CENG 222
Statistical Methods for Computer Engineering

Week 9

Chapter 9

Statistical Inference I

Outline

- Parameter estimation
  - Method of moments
  - Method of maximum likelihood
- Confidence intervals
- Unknown standard deviation
- Hypothesis testing

Recall from Chapter 8: Estimation of population mean

- $\bar{X} = \frac{X_1 + \cdots + X_n}{n}$
- Sample mean is unbiased, consistent, and asymptotically Normal.
  - $\mathbf{E}(\hat{\theta}) = \theta$
  - $\mathrm{Bias}(\hat{\theta}) = \mathbf{E}(\hat{\theta} - \theta)$

Recall from Chapter 8: Estimation of population variance

- $s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$
- 1/*n*-1 needed for an unbiased estimator
- This estimator is also consistent and asymptotically Normal

Estimation of distribution parameters

- Example:
  - Consider a Poisson variable. How should we estimate the parameter λ?
    - Sample mean?
    - Sample variance?
    - Both of them are equal to λ.
- Two generic methods of estimation will be discussed
  - Method of moments
  - Method of maximum likelihood

Moments

- The k-th population moment is defined as:
  - $\mu_k = \mathbf{E}(X^k)$
- The k-th sample moment is computed as:
  - $m_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$

  and it estimates μ_k from a sample $(X_1, \dots, X_n)$

Central Moments

- For $k \geq 2$,The k-th population central moment is defined as:
  - $\mu'_k = \mathbf{E}(X - \mu_1)^k$
- The k-th sample moment is computed as:
  - $m'_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^k$

  and it estimates μ'_k from a sample $(X_1, \ldots, X_n)$

Method of moments

- To estimate *k* parameters of a distribution, equate the first *k* population and sample moments and solve a system of *k* equations and *k* unknowns.
- $$\begin{cases} \mu_1 & = & m_1 \\ \ldots & \ldots & \ldots \\ \mu_k & = & m_k \end{cases}$$

Example 9.5 Pareto Distribution

- cdf of Pareto distribution
  - $F(x) = 1 - \left(\frac{x}{\sigma}\right)^{-\theta}$ for $x > \sigma$
  - Two parameters
- Solution:
  - Find the equations of the first and second population moments, μ_1 and μ_2
  - Solve for θ and σ in terms of m_1 and m_2.

Example 9.5 Pareto Distribution

- In order to find the moments using expectation, we need the pdf:
  - $f(x) = F'(x) = \frac{\theta}{\sigma}\left(\frac{x}{\sigma}\right)^{-\theta-1} = \theta\sigma^{\theta}x^{-\theta-1}$
- $\mu_1 = \mathbf{E}(X) = \int_{\sigma}^{\infty} xf(x)dx = \frac{\theta\sigma}{\theta-1}$ for $\theta > 1$
- $\mu_2 = \mathbf{E}(X^2) = \int_{\sigma}^{\infty} x^2 f(x)dx = \frac{\theta\sigma^2}{\theta-2}$ for $\theta > 2$

Example 9.5 Pareto Distribution

- $\begin{cases} \mu_1 = \dfrac{\theta\sigma}{\theta-1} = m_1 \\ \mu_2 = \dfrac{\theta\sigma^2}{\theta-2} = m_2 \end{cases}$
- $\hat{\theta}_{mom} = \sqrt{\dfrac{m_2}{m_2-m_1^2}} + 1$
- $\hat{\sigma}_{mom} = \dfrac{m_1(\hat{\theta}-1)}{\hat{\theta}}$

Method of maximum likelihood

- Maximum likelihood estimator is the parameter value that maximizes the likelihood of the observed sample.
- For a discrete distribution, maximize the joint pmf of the data $f(X_1, \dots, X_n)$
- For a continuous distribution, maximize the joint pdf of the data $f(X_1, \dots, X_n)$

