The Central Limit Theorem (CLT)

Let $X_1, X_2, ..., X_n$ be a sequence of independent and identically distributed $(i.i.d)$ random variables with $E[X] = \mu$, $< \infty$ (finite first moment), and $Var(X) = \sigma^2$ exists.

(CLT) As n becomes large, then the distribution of $\frac{1}{n}\sum_{i=1}^{n} x_i$ (the sample mean) is approximately a Normal distribution $N\left(\mu, \frac{\sigma^2}{n}\right)$.

Useful direct consequence

As a consequence, we also have that

$$\frac{\sum_{i=1}^{n}(x_i - \mu)}{\sqrt{n}\ \sigma} \xrightarrow{L} N(0,1)$$

Indeed, we have that

$$\sum_{i=1}^{n}(x_i - \mu) \xrightarrow{L} N(0, n\sigma^2)$$

$$\sum_{i=1}^{n} x_i \xrightarrow{L} N(n\mu, n\sigma^2)$$

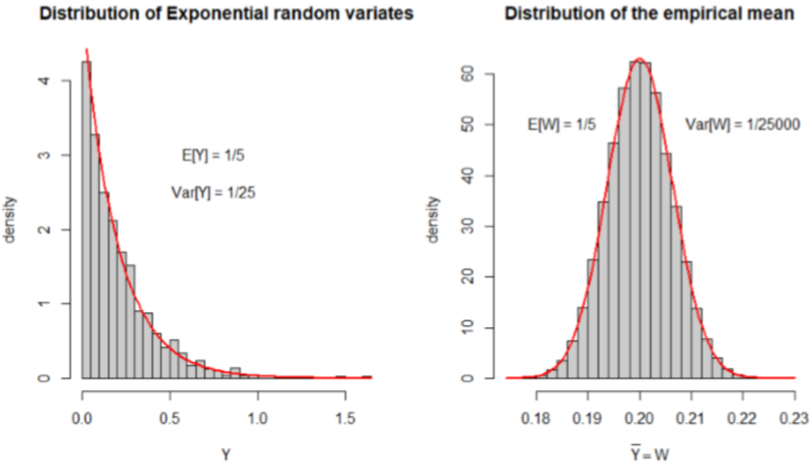
$$\frac{1}{n}\sum_{i=1}^{n} x_i \xrightarrow{L} N\left(\mu, \frac{\sigma^2}{n}\right)$$

Example 1: Exponential variables

Example 1 Let $y_1, y_2, ..., y_n$ be $i.i.d$ exponential realizations of $Y_i \sim \mathcal{E}(\lambda = 5)$ and $n = 1,000$. The theoretical mean is $E[Y] = 1/\lambda = 1/5 = 0.2$ and the theoretical variance is $Var(Y) = 1/5^2$. Then we expect the sample average $\frac{1}{n}\sum_{i=1}^{n} y_i$ to converge to a Normal distribution $N\left(E[Y] = \mu = \frac{1}{5}, Var(Y) = \frac{\sigma^2}{n} = \frac{1}{25{,}000}\right)$

In other words: The mean of the realizations of our Exponential random variables will obey a Normal distribution.

Example 1: Visualization



Example 2: Application 1/3

Example 2 According to a certain maritime organization, the distribution of the length of a specific fish species at maturity in the Mekong delta has mean $\mu = 83$ cm. and standard deviation $\sigma = 7$ cm. The random variable Y denotes the length of these fishes.

(i) Suppose we sample 25 individuals. What is the probability that the sample average is above 86 cm. ?

(i) From the CLT, we know that the distribution of the sample average $\frac{1}{25}\sum_{i=1}^{25} y_i$ has a Normal distribution
$N\left(\mu = 83, \frac{\sigma}{\sqrt{n}} = \frac{7}{5}\right)$.

