STAT417 lecture NOTE 9.0

When sample size is small. only focus on normal models.

<u>Eg.1.</u> If $X_1, X_2, \dots, X_{16}$ are iid samples from $N(\mu, \sigma^2)$.

σ^2 is known. then. the 95%-CI for μ is

$$\bar{X} \pm 1.96 \frac{\sigma}{\sqrt{16}} = \bar{X} \pm 1.96 \left(\frac{\sigma}{4}\right).$$

In general, γ-CI for μ is

$$\bar{X} \pm z_{\frac{\gamma+1}{2}} \left(\frac{\sigma}{\sqrt{n}}\right).$$

<u>Eg.2</u> If $X_1, \dots, X_{16}$ are iid samples from $N(\mu, \sigma^2)$. σ is not known.

then the γ-CI for μ is.

$$\bar{X} \pm t_{\frac{\gamma+1}{2}}(n-1) \frac{S}{\sqrt{n}}, \qquad S = \text{the sample standard deviation.}$$

For example. when $\gamma = 95\%$, then. the 95%-CI is.

$$\bar{X} \pm 2.131 \left(\frac{S}{\sqrt{16}}\right).$$

<u>Eg.3</u> If $X_1, \dots, X_{25}$ are iid samples from $N(\mu, \sigma^2)$. σ is unknown.

then the 99%-CI for μ is

$$\bar{X} \pm t_{0.995}(24) \frac{S}{\sqrt{25}} = \bar{X} \pm 2.797 \left(\frac{S}{5}\right).$$

Chi-square distribution:

The __chi-square distribution with n degrees__ of freedom, or $\chi^2(n)$.

is the distribution of $X_1^2+X_2^2+\cdots+X_n^2$, where $X_1, X_2, \ldots, X_n$ are i.i.d

$\sim$ N(0,1).

Confidence interval for σ^2:

If $X_1, X_2, \ldots, X_n$ are i.i.d samples from $N(\mu, \sigma^2)$, where σ^2 is unknown.

we already know that $\hat{\sigma}^2_{ML} = s^2$, the sample variance. How to establish

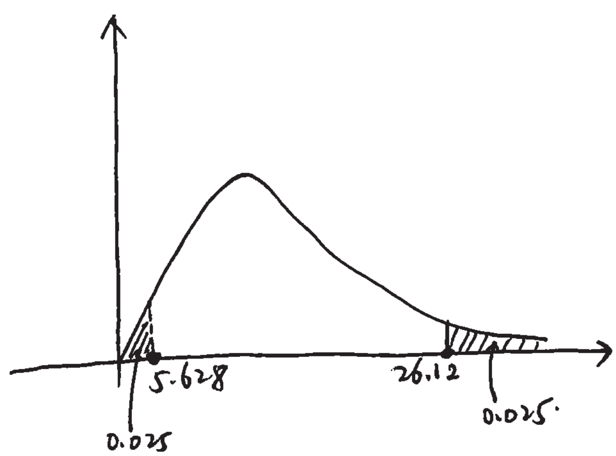
a confidence interval for σ^2?

Theorem 1: $\frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$.

Next we show how to use theorem 1 to establish CI for σ^2.

__E.g. 4.__ If $n=15$, $s^2=2$, establish a 95% CI for σ^2.

Sol: Consider $\chi^2(14)$, which has the following pdf.

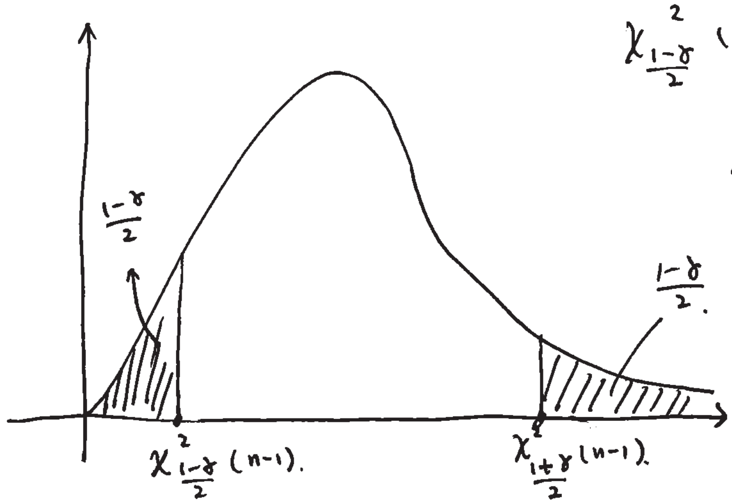


so. $5.628 \le \frac{(n-1)s^2}{\sigma^2} \le 26.12$

we get $\frac{14s^2}{26.12} \le \sigma^2 \le \frac{14s^2}{5.628}$, that is $[0.54, 2.49]$.

In general, the γ-CI for σ^2 is obtained by

$$\chi^2_{\frac{1-\gamma}{2}}(n-1) \le \frac{(n-1)s^2}{\sigma^2} \le \chi^2_{\frac{1+\gamma}{2}}(n-1)$$



so

$$\frac{(n-1)s^2}{\chi^2_{\frac{1+\gamma}{2}}(n-1)} \le \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{\frac{1-\gamma}{2}}(n-1)}$$

<u>EX2</u>. If $n=25$, $s^2=4$, $\gamma=0.99$. $\chi^2_{0.005}(24) = 9.89$.

