Bootstrap Project

Yunjun Xia

2018/6/8

Project Description

There are 3 parts in this project: the first one is to test whether the true mean of independent and identically distributed data is equal to some particular value; the second part is to compute the bootstrap pivotal confidence interval for the mean of the sample data; the third part is to compute the bootstrap studentized pivotal confidence interval for the mean of the sample data.

Part 1

In the setting of independt and identically distributed data with unknown mean but known variance, I use a bootstrap method to carry out the hypothesis test that the true mean equals some particular value. The method of bootstrap uses resampling to approximate the distribution of the test statistic under the null hypothesis in order to estimate the quantile of the null distribution that is required for the test.

Assume that we have X1,X2,X3…Xn samples of size n iid from normal distribution with unknown mean and variance 1.
We want to test that the mean of the sample is equal to 5.
We have the hypothesis test:

H0: true mean is equal to 5
H1: true mean is not equal to 5

```r
# Let n=100, alpha-level=0.05
# Let the original sample be:
SampleOri<-rnorm(n=100,mean=4.8,sd=1)
# Bootstrap test of mean for normal distribution with known variance=1
bootMeanTest<-function(X,mu0,alpha=0.05,B=1000){
  n<-length(X)
  sampleMean<-mean(X)
  TestStatistic<-numeric(B)
  for(i in 1:B){
    #generate samples from null distribution
    sampleX<-rnorm(n,mean=mu0,sd=1)
    TestStatistic[i]<-abs(mean(sampleX)-mu0)
  }
  CriticalValue<-quantile(TestStatistic,probs=1-alpha)
  if(abs(sampleMean-mu0)>CriticalValue){return("We reject H0")}
  if(abs(sampleMean-mu0)<=CriticalValue){return("We fail to reject H0")}
}

# Apply the test to the original sample
bootMeanTest(SampleOri,5)
```

```
## [1] "We fail to reject H0"
```

Conclusion of Part 1
In this case, we choose an original sample data set of size 100 from Normal distribution of mean=4.8 and variance=1, and we can see the result is that we reject H0. That is, the mean of the original sample by the bootstrap test is not equal to 5.

Now we apply our function to test multiple sets of random sample data all with mean=4.9 and variance=1. There are three groups of data sets. The first one has sample size 100 for each set; the second one has 500; the third one has 1000. Then we calculate the probability that the function return "We reject H0" which is supposed to be the true answer. Each of the the three groups has 100 sets of random sample data.

```
ProbList<-numeric(3)
j<-0
count<-0
for(n in c(100,500,1000)){
  j=j+1
  for(i in 1:100){
    Sample<-rnorm(n,mean=4.9,sd=1)
    if(bootMeanTest(Sample,5)=="We reject H0"){
      count=count+1
    }
  }
  ProbList[j]<-count/100
  count<-0
}
ProbList
```

```
## [1] 0.08 0.61 0.85
```

Now, we can observe that as the sample size increases, the accuracy for our bootstrap method to test the true mean increases.

Part 2: Computing bootstrap pivotal confidence interval for mean

The general formula for bootstrap pivotal confidence interval is $(2\hat{\theta} - \hat{\theta}((1-\frac{\alpha}{2})bootQuantile), 2\hat{\theta} - \hat{\theta}(\frac{\alpha}{2}bootQuantile))$

```
# returns bootstrap pivotal confidence interval for mean of the data
mean.bootCI<-function(X,alpha=0.05,B=1000){
  Meanori<-mean(X) #mean of the original sample
  Meanb<-numeric(B)
  #bootstrap process
  for(i in 1:B){
    sampleX<-sample(X,replace=TRUE)
    Meanb[i]<-mean(sampleX) #compute mean of the ith bootstrap sample
  }
  #use formula to get confidence interval
  lower<-2*Meanori-quantile(Meanb,1-alpha/2)
  upper<-2*Meanori-quantile(Meanb,alpha/2)
  return(c(lower,upper))
}

# Now apply the function to the original sample we generated in Part 1
mean.bootCI(SampleOri)
```

```
##    97.5%     2.5%
## 4.611899 4.993950
```

Part 3: Computing bootstrap studentized pivotal confidence interval for mean

```r
# returns bootstrap studentized pivotal confidence interval for mean of the data
mean.bootCI.student<-function(X,alpha=0.05,B=1000){
 C<-500
 Meanori<-mean(X)
 Meanb<-numeric(B)
 Meanbbar<-numeric(B)
 SEb<-numeric(B)
 T<-numeric(B)
 Meanbc<-numeric(C)
 for(i in 1:B){
  #generate samples from empirical distribution of original sample
  sampleX<-sample(X,replace=TRUE)
  Meanb[i]<-mean(sampleX)
  # We want to compute the standard error for mean of each bootstrap sample
  # So we have a nested bootstrap(For the second bootstrap we choose a large integer C=500)
  for(j in 1:C){
   #generate samples from empirical distribution of the bth first boot
   sampleXc<-sample(sampleX,replace=TRUE)
   Meanbc[j]<-mean(sampleXc)
  }
  Meanbbar[i]<-sum(Meanbc)/C
  SEb[i]<-sqrt(sum((Meanbc-Meanbbar[i]/C)^2)/(C-1))
  # compute the t-ratio
  T[i]<-(Meanb[i]-Meanori)/SEb[i]
 }
 SEboot<-sd(Meanb)
 lower<-Meanori-quantile(T,1-alpha/2)*SEboot
 upper<-Meanori-quantile(T,alpha/2)*SEboot
 return(c(lower,upper))
}

# Now apply the function to the original sample we generated in Part 1
mean.bootCI.student(SampleOri)
```

```
##    97.5%     2.5%
## 4.799719 4.807338
```

Conclusion of Part 2 and Part 3

For our original sample from Normal distribution with mean=4.8 and variance=1, we find that the bootstrap studentized pivotalconfidence interval is more accurate with a smaller confidence interval range than the bootstrap pivotal confidence interval, both for testing the mean.