Discrete distributions

- Since we use simple random sampling, each observed X_i is independent of the others. Therefore, the joint pmf is equal to:
  - $\prod_{i=1}^{n} f(X_i) = \prod_{i=1}^{n} P(X = X_i)$
- In order to maximize this, with respect to a parameter. We take the derivative of this wrt that parameter and equate to 0.
- Taking logarithms of the joint pmf is helpful (the maximizing value will be the same)
  - $\ln \prod_{i=1}^{n} f(X_i) = \sum_{i=1}^{n} \ln f(X_i)$

Example 9.7 Poisson distribution

- pmf of Poisson is:
  - $f(x) = e^{-\lambda}\frac{\lambda^x}{x!}$
- $\ln f(x) = -\lambda + x \ln \lambda - \ln(x!)$
- The joint pmf is:
- $\sum_{i=1}^{n} -\lambda + X_i \ln \lambda - \ln(X_i!) =$
- $= -n\lambda + \ln \lambda \sum_{i=1}^{n} X_i + C$

Example 9.7 Poisson distribution

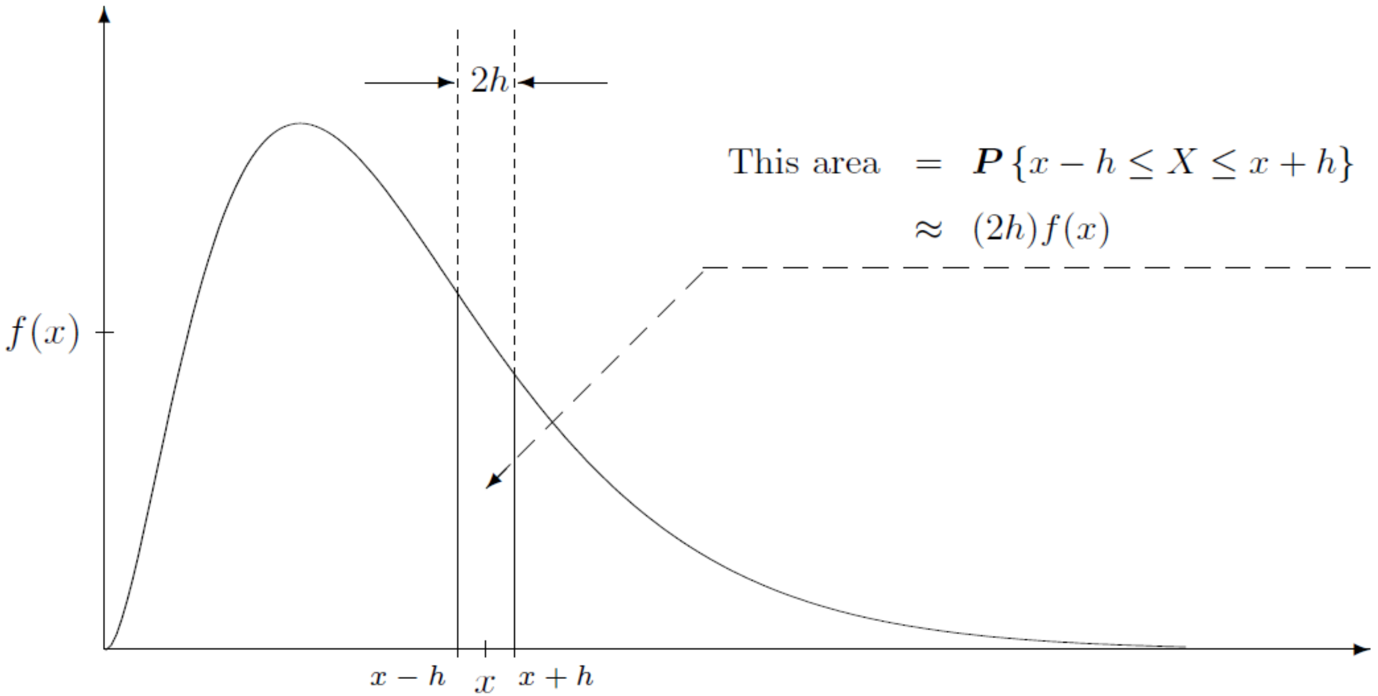
- Differentiate wrt λ and equate to 0
  - $-n + \frac{1}{\lambda}\sum_{i=1}^{n} X_i = 0$
- Only one solution:
  - $\hat{\lambda}_{mle} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$
- Method of moments and the method of maximum likelihood have the same estimator for λ.

