Problem Set 1

Joseph Jourden

1/28/2020

Part 1.

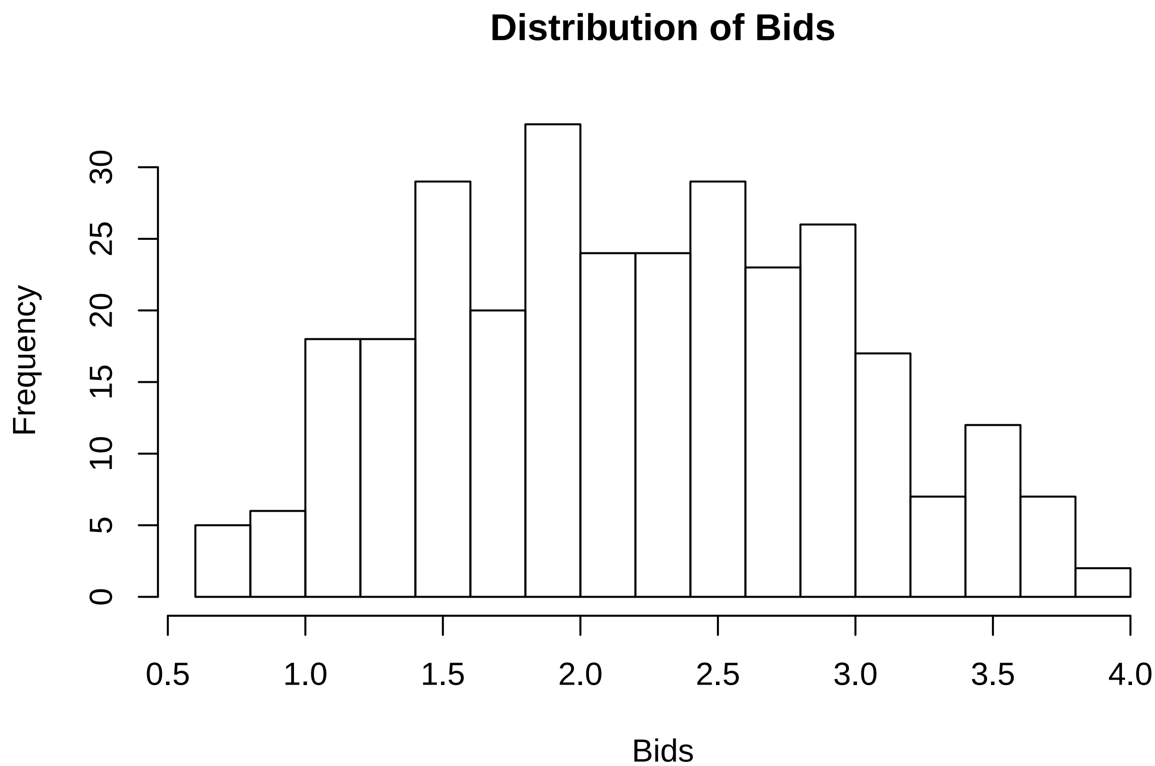
First import the bids dataset. The first observation was read as a column name, so I added it back as an observation, and I renamed the column. Then I can estimate the mean and sample variance for the assumed normal density. I compute the other densities as well, specifying the Silverman plug-in bandwidth estimate with bw="SJ".

```
setwd("C:/Users/Joe/github/empirical-methods")

library(kedd)
library(cubature)
```

```
## Warning: package 'cubature' was built under R version 3.6.2
```

```
dt <- read.delim("ProblemSets/pset1/bids1.csv")
dt <- rbind(2.9679,dt)
names(dt)<- "bids"
  # assumed normal density
hist(dt$bids, xlab="Bids", main="Distribution of Bids", breaks=20)
```




```
mean1 <- mean(dt$bids); sd1 <-sd(dt$bids)
x <- seq(-6,6,by=.01)
assumeN <- data.frame(x=x,y=dnorm(x, mean = mean1,sd = sd1, log = FALSE))
  # density estimates with Gaussian and Epanechnikov kernels
normal <- density(dt$bids, kernel= "gaussian", bw = "SJ", adjust = 1)
epan <- density(dt$bids, kernel= "epanechnikov", bw = "SJ", adjust = 1)
```

Part 2.

