Central Limit Theorem Examples

Probability Theory: Foundation for Data Science

with **Anne Dougherty**

Data Science
UNIVERSITY OF COLORADO **BOULDER**

Central Limit Theorem

At the end of this module, students should be able to

- Understand the definition of a random sample.
- Understand the Law of Large Numbers.
- Understand and use the Central Limit Theorem (CLT).
- Explain the implications of the CLT to the calculation and estimation of the mean.

Proposition: If $X_1, X_2, \ldots, X_n$ are iid with $X_i \sim N(\mu, \sigma^2)$ then $\bar{X} \sim N(\mu, \sigma^2/n)$.

Proposition: If $X_1, X_2, \ldots, X_n$ are independent with $X_i \sim N(\mu_i, \sigma_i^2)$ then $\sum_{i=1}^{n} X_i \sim N(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2)$.

Suppose you have 3 errands to do in three different stores. Let T_i be the time to make the i^{th} purchase for $i = 1, 2, 3$. Let T_4 be the total walking time between stores. Suppose $T_1 \sim N(15, 16)$, $T_2 \sim N(5, 1)$, $T_3 \sim N(8, 4)$, and $T_4 \sim N(12, 9)$. Assume T_1, T_2, T_3, T_4 are independent . If you leave at 10 in the morning and you want tell a colleague, "I'll be back by time t", what should t be so that you will return by that time with probability .99?

Central Limit Theorem Let $X_1, X_2, \ldots, X_n$ be a random sample with $E(X_i) = \mu$ and $V(X_i) = \sigma^2$. If n is sufficiently large, $\bar{X}$ has approximately a normal distribution with mean $\mu_{\bar{X}} = \mu$ and variance $\sigma^2_{\bar{X}} = \sigma^2/n$.

You want to verify that 25-kg bags of fertilizer are being filled to the appropriate amount. You select a random sample of 50 bags of fertilizer and weigh them. Let X_i be the weight of the i^{th} bag for $i = 1, 2, \ldots 50$. You expect $E(X_i) = 25$ and $V(X_i) = .5$. Let $\bar{X} = (1/50)\sum_{i=1}^{50} X_i$.

Find $P(24.75 \leq \bar{X} \leq 25.25)$

Suppose $E(X_i) = 24.5$, that is, the bags are underfilled, and $V(X) = .5$. Now, find $P(24.75 \leq \bar{X} \leq 25.25)$.

In a statistics class of 36 students, past experience indicates that 53% of the students will score at or above 80%. For a randomly selected exam, find the probability at least 20 students will score above 80%.

Example: Normal approximation to the binomial. If $X \sim Bin(n, p)$ then X counts the number of successes in n independent Bernoulli trials, each with probability of success p. We know:

$$E(X) = np \;\; V(X) = np(1 - p)$$

So, by CLT, $\dfrac{X - np}{\sqrt{np(1 - p)}} \approx N(\mu = np, \sigma^2 = np(1 - p)$.

The CLT provides insight into why many random variables have probability distributions that are approximately normal.

For example, the measurement error in a scientific experiment. can be thought of as the sum of a number of underlying perturbations and errors of small magnitude.

A practical difficulty in applying the CLT is in knowing when n is sufficiently large. The problem is that the accuracy of the approximation for a particular n depends on the shape of the original underlying distribution being sampled.