Statistics I – Chapter 8 Estimation for One Population (Part 2)

Ling-Chieh Kung

Department of Information Management
National Taiwan University

December 5, 2012

Introduction

- Last time we introduced the idea of **interval estimation**.
  - Instead of suggesting a single value, we suggest an interval.
  - It is an interval that we know **how good it is**: the probability for the interval to cover the parameter.
  - We can measure the probability because we know the **sampling distribution** of the estimator.
- We introduced how to estimate the **population mean** when the population variance is **known**.
- Today we discuss some other parameters.

Road map

- **Estimating the population mean.**
  - When the variance is unknown.
- Estimating population proportion.
- Estimating population variance.

Review

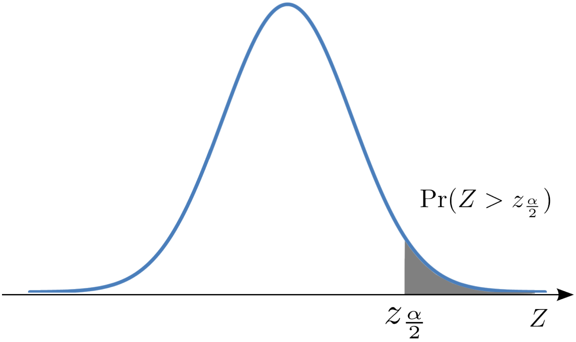
- To estimate the population mean μ when the population variance σ^2 is known:
  - If applicable, we use **the z distribution**.
  - Calculate the sample mean $\bar{x}$.
  - Calculate the standard error $\sigma_{\overline{X}}$: $\frac{\sigma}{\sqrt{n}}$ or $\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$.
  - Calculate the critical value $z_{\frac{\alpha}{2}}$ based on the z distribution and the confidence level $1-\alpha$.
  - The bounds of the interval is $\bar{x} \pm z_{\frac{\alpha}{2}}\sigma_{\overline{X}}$.

Review

- The critical value $z_{\frac{\alpha}{2}}$ satisfies

$$\Pr(Z > z_{\frac{\alpha}{2}}) = \frac{\alpha}{2},$$

where Z follows the standard normal distribution.



Review

- The **standard error** is the standard deviation of the estimator: in this case, $\overline{X}$.
- When we apply the z distribution, we use the fact that

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim \mathrm{ND}(0, 1)$$

  if the population is infinite.
  - What if the population is finite?

Review

- What are the conditions for applying the z distribution?

Sample size	Population distribution	
	Normal	Nonnormal
$n \geq 30$	z distribution	z distribution (CLT)
$n < 30$	z distribution	Nonparametric

When the variance is unknown

- In most cases in practice, the population mean is **unknown**.
- In estimating the population mean, what is the difficulty?
- We have no way to calculate the standard error!
  - $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$ or $\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$.
- In this case, a natural way is to substitute σ by S, the sample standard deviation.
- While $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim \text{ND}(0,1)$, do we know the distribution of

$$\frac{\overline{X}-\mu}{S/\sqrt{n}}?$$

The t distribution

- When we replace σ by S, we rely on the following fact:

Proposition 1

For a normal population, the quantity

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

follows the t distribution with degree of freedom $n - 1$.

- We know the sampling distribution of T (the population must be **normal**). We call it **the t distribution**.
- The only parameter is the **degree of freedom**.
- Its pdf is known. Its cdf can be found by tables or software.

The t distribution

- The t distribution is defined as follows:

Definition 1

A random variable X follows the t distribution with degree of freedom n, denoted as $X \sim t(n)$, if

