APSTA Week 10 Exercises

1. Unbiased estimators (Theory)

Consider a random sample $X_1, \dots, X_n$ from a uniform distribution in the interval $-\theta, \theta$, where θ is an unknwon parameter. You are interested in estimating the values of θ.
a. Show that

$$\hat{\Theta} = \frac{2}{n}(|X_1| + |X_2| + \cdots + |X_n|)$$

is an unbiased estimator for θ. *Hint*: you may need to use the *change of variable formula* (cfr. Chapter 7 of the book).
b. Consider instead the problem of estimating θ^2. Show that

$$T = \frac{3}{n}(X_1^2 + X_1^2 + \cdots + X_n^2)$$

is an unbiased estimator for θ^2
c. Is $\sqrt{T}$ an unbiased estimator for θ? If not, discuss whether it has positive or negative bias.

Solution:
a.
To check whether $\hat{\Theta}$ is unbiased, we need to find its expectation.

$$\begin{aligned} \mathrm{E}\left[\hat{\Theta}\right] &= \mathrm{E}\left[\frac{2}{n}(|X_1| + |X_2| + \cdots + |X_n|)\right] \\ &= \mathrm{E}\left[\frac{2}{n}|X_1| + \frac{2}{n}|X_2| + \cdots + \frac{2}{n}|X_n|\right] \end{aligned}$$

I use linearity of expectations:

$$\begin{aligned} &= \mathrm{E}\left[\frac{2}{n}|X_1|\right] + \mathrm{E}\left[\frac{2}{n}|X_2|\right] + \cdots + \mathrm{E}\left[\frac{2}{n}|X_2|\right] \\ &= \frac{2}{n}(\mathrm{E}[|X_1|] + \mathrm{E}[|X_2|] + \cdots + \mathrm{E}[|X_n|]) \end{aligned}$$

All the X_i are from the same $U(-\theta, \theta)$ distribution. We want the expectation of $|X_i|$, so we need to use the change of variable formula. The probability density function $f(x)$ of the uniform distribution is $f(x) = 1/(\beta - \alpha)$ where α is the lower bound, here $-\theta$, and β the upper bound, here θ.

$$\begin{aligned} \mathrm{E}[|X_i|] &= \int_{-\theta}^{\theta} |x| f(x) \, \mathrm{d}x \\ &= \int_{-\theta}^{\theta} |x| \frac{1}{\theta - (-\theta)} \, \mathrm{d}x \\ &= \int_{-\theta}^{\theta} \frac{|x|}{2\theta} \, \mathrm{d}x \\ &= \frac{1}{2\theta} \int_{-\theta}^{\theta} |x| \, \mathrm{d}x \end{aligned}$$

This integral is a bit tricky, since it includes an absolute value. We can split it up into two cases, the case where x is negative, and the case where it isn't:

$$\begin{aligned}
\int |x| \text{ d}x &= \begin{cases} \int x \text{ d}x & \text{if } x \geq 0 \\ \int -x \text{ d}x & \text{if } x < 0 \end{cases} \\
&= \begin{cases} \frac{x^2}{2} + C & \text{if } x \geq 0 \\ \frac{-x^2}{2} + C & \text{if } x < 0 \end{cases} \\
&= \begin{cases} \frac{x \cdot |x|}{2} + C & \text{if } x \geq 0 \\ \frac{x \cdot |x|}{2} + C & \text{if } x < 0 \end{cases} \\
&= \frac{x|x|}{2} + C
\end{aligned}$$

We insert this into the previous integral.

$$\begin{aligned}
\frac{1}{2\theta} \int_{-\theta}^{\theta} |x| \text{ d}x &= \frac{1}{2\theta} \left[\frac{x|x|}{2} \right]_{x=-\theta}^{x=\theta} \\
&= \frac{1}{2\theta} \left(\left(\frac{\theta|\theta|}{2} \right) - \left(\frac{-\theta| - \theta|}{2} \right) \right)
\end{aligned}$$

θ may be negative, so I can't remove the absolute function, but I can remove the negative sign from the $| - \theta|$, as it will end up positive no matter what.

