Estimation and Sampling Distributions

Loukia Meligkotsidou
Associate Professor of Statistics
Department of Mathematics
National and Kapodistrian University of Athens

Populations and Samples

- A **Population** is the set of all items or individuals of interest

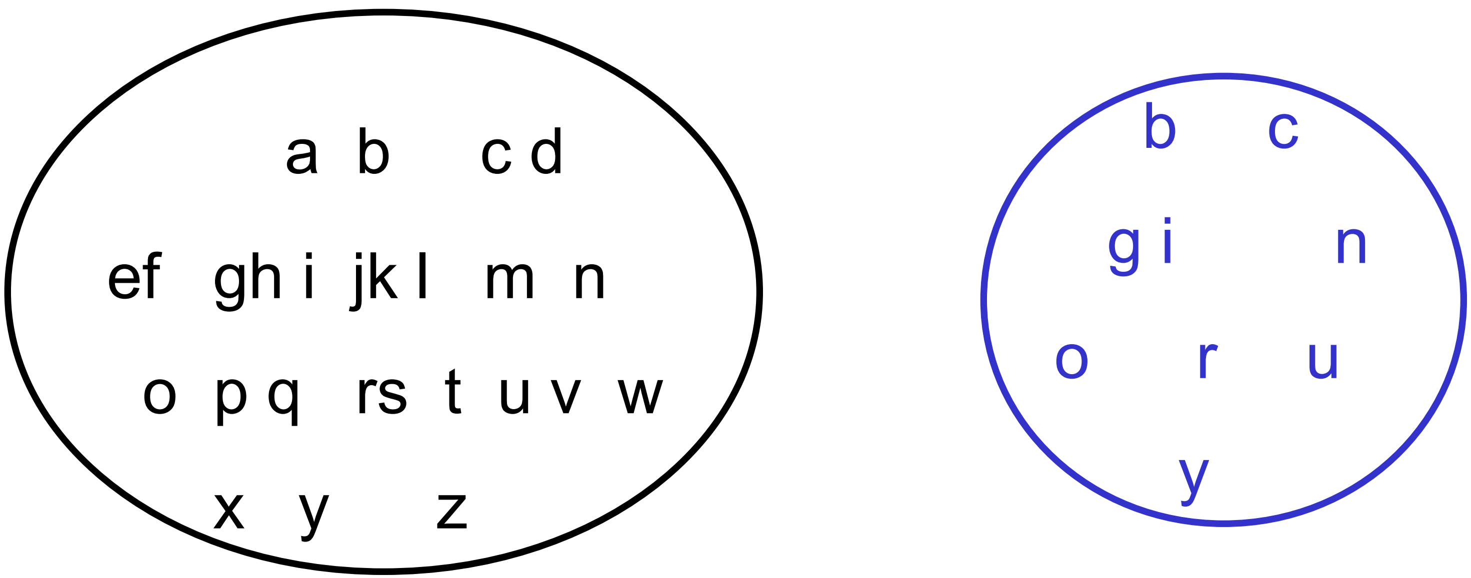
Examples:	All likely voters in the next election
	All parts produced today
	All sales receipts for November

- A **Sample** is a subset of the population

Examples:	1000 voters selected at random for interview
	A few parts selected for destructive testing
	Random receipts selected for audit

Population vs. Sample

Population **Sample**



Why Sample?

- Less time consuming than a census
- Less costly to administer than a census
- It is possible to obtain statistical results of a sufficiently high precision based on samples.

Simple Random Samples

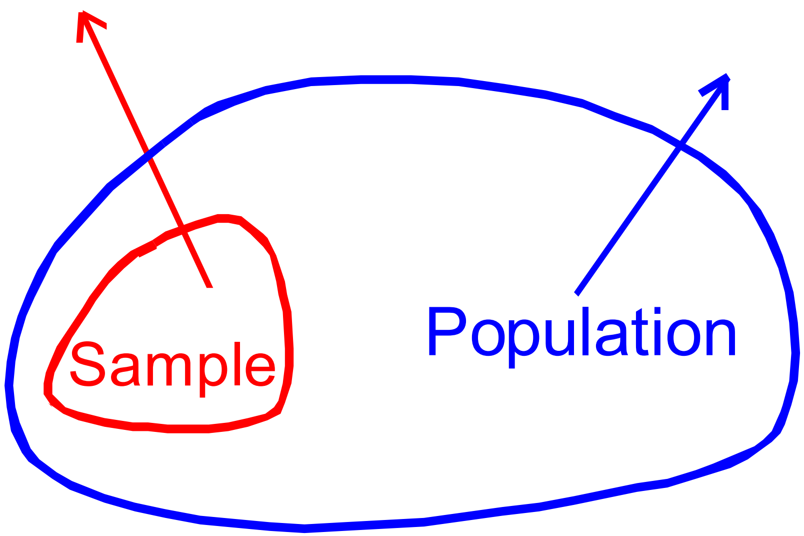
- Every object in the population has an equal chance of being selected
- Objects are selected independently
- Samples can be obtained from a table of random numbers or computer random number generators
- A simple random sample is the benchmark against which other sample methods are compared

Inferential Statistics

- Making statements about a population by examining sample results

Sample statistics → Population parameters

(known) **Inference** (unknown, but can be estimated from sample evidence)



Inferential Statistics

Drawing conclusions and/or making decisions concerning a population based on sample results.