$$f(x|n) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}},$$

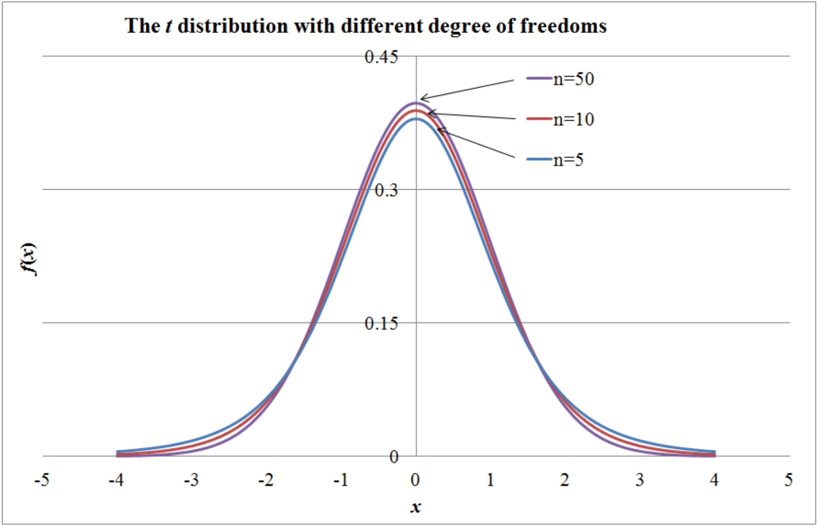
for all $x \in (-\infty, \infty)$.

- $\Gamma(x) = \int_0^\infty z^{x-1}e^{-z}dz$ is the gamma function.

The z and t distributions

- Let's compare $Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ and $T = \frac{\overline{X}-\mu}{S/\sqrt{n}}$.
  - Because we do not know σ, we use S to substitute it.
  - $Z \sim \text{ND}(0,1)$ and $T \sim t(n-1)$.
  - As the t distribution is a substitution of the z distribution, it is designed to be also **centered at 0**: $\mathbb{E}[T] = \mathbb{E}[Z] = 0$.
  - However, as we add one more random variable into the formula (σ is a known constant), T will be "**more random**" than Z, i.e., $\text{Var}(T) > \text{Var}(Z)$.
  - Graphically, t curves will be **flatter** than the z curve.
  - Fact: $t(n) \to \text{ND}(0,1)$ as $n \to \infty$.

The z and t distributions



Using the t distribution

- As we know that $T = \frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$, we may construct the confidence interval as follows:
  - Calculate the **sample mean** $\bar{x}$.
  - Calculate the **multiplier** $\frac{s}{\sqrt{n}}$.
  - Calculate the **critical value** $t_{\frac{\alpha}{2},n-1}$ based on the t distribution and the confidence level $1-\alpha$:

$$\Pr(T > t_{\frac{\alpha}{2},n-1}) = \frac{\alpha}{2}.$$

  - The bounds of the interval is

$$\bar{x} \pm t_{\frac{\alpha}{2},n-1}\frac{s}{\sqrt{n}}.$$

Using the t distribution

- In calculate $t_{\frac{\alpha}{2},n-1}$, all we need is a probability table or an Excel function.
- We do not even need to know the pdf of the t distribution.
- We also do not know:
  - Why T follows the t distribution?
  - How did statisticians define/design the t distribution?
  - The physical meaning of the t distribution.
- Anyway, let's use it to do some estimations.

Example 1

- I did not announce the average score of the midterm.
- You want to estimate the average of the 57 scores with a $1 - \alpha = 95\%$ confidence level.
- Your sample is $\{69, 72, 92, 78, 81, 76, 54, 51, 91\}$.
  - Sample size $n = 9$.
  - Sample mean $\bar{x} = 73.78$.
  - Sample standard deviation $s = 14.32$.

Example 1

- If you do not have the population standard deviation σ:
  - Because the population is normal and the population variance is unknown, we use the t distribution to construct the interval.
  - The sample mean is $\bar{x} = 73.78$.
  - The multiplier is $\frac{s}{\sqrt{n}} = 4.77$.
  - The critical value is $t_{\frac{\alpha}{2},n-1} = t_{0.025} = 2.306$. Note that the degree of freedom is $n - 1 = 8$!
  - The interval bounds are $73.78 \pm 2.306 \times 4.77$.
  - With a 95% confidence level, the mean of the midterm grades is within $[62.77, 84.78]$.

