STAT 403 HW1

Chongyi Xu

April 4, 2018

Questions

1. Let $X_1, \cdots, X_n$ be IID random points from $Exp(1/\beta)$. The PDF of $Exp(1/\beta)$ is

$$p(x) = \frac{1}{\beta} e^{-x/\beta}$$

for $x \geq 0$. Let $\bar{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i$ be the sample average. Let β be the parameter of interest that we want to estimate.

(a) What is the bias and variance of using the sample average $\bar{X_n}$ as the estimator of β?

The population mean will be $\mu = \mathbb{E}(X_n) = \beta$. Since we are using the sample average as the estimator, $\hat{\beta} = \bar{X_n}$, then

$$\mathbf{bias}(\hat{\beta_n}) = \mu - \mu = 0,\ \mathbf{Var}(\hat{\beta_n}) = \frac{\beta^2}{n}$$

(b) What is the mean square error of using $\bar{X_n}$ as the estimator of β?

$$\mathbf{MSE}(\hat{\beta_n}) = \mathbf{Var}(\hat{\beta_n}) + \mathbf{bias}^2(\hat{\beta_n}) = \mathbf{Var}(\hat{\beta_n}) = \frac{\beta^2}{n}$$

(c) Does $\bar{X_n}$ converges to β? Why?

From the calculation above, we can see that $\mathbf{bias}(\hat{\beta_n}) = 0$ for all n and $\mathbf{Var}(\hat{\beta_n}) = \frac{\beta^2}{n}$ converge to 0 as $n \to \infty$. Therefore, the estimator $\hat{\beta_n} = \bar{X_n}$ converges to β.

(d) Now consider a new estimator $\hat{\beta_n} = a * \bar{X_n}$, where $a \in \mathbb{R}$ is a real number. What is the mean square error of $\hat{\beta_n}$?

Similar to what we did in part(a), and the variance will not change since the estimator just gets scaled.

$$\mathbf{bias}(\hat{\beta_n}) = a * \mu - \mu = (a-1)\mu,\ \mathbf{Var}(\hat{\beta_n}) = \frac{\beta^2}{n}$$

Therefore, the mean square error will be

$$\mathbf{MSE}(\hat{\beta_n}) = \mathbf{Var}(\hat{\beta_n}) + \mathbf{bias}^2(\hat{\beta_n}) = \frac{\beta^2}{n} + (a-1)^2\mu^2$$

(e) To mimimize the mean square error, which value of a should we take? Does this give us an estimator that has a lower mean square error than the sample mean $\bar{X_n}$?

From the equation of MSE above, we can see that the minimum value of MSE according to a happens at

