# Homework 7

MSSC 6010- Computational Probability

Henri Medeiros Dos Reis

November 2, 2023

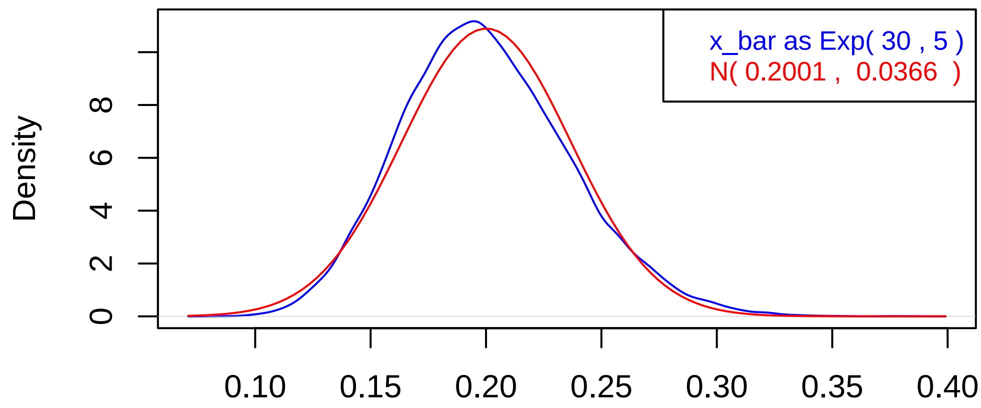
**Question 1.** (6.7.17) From book. Simulate 20,000 random samples of sizes 30,100,300, and 500 from an exponential distribution with a mean of $\frac{1}{5}$. Estimate the density of the distribution of sample mean with the function `density()`. Superimpose a theoretical normal density with appropriate mean and standard deviation. What sample size is needed to get an estimated density close to a normal density?

```
n <- c(30, 100, 300, 500)
n_sims <- 20000

result_matrix <- sapply(n, function(n_val) {
  replicate(n_sims, mean(rexp(n_val, 5)))
})
```

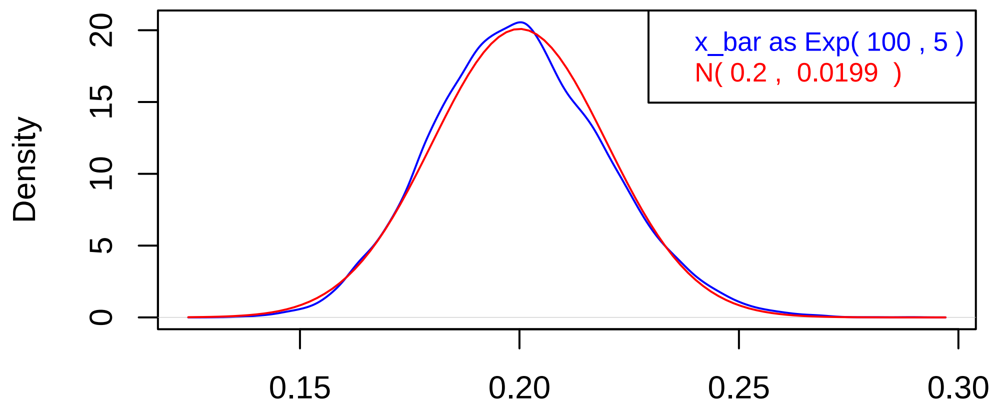
```
for(i in 1:4)
{
plot(density(result_matrix[,i]), col="blue",
     main = paste("Simulated Distribution of x_bar"))
curve(dnorm(x, mean(result_matrix[,i]),
            sd(result_matrix[,i])),col="red", add = TRUE)
legend(x="topright",
       legend=c(
         paste("x_bar as Exp(",n[i],", 5 )"),
         paste("N(",round(mean(result_matrix[,i]),4),
               ", ",round(sd(result_matrix[,i]),4)," )")),
       text.col = c("blue", "red"),
       cex = 0.8)
}
```

