightarrow$ 0,$\alpha\rightarrow$0,$\dot{f}(0)$can be presented as below:

$$\dot{f}(0)=\frac{dL(\theta-\lambda[\ddot{L}(\theta)]^{-1}\dot{L}(\theta))}{d\lambda}=\dot{L}(\theta)^T[-\ddot{L}(\theta)]^{-1}\dot{L}(\theta)>0$$

thus when λ sufficiently small, we can guarantee that $\dot{f}(0)+\alpha>0$, then we obtain that

$$L(\tilde{\theta})-L(\theta)=[\dot{f}(0)+\alpha]>0$$

Since we have proved the existence of the λ,we can change the iteration into the procedure like this
First, we take the initial θ^0,choose the λ_0 so that

$$L(\theta^0-\lambda^0[\ddot{L}(\theta^0)]^{-1}\dot{L}(\theta^0))=maxL(\theta^0-\lambda[\ddot{L}(\theta^0)]^{-1}\dot{L}(\theta^0))>L(\theta^0).$$

Second, we take θ^1=$\theta^0-\lambda^0[\ddot{L}(\theta^0)]^{-1}\dot{L}(\theta^0)$,do the same thing as before.
We continue until $|\theta^{i+1}-\theta^i|<\epsilon$,$\epsilon$ is some chosen positive number which is sufficiently small. In fact, taking into account that λ is rather near 0,we can just divide the space between 0 and 0.1 to search the approximation of the best λ for each iteration. The further the initial point to the 'real' value, the bigger the problem of the traditional Newton-Raphson method is.
To show the difference of these two methods, we introduce some 'bad' case to compare, we can see that the traditional one will have the problem of being not convergent and costing more times.

(α,β)	Initial point	Newton-Raphson	Times	Improved one with λ	Times
(2,6)	(6,9)	(-87588,10061)	48127	(2.578,5.525)	7021
(0.4, 2)	(2,3)	(5.02,-179.609)	1673	(0.488,2.180)	299
(0.5,0.5)	(3,5)	(13.15,-47.63)	68	(0.651,0.624)	825

Table 4: The 'bad' case comparison of two methods with sample size =50.

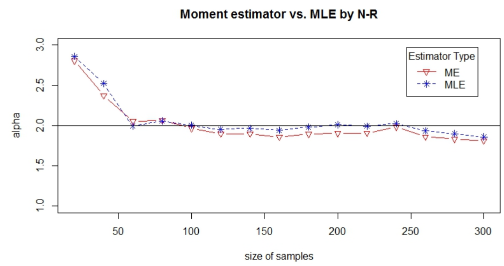

Figure 1: The comparison on α

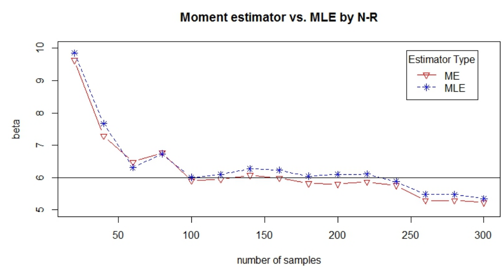

Figure 2: The comparison on β

2.3 Visual comparison of MLEs and moments estimators

In this section,we are going to compare visually the performance facing the different sizes of the sample. We take the parameter combination $(\alpha,\beta)=(2,6)$as the example. Here, we still use the traditional Newton-Raphson method to compare because we use the estimations of the moment estimators as the reliable initial point for the iteration. In that case,we need not to worry about the problem of the traditional Newton-Raphson method for the moment.The size of the sample varies from 20 to 300.

From these pictures, we can see that in the small size samples, the MLEs are not too different from the moment estimators,and they are more unstable than the moment estimators. But as the size of the sample increases, the

MLEs quickly converge,and they perform better than what the moments estimators do. Judging from the result of the calculation, the values of the α are more stable and ideal,but the estimation of β still needs to be improved.

2.4 Wald Statistic

To carry out the Wald test on the hypothesis$\alpha = \beta$,we should first construct the Wald statistic. We can derive a Wald statistic from the score function. Assuming X$\sim f(x,\theta)$,$\theta$is p dimension and we are only interested in part of it which is p_1 dimension,noted as θ_{p_1}, and the rest parameter is p_2 dimension,$p_1+p_2 = p$. Then we consider the hypothesis test:

$$H_0 : \theta_1 = \theta_{10}, H_1 : \theta_1 \neq \theta_{10}$$

We rewrite $\dot{L}(\theta)$and$\ddot{L}(\theta)$into block matrix according to p_0 and p_1:

$$\dot{L}(\theta) = \begin{pmatrix} \dot{L}_1(\theta) \\ \dot{L}_2(\theta) \end{pmatrix} \ddot{L}(\theta) = \begin{pmatrix} L_{11}(\theta) & L_{12}(\theta) \\ L_{21}(\theta) & L_{22}(\theta) \end{pmatrix} \ddot{L}^{-1}(\theta) = \begin{pmatrix} L^{11}(\theta) & L^{12}(\theta) \\ L^{21}(\theta) & L^{22}(\theta) \end{pmatrix}$$

$$\hat{\theta} = \begin{pmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{pmatrix} \hat{\theta}_0 = \begin{pmatrix} \theta_{10} \\ \tilde{\theta}_2 \end{pmatrix}$$

