Recall that the 95% CI of μ is $\left[\bar{X}_n - 1.96\frac{\hat{\sigma}}{\sqrt{n}}, \bar{X}_n + 1.96\frac{\hat{\sigma}}{\sqrt{n}}\right]$

If I want the CI length to be ℓ (e.g. $\ell = 1$), what is the minimum n?

Ans: $2\times 1.96\frac{\hat{\sigma}}{\sqrt{n}} = \ell \Rightarrow \sqrt{n} = \frac{2\times 1.96\hat{\sigma}}{\ell} \Rightarrow n = \left\lceil \frac{(2\times 1.96\,\hat{\sigma})^2}{\ell^2} \right\rceil$

$\lceil x \rceil$ means the smallest integer $\geq x$ — ceiling

Special case 1: 95% CI of σ^2 in a $N(\mu, \sigma^2)$

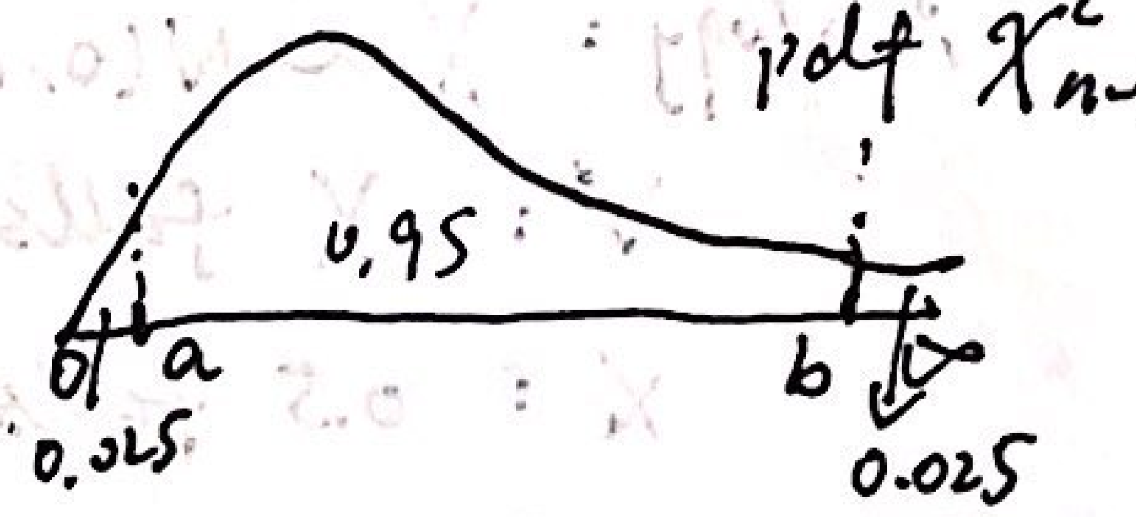
$X_1, \cdots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$

$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$ is an unbiased estimator of σ^2

Use the distribution of S_n^2 to construct a 95% CI of σ^2

Recall: $\frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2_{n-1}$

$$\mathbb{P}\left(a \leq \frac{(n-1)S_n^2}{\sigma^2} \leq b\right) = 0.95$$



$\Updownarrow$

$$\mathbb{P}\left(\frac{(n-1)S_n^2}{b} \leq \sigma^2 \leq \frac{(n-1)S_n^2}{a}\right) = 0.95$$

a is the inverse of the CDF of χ^2_{n-1} at 0.025

b ... 0.975

So the 95% CI of σ^2 is $\left[\frac{(n-1)S_n^2}{b}, \frac{(n-1)S_n^2}{a}\right]$

Special case 2: 95% CI of p in Bernoulli(p)

$X_1, \cdots, X_n \overset{iid}{\sim}$ Bernoulli(p) $\quad E[X_i] = p \quad var[X_i] = p(1-p)$

estimator $\hat{p} = \bar{X}_n$ is both MOM and MLE

by CLT: $\hat{p} \overset{approx}{\sim} N\left(p, \frac{p(1-p)}{n}\right)$

$$\mathbb{P}\left(p - 1.96\sqrt{\frac{p(1-p)}{n}} \leq \hat{p} \leq p + 1.96\sqrt{\frac{p(1-p)}{n}}\right) = 0.95$$

approach 1: $p - 1.96\sqrt{\frac{p(1-p)}{n}} \leq \hat{p} \Rightarrow p \geq$ (only depends on $\hat{p}$)

