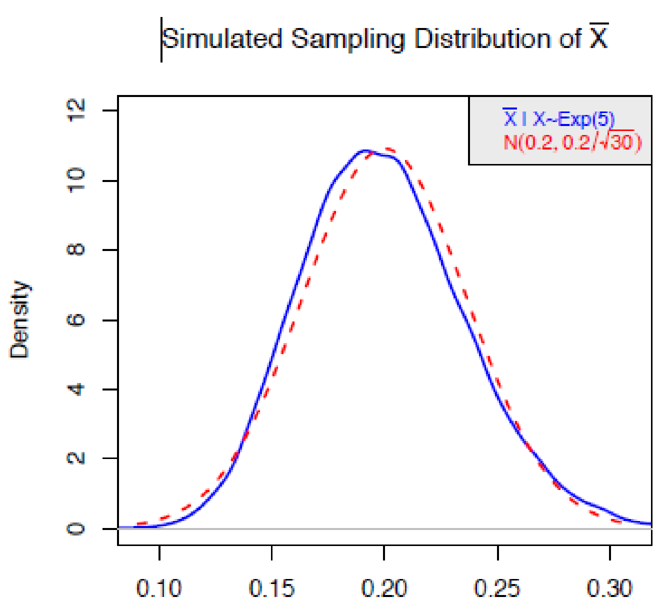
17. Simulate 20,000 random samples of sizes 30, 100, 300, and 500 from an exponential distribution with a mean of $\frac{1}{5}$. Estimate the density of the sampling distribution of the sample mean with the function `density()`. Superimpose a theoretical normal density with appropriate mean and standard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

Solution:

The answer to this question depends on how close you want to be to the normal distribution. What one can observe from the graphs is that there is still a positive skew in the sampling distribution of $\overline{X}$ even for samples of size 300 when sampling from an exponential.

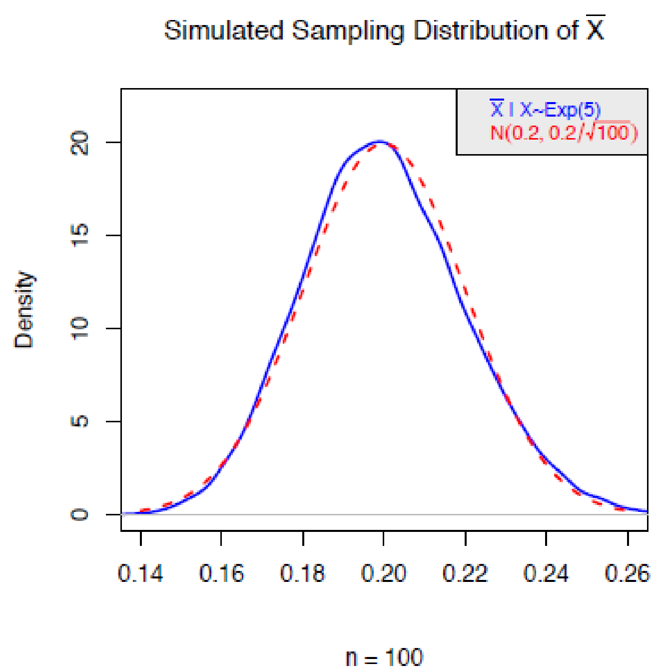
$n = 30$

```
> sims <- 20000
> xbar30 <- numeric(sims)  # storage space
> n <- 30
> for(i in 1:sims){
+   xbar30[i] <- mean(rexp(n, 5))
+ }
> plot(density(xbar30), col = "blue", lwd = 2, xlab = "",
+ ylim=c(0, 1.1*max(density(xbar30)$y)),
+ main = substitute(paste("Simulated Sampling Distribution of ", bar(X))),
+ sub = "n = 30", xlim = c(0.2 - 3*0.2/sqrt(30), 0.2 + 3*0.2/sqrt(30)))
> curve(dnorm(x, 0.2, 0.2/sqrt(30)), add = TRUE, lwd = 2, lty = 2,
+       col = "red")
> legend(x = "topright",
+        legend = c(substitute(paste(bar(X), paste(" | X~Exp(5)"))),
+ expression(N(0.2, 0.2/sqrt(30)))), text.col = c("blue", "red"),
+ bg = "gray92", cex = 0.80)
```



n=100:

