LOOBIC VS BIC

Zheng Yuan

2019/10/15

This simulation study mainly compares the performance of LOOBIC and BIC under the circumstance that there is no true model in the options.

```r
loocv = function(fit) {

  h = lm.influence(fit)$h
  mean((residuals(fit)/(1-h))^2)

}

cvbic = function(fit) {

  dim(fit$model)[1]*loocv(fit)+(fit$rank)*log(dim(fit$model)[1])

}
```

Traditional BIC criterion value

```r
mse <- function(object) {

  mean(residuals(object)^2)
}

bic = function(fit) {
  dim(fit$model)[1]*mse(fit)+(fit$rank)*log(dim(fit$model)[1])
}
```

Define data generator

The true model is non linear: $y = 2x_0 + 9x_1^2 + 4x_3$

```r
def <- defData(varname = "x0", dist = "nonrandom", formula = 1)%>%

defData(,varname = "x1", dist="uniform",formula = "10;20")%>%

defData(,varname = "x2", dist="uniform",formula = "0;3")%>%

defData(,varname = "x3", dist="uniform",formula = "0;5")%>%

defData(,varname = "x4", dist="uniform",formula = "5;10")%>%

defData(, varname = "y1", formula = "2*x0+9*x1^2+4*x3", variance = 1)
```

Test dataset

```r
dt1 <- genData(1000, def)

library(caret)
```

```
## Loading required package: lattice
```

```
## Loading required package: ggplot2
```

```r
md1<-c()
md2<-c()
md<-c()
for(i in 1:1000){

dt <- genData(1000, def)

fit1 <- lm(y1 ~ x3, data = dt)
fit2 <- lm(y1 ~ x2+x3, data = dt)
fit3 <- lm(y1 ~ x1+x3, data = dt)
fit4 <- lm(y1 ~ x4+x3, data = dt)
fit5 <- lm(y1 ~ x1+x3+x4, data = dt)
fit6 <- lm(y1 ~ x2+x3+x4, data = dt)
fit7 <- lm(y1 ~ x1+x2+x3, data = dt)
fit8 <- lm(y1 ~ x1+x2+x3+x4, data = dt)

models = list (fit1,fit2,fit3,fit4,fit5,fit6,fit7,fit8)

md1<-c(md1,which.min(c(cvbic(fit1),cvbic(fit2),cvbic(fit3),cvbic(fit4),cvbic(fit5),cvbic(fit6),cvbic(fi

md2<-c(md2,which.min(c(bic(fit1),bic(fit2),bic(fit3),bic(fit4),bic(fit5),bic(fit6),bic(fit7),bic(fit8))
md<-c(md,ifelse(sum(dt1$y1-unlist(predict(models[md1[i]],dt1)))^2/1000>sum(dt1$y1-unlist(predict(models
}

table(md1)/1000## empirical probability for each model
```

```
## md1
##     3     5     7     8
## 0.705 0.121 0.146 0.028
```

```r
table(md2)/1000
```

```
## md2
##     3     5     7     8
## 0.002 0.032 0.031 0.935
```

```r
table(md)/1000## mean prediction error
```

```
## md
##     0     1
## 0.535 0.465
```

52.3% of 1000 times where LOOBIC selects the model with the smaller prediction error than BIC methods.