Homework 4

Chapter 3.8.

p. 58-61: ex. 3, 10, 19.

3. Let $X_1, \dots, X_n \sim \text{Uniform}(0, 1)$ and let $Y_n = \max\{X_1, \dots, X_n\}$. Find $E(Y_n)$.

Let the random variable $Y_n = \max\{X_1, \dots, X_n\}$ have the PDF $g(y)$. Then,

$$E(Y_n) = \int_a^b y \cdot g(y).$$

We need to find PDF

Consider a continuous random variable X with an absolutely continuous CDF $F(X)$, then PDF $f(x)$ defined by

$$f(x) = \frac{dF(x)}{dx} = F(x)'$$

Then I will find first CDF $F(y)$

$$F(y) = P(Y_n < y) = P(\max\{X_1 \dots X_n\} < y) =$$

$$= P(X_1 < y, X_2 < y, \dots X_n < y) = P(X_1 < y) \cdot \ldots \cdot P(X_n < y)$$

$$= \left(\frac{y-a}{b-a}\right) \cdot \ldots \cdot \left(\frac{y-a}{b-a}\right) = \left(\frac{y-0}{1-0}\right) \cdot \ldots \cdot \left(\frac{y-0}{1-0}\right) =$$

$= y \cdot y \cdot y \cdot \ldots \cdot y = (y)^n$ - since independly uniformly distributed.

Then our PDF will be

$$f(y) = F'(y) = (y^n)' = n \cdot y^{n-1}$$

Expectation will be

$$E(y_n) = \int_0^1 y \cdot f(y) = \int_0^1 y \cdot n \cdot y^{n-1} =$$

$$= \int_0^1 n \cdot y^n = n\left(\frac{y^{n+1}}{n+1}\right)\Big|_0^1 = n\left(\frac{1^{n+1}}{n+1} - \frac{0^{n+1}}{n+1}\right) =$$

$$= n\left(\frac{1}{n+1} - 0\right) = \frac{n}{n+1}$$

$$E(y_n) = \frac{n}{n+1}$$

10. Let $X \sim N(0,1)$ and let $y = e^x$. Find $E(y)$ and $Var(y)$.

Let $y = H(x)$. How do we compute $E(y)$ - ? One way is to find $f_y(y)$ and then compute $E(y) = \int y \cdot f(y)\,dy$.

But there is an easier way:

$$E(y) = E(H(x)) = \int H(x) \cdot dF_x(x)$$

$$E(y) = E(e^x) = \int_{-\infty}^{+\infty} e^x \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx =$$

$$= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{x-\frac{x^2}{2}} dx = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{\frac{2x-x^2}{2}} dx =$$

$$= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2+2x-1+1}{2}} dx = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{\frac{(-x^2+2x-1)}{2}+\frac{1}{2}} dx =$$

$$= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{\frac{1}{2}} \cdot e^{-\frac{(x^2-2x+1)}{2}} dx =$$

$$= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{\frac{1}{2}} \cdot e^{-\frac{(x-1)^2}{2}} dx = e^{\frac{1}{2}} \cdot \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{(x-1)^2}{2}} dx =$$

$= e^{\frac{1}{2}} \cdot 1 = e^{\frac{1}{2}}$ - because total probability is equal to 1 under the curve.

$$Var(y) = E(y^2) - [E(y)]^2$$

$$E(y^2) = E(e^{2x}) = \int_{-\infty}^{+\infty} e^{2x} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx =$$

$$= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{\frac{4x-x^2}{2}} = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2+4x-4+4}{2}} dx =$$

$$= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x^2-4x+4)}{2}} \cdot e^{\frac{4}{2}} = e^2 \cdot \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-2)^2}{2}} =$$

$= e^2 \cdot 1 = e^2$, total area under the curve integrates to 1.

$$Var(y) = e^2 - [e^{\frac{1}{2}}]^2 = e^2 - e^1 = e^2 - e = e(e-1)$$

$$E(y) = e^{\frac{1}{2}}$$

19. Let $X_1, \dots, X_n \sim$ Uniform $(0, 1)$. Let f_x be the density of the Uniform $(0, 1)$.
Plot $f_x(x)$.
Now let $\overline{X_n} = \frac{1}{n} \sum_{i=1}^{n} X_i$. Find $E(\overline{X_n})$, $Var(\overline{X_n})$.

$$f_x(x) = \begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & x < a \text{ or } x > b \end{cases}$$

$$E(\overline{X_n}) = \int_{-\infty}^{+\infty} x \cdot f_x(x)\,dx = \int_0^1 x\left(\frac{1}{b-a}\right)dx =$$

$$= \frac{1}{b-a} \int_0^1 x \cdot dx = \frac{1}{b-a} \int_0^1 x \cdot dx = \frac{1}{b-a}\left(\frac{x^2}{2}\right)\Big|_0^1 =$$

$$= \frac{1}{1-0} \cdot \left(\frac{1}{2} - \frac{0}{2}\right) = \frac{1}{2}$$

$$E(\overline{X_n}) = \mu = \frac{1}{2}$$

$$E(\overline{X}_n^2) = \int_{-\infty}^{+\infty} x^2 \cdot f(x)\,dx = \int_0^1 x^2 \cdot \left(\frac{1}{b-a}\right)dx =$$

$$= \left(\frac{1}{b-a}\right) \cdot \int_0^1 x^2\,dx = \frac{1}{b-a} \cdot \int_0^1 x^2\,dx =$$

$$= \left(\frac{1}{b-a}\right) \cdot \left(\frac{x^3}{3}\right)\Big|_0^1 = 1 \cdot \left(\frac{1}{3} - 0\right) = \frac{1}{3}$$

