MULTIVARIATE STATISTICS, Lesson 4.
ML-estimation of the multivariate normal parameters and the Wishart-distribution

- *Definition*: the $p \times p$ random matrix $\mathbf{W}$ is a (centered) **Wishart-matrix** if it is of the form $\mathbf{W} = \mathbf{X}\mathbf{X}^T$, where the column vectors of the $p \times n$ random matrix $\mathbf{X}$ are i.i.d. $\mathcal{N}_p(\mathbf{0}, \mathbf{C})$-vectors. In other words, the joint distribution of the entries of $\mathbf{W}$ is **Wishart-distribution** with parameters p (dimension), n (degrees of freedom), and $\mathbf{C}$ (covariance matrix). Notation: $\mathbf{W} \sim \mathcal{W}_p(n, \mathbf{C})$.($\mathbf{W}$ is symmetric, positive semidefinite.)
- *Remarks*:
  1. Because of its symmetry $\mathbf{W}$ follows, in fact, a $p(p+1)/2$-dimensional distribution.
  2. Denoting the column vectors of $\mathbf{X}$ by $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$, $\quad \mathbf{W} = \sum_{k=1}^n \mathbf{X}_k \mathbf{X}_k^T$.
  3. If $\mathbf{C} > 0$ and $n > p$, then $\mathbf{W} > 0$ (positive definite) with probability 1.
  4. The $n \times p$ matrix $\mathbf{X}^T$ is called **data matrix**.
  5. The $\mathcal{W}_p(n, \mathbf{I}_p)$-distribution is called *standard Wishart-distribution.* In case of $p = 1$ it is the $\chi^2(n)$-distribution.
- *Standardization*: Let $\mathbf{C} > 0$ be symmetric, positive definite. Then $\mathbf{W} \sim \mathcal{W}_p(n, \mathbf{C})$ holds if and only if $\mathbf{C}^{-1/2}\mathbf{W}\mathbf{C}^{-1/2} \sim \mathcal{W}_p(n, \mathbf{I}_p)$.
- *Additivity*: If $\mathbf{W}_1 \sim \mathcal{W}_p(n, \mathbf{C})$ and $\mathbf{W}_2 \sim \mathcal{W}_p(m, \mathbf{C})$ are independent, then $\mathbf{W}_1 + \mathbf{W}_2 \sim \mathcal{W}_p(n + m, \mathbf{C})$.
- **Theorem (Lukács)**: Let $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n \sim \mathcal{N}_p(\mathbf{m}, \mathbf{C})$ i.i.d. sample, further

$$\bar{\mathbf{X}} := \frac{1}{n}\sum_{k=1}^n \mathbf{X}_k \quad \text{and} \quad \mathbf{S} := \sum_{k=1}^n (\mathbf{X}_k - \bar{\mathbf{X}})(\mathbf{X}_k - \bar{\mathbf{X}})^T. \qquad \text{Then}$$

  1. $\bar{\mathbf{X}} \sim \mathcal{N}_p(\mathbf{m}, \frac{1}{n}\mathbf{C})$,
  2. $\mathbf{S} \sim \mathcal{W}_p(n-1, \mathbf{C})$,
  3. $\bar{\mathbf{X}}$ és $\mathbf{S}$ are (stochastically) independent.
- *Definition*: $\mathbf{S}/n$ is the **empirical**, while $\mathbf{S}/(n-1)$ is the **corrected empirical covariance matrix** based on the above i.i.d. sample.
- **ML-estimation** based on the $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n \sim \mathcal{N}_p(\mathbf{m}, \mathbf{C})$ i.i.d. sample: $\hat{\mathbf{m}} = \bar{\mathbf{X}}$, $\hat{\mathbf{C}} = \mathbf{S}/n$. It follows from the following form of the likelihood function:

$$L_{\mathbf{m},\mathbf{C}}(\mathbf{X}_1, \dots, \mathbf{X}_n) = \frac{1}{(2\pi)^{np/2}|\mathbf{C}|^{n/2}} e^{-\frac{1}{2}\operatorname{tr}\mathbf{C}^{-1}\mathbf{S}} \cdot e^{-\frac{1}{2}n(\bar{\mathbf{X}}-\mathbf{m})^T\mathbf{C}^{-1}(\bar{\mathbf{X}}-\mathbf{m})}$$

- **Theorem**: The density of the standard Wishart matrix $\mathbf{W} \sim \mathcal{W}_p(\mathbf{0}, \mathbf{I}_p)$ and that of its eigenvalues is

$$c_{np}|\mathbf{W}|^{\frac{n-p-1}{2}} e^{-\frac{1}{2}\operatorname{tr}\mathbf{W}} \quad \text{and} \quad \kappa_{np}(\prod_{j=1}^p \lambda_j)^{\frac{n-p-1}{2}} e^{-\frac{1}{2}\sum_{j=1}^p \lambda_j} \prod_{j \neq k} |\lambda_j - \lambda_k|,$$

  where the normalizing constants c_{np} and κ_{np} only depend on p and n ($n > p$).
- **Theorem**: The density of the Wishart-matrix $\mathbf{W} \sim \mathcal{W}_p(\mathbf{0}, \mathbf{C})$ is

$$c_{np}|\mathbf{W}|^{\frac{n-p-1}{2}} |\mathbf{C}|^{-\frac{n}{2}} e^{-\frac{1}{2}\operatorname{tr}\mathbf{C}^{-1}\mathbf{W}}.$$