Unit 2

1. Introduction

Let $X = (X_1, \dots, X_p)'$ be a r.v. with non-singular MND $(\Sigma > 0)$

We are interested on performing inference on the distribution, assuming its parameters being totally or partially ~~[illegible]~~ unknown, based on a simple random sample obtained from it ("observations" given in terms of independent and identically distributed r.v. with same distribution as X)

$$\{X_\alpha : \alpha = 1, \dots, N\}, \quad (N = \text{"sample size"})$$

with $X_{\alpha i}$ being the i-th component of α-th observation.

$$X = \begin{pmatrix} X_1' \\ \vdots \\ X_p' \end{pmatrix} = \begin{pmatrix} X_{11} & \cdots & X_{1p} \\ \vdots & \ddots & \vdots \\ X_{N1} & \cdots & X_{Np} \end{pmatrix}$$

- Sample mean vector: $\bar{X} := \frac{1}{N} \sum_{\alpha=1}^{N} X_\alpha = \begin{pmatrix} \frac{1}{N} \sum_{\alpha=1}^{N} X_{\alpha 1} \\ \vdots \\ \frac{1}{N} \sum_{\alpha=1}^{N} X_{\alpha p} \end{pmatrix} = \begin{pmatrix} \bar{X}_1 \\ \vdots \\ \bar{X}_p \end{pmatrix}$

$\bar{X}$ can be seen as the "centroid" (random) of the sample in $\mathbb{R}^p$.

- Sample dispersion matrix: $A := \sum_{\alpha=1}^{N} (X_\alpha - \bar{X})(X_\alpha - \bar{X})' =$

$$= \begin{pmatrix} \sum_{\alpha=1}^{N} (X_{\alpha 1} - \bar{X}_1)^2 & \sum_{\alpha=1}^{N} (X_{\alpha 1} - \bar{X}_1)(X_{\alpha 2} - \bar{X}_2) & \cdots & \sum_{\alpha=1}^{N} (X_{\alpha 1} - \bar{X}_1)(X_{\alpha p} - \bar{X}_p) \\ \vdots & & \ddots & \\ \sum_{\alpha=1}^{N} (X_{\alpha p} - \bar{X}_p)(X_{\alpha 1} - \bar{X}_1) & \cdots & & \sum_{\alpha=1}^{N} (X_{\alpha p} - \bar{X}_p)^2 \end{pmatrix}$$

- ▸ Sample covariance matrix: $S_N = \frac{A}{N} = (s_{ij})_{i,j=1,\dots,p}$
- ▸ Sample quasi-covariance matrix: $S_{N-1} = \frac{A}{N-1}$
- ▸ Sample correlation matrix: $R = D^{-1/2} S_N D^{-1/2} = \left(\frac{s_{ij}}{s_{ii}^{1/2} s_{jj}^{1/2}} \right)$

  $D = \operatorname{diag}(s_{11}, \dots, s_{pp})$

<u>Lemma</u>: Let $\{X_\alpha : \alpha = 1, \dots, N\}$ be a r.s. of p-dimensional distribution, and let $\bar{X}$ be the sample mean vector. Then, $\forall b \in \mathbb{R}^p$

$$\sum_{\alpha=1}^{N} (X_\alpha - b)(X_\alpha - b)' = \sum_{\alpha=1}^{N} (X_\alpha - \bar{X}) + N(\bar{X} - b)(\bar{X} - b)' = A + N(\bar{X} - b)(\bar{X} - b)'$$

Dem: At first, remember that the sample mean vector is

$$\cdot\ \bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_\alpha \Rightarrow N\bar{x} = \sum_{\alpha=1}^{N} x_\alpha$$

$$\cdot\ A = \sum_{\alpha=1}^{N} (x_\alpha - \bar{x})(x_\alpha - \bar{x})'$$

so, we have that:

$$\sum_{\alpha=1}^{N} (x_\alpha - b)(x_\alpha - b)' =$$

$$= \sum_{\alpha=1}^{N} [(x_\alpha - \bar{x}) + \underbrace{(\bar{x} - b)]}_{\text{"0"}}[(x_\alpha - \bar{x}) + (\bar{x} - b)]' =$$

$$= \sum_{\alpha=1}^{N} [(x_\alpha - \bar{x}) + (\bar{x} - b)][(x_\alpha - \bar{x})' + (\bar{x} - b)'] =$$

$$= \sum_{\alpha=1}^{N} (x_\alpha - \bar{x})(x_\alpha - \bar{x})' + \sum_{\alpha=1}^{N} (\bar{x} - b)(\bar{x} - b)' +$$
① ②

$$+ \sum_{\alpha=1}^{N} (x_\alpha - \bar{x})(\bar{x} - b)' + \sum_{\alpha=1}^{N} (x_\alpha - \bar{x})'(\bar{x} - b)$$
③ ④

① $\sum_{\alpha=1}^{N} (x_\alpha - \bar{x})(x_\alpha - \bar{x})' = A$

② $\sum_{\alpha=1}^{N} (\bar{x} - b)(\bar{x} - b)' = N(\bar{x} - b)(\bar{x} - b)'$

③ $\sum_{\alpha=1}^{N} (x_\alpha - \bar{x})(\bar{x} - b)' = \sum_{\alpha=1}^{N} (x_\alpha - \bar{x})(\bar{x}' - b') =$

$$= \underbrace{\sum_{\alpha=1}^{N} x_\alpha \bar{x}'}_{N\bar{x}} - \sum_{\alpha=1}^{N} x_\alpha b' - \sum_{\alpha=1}^{N} \bar{x}\bar{x}' + \sum_{\alpha=1}^{N} \bar{x} b' =$$

$\downarrow$ $\bar{x}$ and b does not depends of α

$$= \cancel{N\bar{x}\bar{x}'} - \cancel{N\bar{x}b'} - \cancel{N\bar{x}\bar{x}'} + \cancel{N\bar{x}b'} = 0$$

④ Analogus

$$\Rightarrow \sum_{\alpha=1}^{N} (x_\alpha - b)(x_\alpha - b)' = A + N(\bar{x} - b)(\bar{x} - b)' \quad \blacksquare$$

Note: $\cdot\ \bar{x} = \frac{1}{N}\sum_{\alpha=1}^{N} x_\alpha \Rightarrow N\bar{x} = \sum_{\alpha=1}^{N} x_\alpha$

$\cdot\ (x_\alpha - b) = (x_\alpha + \bar{x} - \bar{x} + b) = (x_\alpha - \bar{x})(\bar{x} - b)$

$\cdot$ $\bar{x}$ and b does not depends of α.

2. Maximum - likelihood estimators of μ and Σ in MND (I)

Let $X \sim N_p(\mu, \Sigma)$, $\Sigma > 0$, $\{X_\alpha : \alpha = 1, \dots, N\}$ be a simple random sample of this distribution. The likelihood function is expressed as

$$L(\mu, \Sigma, x_1, \dots, x_N) = f_{\mu,\Sigma}(x_1, \dots, x_N) = \prod_{\alpha=1}^{N} f_{\mu,\Sigma}(x_\alpha) =$$

$$= \prod_{\alpha=1}^{N} \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(x_\alpha - \mu)' \Sigma^{-1} (x_\alpha - \mu)\right\} =$$

$$= \frac{1}{(2\pi)^{pN/2} |\Sigma|^{N/2}} \exp\left\{-\frac{1}{2}\sum_{\alpha=1}^{N} (x_\alpha - \mu)' \Sigma^{-1} (x_\alpha - \mu)\right\}$$

Through algebraic operations and using the generalized multivariate moments formula ($b = \mu$, $\sum_{\alpha=1}^{N} (X_\alpha - \mu)(X_\alpha - \mu)' = A + N(\bar{X} - \mu)(\bar{X} - \mu)'$), we have that:

$$L(\mu, \Sigma) = \frac{1}{(2\pi)^{p\frac{N}{2}} |\Sigma|^{N/2}} \exp\left\{-\frac{1}{2} tr(\Sigma^{-1} \tilde{A}) - \frac{N}{2} (\bar{x} - \mu)' \Sigma^{-1} (\bar{x} - \mu)\right\}$$

with $\tilde{A} = \sum_{\alpha=1}^{N} (x_\alpha - \bar{x})(x_\alpha - \bar{x})'$ (sample dispersion matrix evaluated at the given realization of the sample)

Dem: From the likelihood function expression, we have that:

$$L(\mu, \Sigma) = \frac{1}{(2\pi)^{\frac{Np}{2}} |\Sigma|^{N/2}} \exp\left\{-\frac{1}{2} \sum_{\alpha=1}^{N} (x_\alpha - \mu)' \Sigma^{-1} (x_\alpha - \mu)\right\}$$