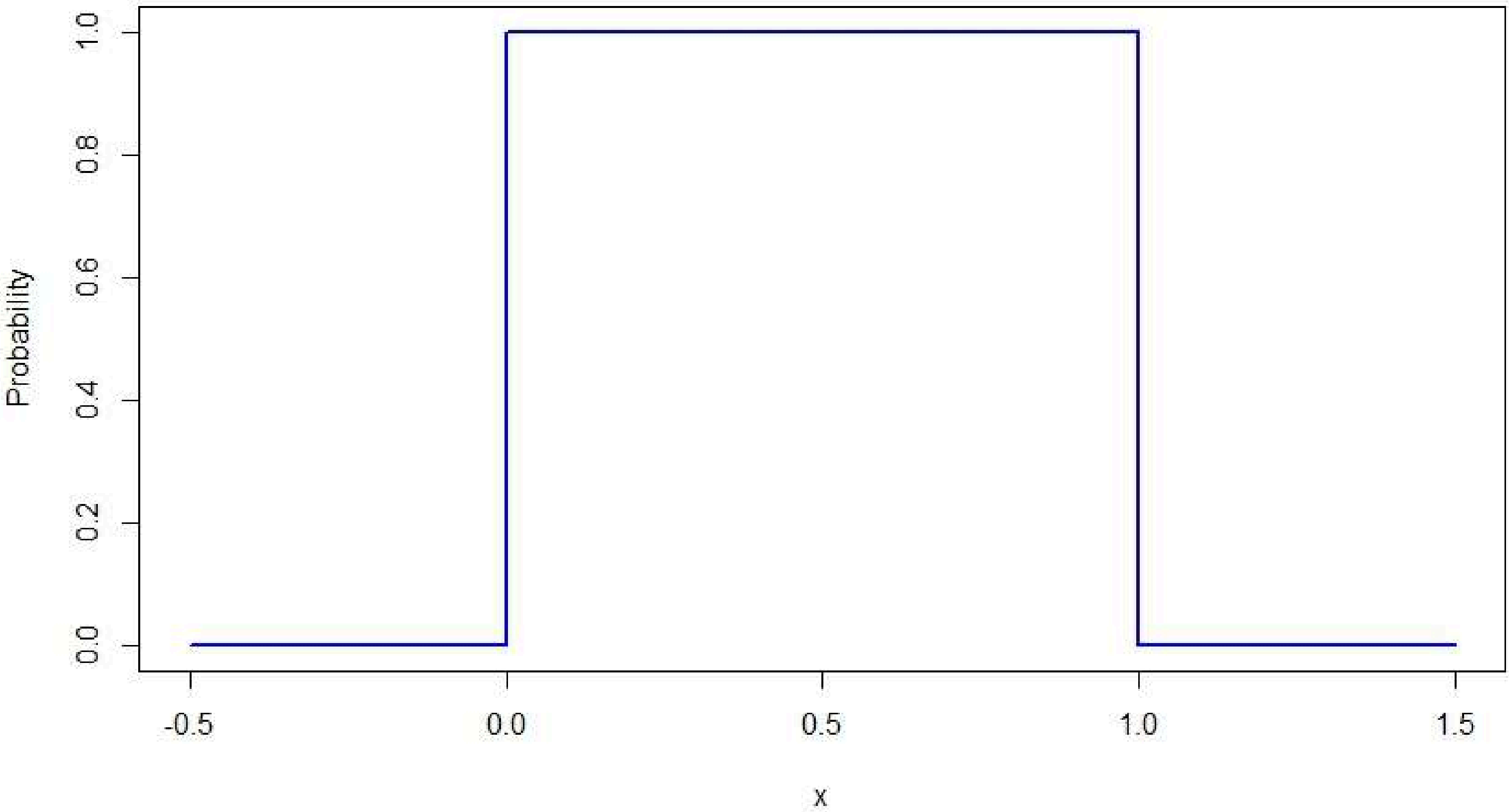
$$Var(X) = E(\overline{X}_n^2) - \left[E(\overline{X}_n)\right]^2 = \frac{1}{3} - \left(\frac{1}{2}\right)^2 =$$

