Exercise 2.7.5. *Let $X_1, X_2 \sim N_m(\mu, \Sigma)$ with $X_1 \perp\!\!\!\perp X_2$, and define $Y = X_1 + X_2$. Find the distribution of Y.*

Exercise 2.7.10. *Let $X_1, X_2 \sim N_m(\mu, \Sigma)$ be independent. Consider the random vector $Z = X_1 + X_2$. Prove that the covariance matrix of Z is 2Σ.*

Exercise 2.7.16. *Suppose $X_1, X_2 \sim N_m(0, \Sigma)$ are independent and $Y = X_1 - X_2$. Derive the distribution of Y. How does the covariance structure of Y compare to that of X_1 and X_2 individually?*

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N_{2m}\left(\begin{pmatrix} \mu \\ \mu \end{pmatrix}, \begin{pmatrix} \Sigma & 0 \\ 0 & \Sigma \end{pmatrix} \right), \quad A \underbrace{\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}}_{\in \mathbb{R}^{2m}} = \underbrace{X_1 + X_2}_{\in \mathbb{R}^m}$$

$$A = \begin{pmatrix} I_m & I_m \end{pmatrix} \in \mathbb{R}^{m \times 2m}, \quad A \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = X_1 + X_2$$

$$X_1 + X_2 \sim N_m \left(A \begin{pmatrix} \mu \\ \mu \end{pmatrix}, A \begin{pmatrix} \Sigma & 0 \\ 0 & \Sigma \end{pmatrix} A^T \right) = N_m(2\mu, 2\Sigma)$$

$$A = \begin{pmatrix} I_m & -I_m \end{pmatrix} \qquad A \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = X_1 - X_2$$

$$A \begin{pmatrix} \mu \\ \mu \end{pmatrix} = \mu - \mu = 0, \quad A \begin{pmatrix} \Sigma & 0 \\ 0 & \Sigma \end{pmatrix} A^T = \Sigma + \Sigma = 2\Sigma$$

$$X_1 - X_2 \sim N_m(0, 2\Sigma)$$

Exercise 2.7.9. *Let $X \sim N_m(0, \sigma^2 I_m)$ and let U be any orthogonal matrix. Show that the distribution X is the same as the distribution of UX. Would the answer be the same if the mean of X was not zero?*

UX , U is a linear transformation , U orthogonal : $U^T U = U U^T = I_m$.

$\mathbb{E}[UX] = U\,\mathbb{E}[X] = U \cdot 0 = 0$

$\mathrm{Cov}(UX, UX) = U\,\mathrm{Cov}(X, X)\,U^T = U \sigma^2 I_m U^T = \sigma^2 I_m$

$UX \overset{d}{=} X \longrightarrow$ "spherical distribution".

Exercise 2.7.11. *Suppose* $X \sim N_m(\mu, \Sigma)$*. Let* $X_{\setminus 1}$ *denote* X *with the first entry removed. Consider the linear regression of* X_1 *on the remaining variables*

$$X_1 = \boldsymbol{w}^\top X_{\setminus 1} + \varepsilon \qquad \textit{with} \quad \varepsilon \perp\!\!\!\perp X_{\setminus i}.$$

Show that the linear regression coefficients $\boldsymbol{w}$ *can be expressed in terms of the blocks of the covariance matrix* Σ*. Hint: What is the relationship between these coefficients and conditional variances?*

$$X \sim N_m(\mu, \Sigma) = N_m\left(\begin{pmatrix}\mu_1 \\ \mu_{\setminus 1}\end{pmatrix}, \begin{pmatrix}\Sigma_{11} & \Sigma_{1,\setminus 1} \\ \Sigma_{\setminus 1,1} & \Sigma_{\setminus 1,\setminus 1}\end{pmatrix}\right)$$

$$X_1 \mid X_{\setminus 1} = x \qquad z = X_1 + \Sigma_{1,\setminus 1}\Sigma_{\setminus 1,\setminus 1}^{-1}(x - X_{\setminus 1})$$

$$\mathbb{E}[z] = \mu_1 + \Sigma_{1,\setminus 1}\Sigma_{\setminus 1,\setminus 1}^{-1}(x - \mu_{\setminus 1})$$

$$\mathrm{Cov}(z, z) = \Sigma_{1,1} - \Sigma_{1,\setminus 1}\Sigma_{\setminus 1,\setminus 1}^{-1}\cancel{\Sigma_{\setminus 1,\setminus 1}}\cancel{\Sigma_{\setminus 1,\setminus 1}^{-1 \top}}\Sigma_{1,\setminus 1}^{\top}$$

$$= \Sigma_{1,1} - \Sigma_{1,\setminus 1}\Sigma_{\setminus 1,\setminus 1}^{-1}\Sigma_{\setminus 1,1}$$

$$Y = \beta_0 + \beta X$$

$$w = (\beta_0, \beta), \quad \beta = \Sigma_{1,\setminus 1}\Sigma_{\setminus 1,\setminus 1}^{-1}$$

$$\beta = S_{xy} S_x^{-1}$$

Exercise 2.7.15. *Prove Lemma 2.2.2.*

Lemma 2.2.2. *Suppose $X \sim N_m(\mu, \Sigma)$. Then $AX \perp\!\!\!\perp BX$ if and only if $A\Sigma B^\top = 0$.*