## Simulated Distribution of x_bar

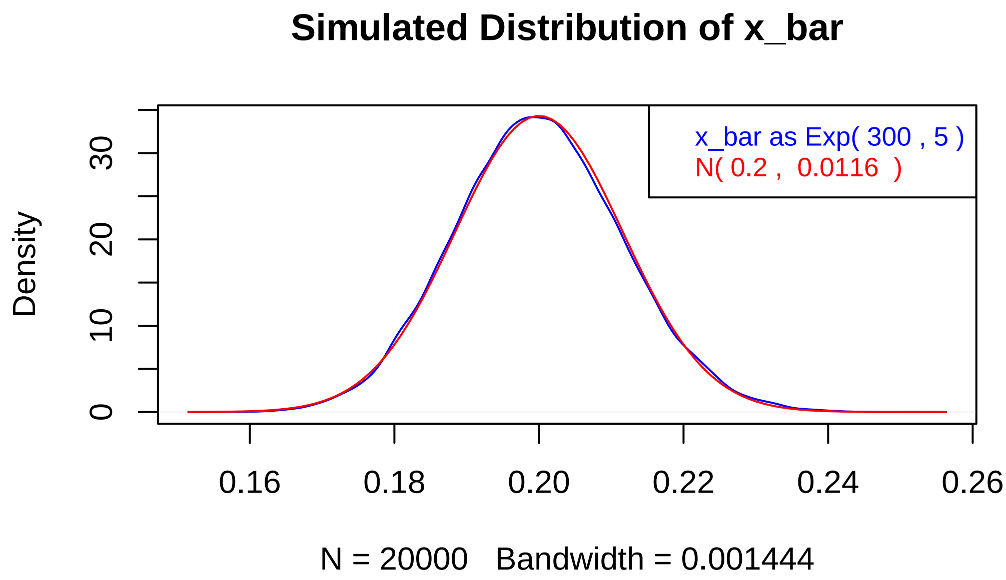


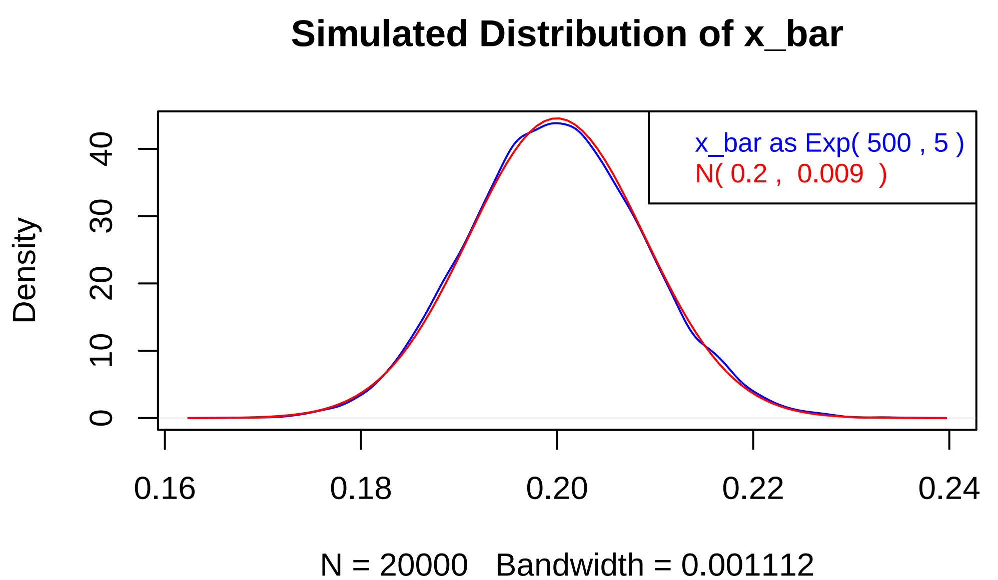

N = 20000 Bandwidth = 0.004547

## Simulated Distribution of x_bar




N = 20000 Bandwidth = 0.002466




All the simulations are close to being normal, but all simulations but $n = 500$ have some positive skewness attached to it. For it to be almost a perfect normal, we want $n = 500$.

**Question 2.** (6.7.23) From book.Consider a random sample of size $n$ from an exponential distribution with parameter $\lambda$. Use moment generating functions to show that the sample mean follows a $\Gamma(n, \lambda n)$. Graph the theoretical sampling distribution of $\bar{X}$ when sampling from an $Exp(\lambda = 1)$for $n = 30, 100, 300, and\ 500$. Superimpose an appropriate normal density for each

$\Gamma(n, \lambda n)$. At what sample size do the sampling distribution and superimposed density virtually coincide?

$$X \sim Exp(\lambda), \text{ then } M_X(t) = (1 - \lambda^{-1}t)^{-1}$$
$$Y \sim \Gamma(\alpha, \lambda), \text{ then } M_Y(t) = (1 - \lambda^{-1}t)^{-n}$$
$$\hat{Y} \sim \Gamma(n, n\lambda), \text{ then } M_{\hat{Y}}(t) = (1 - (n\lambda)^{-1}t)^{-n}$$

