

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 $\text{Var}[X] = \sigma_X^2 < \infty$. For every $n = 1, 2, \dots$ define

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

and

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

Problem 10.1. Find the expected value of S_n and \bar{X}_n for every n .

Solution:

$$\begin{aligned}\mathbb{E}[S_n] &= \mathbb{E}[X_1 + X_2 + \dots + X_n] \text{ (by definition of } S_n) \\ &= \mathbb{E}[X_1] + \mathbb{E}[X_2] + \dots + \mathbb{E}[X_n] \text{ (by linearity of the expectation)} \\ &= n\mathbb{E}[X_1] \text{ (because all } X_i \text{ are identically distributed)} \\ &= n\mu_X.\end{aligned}$$

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

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

Solution:

$$\begin{aligned}\text{Var}[S_n] &= \text{Var}[X_1 + X_2 + \dots + X_n] \text{ (by definition of } S_n) \\ &= \text{Var}[X_1] + \text{Var}[X_2] + \dots + \text{Var}[X_n] \text{ (because } X_i \text{ are all independent)} \\ &= n\text{Var}[X_1] \text{ (because } X_i \text{ are all identically distributed)} \\ &= n\sigma_X^2.\end{aligned}$$

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

Similarly,

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

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

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

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

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 $l < r$,

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

Similarly, 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 10.3. *The Really Terrible Orchestra¹ 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?*

Solution: Let the individual players' weights be Y_1, \dots, Y_{169} and let's define

$$S = Y_1 + \cdots + Y_{169}.$$

¹<http://thereallyterribleorchestra.com/wordpress/>

Then, by the CLT, S is approximately normal with mean $100(169) = 16900$ and standard deviation $10\sqrt{169} = 130$. So, we have that

$$\mathbb{P}[S \leq 180000] \approx 1 - \Phi\left(\frac{170000 - 169000}{130}\right) = 1 - \Phi(0.77) \approx 0.2206.$$

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 90th percentile for the distribution of the total contributions received.

Solution: Let Y_1, \dots, Y_n correspond to individual contributions with $n = 2025$. We know that, by the CLT,

$$S = Y_1 + \dots + Y_n$$

is approximately normal with mean $\mu_S = (2025)(3125)$ and standard deviation $\sigma_S = 250\sqrt{2025} = 250(45)$. We are looking for the approximate 90th percentile of S , i.e., the value π such that, with the CLT approximation, we have

$$\mathbb{P}[S \leq \pi] \approx 0.90.$$

The 90th percentile of the standard normal is about 1.28, i.e., with $Z \sim N(0, 1)$, we have that

$$\begin{aligned} \mathbb{P}[Z \leq 1.28] = 0.90 &\Leftrightarrow \mathbb{P}[\mu_S + \sigma_S Z \leq (2025)(3125) + 250(45)(1.28)] = 0.90 \\ &\Leftrightarrow \mathbb{P}[S \leq 6342525] = 0.90 \end{aligned}$$

Our answer is about 6,342,525.