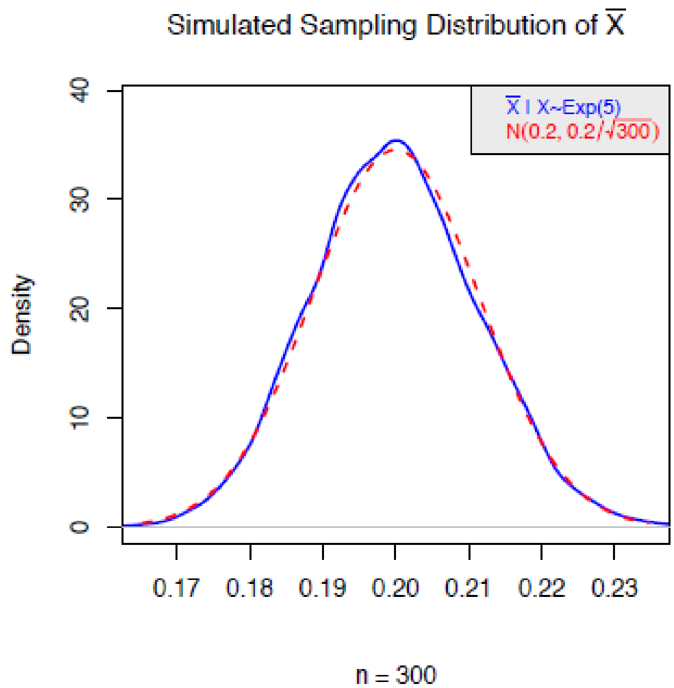
```
> xbar100 <- numeric(sims)  # storage space
> n <- 100
> for(i in 1:sims){
+   xbar100[i] <- mean(rexp(n, 5))
+ }
> plot(density(xbar100), col = "blue", lwd = 2, xlab = "",
+ ylim=c(0, 1.1*max(density(xbar100)$y)),
+ main = substitute(paste("Simulated Sampling Distribution of ", bar(X))),
+ sub = "n = 100", xlim = c(0.2 - 3*0.2/sqrt(100), 0.2 + 3*0.2/sqrt(100)))
> curve(dnorm(x, 0.2, 0.2/sqrt(100)), add = TRUE, lwd = 2, lty = 2,
+       col = "red")
> legend(x = "topright",
+        legend = c(substitute(paste(bar(X), paste(" | X~Exp(5)"))),
+ expression(N(0.2, 0.2/sqrt(100)))), text.col = c("blue", "red"),
+ bg = "gray92", cex = 0.80)
```



$n = 300$

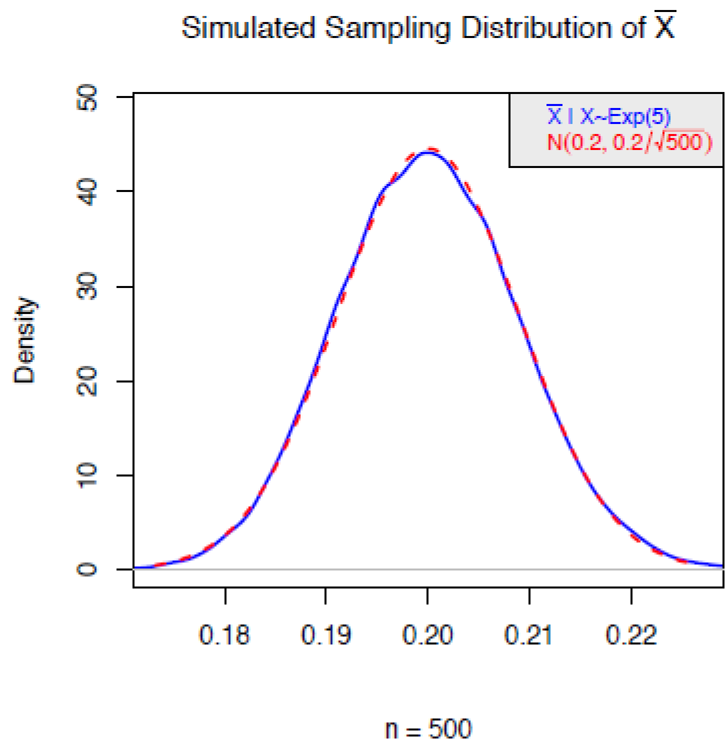
```
> xbar300 <- numeric(sims)  # storage space
> n <- 300
> for(i in 1:sims){
+   xbar300[i] <- mean(rexp(n, 5))
+ }
> plot(density(xbar300), col = "blue", lwd = 2, xlab = "",
+ ylim=c(0, 1.1*max(density(xbar300)$y)),
+ main = substitute(paste("Simulated Sampling Distribution of ", bar(X))),
+ sub = "n = 300", xlim = c(0.2 - 3*0.2/sqrt(300), 0.2 + 3*0.2/sqrt(300)))
> curve(dnorm(x, 0.2, 0.2/sqrt(300)), add = TRUE, lwd = 2, lty = 2,
+       col = "red")
> legend(x = "topright",
```

```
+         legend = c(substitute(paste(bar(X), paste(" | X~Exp(5)"))),
+ expression(N(0.2, 0.2/sqrt(300)))), text.col = c("blue", "red"),
+ bg = "gray92", cex = 0.80)
```