- **Estimation**
  - e.g., Estimate the population mean weight using the sample mean weight
- **Hypothesis Testing**
  - e.g., Use sample evidence to test the claim that the population mean weight is 120 pounds
- **Confidence Intervals**
  - e.g., Construct a 95% confidence interval for the mean

How is inference made?

- Point estimation: Find the “best” approximations of an unknown population parameter
- Interval estimation: Find a range of values that with high probability covers the unknown population parameter
- Hypothesis testing: Give statements about the population (values of parameters, probability distributions, issues of independence,…) and examine their validity

Estimation methods

- Maximum likelihood
- Least squares
- Method of moments

Method of Moments

- Let X1, X2, . . . Xn represent a random sample from a population with mean μ and variance σ^2
- Use as an estimate for μ the sample mean value defined as

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

- Use as an estimate for σ^2 the sample variance

Maximum likelihood

- The population to be investigated is such that the values that comes out in a sample x_1, x_2 , ... are governed by a probability distribution
- The probability distribution is represented by a probability density (or mass) function $f(x)$
- Alternatively, the sample values can be seen as the outcomes of independent random variables X_1, X_2, … all with probability density (or mass) function $f(x)$

Maximum likelihood

- We have a sample $\boldsymbol{x} = (x_1 , \ldots , x_n)$ from a population
- The population contains an unknown parameter θ
- The functional forms of the distributional functions is known, but depend on the unknown θ.
- Denote generally by $f(x;\theta)$ the probability density or mass function of the distribution
- A *point estimate* of θ is a function of the sample values

Maximum Likelihood

- Principle: Estimate a parameter such that for this value the probability of obtaining an actually observed sample is as large as possible.
- The probability (or density) of an observed sample depends on a parameter which will be adjusted to give it a maximum possible value.
- The maximum likelihood estimate of the unknown population parameter is the value of the parameter that maximizes the probability (or density) of the observed sample.

Maximum likelihood

- The joint pdf or pmf of a sample of n random variables is given by

$$L(\theta \mid x) = f(x_1, x_2, \ldots, x_n \mid \theta) = \prod_{i=1}^{n} f_X(x_i \mid \theta)$$

if the random variables are independent

- The above function $L(\theta \mid x)$ is known as the likelihood of the observed sample

Maximum likelihood

- We call the value of θ that maximizes the likelihood function $L(\theta|x)$, the maximum likelihood estimator (MLE) of θ:

$$\hat{\theta} = \arg\max_{\theta}[L(\theta \mid x)]$$

- For convenience, we compute the logarithm of the likelihood function: $\ln[L(\theta|x)]$

- The maximum can be found by taking:

$$\frac{\partial \ln L(\theta \mid x)}{\partial \theta} = 0$$

- Solve the system of partial derivatives

Example 1: Bernoulli

- Let $x_1, x_2, \ldots, x_n$ be a random sample from a Bernoulli distribution with parameter p, i.e.

$$f(x_i; p) = p^{x_i}(1-p)^{1-x_i}, \quad x_i = 0,\ 1$$

Find the MLE of p

- Compute likelihood

$$L(\theta \mid x) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i}$$

- Compute log-likelihood function

$$\ln L(\theta \mid x) = \sum_{i=1}^{n} x_i \ln(p) + \left(n - \sum_{i=1}^{n} x_i \right) \ln(1-p)$$

- Compute the derivative

$$\frac{\partial \ln L(\theta \mid x)}{\partial p} = \frac{1}{p} \sum_{i=1}^{n} x_i - \frac{1}{1-p} \left(n - \sum_{i=1}^{n} x_i \right)$$

- Solve the equation (=0) and find the estimator:

$$\hat{p}_{MLE} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Example 2: Normal

- Let $x_1, x_2, \ldots, x_n$ be a random sample from a normal distribution with mean μ and variance σ^2 Find the MLE of μ and σ^2

- Compute likelihood

$$L(\theta \mid x) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x_i - \mu)^2}{2\sigma^2} \right]$$

$$= \frac{1}{\left(2\pi\sigma^2\right)^{n/2}} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^{n} \left(x_i - \mu\right)^2 \right]$$