$$\frac{d}{da}\mathbf{MSE}(\hat{\beta_n}) = 2\mu^2(a-1) = 0 \Rightarrow a = 1$$

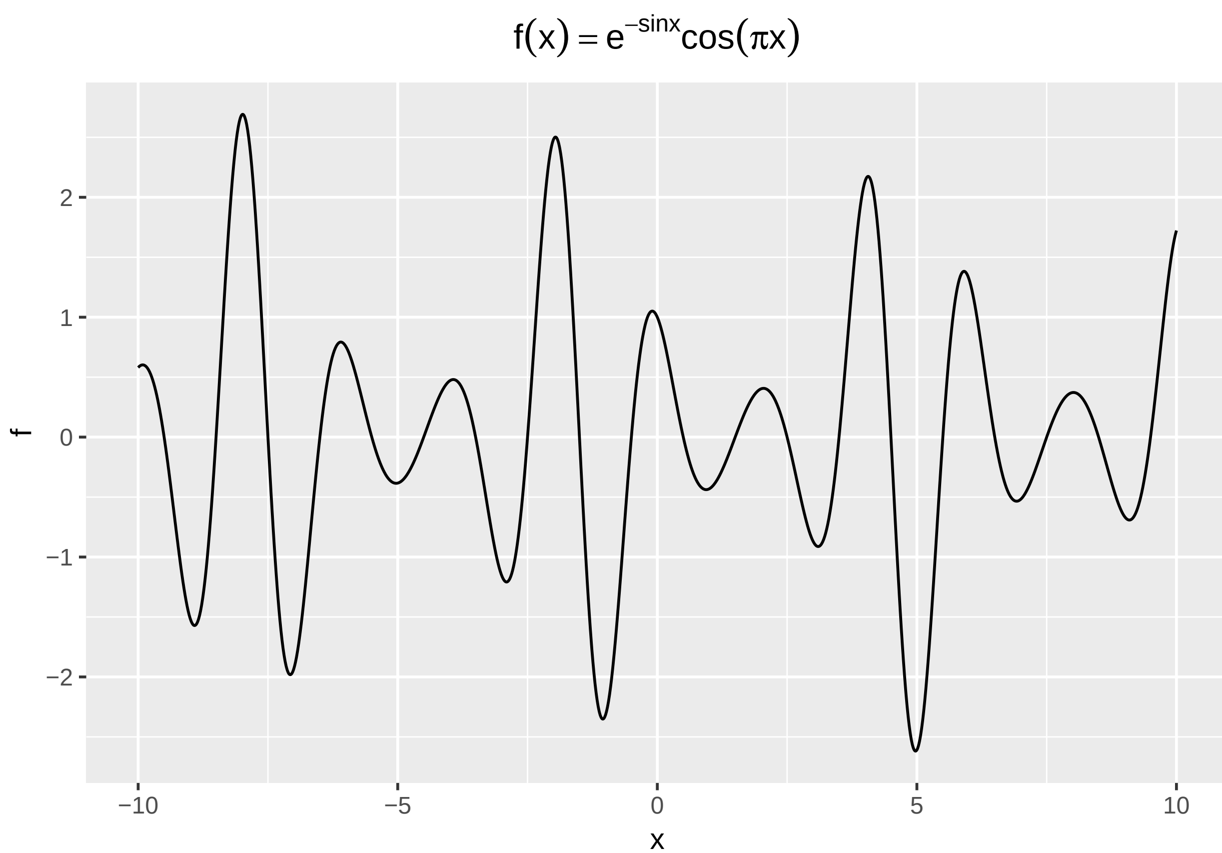
So the value of a to minimize the mean square error is 1, which is the same as the estimator using sample mean $\bar{X_n}$.

2. Use R to plot the function $f(x) = e^{-sin(x)} * cos(\pi x)$ for $x \in [-10, 10]$

```r
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.4.4
```

