STA237 - Assignment 3

Madeline Ahn

2022-12-08

Question 1

1(a)

We want to find the probability density function for $U = Y^2$. We have:

$$\begin{aligned} f_U(u) &= P(U \leq u) \\ &= P(Y^2 \leq u) \\ &= P(-\sqrt{u} \leq Y \leq \sqrt{u}) \end{aligned}$$

Note that $y > 0$, so $P(Y \leq -\sqrt{u})$ would be invalid. Thus:

$$F_U(u) = P(Y \leq \sqrt{u})$$

Now, we take the derivative with respect to u, and have $\frac{dy}{du} = \frac{1}{2\sqrt{u}}$. Then,

$$\begin{aligned} f_U(u) &= f_Y(\sqrt{u})\frac{1}{2\sqrt{u}} \\ &= \frac{2\sqrt{u}}{\theta}e^{-(\sqrt{u})^2/\theta} \\ &= \frac{1}{\theta}e^{-u/\theta} \end{aligned}$$

Therefore, the probability density function is:

$$f_U(u) = \begin{cases} \frac{1}{\theta}e^{-u/\theta}, & y > 0 \\ 0, & \text{otherwise} \end{cases}$$

1(b)

We want to find $E(Y)$ and $V(Y)$ from (a). Knowing that $Y = \sqrt{u}$, we have $E(Y) = E(\sqrt{u})$, and $V(Y) = E(\sqrt{u})$. First, we calculate $E(Y)$:

$$\begin{aligned}
E(\sqrt{u}) &= \int_0^\infty \sqrt{u}\frac{1}{\theta}e^{-u/\theta} \quad du \\
\text{By substitution: } & \text{Let } v = \frac{u}{\theta},\ dv = \frac{du}{\theta} \\
&= \int_0^\infty \sqrt{v\theta} \cdot \frac{1}{\theta}e^{-v} \cdot \theta \quad dv \\
&= \sqrt{\theta}\int_0^\infty \sqrt{v}e^{-v} \quad dv \\
&= \sqrt{\theta}\int_0^\infty v^{3/2-1}e^{-v} \quad dv \\
&= \sqrt{\theta}\ \Gamma(\frac{3}{2})
\end{aligned}$$

Then, notice that $\Gamma(\frac{3}{2}) = \frac{\sqrt{\pi}}{2}$. Therefore,

$$E(Y) = E(\sqrt{u}) = \sqrt{\theta} \cdot \frac{\sqrt{\pi}}{2} = \frac{\sqrt{\pi\theta}}{2}$$

Now, we solve $V(Y)$:

$$\begin{aligned}
V(\sqrt{u}) &= E(\sqrt{u}^2) - E(\sqrt{u})^2 \\
&= E(u) - \left(\frac{\sqrt{\pi\theta}}{2}\right)^2 \\
&= E(u) - \frac{\pi\theta}{4}
\end{aligned}$$

Solve $E(u)$:

$$\begin{aligned}
E(u) &= \int_0^\infty u\frac{1}{\theta}e^{-u/\theta} \quad du \\
\text{Let } v &= \frac{u}{\theta},\ dv = \frac{du}{\theta} \\
&= \int_0^\infty v\theta \cdot \frac{1}{\theta}e^{-v} \cdot \theta \quad dv \\
&= \theta\int_0^\infty v \cdot e^{-v} \quad dv \\
&= \theta\int_0^\infty v^{2-1}e^{-v} \quad dv \\
&= \theta \cdot \Gamma(2) \\
&= \theta, \quad \text{since } \Gamma(2) = 1
\end{aligned}$$

So:

$$V(\sqrt{u}) = \theta - \frac{\pi\theta}{4}$$

Question 2

We want to find the probability density function for $U = \frac{Y_1}{Y_1+Y_2}$. We have:

$$U_1 = \frac{Y_1}{Y_1 + Y_2}, \quad U_2 = Y_1 + Y_2$$

Then:

$$\begin{aligned} u_1 = \frac{y_1}{y_1 + y_2} &\implies y_1 = u_1 u_2 \\ u_2 = y_1 + y_2 &\implies y_2 = u_2 - u_1 u_2 = u_2(1 - u_1) \end{aligned}$$

Then, knowing that Y_1, Y_2 are independent, we have:

$$\begin{aligned} f(y_1, y_2) &= f(y_1) f(y_2) \\ &= e^{-y_1} \cdot e^{-y_2} \\ &= \begin{cases} e^{-(y_1+y_2)};\ y_1 > 0, y_2 > 0 \\ 0;\ \text{otherwise} \end{cases} \end{aligned}$$