$p + 1.96\sqrt{\frac{p(1-p)}{n}} \geq \hat{p} \Rightarrow p \leq$ ($\hat{p}$) (tedious)

approach 2 (plug-in): $\mathbb{P}\left(\hat{p} - 1.96\sqrt{\frac{p(1-p)}{n}} \leq p \leq \hat{p} + 1.96\sqrt{\frac{p(1-p)}{n}}\right) = 0.95$

plug-in $\hat{p}$ for p in the boundaries:

$$\mathbb{P}\left(\hat{p} - 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right) \approx 0.95$$

So the 95% CI of p is $\left[\hat{p} - 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$.

special case 3: $X_1, \cdots, X_n \overset{iid}{\sim} N(\mu_X, \sigma_X^2)$. $Y_1, \cdots, Y_m \overset{iid}{\sim} N(\mu_Y, \sigma_Y^2)$

$X_1, \cdots, X_n, Y_1, \cdots, Y_m$ are independent.

Find a 95% CI of $\mu_X - \mu_Y$.

The estimator $\bar{X}_n - \bar{Y}_m \sim N(\mu_X - \mu_Y, \frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m})$

$$\mathbb{P}\left(\mu_X - \mu_Y - 1.96\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}} \le \bar{X}_n - \bar{Y}_n \le \mu_X - \mu_Y + 1.96\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}\right) = 0.95$$

$$\Updownarrow$$

$$\mathbb{P}\left(\bar{X}_n - \bar{Y}_m - 1.96\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}} \le \mu_X - \mu_Y \le \bar{X}_n - \bar{Y}_m + 1.96\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}\right) = 0.95$$

do plug-in: use $\hat{\sigma}_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$, $\hat{\sigma}_Y^2 = \frac{1}{m-1}\sum_{i=1}^{m}(Y_i - \bar{Y}_m)^2$

the 95% CI of $\mu_X - \mu_Y$ is

$$\left[\bar{X}_n - \bar{Y}_m - 1.96\sqrt{\frac{\hat{\sigma}_X^2}{n} + \frac{\hat{\sigma}_Y^2}{m}},\ \bar{X}_n - \bar{Y}_m + 1.96\sqrt{\frac{\hat{\sigma}_X^2}{n} + \frac{\hat{\sigma}_Y^2}{m}}\right]$$

special case 4: Find a 95% CI of $\frac{\sigma_X^2}{\sigma_Y^2}$.

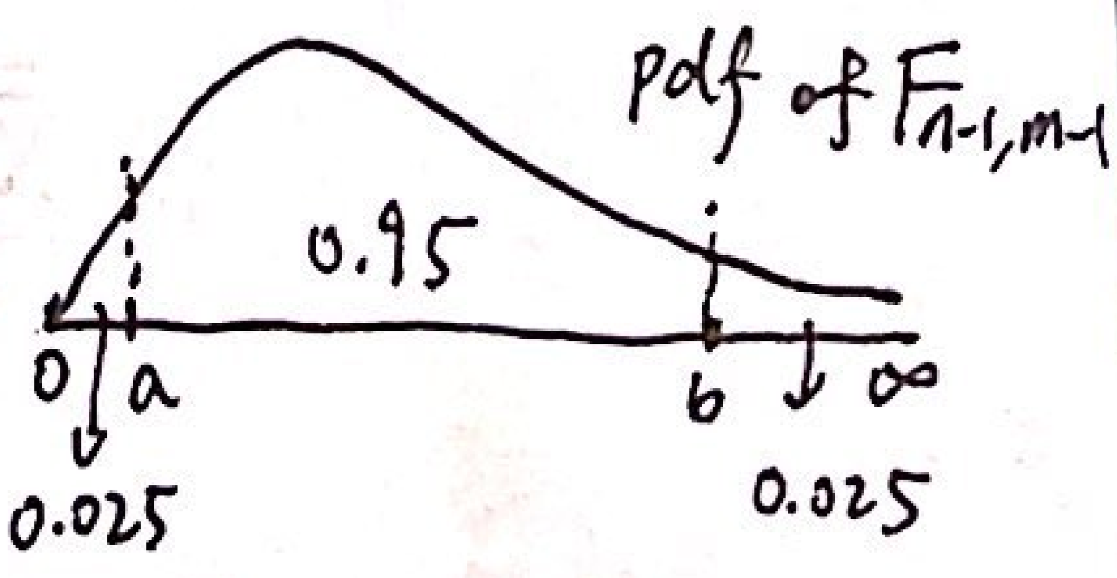
What is the dist of $\frac{\hat{\sigma}_X^2}{\hat{\sigma}_Y^2}$?