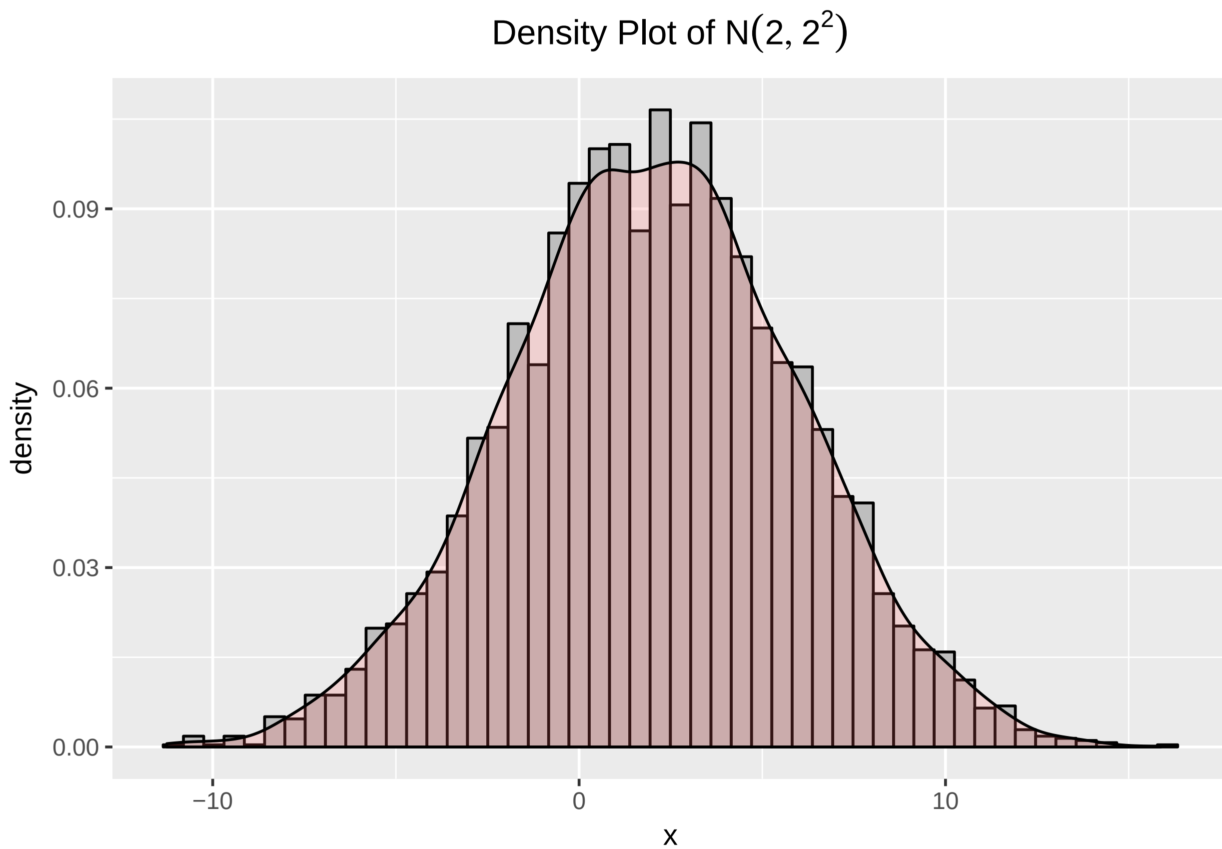
```r
x<- seq(from=-10, to=10, by=0.01)
f <- exp(-sin(x)) * cos(pi * x)
qplot(x, f, geom='line') +
     ggtitle(expression(f(x)==e^{-sinx}*cos(pi*x))) +
  theme(plot.title = element_text(hjust = 0.5))
```



3. Use R to generate 5000 data points from $N(2, 2^2)$. Plot the density histogram with 50 bins. Attach a density curve of $N(2, 2^2)$ to the histogram.

```r
dat <- data.frame(x=rnorm(n=5000, mean=2, sd=4))

ggplot(dat, aes(x=x)) + geom_histogram(aes(y=..density..),
                              bins=50,
                              color="black",
                              fill="grey") +
  geom_density(alpha=.2, fill="#FF6666") +
  ggtitle(expression(paste("Density Plot of ", N(2, 2^2)))) +
  theme(plot.title = element_text(hjust = 0.5))
```