$$= \frac{1}{3} - \frac{1}{4} = \frac{4-3}{12} = \frac{1}{12}$$

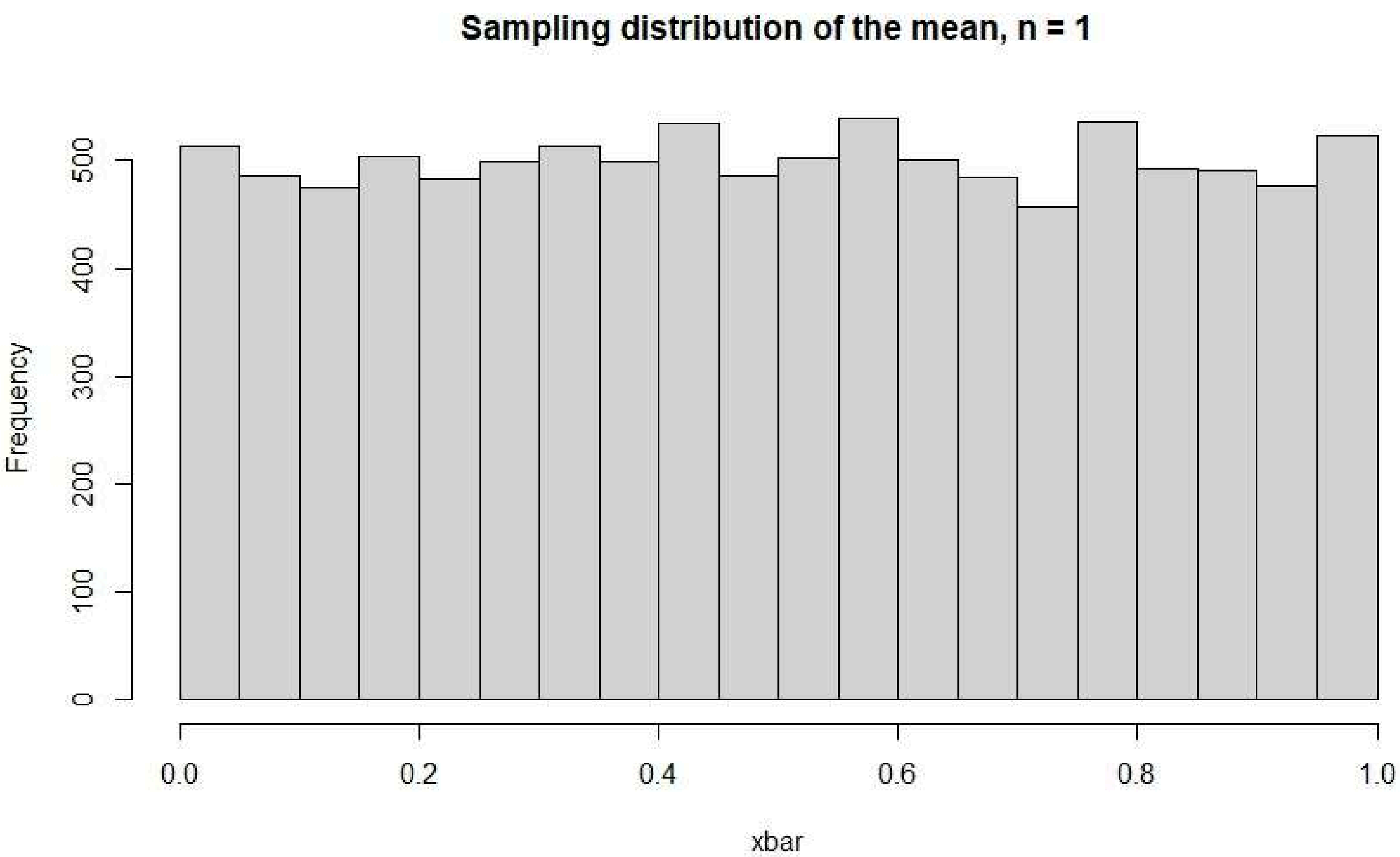
$$Var(\overline{X}_n) = \frac{\sigma^2}{n} = \frac{1}{12 \cdot n}$$