Now, we can use the Jacobian:

$$\begin{aligned} |J| &= \begin{bmatrix} \frac{\partial y_1}{\partial u_1} & \frac{\partial y_1}{\partial u_2} \\ \frac{\partial y_2}{\partial u_1} & \frac{\partial y_2}{\partial u_2} \end{bmatrix} \\ &= \begin{bmatrix} u_2 & u_1 \\ -u_2 & 1 - u_1 \end{bmatrix} \\ &= |(u_2 - u_1 u_2) + u_1 u_2| \\ &= |u_2| \end{aligned}$$

Hence:

$$\begin{aligned} f(u_1, u_2) &= f(y_1, y_2)|J| \\ &= e^{-(y_1+y_2)}|J| \\ &= e^{-(u_2)}|u_2| \\ &= u_2\ e^{-u_2} \end{aligned}$$

For our domain, note that $y_1 = u_1 u_2$, and $y_2 = u_2(1 - u_1)$, so our domain is $u_2 > 0$, and $0 < u_1 < 1$. Therefore, the probability density function is:

$$f_{u_1,u_2}(u_1, u_2) = u_2\ \ e^{-u_2};\ 0 < u_2,\ 0 < u_1 < 1$$

Question 3

3(a)

We want to calculate $\mu_{\overline{x}}$ and $\sigma_{\overline{x}}$. We have $n = 100, \sigma = 0.15, \mu = 3.65$. Then, $n = 100 > 30$, so we also have $\frac{0.15}{\sqrt{100}} = 0.015$. Then, by Central Limit Theorem, we have:

$$\overline{x} \sim N(3.65,\ 0.015)$$

Hence, $\mu_{\overline{x}} = 3.65$, and $\sigma_{\overline{x}} = 0.015$.

3(b)

We want to find the probability of a mean fuel cost between \$3.65 and \$3.67. We have:

$$\begin{aligned}
P(3.65 \leq \overline{y} \leq 3.67) &= P(\overline{y} \leq 3.67) - P(\overline{y} \leq 3.65) \\
&= P(\frac{\overline{x} - \mu_{\overline{x}}}{\sigma_{\overline{x}}} \leq \frac{3.67 - 3.65}{0.015}) - P(\frac{\overline{x} - \mu_{\overline{x}}}{\sigma_{\overline{x}}} \leq \frac{3.65 - 3.65}{0.015}) \\
&= P(z \leq 1.33) - P(z \leq 0) \\
&\approx 00.9082 - 0.5 \\
&= 0.4082
\end{aligned}$$

3(c)

We want to find the probability of a mean fuel cost above \$3.67. We have:

$$\begin{aligned}
P(\overline{y} > 3.67) &= 1 - P(\overline{y} \leq 3.67) \\
&= 1 - P(Z \leq 1.33) \\
&\approx 1 - 0.9082 \\
&= 0.0918
\end{aligned}$$

3(d)

If the sample size doubled to 200, then our value for $\sigma_{\overline{x}}$ would change such that $\sigma_{\overline{x}} = \frac{0.15}{\sqrt{200}} \approx 0.011$. Then, in (b) and (c), because we are dividing by a lower number (i.e., for calculating for the normal distributions), the probabilities then increase. This means there would be a higher probability for the mean fuel cost to be between \$3.65 and \$3.67 in part (b), and hence a lower probability for the mean fuel cost above \$3.67 in part (c) (since we subtract 1 from the probability of a cost equal to 3.67).

Question 4

4(a)

We want to show the mean of the sample distribution of $\hat{p}$. We have $p = 0.67$ and $n = 1000$. First, we check:

$$\begin{aligned} np &= 1000 \cdot 0.67 \\ &= 670 > 10 \\ n(1-p) &= 1000 \cdot 0.33 \\ &= 330 > 10 \end{aligned}$$

Then, the mean $E(\hat{p})$ is calculated by:

$$\begin{aligned} E(\hat{p}) &= p \\ &= 0.67 \end{aligned}$$

4(b)

We want to show the standard deviation of the sampling distribution. From (a), we can just calculate the standard deviation as:

$$\begin{aligned} \sigma_{\hat{p}} &= \sqrt{\frac{p(1-p)}{n}} \\ &= \sqrt{\frac{0.67(0.33)}{1000}} \\ &= \sqrt{\frac{0.2211}{1000}} \\ &\approx 0.01487 \end{aligned}$$

4(c)

By Central Limit Theorem, the shape of the sampling distribution will follow a normal distribution, since $np \geq 10$, and $n(1-p) \geq 10$.