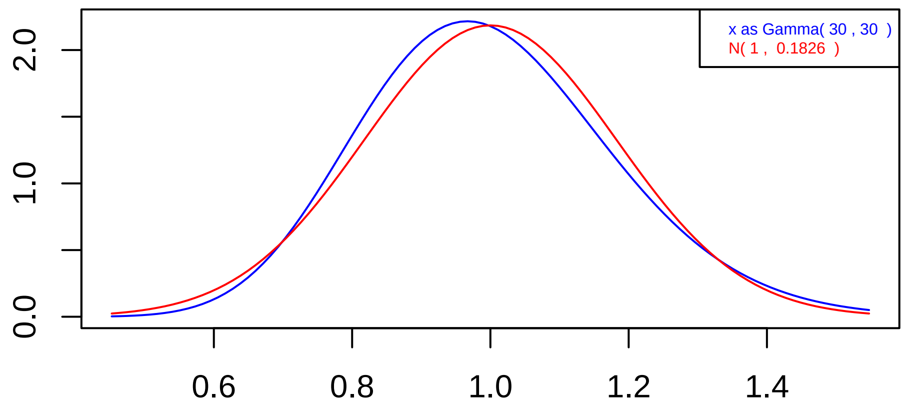
Now finding the moment generating function of $\bar{X}$ gives

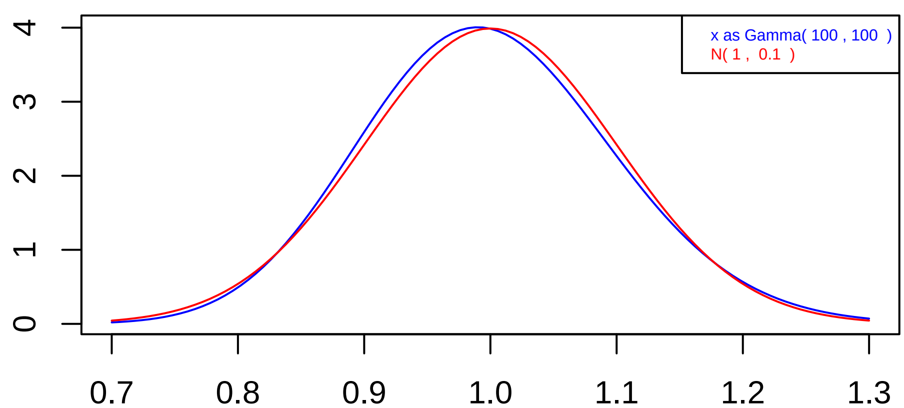
$$M_{\bar{X}}(t) = E[e^{t\bar{X}}] = E[e^{t\sum_{i=1}^{n} X_i/n}] = E[\prod_{i=1}^{n} e^{tX_i/n}] = \prod_{i=1}^{n} E[e^{tX_i/n}] = \prod_{i=1}^{n} M_{X_i}(\frac{t}{n})$$

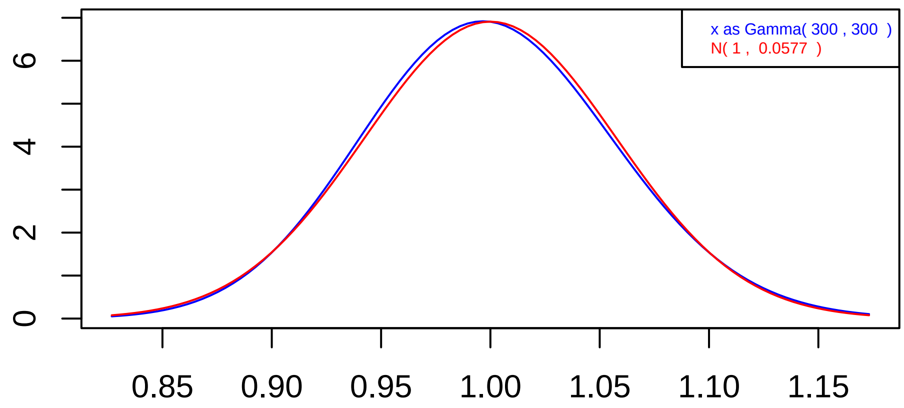
since all $X_i$ are iid. Now plug back to the moment generating function.

