UFOP/ICEB/DEEST
EST022 - Estatística Multivariada II
Prof. Tiago Martins Pereira

Estimadores de Máxima Verossimilhança para o vetor de médias e matriz de covariâncias

Antes de encontrarmos os estimadores de máxima verossimilhança para os parâmetros da distribuição normal multivariada, temos que nos atentar aos seguintes resultados:

Resultado 01: Regras de derivação em relação à vetores e matrizes.

a) $$\frac{\partial\left[\sum_{j=1}^{n}\left(\boldsymbol{X}_j-\boldsymbol{\mu}\right)^t\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{X}_j-\boldsymbol{\mu}\right)\right]}{\partial\boldsymbol{\mu}}=-2\sum_{j=1}^{n}\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{X}_j-\boldsymbol{\mu}\right)$$

b) $$\frac{\partial\ln|\boldsymbol{\Sigma}|}{\partial\boldsymbol{\Sigma}}=\left(\boldsymbol{\Sigma}^{-1}\right)^t$$

c) $$\frac{\partial\mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{A}\right)}{\partial\boldsymbol{\Sigma}}=-\boldsymbol{\Sigma}^{-1}\boldsymbol{A}\boldsymbol{\Sigma}^{-1}$$

Resultado 02: Sejam $\boldsymbol{A}$ uma matriz simétrica $(k\times k)$ e $\boldsymbol{x}$ um vetor $(k\times 1)$. Então,

a) $\boldsymbol{x}^t\boldsymbol{A}\boldsymbol{x}=\mathrm{tr}(\boldsymbol{x}^t\boldsymbol{A}\boldsymbol{x})=\mathrm{tr}(\boldsymbol{A}\boldsymbol{x}\boldsymbol{x}^t)$

b) $\mathrm{tr}(\boldsymbol{A})=\displaystyle\sum_{i=1}^{k}\lambda_i$, em que λ_i são os autovalores de $\boldsymbol{A}$.

Demonstração: Seja $\boldsymbol{X}_1,\ \boldsymbol{X}_2,\ \cdots,\ \boldsymbol{X}_n$ uma amostra aleatória de uma distribuição $N_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$. A função de verossimilhança para o modelo normal multivariado é dado por:

$$\boldsymbol{L}(\boldsymbol{X};\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{np/2}|\boldsymbol{\Sigma}|^{n/2}}\exp\left\{-\frac{1}{2}\sum_{j=1}^{n}\left(\boldsymbol{X}_j-\boldsymbol{\mu}\right)^t\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{X}_j-\boldsymbol{\mu}\right)\right\}$$

Para encontrarmos os estimadores de máxima verossimilhança para o vetor de médias e matriz de covariâncias, basta tomarmos as derivadas parciais de primeira ordem da função de verossimilhança em relação a cada parâmetro e igualarmos as funções obtidas a zero. Para isso, vamos tomar o logaritmo da função de verossimilhança e maximizar a função resultante, conhecida como função suporte. Este artifício é utilizado para facilitar o encontro dos estimadores. Uma vez que a função logarítmica é monótona crescente, os máximos das funções de verossimilhança e suporte são exatamente os mesmos.

Seja então a função suporte:

$$\begin{aligned}
g(\boldsymbol{X};\boldsymbol{\mu},\boldsymbol{\Sigma}) &= \ln[L(\boldsymbol{X};\boldsymbol{\mu},\boldsymbol{\Sigma})] \\
&= \ln\left[\prod_{j=1}^{n} f_{\boldsymbol{X}}(\boldsymbol{X}_j)\right] \\
&= \ln\left[(2\pi)^{-\frac{np}{2}}\,|\boldsymbol{\Sigma}|^{-\frac{n}{2}}\exp\left\{-\frac{1}{2}\sum_{j=1}^{n}(\boldsymbol{X}_j-\boldsymbol{\mu})^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_j-\boldsymbol{\mu})\right\}\right] \\
&= \ln(2\pi)^{-\frac{np}{2}} + \ln|\boldsymbol{\Sigma}|^{-\frac{n}{2}} + \ln\left[\exp\left\{-\frac{1}{2}\sum_{j=1}^{n}(\boldsymbol{X}_j-\boldsymbol{\mu})^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_j-\boldsymbol{\mu})\right\}\right] \\
&= -\frac{np}{2}\ln(2\pi) - \frac{n}{2}\ln|\boldsymbol{\Sigma}| - \frac{1}{2}\sum_{j=1}^{n}(\boldsymbol{X}_j-\boldsymbol{\mu})^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_j-\boldsymbol{\mu})
\end{aligned}$$

