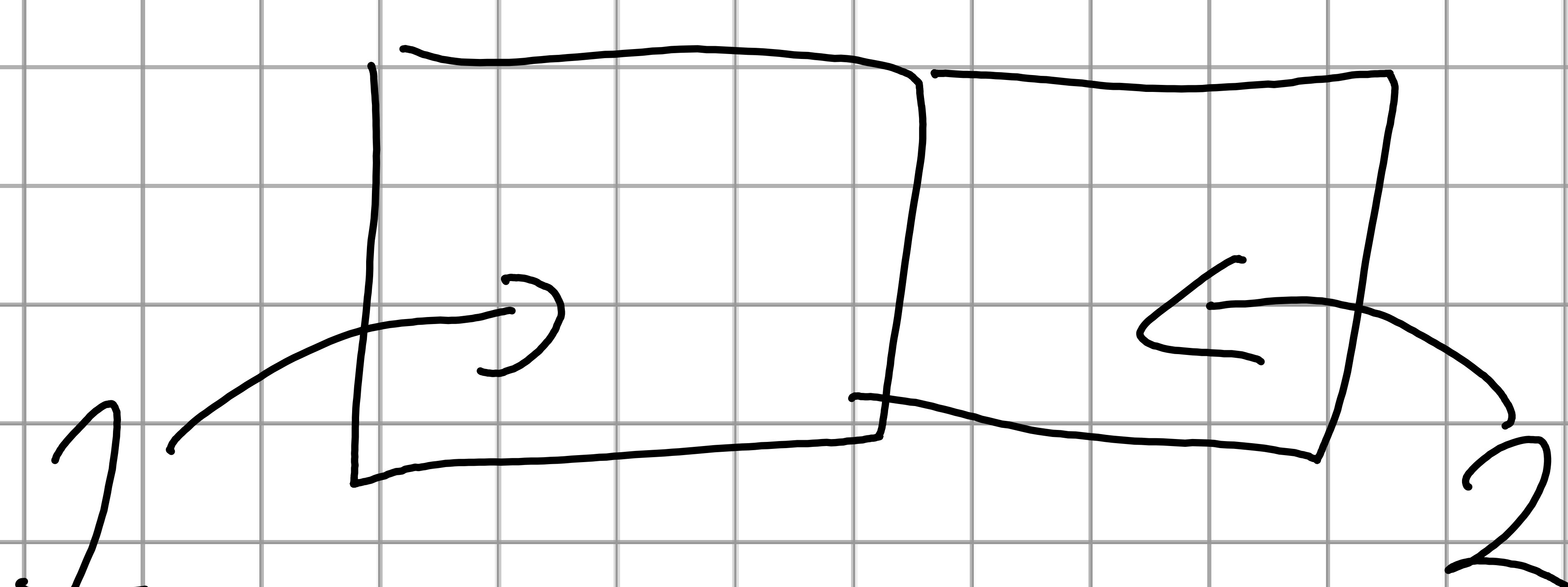
Comparing mean vectors from 2 MVN sample

Paired test

requires each test object split in 2 parts



example: before and after

pairwise observation

$(X_{1j}, X_{2j}), \ j=1,\ldots,n$

$(p \times j) \quad (p \times j)$

$$\left.\begin{array}{l} X_{1j} \sim N_p(\mu_1, \Sigma_1) \\ X_{2j} \sim N_p(\mu_2, \Sigma_2) \end{array}\right\} \text{ Model Check}$$

Difference model

$D_j \overset{def}{=} X_{1j} - X_{2j}$ ← the differene between the 2 vectors

$\sim N_p(\mu_1 - \mu_2, \Sigma_d)$

test if the difference is a value

<u>H_0</u>: $\underset{(p\times 1)}{\delta} \overset{def}{=} \mu_1 - \mu_2 \overset{?}{=} \underset{(p\times 1)}{\delta_0}$ (Often set to $\delta_0 = 0$)

<u>H_1</u>: $\delta \neq \delta_0$ ← not plausable

<u>Estimates</u> $\hat{\mu}_1 - \hat{\mu}_2 = \overset{\text{average}}{\bar{D}} = \frac{1}{n}\sum_{j=1}^{n} D_j \sim$