- Compute log-likelihood function

$$\ln L(\theta, x) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2$$

- Compute partial derivatives

$$\frac{\partial \ln L(\theta \mid x)}{\partial \mu} = -\frac{1}{2\sigma^2}\sum_{i=1}^{n} 2(x_i - \mu)(-1) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu) = 0$$

$$\frac{\partial \ln L(\theta \mid x)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} - \frac{1}{2}\sum_{i=1}^{n}(x_i - \mu)^2(-(\sigma^2)^{-2}) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i - \mu)^2 = 0$$

- Solve the system and find the estimators:

$$\hat{\mu}_{MLE} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \hat{\sigma}^2_{MLE} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu}_{MLE})^2$$

Properties of Estimators: Unbiasedness

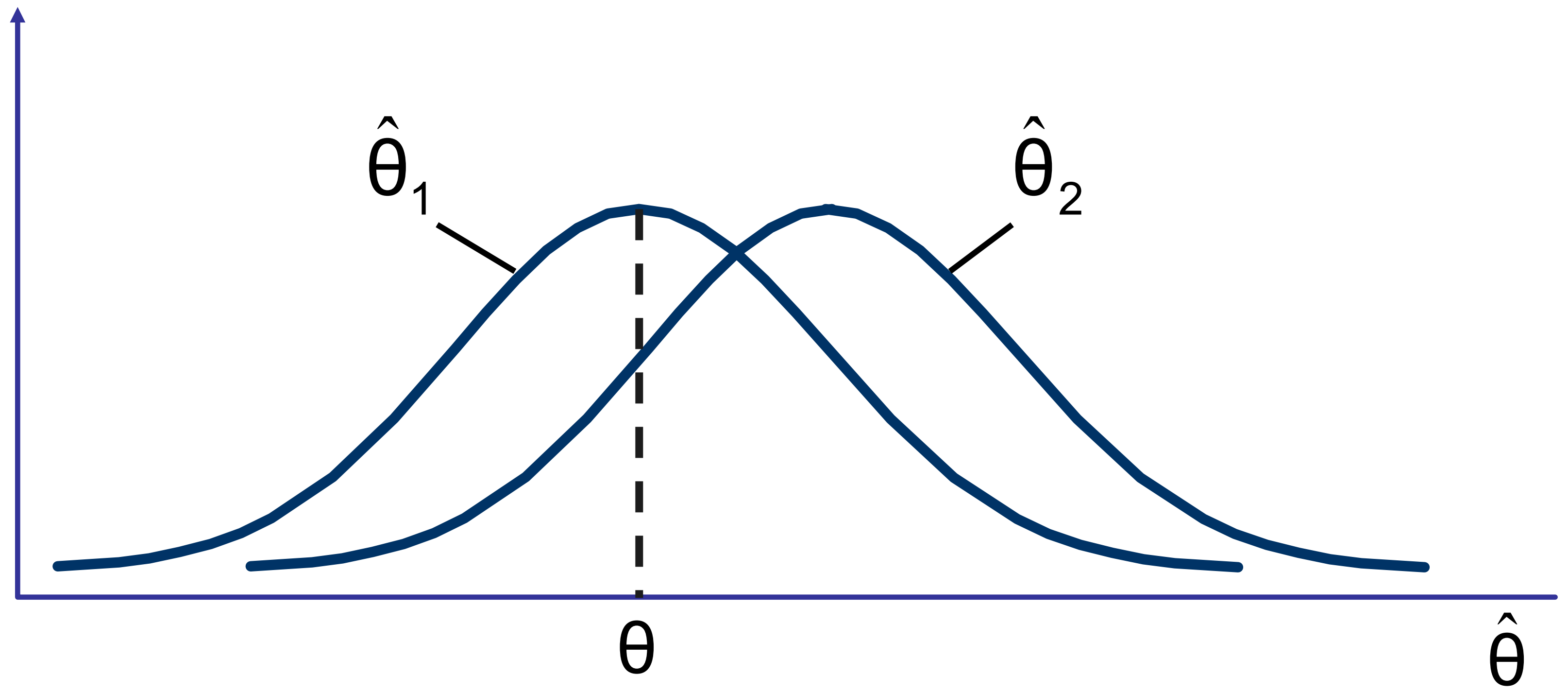
- A point estimator $\hat{\theta}$ is said to be an unbiased estimator of the parameter θ if the expected value, or mean, of the sampling distribution of $\hat{\theta}$ is θ,

$$E(\hat{\theta}) = \theta$$

- Examples:
  - The sample mean is an unbiased estimator of μ
  - The sample proportion is an unbiased estimator of P