We only need to prove that the exponential part is the same:

$$\sum_{\alpha=1}^{N} (x_\alpha - \mu)' \Sigma^{-1} (x_\alpha - \mu) =$$

$$= \sum_{\alpha=1}^{N} [(x_\alpha - \bar{x}) + (\bar{x} - \mu)]' \Sigma^{-1} [(x_\alpha - \bar{x}) + (\bar{x} - \mu)] =$$

$$= \sum_{\alpha=1}^{N} (x_\alpha - \bar{x})' \Sigma^{-1} (x_\alpha - \bar{x}) + \sum_{\alpha=1}^{N} (x_\alpha - \bar{x})' \Sigma^{-1} (\bar{x} - \mu) +$$

$$+ \sum_{\alpha=1}^{N} (\bar{x} - \mu)' \Sigma^{-1} (x_\alpha - \bar{x}) + \sum_{\alpha=1}^{N} (\bar{x} - \mu) \Sigma^{-1} (\bar{x} - \mu) = (*)$$

$$\cdot \sum_{\alpha=1}^{N} (\bar{x} - \mu)' \Sigma^{-1} (x_\alpha - \bar{x}) = \sum_{\alpha=1}^{N} \bar{x}' \Sigma^{-1} x_\alpha - \sum_{\alpha=1}^{N} \mu' \Sigma^{-1} x_\alpha - \sum_{\alpha=1}^{N} \bar{x}' \Sigma^{-1} \bar{x} +$$

$$+ \sum_{\alpha=1}^{N} \mu' \Sigma^{-1} \bar{x} = \cancel{N\bar{x}' \Sigma^{-1} \bar{x}} - \cancel{N\mu' \Sigma^{-1} \bar{x}} - \cancel{N\bar{x}' \Sigma^{-1} \bar{x}} + \cancel{N\mu' \Sigma^{-1} \bar{x}} = 0$$

$$\sum_{\alpha=1}^{N} x_\alpha = N\bar{x}$$

- $\sum_{\alpha=1}^{N}(x_\alpha-\bar{x})'\Sigma^{-1}(\bar{x}-\mu) = \left[\sum_{\alpha=1}^{N}(\Sigma^{-1}(\bar{x}-\mu))'(x_\alpha-\bar{x})\right]' =$

  $\Sigma=\Sigma' \Rightarrow \Sigma^{-1}=(\Sigma^{-1})'$
  Σ symmetric

$$= \sum_{\alpha=1}^{N}\left[(\bar{x}-\mu)'\Sigma^{-1}(x_\alpha-\bar{x})\right]' = 0$$

- $(x_\alpha-\bar{x})'\Sigma^{-1}(x_\alpha-\bar{x})$ is an scalar, so it's equal to its trace $\forall \alpha=1,\dots,N$

$$(*) = \sum_{\alpha=1}^{N} tr\big(\underbrace{(x_\alpha-\bar{x})'}_{A}\underbrace{\Sigma^{-1}(x_\alpha-\bar{x})}_{B}\big) + N(\bar{x}-\mu)'\Sigma^{-1}(\bar{x}-\mu) \underset{tr(AB)=tr(BA)}{=}$$

$$= \sum_{\alpha=1}^{N} tr\left(\Sigma^{-1}(x_\alpha-\bar{x})(x_\alpha-\bar{x})'\right) + N(\bar{x}-\mu)'\Sigma^{-1}(\bar{x}-\mu) =$$

$$= tr\left(\sum_{\alpha=1}^{N}\Sigma^{-1}(x_\alpha-\bar{x})(x_\alpha-\bar{x})'\right) + N(\bar{x}-\mu)'\Sigma^{-1}(\bar{x}-\mu) =$$

$$= tr\big(\Sigma^{-1}\underbrace{\sum_{\alpha=1}^{N}(x_\alpha-\bar{x})(x_\alpha-\bar{x})'}_{\tilde{A}}\big) + N(\bar{x}-\mu)'\Sigma^{-1}(\bar{x}-\mu) \quad \boxtimes$$

<u>Remember</u>:
- Starts with likelihood function.
- $(x_\alpha-\mu) = (x_\alpha-\bar{x})+(\bar{x}-\mu)$ and develope
- $(x_\alpha-\bar{x})'\Sigma^{-1}(x_\alpha-\bar{x})$ is an scalar so its equal to its trace.
- $tr(AB) = tr(BA)$

More conveniently, for maximization purposes, is common to use

$$\ln(L(\mu,\Sigma)) = -\frac{pN}{2}\ln(2\pi) - \frac{N}{2}\ln(|\Sigma|) - \frac{1}{2}tr(\Sigma^{-1}\tilde{A}) - \frac{N}{2}(\bar{x}-\mu)\Sigma^{-1}(\bar{x}-\mu)$$

($-\ln(L(\mu,\Sigma))$ for minimization).

Result: Let $X \sim N_p(\mu, \Sigma)$, $\Sigma > 0$, and let $\{X_\alpha : \alpha = 1, \dots, N\}$ be a simple random sample of the distribution. Then, the maximum-likelihood estimators (MLE) of μ and Σ are, respectivily, $\bar{X}$ and $\frac{A}{N} = S_N$ (under the condition of A be positive definite)

Dem.

- Maximization of $\ln(L(\mu, \Sigma))$ in μ

From the form obtained of $\ln(L(\mu, \Sigma))$, it follows that, regardless the value of Σ, the maximum is reached when the quadratic form $(\bar{x}-\mu)'\Sigma^{-1}(\bar{x}-\mu)$ is minimized. Now, since Σ^{-1} is positive definite ($\Sigma > 0$), the quadratic form reaches the minimum value 0 for $\bar{x} - \mu = 0 \Rightarrow \bar{x} = \mu$. Therefore, the static defined by $\hat{\mu} := \bar{X}$ is the (unique) MLE of μ.

- Maximization of $\ln(L(\mu, \Sigma))$ in Σ:

Before it, we need the watson's lemma.

Watson's lemma: Let $f(G) = -N \ln(|G|) - tr(G^{-1}D)$, with D a given symmetric positive definite matrix, and argument G a symmetric positive definite matrix, both $p \times p$. Then, $\exists_1$ maximum of f with respect to G and it's reached at $G = \frac{1}{N} D$.

$$f\left(\frac{1}{N} D\right) = pN \ln(N) - N \ln(|D|) - pN$$

Dem: Take $D = EE'$, with E non-singular and define $H = E'G^{-1}E \Rightarrow G^{-1} = (E')^{-1} H E^{-1} \Rightarrow$

$$\Rightarrow G = E \underbrace{H^{-1}}_{\text{non-singular}} E' \Rightarrow |G| = |E||H^{-1}||E'| = |H^{-1}||EE'| = \frac{|D|}{|H|}$$

$$tr(G^{-1}D) = tr(\underbrace{G^{-1}}_{A}\underbrace{EE'}_{B}) = tr(E'G^{-1}E) = tr(H)$$

$$tr(AB) = tr(BA)$$

Take $f(G) = g(H) = -N \ln(|D|) + N \ln(|H|) - tr(H)$
$\longmapsto$ using previous results.

Descompose H with cholesky's descomposition : $H = T'T$,

T upper triangular matrix, with positive diagonal elements.

Now, $f(G) = g(H) = h(T) = -N\ln(|D|) + N\ln(|T|^2) - tr(T'T)$

- $|T|^2 = \prod_{i=1}^{p} t_{ii} \Rightarrow \ln(|T|^2) = \sum_{i=1}^{p} \ln(t_{ii})^2 \Rightarrow N\ln(|T|^2) = N\sum_{i=1}^{p} \ln(t_{ii}^2)$ =

$\downarrow$

upper triangular

- $tr(T'T) = \sum_{i=1}^{p} t_{ii}^{2} + \sum_{i<j} t_{ij}$

As a consequence, we have that:

$$h(T) = -N\ln(|D|) + \sum_{i=1}^{p} [N\ln(t_{ii}^2) - t_{ii}^2] - \sum_{i<j} t_{ij}^2$$

The function reaches maximum when:

- $t_{ij} = 0, \ \forall i<j \Rightarrow T$ diagonal
- $\underbrace{t_{ii}^2 = N}$, $\forall i \Rightarrow T = \sqrt{N} I_p$

$\longmapsto$ $y = N\ln x - x$
$y' = \frac{N}{x} - 1 = 0 \Rightarrow x = N$
$y'' = -\frac{N}{x^2} < 0, \ \forall x$
$\Bigg\} \Rightarrow x = N$ maximum

So, we get the maximum at $H = T'T = (\sqrt{N}I_p)'(\sqrt{N}I_p) = N I_p$.