De forma que, derivando a função suporte em relação ao parâmetro $\boldsymbol{\mu}$, utilizando a regra apresentada na alínea a) do Resultado 01:

$$\begin{aligned}
\frac{\partial g\left(\boldsymbol{X};\boldsymbol{\mu},\boldsymbol{\Sigma}\right)}{\partial\boldsymbol{\mu}} &= \frac{\partial\left\{-\frac{np}{2}\ln(2\pi) - \frac{n}{2}\ln|\boldsymbol{\Sigma}| - \frac{1}{2}\sum_{j=1}^{n}\left[\left(\boldsymbol{X}_j-\boldsymbol{\mu}^t\right)\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{X}_j-\boldsymbol{\mu}^t\right)\right]\right\}}{\partial\boldsymbol{\mu}} \\
&= -\frac{1}{2}\left(-2\right)\sum_{j=1}^{n}\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{X}_j-\boldsymbol{\mu}\right) \\
&= \sum_{j=1}^{n}\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{X}_j-\boldsymbol{\mu}\right) \\
&= \boldsymbol{\Sigma}^{-1}\sum_{j=1}^{n}\left(\boldsymbol{X}_j-\boldsymbol{\mu}\right)
\end{aligned}$$

Igualando a zero:

$$
\begin{aligned}
\hat{\boldsymbol{\Sigma}}^{-1}\sum_{j=1}^{n}(\boldsymbol{X}_j-\hat{\boldsymbol{\mu}})=\mathbf{0} \quad &\Rightarrow \quad \sum_{j=1}^{n}(\boldsymbol{X}_j-\hat{\boldsymbol{\mu}})=\mathbf{0}\\
&\Rightarrow \quad \sum_{j=1}^{n}(\boldsymbol{X}_j-\hat{\boldsymbol{\mu}})=\mathbf{0}\\
&\Rightarrow \quad \sum_{j=1}^{n}\boldsymbol{X}_j-\sum_{j=1}^{n}\hat{\boldsymbol{\mu}}=\mathbf{0}\\
&\Rightarrow \quad \sum_{j=1}^{n}\boldsymbol{X}_j-n\hat{\boldsymbol{\mu}}=\mathbf{0}\\
&\Rightarrow \quad \sum_{j=1}^{n}\boldsymbol{X}_j=n\hat{\boldsymbol{\mu}} \quad \therefore \quad \boldsymbol{\mu}=\frac{\sum_{j=1}^{n}\boldsymbol{X}_j}{n}=\bar{\boldsymbol{X}}
\end{aligned}
$$

$\blacksquare$

O expoente da função de verossimilhança pode ser simplificado. De acordo com a alínea a) do Resultado 02, temos,

$$
\begin{aligned}
(\boldsymbol{X}_j-\boldsymbol{\mu})^t\,\boldsymbol{\Sigma}^{-1}\,(\boldsymbol{X}_j-\boldsymbol{\mu}) &= \operatorname{tr}\left[(\boldsymbol{X}_j-\boldsymbol{\mu})^t\,\boldsymbol{\Sigma}^{-1}\,(\boldsymbol{X}_j-\boldsymbol{\mu})\right]\\
&= \operatorname{tr}\left[\boldsymbol{\Sigma}^{-1}\,(\boldsymbol{X}_j-\boldsymbol{\mu})\,(\boldsymbol{X}_j-\boldsymbol{\mu})^t\right]
\end{aligned}
$$

