PS1_Estimating_Auctions

Fangxu Duan

Import the bids dataset.

```
library(readr)
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.6.3
```

```
library("kdensity")
```

```
## Warning: package 'kdensity' was built under R version 3.6.3
```

```
bids <- read_csv("bids.csv", col_names = c("bid"), cols(bid = col_double()) )
```

Question 1. Estimate the density of bids:

a. an assumed normal distribution
b. a Gaussian kernel
c. Epanechnikov kernel

```
normal_density = dnorm(bids$bid, mean = mean(bids$bid), sd = sd(bids$bid), log = FALSE)
gau_kernel = density(bids$bid, bw = "nrd0", kernel = c("gaussian"))
epa_kernel = density(bids$bid, bw = "nrd0", kernel = c("epanechnikov"))
```

Question 2. Least-Squares Cross-Validation

```
set.seed(1)

X = bids$bid

J<- function(h){
  fhat=Vectorize(function(x) density(X,from=x,to=x,n=1,bw=h, kernel ="epanechnikov" )$y)
  fhati=Vectorize(function(i) density(X[-i],from=X[i],to=X[i],n=1,bw=h, kernel ="epanechnikov" )$y)
  F=fhati(1:length(X))
  return(integrate(function(x) fhat(x)^2,-(10)^10,(10)^10)$value-2*mean(F))
}

vx=seq(.05,.2,by=.01)
vy=Vectorize(J)(vx)
df=data.frame(vx,vy)


(myopt <- optimize(J,interval=c(.01,.8)) )
```

```
## $minimum
## [1] 0.03394882
##
## $objective
## [1] -0.7308726
```

```
bw_cv <- myopt$minimum
cv_epa_kernel = density(bids$bid, bw = bw_cv, kernel = c("epanechnikov"))
```

Question 3. Plot the estimated density

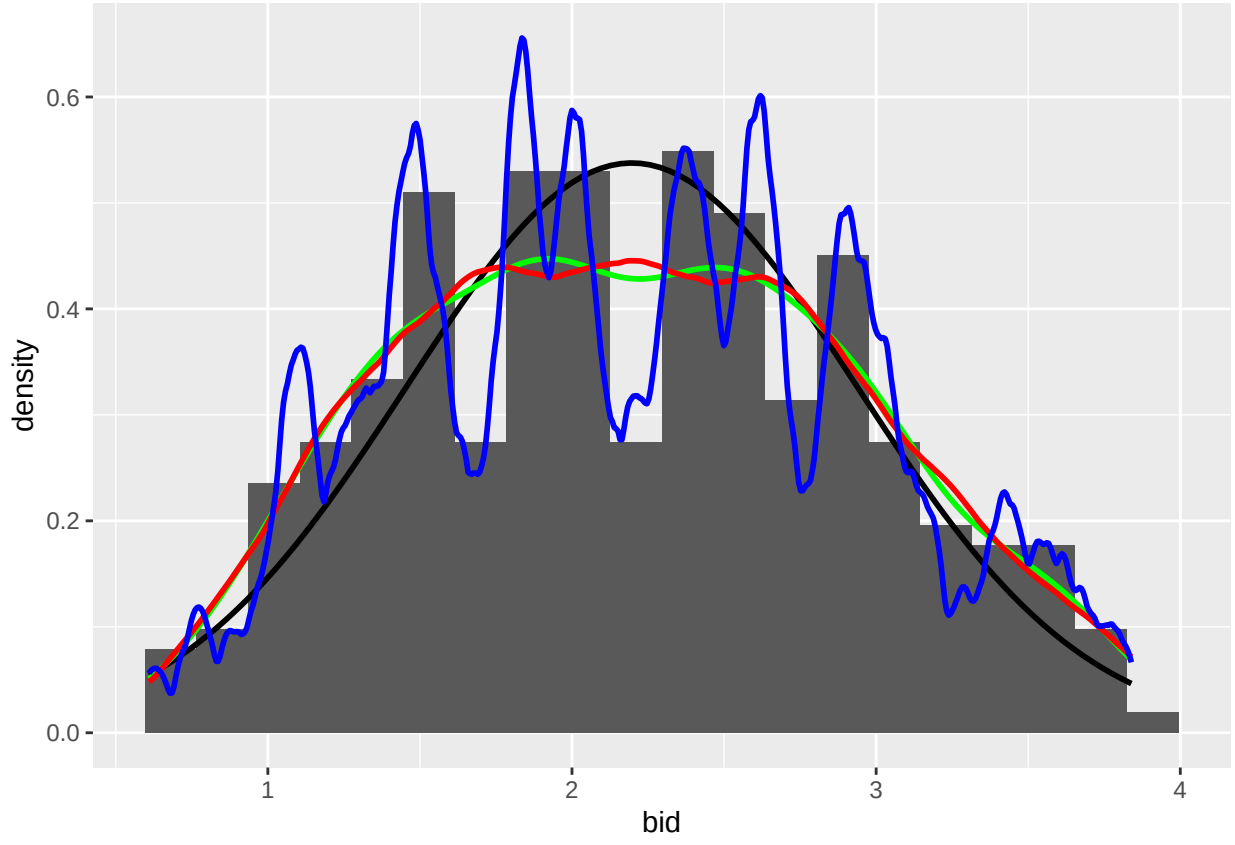
It appears that the least-squares cross-validation bandwidth in Epanechnikov Kernel fits the best.