Then, $G = EH^{-1}E' = \frac{1}{N} EE' = \frac{D}{N}$ is the maximum of

f respect G, $f\left(\frac{D}{N}\right) = -N\ln\left(\left|\frac{D}{N}\right|\right) - tr\left(\left(\frac{D}{N}\right)^{-1} D\right) =$

$= pN\ln N - N\ln(|D|) - pN$ ■

Remember: Take $D = EE'$ non singular and $H = E'G^{-1}E$ and make $f(G) = g(H)$. Then, take $H = T'T$, T upper triangular matrix with positive diagonal and $g(H) = h(T)$

• Maximization of $\ln(L(\bar{x},\Sigma))$ in Σ:

We try to maximize the function

$$\ln(L(\bar{x},\Sigma)) = -\frac{pN}{2}\ln(2\pi) - \frac{N}{2}\ln(|\Sigma|) - \frac{1}{2}tr(\Sigma^{-1}\tilde{A})$$

Identify the notation of "Watson's lemma":

$$G := \Sigma \quad , \quad D := \tilde{A}$$

$$f(\Sigma) := 2\left[\ln(L(\bar{x},\Sigma)) + \frac{pN}{2}\ln(2\pi)\right] = pN\ln(2\pi) + 2\ln(L(\bar{x},\Sigma)) =$$

$$= -N\ln(|\Sigma|) - tr(\Sigma^{-1}\tilde{A})$$

Therefore, the maximum of f in Σ is reached for $\Sigma = \frac{\tilde{A}}{N} = \tilde{S}_N$. Then, the statistic defined by $\hat{\Sigma} = \frac{A}{N} = S_N$ is the MLE of Σ.

It is verified that the maximum value reached by the likelihood function at the point $(\mu,\Sigma) = (\bar{x},\tilde{S}_N)$ of the parametric space is

$$L(\mu,\Sigma) = \frac{1}{(2\pi e)^{\frac{pN}{2}}|\tilde{S}_N|^{N/2}} = \left[\frac{N}{(2\pi e)|\tilde{A}|^{1/p}}\right]^{\frac{pN}{2}}$$

(Pendiente de probar)

In relation to the previous result on the MLE of the parametric pair (μ,Σ), the following question arises: Formally, it may happen that for a particular realization of the random sample the matrix $\tilde{A}$ turns out to be singular (positive - semidefinite). With what probability can this case occur?

<u>Theorem (Dykstra)</u>: Let $X \sim N_p(\mu,\Sigma)$ $(\Sigma > 0)$. Let $\{X_\alpha : \alpha = 1,\ldots,N\}$ be a simple random sample of this distribution and let $A = \sum_{\alpha=1}^{N}(X_\alpha - \bar{X})(X_\alpha - \bar{X})'$ be the corresponding sample dispersion matrix. Then, A is positive definite, with probability 1, if and only if $N > p$.

Lemma 1: Let $\{X_\alpha : \alpha = 1, \dots, N\}$ be vectors (random or not) of dimension p. Let $C = (c_{\alpha\beta})_{\alpha,\beta = 1,\dots,N}$ be an orthogonal matrix. Then, defining $Y_\alpha = \sum_{\beta=1}^{N} c_{\alpha\beta} X_\beta$, $\alpha = 1, \dots, N$ we have that.

$$\sum_{\alpha=1}^{N} X_\alpha X_\alpha' = \sum_{\alpha=1}^{N} Y_\alpha Y_\alpha'$$

Dem:

$$\sum_{\alpha=1}^{N} Y_\alpha Y_\alpha' = \sum_{\alpha=1}^{N} \left[\left(\sum_{\beta=1}^{N} c_{\alpha\beta} X_\beta \right) \boxed{\left(\sum_{\gamma=1}^{N} c_{\alpha\gamma} X_\gamma \right)'} \right] =$$

Lo tomas en función de otra variable que la refleja ya traspuesta

$$= \sum_{\alpha=1}^{N} \sum_{\beta=1}^{N} \sum_{\gamma=1}^{N} c_{\alpha\beta} X_\beta c_{\alpha\gamma} X_\beta' = \sum_{\beta=1}^{N} \sum_{\gamma=1}^{N} X_\beta X_\beta' \underbrace{\left(\sum_{\alpha=1}^{N} c_{\alpha\beta} c_{\alpha\gamma} \right)}_{\langle c_{\cdot\beta}, c_{\cdot\gamma} \rangle = \delta_{\beta\gamma}} =$$

$$= \sum_{\beta=1}^{N} X_\beta X_\beta'$$

La devuelves a la variable sin reflejar el valor traspuesto

(the cuadratric sum is 1) ☒

Remember: It's directly.

Lemma 2: Let $\{X_\alpha : \alpha = 1, \dots, N\}$ be random vectors of dimension p, with $X_\alpha \sim N_p(\mu_\alpha, \Sigma)$, $\alpha = 1, \dots, N$ (same covariance, but maybe different mean vector), independents. Let $C = (c_{\alpha\beta})_{\alpha,\beta = 1,\dots,N}$ be an orthogonal matrix. Defining $Y_\alpha = \sum_{\beta=1}^{N} c_{\alpha\beta} X_\beta$, $\alpha = 1, \dots, N$, we have that:

- $Y_\alpha \sim N_p(\nu_\alpha, \Sigma)$ with $\nu_\alpha = \sum_{\beta=1}^{N} c_{\alpha\beta} \mu_\beta$, $\alpha = 1, \dots, N$
- The vectors $\{Y_\alpha : \alpha = 1, \dots, N\}$ are independents.

Dem: $Y_\alpha = \sum_{\beta=1}^{N} c_{\alpha\beta} X_\beta = \sum_{\beta=1}^{N} c_{\alpha\beta} I_p X_\beta = \sum_{\beta=1}^{N} A_{\alpha\beta} X_\beta$ where

$$A_{\alpha\beta} = \begin{pmatrix} c_{\alpha\beta} & & 0 \\ & c_{\alpha\beta} & \\ 0 & & \ddots \; c_{\alpha\beta} \end{pmatrix}$$

and, because the X_β are independents, we have that $Y_\alpha \sim N_p\left(\sum_{\beta=1}^{N} (A_{\alpha\beta} \mu_\beta), \sum_{\beta=1}^{N} (A_{\alpha\beta} \Sigma A_{\alpha\beta}') \right)$ and, because $A_{\alpha\beta}$ diagonal, we get that:

$$Y_\alpha \sim N_p\left(\sum_{\beta=1}^{N} c_{\alpha\beta} \mu_\beta, \sum_{\beta=1}^{N} c_{\alpha\beta}^2 \Sigma \right) = N_p(\nu_\alpha, \Sigma)$$

Because is orthogonal, the ~~[illegible]~~ square sum is 1.

Consider the vector $\underline{Y} = (Y_1, \dots, Y_N)'$ and $(t_1, \dots, t_N)' \in \mathbb{R}^{pN}$ with $t_\alpha \in \mathbb{R}^p$, $\alpha = 1, \dots, N$

$$\phi_Y(t) = E[e^{it'Y}] = E[e^{i\sum_{\alpha=1}^N t_\alpha' Y_\alpha}] = E[e^{i\sum_{\alpha=1}^N t_\alpha'(\sum_{\beta=1}^N c_{\alpha\beta} X_\beta)}] =$$

$Y_\alpha = \sum_{\beta=1}^N c_{\alpha\beta} X_\beta$ $\qquad$ $X_1, \dots, X_n$ independents

$$= E[e^{i\sum_{\beta=1}^N (\sum_{\alpha=1}^N t_\alpha' c_{\alpha\beta}) X_\beta}] = E[\prod_{\beta=1}^N e^{i(\sum_{\alpha=1}^N t_\alpha' c_{\alpha\beta}) X_\beta}] =$$

