University of Texas at Austin

Problem Set # 7

The Central Limit Theorem.

Let $\{X_n, n = 1, 2, 3, \dots\}$ be a sequence of independent, identically distributed random variables such that $\mu_X = \mathbb{E}[X_1] < \infty$ and $Var[X] = \sigma_X^2 < \infty$. For every $n = 1, 2, \dots$ define

$$\bar{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n}.$$

Problem 7.1. Find the expected value of $\bar{X}_n$ for every n.

Solution:

$$\begin{aligned}
\mathbb{E}[\bar{X}_n] &= \mathbb{E}\left[\frac{X_1 + X_2 + \cdots + X_n}{n}\right] \textbf{ (by definition of } \bar{X}_n\textbf{)}\\
&= \frac{1}{n}\mathbb{E}\left[X_1 + X_2 + \cdots + X_n\right] \textbf{ (by linearity of the expectation)}\\
&= \frac{1}{n}(\mathbb{E}[X_1] + \mathbb{E}[X_2] + \cdots + \mathbb{E}[X_n])\textbf{(by linearity of the expectation)}\\
&= \frac{1}{n}(n\mathbb{E}[X_1])\textbf{(because all } X_i \textbf{ are identically distributed)}\\
&= \mu_X.
\end{aligned}$$

Problem 7.2. Find the variance and standard deviation of $\bar{X}_n$ for every n.

Solution:

$$\begin{aligned}
Var[\bar{X}_n] &= Var\left[\frac{X_1 + X_2 + \cdots + X_n}{n}\right] \textbf{ (by definition of } \bar{X}_n\textbf{)}\\
&= \frac{1}{n^2}Var\left[X_1 + X_2 + \cdots + X_n\right] \textbf{ (straight from the definition of the variance)}\\
&= \frac{1}{n^2}(Var[X_1] + Var[X_2] + \cdots + Var[X_n])\textbf{(because } X_i \textbf{ are all independent)}\\
&= \frac{1}{n^2}(nVar[X_1])\textbf{(because } X_i \textbf{ are all identically distributed)}\\
&= \frac{\sigma_X^2}{n}.
\end{aligned}$$

So, $SD[\bar{X}_n] = \frac{\sigma_X}{\sqrt{n}}$.

Theorem 7.1. The Central Limit Theorem (CLT). *If the above conditions are satisfied, we have that*

$$\frac{\bar{X}_n - \mu_X}{\frac{\sigma_X}{\sqrt{n}}} \overset{\mathcal{D}}{\Rightarrow} N(0,1) \quad \textit{as } n \to \infty.$$

Practically, for "large enough" n, $\bar{X}_n$ is approximately normal with mean μ_X and variance $\frac{\sigma_X^2}{n}$. The rule of thumb is that we use the theorem for $n \geq 30$. If that is the case, we have that for any real $a < b$,

$$\mathbb{P}[a < \bar{X}_n \leq b] = \mathbb{P}\left[\frac{a - \mu_X}{\frac{\sigma_X}{\sqrt{n}}} < \frac{\bar{X}_n - \mu_X}{\frac{\sigma_X}{\sqrt{n}}} \leq \frac{b - \mu_X}{\frac{\sigma_X}{\sqrt{n}}}\right] \approx \Phi\left(\frac{b - \mu_X}{\frac{\sigma_X}{\sqrt{n}}}\right) - \Phi\left(\frac{a - \mu_X}{\frac{\sigma_X}{\sqrt{n}}}\right).$$

Problem 7.3. Travel time by sled between Whoville and Whoburgh takes on average 36 minutes with a standard deviation of 6 minutes. Over a particular weekend, 64 sled trips take place. What is the (approximate) probability that the average sled trip took more than 38 minutes?

Solution: Let $\bar{X}_n$ denote the average sled trip with $n = 64$. We know that $\mathbb{E}[\bar{X}_n] = 36$ and $SD[\bar{X}_n] = \frac{6}{\sqrt{64}} = 0.75$. By the CLT, we have

$$\mathbb{P}[\bar{X}_n > 38] \approx 1 - \Phi\left(\frac{38-36}{0.75}\right) = 1 - \Phi(\tfrac{8}{3}).$$

Using **pnorm** in **R**, we get 0.003830381.

Problem 7.4. The amount of time your friendly taquero at *Torchy's Tacos* spends to assemble any one tasty taco is a random variable with mean 3 minutes and 15 seconds and standard deviation of thirty seconds. You and your 31 friends from *Applied Statistics* celebrate by ordering two tacos each. What is the probability that the average taco-assembly time is:

- less than 2 minutes and 30 seconds;
- more than 3 minutes and 15 seconds;
- at least 3 minutes but at most 3 minutes and 30 seconds?

Solution: Let $\bar{X}_n$ with $n = 64$ be the random variable representing the average taco-assembly time. Then,

$$\mathbb{E}[\bar{X}_n] = 3.25 \quad \text{and} \quad SD[\bar{X}_n] = \frac{0.5}{\sqrt{64}} = 0.0625.$$

Using the CLT, we get the following:

- For $\mathbb{P}[\bar{X}_n < 2.5]$, in **R** we write **pnorm(2.5,3.25,0.0625)** to get 1.776482×10^{-33}.
- For $\mathbb{P}[\bar{X}_n > 3.25]$, we immediately get 1/2.
- For $\mathbb{P}[3 < \bar{X}_n < 3.5]$, in **R** we write **pnorm(3.5,3.25,0.0625)-pnorm(3,3.25,0.0625)** to get 0.9999367.