$n = 500$

```
> xbar500 <- numeric(sims)  # storage space
> n <- 500
> for(i in 1:sims){
+   xbar500[i] <- mean(rexp(n, 5))
+ }
> plot(density(xbar500), col = "blue", lwd = 2, xlab = "",
+ ylim=c(0, 1.1*max(density(xbar500)$y)),
+ main = substitute(paste("Simulated Sampling Distribution of ", bar(X))),
+ sub = "n = 500", xlim = c(0.2 - 3*0.2/sqrt(500), 0.2 + 3*0.2/sqrt(500)))
> curve(dnorm(x, 0.2, 0.2/sqrt(500)), add = TRUE, lwd = 2, lty = 2,
+       col = "red")
> legend(x = "topright",
+         legend = c(substitute(paste(bar(X), paste(" | X~Exp(5)"))),
+ expression(N(0.2, 0.2/sqrt(500)))), text.col = c("blue", "red"),
+ bg = "gray92", cex = 0.80)
> par(opar)  # restore original settings
```



23. Consider a random sample of size n from an exponential distribution with parameter λ. Use moment generating functions to show that the sample mean follows a $\Gamma(n, \lambda n)$. Graph the theoretical sampling distribution of $\overline{X}$ 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?

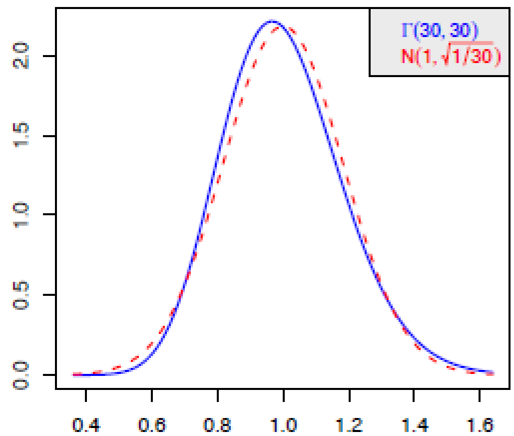
Solution:

For $X \sim Exp(\lambda)$, $M_X(t) = \left(1 - \frac{t}{\lambda}\right)^{-1}$. Also, the moment generating function of a $Y \sim \Gamma(n, \lambda n)$ is $M_Y(t) = \left(1 - \frac{t}{\lambda n}\right)^{-n}$.
Since $\overline{X} = \sum_{i=1}^{n} \frac{X_i}{n}$, $M_{\overline{X}}(t) = E\left[e^{t\overline{X}}\right] = E\left[e^{t\left(\sum_{i=1}^{n} \frac{X_i}{n}\right)}\right] = E\left[\prod_{i=1}^{n} e^{t \cdot \frac{X_i}{n}}\right]$
Because the X_is are independent and identically distributed,

$$M_{\overline{X}}(t) = \prod_{i=1}^{n} E\left[e^{t \cdot \frac{X_i}{n}}\right] = \prod_{i=1}^{n} M_{X_i}\left(\frac{t}{n}\right) = \prod_{i=1}^{n} \left(1 - \frac{t}{n\lambda}\right)^{-1} = \left(1 - \frac{t}{n\lambda}\right)^{-n} = M_Y(t).$$