$X_\alpha \sim N(\mu, \Sigma)$

$$= \prod_{\beta=1}^N E[\exp(i(\sum_{\alpha=1}^N t_\alpha c_{\alpha\beta}')' X_\beta)] = \prod_{\beta=1}^N \phi_{X_\beta}(\sum_{\alpha=1}^N t_\alpha c_{\alpha\beta}') =$$

$$= \prod_{\beta=1}^N \exp\left\{ i(\sum_{\alpha=1}^N t_\alpha c_{\alpha\beta}')' \mu_\beta - \frac{1}{2}\left(\sum_{\alpha=1}^N t_\alpha c_{\alpha\beta}'\right)' \Sigma \left(\sum_{\gamma=1}^N t_\gamma c_{\gamma\beta}'\right)\right\} =$$

$$= \exp\left\{ i\sum_{\beta=1}^N (\sum_{\alpha=1}^N t_\alpha c_{\alpha\beta}')' \mu_\beta - \frac{1}{2}\sum_{\beta=1}^N \left(\sum_{\alpha=1}^N t_\alpha c_{\alpha\beta}'\right)' \Sigma \left(\sum_{\gamma=1}^N t_\gamma c_{\gamma\beta}'\right)\right\} =$$

$$= \exp\left\{ i\sum_{\alpha=1}^N t_\alpha'\left(\sum_{\beta=1}^N c_{\alpha\beta}\mu_\beta\right) - \frac{1}{2}\sum_{\alpha=1}^N \sum_{\gamma=1}^N \left[(t_\alpha' \Sigma t_\gamma) \underbrace{\left(\sum_{\beta=1}^N c_{\alpha\beta} c_{\gamma\beta}\right)}_{\substack{1 \text{ if } \gamma = \alpha \\ 0 \text{ if } \gamma \neq \alpha}}\right]\right\} .$$

$$= e^{i\sum_{\alpha=1}^N (t_\alpha' v_\alpha - \frac{1}{2} t_\alpha' \Sigma t_\alpha)} =$$

$$= \prod_{\alpha=1}^N e^{i t_\alpha' v_\alpha - \frac{1}{2} t_\alpha' \Sigma t_\alpha} = \prod_{\alpha=1}^N \phi_{Y_\alpha}(t_\alpha) \qquad \blacksquare$$

Remember: i) $c_{\alpha\beta} I_p = A_{\alpha\beta}$ and apply linearity

ii) Use the characteristic function and try to calculate. $\underline{Y} = (Y_1, \dots, Y_N)'$, $(t_1, \dots, t_N) \in \mathbb{R}^{pN}$, $t_i \in \mathbb{R}^p$

Let's go back with Dykstra's theorem proof.

Dem: Let us consider, in our case, a matrix $B = (b_{\alpha\beta})_{\alpha,\beta = 1, \dots, N}$ orthogonal, whose last row is $(\frac{1}{\sqrt{N}}, \dots, \frac{1}{\sqrt{N}})$. That's, the rest of the rows must be vectors orthogonal with this one and with each other with norm equal 1. Define

$$Z_\alpha = \sum_{\beta=1}^N b_{\alpha\beta} X_\beta, \quad \alpha = 1, \dots, N$$

From the previous lemmas, we get that:

Si no es el mismo α es 0

- $Z_\alpha \sim N_p(v_\alpha, \Sigma)$

$$v_\alpha = \sum_{\beta=1}^N b_{\alpha\beta}\mu = \left(\sum_{\beta=1}^N b_{\alpha\beta}\right)\mu = \begin{cases} 0 & \text{if } \alpha = 1, \dots, N-1 \quad (*) \\ \frac{N}{\sqrt{N}}\mu = \sqrt{N}\mu & \text{if } \alpha = N \end{cases}$$

Si es el mismo es sumar la última fila que es $(\frac{1}{\sqrt{N}} \cdots \frac{1}{\sqrt{N}})$

(*) Because rows of B are orthogonal, $\langle b_\alpha, b_N \rangle = 0$ $\forall \alpha = 1, \dots, N$, then $\langle b_\alpha, b_N \rangle = \sum_{\beta=1}^N b_{\alpha\beta} \frac{1}{\sqrt{N}} = 0 \Rightarrow \sum_{\beta=1}^N b_{\alpha\beta} = 0$

- The vectors are independents ($\{Z_\alpha : \alpha = 1, \dots, N\}$ independents)

- $$Z_N = \sum_{\beta=1}^{N} \frac{1}{\sqrt{N}} X_\beta = \sqrt{N}\left(\frac{1}{N}\sum_{\beta=1}^{N} X_\beta\right) = \sqrt{N}\,\bar{X} \Rightarrow \bar{X} = \frac{1}{\sqrt{N}} Z_N$$

rationalazing

- $$A = \sum_{\alpha=1}^{N} (X_\alpha - \bar{X})(X_\alpha - \bar{X})' = \sum_{\alpha=1}^{N} X_\alpha X_\alpha' - \sum_{\alpha=1}^{N} X_\alpha \bar{X}' - \sum_{\alpha=1}^{N} \bar{X} X_\alpha + \sum_{\alpha=1}^{N} \bar{X}\bar{X}' =$$

$$= \sum_{\alpha=1}^{N} X_\alpha X_\alpha' - N\bar{X}\bar{X}' - \cancel{N\bar{X}\bar{X}'} + \cancel{N\bar{X}\bar{X}'} = \sum_{\alpha=1}^{N} X_\alpha X_\alpha' - N\bar{X}\bar{X}' =$$

$$= \sum_{\alpha=1}^{N} Z_\alpha Z_\alpha' - \cancel{N} \cdot \frac{1}{\cancel{\sqrt{N}}} Z_N \cdot \frac{1}{\cancel{\sqrt{N}}} Z_N' = \sum_{\alpha=1}^{N-1} Z_\alpha Z_\alpha'$$

(Lemma 2)

$$Z_\alpha = \sum_{\beta=1}^{N} b_{\alpha\beta} X_\beta$$

$$Z_\alpha' = \sum_{\beta=1}^{N} (b_{\alpha\beta} X_\beta)' \quad \Big\} \text{(product)} = \sum_{\beta=1}^{N} \underbrace{b_{\alpha\beta}^2}_{1 \text{ because } B \text{ orthogonal}} X_\beta X_\beta'$$

Because A only depends on Z_α, $\alpha = 1, \dots, N-1$ and $\bar{X}$ only depends on Z_N and, because of Z_α are independents, we get that A and $\bar{X}$ are independents.

Finally, let us note that we can write $A = Z'Z$, with $Z = \begin{pmatrix} Z_1' \\ \vdots \\ Z_{N-1}' \end{pmatrix}$ a $(N-1)\times p$ matrix with Z' a $p \times (N-1)$ matrix.

This is equivalent to the matrix A being non-negative definite. On the other hand, it holds that $\text{rank}(Z) = \text{rank}(Z') = \text{rank}(A)$. So we only need to prove that $\text{rank}(Z) = p$ with probability 1, if and only if, $N > 1$.

Remember up to this point: $B = (b_{\alpha\beta})$ $\alpha,\beta = 1, \dots, N$ orthogonal and the last row is $\left(\frac{1}{\sqrt{N}}, \dots, \frac{1}{\sqrt{N}}\right)$. Let $Z_\alpha = \sum_{\beta=1}^{N} b_{\alpha\beta} X_\beta$. Apply lemma 1 and 2 (Remember that $\langle b_\alpha, b_N \rangle = \sum_{\beta=1}^{N} b_{\alpha\beta} \frac{1}{\sqrt{N}} = 0 \Rightarrow \sum_{\beta=1}^{N} b_{\alpha\beta} = 0$) if $\alpha \neq N$

Calculate A and $\bar{X}$ and see they are independents. You can write $A = Z'Z$ (that's A non-negative definite) and it holds $\text{rank}(Z) = \text{rank}(Z') = \text{rank}(A)$.

We have to prove that: $\operatorname{rank}(Z) = p$ with probability $1 \iff p < N$

$\Rightarrow]$ if $N \leq p \Rightarrow N-1 < p \Rightarrow \operatorname{rank}(Z) \leq N-1 < p$!!

$\Leftarrow]$ suppose $p < N$ and suppose $N = p+1 \Rightarrow N-1 = p$. In this case, Z' is of dimension $p \times (N-1) = p \times p$, $Z' = (z_1, \dots, z_{\underset{N-1}{\overset{\shortparallel}{p}}})$ with $z_i \sim N_p(0, \Sigma)$, $\Sigma > 0$, independents.