4(d)

We want to find the probability of $\hat{p} < 0.75$. We know that $\hat{p} \sim N(0.67, 0.01487)$, so we have:

$$\begin{aligned} P(\hat{p} < 0.75) &= P\left(z < \frac{0.75 - 0.67}{0.01487}\right) \\ &\approx P(z < 5.38) \\ &= 1 \end{aligned}$$

4(e)

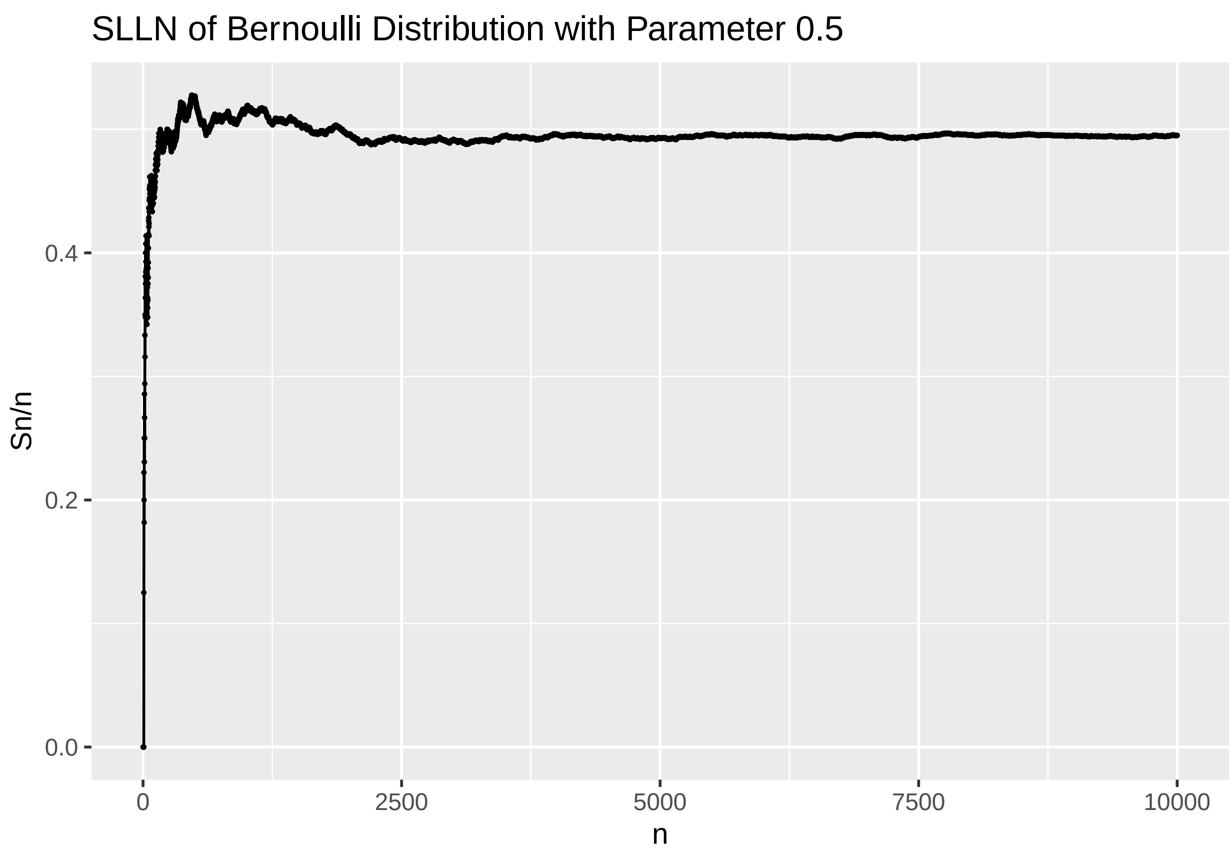
We want to find the probability of $\hat{p} > 0.5$. We have:

$$\begin{aligned} P(\hat{p} > 0.5) &= 1 - P(\hat{p} < 0.5) \\ &= 1 - P(z < \frac{0.5 - 0.67}{0.01487}) \\ &\approx 1 - P(z < -11.43) \\ &= 1 - 0 \\ &= 1 \end{aligned}$$