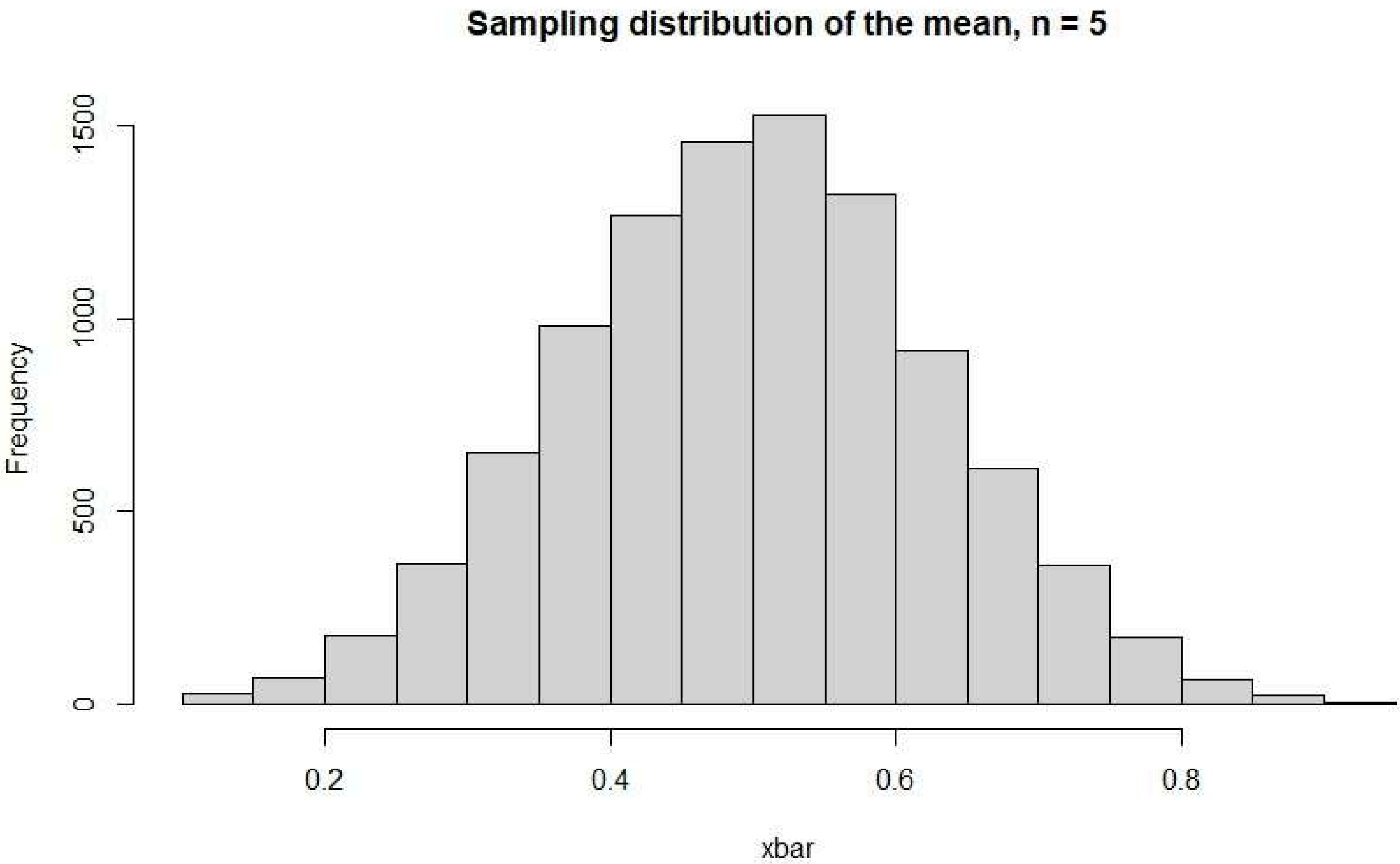
$$E(\overline{X}_n) = \frac{1}{2}$$

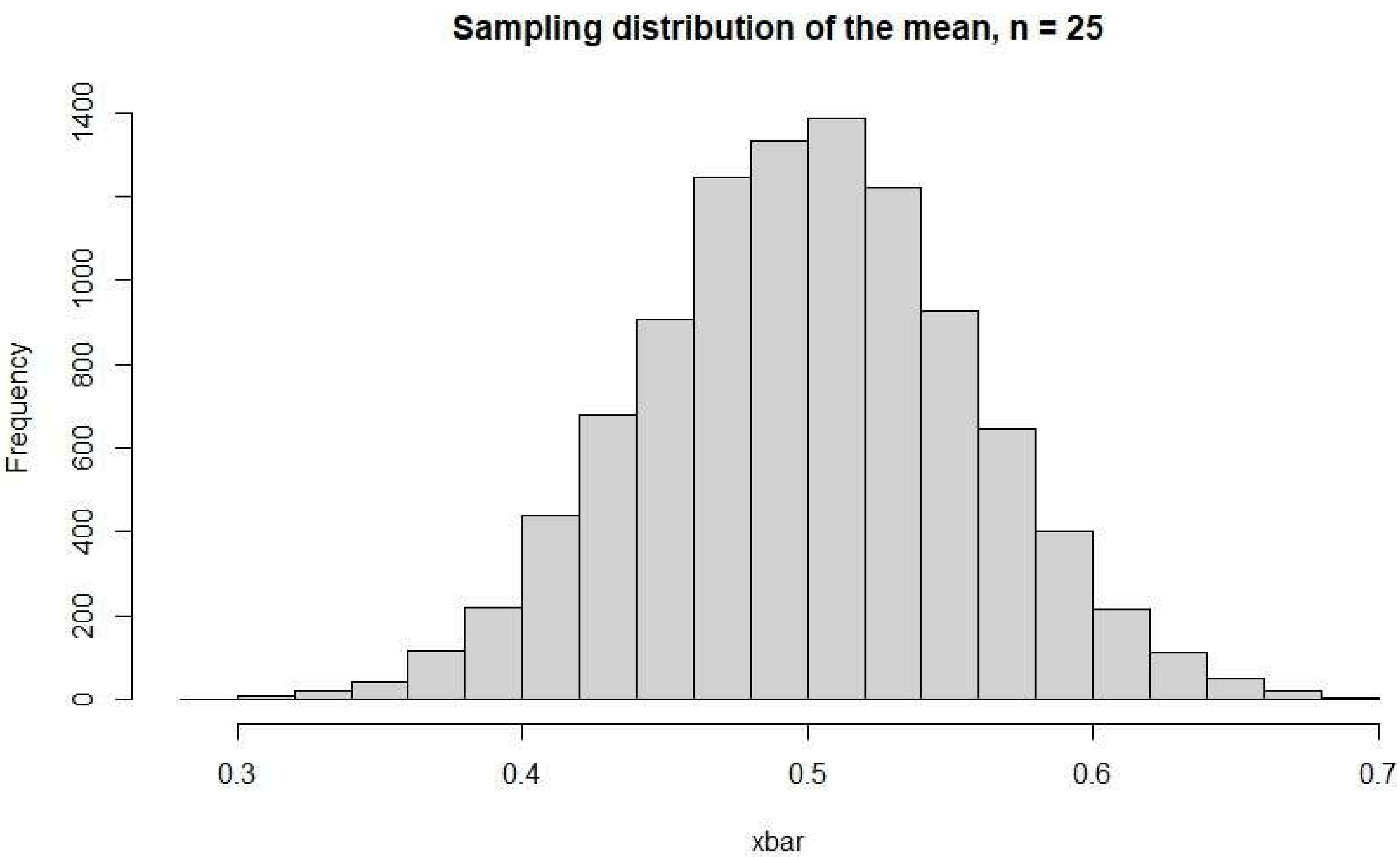
The mean stays similar $\left(\frac{1}{2}\right)$, but on the other hand the variance decreases by the increasement of n. Distribution gets around the $\mu = \frac{1}{2}$, which make sampling distribution look Normal.



