Sean $\mathbb{X}_1, \ldots, \mathbb{X}_n$ vectores aleatorios independientes e idénticamente distribuidos $N_p(\mu, \Sigma)$, la distribución conjunta de $\mathbb{X}_1, \ldots, \mathbb{X}_n$ es

$$f_{\mathbb{X}_1,\ldots\mathbb{X}_n}(x_1,\ldots,x_n) = \frac{1}{(2\pi)^{\frac{np}{2}}\ |\Sigma|^{n/2}} e^{-\frac{1}{2}\sum_{j=1}^{n}(x_j-\mu)'\Sigma^{-1}(x_j-\mu)} .$$

Podemos aplicar la identidad[2] (x) en el exponente para obtener

$$f_{\mathbb{X}_1,\ldots\mathbb{X}_n}(x_1,\ldots,x_n) = \frac{1}{(2\pi)^{\frac{np}{2}}\ |\Sigma|^{n/2}} e^{-\frac{1}{2}tr\left[\Sigma^{-1}\left\{\sum_{j=1}^{n}(x_j-\mu)(x_j-\mu)'\right\}\right]}$$

En el problema de estimar[1] μ y Σ, tenemos que encontrar valores de estos dos argumentos, digamos $\hat{\mu}$ y $\hat{\Sigma}$, donde

$$L(\mu, \Sigma, \underline{x}) \equiv f_{\mathbb{X}_1\ldots\mathbb{X}_n}(x_1,\ldots,x_n, \mu, \Sigma)$$

alcance su ~~máximo~~ máximo, por monotonía de la función logaritmo, si $\hat{\mu}$ y $\hat{\Sigma}$ son los valores de μ y Σ donde $L(\mu, \Sigma, x)$ alcanza su

(1) estimación por máxima verosimilitud

(2) Identidad (x), nota 2

máximo, entonces $\hat{\mu}$ y $\hat{\Sigma}$ también son los valores de μ y Σ donde

$$l(\mu, \Sigma, \underline{x}) \equiv \log L(\mu, \Sigma, \underline{x})$$

alcanza su máximo (y viceversa).

Sea $W = \Sigma^{-1}$, entonces

$$l(\mu, W, \underline{x}) = -\frac{np}{2}\log(2\pi) + \frac{n}{2}\log(|W|) - \frac{1}{2}\operatorname{tr}\left[W \sum_{j=1}^{n}(x_j-\mu)(x_j-\mu)'\right].$$

Pero usando (v) de la nota 3

$$l(\mu, W, \underline{x}) = -\frac{np}{2}\log(2\pi) + \frac{n}{2}\log(|W|) - \frac{1}{2}\operatorname{tr}\left[W V\right] - \frac{1}{2}\operatorname{tr}\left[n W P(\mu)\right];$$

con

$$P(\mu) = (\bar{x}_{(n)} - \mu)(\bar{x}_{(n)} - \mu)'$$

$$V = \sum_{j=1}^{n}(x_j - \bar{x}_{(n)})(x_j - \bar{x}_{(n)})'$$

Para encontrar valores de μ y W que maximizan $l(\mu, W, \underline{x})$ usaremos cálculo diferencial matricial, así como el principio de supremos iterados[1].

(1) Véase nota 5

Por el principio de los supremos iterados, podemos optimizar $l(\underline{\mu}, \Sigma, \underline{x}_G)$ primero en la dirección de $\underline{\mu}$ y después en la dirección de Σ.

$$\frac{\partial l(\underline{\mu}, \Sigma, \underline{x}_G)}{\partial \underline{\mu}} = -\frac{n}{2} \frac{\partial}{\partial \underline{\mu}} \left\{ tr\left(W \left(\overline{\underline{x}}_{G(n)} - \underline{\mu}\right)\left(\overline{\underline{x}}_{G(n)} - \underline{\mu}\right)' \right) \right\}$$

$$= -\frac{n}{2} \times \left\{ 2 W \left(\underline{\mu} - \overline{\underline{x}}_{G(n)} \right) \right\}$$

la última igualdad se sigue de la ecuación (u), nota 4 sobre cálculo matricial.

$$\left.\frac{\partial l}{\partial \underline{\mu}}\right|_{\underline{\mu} = \hat{\underline{\mu}}} = 0 \iff \hat{\underline{\mu}} = \overline{\underline{x}}_{G(n)} .$$

Además, $\left.\frac{\partial^2 l}{\partial \underline{\mu} \partial \underline{\mu}'}\right|_{\underline{\mu} = \hat{\underline{\mu}}} = -n \Sigma^{-1}$ y como Σ es positivo definida ($\Rightarrow \Sigma^{-1}$ es positivo definida), la matriz $-n \Sigma^{-1}$ es negativo definida

$\therefore$ $\hat{\underline{\mu}} = \overline{\underline{x}}_{G(n)}$ es un máximo de $l(\underline{\mu}, \Sigma, \underline{x}_G)$
(manteniendo Σ fijo)

