Computer Science

CSC196: Analyzing Data

Sampling Distributions

Jason Pacheco and Cesim Erten

Review of Previous Lecture

Let $X_1, \cdots, X_n$ iid with mean μ and variance σ^2.

Probability distribution of $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is sampling distribution of the mean.

$E(\overline{X}) =$

$Var(\overline{X}) =$

Outline

- ➢ Sampling Distributions (continued)
- ➢ Central Limit Theorem

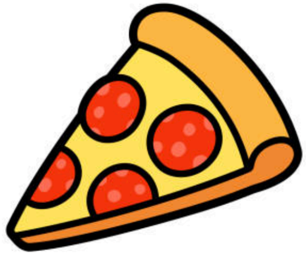
Examples Combining Sample Mean with Chebyshev's

Example: Want to take a random sample from a distribution with unknown μ and variance σ^2 is 8.

Size of sample such that $\overline{X}$ is within 2 units of μ with ≥ 0.99 probability?

Central Limit Theorem

Example: Time it takes to deliver a pizza: exponential distribution



Mean delivery time: 15 minutes

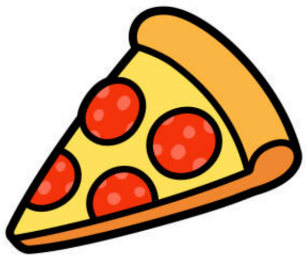
Consider the average delivery times of 30 drivers over a week:

- Individual delivery times might be erratic.
- BUT average of 30 deliveries has a pattern.

Central Limit Theorem

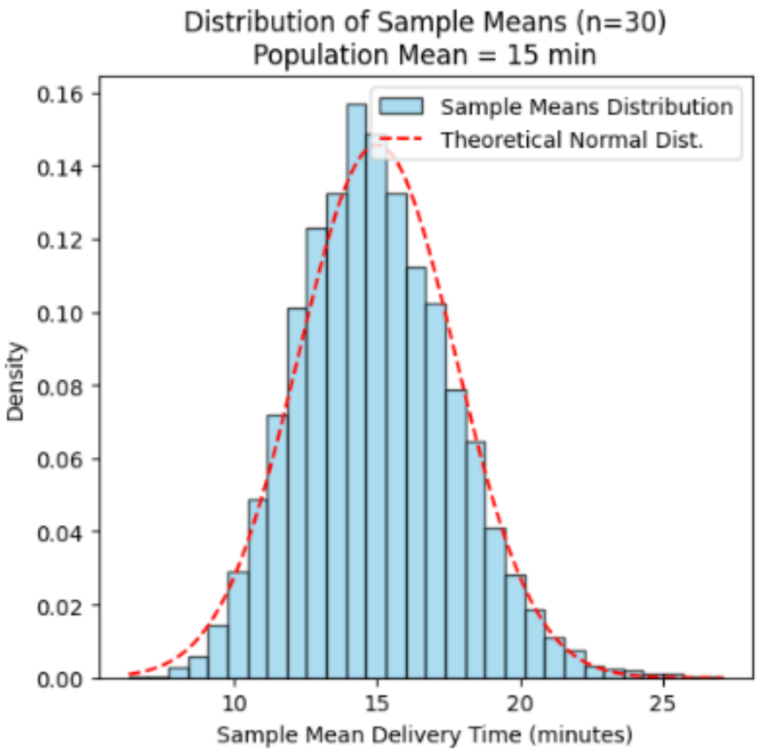
Example: Time it takes to deliver a pizza: exponential distribution

Mean delivery time: 15 minutes



Consider the average delivery times of 30 drivers over a week:

- Individual delivery times might be erratic.
- BUT average of 30 deliveries has a pattern.



Check: https://colab.research.google.com/drive/1DOHaH6GgxSQkgLQLkR1xpNBcw98IEcp4?usp=sharing

Central Limit Theorem

Informally:

For a large random sample, sample mean has an approximately normal distribution with mean μ, variance σ^2/n.

Note: Nothing is assumed regarding the distribution of X_i.

Central Limit Theorem

Informally:

For a large random sample, sample mean has an approximately normal distribution with mean μ, variance σ^2/n.

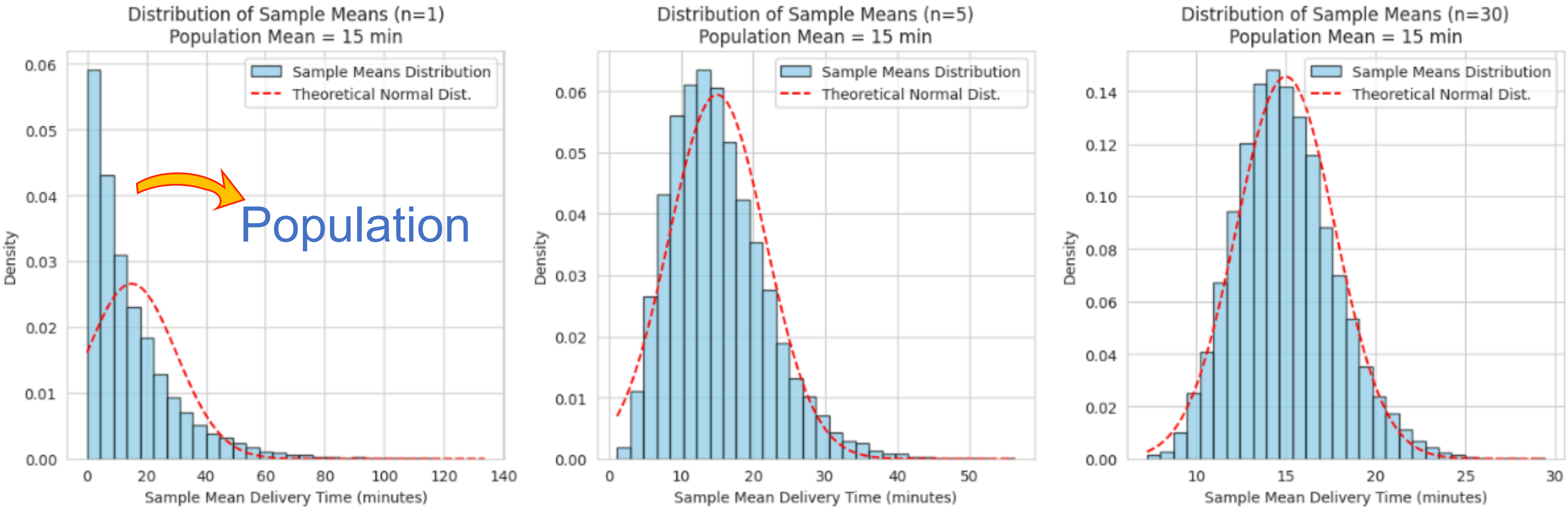
Note: Nothing is assumed regarding the distribution of X_i.

