Resultado 1 : Si $\mathbf{x}_1, \ldots, \mathbf{x}_n \overset{i.i.d.}{\sim} N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

entonces los estimadores de máxima verosimilitud para $\boldsymbol{\mu}$ y $\boldsymbol{\Sigma}$ son

$$\hat{\boldsymbol{\mu}} = \overline{\mathbf{x}}_{(n)} \quad \text{y} \quad \hat{\boldsymbol{\Sigma}} = \frac{1}{n} \sum_{j=1}^{n} (\mathbf{x}_j - \overline{\mathbf{x}}_{(n)})(\mathbf{x}_j - \overline{\mathbf{x}}_{(n)})'$$

Dem ver las notas MLE-MULTVAR-NORMAL

Resultado 2: Si $\mathbf{x}_1, \ldots, \mathbf{x}_n \overset{i.i.d.}{\sim} N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

entonces $n\hat{\boldsymbol{\Sigma}} \sim \text{Wishart}_p(\boldsymbol{\Sigma}, n-1)$

Este Teorema se revisó en la página 55 de las notas para componentes principales

Resultado 3: Si $\mathbf{x}_1, \ldots, \mathbf{x}_n \overset{i.i.d.}{\sim} N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

entonces $\hat{\boldsymbol{\mu}} = \overline{\mathbf{x}}_{(n)}$ y $n\hat{\boldsymbol{\Sigma}}$ son independientes

Dem Véase Press, S. J., (2005), "Applied Multivariate Analysis, Using Bayesian and Frequentist Methods of Inference", 2nd Edition, Dover, Teorema (7.1.2), pag 181, sección 7.1.

Resultado 4: Si $\mathbb{X}_1, \dots, \mathbb{X}_n \overset{i.i.d.}{\sim} N_p(\mu, \Sigma)$

entonces $(\hat{\mu}, n\hat{\Sigma})$ es suficiente para

$\Theta = (\mu, \Sigma)$.

Dem Véase Press, S. J. (2005) ...,

Teorema (7.1.1), página 181, sección 7.1

Estimación de máxima verosimilitud para

Λ y Ψ (Lawley y Maxwell, 1963. Anderson

y Rubin, 1956).

Argumento (caso variables discretas):

Sean $f(x_1, \dots, x_n, \theta)$ la densidad de $X_1, \dots, X_n$

y $f(x_1, \dots, x_n, t, \theta)$ la densidad conjunta de

$x_1, \dots, x_n, T(x_1, \dots, x_n)$, donde T es una

estadística suficiente para θ.

$f(x_1, \dots, x_n, t, \theta) = \mathbb{P}(X_1 = x_1, \dots, X_n = x_n, T = t; \theta)$

$= \mathbb{P}(X_1 = x_1; \dots; X_n = x_n; \theta) = f(x_1, \dots, x_n, \theta)$

$\uparrow$

si t es tal que $T(x_1, \dots, x_n) = t$

$$f(x_1,\dots,x_n,\theta) = f(x_1,\dots,x_n,t,\theta) =$$

$$f(x_1,\dots,x_n \mid t,\theta)\, f(t,\theta)$$

$$\underset{\substack{\uparrow \\ t \text{ es} \\ \text{suficiente}}}{=} f(x_1,\dots,x_n \mid t)\, f(t,\theta)$$

Como el primer factor <u>no</u> depende de θ, para maximizar $f(x_1,\dots,x_n,\theta)$ c.r. a θ debemos maximizar $f(t,\theta)$ c.r. a θ.

Para poder aplicar este argumento al caso(1), recordemos que $\overline{x}_{(n)} \sim N_p(\mu, \frac{1}{n}\Sigma)$, entonces por el resultado 3

$$f_{\hat{\mu},\, n\hat{\Sigma}}(t_1, t_2) = f_{\hat{\mu}}(t_1)\, f_{n\hat{\Sigma}}(t_2)$$

$$\overset{(2)}{=} N_p(t_1; \mu, \tfrac{1}{n}\Sigma) \cdot \text{Wishart}_p(t_2; \Sigma, n-1)$$

que es la densidad conjunta de $T(x_1,\dots,x_n) = (\hat{\mu}, n\hat{\Sigma})$(3)

(1) $x_1,\dots,x_n \overset{iid}{\sim} N_p(\mu,\Sigma)$

(2) Usando: Resultado 2.