Facts: $\frac{(n-1)\hat{\sigma}_X^2}{\sigma_X^2} \sim \chi^2_{n-1}$, $\frac{(m-1)\hat{\sigma}_Y^2}{\sigma_Y^2} \sim \chi^2_{m-1}$

$\hat{\sigma}_X^2$ and $\hat{\sigma}_Y^2$ are independent

$$\Rightarrow \frac{\frac{(n-1)\hat{\sigma}_X^2}{\sigma_X^2} / (n-1)}{\frac{(m-1)\hat{\sigma}_Y^2}{\sigma_Y^2} / (m-1)} \sim F_{n-1,m-1}$$

$$\Rightarrow \frac{\hat{\sigma}_X^2}{\hat{\sigma}_Y^2} \cdot \frac{\sigma_Y^2}{\sigma_X^2} \sim F_{n-1,m-1}$$



so $\mathbb{P}\left(a \le \frac{\hat{\sigma}_X^2}{\hat{\sigma}_Y^2} \cdot \frac{\sigma_Y^2}{\sigma_X^2} \le b\right) = 0.95$

$$\Updownarrow$$

$$\mathbb{P}\left(\frac{\hat{\sigma}_X^2}{\hat{\sigma}_Y^2} \cdot \frac{1}{b} \le \frac{\sigma_X^2}{\sigma_Y^2} \le \frac{\hat{\sigma}_X^2}{\hat{\sigma}_Y^2} \cdot \frac{1}{a}\right) = 0.95$$

so the 95% CI of $\frac{\sigma_X^2}{\sigma_Y^2}$ is $\left[\frac{\hat{\sigma}_X^2}{\hat{\sigma}_Y^2} \cdot \frac{1}{b}, \frac{\hat{\sigma}_X^2}{\hat{\sigma}_Y^2} \cdot \frac{1}{a}\right]$.

Hypothesis Testing $\Longleftrightarrow$ confidence intervals   goal: inference of θ

inference = point estimate + estimation uncertainty

$\hat{\theta}$ bias$(\hat{\theta})$, var$(\hat{\theta})$, etc

Four elements of a test:

1. Null hypothesis (default guess, status-quo): $H_0: \theta = \theta_0$ (eg. $\mu = 0$)
2. Alternative hypothesis (interesting direction): H_1 or H_a: $\theta \neq \theta_0$ (two-sided)

$\left.\begin{array}{l}\theta > \theta_0 \\ \theta < \theta_0\end{array}\right\}$ (one-sided)

3. Test statistic (constructed from $\hat{\theta}$, and it follows a distibution under H_0 that does not depend on unknow parameters)
4. Significance level $\alpha \in (0,1)$. e.g. $\alpha = 0.05$, 0.01

$\Rightarrow$ binary decision $\begin{cases} \text{reject } H_0 \\ \text{do not reject } H_0 \end{cases}$ (smaller α makes rejection more difficult) reject H_0 at a smaller α means greater confidence

using the same $\hat{\theta}$, α-level test rejects $H_0: \theta = \theta_0$

$\Updownarrow$

$(1-\alpha)$ CI does not cover θ_0

Example: $X_1, \cdots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$. $H_0: \mu = \mu_0$ vs. $H_1: \mu \neq \mu_0$

estimator of μ is $\hat{\mu} = \bar{X}_n$

under H_0: $\bar{X}_n \sim N(\mu_0, \frac{\sigma^2}{n})$ — depends on unknown σ^2

$$\boxed{\frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}}} \sim N(0,1)$$

involves unknow σ (cannot be directly used as a test statistic)

$S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$ is an estimator of σ^2

$$\frac{\bar{X}_n - \mu_0}{S_n/\sqrt{n}} \overset{approx}{\sim} N(0,1)$$

we know

test statistic $\boxed{\frac{\bar{X}_n - \mu_0}{S_n/\sqrt{n}}} \overset{exact}{\sim} t_{n-1}$ — doesn't depend on unknown parameters