Question 5

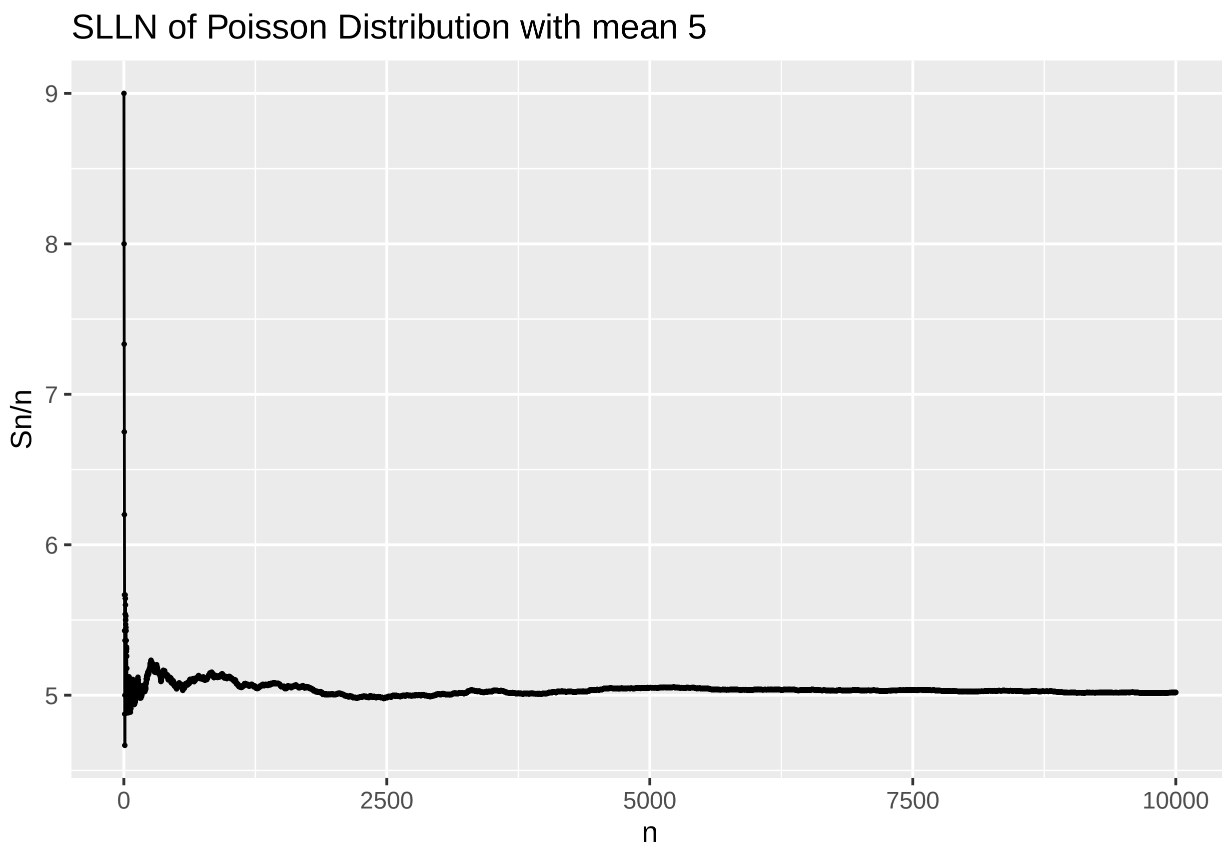
5(i)

```
p = 0.5  # parameter
n = 10000  # simulations
result = numeric(n)  # list of results
sn = 0
for (i in 1:n) {
  sn = sn + rbinom(n, 1, p)  # bernoulli distribution
  result[i] = sn/i
}

i = 1:n  # index
data = data.frame(Index = i, result = result)
ggplot(data, aes(x = Index, y = result)) + geom_point(size = 0.3) + ylab("Sn/n") + xlab("n") + labs(tit
```