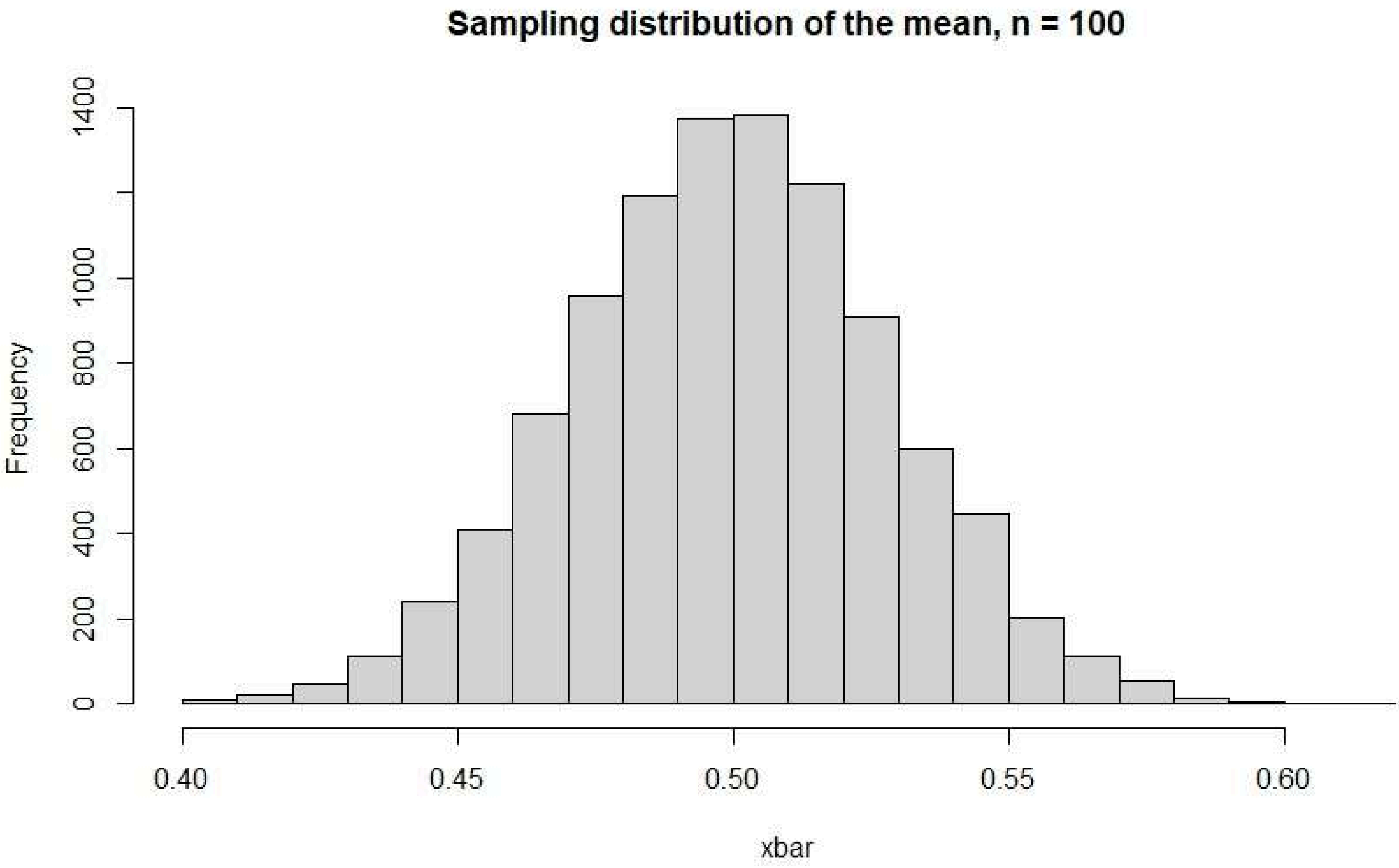
1.0
0.8
0.6
0.4
0.2
0.0
Probability
-0.5
0.0
0.5
1.0
1.5
x


Sampling distribution of the mean, n = 1
Frequency
0
100
200
300
400
500
0.0
0.2
0.4
0.6
0.8
1.0
xbar


Sampling distribution of the mean, n = 5
Frequency
0
500
1000
1500
0.2
0.4
0.6
0.8
xbar



Sampling distribution of the mean, n = 25
Frequency
0
200
400
600
800
1000
1400
0.3
0.4
0.5
0.6
0.7
xbar

