MULTIVARIATE STATISTICS, Lesson 5.
Estimation and hypothesis testing in multivariate normal model

- *Definition*: Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be i.i.d. sample from a p-variate distribution with parameter (vector) $\underline{\theta} \in \mathbb{R}^k$. The expectation $\mathbf{m}$ and covariance matrix $\mathbf{C} > 0$ of $\mathbf{X}_1$ exist. The **Fisher-information matrix** of the above sample is the $k \times k$ (symmetric, positive definite) matrix $\mathbf{I}_n(\underline{\theta})$ that – under general regularity conditions – is equal to $n\mathbf{I}_1(\underline{\theta})$, where
$$\mathbf{I}_1(\underline{\theta}) = \mathbb{E}_{\underline{\theta}} \left(\frac{\partial}{\partial \underline{\theta}} \ln f_{\underline{\theta}}(\mathbf{X}_1) \right) \left(\frac{\partial}{\partial \underline{\theta}} \ln f_{\underline{\theta}}(\mathbf{X}_1) \right)^T = \mathbb{D}^2_{\underline{\theta}} \left(\frac{\partial}{\partial \underline{\theta}} \ln f_{\underline{\theta}}(\mathbf{X}_1) \right)$$
is the Fisher-information matrix of the 1-element sample, and f is the p.d.f. of the underlying p-variate distribution. Remark that $\mathbb{E} \left(\frac{\partial}{\partial \underline{\theta}} \ln f_{\underline{\theta}}(\mathbf{X}_1) \right) = \mathbf{0} \in \mathbb{R}^k$ (under regularity conditions).

- **Theorem (Cramér–Rao Inequality)**: Let $\mathbf{T} = \mathbf{T}(\mathbf{X}_1, \ldots, \mathbf{X}_n) \in \mathbb{R}^k$ be unbiased estimator of the parameter vector $\underline{\theta} \in \mathbb{R}^k$ based on the above sample. Under the usual regularity conditions, for the covariance matrix of $\mathbf{T}$ the following inequality holds:
$$\mathbb{D}^2_{\underline{\theta}}(\mathbf{T}) \geq \frac{1}{n} \mathbf{I}_1^{-1}(\underline{\theta}) = \mathbf{I}_n^{-1}(\underline{\theta}), \qquad \forall \underline{\theta} \in \Theta.$$
($\mathbf{A} \geq \mathbf{B}$ means that $\mathbf{A} - \mathbf{B}$ is positive semidefinite, and equality holds if and only if $\mathbf{A} = \mathbf{B}$.)

- *Remark*: If equality is attained by an unbiased $\mathbf{T}$, than $\mathbf{T}$ also provides an efficient estimator for $\underline{\theta}$. However, $\mathbf{T}$ may be an efficient estimator (unique with prob. 1) even if it does not reach the information bound, e.g., in the multivariate Gaussian case. If $\mathbf{T}$ is unbiased estimator for $\underline{\theta}$, further it is a sufficient and complete statistic ($\mathbb{E}_{\underline{\theta}}(g(\mathbf{T})) = \mathbf{0}, \forall \underline{\theta} \Longrightarrow g = \mathbf{0}$ a.s.), then it is also efficient. This is an easy consequence of the **Rao–Blackwell–Kolmogorov Theorem** that works in the same way for multivariate distributions.

I. Testing the multivariate normal mean in case of known covariance matrix

1. *1-sample case*: Let $\mathbf{X}_1, \ldots, \mathbf{X}_n \sim \mathcal{N}_p(\mathbf{m}, \mathbf{C})$ be i.i.d. sample with $n > p$ and $\mathbf{C} > 0$ known. For testing
$$H_0 \,:\, \mathbf{m} = \mathbf{m}_0 \quad \text{versus} \quad H_1 \,:\, \mathbf{m} \neq \mathbf{m}_0$$
the statistic
$$U_1 = (\bar{\mathbf{X}} - \mathbf{m}_0)^T \left(\frac{\mathbf{C}}{n} \right)^{-1} (\bar{\mathbf{X}} - \mathbf{m}_0) = n(\bar{\mathbf{X}} - \mathbf{m}_0)^T \mathbf{C}^{-1} (\bar{\mathbf{X}} - \mathbf{m}_0)$$
is used that follows $\chi^2(p)$-distribution under H_0 (generalization of the u-test).

2. *2-sample case*: Let $\mathbf{X}_1, \ldots, \mathbf{X}_n \sim \mathcal{N}_p(\mathbf{m}_1, \mathbf{C}_1)$ and $\mathbf{Y}_1, \ldots, \mathbf{Y}_m \sim \mathcal{N}_p(\mathbf{m}_2, \mathbf{C}_2)$ be i.i.d. samples ($\mathbf{X}_i$ is not necessarily identically distributed with $\mathbf{Y}_j$, but they are independent $\forall i, j$). Suppose that $n, m > p$ and $\mathbf{C}_1 > 0$, $\mathbf{C}_2 > 0$ are known covariance matrices. For testing
$$H_0 \,:\, \mathbf{m}_1 = \mathbf{m}_2 \quad \text{versus} \quad H_1 \,:\, \mathbf{m}_1 \neq \mathbf{m}_2$$
the statistic
$$U_2 = (\bar{\mathbf{X}} - \bar{\mathbf{Y}})^T \left(\frac{\mathbf{C}_1}{n} + \frac{\mathbf{C}_2}{m} \right)^{-1} (\bar{\mathbf{X}} - \bar{\mathbf{Y}}) = (\bar{\mathbf{X}} - \bar{\mathbf{Y}})^T \left(\frac{m\mathbf{C}_1 + n\mathbf{C}_2}{nm} \right)^{-1} (\bar{\mathbf{X}} - \bar{\mathbf{Y}})$$
is used that follows $\chi^2(p)$-distribution under H_0. In the special case $\mathbf{C}_1 = \mathbf{C}_2 = \mathbf{C}$:
$$U_2 = \frac{nm}{n+m} (\bar{\mathbf{X}} - \bar{\mathbf{Y}})^T \mathbf{C}^{-1} (\bar{\mathbf{X}} - \bar{\mathbf{Y}}) = \frac{nm}{n+m} \cdot D^2(\mathbf{X}, \mathbf{Y}),$$
where D^2 denotes the **Mahalanobis-distance** between the two populations.