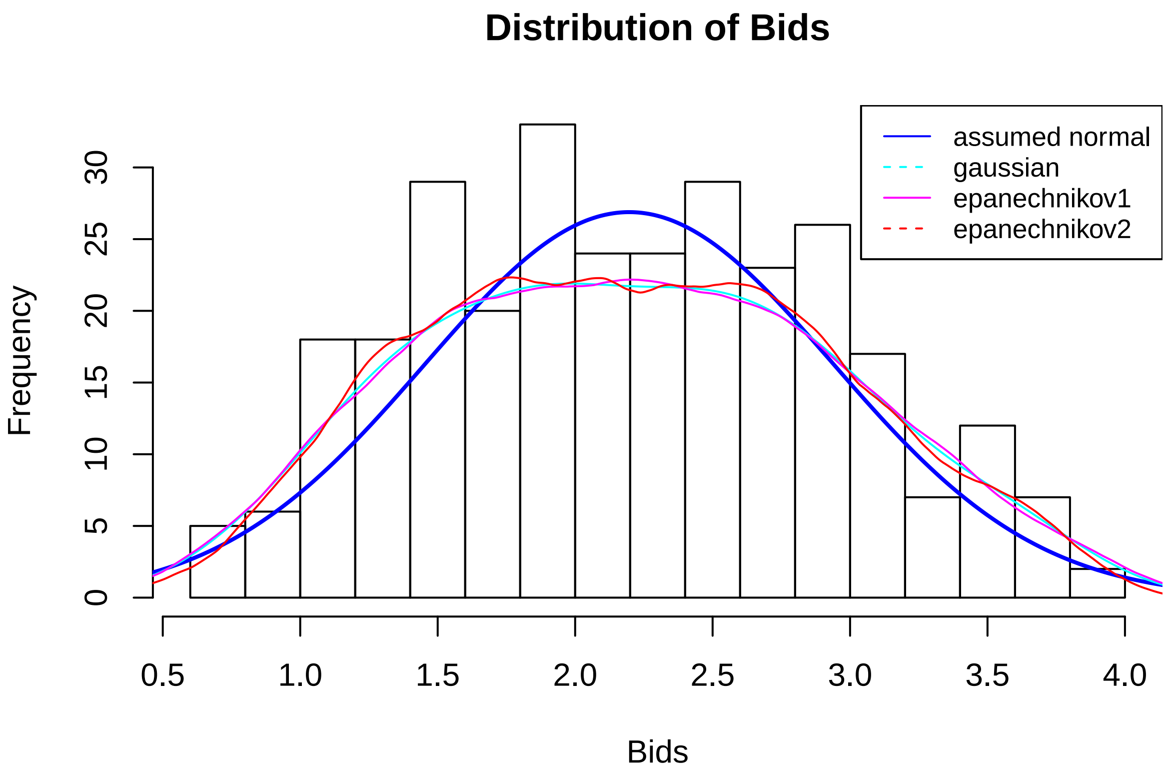
Now I use the "kedd" package to find the best bandwidth, using maximum likelihood LOOCV, and I find the density with that bandwidth.

```
h_best <- h.mlcv(dt$bids, kernel = "epanechnikov")$h
epan_best <- density(dt$bids, kernel= "epanechnikov", bw = h_best, adjust = 1)
```

Part 3.

I overlay the density estimates on top of the histogram. I choose around 20 bins, because after considering other numbers of bins, this choice seems to capture the features of the dataset well, also considering the number of observations. I have to scale up the distribution plots to fit in the same plot as the histogram.

```
hist(dt$bids, xlab="Bids", main="Distribution of Bids", breaks=20)
lines(assumeN$x,50*assumeN$y,col="blue",lwd=2)
lines(normal$x, 50*normal$y, col="cyan")
lines(epan$x, 50*epan$y, col="magenta")
lines(epan_best$x, 50*epan_best$y, col="red")
legend(x="topright", inset=0, legend=c("assumed normal", "gaussian", "epanechnikov1", "epanechnikov2"),
```



Part 4.