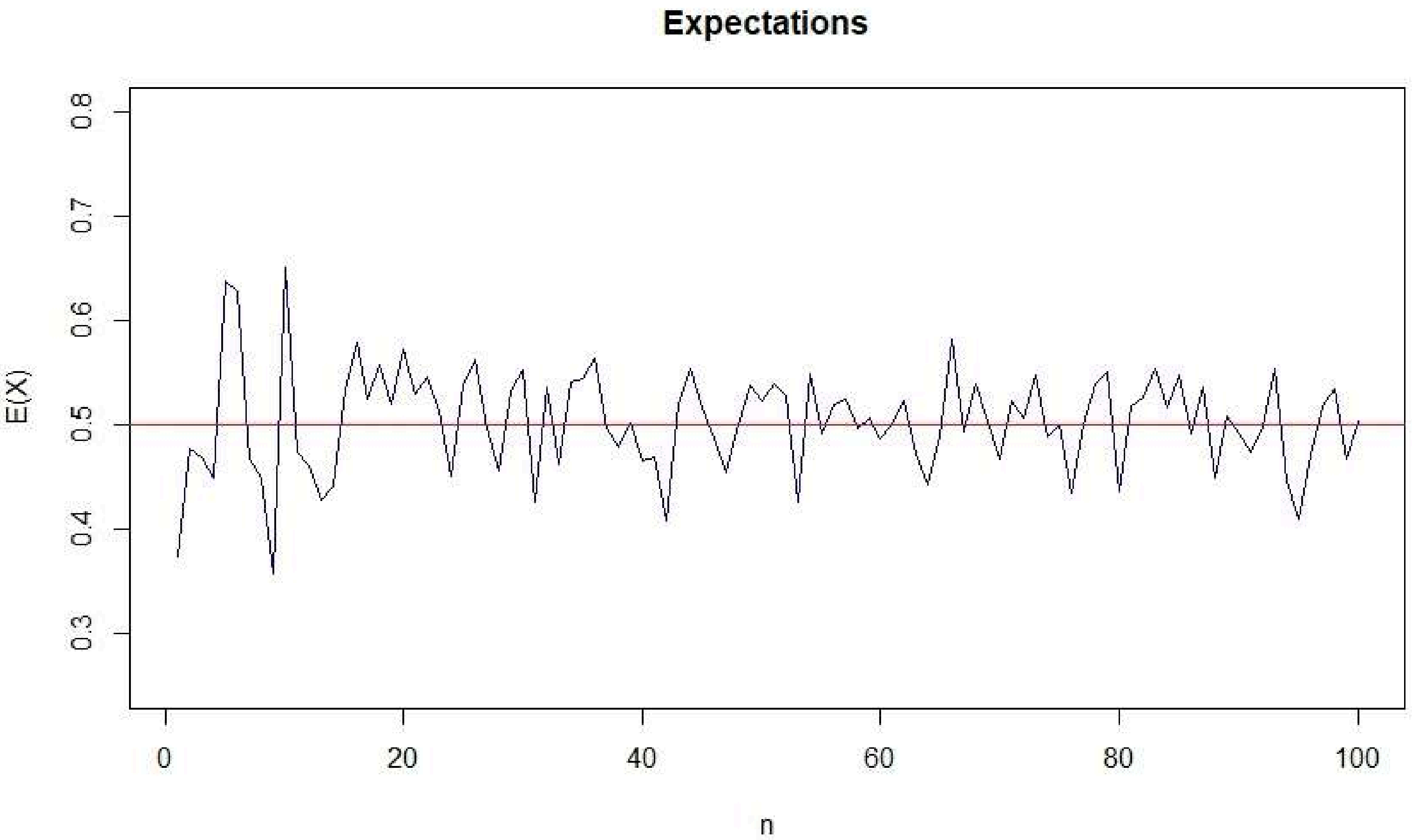


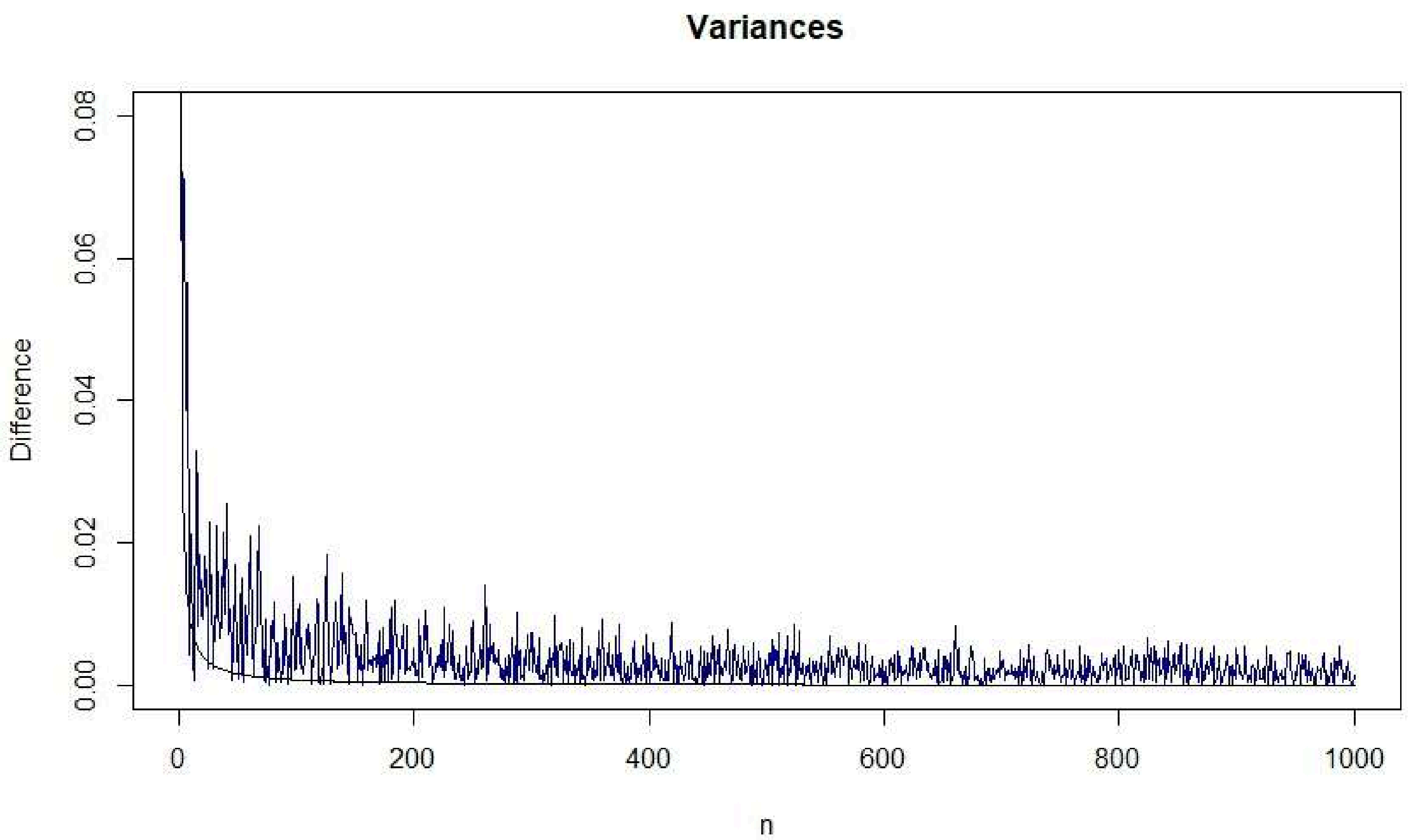
Sampling distribution of the mean, n = 100
Frequency
0
200
400
600
800
1000
1400
0.40
0.45
0.50
0.55
0.60
xbar