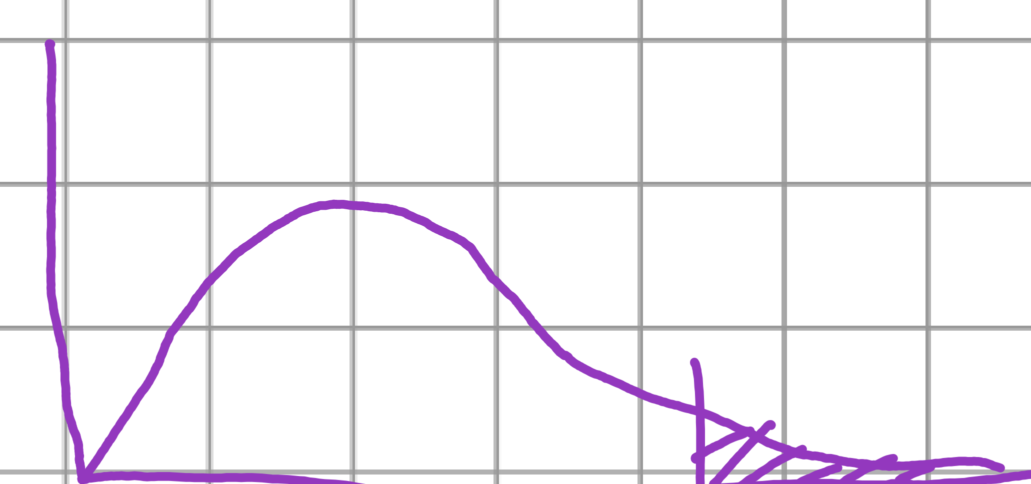
$\sim N_p\left(\mu_1 - \mu_2, \frac{\Sigma_d}{n}\right)$

$$\underset{(p\times p)}{\hat{\Sigma}_d} = S_d = \frac{1}{n-1}\sum_{j=1}^{n}(D_j-\bar{D})(D_j-\bar{D})^T$$

Hotelling T^2 test stat

$$\underset{(1\times1)}{T^2} = \bar{D} - \underbrace{(\mu_1-\mu_2)}_{\text{true mean}}\left(\frac{S_d}{n}\right)^{-1}\underbrace{(\bar{D}-(\mu_1-\mu_2))}\sim \frac{p(n-1)}{n-p}F_{p,n-p}$$

Test



$$T_0^2 = (\bar{D}-\delta_0)^T\left(\frac{S_d}{n}\right)^{-1}(\bar{D}-\delta_0)\sim\frac{p(n-1)}{n-p}F_{p,n-p}$$

finv

$T_0^{2\,\text{data}} = t_0^2$ choose α: reject Ho IFF

$$t_0^2 > \frac{p(n-1)}{n-p}F_{p,n-p,\alpha}$$

$$\text{p-value} = P\left(\frac{p(n-1)}{n-p}F_{p,n-p} > t_0^2\right)$$

fcdf

$100(1-\alpha)\%$ CR for $\mu_1-\mu_2$

$\forall(\mu_1-\mu_2)\in\mathbb{R}^p$: $T^2 \le \frac{p(n-1)}{n-p}F_{p,n-p,\alpha}$

Hyperellipsoid in $\mathbb{R}^p$

Confidence intervals

Bonferroni CI for each variable

Student-t dist

$$\left[\overline{D}_i - t_{n-1,\frac{\alpha}{2p}}\cdot\sqrt{\frac{S_{D,ii}}{n}}\ ,\ \overline{D}_i + t_{n-1,\frac{\alpha}{2p}}\cdot\sqrt{\frac{S_{D,ii}}{n}}\right]$$

variable

$i=1,p$

Non-paired

always possible

Model $X_{1j} \sim N_p(\mu_1, \Sigma_1)$, $j=1,..n_1$

$X_{2j} \sim N_p(\mu_2, \Sigma_2)$, $j=1,..n_2$

model check

$H_0 : \delta \overset{def}{=} \mu_1 - \mu_2 \overset{?}{=} \delta_0$ often $\delta_0 = 0$

$H_1 : \delta \neq \delta_0$

Estimates $\underset{(p\times 1)}{\hat{\mu}_1} = \bar{X}_1 \sim N_p\left(\mu_1, \frac{\Sigma_1}{n_1}\right)$

$\underset{(p\times 1)}{\hat{\mu}_2} = - \; // \; -$

Case 1 Homoscedacity

$\underline{\Sigma_1 = \Sigma_2} = \Sigma$ (test with Bartlett test)

$$S_1 = \frac{1}{n_1 - 1} \sum_{j=1}^{n_1} (X_{1j} - \bar{X}_1)(X_{2j} - \bar{X}_2)^T$$

$$S_2 = \frac{1}{n_2 - 1} \sum_{j=1}^{n_2} (X_{2j} - \bar{X}_2)(X_{2j} - \bar{X}_2)^T$$

if $n_1 = n_2 \rightarrow$ use average

else:

pooled estimate $\rightarrow$ weighted avg

$$\hat{\Sigma} = S_p = \frac{(n_1 - 1) S_1 + (n_2 - 1) S_2}{n_1 + n_2 - 2}$$

↑

doable only because we asume

$\Sigma_1 = \Sigma_2 \rightarrow$ tested with Bartlett

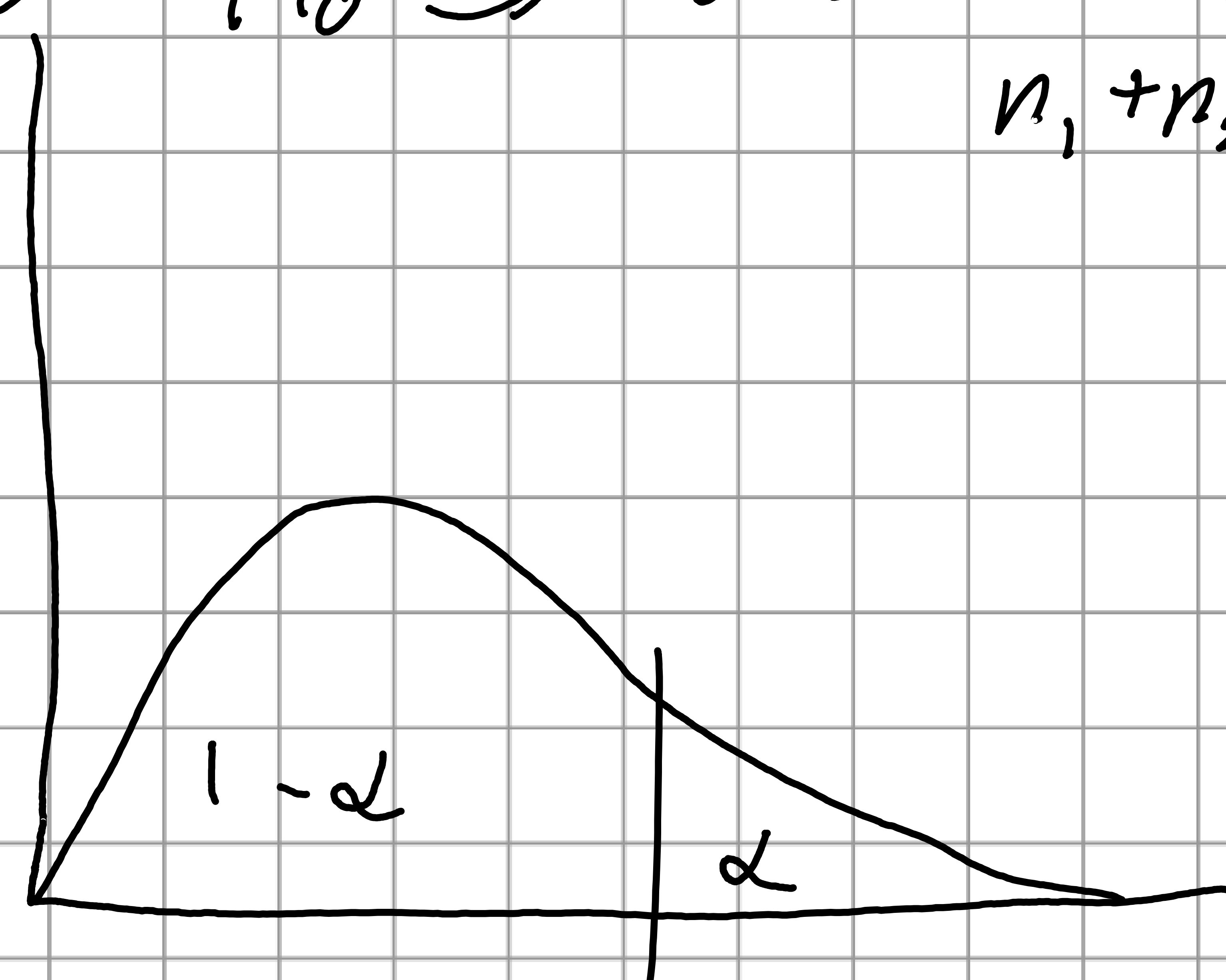
Hotelling T^2 TS

$$T^2 = \underbrace{\left((\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)\right)\left[S_p\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\right]^{-1}\left((\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)\right)}_{\text{test Statistics}} \sim \underbrace{\frac{p(n_1+n_2-2)}{n_1+n_2-p-1}F_{p,n_1+n_2-p-1}}_{\text{distribution}}$$

test $T^2 \overset{\mu_1-\mu_2=\delta_0}{=} T_0^2 \overset{data}{=} t^2$

insert data to
set number t^2

Reject $H_0 \Rightarrow t^2 > \dfrac{p(n_1+n_2-2)}{n_1+n_2-p-1}F_{p,n_1+n_2-p-1,\alpha}$