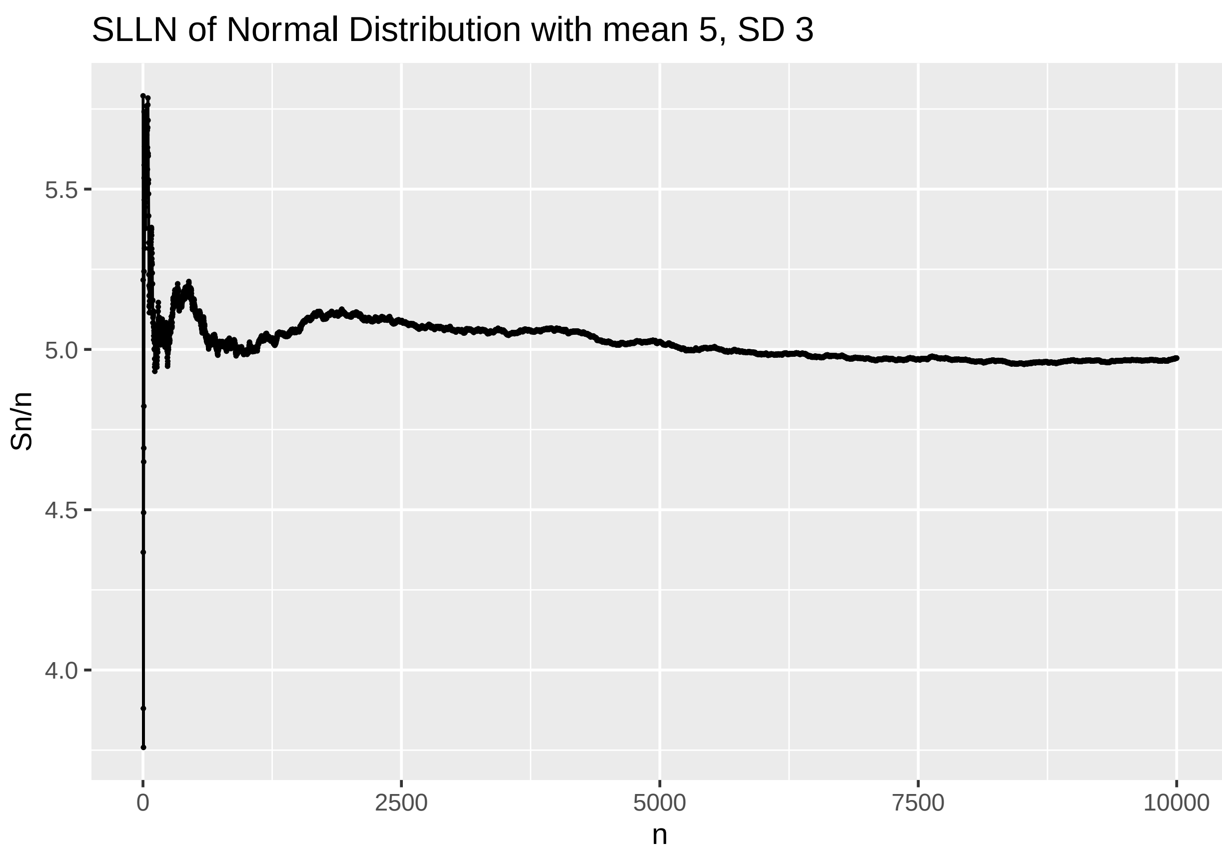
5(ii)

```
mean = 5  # mean
n = 10000  # simulations
result = numeric(n)  # list of results
sn = 0
for (i in 1:n) {
  sn = sn + rpois(n, mean)  # poisson distribution
  result[i] = sn/i
}

i = 1:n  # index
data = data.frame(Index = i, result = result)
ggplot(data, aes(x = Index, y = result)) + geom_point(size = 0.3) + ylab("Sn/n") + xlab("n") + labs(titl
```