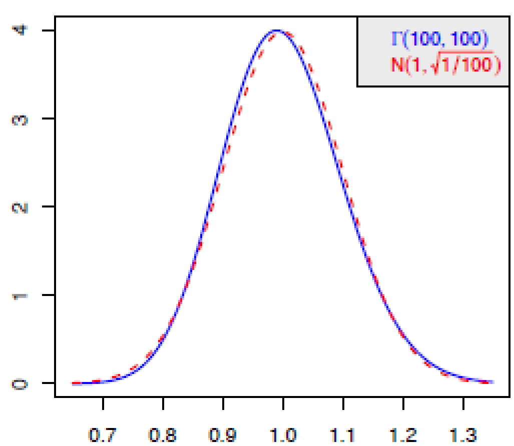
Note that the sampling distributions of the sample mean are a $\Gamma(30, 30)$, $\Gamma(100, 100)$, $\Gamma(300, 300)$, and $\Gamma(500, 500)$ for the sample sizes $n = 30$, 100, 300, and 500, respectively. The normal distribution superimposed over the gamma distributions are $N(1, \sqrt{1/30})$, $N(1, \sqrt{1/100})$, $N(1, \sqrt{1/300})$, and $N(1, \sqrt{1/500})$, respectively, since the mean of the gamma is α/λ and the variance is α/λ^2.
Code for the graph with $\Gamma(30, 30)$:

```
> curve(dgamma(x, 30, 30), from = 1 - 3.5*sqrt(1/30),
+   to = 1 + 3.5*sqrt(1/30), ylab = "", lwd = 2, col = "blue", xlab = "")
> curve(dnorm(x, 1, sqrt(1/30)), from = 1 - 3.5*sqrt(1/30),
+   to = 1 + 3.5*sqrt(1/30), ylab = "", lwd = 2, lty = 2, col = "red",
+   add = TRUE, xlab = "")
> legend(x = "topright", legend = c(expression(Gamma(list(30, 30))),
+                                   expression(N(list(1, sqrt(1/30))))),
+   text.col=c("blue", "red"), bg = "gray92", cex = 0.90)
```



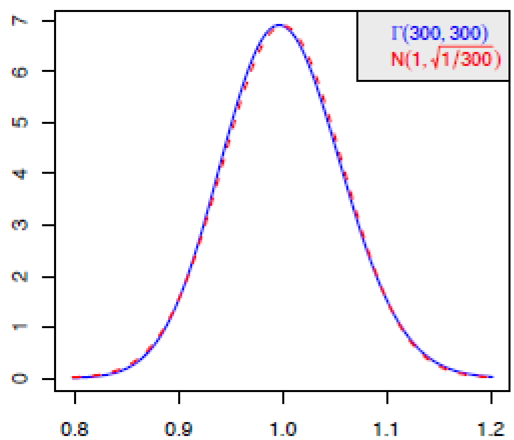
Code for the graph with $\Gamma(100, 100)$:

```
> curve(dgamma(x, 100, 100), from = 1 - 3.5*sqrt(1/100),
+   to = 1 + 3.5*sqrt(1/100), ylab = "", lwd = 2, col = "blue", xlab = "")
> curve(dnorm(x, 1, sqrt(1/100)), from = 1 - 3.5*sqrt(1/100),
+   to = 1 + 3.5*sqrt(1/100), ylab = "", lwd = 2, lty = 2, col = "red",
+   add = TRUE, xlab = "")
> legend(x = "topright", legend = c(expression(Gamma(list(100, 100))),
+                                   expression(N(list(1, sqrt(1/100))))),
+   text.col=c("blue", "red"), bg = "gray92", cex = 0.90)
```