Properties of Estimators: Unbiasedness

- $\hat{\theta}_1$ is an unbiased estimator, $\hat{\theta}_2$ is biased:



Properties of Estimators: Bias

- Let $\hat{\theta}$ be an estimator of θ
- The bias in $\hat{\theta}$ is defined as the difference between its mean and θ

$$\text{Bias}(\hat{\theta}) = E(\hat{\theta}) - \theta$$

- The bias of an unbiased estimator is 0

Properties of Estimators: Efficiency

- Suppose there are several unbiased estimators of θ
- The most efficient estimator or the minimum variance unbiased estimator of θ is the unbiased estimator with the smallest variance
- Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two unbiased estimators of θ, based on the same number of sample observations. Then,
  - $\hat{\theta}_1$ is said to be more efficient than $\hat{\theta}_2$ if $\mathrm{Var}(\hat{\theta}_1) < \mathrm{Var}(\hat{\theta}_2)$

Example 1:Unbiasedness of Sample Mean

- Let $X_1, X_2, \dots X_n$ represent a random sample from a population with mean μ and variance σ^2
- The sample mean value of these observations is defined as

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

- The Expected value of the sample mean is:

$$E(\overline{X}) = E(\frac{1}{n}\sum_{i=1}^{n} X_i) = \mu$$

- The sample mean is an unbiased estimator of μ

Example 1: Variance of the sample mean

- The variance of the sample mean is:

$$V(\overline{x}) = V\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{\sigma^2}{n}$$

- A measure of the variability in the mean from sample to sample is given by the Standard Error of the Mean:

$$\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}}$$

- Note that the standard error of the mean decreases as the sample size increases

Example 2: Unbiasedness of proportion - Variance of proportion

Let X_1, X_2, . . . X_n be a random sample from a Bernoulli distribution with parameter p, i.e. A Binomial experiment with n trials and probability of success p.

The maximum likelihood estimate of proportion is unbiased:

$$E(\hat{P}) = p$$

The variance of the proportion is given by:

$$V(\hat{p}) = V(\frac{x}{n}) = \frac{1}{n^2} V(x) = \frac{1}{n^2} np(1-p) = \frac{p(1-p)}{n}$$

The standard error of the estimate of proportion is given by:

$$\sigma_{\hat{p}} = \sqrt{V(\hat{p})} = \sqrt{\frac{p(1-p)}{n}}$$

Law of Large Numbers

- Informally: An average of many measurements is more accurate than a single measurement.
- Formally: Let $X_1, X_2, \ldots$ be i.i.d. random variables all with mean μ and standard deviation σ.

  Let

$$\bar{x}_n = \frac{X_1 + X_2 + \ldots + X_n}{n} = \frac{\sum_{i=1}^{n} X_i}{n}$$

  Then, for any (small number) a, we have that

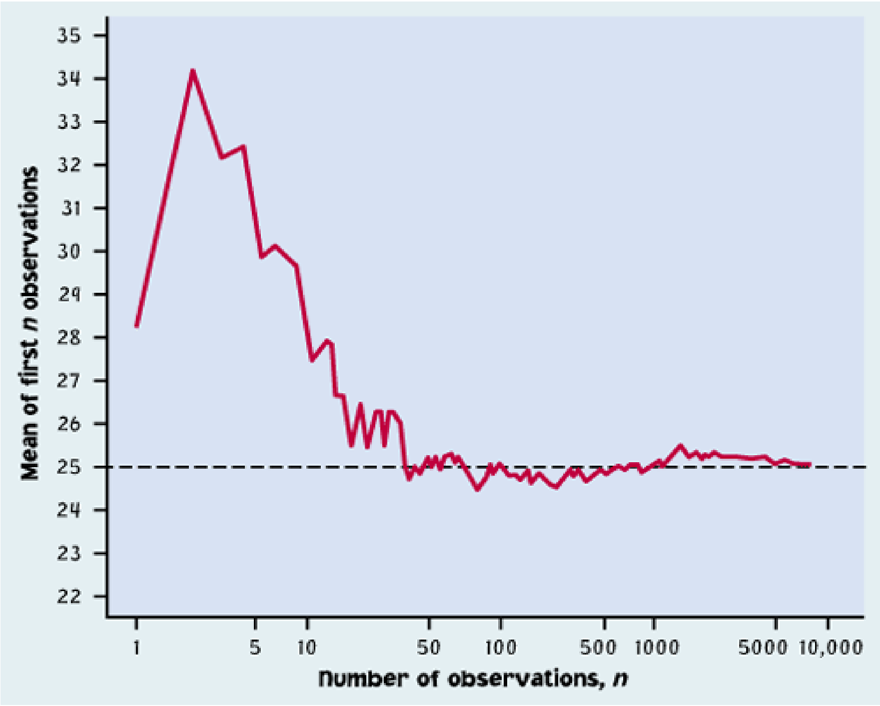
$$\lim_{n \to \infty} P\left(\left| \bar{x}_n - \mu \right| < a \right) = 1$$