$$\begin{aligned}
\text{E}[|X_i|] &= \frac{1}{2\theta} \left(\left(\frac{\theta|\theta|}{2} \right) - \left(\frac{-\theta|\theta|}{2} \right) \right) \\
&= \frac{1}{2\theta} \left(\frac{\theta|\theta|}{2} + \frac{\theta|\theta|}{2} \right) \\
&= \frac{1}{2\theta} \left(\frac{2\theta|\theta|}{2} \right) \\
&= \frac{\theta|\theta|}{2\theta} \\
&= \frac{|\theta|}{2}
\end{aligned}$$

Now that we know the expectation of X_i, we can insert it into the expectation for $\hat{\Theta}$ function.

$$\begin{aligned}
\text{E}[\hat{\Theta}] &= \frac{2}{n} (\text{E}[|X_1|] + \text{E}[|X_2|] + \cdots + \text{E}[|X_n|]) \\
&= \frac{2}{n} \left(\frac{|\theta|}{2} + \frac{|\theta|}{2} + \cdots + \frac{|\theta|}{2} \right) \\
&= \frac{2}{n} \cdot \frac{|\theta|}{2} + \frac{2}{n} \cdot \frac{|\theta|}{2} + \cdots + \frac{2}{n} \cdot \frac{|\theta|}{2}) \\
&= \frac{2|\theta|}{2n} + \frac{2|\theta|}{2n} + \cdots + \frac{2|\theta|}{2n} \\
&= \frac{|\theta|}{n} + \frac{|\theta|}{n} + \cdots + \frac{|\theta|}{n} \\
&= n \cdot \frac{|\theta|}{n} \\
&= |\theta|
\end{aligned}$$

Now technically we don't know if it actually gives θ, because if θ is negative, it gives $-\theta$. This however, implies that $-\theta > \theta$, which is not allowed for the uniform distribution, as the lower bound *has* to be lower than the upper bound, meaning θ cannot be negative. Thus we know that $\theta > -\theta$, and $\text{E}[\hat{\Theta}] = \theta$.

b. Now we ant to prove that

$$T = \frac{3}{n}(X_1^2 + X_2^2 + \cdots + X_n^2)$$

is an unbiased estimator for θ^2.

$$\mathrm{E}[T] = \mathrm{E}[\frac{3}{n}(X_1^2 + X_2^2 + \cdots + X_n^2)]$$

Use linearity of expectations to write

$$\begin{aligned}
\mathrm{E}[T] &= \mathrm{E}[\frac{3}{n}(X_1^2 + X_2^2 + \cdots + X_n^2)] \\
&= \frac{3}{n}(\mathrm{E}[X_1^2] + \mathrm{E}[X_2^2] + \cdots + \mathrm{E}[X_n^2])
\end{aligned}$$

Now we find the expectation of X_i^2, again using the change of variable formula.

$$\begin{aligned}
\mathrm{E}[X_i^2] &= \int_{-\theta}^{\theta} x^2 \frac{1}{2\theta}\ \mathrm{d}x \\
&= \frac{1}{2\theta}\int_{-\theta}^{\theta} x^2\ \mathrm{d}x \\
&= \frac{1}{2\theta}\left[\frac{x^3}{3}\right]_{-\theta}^{\theta} \\
&= \frac{1}{2\theta}\left(\frac{\theta^3}{3} + \frac{\theta^3}{3}\right) \\
&= \frac{1}{2\theta}\cdot\frac{2\theta^3}{3} \\
&= \frac{\theta^2}{3}
\end{aligned}$$

Inserting in the previous equation...

$$\begin{aligned}
\mathrm{E}[T] &= \frac{3}{n}(\mathrm{E}[X_1^2] + \mathrm{E}[X_2^2] + \cdots + \mathrm{E}[X_n^2]) \\
&= \frac{3}{n}(\frac{\theta^2}{3} + \frac{\theta^2}{3} + \cdots + \frac{\theta^2}{3}) \\
&= \frac{3}{n}\cdot\frac{n\theta^2}{3} \\
&= \theta^2
\end{aligned}$$

So it is unbiased.

c. $\sqrt{T}$ is biased, as unbiasedness does not carry over in the square root operation, but I'm too tired to prove it right now.
I would assume it has a negative bias.