So, $P\left(\overline{y} > 86\right) = P\left(Z > \frac{\overline{y}-\mu}{\sigma/\sqrt{n}}\right) = P\left(Z > \frac{86-83}{1.4}\right) = P\left(Z > 2.142857\right) = \mathbf{0.01606 \approx 1.6\%}$

```
pnorm(q=86, mean=83, sd=7/5, lower.tail = FALSE)
# 0.01606
```

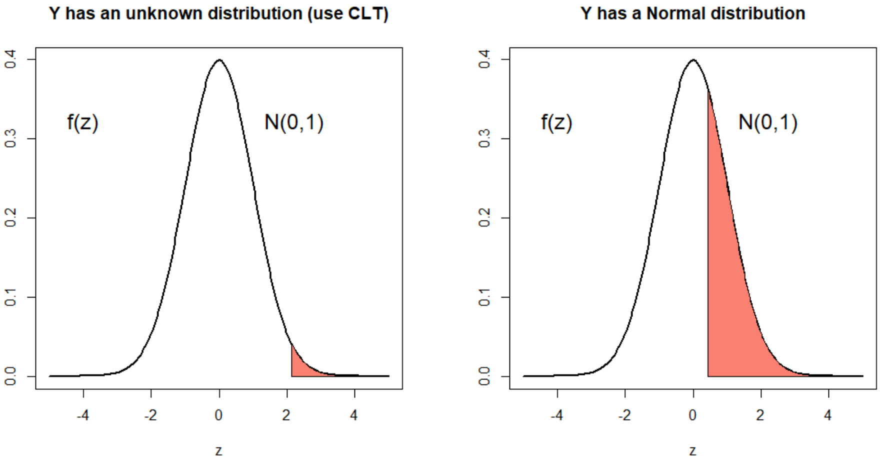
Bayesian Questionnaire: Question 3

According to a certain maritime organization, the distribution of the length of a specific fish species at maturity in the Mekong delta has mean $\mu = 83$ cm. and standard deviation $\sigma = 7$ cm. The random variable Y denotes the length of these fishes.

If Y had a Normal distribution, we would have $P\Big(\overline{y} > 86\Big) = P\Big(Z > \frac{86-83}{7}\Big)$. What is this probability ? You can do it in R or use any other ressource.

A: 25.09 %
B: 33.41 %
C: 6.54 %
D: 60 %

Example 2: Application 1/3



Example 2: Application 2/3

Example 2 (continued) According to a certain maritime organization, the distribution of the length of a specific fish species at maturity in the Mekong delta has mean $\mu = 83$ cm. and standard deviation $\sigma = 7$ cm. The random variable Y denotes the length of these fishes.

(ii) Give a 95% Confidence Interval for the sample average length.

(ii) From the CLT, we know that the distribution of the sample average $\frac{1}{25}\sum_{i=1}^{25} y_i$ has a Normal distribution
$N\left(\mu = 83, \frac{\sigma}{\sqrt{n}} = \frac{7}{5}\right)$.

So, $P\left(-1.96 > Z > 1.96\right) = 95\% \quad \Leftrightarrow \quad P\left(\mu - 1.96 * \sigma/\sqrt{n} > \overline{y} > \mu + 1.96 * \sigma/\sqrt{n}\right) = 95\% \quad \Leftrightarrow \quad \left[\mathbf{80.26, 85.74}\right]$

```
c((83 - 1.96*(7/5)), (83 + 1.96*(7/5)))
# [1] 80.256 85.744
```

Example 2: Application 3/3

Example 2 (continued) According to a certain maritime organization, the distribution of the length of a specific fish species at maturity in the Mekong delta has mean $\mu = 83$ cm. and standard deviation $\sigma = 7$ cm. The random variable Y denotes the length of these fishes.

(iii) How many mature tuna must we sample if we want a 95 % confidence that the sample average is within 1 cm. from the population mean ?

(iii) From the CLT, we know that the distribution of the sample average $\frac{1}{25}\sum_{i=1}^{n} y_i$ has a Normal distribution
$N\left(\mu = 83, \frac{\sigma}{\sqrt{n}} = \frac{7}{\sqrt{n}}\right)$.

So, $Z = \frac{\overline{y} - \mu}{\sigma/\sqrt{n}} \quad \Leftrightarrow \quad 1.96 = \frac{1}{7/\sqrt{n}} \quad \Leftrightarrow \quad n = (1.96 * 7)^2 \approx \mathbf{189}$ fishes

So we replace all we know and solve for n, the sample size.