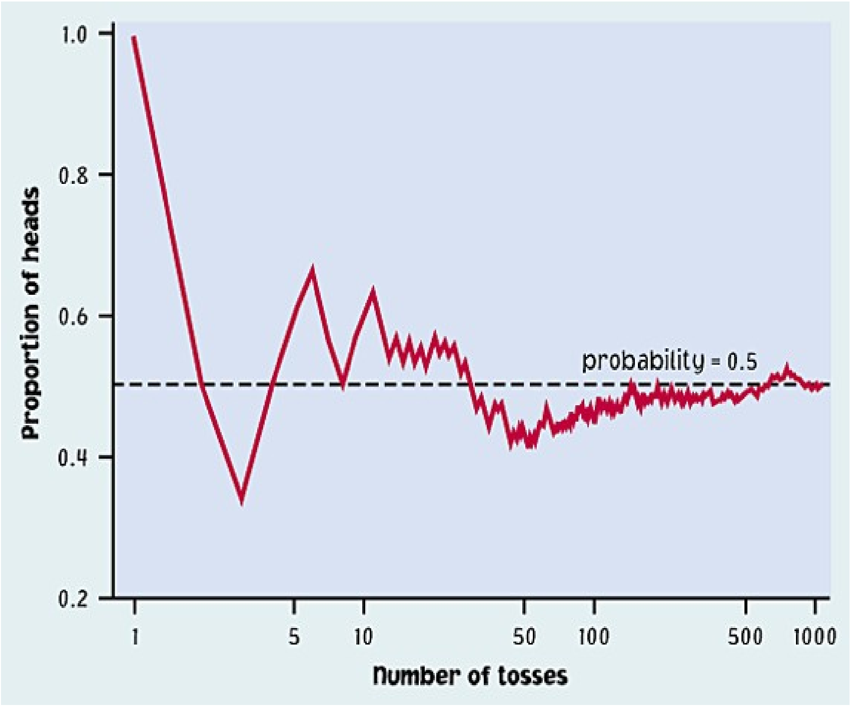
The law of large numbers

Law of large numbers: As the number of randomly-drawn observations (n) in a sample increases

the mean of the sample ($\bar{x}$) gets closer and closer to the population mean μ (quantitative variable).

the sample proportion ($\hat{p}$) gets closer and closer to the population proportion p (categorical variable).





Sampling distribution of the mean case1 : the Population is Normal

- If a population is normal with mean μ and standard deviation σ, the sampling distribution of $\bar{X}$ is also normally distributed with