Z, Y jointly Gaussian

$AX, BX, \quad Z \perp\!\!\!\perp Y, \quad \longleftrightarrow \quad \mathrm{Cov}(Z, Y) = 0$

$Y \in \{-1, 1\}$ with equal probability, $X \sim N(0,1)$

$\mathrm{Cov}(Y \cdot X, X) = 0.$

$\mathrm{Cov}(AX, BX) = A\,\mathrm{Cov}(X, X) B^\top = A\Sigma B^\top = 0.$

Exercise 2.7.7. *Let $X \sim N_m(\mu, \Sigma)$. Use the spectral decomposition of Σ to transform X into independent standard normal variables.*

Exercise 2.7.3. *Suppose we want to generate a sample of n observations from $N_m(0, \Sigma)$. Using R or Python, write a code that does it by sampling independently a bunch of univariate $N(0,1)$ variables and transforming them approprietly.*

Σ symmetric, $\Sigma = \Sigma^T$, $\Sigma = U \Lambda U^T$

$\Lambda = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$, $\lambda_i > 0$, $\underline{\Sigma^{1/2} = U \Lambda^{1/2}}$

$$\Sigma^{1/2} \cdot (\Sigma^{1/2})^T = U \Lambda^{1/2} (U \Lambda^{1/2})^T = U \Lambda^{1/2} \Lambda^{1/2} U^T = \Sigma$$

$$\underline{z = \Sigma^{-1/2}}(X - \mu)$$

$$\mathbb{E}[z] = \Sigma^{-1/2} \mathbb{E}[X - \mu] = 0$$

$$\mathrm{Cov}(z, z) = \Sigma^{-1/2} \mathrm{Cov}(X - \mu, X - \mu)(\Sigma^{-1/2})^T$$

$$= \Sigma^{-1/2} \quad \Sigma \quad \cdot (\Sigma^{-1/2})^T$$

$$= U \cancel{\Lambda^{-1/2}} U^T \cancel{\Lambda} U \cancel{\Lambda^{-1/2}} U^T = U U^T \underline{U U^T = I_m}.$$

$\Longrightarrow z \sim N_m(0, I_m)$

$X = \mathrm{rnorm}(m)$ $\quad \Sigma \longrightarrow \mathrm{chol}(\Sigma)$, $\quad \mu + \Sigma^{1/2} z = \underline{X}$.

Exercise 2.7.17. *Let $X \sim N_m(\mu, \Sigma)$. Show that for any $a \in \mathbb{R}^m$, the probability $\mathbb{P}(a^\top X > c)$ depends on both $a^\top \mu$ and $a^\top \Sigma a$. Derive a formula for this probability in terms of the c.d.f. of the standard normal distribution.*

$X \sim N_m(\mu, \Sigma)$, density $f(x) = (2\pi)^{-m/2} |\Sigma|^{-1/2} \exp\left(-\frac{1}{2} \underbrace{(x-\mu)^T \Sigma^{-1} (x-\mu)}\right)$

$(x-\mu)^T \Sigma^{-1} (x-\mu) = c$ "Mahanalobis distance"

$\mathbb{P}(a^T X > c)$, $a^T X \sim N(a^T\mu, \underline{a^T \Sigma a})$

$Z = \dfrac{a^T X - a^T \mu}{\sqrt{a^T \Sigma a}}$ $\mathbb{P}\left(Z > \dfrac{c - a^T\mu}{\sqrt{a^T \Sigma a}}\right)$, $Z \sim N(0,1)$

Exercise 2.7.18. *Consider the functions $f, g : \mathbb{R}^n \to \mathbb{R}$ given by $f(\mathbf{x}) = \boldsymbol{a}^\top \mathbf{x}$ and $g(\mathbf{x}) = \mathbf{x}^\top A\mathbf{x}$ for $\boldsymbol{a} \in \mathbb{R}^n$ and $A \in \mathbb{S}^n$. Show that $\nabla f(\mathbf{x}) = \boldsymbol{a}$ and $\nabla g(\mathbf{x}) = 2A\mathbf{x}$.*

$$\nabla f = \begin{pmatrix} \partial f/\partial x_1 \\ \vdots \\ \partial f/\partial x_n \end{pmatrix}, \quad f(x) = a^T x = \sum_{i=1}^{n} a_i x_i$$

$$\frac{\partial f}{\partial x_i} = a_i, \quad \nabla f(x) = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} = a$$

$$\nabla(a^T x) = \nabla(x^T a) = a.$$

$$g(x) = x^T A x = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} x_i x_j$$

$$\frac{\partial g}{\partial x_k} = A_{ik} x_i + A_{kj} x_j = A_{ik} x_i + A^T_{jk} x_j$$

$$\nabla g = Ax + A^T x \overset{A \in \mathbb{S}^n}{=} 2Ax.$$