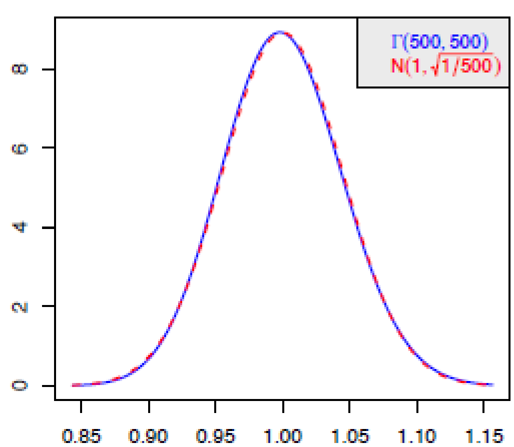
Code for the graph with $\Gamma(300, 300)$:

```
> curve(dgamma(x, 300, 300), from = 1 - 3.5*sqrt(1/300),
+   to = 1 + 3.5*sqrt(1/300), ylab = "", lwd = 2, col = "blue", xlab = "")
> curve(dnorm(x, 1, sqrt(1/300)), from = 1 - 3.5*sqrt(1/300),
+   to = 1 + 3.5*sqrt(1/300), ylab = "", lwd = 2, lty = 2, col = "red",
+   add = TRUE, xlab = "")
> legend(x = "topright", legend = c(expression(Gamma(list(300, 300))),
+                                   expression(N(list(1, sqrt(1/300))))),
+   text.col=c("blue", "red"), bg = "gray92", cex = 0.90)
```



Code for the graph with $\Gamma(500, 500)$:

```
> curve(dgamma(x, 500, 500), from = 1 - 3.5*sqrt(1/500),
+   to = 1 + 3.5*sqrt(1/500), ylab = "", lwd = 2, col = "blue", xlab = "")
> curve(dnorm(x, 1, sqrt(1/500)), from = 1 - 3.5*sqrt(1/500),
+   to = 1 + 3.5*sqrt(1/500), ylab = "", lwd = 2, lty = 2, col = "red",
+   add = TRUE, xlab = "")
> legend(x = "topright", legend = c(expression(Gamma(list(500, 500))),
+                                   expression(N(list(1, sqrt(1/500))))),
+   text.col=c("blue", "red"), bg = "gray92", cex = 0.90)
```



31. 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?

(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 50 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.

Solution:

(a) $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{500}{\sqrt{n}} \leq 100 \Longrightarrow \sqrt{n} \geq 5 \Longrightarrow n \geq 25$, so at least 25 chickens must be sampled to ensure the a standard deviation of the mean is no more than 100 grams with a confidence level of 0.95.

(b) Let $\overline{C}_i$ = mean weight from chicken coop i in grams where $i = 1, 2, 3$ and C_i be the event chickens coming from coop i.

Let W = overall average weight.

Then, $\overline{C}_1 \sim N(1800, 500/\sqrt{50})$, $\overline{C}_2 \sim N(1900, 500/\sqrt{50})$, and ; $\overline{C}_3 \sim N(2000, 500/\sqrt{50})$.

The problem wishes to discover $\mathbb{P}(C_1 \,|\, W > 1975)$

$$\begin{aligned}
\mathbb{P}(C_1 \,|\, W > 1975) &= \frac{\mathbb{P}(C_1, W > 1975)}{\mathbb{P}(W > 1975)} \\
&= \frac{\mathbb{P}(W > 1975 \,|\, C_1)\,\mathbb{P}(C_1)}{\sum_{i=1}^{3} \mathbb{P}(W > 1975 \,|\, C_i)\,\mathbb{P}(C_i)} \\
&= \frac{0.0022}{0.2631} = 0.0084
\end{aligned}$$

```
> num <- (1 - pnorm(1975, 1800, 500/sqrt(50)))*1/3
> num

[1] 0.002221388

> den <- (1 - pnorm(1975, 1800, 500/sqrt(50)))*1/3 +
+         (1 - pnorm(1975, 1900, 500/sqrt(50)))*1/3 +
+         (1 - pnorm(1975, 2000, 500/sqrt(50)))*1/3
> den

[1] 0.2630832

> PC1givnW <- num/den
> PC1givnW

[1] 0.008443672
```