$\tilde{\theta}_2$ is the MLE under the θ_{10},so $\dot{L}_2(\tilde{\theta}_2) = 0$.
extend $\dot{L}(\hat{\theta}) = 0$ at $\hat{\theta}_0$:

$$\dot{L}(\hat{\theta}) = \dot{L}(\hat{\theta}_0) + \ddot{L}(\hat{\theta}_0)(\hat{\theta} - \hat{\theta}_0) + remainder = 0$$

then we can get the following

$$\begin{pmatrix} \hat{\theta}_1 - \hat{\theta}_{10} \\ \hat{\theta}_2 - \tilde{\theta}_2 \end{pmatrix} = - \begin{pmatrix} L^{11}(\hat{\theta}_0) & L^{12}(\hat{\theta}_0) \\ L^{21}(\hat{\theta}_0) & L^{22}(\hat{\theta}_0) \end{pmatrix} \begin{pmatrix} \dot{L}_1(\hat{\theta}_0) \\ 0 \end{pmatrix}$$

According to the property of the score function,we have $\frac{1}{\sqrt{n}}\dot{L}_1(\hat{\theta}) \xrightarrow{L} N(0, i(\theta))$. So we can construct a Wald statistic based on the following form,

$$WD(\theta_1 0) = (\hat{\theta}_1 - \hat{\theta}_{10})^T[-L^{11}(\hat{\theta})]^{-1}(\hat{\theta}_1 - \hat{\theta}_{10}) \rightarrow \chi^2(p_1)$$

We now use the $\alpha - \beta$ as the parameter interested θ_1 and to be calculated,the rest parameter remains β.The chi-squared distribution is to be used,and the degree of freedom here is 1 decided by the size of the parameter interested.We do the what we have done before using the Newton-Raphson method to get

the approximation of the $\widehat{\alpha - \beta}$ and$-L^{11}(\widehat{\alpha - \beta})$. Thus we can calculate the p-values of the samples with the different size. The Wald Statistic is constructed based on the MLEs and the inverse matrix of the second derivative matrix.
Besides, we can use the parametric bootstrap to estimate the p-value without the assistance of the asymptotic distribution. The procedure is as below:

a)Under the MLEs ,reconstruct a sampling distribution which obeys the null hypothesis, then resample from this sampling distribution $\hat{F}_n$

b)Calculate the MLE $\widehat{\alpha - \beta}$ of the each resampled samples

c)Repeat a) and b) B times,B should be big enough to obey the grand number law,then we obtain $(\widehat{\alpha - \beta})_1$,...$(\widehat{\alpha - \beta})_B$

d)Sort the $(\widehat{\alpha - \beta})_1$,...$(\widehat{\alpha - \beta})_B$,and determine the position of the original $(\widehat{\alpha - \beta})$ among the $(\widehat{\alpha - \beta})_1$,...$(\widehat{\alpha - \beta})_B$,then calculate the corresponding two-side p-value by using $\frac{\sum_{i=k}^{B} I(|((\widehat{\alpha-\beta})_i|>|(\widehat{\alpha-\beta})|)}{B}$,I($\bullet$) is the indicator function. The results are as followed:

	asympotic distribution			non parametric bootstrap		
real value	100	200	300	100	200	300
(2,2)	0.829571	0.610137	0.822464	0.573	0.692	0.606
(0.4,2)	5.20E-07	7.12E-13	0	0	0	0
(1,1)	0.4963	0.2665	0.588	0.220	0.114	0.699
(2,6)	1.58E-09	0	0	0	0	0
(0.5,0.5)	0.739	0.332	0.276	0.636	0.826	0.864

Table 5: The P-value of Wald test under the different sizes

We can observe from the table that facing the same size, the asymptotic distribution works well, it rejects the null hypothesis when the parameters are unequal under the 5 percent level and it accepts the null hypothesis with the real parameters are equal. So does the bootstrap, without the aid of the specific distribution,it gets the same result as the asymptotic one do.
In general, the Wald test has an effect in determining the equality of two parameters. Judging from the result under the 5 percent level, the asymptotic

one is as good as the bootstrap one while the later one doesn't require a specific distribution. As the two methods don't differ to much, so for the beta distribution, Wald statistic can be used to determine whether to accept the null hypothesis although it is a asymptotic case.