$\chi^2_{0.995}(24) = 45.56$

Find the γ-CI for σ^2. $[\ 2.11\ ,\ 9.76\ ]$

Hypothesis testing:

If $X_1, X_2, \dots, X_n$ are iid samples from $N(\mu, \sigma^2)$. test

H_0: $\mu = \mu_0$. ~~we use Z test if~~

If σ^2 is known, we use Z test: $z = \frac{\sqrt{n}(\bar{x}-\mu_0)}{\sigma}$.

the p-value, under H_0, is

$$P(|Z| > |z|) = 2(1 - \Phi(|z|)).$$

If σ^2 is unknown. ~~but n is big~~. we use T-test: $t = \frac{\sqrt{n}(\bar{x}-\mu_0)}{s}$.

the p-value is $P(|T| > |t|) = 2(1 - G(|t|; n-1))$.

$G(\cdot; n-1)$ is the $t(n-1)$-distribution.

E.g. 5:

If $X_1, X_2, \ldots, X_{30}$ are iid-samples from $N(\mu, \sigma^2)$.

If $\sigma^2 = 4$. $\bar{X} = 5$. test $H_0: \mu = 4$

the z-value is $z = \sqrt{30}\left(\frac{5-4}{2}\right) = 2.74$.

So the p-value is $2(1 - \Phi(2.74)) = 0.012 < 0.05$.

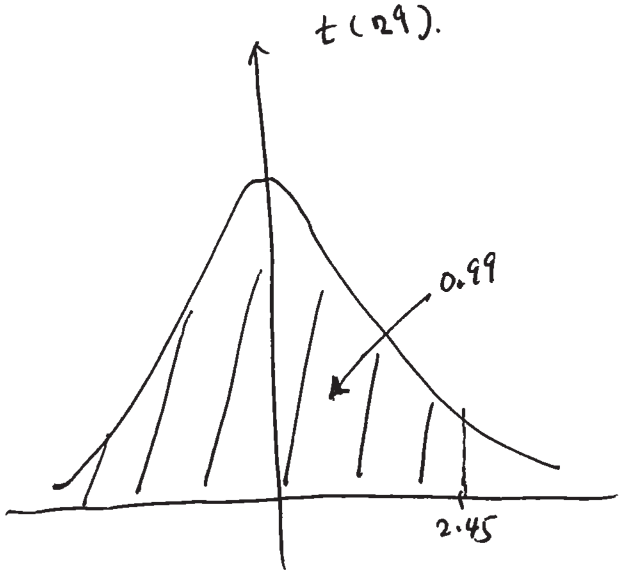
So reject H_0.

E.g. 6. Following Eg5. but drop $\sigma^2 = 4$. instead, we know. $s^2 = 5$.

the t-value is $t = \sqrt{30}\left(\frac{5-4}{\sqrt{5}}\right) = 2.45$.

so the p-value is $2(1 - G(2.45; 29)) = 2(1-0.99) = 0.02 < 0.05$

So reject H_0.



Test proportions:

E.g. 7 A sample of 2500 voters. θ = the proportion of voters that support obamacare. If we observe that 2000 of 2500 supports ~~ob~~.

Test $H_0: \theta = 0.78$ $\qquad \bar{X} = \frac{2000}{2500} = 0.8.$

use z-test: $z = \frac{\sqrt{n}(\bar{X} - \theta_0)}{\sqrt{\theta_0(1-\theta_0)}} = \frac{\sqrt{2500}(0.8-0.78)}{\sqrt{0.78 \times 0.22}} = 2.41.$

so the p-value is $2(1 - \Phi(2.41)) = 0.016 < 0.05.$

reject H_0:

Sample size calculation:

For CI: ~~Question~~: A short CI is prefered. but may need more samples.
Question is, how much sample is needed for the length of CI not greater than any preselected threshold?

E.g. 8. If $X_1, \dots, X_n$ are iid $\sim N(\mu, \sigma^2)$. the 95%-CI for μ is

$\bar{X} \pm 1.96 \frac{\sigma}{\sqrt{n}}$. $\qquad 1.96 \frac{\sigma}{\sqrt{n}}$ is called the margin error.

How ~~many sam~~ large is n so that the ~~length~~ margin error is less than 0.01. ?

Sol: set $1.96 \frac{\sigma}{\sqrt{n}} \leq 0.01$. then. $n \geq \sqrt{196\sigma}$.

In general. If we want the margin error of γ-CI to be less than δ.
then $z_{\frac{\gamma+1}{2}} \cdot \frac{\sigma}{\sqrt{n}} \leq \delta$, so $n \geq \left(\frac{z_{\frac{\gamma+1}{2}} \cdot \sigma}{\delta}\right)^2.$

E.g. 9 How many times must we toss a coin such that the 95%-CI for $\theta = P(H)$ has length ~~than~~ less than 0.01 ?

recall the CI for θ is

$$\bar{X} \pm 1.96 \sqrt{\frac{\bar{X}(1-\bar{X})}{n}}.$$

so. $2 \times 1.96 \times \sqrt{\frac{\bar{X}(1-\bar{X})}{n}} \le 0.01.$

while. $\sqrt{\bar{X}(1-\bar{X})} \le \frac{1}{2}$. so. we only to make sure.

$$\frac{1.96}{\sqrt{n}} \le 0.01. \quad \text{so.} \quad n \ge 196^2 = 38416.$$

Textbook 4.3.1.