2. Maximum likelihood estimator for geometric random variables (Theory)

The geometric random variable, as presented in the textbook, has the following probability mass function

$$\Pr[X = k] = (1-p)^{k-1}\cdot p$$

which can be described as the probability of requiring k trials to obtain the first success in a sequence of Bernoulli trials. For this random variable, we have seen that the maximum likelihood estimator for the parameter p is

$$\hat{p} = \frac{n}{\sum_{i=0}^{n} x_i} = \frac{1}{x}$$

However, in some contexts[1] a slightly different definition of geometric random variable is used:

$$\Pr[X = k] = (1-p)^k \cdot p$$

This second formulation can be described as the probability of experiencing k consecutive failures before the first success.

We shall see, with this exercise, that this small change leads to a different maximum likelihood estimator for p!

a. Derive the loglikelihood function $\ell(p)$.

b. Compute the derivative $\ell'(p)$ of the loglikelihood function.

c. Show that the maximum likelihood estimator for p is

$$\hat{p} = \frac{n}{n + \sum_{i=0}^{n} x_i}$$

Therefore, *pay attention* to the distribution you are dealing with, always read carefully the definitions and the documentation.

Solution:
a. The distribution is defined by

$$\mathrm{P}(X = k) = (1-p)^k \cdot p = f(x)$$

The likelihood function of the estimator p is then

$$\begin{aligned}
L(p) &= p_p(x_1)p_p(x_2)\dots p_p(x_n) \\
&= ((1-p)^{x_1} \cdot p)((1-p)^{x_2} \cdot p)\dots((1-p)^{x_n} \cdot p) \\
&= (1-p)^{x_1}p(1-p)^{x_2}p\dots(1-p)^{x_n}p \\
&= p^n(1-p)^{x_1}(1-p)^{x_2}\dots(1-p)^{x_n} \\
&= p^n(1-p)^{x_1+x_2+\dots+x_n}
\end{aligned}$$

The loglikelihood is just the log-transformed likelihood:

$$\begin{aligned}
\ell(p) &= \ln L(p) \\
&= \ln\left(p^n(1-p)^{x_1+x_2+\dots+x_n}\right) \\
&= n\ln(p) + (x_1 + x_2 + \dots + x_n)\ln(1-p) \\
&= n\ln(p) + \left(\sum_{i=0}^{n} x_i\right)\ln(1-p)
\end{aligned}$$

[1] Including the R implementation of the geometric random distribution.

b. Compute derivative:

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}p}[\ell(p)] &= \frac{\mathrm{d}}{\mathrm{d}p}\left[n\ln(p) + \left(\sum_{i=0}^{n} x_i\right)\ln(1-p)\right] \\
&= n\frac{\mathrm{d}}{\mathrm{d}p}[\ln(p)] + \frac{\mathrm{d}}{\mathrm{d}p}\left[\left(\sum_{i=0}^{n} x_i\right)\ln(1-p)\right] \\
&= \frac{n}{p} + \left(\sum_{i=0}^{n} x_i\right)\frac{1}{p-1}
\end{aligned}$$

c. To find the estimator, we set the above derivative to equal zero and isolate p:

$$\frac{n}{p} + \left(\sum_{i=0}^{n} x_i\right)\frac{1}{p-1} = 0 \Leftrightarrow p = \frac{n}{n+\sum_{i=0}^{n} x_i}$$

3. Maximum likelihood estimator for geometric random variables (R)

In Problem 2, you showed that the geometric distribution defined as

$$\Pr[X = k] = (1-p)^k \cdot p$$

has the following maximum likelihood estimator for p:

$$\hat{p}^* = \frac{n}{n+\sum_{i=0}^{n} x_i}.$$

This definition of geometric random variable is the one use by R, as stated at the beginning of the "Details" section of `help(rgeom)`. In this exercise you will verify that, in this case, using the inverse of the sample mean as the estimator for p leads to heavily biased estimations.