Ahora, procedemos a maximizar $\ell(\hat{\mu}, W, \underline{x})$ como función de W. Con este fin, notemos que

$$\frac{\partial}{\partial W}\log(|W|) = |W|^{-1}\frac{\partial}{\partial W}|W|$$

$$= |W|^{-1}\left(2|W|\,W^{-1} - \mathrm{Diag}(|W|\,W^{-1})\right),$$

donde la última igualdad se sigue de 2 en la nota 4 sobre cálculo matricial. Entonces

$$\frac{\partial}{\partial W}\log(|W|) \overset{(1)}{=} 2W^{-1} - \mathrm{Diag}(W^{-1}).$$

Además,

$$\frac{\partial\, tr(WV)}{\partial W} = \frac{\partial\, tr(VW)}{\partial W}$$

$$= 2V - \mathrm{Diag}(V),$$

la última igualdad se sigue de 1, en la nota 4.

si $p=2$ y $a\in\mathbb{R}$:

(1) $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ $\mathrm{Diag}(A) = \begin{pmatrix} a_{11} & 0 \\ 0 & a_{22} \end{pmatrix}$ $\mathrm{Diag}(aA) = a\cdot\mathrm{Diag}(A)$

Para el último término en $l(\hat{\mu}, W, \underline{x})$

$$tr\left(n W P(\hat{\mu})\right) = n\, tr\left(W P(\hat{\mu})\right)$$

$$= n\, tr\left(P(\hat{\mu}) W\right), \quad \text{entonces usando (1) de nuevo}$$

$$\frac{\partial}{\partial W} tr\left(n W P(\hat{\mu})\right) = n \frac{\partial}{\partial W} tr\left(P(\hat{\mu}) W\right)$$

$$= n\left[2P(\hat{\mu}) - \mathrm{Diag}\left(P(\hat{\mu})\right)\right].$$

Por tanto

$$\frac{\partial}{\partial W} l(\hat{\mu}, W, \underline{x}) = \frac{n}{2}\left[2W^{-1} - \mathrm{Diag}\left(W^{-1}\right)\right]$$

$$-\frac{1}{2}\left[2V - \mathrm{Diag}(V)\right] - \frac{n}{2}\left[2P(\hat{\mu}) - \mathrm{Diag}\left(P(\hat{\mu})\right)\right]$$

$$= \frac{n}{2}\left[2W^{-1} - \mathrm{Diag}\left(W^{-1}\right)\right] - \frac{1}{2}\left[2A - \mathrm{Diag}(A)\right]. \quad \cdots\cdots (d1)$$

Para la última igualdad usamos (v) en la nota 3, donde $A = \sum_{j=1}^{n}(x_j - \hat{\mu})(x_j - \hat{\mu})' =$

$$= \sum_{j=1}^{n}(x_j - \overline{x}_{(n)})(x_j - \overline{x}_{(n)})'.$$

Al igualar la derivada en (d1) con 0 obtenemos $(W^{-1} = \Sigma)$ la identidad

$$2\Sigma - Diag(\Sigma) = 2\left(\frac{\mathbb{A}}{n}\right) - Diag\left(\frac{\mathbb{A}}{n}\right)$$

Escribiendo las entradas de ambas matrices tendríamos

$$\begin{pmatrix} \Sigma_{11} & 2\Sigma_{12} & 2\Sigma_{13} & \cdots & 2\Sigma_{1p} \\ 2\Sigma_{21} & \Sigma_{22} & 2\Sigma_{23} & \cdots & 2\Sigma_{2p} \\ \vdots & & & & \vdots \\ 2\Sigma_{p1} & 2\Sigma_{p2} & 2\Sigma_{p3} & \cdots & \Sigma_{pp} \end{pmatrix} = \begin{pmatrix} \frac{A_{11}}{n} & \frac{2A_{12}}{n} & \frac{2A_{13}}{n} & \cdots & \frac{2A_{1p}}{n} \\ \frac{2A_{21}}{n} & \frac{A_{22}}{n} & \frac{2A_{23}}{n} & \cdots & \frac{2A_{2p}}{n} \\ \vdots & \vdots & & & \vdots \\ \frac{2A_{p1}}{n} & \frac{2A_{p2}}{n} & \frac{2A_{p3}}{n} & \cdots & \frac{A_{pp}}{n} \end{pmatrix}$$

lo cual implicaría que

$$i=j \qquad \Sigma_{ii} = \frac{A_{ii}}{n}$$

$$i \neq j \qquad \Sigma_{ij} = \frac{A_{ij}}{n}$$

$$\therefore \quad \hat{\Sigma} = \frac{1}{n}\mathbb{A} = \frac{1}{n}\sum_{j=1}^{n} (x_{G_j} - \bar{x}_{G(n)})(x_{G_j} - \bar{x}_{G_n})'$$