S2S Lab 6 Task Solutions

1 Welcome!

2 Sampling Distribution of $\bar{X}$

2.1 Central Limit Theorem

If we were to take samples of size $n = 2$ from the Unif$(-11, 41)$ distribution, what distribution would $\bar{X}$ follow, according to the Central Limit Theorem?

If $X \sim \text{Unif}(-11, 41)$, then we can show that,

- $E[X] = 15 = \mu$
- $sd(X) \approx 15 = \sigma$

We can then use these values to show that, according to the Central Limit Theorem,

$$\bar{X} \sim N\left(15, \frac{15}{\sqrt{2}}\right)$$

If we were to take samples of size $n = 2$ from the Expo$\left(\frac{1}{15}\right)$ distribution, what distribution would $\bar{X}$ follow, according to the Central Limit Theorem?

If $X \sim \text{Expo}\left(\frac{1}{15}\right)$, then we can show that,

- $E[X] = 15 = \mu$
- $sd(X) = 15 = \sigma$

We can then use these values to show that, according to the Central Limit Theorem,

$$\bar{X} \sim N\left(15, \frac{15}{\sqrt{2}}\right)$$

Do you think the the sampling distribution of $\bar{X}$, when samples of size $n = 2$ are taken from the Unif$(-11, 41)$ distribution, is approximately normal?

- Yes

What about the sampling distribution of $\bar{X}$ when samples of size $n = 2$ are taken from the Expo$\left(\frac{1}{15}\right)$ distribution - is it approximately normal?

- No

Complete the code below to draw 10,000 samples of size $n = 100$ from the Unif$(-11, 41)$ distribution, and another 10,000 samples of size $n = 100$ from the Expo$\left(\frac{1}{15}\right)$ distribution.

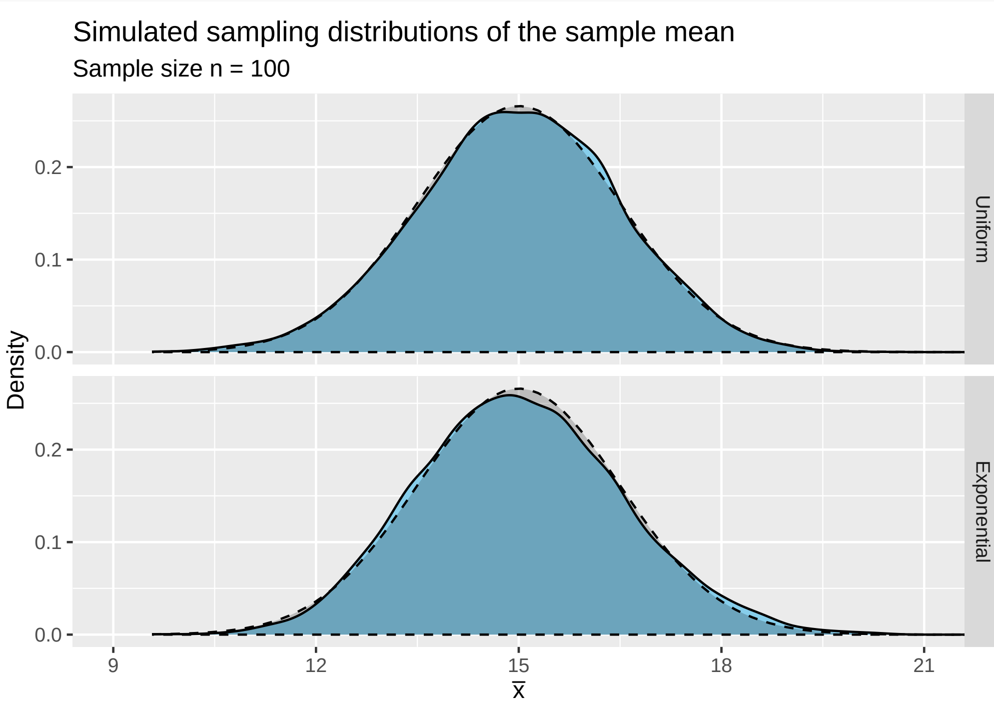
```
means_unif_100 <- numeric(m)
means_expo_100 <- numeric(m)

for(i in 1:m){
  means_unif_100[i] <- mean(runif(n = 100, min = -11, max = 41))
}

for(i in 1:m){
  means_expo_100[i] <- mean(rexp(n = 100, rate = 1/15))
}
```

Complete the code below to plot the sampling distribution of $\bar{X}$ for the samples of size $n = 100$ taken from the Unif$(-11, 41)$ and Expo$\left(\frac{1}{15}\right)$ distributions. Superimpose the normal distribution that each sampling distribution follows according to the Central Limit Theorem.

```
ggplot(data = subset(means, subset = (size == "n = 100"))) +
  geom_density(aes(x = mean), fill = "skyblue") +
  facet_grid(distribution ~ .) +
  coord_cartesian(xlim = c(9, 21)) +
  labs(title = "Simulated sampling distributions of the sample mean",
       subtitle = "Sample size n = 100",
       x = expression(bar(x)), y = "Density") +
  stat_function(fun = dnorm, args = list(mean = 15, sd = 15/sqrt(100)),
                geom = "area", col = "black", linetype = 2, fill = "black",
                alpha = 0.2)
```



3 Sampling distribution for $\bar{X} - \bar{Y}$

Using the statement above about the distribution of $\bar{X} - \bar{Y}$, what distribution would you expect the differences in sample means to follow?

We know that $X \sim N(90, 12)$ and that we want to take samples of size 120 from this distribution. Therefore, we can say that $\mu_X = 90$, $\sigma_X = 12$ and $n_X = 120$.

We also know that $Y \sim N(40, 10)$ and that we want to take samples of size 96 from this distribution, so $\mu_Y = 40$, $\sigma_Y = 10$ and $n_Y = 96$.

Plugging these values into the given distribution for $\bar{X} - \bar{Y}$ tells us that,

$$\bar{X} - \bar{Y} \sim N\left(90 - 40 = 50, \sqrt{\tfrac{12^2}{120} + \tfrac{10^2}{96}}\right)$$

Complete the code below draw 20,000 samples of size 120 from the $N(90, 12)$ distribution and save the mean of each sample in a vector called `meansX`.

Let `m` represent the number of samples you want to draw, and `nx` represent the size each sample should be i.e. n_X.

```
m <- 20000
nx <- 120

meansX <- numeric(m)

for(i in 1:m){
  meansX[i] <- mean(rnorm(nx, mean = 90, sd = 12))
}
```

Do you think the kernel density estimate of the distribution of the differences between pairs of sample means is approximately equal to the $N\left(50, \sqrt{\tfrac{12^2}{120} + \tfrac{10^2}{96}}\right)$ distribution?

- Yes