Let $n = 200$. First of all, define a function `estimate_p` that, given the realization of a random sample of n elements it returns the estimate of p using the estimator $\hat{p}^* = \frac{n}{n+\sum_{i=0}^{n} x_i}$.

Then, define $p = 0.3$, and take a random sample of n elements using the `rgeom` function. From this random sample, estimate p using first the estimator $\hat{p} = \frac{1}{\bar{x}}$ and then using the estimator $\hat{p}^* = \frac{n}{n+\sum_{i=0}^{n} x_i}$. Compute the two values $\hat{p} - p$ and $\hat{p}^* - p$. What do the resulting numbers suggest?

Repeat the above sampling and estimation procedure 1000 times, accumulating the values $\hat{p} - p$ and $\hat{p}^* - p$ in two separate lists. Plot the two distributions, possibly overlaying them on the same plot. What can you conclude by observing the plot?

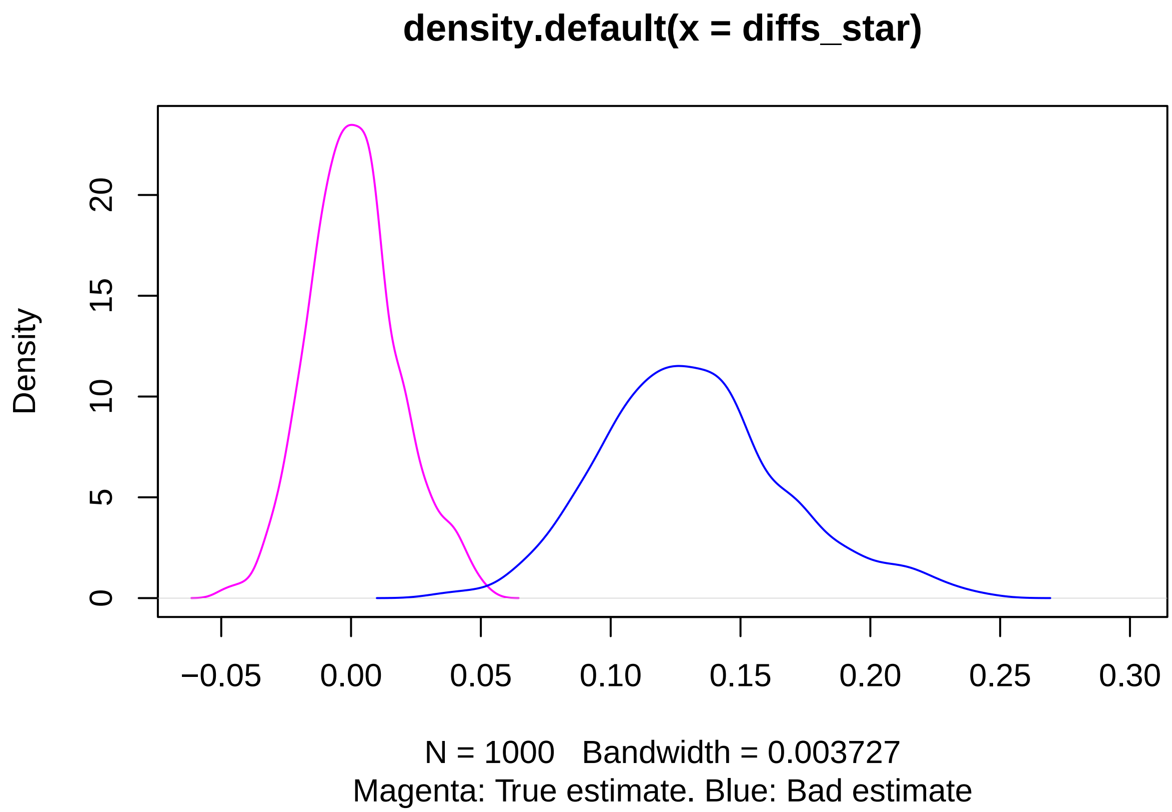
Solution:

```
n <- 200
estimate_p <- function(X) {
  return(length(X)/(length(X)+sum(X)))
}
p <- 0.3

diffs <- c()
diffs_star <- c()
```

```r
for (i in 1:1000) {
  X <- rgeom(n, p)
  p_est <- 1/mean(X)
  p_est_star <- estimate_p(X)
  diff <- p_est - p
  diff_star <- p_est_star - p

  diffs <- c(diffs, diff)
  diffs_star <- c(diffs_star, diff_star)
}
plot(density(diffs_star), xlim=c(-0.06, 0.3), col="magenta", sub="Magenta: True estimate. Blue: Bad esti
lines(density(diffs), col="blue")
```



