M378K: October 9th, 2024.

M378K Introduction to Mathematical Statistics
Problem Set #10
The Central Limit Theorem (CLT).

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

$$S_n = X_1 + X_2 + \cdots + X_n$$

and

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

✓

Problem 10.1. *Find the expected value of S_n and X_n for every n.*

$\mathbb{E}[\bar{X}_n] = \mu_X$ $\boxed{\text{accuracy}}$

Problem 10.2. *Find the variance and standard deviation of S_n and $\bar{X}_n$ for every n.*

$SD[S_n] = \sigma_X \sqrt{n}$

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

Theorem 10.1. **The Central Limit Theorem (CLT).** *If the above conditions are satisfied, we have that*

$$\frac{S_n - n\mu_X}{\sigma_X\sqrt{n}} = \frac{\bar{X}_n - \mu_X}{\frac{\sigma_X}{\sqrt{n}}} \xrightarrow{\mathcal{D}} N(0,1) \quad \text{as } n \to \infty.$$

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 $l < r$,

$$\mathbb{P}[l < S_n \leq r] = \mathbb{P}\left[\frac{l - n\mu_X}{\sigma_X\sqrt{n}} < \boxed{\frac{S_n - n\mu_X}{\sigma_X\sqrt{n}}} \leq \frac{r - n\mu_X}{\sigma_X\sqrt{n}}\right] \approx \Phi\left(\frac{r - n\mu_X}{\sigma_X\sqrt{n}}\right) - \Phi\left(\frac{l - n\mu_X}{\sigma_X\sqrt{n}}\right).$$

$\overset{??}{Z} \sim N(0,1)$

Similarly, for any real $a < b$,

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

Problem 10.3. *The* Really Terrible Orchestra[1] *plans a concert at a gazebo in a local park. The orchestra has* 169 *members whose weights are assumed to be independent and identically distributed with mean* 100 *kilos and standard deviation of* 10 *kilos (the weight of the instruments is taken into account here). The gazebo can safely support up to* 17 *tons (each ton is* 1000 *kilos). What is the approximate probability that the gazebo will collapse?*

→: $S = Y_1 + Y_2 + \cdots + Y_{169}$ w/ Y_i has $\mu_Y = 100$ and $\sigma_Y = 10$

$\mathbb{E}[S] = 169 \cdot 100$ and $SD[S] = 10\sqrt{169} = 130$

0.77

$\mathbb{P}[S > 17000] = 1 - \mathbb{P}[S \le 17000] = 1 - \mathbb{P}\left[\underbrace{\frac{S-16900}{130}}_{\approx N(0,1)} \le \frac{17000-16900}{130}\right]$

In R: z = (17000-16900)/130
1-pnorm(z) ... 0.22

In the tables:
$1 - 0.7794 \approx 0.2206$

Problem 10.4. Source: Sample P exam, Problem #65.
A charity receives 2025 *contributions. Contributions are assumed to be mutually independent and identically distributed with mean* 3125 *and standard deviation* 250. *Calculate the approximate* 90^{th} *percentile for the distribution of the total contributions received.*

→: $S = Y_1 + \cdots + Y_n$ $n = 2025$ w/ Y_i s.t. $\mu_Y = 3125$, $\sigma_Y = 250$

$\mathbb{E}[S] = (2025)(3125) = \mu_S$

$SD[S] = 250\sqrt{2025} = 250(45) = \sigma_S$

$\pi = ?$ such that $\mathbb{P}[S \le \pi] \approx 0.90$

1st Find the 90th percentile of $Z \sim N(0,1)$

$\boxed{z^* = 1.28}$ from the std normal tables

2nd $\mathbb{P}[Z \le z^*] = 0.90$

[1] http://thereallyterribleorchestra.com/wordpress/

$\mathbb{P}[\boxed{\mu_S + \sigma_S \cdot Z} \le \mu_S + \sigma_S \cdot z^*] = 0.90$

$\mathbb{P}[S \le \boxed{2025(3125) + 250(45) \cdot 1.28}] \approx 0.90$

$\boxed{6342525}$
answer