Lecture 15

Central Limit Theorem

STAT 330 - Iowa State University

Central Limit Theorem (CLT)

Suppose $X_1, X_2, \ldots, X_n$ are iid random variables. For $i = 1, \ldots, n$,

$$X_i \overset{iid}{\sim} \text{distribution}$$

Any function of $\{X_i\}$ is also a random variable. Specifically,

- $S_n = \sum_{i=1}^{n} X_i$ is a R.V (with some distribution)
- $\overline{X_n} = \frac{\sum_{i=1}^{n} X_i}{n}$ is a R.V (with some distribution)

For large sample size n, the distribution of S_n and $\overline{X}$ both follow normal distributions!

Even without knowing the distribution of $\{X_i\}$, we can calculate probabilities for its sample mean and sample sum using the normal distribution. (extremely useful for real life problems)!

Central Limit Theorem (CLT)

- Sums and averages of RVs from *any* distribution have approximately normal distributions for large sample sizes

Central Limit Theorem (CLT)

Suppose $X_1, X_2, \ldots, X_n$ are iid random variables with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$ for $i = 1, \ldots, n$.
Define:

1. sample mean: $\overline{X_n} = \frac{\sum_{i=1}^n X_i}{n}$
2. sample sum: $S_n = \sum_{i=1}^n X_i$

Then, for *large n*,

$$\overline{X_n} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

$$S_n \sim N(n\mu, n\sigma^2)$$

How to Use CLT for Means

- For large n,

$$\overline{X_n} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

- How to calculate probabilities involving $\overline{X_n}$?
- Standardize $\overline{X_n}$ to turn it into a standard normal random variable Z, and use the $z-$table! (lecture notes 14)
- Standardize any normal random variable by subtracting its mean, and dividing by its standard deviation.

$$Z = \frac{\overline{X_n} - \mu}{\sigma/\sqrt{n}}$$

$$Z \sim N(0, 1)$$

How to Use CLT for Means Cont.

- Ex: $P(a < \overline{X_n} < b) =?$
- Standardize all of the quantities involved in the above probability. Then use Z-table to obtain probabilities.

$$\begin{aligned} P(a < \overline{X_n} < b) &= P\left(\frac{a-\mu}{\sigma/\sqrt{n}} < \frac{\overline{X_n}-\mu}{\sigma/\sqrt{n}} < \frac{b-\mu}{\sigma/\sqrt{n}}\right) \\ &= P\left(\frac{a-\mu}{\sigma/\sqrt{n}} < Z < \frac{b-\mu}{\sigma/\sqrt{n}}\right) \\ &= P\left(Z < \frac{b-\mu}{\sigma/\sqrt{n}}\right) - P\left(Z < \frac{a-\mu}{\sigma/\sqrt{n}}\right) \\ &= \Phi\left(\frac{b-\mu}{\sigma/\sqrt{n}}\right) - \Phi\left(\frac{a-\mu}{\sigma/\sqrt{n}}\right) \end{aligned}$$

How to Use CLT for Sums

- For large n,

$$S_n \sim N(n\mu, n\sigma^2)$$

- Standardize S_n by subtracting its mean, and dividing by its standard deviation.

$$Z = \frac{S_n - n\mu}{\sqrt{n\sigma^2}} = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

$$Z \sim N(0, 1)$$

- Then, use the $Z-$table to obtain desired probabilities.
- Ex:

$$\begin{aligned} P(S_n < a) &= P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} < \frac{a - n\mu}{\sigma\sqrt{n}}\right) \\ &= P\left(Z < \frac{a - n\mu}{\sigma\sqrt{n}}\right) \\ &= \Phi\left(\frac{a - n\mu}{\sigma\sqrt{n}}\right) \end{aligned}$$

Examples

Examples

Example 1: The time you spend waiting for the bus each day has a uniform distribution between 2 minutes and 5 minutes. Suppose you wait for the bus every day for a month (30 days).

1. Let $X_i =$ time spent waiting for the bus on the i^{th} day for $i = 1, \ldots, 30$.

   What is the distribution of each X_i?

   What is it's expected value and variance?

Examples

2. Let $\overline{X_n}$ be the average time spent waiting for the bus over the month. $\overline{X_n} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{\sum_{i=1}^{30} X_i}{30}$

   What is the (approximate) probability that the average time you spent waiting for the bus is less than 4 min?

Examples

Examples

3. How much time do you expect to spend waiting for the bus in total for a month?

4. What is the (approximate) probability that you spend more than 2 hours waiting for a bus in total for a month?

Examples

Examples

Example 2: Suppose an image has an expected size 1 megabyte with a standard deviation of 0.5 megabytes. A disk has 330 megabytes of free space. Is this disk likely to be sufficient for 300 independent images?

Examples

Example 3: An astronomer wants to measure the distance, d, from the observatory to a star. The astronomer plans to take n measurements of the distance and use the sample mean to estimate the true distance. From past records of these measurements the astronomer knows the standard deviation of a single measurement is 2 parsecs. How many measurements should the astronomer take so that the chance that his estimate differs by d by more than 0.5 parsecs is at most 0.05?

Examples

Examples