However, it tends to overfit the data.

Both Guassian kernel and Epanechnikov with Silverman's plug-in are goood in my opinion.

```
data <- as.data.frame(bids)

ggplot(data,aes(bid))  + geom_histogram(aes(y = stat(density)), bins = 20) +
  stat_function(fun = dnorm, args = list(mean = mean(bids$bid), sd = sd(bids$bid)), lwd = 1)+
  geom_line(stat="density", bw = "nrd0", kernel = c("gaussian"), lwd = 1, col= 'green')+
  geom_line(stat="density",bw="nrd0",kernel = c("epanechnikov"), lwd = 1, col = 'red') +
  geom_line(stat="density", bw=bw_cv,kernel = c("epanechnikov"), lwd = 1, col = 'blue')
```



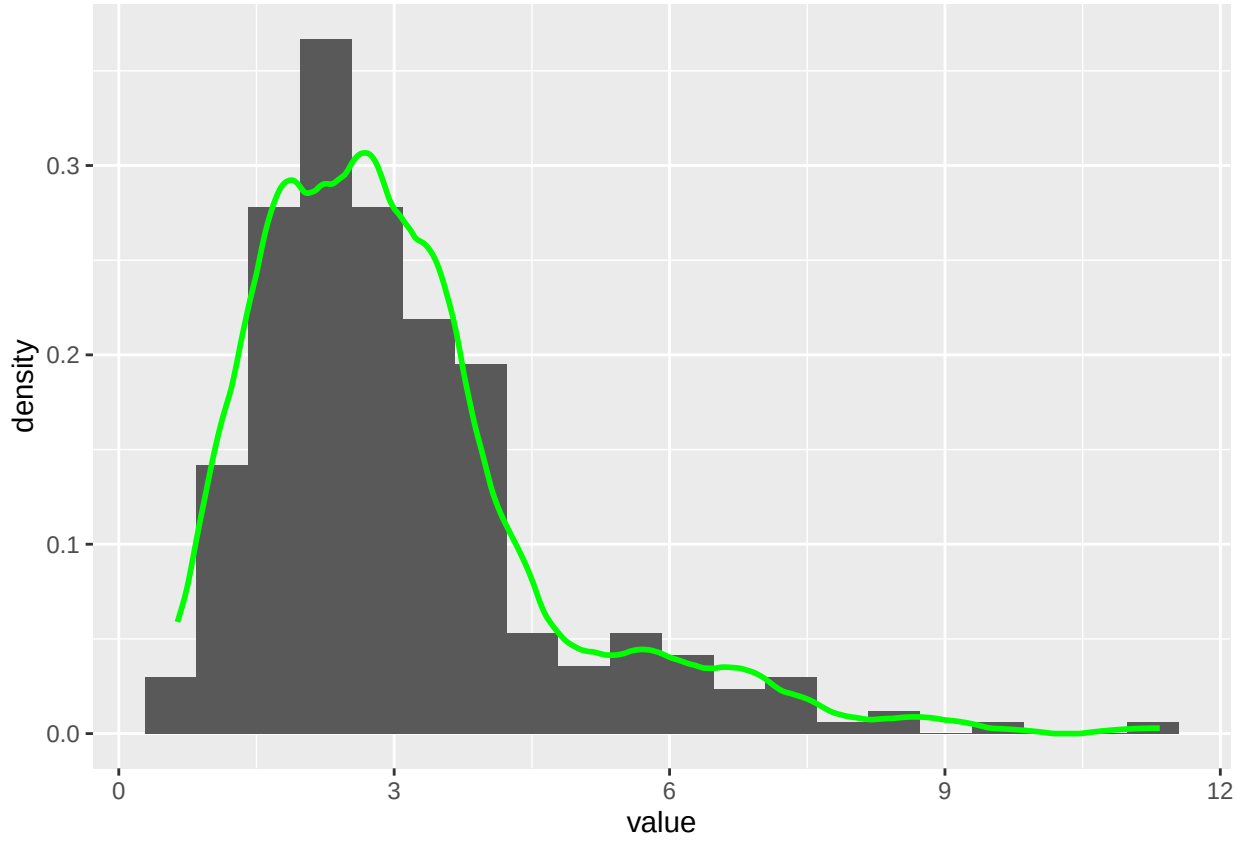
Question 4. Estimated the valuation v

```
library("kdensity")
G_B <- ecdf(bids$bid)
g_b <- kdensity(bids$bid,bw = bw_cv , kernel = "epanechnikov",normalized = FALSE)
```

```
x <- bids$bid
value = x + G_B(x)/(2*g_b(x))
```