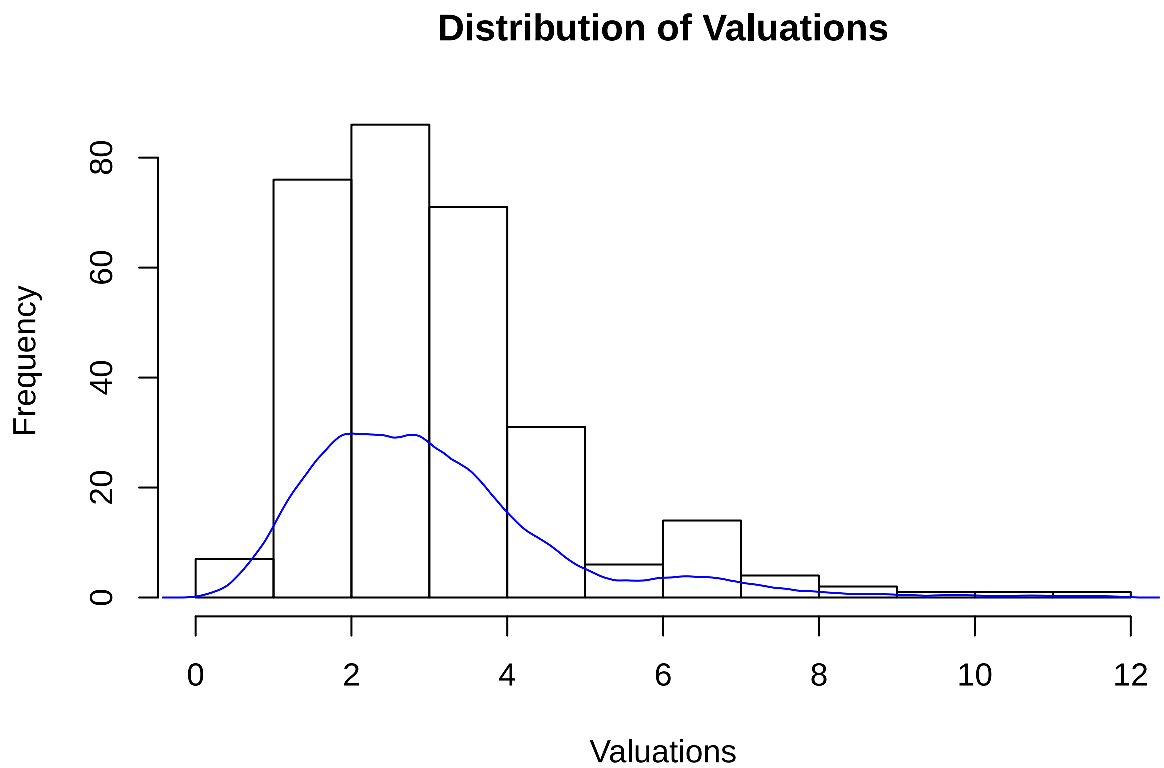
First I compute an approximate g, the PDF of bids, using the cross-validated density. Then, I use the package “cubature” to integrate g and find G, the CDF for the cross-validated Epanechnikov kernel. Finally, I use the formula from GPV to estimate valuations for each bidder.

```
g <- approxfun(epan_best$x, epan_best$y, yleft=0, yright=0)
G <- function(x) cubintegrate(g, -Inf, x, method = "pcubature", nVec = 1L)$integral
valuation_hat <- function(b) b + G(b)/((3-1)*g(b))
valuations <- sapply(dt$bids, FUN=valuation_hat)
```

Part 5.

Use Epanechnikov kernel with plug-in bandwidth to estimate valuation distribution, then plot it overlaid on a histogram of valuations.

```
val_distr <- density(valuations, kernel= "epanechnikov", bw = "SJ", adjust = 1)
hist(valuations,xlab="Valuations", main="Distribution of Valuations")
lines(val_distr$x, 100*val_distr$y, col="blue")
```



Part 6.

The valuations appear to fit a chi-square distribution because of the shape of the distribution, with a right tail and no negative values. I test goodness of fit of a chi-square distribution, and estimate its parameter, getting an estimate of 299 degrees of freedom.

```r
chisq.test(valuations)
```

```
## Warning in chisq.test(valuations): Chi-squared approximation may be incorrect
```

```
##
##  Chi-squared test for given probabilities
##
## data:  valuations
## X-squared = 263.33, df = 299, p-value = 0.9324
```