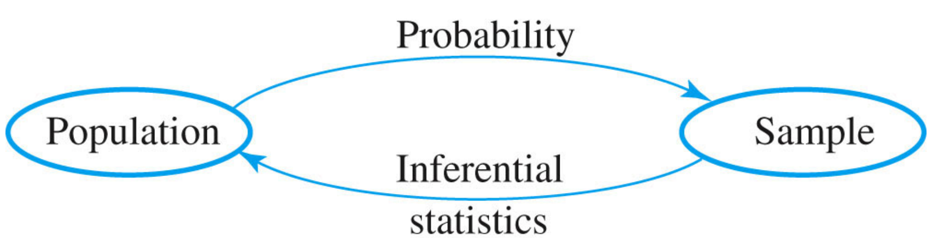
Sample Distribution

Today's Class

- Sample Distribution
- Central Limit Theorem

Statistics and Their Distributions

- A **statistic** is any quantity whose value can be calculated from sample data.
- Prior to obtaining data, there is uncertainty as to what value of any particular statistic will result.
- Therefore, a statistic is a random variable and will be denoted by an uppercase letter; a lowercase letter is used to represent the calculated or observed value of the statistic



Random Samples

- The probability distribution of any particular statistic depends not only on the population distribution and the sample size *n* but also on the method of sampling.
- Frequently, we make the simplifying assumption that our data constitute a random sample $X_1, X_2, \ldots, X_n$ from a distribution. This means that
  - The X is are independent
  - All the X is have the same probability distribution

Deriving a Sampling Distribution

- Given such a sample, we can evaluate a statistic of interest
- The values of the statistic calculated from these samples allow us to examine the distribution of the statistic
- By changing settings, we can examine how the distribution of the statistic changes

Example

- A large automobile service center charges \$40, \$45, and \$50 for a tune-up of four-, six, and eight-cylinder cars, respectively. If 20% of its tune-ups are done on four-cylinder cars, 30% on six-cylinder cars, and 50% on eight-cylinder cars, then the probability distribution of revenue from a single randomly selected tune-up is given by

x	40	45	50
P(x)	0.2	0.3	0.5

- Suppose on a particular day, only two servicing jobs involve tune-ups. Let X_1 = the revenue from the first tune-up and X_2 = the revenue from the second

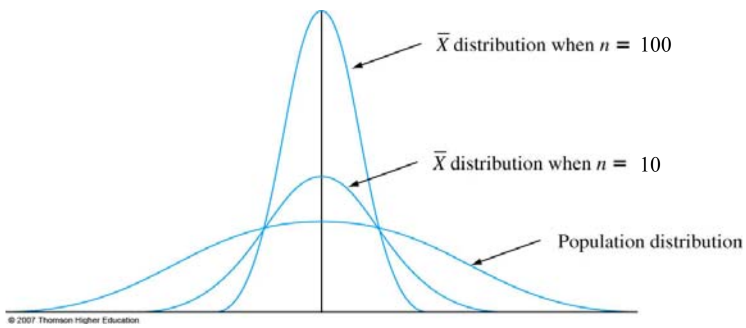
Distribution of the Sample Mean

- Let $X_1, X_2, .. X_n$ be a random sample from a distribution with mean value μ and standard deviation σ. Then the expected value of sample mean is

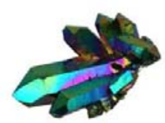
$$E(\bar{X}) = \mu_{\bar{X}} = \mu_X$$

and the variance of sample mean is

$$V(\bar{X}) = \sigma^2_{\bar{X}} = \frac{\sigma^2}{n}$$



Example 5.24



- In a notched tensile fatigue test on a titanium specimen, the expected number of cycles to first acoustic emission (used to indicate crack initiation) is μ = 28,000, and the standard deviation of the number of cycles is σ = 5000. Let $X_1, X_2, \ldots, X_{25}$ be a random sample of size 25, where each X_i is the number of cycles on a different randomly selected specimen.
  - What are the expected value and STD of the sample mean number of cycles until first emission?
  - What are the expected value and STD when n=100?

Example: Sample Mean

- The height of men can be represented by a normal distribution with mean 72 inches and standard deviation 2 inches. Suppose you measure the height of 10 randomly chosen men.
  - What are the expected value of the sample mean?

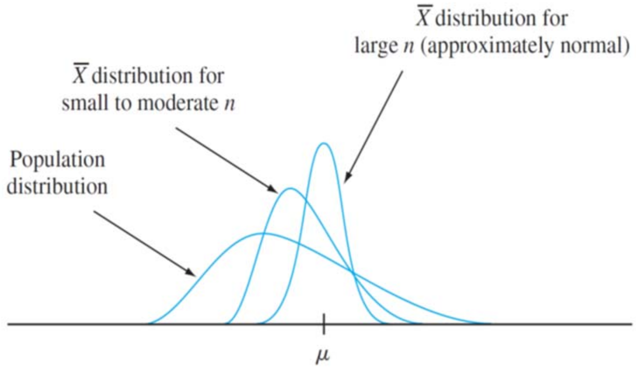
Central Limit Theorem (CLT)

- Let $X_1, X_2, .. X_n$ be a random sample from a distribution with mean value μ and variance σ^2
- Then if n is sufficiently large,

$$\overline{X} \sim N(\mu, \sigma^2/n)$$

- The larger the value of n, the better the approximation (n>30)

CLT



CLT Example

- Let X denote the number of flaws in copper wire. The pmf is as follows:

X	0	1	2	3
P(X=x)	0.48	0.39	0.12	0.01

  - If one hundred wires are sampled from this population, what is the probability that the average number of flaws per wire in this sample is less than 0.5?