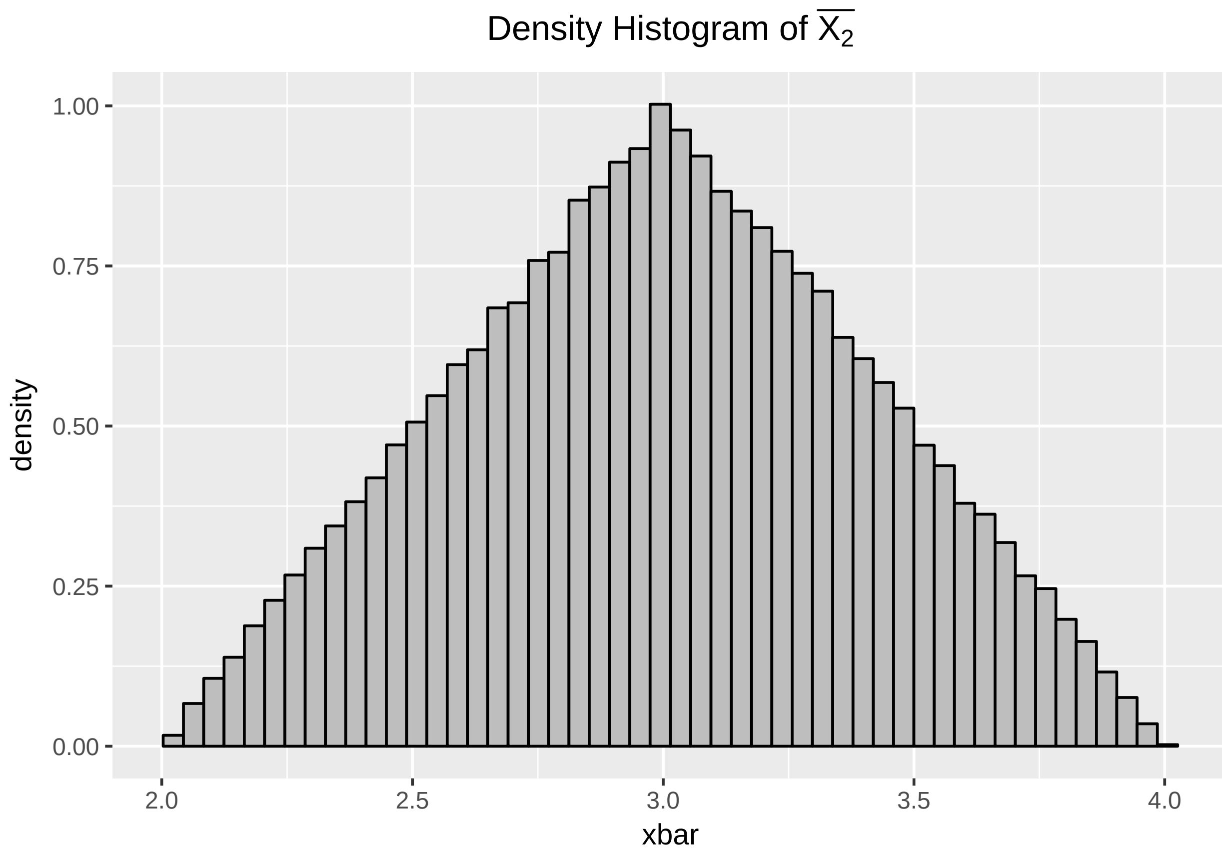
4. Let X_1, X_2 $Uni[2,4]$. Let $\bar{X}_2$ be the sample mean.

(a) Use the for loop to generate at least 100000 realizations of $\bar{X}_2$ and plot the corresponding histogram.

```
xbar <- vector(length=100000)
for (n in 1:100000) {
  x1 <- runif(n=1, min=2, max=4)
  x2 <- runif(n=1, min=2, max=4)
  xbar[n] <- mean(c(x1, x2))
}

dat <- data.frame(xbar=xbar)

p <- ggplot(dat, aes(x=xbar)) +
  geom_histogram(aes(y=..density..), colour="black", fill="grey", bins=50) +
  ggtitle(expression(paste("Density Histogram of ", bar(X[2])))) +
  theme(plot.title = element_text(hjust = 0.5))
p
```