Let's consider $P_{z_1}, \dots, P_{z_p}$ the probability matrix of the joint vector $(z_1' \dots z_p')$

$$P[\operatorname{rank}(Z) < p] = P[z_1, \dots, z_p \text{ linearly - dependents}] \leq$$

$$\leq \sum_i P[z_i \in \underset{\substack{\uparrow \\ \text{linearly envolment}}}{S}(z_1, \dots, z_{i-1}, z_{i+1}, \dots, z_p)]$$

Because $z_1, \dots, z_p$ are identically distributed and independents (prob. way),

$P[z_i \in S(z_1, \dots, z_{i-1}, z_{i+1}, \dots, z_p)] = \overset{\text{"same value"}}{\varphi}$, $\forall i = 1, \dots, p$, therefore,

$$P[\operatorname{rank}(Z) < p] \leq p \cdot P[z_1 \in S(z_2, \dots, z_p)]$$

$$P[z_1 \in S(z_2, \dots, z_p)] = E_p[\underset{\substack{\downarrow \\ \text{c.f.}}}{I}_{[z_1 \in S(z_2 \dots z_p)]}] =$$

$$\overset{z_i \text{ independents}}{=} \int_{\underset{(n)}{\mathbb{R}^p \times \dots \times \mathbb{R}^p}} I_{[z_1 \in S(z_2 \dots z_p)]}(z_1, \dots z_p) P_{z_1, \dots, z_p}(dz_1 \dots dz_p) =$$

$$= \int_{\underset{(n-1)}{\mathbb{R}^p \times \dots \times \mathbb{R}^p}} \left(\int_{\mathbb{R}^p} I_{[z_1 \in S(z_2 \dots z_p)]}(z_1, \dots, z_p) P_{z_1}(dz_1) \right) P_{z_2, \dots, z_p}(dz_2 \dots dz_p) =$$

$$= \int_{\underset{(n-1)}{\mathbb{R}^p \times \dots \mathbb{R}^p}} \underbrace{P_z [z_1 \in S(z_2 \dots, z_p)]}_{0} P_{z_2 \dots z_p}(dz_2 \dots dz_p) = 0$$

0 because z is a continuous distribution and $S(z_1 \dots z_p)$ is an hiperplane with null - lebesgue - measure.

Then, $P[\operatorname{rank}(Z) < p] = 0 \Rightarrow P[\operatorname{rank}(Z) = p] = 1$.

Suppose now true for $N \geq p+1$. Let's see it holds for $N+1$.

Let Z^N be the matrix based on the sample $x_1, \dots, x_N$ and Z^{N+1} the matrix based on the datas $x_1, \dots, x_N, x_{N+1}$ with x_{N+1} added independently from the rest.

Because Z^N has been built through an orthogonal matrix B^N with b_α independents vectors and orthogonal each others,

Define B^{N+1} for Z^{N+1} as: $B^{N+1} = \left(\begin{array}{ccc|c} b_{11} & \cdots & b_{1N} & 0 \\ \vdots & \ddots & \vdots & \vdots \\ b_{(N-1)1} & \cdots & b_{(N-1)N} & 0 \\ \hline b_{N1} & \cdots & b_{NN} & b_{N(N+1)} \\ \frac{1}{\sqrt{N+1}} & \cdots & \frac{1}{\sqrt{N+1}} & \frac{1}{\sqrt{N+1}} \end{array}\right)$

independent and orthogonal vector

Therefore, $Z^{n+1} = \begin{pmatrix} Z^N \\ z_{N+1} \end{pmatrix}$ hypothesis matrix, last row added

So, we get that $P[\operatorname{rank}(Z^n) = p] = 1 \implies$

$$\implies P[\operatorname{rank}(Z^{n+1}) = p] = 1.$$

$\boxtimes$

Remember:

Theorem (Fisher): (From the previous proof, we also get the proof for this theorem). Given a simple random sample $x_1, \dots, x_N$ of distribution $N_p(\mu, \Sigma)$, the sample mean vector $\bar{x}$ is distributed according to a $N_p(\mu, \frac{\Sigma}{N})$, and the sample dispersion matrix is the same way as $\sum_{\alpha=1}^{N-1} z_\alpha z_\alpha'$ with z_α independent and identically distributed to $N_p(0, \Sigma)$, $\alpha = 1, \dots, N-1$, with the vector $\bar{x}$ and the matrix A being independent.

$$z_N \sim N_p(\sqrt{N}\mu, \Sigma) \xrightarrow{1/\sqrt{N}} N_p(\mu, \frac{\Sigma}{N})$$

3. Maximum - likelihood estimators of μ and Σ in a MND (II)

Def: Let $\{X_\alpha : \alpha = 1, \dots, N\}$ be a random sample of a distribution depending on a parameter $\theta = (\theta_1, \dots, \theta_k)' \in \Theta \subseteq \mathbb{R}^k$. An estimator $\hat{\theta}$ is a measurable function of the sample, $\hat{\theta} = T(X)$. It is said to be unbiased if $E_\theta[\hat{\theta} - \theta] = 0, \forall \theta \in \Theta$, i.e., $E_\theta[\hat{\theta}] = \theta$.

Let $X \sim N_p(\mu, \Sigma)$, $\Sigma > 0$.

- Sample mean vector: $\hat{\mu} := \bar{X}$ is an unbiased estimator of μ.

Dem: $\bar{X} = \frac{1}{N} \sum_{\alpha=1}^{N} X_\alpha \Rightarrow E[\bar{X}] = E\left[\frac{1}{N} \sum_{\alpha=1}^{N} X_\alpha\right] \overset{\text{linearity}}{=}$

$= \frac{1}{N} \sum_{\alpha=1}^{N} E[X_\alpha] = \frac{1}{N} \sum_{\alpha=1}^{N} \mu = \mu$ ∎

$\rightarrow E[X_\alpha] = \mu$, because $X \sim N(\mu, \Sigma)$ and X_α is a random sample of X, independents and identically distributed.

Remember: $E[\bar{X}] = \dots \underset{\text{linearity}}{=} \dots \underset{X_\alpha \sim N_p(\mu, \Sigma)}{=} \dots = \mu$

- Sample covariance matrix: $\hat{\Sigma} := S_N$ is not an unbiased estimator of Σ.

Dem: $S_N = \frac{A}{N} \overset{\text{Th. Fisher}}{=} \frac{1}{N} \sum_{\alpha=1}^{N-1} Z_\alpha Z_\alpha'$ with $Z_\alpha \sim N_p(0, \Sigma)$ independents

Therefore,

$E[S_N] = E\left[\frac{1}{N} \sum_{\alpha=1}^{N-1} Z_\alpha Z_\alpha'\right] \overset{\text{linearity}}{=} \frac{1}{N} \sum_{\alpha=1}^{N-1} E[Z_\alpha Z_\alpha'] =$

$= \frac{1}{N} \sum_{\alpha=1}^{N-1} \Sigma = \frac{(N-1)}{N} \Sigma \neq \Sigma \Rightarrow$ Not unbiased.

$\rightarrow Z_\alpha \sim N_p(0, \Sigma)$
$E[(Z_\alpha - 0)(Z_\alpha - 0)'] = Var(Z_\alpha) = \Sigma = E[Z_\alpha Z_\alpha']$ ∎

- $\hat{S}_{N-1} := \frac{N}{N-1} S_N$ is an ~~unbiase~~ unbiased estimator of Σ

Dem: The same as before.

Remember: Th. Fisher $\Rightarrow A = \sum_{\alpha=1}^{N-1} Z_\alpha Z_\alpha'$, $Z_\alpha \sim N_p(0, \Sigma)$ independents and follow.

Def: Let $\{\mathbf{X}_\alpha : \alpha = 1, \dots, N\}$ be a simple random sample of a distribution depending on a parameter $\theta = (\theta_1, \dots, \theta_K)' \in \Theta \subseteq \mathbb{R}^K$. An estimator $\hat{\theta}_N$ ($\hat{\theta}_N = T_N(\mathbf{X})$) of the parameter θ is said to be:

- weakly consistent if $\hat{\theta}_N$ converges in probability to θ, i.e.,
$$\forall \varepsilon > 0, \lim_{N \to \infty} P_\theta [\| \hat{\theta}_N - \theta \| < \varepsilon] = 1$$
- strongly consistent if $\hat{\theta}_N$ converges almost surely to θ, i.e.,
$$P_\theta [\lim_{N \to \infty} \hat{\theta}_N = \theta] = 1$$

Let $\mathbf{X} \sim N_p(\mu, \Sigma)$, $\Sigma > 0$. It's proved that $\hat{\mu} := \bar{\mathbf{X}}$, $\hat{\Sigma} := S_N$ and $\hat{\Sigma} := S_{N-1}$ are strongly consistent estimators of μ, Σ, Σ, respectivily.

Def: Let $T = (T_1, \dots, T_K)'$ be an unbiased estimator of the parameter $\theta = (\theta_1, \dots, \theta_K)' \in \Theta$, with $\Theta = \mathbb{R}^K$ or a "rectangle" in $\mathbb{R}^K$. T is said to be efficient for $\theta = (\theta_1, \dots, \theta_K)'$ if, for any other unbiased estimator U of θ, it holds that the difference between the covariance matrices $Cov_\theta(U) - Cov_\theta(T)$ ($\forall \theta \in \Theta$) is a non-negative definite matrix ($Cov_\theta(U) \geq Cov_\theta(T)$)

Let $\mathbf{X} \sim N_p(\mu, \Sigma)$, $\Sigma > 0$. It's proved that:

- $\to$ $\bar{\mathbf{X}}$ is an efficient estimator of μ in the space $\mathbb{R}^p$
- $\to$ S_{N-1} is an efficient estimator of Σ in the space of symmetric positive-definite matrices of dim $p \times p$.