Formally: Let $X_1, \cdots, X_n$ be iid with mean μ, standard deviation σ.

$$\text{As } n \to \infty, \quad Z = \frac{\bar{X} - \mu}{\sqrt{Var(\bar{X})}} = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \sim N(0,1)$$

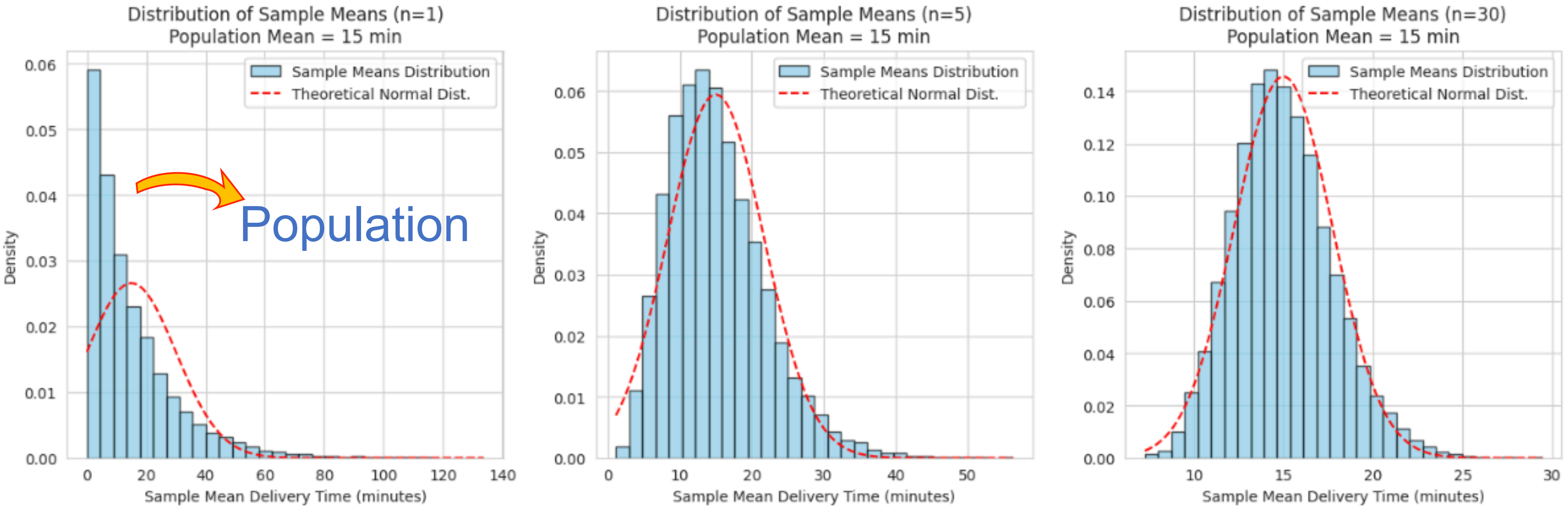
Central Limit Theorem

What sample size n? In general $n \geq 30$ should be good.



Central Limit Theorem

What sample size n? In general $n \geq 30$ should be good.



To observe the power of CLT, as an exercise: plugin a value for p in our polling example and apply the CLT.

Examples on Central Limit Theorem

Example 8.4: An electrical firm manufactures light bulbs that have a length of life that is approximately normally distributed, with mean equal to 800 hours and a standard deviation of 40 hours. Find the probability that a random sample of 16 bulbs will have an average life of less than 775 hours. Use norm.cdf(-2.5)=0.0062.

Examples on Central Limit Theorem

Example 8.5: Traveling between two campuses of a university in a city via shuttle bus takes, on average, 28 minutes with a standard deviation of 4 minutes. In a given week, a bus transported passengers 36 times. What is the probability that the average transport time was more than 30 minutes? Assume the mean time is measured to the nearest minute. Use 1-norm.cdf(3.75)=0.000233.

Examples on Central Limit Theorem

Example: Number of errors per program has Poisson distribution with mean 4. Let $X_1, \cdots, X_{100}$ be the numbers of errors in 100 programs. How to approximate $P(\overline{X} < 4.2)$? Use norm.cdf(1)=0.84.