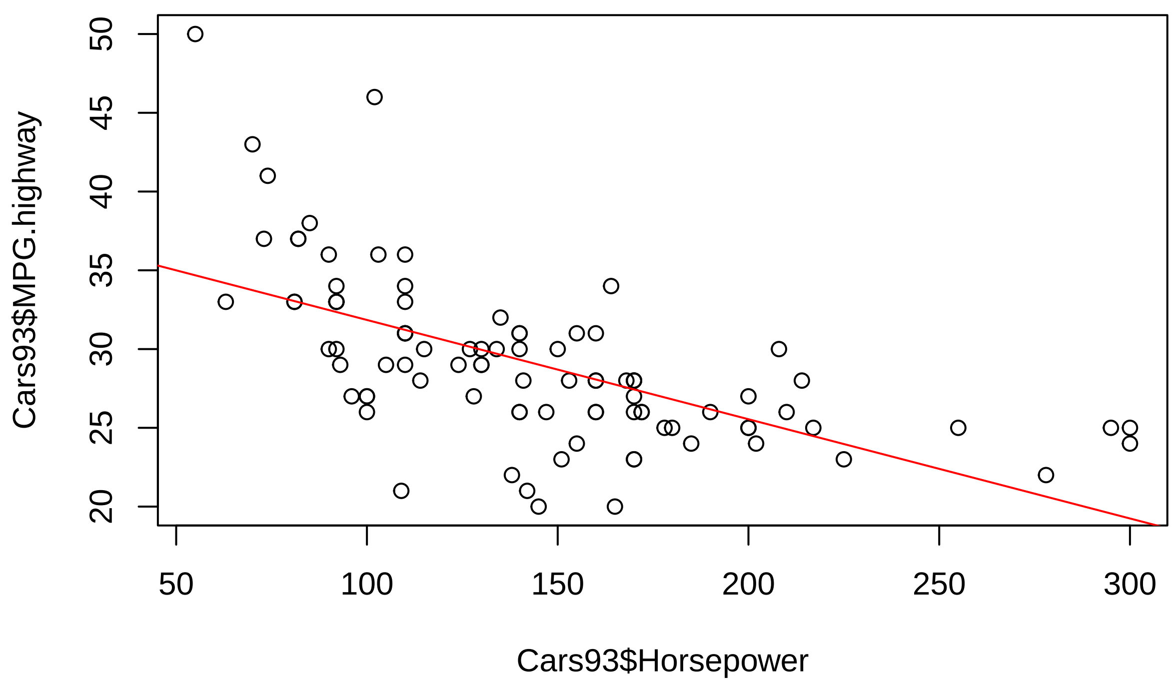
As we can see the true estimate is centered around 0, meaning its expected value is that of the true p, while the bad estimate is positively biased and also has a wider standard deviation.

4. Linear regression model and residuals (R)

Let us take a look at the `Cars93 (MASS)` data set.
(a) Plot the mileage `MPG.highway` in the function of `Horsepower`. Compute the least-squares estimate for the regression line and add it to the plot.
(b) What the predicted mileage for a car with 225 horsepower?
(c) Compute and plot the residuals in the function of horsepower. On the basis of the residuals, is the linear model assumption reasonable?