Theorem (Zehna): Let $P = \{P_\theta : \theta \in \Theta\}$ be a family of probability distribution on the space $(\mathbb{R}^p, \mathcal{B}^p)$. Let $g: \Theta \to \Lambda$ be a given arbitrary function. If $\hat{\Theta}$ is a maximum-likelihood estimator of Θ $\Rightarrow$
$\Rightarrow g(\hat{\theta})$ is a maximum-likelihood estimator of $g(\theta)$.

Remark: This theorem makes use of the concept of "induced likelihood function" by g. For each $\lambda \in g(\Theta) \subseteq \Lambda$, we define
$$M(\lambda) = \sup_{\theta \in \Theta_\lambda} L(\theta), \text{ with } \Theta_\lambda = \{\theta \in \Theta : g(\theta) = \lambda\}$$
It's then understood that $g(\hat{\theta})$ is a MLE of $g(\theta)$ in the sense that $g(\hat{\theta})$ is a maximum for M.

Now, from Fisher's theorem we obtained that for a simple random sample $X_1, \dots, X_N$ of a distribution $N_p(\mu, \Sigma)$ we had that $\bar{X} \sim N_p\left(\mu, \frac{\Sigma}{N}\right)$, t.i, normalizing in mean and sample size:

$$N^{1/2}(\bar{X}-\mu) = N^{-1/2}\sum_{\alpha=1}^{N}(X_\alpha - \mu) \sim N_p(0, \Sigma)$$

Now, the limiting distribution no longer depends on N.

- Asymptotic distribution of the sample mean vector, $\bar{X}$

Result: Let $X_1, \dots, X_N \dots$ be a sequence of p-dimensional independent and identically distributed random vectors, with mean vector μ and covariance matrix Σ (i.e. $X_\alpha \sim (\mu, \Sigma)$)

Let $\bar{X}_N = \frac{1}{N}\sum_{\alpha=1}^{N} X_\alpha$, $\forall N \geq 1$. Then, when $N \to \infty$, we have the asymptotic distribution

$$N^{1/2}(\bar{X}_N - \mu) = N^{-1/2}\sum_{\alpha=1}^{N}(X_\alpha - \mu) \underset{N\to\infty}{\sim} N_p(0, \Sigma)$$

- Asymptotic distribution for the sample dispersion matrix, A

Result: Let $X_1, \dots, X_N \dots$ be a sequence of p-dimensional independent and identically distributed random vectors, with mean vector μ and covariance matrix Σ (i.e. $X_\alpha \sim (\mu, \Sigma)$, $\alpha = 1, \dots, N, \dots$) and with finite fourth-order moments.

Let $\bar{X}_N = \frac{1}{N}\sum_{\alpha=1}^{N} X_\alpha$ and $A_N = \sum_{\alpha=1}^{N}(X_\alpha - \bar{X}_N)(X_\alpha - \bar{X}_N)'$, $\forall N \geq 1$.

Then, when $N \to \infty$, we have the asymptotic distribution

$$N^{-1/2}(A_N - N\Sigma) \underset{N\to\infty}{\sim} N_{p^2}(0, V)$$ ↳ necessary singular

(in the sense that $N^{-1/2}(\mathrm{Vec}(A_N) - N\,\mathrm{Vec}(\Sigma)) \underset{N\to\infty}{\sim} N_p(0, V)$)

with $V = \mathrm{Cov}(\mathrm{Vec}((X_\alpha - \mu)(X_\alpha - \mu)'))$.

Dem asymptotic distribution of the sample mean vector, $\bar{X}$:

Define $Y_N = N^{-1/2} \sum_{\alpha=1}^{N} (X_\alpha - \mu)$. We need to prove that

$\phi_{Y_N}(t)$ converges to $\exp\{-\frac{1}{2} t' \Sigma t\}$, $\forall t \in \mathbb{R}^p$ (converges to the c.f. of MND but with $\mu = 0$ because $\bar{X} - \mu$ is "normalized")

We know that for each $t \in \mathbb{R}^p$: $\phi_{Y_N}(t) = \phi_{t'Y_N}(1)$,

with $t'Y_N = N^{-1/2} \sum_{\alpha=1}^{N} (t'X_\alpha - t'\mu)$ with $t'X_\alpha - t'\mu$

following a $(0, t'\Sigma t)$, independents between them.
↳ distribution of parameters:

By the theorem of the central limit: $t'Y_N \underset{N\to\infty}{\sim} N_1(0, t'\Sigma t)$

By Th. continuity of c.f. in univariate case, we get that $\forall s \in \mathbb{R}$:

$$\phi_{t'Y_N}(s) \xrightarrow[N\to\infty]{} \exp\{-\frac{1}{2} s^2 (t' \Sigma t)\}$$

Take $s = 1$:

$$\phi_{t'Y_N}(1) \xrightarrow[N\to\infty]{} \exp\{-\frac{1}{2} t' \Sigma t\}$$

$$\parallel$$

$$\phi_{Y_N}(t) \qquad \boxtimes$$

Theorem of central limit: " The sampling distribution of the mean will always be normally distributed as long as the sample size is large enough". $(N \to \infty, \bar{X} \sim N)$

Th. continuity of c.f.: If $\varphi_n(t) \xrightarrow[n\to\infty]{} \varphi(t), \forall t \in \mathbb{R} \Rightarrow$

$\Rightarrow X_n \xrightarrow{D} X$

4. Wishart distribution

Def: Let $Z_1, \dots, Z_n$ be independent and identically distributed random vectors, according to a $N_p(0, \Sigma)$ $(\Sigma \geq 0)$, with $n \geq p$!!. We define the Wishart* centered distribution with n degrees of freedom as the distribution of the random matrix $\sum_{\alpha=1}^{N} Z_\alpha Z_\alpha'$ and it's denoted $W_p(n, \Sigma)$

As a consequence, in the case of the dispersion matrix derived from simple random sample of a population $N_p(\mu, \Sigma)$ $(\Sigma > 0)$ (it can be extended to $\Sigma \geq 0$), we will have that:

$$A = N S_N = (N-1) S_{N-1} \sim W_p(n, \Sigma), \text{ with } n = N-1$$

Result: If $a \sim W_1(n, \sigma^2) \Rightarrow \frac{a}{\sigma^2} \sim \chi_n^2$.

Dem: $$f(a) = \frac{|a|^{\frac{n-1-1}{2}} \exp\{-\frac{1}{2} tr(\sigma^{-2} \cdot a)\}}{2^{\frac{n \cdot 1}{2}} |\sigma^2|^{n/2} \prod_{i=1}^{1} \Gamma\left(\frac{n+1-i}{2}\right)} =$$

$$= \frac{a^{\frac{n-2}{2}} e^{-\frac{a}{2\sigma^2}}}{2^{n/2} (\sigma^2)^{n/2} \Gamma\left(\frac{n}{2}\right)} = \begin{bmatrix} v = \frac{a}{\sigma^2} \\ dv = \frac{1}{\sigma^2} da \end{bmatrix} =$$

$$= f(v) = \frac{(\sigma^2 v)^{\frac{n-2}{2}} e^{-v/2}}{2^{n/2} (\sigma^2)^{n/2} \Gamma\left(\frac{n}{2}\right)} \boxed{\cdot \sigma^2} \text{ !!} =$$

$$= \frac{\cancel{(\sigma^2)^{\frac{n}{2}-1}} \cdot \cancel{\sigma^2} \, e^{-v/2} \cdot (v)^{\frac{n-2}{2}}}{2^{n/2} \cancel{(\sigma^2)^{n/2}} \Gamma\left(\frac{n}{2}\right)}, v > 0 \Rightarrow v \sim \chi_n^2 \Rightarrow \frac{a}{\sigma^2} \sim \chi_n^2 \quad \blacksquare$$

Remember: You don't need to study the formulas, so just substitute and apply $v = \frac{a}{\sigma^2}$ as change.