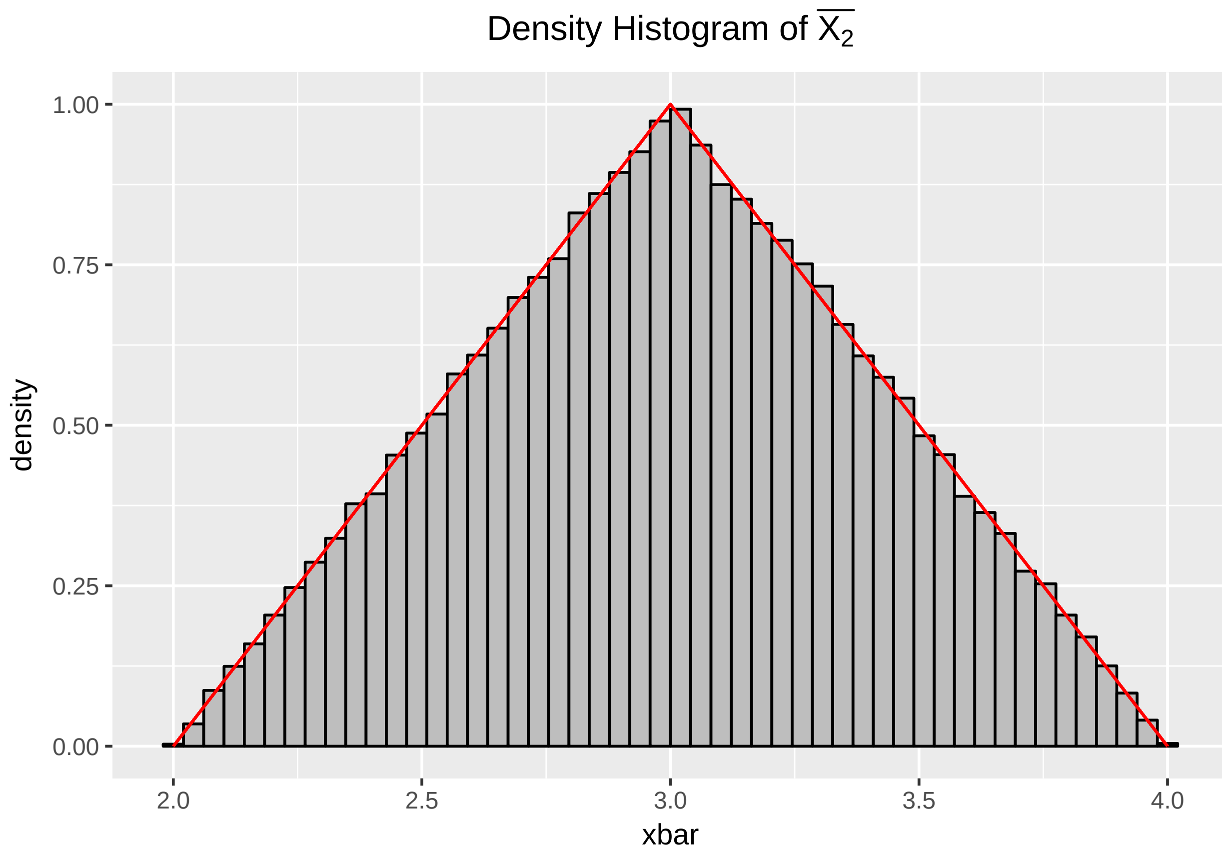


(b) The density curve of $\bar{X}_2$ is

$$p(x) = \begin{cases} x-2, \text{when } 2\leq \text{ x} \leq 3 \\ 4-x, \text{when } 3\leq \text{ x} \leq 4 \\ 0, \text{otherwise} \end{cases}$$

Attach the corresponding density curve of $\bar{X}_2$ to the histogram.

```
size=nrow(dat)
px <- vector(length=size)
x <- seq(from=2, to=4, by=(4-2)/size)
for (i in 1:length(x)) {
  if (x[i] <= 3) {
    px[i] <- x[i] - 2
  } else {
    px[i] <- 4 - x[i]
  }
}
p + geom_line(aes(x=x[1:length(x)-1],y=px[1:length(px)-1]), color="red")
```



Density Histogram of $\overline{X_2}$
density
1.00
0.75
0.50
0.25
0.00
2.0
2.5
3.0
3.5
4.0
xbar