em seguida,

$$
\begin{aligned}
\sum_{j=1}^{n}(\boldsymbol{X}_j-\boldsymbol{\mu})^t\,\boldsymbol{\Sigma}^{-1}\,(\boldsymbol{X}_j-\boldsymbol{\mu}) &= \sum_{j=1}^{n}\operatorname{tr}\left[(\boldsymbol{X}_j-\boldsymbol{\mu})^t\,\boldsymbol{\Sigma}^{-1}\,(\boldsymbol{X}_j-\boldsymbol{\mu})\right]\\
&= \sum_{j=1}^{n}\operatorname{tr}\left[\boldsymbol{\Sigma}^{-1}\,(\boldsymbol{X}_j-\boldsymbol{\mu})\,(\boldsymbol{X}_j-\boldsymbol{\mu})^t\right]\\
&= \operatorname{tr}\left[\boldsymbol{\Sigma}^{-1}\left(\sum_{j=1}^{n}(\boldsymbol{X}_j-\boldsymbol{\mu})\,(\boldsymbol{X}_j-\boldsymbol{\mu})^t\right)\right]
\end{aligned}
$$

uma vez que o traço da soma é igual à soma dos traços. Somando e subtraindo $\bar{\boldsymbol{X}}$ de cada termo $(\boldsymbol{X}_j-\boldsymbol{\mu})$ em $\sum_{j=1}^{n}(\boldsymbol{X}_j-\boldsymbol{\mu})\,(\boldsymbol{X}_j-\boldsymbol{\mu})^t$, temos,

$$
\begin{aligned}
\sum_{j=1}^{n}\left(\boldsymbol{X}_j-\bar{\boldsymbol{X}}+\bar{\boldsymbol{X}}-\boldsymbol{\mu}\right)\left(\boldsymbol{X}_j-\bar{\boldsymbol{X}}+\bar{\boldsymbol{X}}-\boldsymbol{\mu}\right)^t &= \sum_{j=1}^{n}\left(\boldsymbol{X}_j-\bar{\boldsymbol{X}}\right)\left(\boldsymbol{X}_j-\bar{\boldsymbol{X}}\right)^t \\
&+ 2\sum_{j=1}^{n}\left(\boldsymbol{X}_j-\bar{\boldsymbol{X}}\right)\left(\bar{\boldsymbol{X}}-\boldsymbol{\mu}\right)^t \\
&+ \sum_{j=1}^{n}\left(\bar{\boldsymbol{X}}-\boldsymbol{\mu}\right)\left(\bar{\boldsymbol{X}}-\boldsymbol{\mu}\right)^t
\end{aligned}
$$

Uma vez que a matriz de produtos cruzados $\sum_{j=1}^{n}\left(\boldsymbol{X}_j-\bar{\boldsymbol{X}}\right)\left(\bar{\boldsymbol{X}}-\boldsymbol{\mu}\right)^t$ é uma matriz nula, temos,

$$
\sum_{j=1}^{n}\left(\boldsymbol{X}_j-\boldsymbol{\mu}\right)\left(\boldsymbol{X}_j-\boldsymbol{\mu}\right)^t = \sum_{j=1}^{n}\left(\boldsymbol{X}_j-\bar{\boldsymbol{X}}\right)\left(\boldsymbol{X}_j-\bar{\boldsymbol{X}}\right)^t + n\left(\bar{\boldsymbol{X}}-\boldsymbol{\mu}\right)\left(\bar{\boldsymbol{X}}-\boldsymbol{\mu}\right)^t
$$

Assim, podemos reescrever a função suporte,

$$
\begin{aligned}
g(\boldsymbol{X};\boldsymbol{\mu},\boldsymbol{\Sigma}) &= -\frac{np}{2}\ln(2\pi)-\frac{n}{2}\ln|\boldsymbol{\Sigma}| \\
&- \frac{1}{2}\mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}\left(\sum_{j=1}^{n}\left(\boldsymbol{X}_j-\bar{\boldsymbol{X}}\right)\left(\boldsymbol{X}_j-\bar{\boldsymbol{X}}\right)^t + n\left(\bar{\boldsymbol{X}}-\boldsymbol{\mu}\right)\left(\bar{\boldsymbol{X}}-\boldsymbol{\mu}\right)^t\right)\right] \\
&= -\frac{np}{2}\ln(2\pi)-\frac{n}{2}\ln|\boldsymbol{\Sigma}|-\frac{1}{2}\mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}\sum_{j=1}^{n}\left(\boldsymbol{X}_j-\bar{\boldsymbol{X}}\right)\left(\boldsymbol{X}_j-\bar{\boldsymbol{X}}\right)^t\right] \\
&- \frac{1}{2}\mathrm{tr}\left[n\boldsymbol{\Sigma}^{-1}\left(\bar{\boldsymbol{X}}-\boldsymbol{\mu}\right)\left(\bar{\boldsymbol{X}}-\boldsymbol{\mu}\right)^t\right]
\end{aligned}
$$