$$\mu_{\bar{X}} = \mu$$

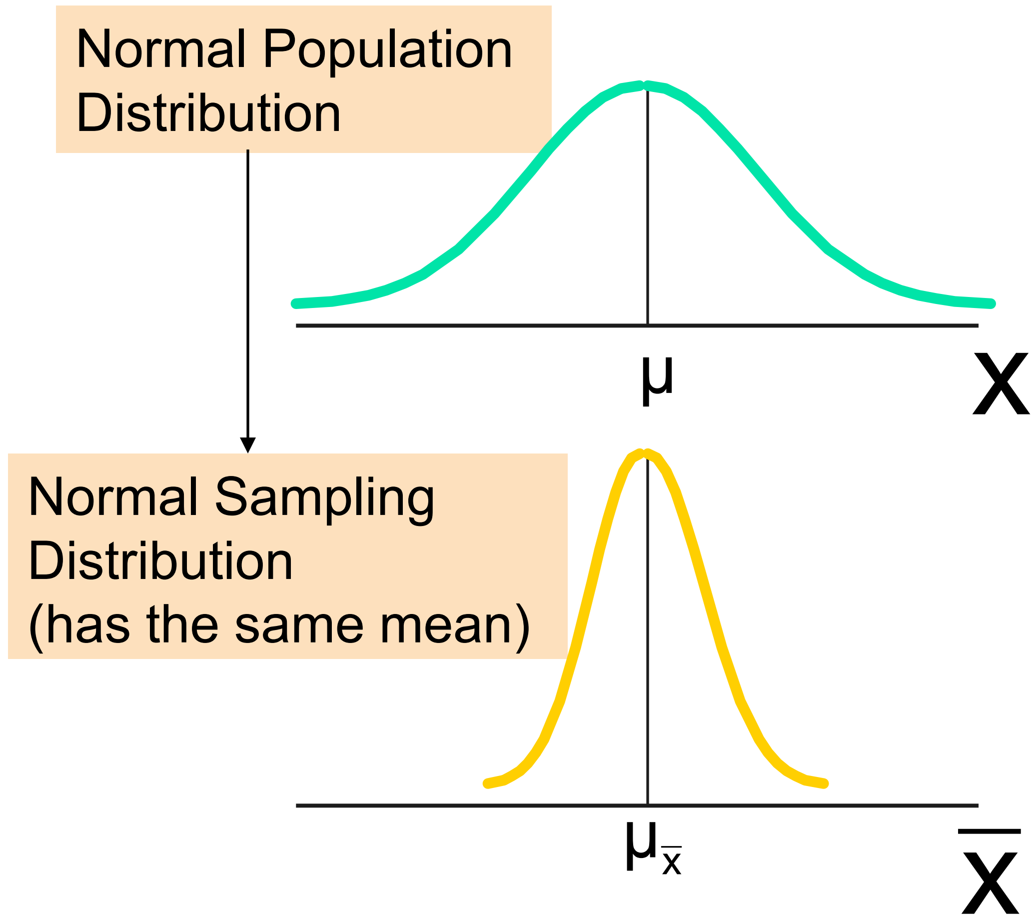
and

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

Sampling distribution of the mean
case1 : the Population is Normal

- $$E(\overline{x}) = \mu_{\overline{x}} = \mu$$

(i.e. $\overline{X}$ is unbiased)

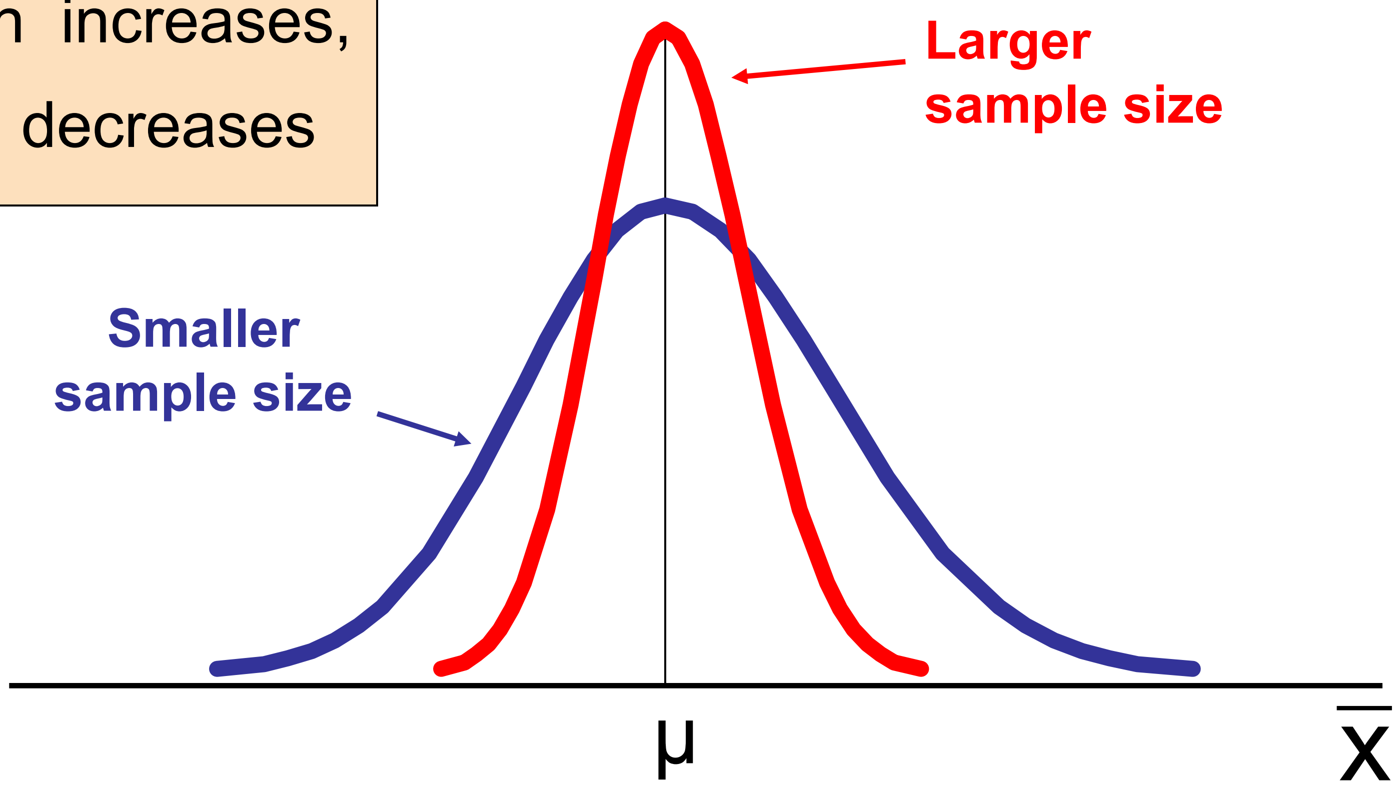


Sampling distribution of the mean case1 : the Population is Normal

(continued)

- For sampling with replacement:

As n increases, $\sigma_{\bar{x}}$ decreases



Sampling distribution of the mean case2 : the Population is not Normal

- We can apply the Central Limit Theorem:
  - Even if the population is not normal,
  - …sample means from the population will be approximately normal as long as the sample size is large enough.