Example 1

- If we know the population standard deviation is $\sigma = 16.72$, we may use the z distribution and get $[62.85, 84.70]$.
- Comparisons:

Population variance σ^2	Unknown	Known
Distribution to use	t distribution	z distribution
Sample mean $\bar{x}$	73.78	73.78
Critical value	$t_{0.025,8} = 2.306$	$z_{0.025} = 1.96$
Multiplier	$\frac{s}{\sqrt{n}} = 4.77$	$\frac{\sigma}{\sqrt{n}} = 5.57$
Confidence interval	$[62.77, 84.78]$	$[62.85, 84.7]$

Using the t distribution

- Only when the population is **normal**, the quantity $T = \frac{\overline{X}-\mu}{S/\sqrt{n}}$ follows the t distribution.
- If the population is nonnormal, we do not know the distribution of T.
- Fortunately, if the **sample size is large** ($n \geq 30$):
    - We may apply the **central limit theorem** and conclude that $\overline{X} \sim ND(\mu, \frac{\sigma}{\sqrt{n}})$. But how to deal with the unknown σ?
    - The sample variance S^2 will be close to the population variance σ^2 (i.e., S^2 is a consistent estimator of σ^2).
    - We may use $\frac{s}{\sqrt{n}}$ as an substitute of $\frac{\sigma}{\sqrt{n}}$.
- We then use the z distribution to construct the interval.

Example 2

- A survey is conducted to study the average number of months a Taiwanese college graduate spends on finding the first job after graduation.
- 100 persons are randomly selected and the data are recorded:

$$6 \quad 2 \quad 2 \quad 3 \quad 1 \quad 0 \quad 15 \quad 11 \quad \cdots \quad 4.$$

- Construct a confidence interval with a 99% confidence level.

Example 2

- Answer:
  - Because the population size is large, we use the z distribution to construct the interval. Because the population variance σ^2 is unknown, we will use the sample variance s^2 as a substitute.
  - The sample mean is $\bar{x} = 2.55$.
  - The sample standard deviation is $s = 2.09$.
  - The standard error is (approximately) $\frac{s}{\sqrt{n}} = 0.209$.
  - The critical value is $z_{\frac{\alpha}{2}} = z_{0.005} = 2.576$.
  - The interval bounds are $2.55 \pm 2.576 \times 0.209$.
  - With a 99% confidence level, the average months for a Taiwanese college graduate to find the first job is within $[2.01, 3.09]$.

Remarks

- We may ignore the finite population issue.
  - The existence of a finite population has somewhat affected the calculation of the sample standard deviation S.
- If the population is normal and the sample size is large, it is also fine to use the z distribution (with s substituting σ).
  - This is also due to the central limit theorem.
- If the population is nonnormal and the sample size is small, we must relegate to nonparametric methods.
  - However, the t distribution for estimating the population mean is **robust** to the normal population assumption: Having nonnormal population does not harm a lot.

Summary

- To estimate the population mean μ:

σ^2	Sample size	Population distribution	
		Normal	Nonnormal
Known	$n \geq 30$	z	z
	$n < 30$	z	Nonparametric
Unknown	$n \geq 30$	t or z	z
	$n < 30$	t	Nonparametric

  - If z distribution, do finite population correction if $n > 0.05N$.
  - If t distribution, no need to do this.

Road map

- Estimating the population mean.
  - When the variance is unknown.
- **Estimating population proportion**.
- Estimating population variance.

Estimating population proportion

- For a population $\{x_i\}_{i=1,\ldots,N}$, we label each entity as 1 or 0.
  - 1 for boys, 0 for girls.
  - 1 for defects, 0 for good products.
  - 1 for having monthly income higher than \$30000, 0 or not.
- The **population proportion** is $p = \frac{1}{N}\sum_{i=1}^{N} x_i$.
- Let $X = \sum_{i=1}^{n} X_i$. the **sample proportion**

$$\hat{p} = \frac{X}{n}$$

  is an **unbiased** estimator of p (why)?
  - It is also consistent and more efficient than most other unbiased estimators of the population proportion.

Estimating population proportion

- To conduct interval estimation for the population proportion p, we will use the sample proportion $\hat{p}$ as the **center** of the interval.
- How to decide the leg length based on the confidence level?
- Suppose the **sample size is large** ($n \geq 30$).
  - The **central limit theorem** implies that the sample proportion follows the normal distribution.
  - The mean is p. The standard error is $\sqrt{\frac{p(1-p)}{n}}$. We have

$$\frac{\hat{p} - p}{\sqrt{p(1-p)/n}} \sim \text{ND}(0, 1).$$

- Then we may use the z distribution... May we?