Continuous distributions

- $P(X_i = x)$ is 0 for continuous distributions, so the joint pdf will be 0. We will use

$$P(x - h < X_i < x + h)$$

instead



Continuous distributions

- Probability of almost observing a point is proportional to the pdf at that point; therefore, as in the discrete case, we will maximize the product of individual pdfs.
  - $\prod_{i=1}^{n} f(X_i)$

Example 9.8 Exponential

- pdf of Exponential distribution is:
  - $f(x) = \lambda e^{-\lambda x}$
- $\ln f(x) = \ln \lambda - \lambda x$
- The joint pdf is:
- $\sum_{i=1}^{n} \ln \lambda - \lambda X_i = n \ln \lambda - \lambda \sum_{i=1}^{n} X_i$

Example 9.8 Exponential

- Differentiate wrt λ and equate to 0

$$\frac{n}{\lambda} - \sum_{i=1}^{n} X_i = 0$$

- Only one solution:
  - $\hat{\lambda}_{mle} = \frac{n}{\sum_{i=1}^{n} X_i} = \frac{1}{\bar{X}}$

Estimation of standard errors

- What is the standard error of $\hat{\lambda}_{mle} = \frac{1}{\bar{X}}$ we found on Example 9.8?
  - I.e., $\sigma(\hat{\lambda}_{mle})$=?
- The k-th moment of $\hat{\lambda}_{mle}$ can be computed by using the fact that $\hat{\lambda}_{mle} = 1/\bar{X}$ and that $\sum_{i=1}^{n} X_i$ is a Gamma rv.
- First moment: $\mathbf{E}(\hat{\lambda}_{mle}) = \frac{n\lambda}{n-1}$
- Second moment: $\mathbf{E}(\hat{\lambda}_{mle}^2) = \frac{n^2\lambda^2}{(n-1)(n-2)}$