Properties of the sampling distribution:

$$\mu_{\bar{x}} = \mu$$

and

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

The Central Limit Theorem

Formally: Let $X_1, X_2, \ldots$ be i.i.d. random variables all with mean μ and standard deviation σ.

Let

$$\bar{x}_n = \frac{X_1 + X_2 + \ldots + X_n}{n} = \frac{\sum_{i=1}^{n} X_i}{n}$$

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

For large n,

$$\bar{x}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

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

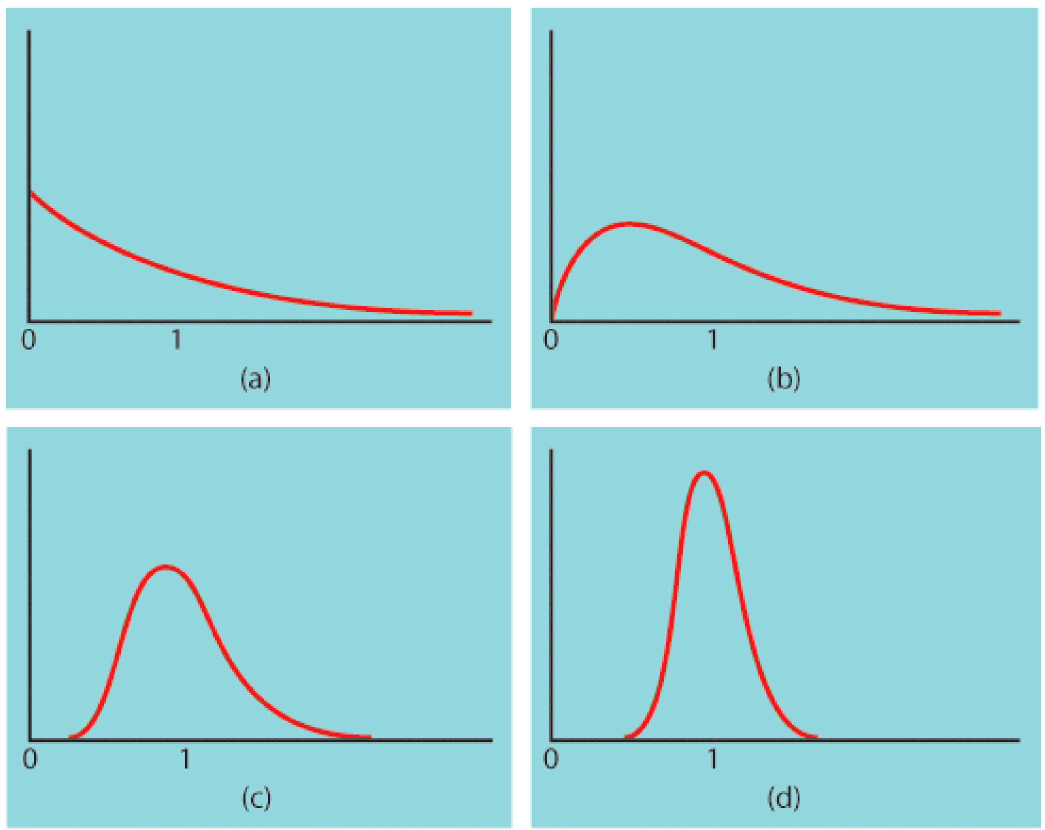
Standardized S_n or $\bar{x}_n \sim N(0,1)$

The central limit theorem

Central Limit Theorem: When randomly sampling from any population with mean μ and standard deviation σ, **when *n* is large enough,** the sampling distribution of $\bar{x}$ is approximately normal: $N(\mu, \sigma/\sqrt{n})$.