Solution:
a.

```
library(MASS)
fit <- lm(Cars93$MPG.highway ~ Cars93$Horsepower)
plot(Cars93$Horsepower, Cars93$MPG.highway)
abline(fit, col="red")
```



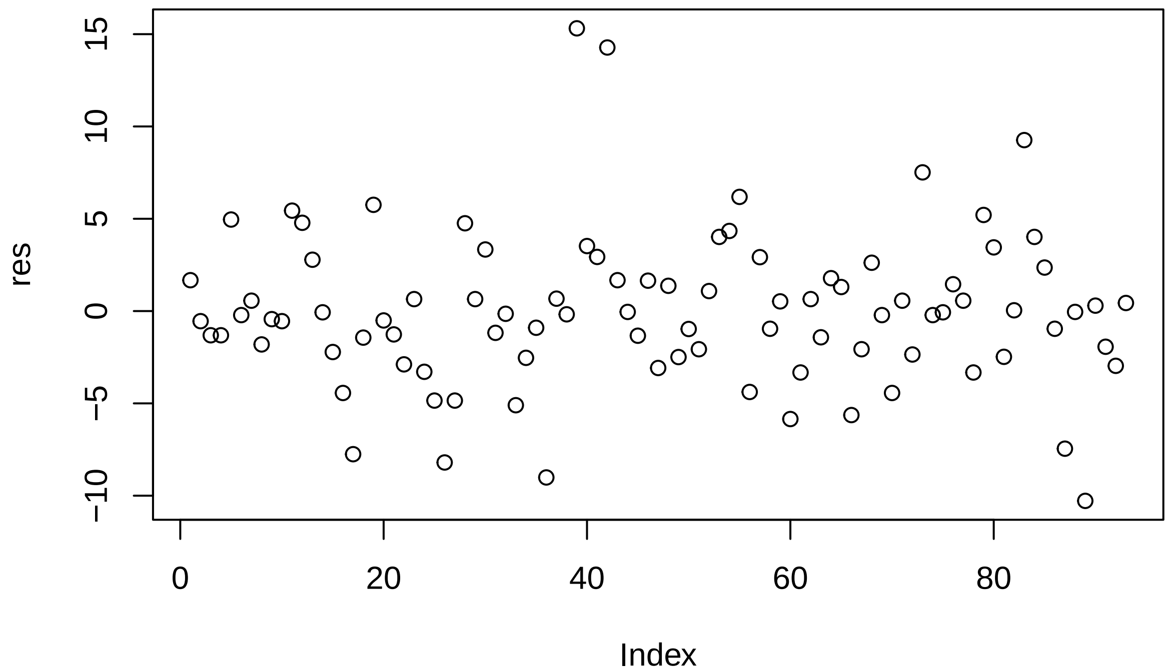
b.

```
coeffs <- coefficients(fit)
estimate_mpg <- function(x) {
  return(coeffs[1] + coeffs[2]*x)
}
hp <- 225
cat("The estimated mileage for a car with", hp, "horsepower is", estimate_mpg(hp))
```

```
## The estimated mileage for a car with 225 horsepower is 23.97066
```

c.

```
res <- resid(fit)
plot(res)
```



It is homoscedastic and looks evenly distributed.

5. Linear models (Theory)

In some situations we may know that the linear model should have some peculiarities, like having no slope, or having intercept equals to zero[2]. Answer to the two following separate questions (i.e. the answer to one doesn't depend on the answer to the other). Let U_i be random variables with expectation zero and variance σ^2.

a. Consider the case $\alpha = 0$. The model then becomes $Y_i = \beta x_i + U_i$, for $i = 1, 2, \ldots, n$. Find the least squares estimate $\hat{\beta}$ for β.

b. Consider the case $\beta = 0$. The mode is then $Y_i = \alpha + U_i$, for $i = 1, 2, \ldots, n$. Find the least squares estimate $\hat{\alpha}$ for α.

Solution:
a.
The formulas for both estimators are:

$$n\alpha + \beta \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$$
$$\alpha \sum_{i=1}^{n} x_i + \beta \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i.$$

[2]For instance we may know that when one quantity of the bivariate dataset is 0 then the other must be zero.

When $\alpha = 0$, $\hat{\beta}$ is

$$\beta \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \Leftrightarrow$$

$$\hat{\beta} = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i}.$$

When $\beta = 0$, $\hat{\alpha}$ is

$$\alpha \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i y_i \Leftrightarrow$$

$$\hat{\alpha} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i}.$$