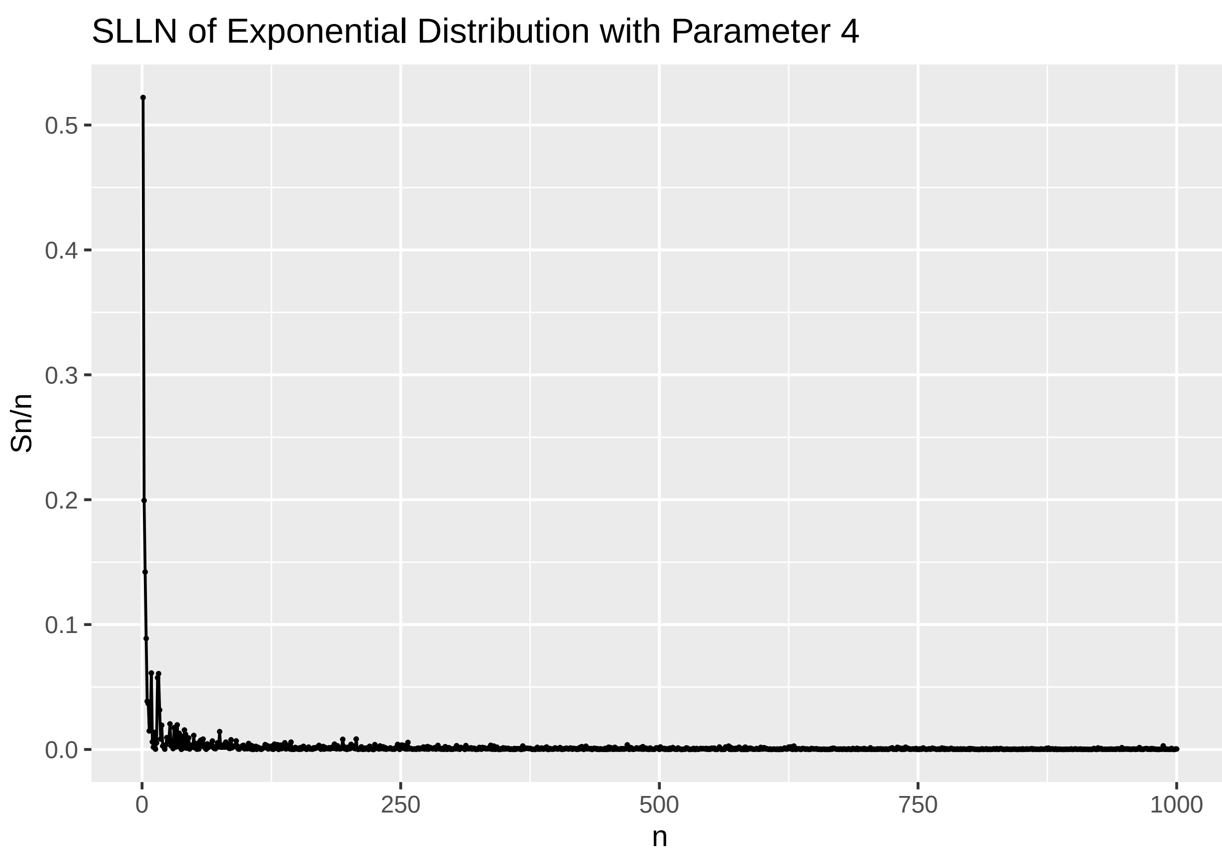
5(iii)

```
mean = 5  # mean
sd = 3  # standard deviation
n = 10000  # simulations
result = numeric(n)  # list of results
sn = 0
for (i in 1:n) {
  sn = sn + rnorm(n, mean, sd)  # normal distribution
  result[i] = sn/i
}

i = 1:n  # index
data = data.frame(Index = i, result = result)
ggplot(data, aes(x = Index, y = result)) + geom_point(size = 0.3) + ylab("Sn/n") + xlab("n")  + labs(ti
```



5(iv)

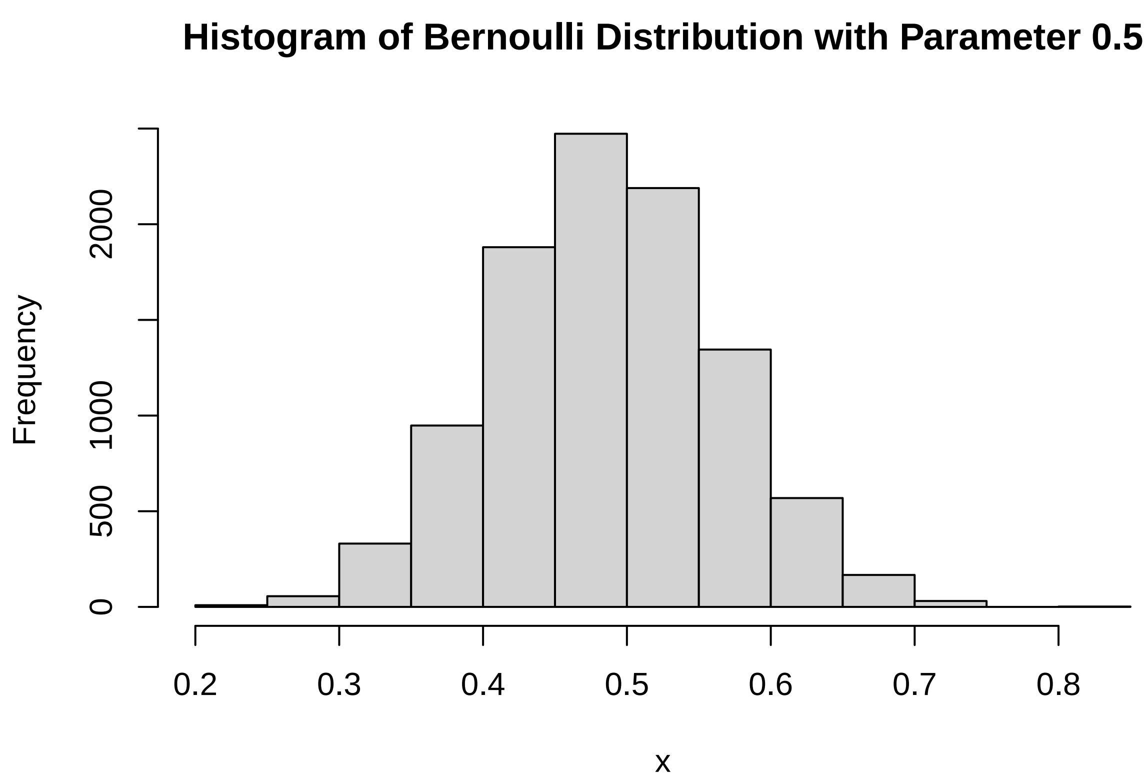
```
p = 4  # parameter
n = 1000  # simulations
result = numeric(n)  # list of results
sn = 0
for (i in 1:n) {
  sn = rexp(n, p)  # exponential distribution
  result[i] = sn/i
}

i = 1:n
data = data.frame(Index = i, result = result)
ggplot(data, aes(x = Index, y = result)) + geom_point(size = 0.3) + xlab("n") + ylab("Sn/n") + labs(tit
```