Population with strongly skewed distribution

Sampling distribution of $\bar{x}$ for $n = 2$ observations



Sampling distribution of $\bar{x}$ for $n = 10$ observations

Sampling distribution of $\bar{x}$ for $n = 25$ observations

If the Population is **not** Normal

(continued)

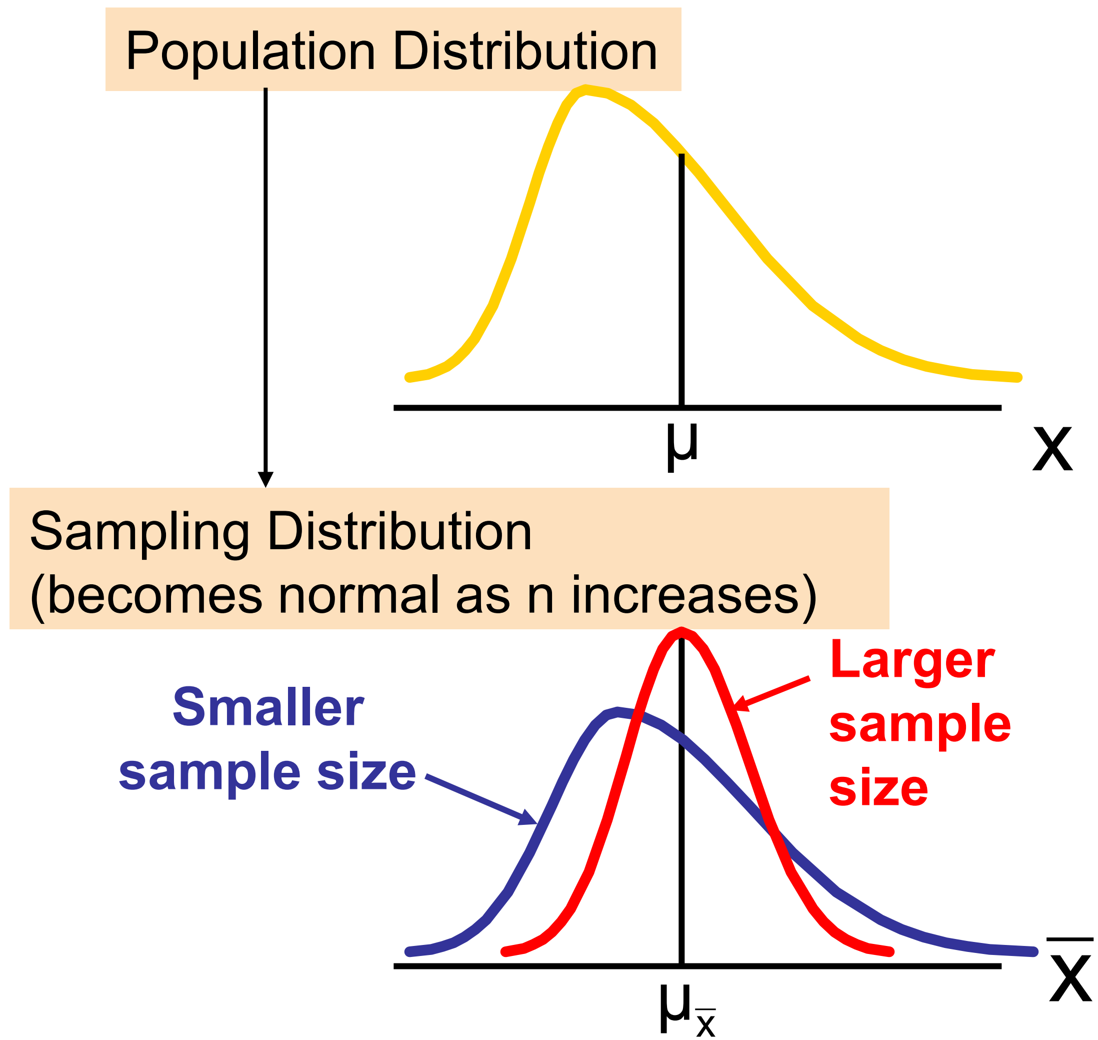
Sampling distribution properties:

Central Tendency

$$\mu_{\bar{x}} = \mu$$

Variation

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$



How Large is Large Enough?

- For most distributions, $n > 25$ will give a sampling distribution that is nearly normal
- For normal population distributions, the sampling distribution of the mean is always normally distributed

Standardization

- If the random variable X has a $N(\mu,\sigma)$ distribution, then the random variable $Z=(X-\mu)/\sigma$ has the standard normal $N(0,1)$ distribution.
- Standardization of the sample mean helps in calculating probabilities related to $\bar{X}$

For example, $P(\bar{X} < a)=P(Z<(a-\mu)/\sigma)$

Example

- The weight of cereal in a box is normally distributed with mean 368gr and standard deviation 15gr. We select a random sample of 25 boxes. What is the probability that the sample mean is below 365gr?
- Xbar~N(368,15/5)

- P(Xbar<365)=P((Xbar-368)/3<(365-368)/3)

  =P(Z<-3/3)=P(Z<1)