Derivando em relação a $\boldsymbol{\Sigma}$, utilizando as alíneas b) e c) do Resultado 02, encontramos o estimador de máxima verossimilhança para $\boldsymbol{\Sigma}$:

$$
\begin{aligned}
\frac{\partial g(\boldsymbol{X};\boldsymbol{\mu},\boldsymbol{\Sigma})}{\partial\boldsymbol{\Sigma}} &= \frac{\partial\left[-\frac{np}{2}\ln(2\pi)-\frac{n}{2}\ln|\boldsymbol{\Sigma}|-\frac{1}{2}\mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}\sum_{j=1}^{n}\left(\boldsymbol{X}_j-\bar{\boldsymbol{X}}\right)\left(\boldsymbol{X}_j-\bar{\boldsymbol{X}}\right)^t\right]-\frac{1}{2}\mathrm{tr}\left[n\boldsymbol{\Sigma}^{-1}\left(\bar{\boldsymbol{X}}-\boldsymbol{\mu}\right)\left(\bar{\boldsymbol{X}}-\boldsymbol{\mu}\right)^t\right]\right]}{\partial\boldsymbol{\Sigma}} \\
&= -\frac{n}{2}(\boldsymbol{\Sigma}^{-1})^t+\frac{1}{2}\boldsymbol{\Sigma}^{-1}\sum_{j=1}^{n}\left(\boldsymbol{X}_j-\bar{\boldsymbol{X}}\right)\left(\boldsymbol{X}_j-\bar{\boldsymbol{X}}\right)^t\boldsymbol{\Sigma}^{-1} \\
&+ \frac{n}{2}\boldsymbol{\Sigma}^{-1}(\bar{\boldsymbol{X}}-\boldsymbol{\mu})(\bar{\boldsymbol{X}}-\boldsymbol{\mu})^t\boldsymbol{\Sigma}^{-1}
\end{aligned}
$$

Igualando a zero:

$$
\begin{aligned}
-\frac{n}{2}(\hat{\boldsymbol{\Sigma}}^{-1})^t &+ \frac{1}{2}\hat{\boldsymbol{\Sigma}}^{-1}\sum_{j=1}^{n}\left(\boldsymbol{X}_j-\bar{\boldsymbol{X}}\right)\left(\boldsymbol{X}_j-\bar{\boldsymbol{X}}\right)^t\hat{\boldsymbol{\Sigma}}^{-1}+\frac{n}{2}\hat{\boldsymbol{\Sigma}}^{-1}(\bar{\boldsymbol{X}}-\hat{\boldsymbol{\mu}})(\bar{\boldsymbol{X}}-\hat{\boldsymbol{\mu}})^t\hat{\boldsymbol{\Sigma}}^{-1}=\mathbf{0}\\
-n\hat{\boldsymbol{\Sigma}}^{-1} &+ \hat{\boldsymbol{\Sigma}}^{-1}\sum_{j=1}^{n}\left(\boldsymbol{X}_j-\bar{\boldsymbol{X}}\right)\left(\boldsymbol{X}_j-\bar{\boldsymbol{X}}\right)^t\hat{\boldsymbol{\Sigma}}^{-1}+n\hat{\boldsymbol{\Sigma}}^{-1}(\bar{\boldsymbol{X}}-\hat{\boldsymbol{\mu}})(\bar{\boldsymbol{X}}-\hat{\boldsymbol{\mu}})^t\hat{\boldsymbol{\Sigma}}^{-1}=\mathbf{0}\\
-n\hat{\boldsymbol{\Sigma}}^{-1} &+ \hat{\boldsymbol{\Sigma}}^{-1}\sum_{j=1}^{n}\left(\boldsymbol{X}_j-\bar{\boldsymbol{X}}\right)\left(\boldsymbol{X}_j-\bar{\boldsymbol{X}}\right)^t\hat{\boldsymbol{\Sigma}}^{-1}+n\hat{\boldsymbol{\Sigma}}^{-1}(\bar{\boldsymbol{X}}-\bar{\boldsymbol{X}})(\bar{\boldsymbol{X}}-\bar{\boldsymbol{X}})^t\hat{\boldsymbol{\Sigma}}^{-1}=\mathbf{0}\\
-n\hat{\boldsymbol{\Sigma}}^{-1} &+ \hat{\boldsymbol{\Sigma}}^{-1}\sum_{j=1}^{n}\left(\boldsymbol{X}_j-\bar{\boldsymbol{X}}\right)\left(\boldsymbol{X}_j-\bar{\boldsymbol{X}}\right)^t\hat{\boldsymbol{\Sigma}}^{-1}=\mathbf{0}\\
-n\hat{\boldsymbol{\Sigma}}\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\Sigma}} &+ \hat{\boldsymbol{\Sigma}}\hat{\boldsymbol{\Sigma}}^{-1}\sum_{j=1}^{n}\left(\boldsymbol{X}_j-\bar{\boldsymbol{X}}\right)\left(\boldsymbol{X}_j-\bar{\boldsymbol{X}}\right)^t\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\Sigma}}=\mathbf{0}\\
n\hat{\boldsymbol{\Sigma}} &= \sum_{j=1}^{n}\left(\boldsymbol{X}_j-\bar{\boldsymbol{X}}\right)\left(\boldsymbol{X}_j-\bar{\boldsymbol{X}}\right)^t\\
\hat{\boldsymbol{\Sigma}} &= \frac{1}{n}\sum_{j=1}^{n}\left(\boldsymbol{X}_j-\bar{\boldsymbol{X}}\right)\left(\boldsymbol{X}_j-\bar{\boldsymbol{X}}\right)^t=\frac{(n-1)}{n}\boldsymbol{S}=\boldsymbol{S}_n
\end{aligned}
$$

■

Podemos reescrever a função de verossimilhança da seguinte maneira:

$$
\boldsymbol{L}(\boldsymbol{X};\boldsymbol{\mu},\boldsymbol{\Sigma}) = (2\pi)^{-\frac{np}{2}}|\boldsymbol{\Sigma}|^{-\frac{n}{2}}\exp\left\{-\frac{1}{2}\mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}\sum_{j=1}^{n}\left(\boldsymbol{X}_j-\bar{\boldsymbol{X}}\right)\left(\boldsymbol{X}_j-\bar{\boldsymbol{X}}\right)^t\right]-\frac{1}{2}\mathrm{tr}\left[n\boldsymbol{\Sigma}^{-1}\left(\bar{\boldsymbol{X}}-\boldsymbol{\mu}\right)\left(\bar{\boldsymbol{X}}-\boldsymbol{\mu}\right)^t\right]\right\}
$$

Substituindo os valores de $\hat{\boldsymbol{\mu}}$ e $\hat{\boldsymbol{\Sigma}}$ na função de verossimilhança, encontramos seu máximo:

$$
\begin{aligned}
\boldsymbol{L}_\Omega(\boldsymbol{X};\hat{\boldsymbol{\mu}},\hat{\boldsymbol{\Sigma}}) &= (2\pi)^{-\frac{np}{2}}|\boldsymbol{S}_n|^{-\frac{n}{2}}\exp\left\{\frac{-\mathrm{tr}\left[\boldsymbol{n}\boldsymbol{S}_n^{-1}S_n\right]-\mathrm{tr}\left[n\boldsymbol{S}_n^{-1}(\bar{\boldsymbol{X}}-\bar{\boldsymbol{X}})(\bar{\boldsymbol{X}}-\bar{\boldsymbol{X}})^t\right]}{2}\right\}\\
&= (2\pi)^{-\frac{np}{2}}|\boldsymbol{S}_n|^{-\frac{n}{2}}\exp\left\{-\frac{np}{2}\right\}
\end{aligned}
$$

em que Ω representa o espaço paramétrico irrestrito.

■