Question 6

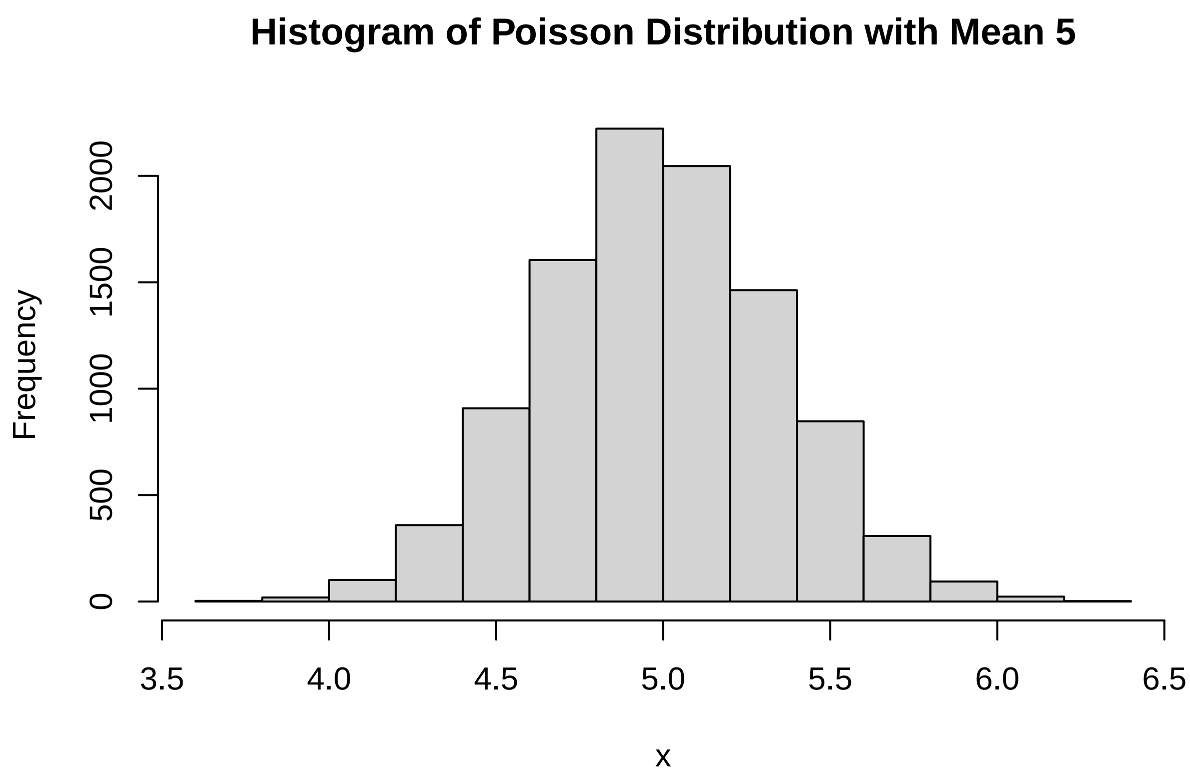
6(i)

```
p = 0.5   # parameter
n = 10000 # simulations
s = 40   # sample size
means = numeric(n)  # list of simulations
values = matrix(0, s, n)  # simulated values

for (i in 1:n) {
  values[,i] = rbinom(s, 1, p)  # bernoulli distribution
  means[i] = mean(values[,i])
}

hist(means, main="Histogram of Bernoulli Distribution with Parameter 0.5", xlab="x")
```