Console Terminal Jobs

R 4.1.2 · ~/

```
> plot(variance, type='l', col="darkblue", main="Variances", xlab="n", ylab="Var(X)", ylim=c(0,0.08))
> lines(nn, 1/(12*nn), type='l')
> curve(dunif(x, min = 0, max = 1),
+       from = -0.5, to = 1.5,
+       n = 100000,
+       col = "blue",
+       lwd = 2,
+       ylab = 'Probability')
> n_1 <- runif(1, min=0, max=1)
> n_5 <- runif(5, min=0, max=1)
> n_25 <- runif(25, min=0, max=1)
> n_100 <- runif(100, min=0, max=1)
> mean(n_1)
[1] 0.3770901
> var(n_1)
[1] NA
> mean(n_5)
[1] 0.4287843
> var(n_5)
[1] 0.0585694
>
> mean(n_25)
[1] 0.4679786
> var(n_25)
[1] 0.08342589
>
> mean(n_100)
[1] 0.5808355
> var(n_100)
[1] 0.07212987
> 
```


Expectations
E(X)
0.3
0.4
0.5
0.6
0.7
0.8
0
20
40
60
80
100
n



Variances
Difference
0.00
0.02
0.04
0.06
0.08
0
200
400
600
800
1000
n

Chapter 6.6.

p. 95 - 96 : ex. 3.

3. Let $X_1, \ldots, X_n \sim Uniform(0, \theta)$ and let $\hat{\theta} = 2 \cdot \overline{X}_n$. Find the bias, SE, MSE of this estimator.

$$E(X_i) = \frac{a+b}{2} = \frac{0+\theta}{2} = \frac{\theta}{2}$$

$$Var(X_i) = \frac{(b-a)^2}{12} = \frac{(\theta - 0)^2}{12} = \frac{\theta^2}{12}$$

$$E_\theta(\hat{\theta}) = E_\theta(2 \cdot \overline{X}_n) = 2 \cdot E_\theta(\overline{X}_n) = 2 \cdot \frac{\theta}{2} = \theta$$

The bias of an estimator is defined by

$bias(\theta) = E_\theta(\hat{\theta}) - \theta$, therefore,

$$bias(\hat{\theta}) = \theta - \theta = 0$$

$$Var_\theta(\hat{\theta}) = Var_\theta(2 \cdot \overline{X}_n) = 4 \cdot Var(\overline{X}_n) = 4 \cdot \frac{Var(X_i)}{n} =$$

$$= \cancel{4} \cdot \frac{\theta^2}{\cancel{12}_3 \cdot n} = \frac{\theta^2}{3n}$$

$$SE = SE(\hat{\theta}) = \sqrt{Var(\hat{\theta})} = \sqrt{\frac{\theta^2}{3n}} = \theta \sqrt{\frac{1}{3n}}$$

MSE can be written as

$$MSE = bias^2(\hat{\theta}) + Var_\theta(\hat{\theta})$$

Therefore,

$$MSE = 0^2 + \frac{\theta^2}{3n} = \frac{\theta^2}{3n}$$

Bias, SE and MSE of $\hat{\theta}$ were found.

Chapter 8.6.

p. 116-118: ex. 1.