Define $T = \dfrac{\bar{X}_n - \mu_0}{S_n/\sqrt{n}}$ $\quad$ t statistic

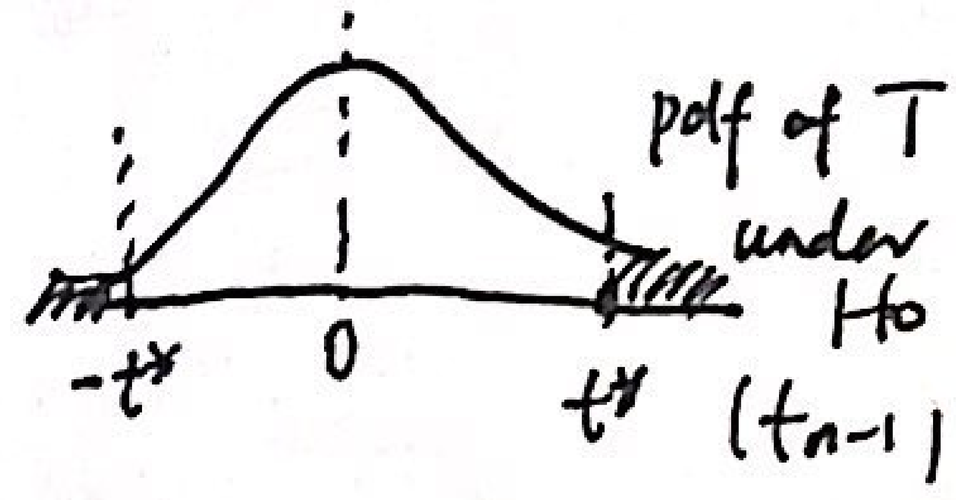
Realization of T on $x_1, \cdots, x_n$: t^* (e.g. $t^* = 3.71$)

Decision : approach 1 : p-value approach : $\left.\begin{array}{l}\text{dist of } T \text{ under } H_0 \\ \text{realization } t^*\end{array}\right\} \Rightarrow$ p-value

p-value is a transformation of t^*

p-value : the prob. that T takes values "more extreme" (determined by H_1) than t^* under H_0

$H_1: \mu \neq \mu_0 \triangleq \mathbb{P}_{H_0}(|T| \geq |t^*|)$



or

$H_1: \mu > \mu_0 \triangleq \mathbb{P}_{H_0}(T \geq t^*)$

or

$H_1: \mu < \mu_0 \triangleq \mathbb{P}_{H_0}(T \leq t^*)$

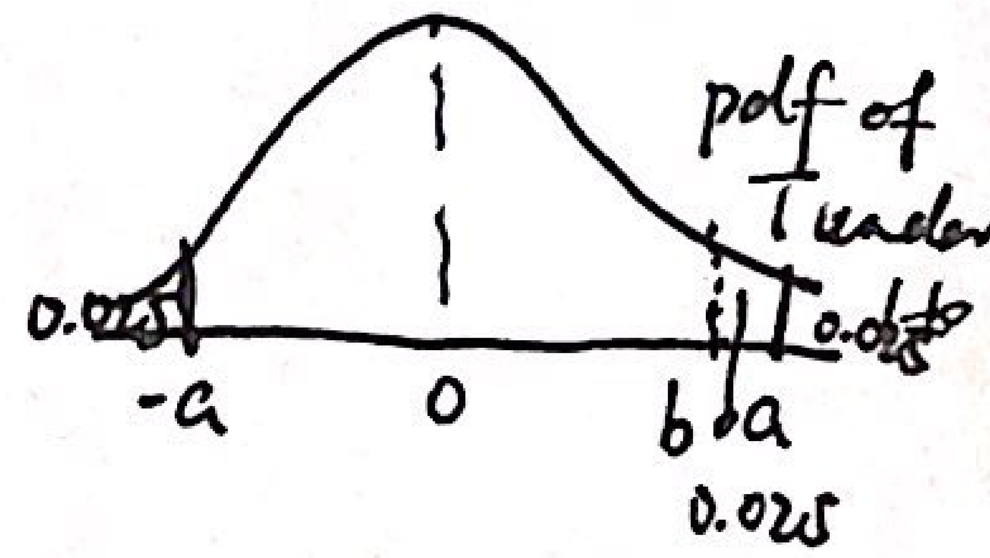
Decision : if p-value $\leq \alpha \Rightarrow$ reject H_0

approach 2 : rejection region based on α and the dist of T under H_0

e.g. $\alpha = 0.05$

$H_1: \mu \neq \mu_0$: rejection region

$\{T \leq -a\} \cup \{T \geq a\}$



or $H_1: \mu > \mu_0$: $\{T \geq b\}$

or $H_1: \mu < \mu_0$: $\{T \leq -b\}$

Decision : if $t^* \in$ rejection region $\Rightarrow$ reject H_0