Remark: In relation with the previous result, if $\hat{\sigma}^2 = S_N^2 = \frac{1}{N} \sum_{\alpha=1}^{N} (x_\alpha - \bar{x})^2$ is the MLE of σ^2 in the one-dimensional case $(p = 1)$, then we have that

$$\frac{N S_N^2}{\sigma^2} = \frac{\sum_{\alpha=1}^{N} (x_\alpha - \bar{x})^2}{\sigma^2} \sim \chi_n^2, \text{ with } n = N-1$$

Firstly, let us see how the c.f. of a random matrix is defined in a general way:

~~Let~~

Let Y be a $(r\times s)$-dimensional matrix. The c.f. is defined as $\phi_Y(\Theta) = E\left[e^{i\,tr(\Theta' Y)}\right]$, $\forall\, \Theta$ in the space of matrices of dim $r\times s$.

This expression comes from the vectorization of Y and Θ

$$\underline{\phi_{vec(Y)}(vec(\Theta))} = E\left[e^{i\,vec(\Theta)' vec(Y)}\right] =$$

$$= E\left[e^{i\sum_{\ell=1}^{s}\sum_{k=1}^{r}\theta_{k\ell}\, Y_{k\ell}}\right] = \underline{E\left[e^{i\,tr(\Theta' Y)}\right]}$$

Now, in the case of a symmetric $(r\times r)$-dimensional matrix Y, the formulation can be restricted to the set of the $\frac{1}{2}r(r+1)$ variables corresponding to the diagonal and the upper (or lower) triangular part of Y, adopting the form

$$\phi^*_Y(\Theta) = E\left[e^{i\sum_{\ell=1}^{s}\sum_{\substack{k=1\\ k\le \ell}}^{r}\theta_{k\ell} Y_{k\ell}}\right],\ \forall\,\Theta \text{ in the space of symmetric matrices dim } r\times r$$

"Esto no entraba"

Result: Let $A\sim W_p(n,\Sigma)$. Then, $E[A] = n\Sigma$

$$Cov(a_{ij}, a_{kl}) = n(\sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk}),\quad (i,j,k,l = 1,\dots,p)$$

Result (Reproductive property with respect to degrees of freedom):

Let $A_1,\dots,A_q$ be independent random matrices, with $A_j\sim W_p(n_j,\Sigma)$, $j=1,\dots,q$. Then:

$$A = \sum_{j=1}^{q} A_j \sim W_p(n,\Sigma),\quad n=\sum_{j=1}^{q} n_j$$

Dem: $\phi_A(\Theta) = \phi_{\sum_{j=1}^q A_j}(\Theta) = E\left[e^{\underline{i\,tr(\Theta'\sum_{j=1}^q A_j)}}\right] = E\left[e^{i\sum_{j=1}^q tr(-)}\right] = \prod E[-]$ (*) A_j independents

(*) $\Rightarrow \phi_A(\Theta) = \prod_{j=1}^{q}\phi_{A_j}(\Theta) = \prod_{j=1}^{q}|I_p - 2i\,\Theta\Sigma|^{-n_j/2} = |I_p - 2i\,\Theta\Sigma|^{-\frac{1}{2}\sum_{j=1}^q n_j} =$

~~independents~~

$= |I_p - 2i\,\Theta\Sigma|^{-\frac{n}{2}} \Rightarrow A\sim W_p(n,\Sigma)$ ∎

Remember: You don't need to know ϕ_{A_j}, so just apply independence of A_j and calculate.

Result (Reproductive property under rectangular linear transformations)

Let $A \sim W_p(n, \Sigma)$ and let $M_{K\times p}$ be a non-random matrix of rank $K \leq p$. Then:

$$MAM' \sim W_K(n, M\Sigma M')$$

Dem: $A \sim W_p(n,\Sigma) \overset{def}{\Longrightarrow} A = \sum_{\alpha=1}^{n} Z_\alpha Z_\alpha'$, with $Z_\alpha \sim N_p(0,\Sigma)$ independents.

Take $Y_\alpha := MZ_\alpha \Rightarrow Z_\alpha \sim N_p(0, M\Sigma M')$ independents

Then, we have that:

$$\sum_{\alpha=1}^{n} Y_\alpha Y_\alpha' = \sum_{\alpha=1}^{n} (MZ_\alpha)(MZ_\alpha)' = \sum_{\alpha=1}^{n} (MZ_\alpha)(Z_\alpha' M') =$$

$$= M\left(\sum_{\alpha=1}^{n} Z_\alpha Z_\alpha'\right)M' = MAM' \Rightarrow \underset{K=\operatorname{rank}(M)}{MAM' \sim W_K(n, M\Sigma M')} \quad \blacksquare$$

Remember: $A \sim W_p(n,\Sigma) \rightarrow$ use def.

$Y_\alpha := MZ_\alpha \Rightarrow Y_\alpha \sim N_p(0, M\Sigma M')$ (linearity)

$\sum_{\alpha=1}^{n} Y_\alpha Y_\alpha' \longrightarrow$ calculate

Result (Distribution of the sample covariance and quasicovariance matrices)

Let $S_N = \frac{A}{N}$ and $S_{N-1} = \frac{A}{N-1}$ be the sample covariance and quasicovariance matrices corresponding to a simple random sample of size N from a distribution $N_p(\mu, \Sigma)$.

Then, $S_N \sim W_p\left(n, \frac{\Sigma}{N}\right)$, $S_{N-1} \sim W_p\left(n, \frac{\Sigma}{n}\right)$

$n = N-1$

Dem: $A \sim W_p(n, \Sigma)$

$S_N = \frac{1}{N}A$. Take $M = \frac{1}{\sqrt{N}} I_p \Rightarrow MAM' = \frac{1}{N}A = S_N$

$MAM' \sim W_p(n, \frac{1}{N}\Sigma)$

$S_{N-1} = \frac{1}{N-1}A$. Take $M = \frac{1}{\sqrt{N-1}} I_p \Rightarrow MAM' \sim W_p(n, \frac{1}{N-1}\Sigma)$ $\blacksquare$

Remember: Just apply reproductive linearity with

$M = \frac{1}{\sqrt{N}} I_p$ and $M = \frac{1}{\sqrt{N-1}} I_p$

Result (Marginalization diagonal and by blocks and conditioning)

I: Let $A \sim W_p(n, \Sigma)$. Then $\frac{a_{ii}}{\sigma^2} \sim \chi_n^2$, $i = 1 \dots p$

Dem: consider $M_i = (0, 0, \dots, 0, \underset{(i)}{1}, 0, \dots, 0)_{1 \times p}$.

- $M_i \Sigma M_i' = \sigma_{ii} = \sigma_i^2$, $\forall i = 1 \dots p$

- $M_i A M_i' = a_{ii}$ and, by reproductive linear transformation's result, we know $a_{ii} \sim W_1(n, \sigma^2)$, $\forall i = 1 \dots p$

By result 1, we have: $\frac{a_{ii}}{\sigma_i^2} \sim \chi^2_n$, $\forall i = 1 \dots p$ ☒

Remember: $M_i = (0, \dots, 0, \underset{(i)}{1}, 0 \dots 0)_{1 \times p}$. Apply linear reproductive and result 1.

II: Let $A \sim W_p(n, \Sigma)$, and consider the partitioning into boxes

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

with A_{11} of dim $q \times q$ and A_{22} of dim $(p-q) \times (p-q)$. Then,

- $A_{11} \sim W_q(n, \Sigma_{11})$ and $A_{22} \sim W_{p-q}(n, \Sigma_{22})$
- If $\Sigma_{12} = 0 \Rightarrow A_{11}$ and A_{22} independents.

Dem: Consider $M_1 = (I_q \mid 0_{q \times (p-q)})$. We have then:

$$M_2 = (0_{(p-q) \times q} \mid I_{p-q})$$

$\rightarrow M_1 A M_1' = A_{11} \sim W_q(n, M_1 \Sigma M_1') = W_q(n, \Sigma_{11})$

$\rightarrow M_2 A M_2' = A_{22} \sim W_{p-q}(n, M_2 \Sigma M_2') = W_{p-q}(n, \Sigma_{22})$

Now, consider $\Sigma_{21} = 0 \Rightarrow \Sigma = \left(\begin{array}{c|c} \Sigma_{11} & 0 \\ \hline 0 & \Sigma_{22} \end{array}\right)$. Because $A \sim W_p(n, \Sigma)$, $A \overset{d}{=} \sum_{\alpha=1}^{n} Z_\alpha Z_\alpha'$, with $Z_\alpha \sim N_p(0, \Sigma)$ independents $\forall \alpha$, consider the partition $Z_\alpha = \begin{pmatrix} Z_\alpha[1] \\ Z_\alpha[2] \end{pmatrix}$ with $Z_\alpha[1]$ of dim q and $Z_\alpha[2]$ of dim $p-q$ and we know $Z_\alpha[1]$, $Z_\alpha[2]$ are independents. Then, $A_{11} \overset{d}{=} \sum_{\alpha=1}^{n} Z_\alpha[1] Z_\alpha[1]'$ and, because

$$A_{22} \overset{d}{=} \sum_{\alpha=1}^{n} Z_\alpha[2] Z_\alpha[2]'$$

of independence of $Z_\alpha[1]$ and $Z_\alpha[2]$ $\Rightarrow A_{11}$ and A_{22} independents ☒

Remember: i) Result linear reproductive ii) Definition of A.