(3) $T(x_1,\dots,x_n)$ suficiente

Resultado 4:

$$\tilde{L}_T(t,\mu,\Sigma) = f_{\hat{\mu},\, n\hat{\Sigma}}(t_1, t_2)$$

De acuerdo al argumento dado arriba, hay que maximizar a $\tilde{L}_T(t,\theta) = \tilde{L}_T(t,\mu,\Sigma)$ como función de μ, Σ, pero como lo que nos interesa son los parámetros Q y Ψ, entonces podemos substituir $\Sigma = QQ' + \Psi$ en $\tilde{L}_T(t,\mu,\Sigma)$ y maximizar esta última c.r.a Q y Ψ. Tomando logaritmos

$$\ell(\mu, Q, \Psi, x) = \log \tilde{L}_T(t,\mu, QQ'+\Psi)$$

$$= \log\left\{ N_p\left(t_1, \mu, \tfrac{1}{n}(QQ'+\Psi)\right) \times \text{Wishart}_p\left(t_2; QQ'+\Psi, n-1\right)\right\}$$

$$= c + \left(\frac{n-p-2}{2}\right)\log|A| - \frac{n}{2}\log|QQ'+\Psi|$$
$$-\frac{1}{2}\operatorname{tr}\left[(QQ'+\Psi)^{-1}A\right] - \frac{n}{2}(\bar{x}_{(n)}-\mu)'(QQ'+\Psi)^{-1}(\bar{x}_{(n)}-\mu),$$

$A = n\hat{\Sigma}$

donde c es una constante que no depende de Q, Ψ y x. Notemos que al hacer $\mu = \hat{\mu}$ el último término se cancela, después de esto, si se deriva con respecto a Q y $\Psi_{11}, \Psi_{22}, \ldots, \Psi_{pp}$ y se igualan estas derivadas a 0, se obtiene un sistema de ecuaciones

$$\mathrm{Diag}(\hat{\Psi} + \hat{Q}\hat{Q}') = \mathrm{Diag}(\hat{\Sigma}),$$

$$\hat{\Sigma}\,\hat{\Psi}^{-1}\hat{Q} = \hat{Q}(\mathbb{I} + \hat{Q}'\hat{\Psi}^{-1}Q),$$

que deben satisfacer los estimadores de máxima verosimilitud $\hat{Q}$ y $\hat{\Psi}$.

Modelo para datos estandarizados

$$\mathbb{X} = \mathbb{Q}\mathbb{F} + \mathbb{U} + \mathbb{M}$$

donde denotando $\mathbb{D} = DIAG(VAR(X_1), \ldots, VAR(X_p))$

tenemos que

$$\mathbb{D}^{-1/2}(\mathbb{X} - \mathbb{M}) = (\mathbb{D}^{-1/2}\mathbb{Q})\mathbb{F} + \mathbb{D}^{-1/2}\mathbb{U} \cdots (\mathrm{I}).$$

$$\mathbb{D}^{-1/2}\mathbb{Q} \equiv \mathbb{Q}_Y \; ; \quad \mathbb{Y} \equiv \mathbb{D}^{-1/2}(\mathbb{X} - \mathbb{M}) \; ; \quad \mathbb{D}^{-1/2}\mathbb{U} = \mathbb{U}_Y$$

$$\mathbb{Y} = \mathbb{Q}_Y \mathbb{F} + \mathbb{U}_Y , \quad \cdots\cdots \quad (\mathrm{II})$$

donde
$$\begin{aligned} VAR(\mathbb{U}_Y) &= COV(\mathbb{D}^{-1/2}\mathbb{U}, \mathbb{D}^{-1/2}\mathbb{U}) \\ &= \mathbb{D}^{-1/2}\, COV(\mathbb{U}, \mathbb{U})\, \mathbb{D}^{-1/2} \\ &= \mathbb{D}^{-1/2}\, \Psi\, \mathbb{D}^{-1/2}. \end{aligned}$$

Al calcular $VAR(\cdot)$ en ambos lados de (II) (usando los supuestos (z) en la página 8 tenemos

matriz de correlaciones de $\mathbb{X}$ $\longrightarrow$

$$\begin{aligned} \mathbb{R} &= COV(\mathbb{D}^{-1/2}\mathbb{Q}\mathbb{F}, \mathbb{D}^{-1/2}\mathbb{Q}\mathbb{F}) + COV(\mathbb{D}^{-1/2}\mathbb{U}, \mathbb{D}^{-1/2}\mathbb{U}) \\ &= \mathbb{D}^{-1/2}\mathbb{Q}\, COV(\mathbb{F}, \mathbb{F})\, \mathbb{Q}'\mathbb{D}^{-1/2} + \mathbb{D}^{-1/2}\, \Psi\, \mathbb{D}^{-1/2} \\ &= \mathbb{Q}_Y \mathbb{Q}_Y' + \Psi_Y \end{aligned}$$

Para estimar Q_Y y Ψ_Y en (II) usamos los MLE's $\hat{Q}$ y $\hat{\Psi}$ obtenidos en las páginas 15-16, de donde

$$\hat{Q}_Y = \hat{D}^{-1/2}\,\hat{Q}$$

$$\hat{\Psi}_Y = \hat{D}^{-1/2}\,\hat{\Psi}\,\hat{D}^{-1/2}$$

$$\hat{D} = \mathrm{DIAG}\left(\widehat{\mathrm{VAR}}(X_1), \ldots, \widehat{\mathrm{VAR}}(X_p)\right)$$