Estimating population proportion

- The standard error $\sqrt{\frac{p(1-p)}{n}}$ contains p, which is **unknown**!
- In fact, as the population variance $p(1-p)$ depends on p, the population variance is unknown.
- So similar to the case of estimating population mean with unknown population variance, we will use $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ as an **substitute** of $\sqrt{\frac{p(1-p)}{n}}$.

Example

- A manufacturer recently got an offer from a downstream retailer. The retailer asked for 8500 units of a newly designed product and will pay \$900 for it. If the manufacturer cannot make it, the retailer is also willing to pay \$400 for 4000 units.
- The capacity of the manufacturer is 10000 units. So whether it can promise the retailer for delivering 9000 units depends on the **yield rate**, the proportion of products passing the quality requirements.
- If its yield rate can reach 85%, it can sign the $(8500, \$900)$ contract. But because the product is new, it does not have past data for the yield rate.

Example

- An inspector was assigned the task of estimating the yield rate. She ran a production run for 100 products and found that 91 are good.
- The manager will accept the offer only if she is 99% sure that the yield rate is above 85%. Should she accept it?
  - What is the parameter to estimate?
  - If we use the sample proportion as the estimator, what is the point estimate?
  - How to construct the confidence interval for different values of confidence level?

Example

- The required confidence level is $1 - \alpha = 0.99$:
  - Because the sample size is large, we may use the z distribution to construct the interval.
  - The **sample proportion** $\hat{p} = \frac{91}{100} = 0.91$.
  - The **multiplier** is $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.0286$.
  - The **critical value** is $z_{0.005} = 2.576$.
  - The interval bounds are $0.91 \pm 2.576 \times 0.0286$.
  - With a 99% confidence level, the yield rate is between 83.6% and 98.4%. The manager should not accept the offer.

Polls for elections

- One main application of estimating the population proportion is the **polls for elections**.
- How to read the results of polls?
  - Read those $\hat{p}$s.
  - Read the maximum error(s).
  - Read the sample size and confidence level.
  - Compare those confidence intervals.
  - Read the sampling method(s).

Polls for elections: example 1

- Proportion of voters supporting candidate 1: $\hat{p}_1 = 0.5$.
- Proportion of voters supporting candidate 2: $\hat{p}_2 = 0.28$.
- Simple random sampling.
- Population: All voters living in Tainan.
- $1 - \alpha = 95\%$.
- Sample size $n = 825$.
- Max error $= 0.034$.

五都選舉 / 大台南民調　賴清德仍以5成支持率勝郭添財

2010年10月15日 00:02

政治中心／綜合報導

距離五都選舉只剩約一個半月時間，根據最新媒體民調顯示，大台南市長選舉，目前仍由民進黨候選人賴清德以5成的支持率領先，國民黨候選人郭添財的支持率則為28%。

這項民調是TVBS民意調查中心，在10月13日晚上6時30分至10時15分，以電話後4碼電腦隨機抽，人員電話訪問住鎺，有效樣本825位20歲以上的台南縣市民眾，在95%的信心水準下，抽樣誤差為正負3.4個百分點。

郭添財指出，他若當選大台南市長福利政策會增加，不會減少。他還主張，台南縣市65歲以上長輩免費裝假牙、免費醫療公車，還會爭取更多的福利。

更多照片

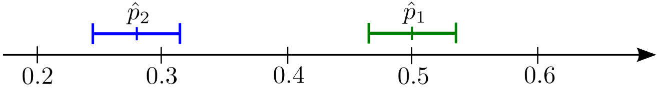
(http://www.nownews.com/2010/10/15/11490-2655158.htm.)

Polls for elections: example 1

- The maximum error is

$$z_{0.025}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n}} \approx 1.96\sqrt{\frac{(0.5)(0.5)}{825}} \approx 0.034.$$

- So for p_1 and p_2 (the true proportions of voters supporting candidates 1 and 2):
  - $\hat{p}_1 = 0.5$ and $\hat{p}_2 = 0.28$.
  - Confidence intervals: $[0.466, 0.534]$ and $[0.246, 0.314]$.



  - The difference is (statistically) **significant**: There is an enough evidence that, with a 95% confidence level, candidate 1 is in an advantage.