Result: Let $A \sim W_p(n, \Sigma)$ and let Y be a p-dimensional random vector, independent of A, s.t. $P[Y=0]=0$. Then,

$$\frac{Y'AY}{Y'\Sigma Y} \sim \chi_n^2, \text{ being independent of } Y$$

Dem: Consider $Y = y \neq 0$ (We can consider it because $P[Y=0]=0$)

Take $M = y' \Rightarrow MAM' = y'Ay \sim W_1(n, y'\Sigma y)$

Because Y and A independents, $(Y'AY / Y=y) \overset{d}{=} y'(A/Y=y)y \overset{d}{=} y'Ay$

By result 1, $\frac{y'Ay}{y'\Sigma y} \sim \chi_n^2$ and, because it doesn't depend on $y \Rightarrow \frac{Y'AY}{Y'\Sigma Y} \sim \chi_n^2$ $\blacksquare$

Remember: $Y = y \neq 0$ ($P[Y=0]=0$) and $M = y'$, apply linear repr.

Y and A independents $\Rightarrow (Y'AY \mid Y=y) \overset{d}{=} y'(A/Y=y)y = y'Ay$.

An interesting consequence is that: Let $\{X_\alpha : \alpha = 1 \dots N\}$ be a simple random sample of distribution $N_p(\mu, \Sigma)$, with $n = N-1$:

$$n\frac{\bar{X}'S_n\bar{X}}{\bar{X}'\Sigma\bar{X}} \sim \chi_n^2, \text{ being independent of } \bar{X}$$

Dem: $A = nS_n \sim W_p(n, \Sigma)$, $n = N-1$. A and $\bar{X}$ independent (Fisher)

$\bar{X} \sim N_p(\mu, \Sigma/N)$

$$\frac{\bar{X}'A\bar{X}}{\bar{X}'\Sigma\bar{X}} = n\frac{\bar{X}'S_n\bar{X}}{\bar{X}'\Sigma\bar{X}} \sim \chi_n^2$$

$\hookrightarrow P[\bar{X}=0]=0$ (Because is a continuous distribution)

5. Hotelling's T^2 distribution

Def: Let X be a p-dimensional random vector with distribution $N_p(\mu, \Sigma)$, $(\Sigma > 0)$. Let B be a $(p \times p)$-dimensional random matrix with distribution $W_p(n, \Sigma)$, $n \geq p$. Assume that X and B are independents. Then, the non-centered Hotelling's T^2 distribution, with p and n degrees of freedom and ~~[illegible]~~ non-centrality parameter $\delta = \mu'\Sigma\mu$ is defined as the distribution of the random variable (one-dimensional) given by the quadratic form

$$T^2 = nX'B^{-1}X$$

Result: under the conditions of the previous definition, we have that:

$$\frac{T^2}{n} \frac{n-p+1}{p} \sim \boxed{F_{p,n-p+1}(\delta)}$$ Non - centered Snedecor's F distribution

with non - centrality parameter $\delta = \mu' \Sigma^{-1} \mu$.

Dem: $X \sim N_p(\mu, \Sigma)$, $\Sigma > 0$, $B \sim W_p(n, \Sigma)$, $n \geq p$, independents.

$$\frac{T^2}{n} \frac{n-p+1}{p} = \frac{\cancel{n} X' B^{-1} X}{\cancel{n}} \cdot \frac{n-p+1}{p} = X'B^{-1}X \frac{X'\Sigma^{-1}X}{X'\Sigma^{-1}X} \cdot \frac{n-p+1}{p} =$$

$$= \frac{(X'\Sigma^{-1}X)/p}{\frac{X'\Sigma^{-1}X}{X'B^{-1}X} \cdot \frac{1}{n-p+1}}$$

We know that $X'\Sigma^{-1}X \sim \chi_p^2(\delta)$ with $\delta = \mu'\Sigma^{-1}\mu$. Therefore,

$\frac{X'\Sigma^{-1}X}{X'B^{-1}X} \sim \chi^2_{n-p+1}$ independent of X, so, it's also

independent from $X'\Sigma^{-1}X \Rightarrow \frac{T^2}{n} \cdot \frac{n-p+1}{p} \sim F_{p,n-p+1}(\delta)$

$\delta = \mu'\Sigma^{-1}\mu$ ∎

Remember:

- If $\mu = 0 \Rightarrow$ We have the centered $F_{p,n-p+1}$
- In the case $p = 1$, we have that $T^2 = F_{1,n}(\delta)$, with $\delta = \left(\frac{\mu}{\sigma}\right)^2$

Def: Let $X_1, \ldots, X_N$ be a simple random sample of a distribution $N_p(\mu, \Sigma)$ $(\Sigma > 0)$ and let $\bar{X}, A, S_N$ and S_n $(n = N-1)$ be the corresponding sample mean vector, dispersion matrix and covariance and quasi-covariance matrix statistics. We define Hotelling's T^2 statistic as

$$T^2 := n\bar{X}'S_N^{-1}\bar{X} = N\bar{X}'S_n^{-1}\bar{X} = nN\bar{X}'A^{-1}\bar{X}$$

(Formally, it would be written $T^2(X_1 \ldots X_N)$, with $T^2: \mathbb{R}^{Np} \to \mathbb{R}_0^+$ measurable, although the argument is usually implicitly understood)

Result: Under the previous conditions, the statistic T^2, is distributed as follows:

$$\frac{T^2}{n} \cdot \frac{n-p+1}{p} \sim F_{p,\, n-p+1}(\delta)$$

or equivalently

$$\frac{T^2}{N-1} \cdot \frac{N-p}{p} \sim F_{p,\, N-p}(\delta)$$

with non-centrality parameter $\delta = N\mu' \Sigma^{-1} \mu$

Approach: Let $X_1 \ldots X_N$ be a simple random sample drawn from a population with distribution $N_p(\mu, \Sigma)$ $(\Sigma > 0)$, $N > p$.

We denote $X = (X_1 \ldots X_N)'$ (matrix corresponding to the random sample) and, correspondingly $x = (x_1 \ldots x_N)'$ (matrix of observed sample data)

Testing problem: We consider the hypothesis

$$\begin{cases} H_0: & \mu = \mu_0 \\ H_1: & \mu \neq \mu_0 \end{cases}, \quad \mu_0 \in \mathbb{R}^p \text{ a given fixed vector}$$

Case Σ known: Test statistic:

$$U := N(\bar{X} - \mu_0)' \Sigma^{-1} (\bar{X} - \mu_0) \sim \chi_p^2(\delta)$$

with $\delta = N(\mu - \mu_0)' \Sigma^{-1} (\mu - \mu_0)$

Test function: with $\phi : \mathbb{R}^{pN} \longrightarrow \{0, 1\}$

$$\phi(x) = \begin{cases} 1 & \text{if} \quad u := U(x) > \chi^2_{p;\alpha} \\ 0 & \text{if} \quad u := U(x) \leq \chi^2_{p;\alpha} \end{cases}$$

for all $x \in \mathbb{R}^{Np}$, with $\chi^2_{p;\alpha}$ being the value of χ^2 (centered) distribution with p degrees of freedom leaving its right a probability mass equal to α.

Case Σ unknown: Test statistic:

$$T^2 := N(\bar{X} - \mu_0)' S_N^{-1} (\bar{X} - \mu_0)$$

satisfying

$$\frac{T^2}{n} \cdot \frac{n-p+1}{p} \sim F_{p,\, n-p+1}(\delta)$$

with $\delta = N(\mu - \mu_0)' \Sigma^{-1} (\mu - \mu_0)$

Test function: with $\phi: \mathbb{R}^{Np} \to \{0,1\}$

$$\phi(x) = \begin{cases} 1 & \text{if } t := T^2(x) > \frac{np}{n-p+1} F_{p,n-p+1,\alpha} \\ 0 & \text{if } t := T^2(x) \leq \frac{np}{n-p+1} F_{p,n-p+1,\alpha} \end{cases}$$

$\forall x \in \mathbb{R}^{pN}$ with $F_{p,n-p+1,\alpha}$ being the value of F (centered) distribution with $(p, n-p+1)$ degrees of freedom leaving on its right a probability mass equal to α.