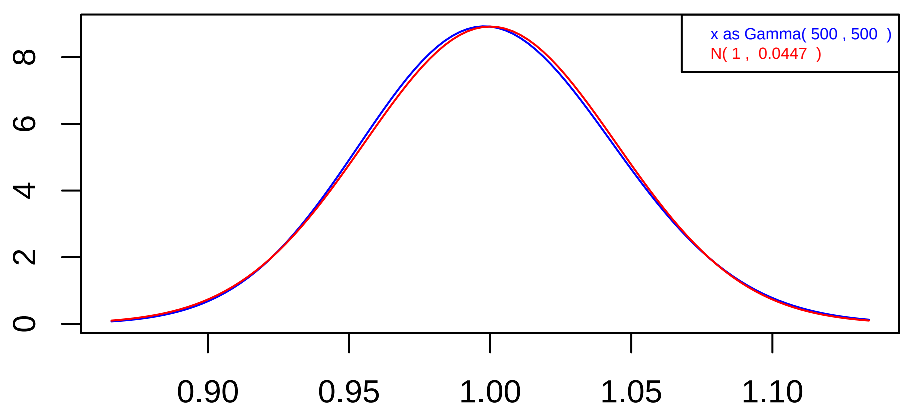
$$\prod_{i=1}^{n}(1 - \lambda^{-1}(t/n))^{-1} = \prod_{i=1}^{n}(1 - (n\lambda)^{-1}t)^{-1} = (1 - (n\lambda)^{-1}t)^{-n} = M_{\hat{Y}}(t)$$

```r
# mean of gamma is n/nlambda var is n/(nlambda)2
n <- c(30,100,300,500)
lambda <- 1
alpha <- n
for(i in 1:4)
{
  curve(dgamma(x,n[i],n[i]*lambda),col="blue",
        from = 1-3*sqrt(n[i]/(n[i]*lambda)2),
        to=1+3*sqrt(n[i]/(n[i]*lambda)2), ylab="",xlab = "")
  curve(dnorm(x, (alpha[i]/lambda)/n[i],sqrt(n[i]/(n[i]*lambda)2)),
        col="red", add=TRUE, ylab="",xlab ="")
  legend(x="topright",
       legend=c(
        paste("x as Gamma(",n[i],",",n[i]," )"),
        paste("N(",1,
              ", ",round(sqrt(n[i]/(n[i]*lambda)2),4)," )")),
       text.col = c("blue", "red"),
       cex = 0.5)
}
```


x as Gamma( 30 , 30 )
N( 1 , 0.1826 )
0.0
1.0
2.0
0.6
0.8
1.0
1.2
1.4


x as Gamma( 100 , 100 )
N( 1 , 0.1 )
0
1
2
3
4
0.7
0.8
0.9
1.0
1.1
1.2
1.3




At $n = 300$, the curves are basically coinciding.

**Question 3.** (6.7.31) From book. A farmer is interested in knowing the mean weight of his chickens when they leave the farm. Suppose that the standard deviation of the chickens' weight is 500 grams.

- *(a)* What is the minimum number of chickens needed to ensure that the standard deviation of the mean is no more than 100 grams?

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{500}{\sqrt{n}} \leq 100 \Rightarrow \sqrt{n} \geq 5 \Rightarrow n \geq 5$$

Therefore, he needs at least 25 chickens to ensure that the standard deviation of the mean in no more than 100g.

- *(b)* If the farm has three coops and the mean chicken weight in each coop is 1.8,1.9, and 2 kg respectively, calculate the probability that a random sample of chickens with an average weight larger than 1.975 kg comes from the first coop. Assume the weight of the chickens follows a normal distribution.

For each coop we will have a normal distribution with mean 1.8,1.9,2 kg and size of 50.

$$\text{Each coop } \sim N(\text{weight}, \frac{\sigma}{\sqrt{n}})$$

We want to find to find $P(\text{coop}|w > 1975)$

$$\mathbb{P}(\text{coop}|w > 1975) = \frac{\mathbb{P}(\text{coop}, w > 1975g)}{\mathbb{P}(w > 1975g)}$$

To find $\mathbb{P}(w > 1975g)$, we need to add the probabilities of the chicken having the necessary weight given a specific coop.

$$\mathbb{P}(w > 1975g) = \sum_{i=1}^{3} \mathbb{P}(w > 1975g|\text{coop}_i)\mathbb{P}(\text{coop}_i)$$

And the probability that a chicken comes from coop 1 and have the necessary weight is the same as the chicken having the necessary weight given coop 1.

$$\mathbb{P}(\text{coop}_1, w > 1975) = \mathbb{P}(w > 1975g|\text{coop}_1)\mathbb{P}(coop_1)$$

Where the probability of any coop is $\frac{1}{3}$

Finally

$$\mathbb{P}(\text{coop}_1|w > 1975) = \frac{\mathbb{P}(w > 1975g|\text{coop}_1)\mathbb{P}(coop_1)}{\sum_{i=1}^{3} \mathbb{P}(w > 1975g|\text{coop}_i)\mathbb{P}(\text{coop}_i)}$$

```r
mean_i <- c(1800,1900,2000)
((1 - pnorm(1975, mean_i[1], 500/sqrt(50)))*1/3)/
  (sum((1 - pnorm(1975, mean_i, 500/sqrt(50)))*1/3))
```

```
[1] 0.008443672
```