1. Estimated correlation coefficient:

$\hat{\theta} = 0.5459$

Estimated SE of correlation coefficient:

$SE_\theta = 0.2674$

95% conf. interval with Normal method

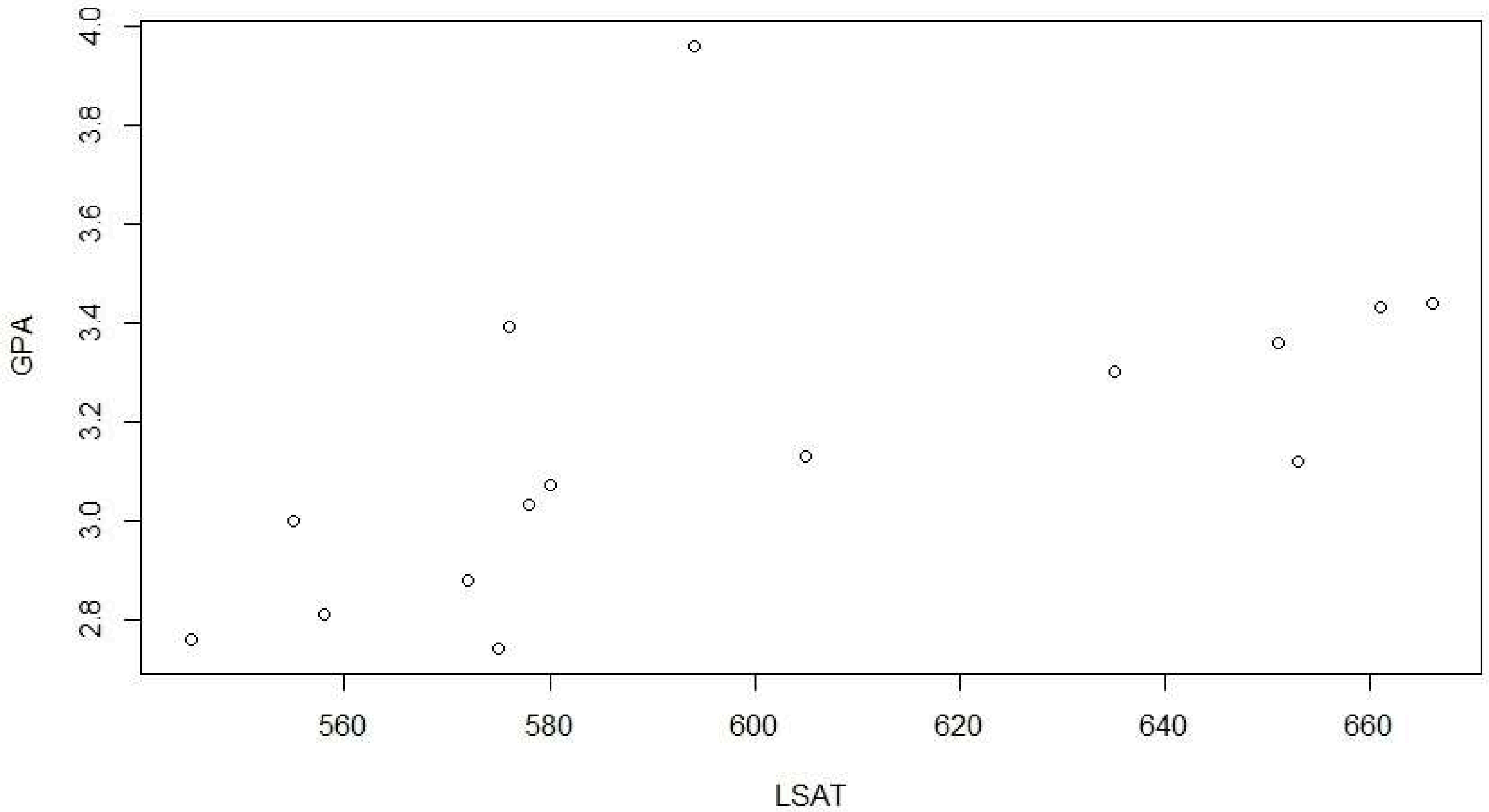
$CI = (0.0218;\ 1.0701)$

95% conf. interval with percentile method

$CI = (-0.5014;\ 0.5244)$

95% conf. interval with pivotal method:

$CI = (0.5675 ; 1.5933)$


4.0
3.8
3.6
3.4
3.2
3.0
2.8
GPA
560
580
600
620
640
660
LSAT

19Ex_Zhetessov.R 8.6-1ExZhetessov.R*

```r
plot(LSAT,GPA, type = 'p')

LSAT_mean = mean(LSAT)
GPA_mean = mean(GPA)

theta = sum((LSAT-LSAT_mean)*(GPA-GPA_mean)) / sqrt(sum((LSAT-LSAT_mean)*(LSAT-LSAT_mean))*sum((GPA-GPA_mean)*(GPA-GPA_

theta

#--------------------------------------------------------#

theta_hat = c()

for(i in 1:1000000){
  boot1 = sample(LSAT, 15, replace = TRUE)
  boot2 = sample(GPA, 15, replace = TRUE)

  mean1 = mean(boot1)
  mean2 = mean(boot2)

  theta_hat[i] = sum((boot1-mean1)*(boot2-mean2)) / sqrt(sum((boot1-mean1)*(boot1-mean1))*sum((boot2-mean2)*(boot2-mean
}

SE = sd(theta_hat)

SE

#--------------------------------------------------------#

normal = c(theta - 1.96 * SE, theta + 1.96 * SE)
percentile = c(quantile(theta_hat,0.025), quantile(theta_hat,0.975))
pivotal = c(2*theta - quantile(theta_hat,0.975), 2*theta - quantile(theta_hat,0.025))

theta
SE
normal
percentile
pivotal
#Zhetessov
```

37:1 (Top Level) R Script

Console Terminal Jobs

R 4.1.2 · ~/

```
> theta
[1] 0.5459189
> SE
[1] 0.2674222
> normal
[1] 0.02177137 1.07006646
> percentile
      2.5%      97.5%
-0.5014273  0.5243704
> pivotal
    97.5%      2.5%
0.5674675 1.5932651
>
```