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.

$$\mathbb{E}[\bar{X}_n] = \mathbb{E}\left[\frac{1}{n}(X_1 + X_2 + \cdots + X_n)\right]$$

$$= \frac{1}{n}\mathbb{E}[X_1 + X_2 + \cdots + X_n] \quad \text{linearity}$$

$$= \frac{1}{n}\left(\mathbb{E}[X_1] + \mathbb{E}[X_2] + \cdots + \mathbb{E}[X_n]\right) \quad \text{identically dist'd.}$$

$$= \frac{1}{\cancel{n}}(\cancel{n}\cdot\mu_X) = \mu_X$$

accuracy

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

$$Var[\bar{X}_n] = Var\left[\frac{1}{n}(X_1 + X_2 + \cdots + X_n)\right] =$$

$$= \frac{1}{n^2}\cdot Var[X_1 + X_2 + \cdots + X_n] \quad \text{independence}$$

$$= \frac{1}{n^2}\left(Var[X_1] + Var[X_2] + \cdots + Var[X_n]\right) \quad \text{identically dist'd}$$

$$= \frac{1}{n^{\cancel{2}}}\cdot(\cancel{n}\cdot\sigma_X^2) = \frac{\sigma_X^2}{n}$$

$$SD[\bar{X}_n] = \boxed{\frac{\sigma_X}{\sqrt{n}}}$$

precision

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

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

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 $\boxed{n \geq 30}$. If that is the case, we have that for any real $a < b$,

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

FTC

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?

$\longrightarrow$: $n = 64 \geq 30$

$\bar{X}_n$ "≈" Normal (mean = $\underline{36}$, sd = $\frac{6}{\sqrt{64}} = \frac{3}{4}$)

CLT

$\approx N(0,1)$

$$\mathbb{P}[\bar{X}_n > 38] = 1 - \mathbb{P}[\bar{X}_n \leq 38] = 1 - \mathbb{P}\left[\frac{\bar{X}_n - 36}{0.75} \leq \frac{38-36}{0.75} = \frac{8}{3}\right]$$

$$= 1 - \Phi(2.67) = 1 - 0.9962 = 0.0038$$

$$= 1 - \text{pnorm}(8/3) = 0.00383$$

Alternatively: 1-pnorm(38, mean=36, sd=0.75) = 0.00383

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?

$\underline{n=64}$: $\bar{X}_n \approx$ Normal (mean = $\underline{3.25}$, sd = $\frac{0.5}{\sqrt{64}} = 0.0625$)

$$\mathbb{P}[\bar{X}_n < 2.5] = ?$$

pnorm(2.5, mean=3.25, sd=0.0625) = $1.776 \cdot 10^{-33}$