Question 5. Estimate the kernel density of the esimated v

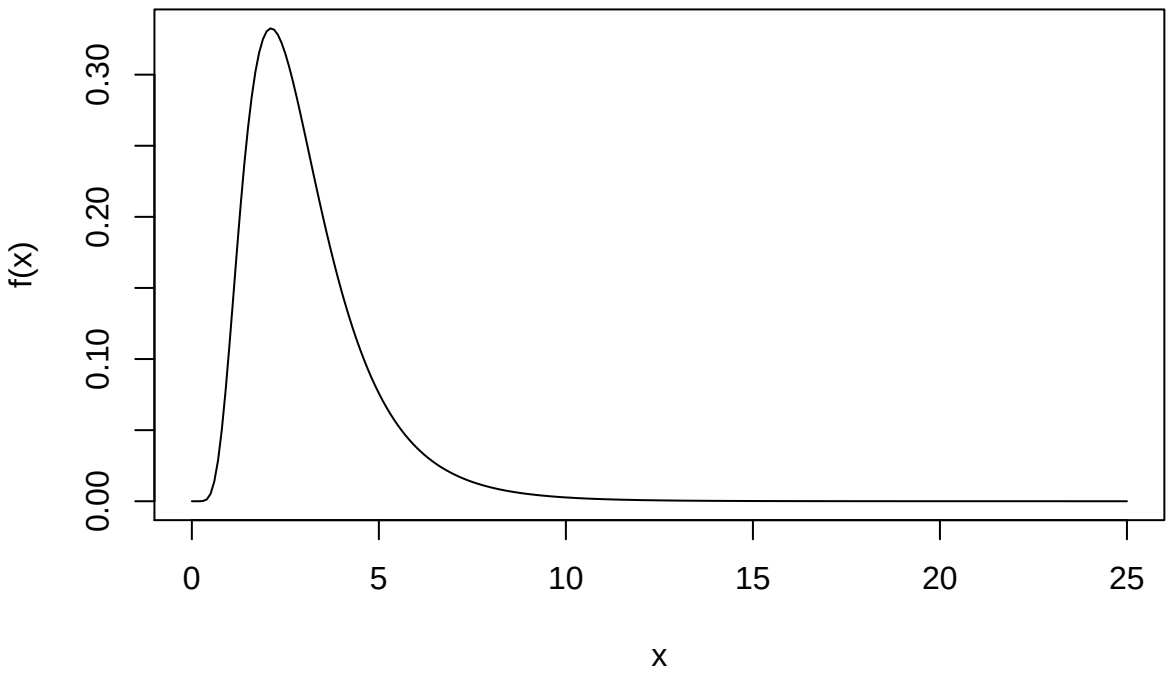
```
kd_value = density(value, bw = "nrd0", kernel = c("epanechnikov"))
data2 <- as.data.frame(value)
ggplot(data2,aes(value))  + geom_histogram(aes(y = stat(density)),bins = 20) +
  geom_line(stat="density", bw = "nrd0", kernel = c("epanechnikov"), lwd = 1, col= 'green')
```



Question 6 Guesss the valuation distribution

My best guess is that it is a lognormal distribution with mean 1 and standard divation 0.5 (see the plot below).

```
grid = seq(0,25,.1)
plot(grid,dlnorm(grid,1,0.5),type="l",xlab="x",ylab="f(x)")
```


f(x)
0.00
0.10
0.20
0.30
0
5
10
15
20
25
x