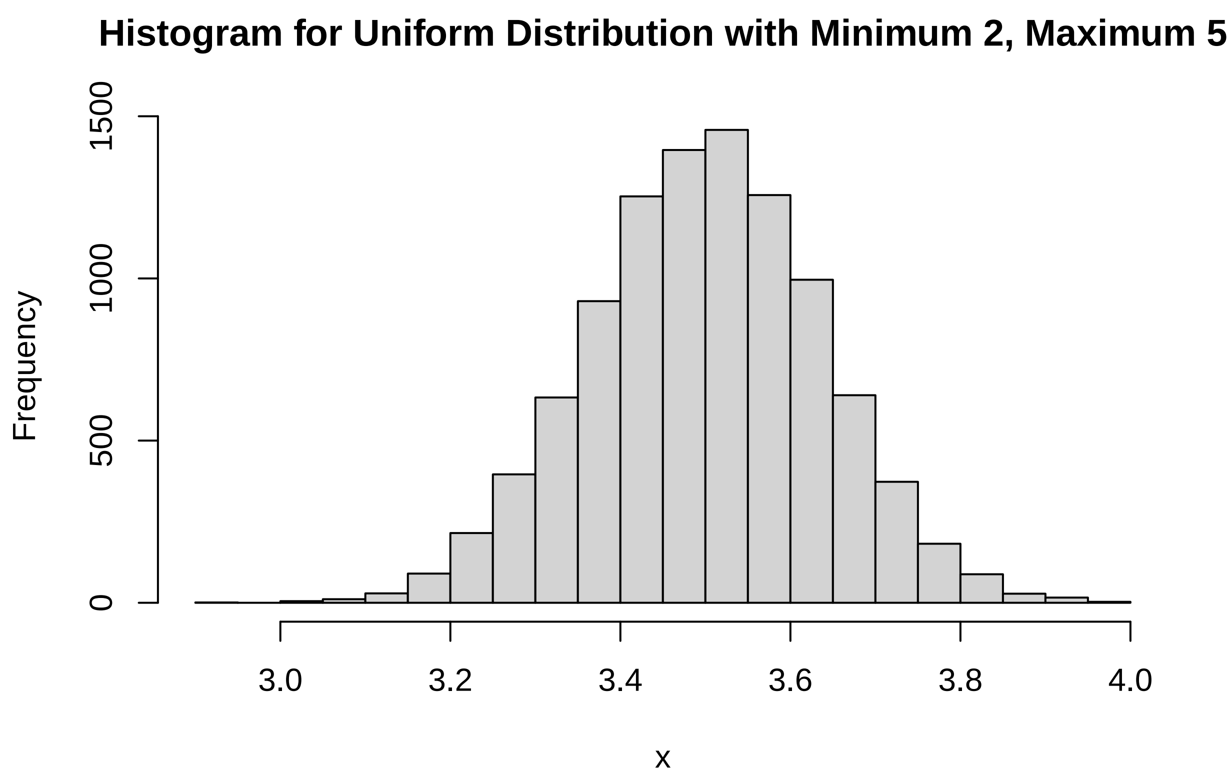
6(ii)

```
n = 10000
s = 40
mean = 5
means = numeric(n)
values = matrix(0, s, n)
for (i in 1:n) {
  values[,i] = rpois(s, mean)  # poisson distribution
  means[i] = mean(values[,i])  # gets the mean for the sample
}

hist(means, main="Histogram of Poisson Distribution with Mean 5", xlab="x")
```



6(iii)

```
n = 10000
s = 40
min = 2
max = 5
means = numeric(n)
values = matrix(0, s, n)

for (i in 1:n) {
  values[,i] = runif(s, min, max)  # uniform distribution
  means[i] = mean(values[,i])
}

hist(means, main="Histogram for Uniform Distribution with Minimum 2, Maximum 5", xlab="x")
```



6(iv)

```r
n = 10000
s = 40
p = 4
means = numeric(n)
values = matrix(0, s, n)

for (i in 1:n) {
  values[,i] = rexp(s, p)  # exponential distribution
  means[i] = mean(values[,i])
}

hist(means, main="Histogram for Exponential Distribution with Parameter 4", xlab="x")
```