finv $\longrightarrow F_\alpha$

CR for $\mu_1 - \mu_2$

$$T^2 \leq \frac{p(n_1+n_2-2)}{n_1+n_2-p-1} F_{p,n_1+n_2-p-1,\alpha}$$

Bonferroni CI's

$$\left[(\bar{X}_{1j} - \bar{X}_{2j}) \pm t_{n-1, \alpha/2p} \cdot \sqrt{S_{p_{ii}} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} \right]$$

covar matrix

Non-paired case 2

$\Sigma_1 \neq \Sigma_2$

Large sample approximation

$\hat{\Sigma}_1 = S_1 \,,\ \hat{\Sigma}_2 = S_2$ $\quad n_1, n_2 \gg p$

$$T^2 = \left((\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)\right)\left(\frac{S_1}{n_1} + \frac{S_2}{n_2}\right)^{-1}\left((\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)\right)$$

$$T^2 \underset{\text{approx}}{\sim} \chi^2_p \to \text{test}$$

$\to$ CR

$\to$ CI's

Case 1 or Case 2 ?

Bartlett test (Box's M-test)

test for homoscedacity

Model $X_{1j} \sim N_p(\mu_1, \Sigma_1)$

$X_{2j} \sim N_p(\mu_2, \Sigma_2)$

$H_0: \underset{p\times p}{\Sigma_1} \overset{?}{=} \underset{p\times p}{\Sigma_2}$

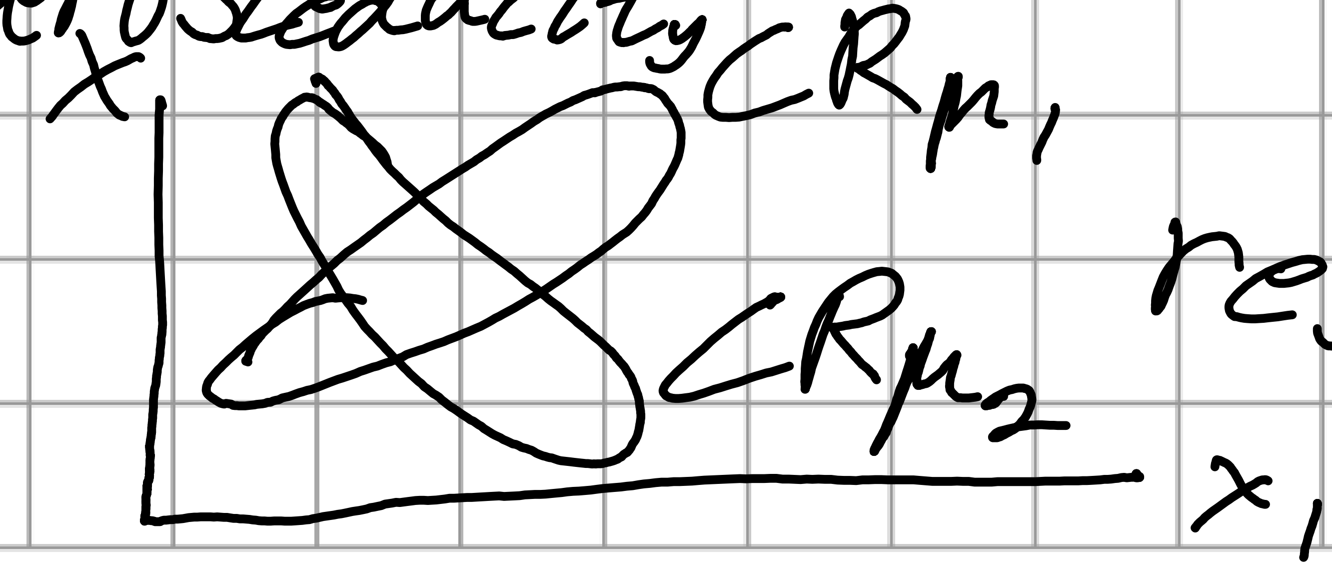
$H_1: \Sigma_1 \neq \Sigma_2$

2D visual

accept H_0

Homoscedacity

Heteroscedacity



reject H_1

Full model, 2 different Σ's:
Likelyhood function
$\Sigma_1 \neq \Sigma_2$

Constrained model

Only 1 common Σ: $\Sigma_1 = \Sigma_2$

Likelyhood function $L_{H_0}(\Sigma)$

Likelyhood ratio $\Lambda \overset{def}{=} \dfrac{\max_{\Sigma} L_{H_0}(\Sigma)}{\max_{\Sigma_1, \Sigma_2} L(\Sigma_1, \Sigma_2)}$

$0 \leq \Lambda \leq 1$

$-2\log(\Lambda) = (n_1+n_2-2)\log|S_p| - (n_1-1)\log|S_1|$

$-($