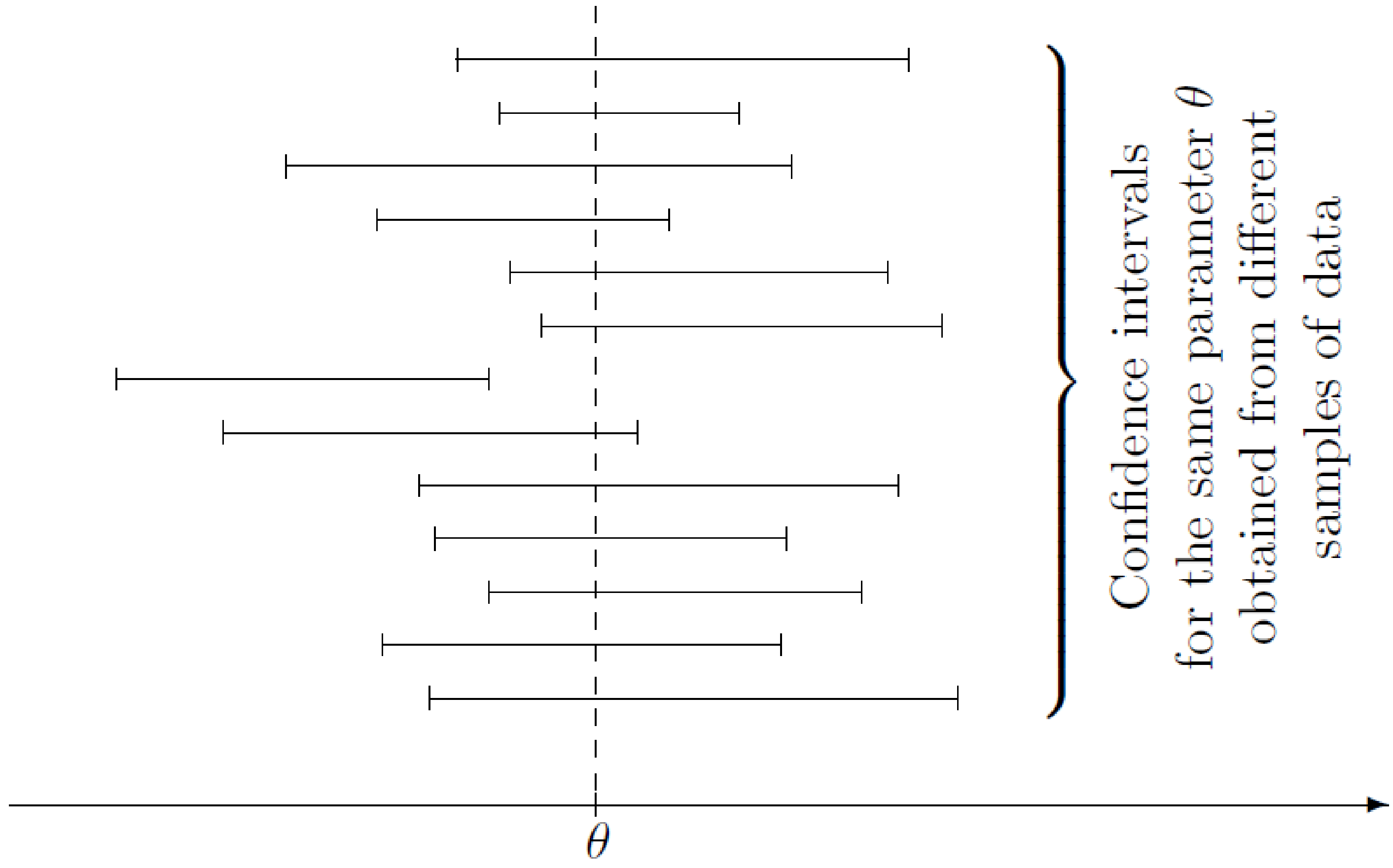
Estimation of standard errors

- $\sigma(\hat{\lambda}_{mle}) = \sqrt{\mathbf{E}(\hat{\lambda}_{mle}^2) - \mathbf{E}^2(\hat{\lambda}_{mle})}$
- $\sigma(\hat{\lambda}_{mle}) = \frac{n\lambda}{(n-1)\sqrt{n-2}}$
- We do not know the parameter λ in this expression; so, use the estimator $1/\bar{X}$ to have an "estimator" for the standard error:
  - $s(\hat{\lambda}_{mle}) = \frac{n}{\bar{X}(n-1)\sqrt{n-2}}$

Confidence intervals

- An interval [a,b] is a $(1-\alpha)$100% confidence interval for the parameter θ, if it contains the parameter with probability $(1-\alpha)$
  - $\boldsymbol{P}(a \le \theta \le b) = 1 - \alpha$
  - The coverage probability $(1-\alpha)$ is also called a confidence level.
  - a and b are computed from sample data and therefore, they are random, but θ is not.

Confidence intervals



A generic methodology to construct confidence intervals

- Find an unbiased estimator for θ.
- Check if the estimator has a Normal distribution.
- Find the standard error of the estimator.
- Obtain the quantiles $\pm z_{\alpha/2}$ from the standard Normal table
- A $(1-\alpha)100\%$ confidence interval for θ is:

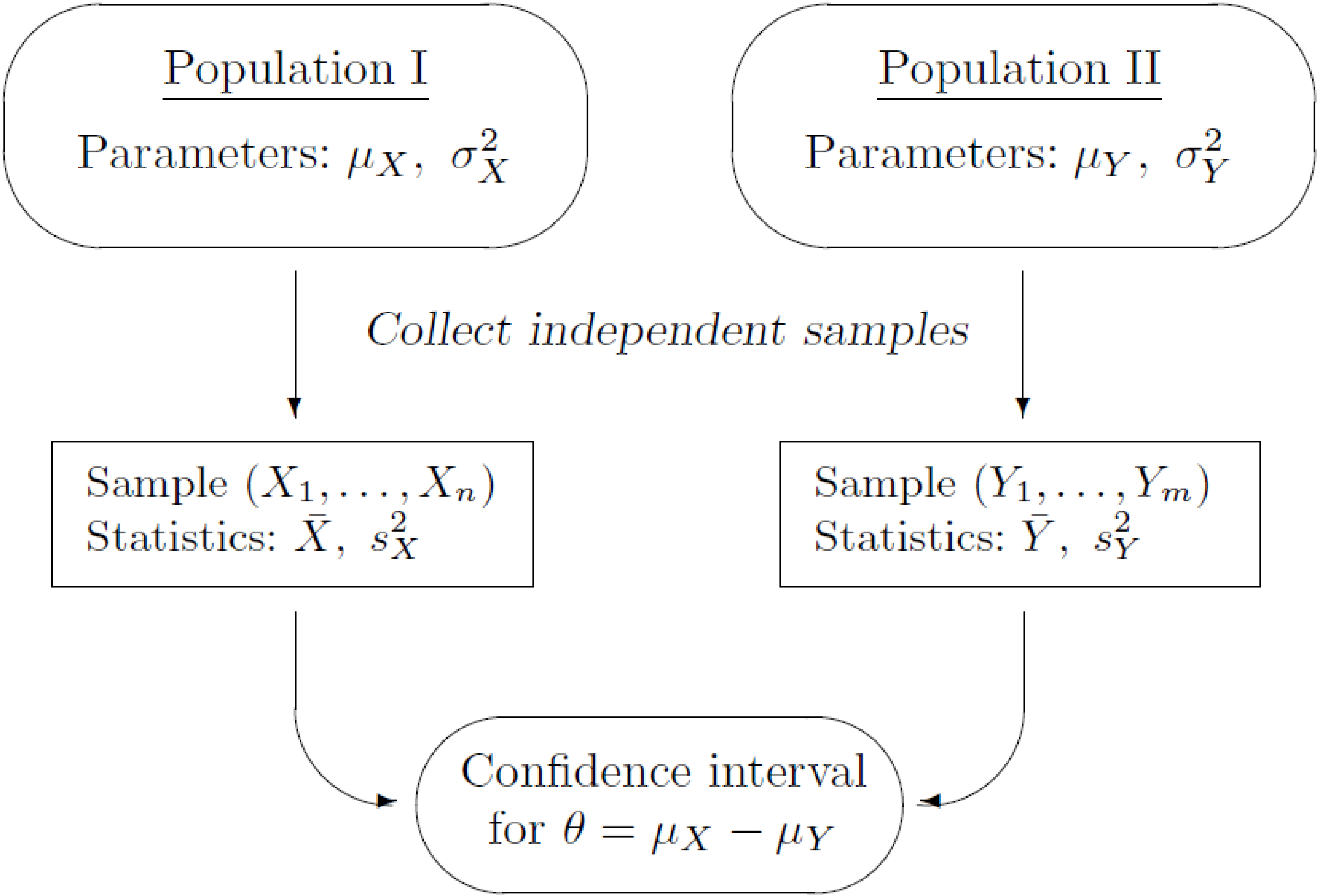
$$\left[\hat{\theta} - z_{\frac{\alpha}{2}} \cdot \sigma\left(\hat{\theta}\right), \hat{\theta} + z_{\frac{\alpha}{2}} \cdot \sigma\left(\hat{\theta}\right)\right]$$

Confidence interval for the population mean

- $\theta = \mu = \mathbf{E}(X)$
- $\hat{\theta} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$
- If the sample comes from Normal distribution, then the estimator is also normal. If the sample comes from any distribution, $\bar{X}$ will be normally distributed if *n* is large.
  - $\mathbf{E}(\bar{X}) = \mu$ (thus it is unbiased)
  - $\sigma(\bar{X}) = \frac{\sigma}{\sqrt{n}}$