Polls for elections: example 2

- Proportion of voters supporting candidate 1: $\hat{p}_1 = 0.43$.
- Proportion of voters supporting candidate 2: $\hat{p}_2 = 0.42$.
- Simple random sampling.
- Population: All voters living in Taipei.
- $1 - \alpha = 95\%$.
- Sample size $n = 824$.
- Max error $= 0.034$.

花風暴真傷？郝、蘇最新民調　43%：42%陷膠著 (2010/10/07 23:14)

政治中心／綜合報導

距離年底五都選舉已剩不到2個月時間，針對日前傳出新生高架橋弊案，檢調單位已數度搜索台北市府相關單位，並約談市府人員，即將在一個月後開幕的花卉博覽會，也因採購價格偏高而爭議不斷，這些是否會影響台北市長選情？根據TVBS民調中心最新的調查結果顯示，目前台北市長郝龍斌的支持度是43%，民進黨的蘇貞昌則是42%，雙方支持度不相上下，陷入膠著狀況。

本次調查是TVBS民意調查中心在10月5、6日晚間18:30~22:00進行的調，共接觸969位20歲上台北市民眾，拒訪人數145位，拒訪率為15.0%，有效樣本數為824位，在95%的信心水準下，抽樣誤差為正負3.4個百分點。抽樣方法採用電話號碼後四碼隨機抽樣。

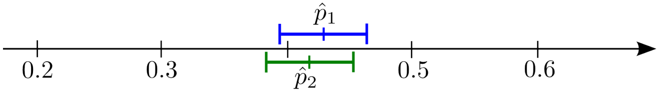
(http://www.nownews.com/2010/10/07/11606-2653111.htm.)

Polls for elections: example 2

- The maximum error is

$$z_{0.025}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \approx 1.96\sqrt{\frac{(0.43)(0.57)}{824}} \approx 0.034.$$

- So for p_1 and p_2 (the true proportions of voters supporting candidates 1 and 2):
  - $\hat{p}_1 = 0.43$ and $\hat{p}_2 = 0.42$.
  - Confidence intervals: $[0.396, 0.464]$ and $[0.386, 0.454]$.



  - The difference is (statistically) **insignificant**: There is no enough evidence that candidate 1 is in an advantage.

When the population is finite

- If we adopt sampling without replacement and the population size is small ($n > 0.05N$), we need to include the **finite population factor** $\sqrt{\frac{N-n}{N-1}}$ in the multiplier:

$$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\sqrt{\frac{N-n}{N-1}}.$$

- Everything then follows.

When the sample size is small

- If the sample size is small ($n < 30$), the sample proportion is no longer normal.
- While we do not know the distribution of the sample proportion $\hat{p} = \frac{X}{n}$, we know $X = \sum_{i=1}^{n} X_i \sim \text{Bi}(n, p)$. We may thus do an interval estimation.
- There are multiple ways of doing the inference:
  - Based on binomial distributions.
  - Based on F distributions.
- We will skip this topic.

Remark

- Instead of substituting $\sqrt{\frac{p(1-p)}{n}}$ by $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$, some people use $\sqrt{\frac{\hat{p}(1-\hat{p})}{n-1}}$ because it is unbiased for $\sqrt{\frac{p(1-p)}{n}}$.
- Instead of substituting $\sqrt{\frac{p(1-p)}{n}}\sqrt{\frac{N-n}{N-1}}$ by $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\sqrt{\frac{N-n}{N-1}}$, some people use $\sqrt{\frac{\hat{p}(1-\hat{p})}{n-1}}\sqrt{\frac{N-n}{N}}$ because it is unbiased for $\sqrt{\frac{p(1-p)}{n}}\sqrt{\frac{N-n}{N-1}}$.
- Both of the two alternative substitutes improve a little when n is large.
- We will adopt the naïve way: Just replace p by $\hat{p}$.

Road map

- Estimating the population mean.
  - When the variance is unknown.
- Estimating population proportion.
- **Estimating population variance**.

Estimating population variance

- Another population parameter that is often of interest is the **population variance** σ^2.
  - Here we go back to discuss a quantitative population rather than a qualitative one.
- The most common estimator is the **sample variance** S^2.
  - The denominator is $n - 1$!
  - As an estimator of σ^2, S^2 is unbiased and consistent.
- Interestingly, the sample standard deviation S is a **biased** estimator of the population standard deviation σ.