2.5 Conclusion

In this article,we have explored the moments estimators and the maximum log likelihood estimators. We compare two methods and make a little improvement in Newton-Raphson iteration to lessen the sensibility of this algorithm to the original point. We also compare visually these two methods, and we finds out that the moments estimators can be rather approachable to the real parameters in the most cases, while the MLEs outperform in the bigger samples. Further,we construct a Wald statistic to determine whether the two parameters are equal. We have tried the method of asymptotic distribution and the parametric bootstrap to get the p-value of different size. Comparing their results, we find that the Wald Statistic is fairly a reliable indicator to determine whether to accept the null hypothesis or not.

3 Annex

```
#Q1 and Q2
betapara<-matrix(c(0.2,0.4,0.5,1,2,2,0.5,2,0.5,1,2,6),nrow=6,ncol=2)#in
theta_hatm<-matrix(0,nrow=6,ncol=2)#initialize the moments estimators
#MNS calculate the moments
MNS<-function(x){
xmean<-mean(x)
#sample variance
xvar<-var(x)
#compute moment estimator alpha and beta
alpham=xmean*(xmean*(1-xmean)/xvar-1)
betam=(1-xmean)*(xmean*(1-xmean)/xvar-1)
theta_hatm<-c(alpham,betam)
return(theta_hatm)
}
#calculate the order 1 derivative
DL<-function(a,x){
```

```
dalpha<-digamma(sum(a))-digamma(a[1])+mean(log(x))
dbeta<-digamma(sum(a))-digamma(a[2])+mean(log(1-x))
dl<-c(dalpha,dbeta)
return(dl)
}
#calculate the order 2 derivative
DDL<-function(a){
ddalpha<-trigamma(sum(a))-trigamma(a[1])
ddbeta<-trigamma(sum(a))-trigamma(a[2])
ddab<-trigamma(sum(a))
ddl<-matrix(c(ddalpha,ddab,ddab,ddbeta),nrow=2)
return(ddl)
}
ns<-6 #define the number of the groups to be compared
mns<-matrix(rep(0,2*ns),nrow=2,ncol=ns*6)#initialize the vector of mome
mlens<-matrix(rep(0,2*ns),nrow=2,ncol=ns*6)#initialize the vector of mo
tn<-50*c(1:ns)#use 50 as one unit of the number of the samples
for(num in 1:6){
for (i in 1:ns){
set.seed(50)
x<-rbeta(tn[i], betapara[num,1],betapara[num,2] )#create the simulated
mns[,ns*num-ns+i]<-MNS(x)
theta_hat<-MNS(x) #initialize the original point
ddL<-DDL(theta_hat)
dL<-DL(theta_hat,x)
n<-0
epi=1 #initialize the episilon
#update the theta_hat
while(epi>10^(-8)){
#record the old theta_hat
old<-theta_hat
#update theta_hat
theta_hat<-theta_hat-solve(ddL)%*%dL
#update dL
dL<-DL(theta_hat,x)
#updata ddL
ddL<-DDL(theta_hat)
epi<-sum((theta_hat-old)^2)
```

```
n<-n+1
}
mlens[,ns*num-ns+i]=theta_hat
}
}
#Q3:the graphs for the different numbers of sample,for the simplicity,ju
#the function calculate the moment estimators
MNS<-function(x){
xmean<-mean(x)
#sample variance
xvar<-var(x)
#compute moment estimator alpha and beta
alpham=xmean*(xmean*(1-xmean)/xvar-1)
betam=(1-xmean)*(xmean*(1-xmean)/xvar-1)
thetachapeaum<-c(alpham,betam)
return(thetachapeaum)
}
ns<-15 #define the number of the groups to be compared
mns<-matrix(rep(0,2*ns),nrow=2,ncol=ns)#initialize the vector of moment
mlens<-matrix(rep(0,2*ns),nrow=2,ncol=ns)#initialize the vector of mome
tn<-20*c(1:ns)#use 50 as one unit of the number of the samples
#the iteration of MLEs
for (i in 1:ns){
set.seed(50)
x<-rbeta(tn[i], 6, 9)
mns[,i]<-MNS(x)
thetachapeau<-MNS(x) #initialize the original point
ddL<-DDL(thetachapeau)
dL<-DL(thetachapeau,x)
n<-0
epi=1