$\rightarrow \left[\bar{X} - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right]$ is a $(1-\alpha)100\%$ confidence interval for μ (See Example 9.13)

Confidence interval for the difference between two means



Confidence interval for the difference between two means

- Propose an estimator:
  - $\hat{\theta} = \bar{X} - \bar{Y}$ (unbiased using linearity of **E**)
- Compute standard error:
  - $\sigma(\hat{\theta}) = \sqrt{\mathrm{Var}(\bar{X} - \bar{Y})} = \sqrt{\mathrm{Var}(\bar{X}) + \mathrm{Var}(\bar{Y})}$
  - $\sigma(\hat{\theta}) = \sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}$
- $\left[\bar{X} - \bar{Y} - z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}, \bar{X} - \bar{Y} + z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}\right]$ is a $(1-\alpha)100\%$ confidence interval for θ

Sample size vs. margin

- Margin (Δ) is the length, our estimator is the center is of the confidence interval.
- $n \geq \left(\frac{z_{\alpha/2} \cdot \sigma}{\Delta}\right)^2$
  - If we want to decrease the margin, we need to increase the sample size
  - If we want to increase the confidence level, we need to increase the sample size
- Example 9.15

When σ is unknown

- Estimate it from the sample
- We will focus on two cases:
  - Large samples from any distribution
  - Samples of any size from a Normal distribution
- We will not consider small non-Normal samples
  - Special methods, such as the *bootstrap* method, are needed for such cases.

Large samples

- Instead of $\sigma\left(\hat{\theta}\right)$ use the estimator $s\left(\hat{\theta}\right)$ and obtain an approximate confidence interval

$$\left[\hat{\theta} - z_{\frac{\alpha}{2}} \cdot s\left(\hat{\theta}\right), \hat{\theta} + z_{\frac{\alpha}{2}} \cdot s\left(\hat{\theta}\right)\right]$$

- Example 9.16
- When estimating proportions, i.e., the success probability of a Bernoulli variable, we do not know the standard deviation (mean and standard deviation are both functions of the parameter to be estimated).
  - Example 9.17

Sample size for estimating proportions

- $n \geq \hat{p}(1-\hat{p})\left(\frac{z_{\alpha/2}}{\Delta}\right)^2$
- But, we cannot compute $\hat{p}$ before deciding on the sample size, n.
- Use the maximum value of $\hat{p}(1-\hat{p})$ instead, which is 0.25.
  - $n \geq 0.25\left(\frac{z_{\alpha/2}}{\Delta}\right)^2$

Small samples

- Use Student's t distribution instead of the normal distribution.
- If the sample $X_1, \dots, X_n$ is from Normal distribution with unknown μ and σ:
  - Estimate σ by $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2}$
  - Use t-distribution with (n-1) degrees of freedom
  - Confidence interval for the mean:
    - $\left[\bar{X} - t_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}, \bar{X} + t_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}\right]$

Small samples: comparing means of two populations

- Equal variances:
  - $\left[\bar{X} - \bar{Y} - t_{\frac{\alpha}{2}} \cdot s_p\sqrt{\frac{1}{n} + \frac{1}{m}}, \bar{X} - \bar{Y} + t_{\frac{\alpha}{2}} \cdot s_p\sqrt{\frac{1}{n} + \frac{1}{m}}\right]$
  - s_p is the pooled standard deviation:
    - $s_p^2 = \frac{(n-1)s_X^2 + (m-1)s_Y^2}{n+m-2}$
- Unequal variances:
  - $\left[\bar{X} - \bar{Y} - t_{\frac{\alpha}{2}}\sqrt{\frac{s_X^2}{n} + \frac{s_Y^2}{m}}, \bar{X} - \bar{Y} + t_{\frac{\alpha}{2}}\sqrt{\frac{s_X^2}{n} + \frac{s_Y^2}{m}}\right]$