Estimating population variance

- To construct the confidence interval for the population variance, we rely on the quantity

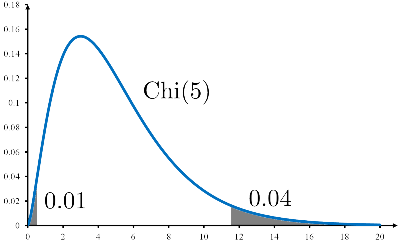
$$\chi^2 = \frac{(n-1)S^2}{\sigma^2},$$

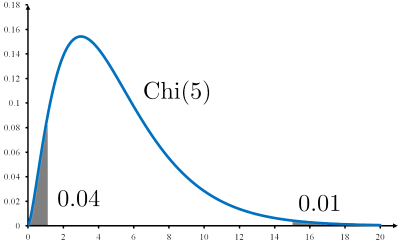
which follows **the chi-square distribution** with degree of freedom $n-1$ if the population is **normal**.

  - The notation χ^2 here is a random variable.
  - The estimation is quite sensitive to the normal population assumption; this method is **not robust**.

The chi-square distribution

- For a random variable $\chi^2 \sim \text{Chi}(n-1)$, how to find two cutoffs a and b such that $\Pr(a \leq \chi^2 \leq b) = 1 - \alpha$?
- There are multiple ways (assuming $1 - \alpha = 0.95$):





Left tail probability = 0.01
Right tail probability = 0.04

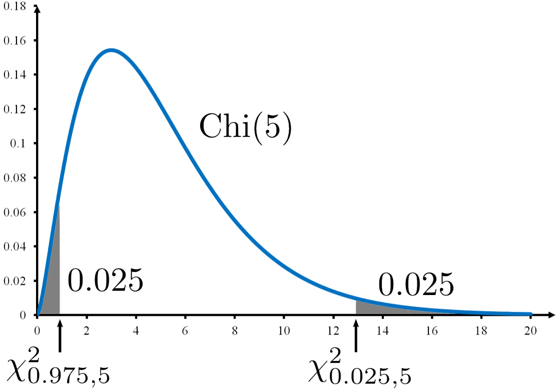
Left tail probability = 0.04
Right tail probability = 0.01

The chi-square distribution

- One particular way is "half-half".
- We may choose the two cutoffs as $\chi^2_{1-\frac{\alpha}{2},n-1}$ and $\chi^2_{\frac{\alpha}{2},n-1}$, where

$$\Pr(\chi^2 > \chi^2_{y,n-1}) = y.$$

  for $\chi^2 \sim \text{Chi}(n-1)$.
- The notation $\chi_{y,n-1}$ is a critical value.



The chi-square distribution

- By using $\chi^2_{1-\frac{\alpha}{2},n-1}$ and $\chi^2_{\frac{\alpha}{2},n-1}$:
  - The interval constructed with them are not the smallest.
  - But because it is hard to find the smallest interval, in practice people use these two cutoffs for convenience.

Estimating population variance

- We then have

$$
\begin{aligned}
1-\alpha &= \Pr\left(\chi^2_{1-\frac{\alpha}{2},n-1} \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi^2_{\frac{\alpha}{2},n-1}\right) \\
&= \Pr\left(\chi^2_{1-\frac{\alpha}{2},n-1}\sigma^2 \leq (n-1)S^2 \leq \chi^2_{\frac{\alpha}{2},n-1}\sigma^2\right) \\
&= \Pr\left(\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2},n-1}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2},n-1}}\right).
\end{aligned}
$$

- Given a realized value of sample variance s^2, with a $1-\alpha$ confidence level, the population variance is between

$$
\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2},n-1}} \quad \text{and} \quad \frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2},n-1}}.
$$

Example

- In a company, the human resource team is estimating how diverse the workers' weekly work hours are.
  - The population: workers working in the company.
  - The parameter: the population variance of all workers' weekly work hours.
- It is known that the population is **normal**.
- The team collect a sample of 20 workers and obtain their weekly work hour.
- The sample variance $s^2 = 18.26$ square hours.
- Estimate the variance of all workers' weekly work hours with a 90% confidence level.