while(epi>10^(-8)){
#record the old thetachapeau
old<-thetachapeau
#update thetachapeau
thetachapeau<-thetachapeau-solve(ddL)%*%dL
#update dL
dL<-DL(thetachapeau,x)
```

```
#updata ddL
ddL<-DDL(thetachapeau)
epi<-sum((thetachapeau-old)^2)
n<-n+1
}
mlens[,i]=thetachapeau
}
#the plot precedure
plot(tn, mns[1,], type="b",
  pch=6, lty=1, col="red", ylim=c(1, 3),
  main="Moment estimator vs. MLE by N-R",
  xlab="number of samples", ylab="alpha")
lines(tn, mlens[1,], type="b",
pch=8, lty=2, col="blue")
abline(h=2)
legend("topright", inset=.05, title="Estimator Type", c("ME","MLE"),
lty=c(1, 2), pch=c(15, 17), col=c("red", "blue"))
plot(tn, mns[2,], type="b",
  pch=6, lty=1, col="red", ylim=c(5, 10),
  main="Moment estimator vs. MLE by N-R",
  xlab="number of samples", ylab="beta")
lines(tn, mlens[2,], type="b",
pch=8, lty=2, col="blue")
abline(h=6)
legend("topright", inset=.05, title="Estimator Type", c("ME","MLE"),
lty=c(1, 2), pch=c(15, 17), col=c("red", "blue"))
#pvalue and pvalueb is the pvalue according to chisq distribution and b
#define the dL ddL for alpha-beta,use alpha-beta and beta as the parame
estimate
DLw<-function(a,x){
dab<-digamma(sum(a))-digamma(sum(a)+a[2])-mean(log(x))
dbeta<-digamma(sum(a))+digamma(a[2])-2*digamma(sum(a)+a[2])-mean(log(1-x
dl<-c(dab,dbeta)
return(dl)
}
DDLw<-function(a){
ddamb<-trigamma(sum(a))-trigamma(sum(a)+a[2])
ddbeta<-trigamma(sum(a))+trigamma(a[2])-4*trigamma(sum(a)+a[2])
```

```
ddab<-trigamma(sum(a))-2*trigamma(sum(a)+a[2])
ddl<-matrix(c(ddamb,ddab,ddab,ddbeta),nrow=2)
return(ddl)
}
ns<-15 #define the number of the groups to be compared
wald<-rep(0,ns)#initialize the vector of wald test statistics
pvalue<-rep(0,ns)
mlensw<-matrix(rep(0,2*ns),nrow=2,ncol=ns)#initialize mle
tn<-20*c(1:ns)#use 20 as one unit of the number of the samples
for (i in 1:ns){
set.seed(50)
x<-rbeta(tn[i],0.5, 0.5) #change parameter here
#initialize the original point
th<-MNS(x)
th[1]<-th[1]-th[2]
thw<-th
ddL<-DDLw(thw)
dL<-DLw(thw,x)
n<-0
epi=1
while(epi>10^(-8)){
#record the old thw
old<-thw
#update thw
thw<-thw-solve(ddL)%*%dL
#update dL
dL<-DLw(thw,x)
#updata ddL
ddL<-DDLw(thw)
epi<-sum((thw-old)^2)
n<-n+1
}
mlensw[,i]=thw
iddL<-solve(tn[i]*ddL)
#calculate the wald satistic
wald[i]<-mlensw[1,i]^2/iddL[1,1]
#calculate the pvalue
pvalue[i]<-2*pchisq(wald[i],1)*(pchisq(wald[i],1)<0.5)+2*(1-pchisq(wald
```

```
}
# parametric bootstrap
mlensb<-c(0,0)#initialize mle
B=1000
pvalueb<-rep(0,ns)
for (i in 1:ns){
theta1<-rep(0,B)
for(j in 1:B){
set.seed(30+10*j)
xb<-rbeta(tn[i],mlensw[2,i] ,mlensw[2,i])#resampling from sampling distr
th<-MNS(xb)
th[1]<-th[1]-th[2]
thw<-th
ddL<-DDLw(thw)
dL<-DLw(thw,xb)
epi=1
while(epi>10^(-8)){
#record the old thw
old<-thw
#update thw
thw<-thw-solve(ddL,tol=1e-20)%*%dL
#update dL
dL<-DLw(thw,xb)
#updata ddL
ddL<-DDLw(thw)
epi<-sum((thw-old)^2)
}
mlensb=thw
iddL<-solve(tn[i]*ddL,tol=1e-20)
theta1[j]<-mlensb[1]
}
p<-sum(abs(theta1)>abs(mlensw[1,i]))/B#calculate the p value
pvalueb[i]<-p
}
```

4 Reference

[1] Casella, George, and Roger L. Berger. Statistical inference.

Vol. 2. Pacific Grove, CA: Duxbury, 2002 62−64
[2] Wei Bocheng Parametric statistics tutorialHigher Education Press, 2006,ISBN 7040200546 214−216
[3] Wasserman L. All of nonparametric statistics [M]. Springer Science & Business Media, 2006. 24−26