MULTIVARIATE STATISTICS, Lesson 1.
Multivariate normal distribution

- **Definition.** The random vector $\mathbf{Y}$ has *p-dimensional standard normal distribution*, if its components are i.i.d. standard Gaussian variables. The p.d.f. of $\mathbf{Y} \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}_p)$ is

$$g(\mathbf{y}) = \prod_{i=1}^{p} \phi(y_i) = \frac{1}{\sqrt{2\pi}^p} e^{-(\sum_{i=1}^{p} y_i^2)/2} = \frac{1}{(2\pi)^{p/2}} e^{-\|\mathbf{y}\|^2/2}$$

where ϕ is the standard Gaussian density and $\mathbf{y} = (y_1, \dots, y_p)^T$.

- **Definition.** The random vector $\mathbf{X} = \mathbf{A}\mathbf{Y} + \mathbf{m}$ has *p-dimensional normal distribution*, where $\mathbf{A}$ is $p \times p$ matrix and $\mathbf{m} \in \mathbb{R}^p$. If $\mathbf{A}$ is singular then the p-dimensional distribution is degenerated (the same holds for rectangular $\mathbf{A}$).

- **Remark.** $\mathbb{E}(\mathbf{X}) = \mathbf{m}$ and the covariance matrix of $\mathbf{X}$ is

$$\mathbf{C} = \mathbb{D}^2\mathbf{X} = \mathbb{E}(\mathbf{X}-\mathbf{m})(\mathbf{X}-\mathbf{m})^T = \mathbb{E}(\mathbf{A}\mathbf{Y})(\mathbf{A}\mathbf{Y})^T = \mathbb{E}(\mathbf{A}\mathbf{Y}\mathbf{Y}^T\mathbf{A}^T) = \mathbf{A}\mathbb{E}(\mathbf{Y}\mathbf{Y}^T)\mathbf{A}^T = \mathbf{A}\mathbf{A}^T.$$

If $\mathbf{C}$ is invertible, the p.d.f. of $\mathbf{X} \sim \mathcal{N}_p(\mathbf{m}, \mathbf{C})$ is

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{p/2}|\mathbf{C}|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\mathbf{m})^T\mathbf{C}^{-1}(\mathbf{x}-\mathbf{m})}, \quad \mathbf{x} \in \mathbb{R}^p.$$

1. **Proposition.** The components of $\mathbf{X} \sim \mathcal{N}_p(\mathbf{m}, \mathbf{C})$ are (completely) independent if and only if $\mathbf{C}$ is diagonal.
2. Find the linear transformation that back-transforms $\mathbf{X} \sim \mathcal{N}_p(\mathbf{m}, \mathbf{C})$ into a p-dimensional standard normal vector.
3. Find the level surfaces of the p.d.f. of $\mathbf{X} \sim \mathcal{N}_p(\mathbf{m}, \mathbf{C})$.
4. Prove that the level surfaces are spheres if and only if the components of $\mathbf{X}$ are independent with equal variances.
5. **Theorem.** If the covariance matrix $\mathbf{C}$ of $\mathbf{X} \sim \mathcal{N}_p(\mathbf{m}, \mathbf{C})$ is positive definite then
$$(\mathbf{X}-\mathbf{m})^T\mathbf{C}^{-1}(\mathbf{X}-\mathbf{m}) \sim \chi^2(p).$$
6. Prove that the characteristic function of $\mathbf{Y} \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}_p)$ is
$$\psi_{\mathbf{Y}}(\mathbf{t}) = \mathbb{E}(e^{i\mathbf{Y}^T\mathbf{t}}) = e^{-\|\mathbf{t}\|^2/2}, \quad \mathbf{t} \in \mathbb{R}^p.$$
7. Prove that the characteristic function of $\mathbf{X} \sim \mathcal{N}_p(\mathbf{m}, \mathbf{C})$ is
$$\psi_{\mathbf{X}}(\mathbf{t}) = \mathbb{E}(e^{i\mathbf{X}^T\mathbf{t}}) = e^{i\mathbf{m}^T\mathbf{t} - \frac{1}{2}\mathbf{t}^T\mathbf{C}\mathbf{t}}, \quad \mathbf{t} \in \mathbb{R}^p,$$
where i is the imaginary unit.
8. **Proposition.** The random vector $\mathbf{X}$ has multivariate normal distribution if and only if any linear combination of its components has one-dimensional normal distribution.
9. **Proposition.** If X and Y are independent and $X + Y$ is normally distributed, then X és Y are also normally distributed.
10. Prove that any subset of the components of $\mathbf{X} \sim \mathcal{N}_p(\mathbf{m}, \mathbf{C})$ has also multivariate normal distribution. Find the parameters of this distribution!