Example

- The required confidence level is $1 - \alpha = 0.9$:
  - Because the population is normal we may use the chi-square distribution to construct the interval.
  - The **sample variance** $s^2 = 18.26$.
  - The **degree of freedom** is $n - 1 = 19$.
  - The **critical values** are

$$\chi^2_{0.95,19} = 10.117 \quad \text{and} \quad \chi^2_{0.05,19} = 30.144.$$

  - The bounds are $\frac{(19)(18.26)}{30.144} = 11.51$ and $\frac{(19)(18.26)}{10.117} = 34.29$.
  - With a 90% confidence level, the variance of all workers' weekly work hour is between 11.51 and 34.29.

Example

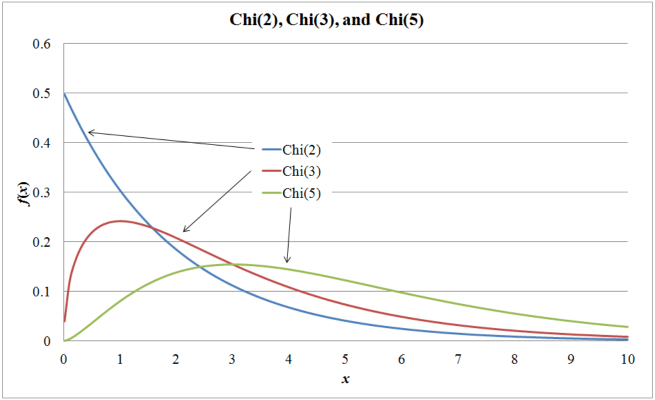
- Note that the bounds are $\frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}}$ and $\frac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}}$.
  - Does that mean a larger sample size results in a larger confidence interval?
  - No! Because the two **critical values** will also increase when n increases. This is because the chi-square distribution becomes **flatter** when n increases.
- In this example:

n	10	20	50	100	200
Lower bound	9.71	11.51	13.49	14.67	15.6
Upper bound	49.42	34.29	26.37	23.46	21.71

- It can be shown that increasing n reduces the interval length.

The chi-square distribution

- The chi-square curve gets flatter when the degree of freedom gets larger.



Population standard deviations

- To estimate the population standard deviation σ:
  - First estimate the population variance.
  - Then take square root for the two bounds.
- E.g., if the 90% confidence interval for σ^2 is $[11.51, 34.29]$, the 90% confidence interval for σ is

$$\left[\sqrt{11.51}, \sqrt{34.29}\right] = [3.39, 5.86].$$

Remarks: sampling distributions

- The foundation for estimating the population variance σ^2 is that $\frac{(n-1)S^2}{\sigma^2} \sim \text{Chi}(n-1)$.
  - We spent a lot of time proving this so we **know** it is true.
- The foundation for estimating the population mean μ when σ^2 is known is that $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim \text{ND}(0,1)$ (for finite populations).
  - We also **know** it is true.
- The foundation for estimating the population mean μ when σ^2 is unknown is that $\frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$.
  - We did not prove this. We can only **believe** that it is true.
- Keep in mind that all these require a normal population. Otherwise we need the central limit theorem.

Remarks: z, t, χ^2, and F distributions

- The z, t, and chi-square distributions are three of the most important sampling distributions.
- The fourth very important sampling distribution, the F distribution, will be introduced in the next semester.
- In using them to do statistical inference, all we need is to **find critical values** based on the parameters and the predetermined **tail probability**.
  - Be familiar with probability tables or software.

Remarks: one-sided estimations

- We have introduced **two-sided confidence intervals**.
  - "With a 95% confidence level, μ is within a and b."
- There are also **one-sided confidence intervals**:
  - "With a 95% confidence level, μ is above c."
- We omitted one-sided confidence intervals as they are less frequently used than two-sided ones.
- As the concepts are similar, now you are able to teach yourself how to do one-sided estimations.

Remarks: sample size

- Increasing sample size reduces interval length.
- Given a predetermined confidence level and an interval length, it is possible to determine the sample size that can achieve the interval length.
  - For estimating population means and proportions, we may derive formulas.
  - For estimating population variances, we need to do trial-and-error.