Multivariate distributions

Question 1: Data and distributional assumptions

(a) Let us assume that we are given data *with all metric columns* of the following form:

$$\begin{array}{c|cccc} & M_1 & M_2 & \cdots & M_m \\ \hline 1: & x_{11} & x_{12} & \cdots & x_{1n} \\ 2: & x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ n: & x_{n1} & x_{n2} & \cdots & x_{nn} \end{array} .$$

How would, i.e. as what mathematical objects and using which probabilistic assumptions, would we model the elements of this data to then be able to make inferences about the behaviour/characteristics of new row-wise observations, like $[x_{(n+1)1}, x_{(n+1)2}, \dots, x_{(n+1)m}]$?

(b) Show that the arithmetic mean is an unbiased estimate of the expected value, *given that we are viewing the points we are averaging over as realizations of random variables whose distributions all have the same expected value.*

Do we additionally need to assume that the random variables of which we have realizations are i.i.d.? Explain your answer.

(c) Given the setting of (a), consider the case $m = 1$, i.e. that we only have the data of column M_1, but are otherwise making the same modelling choices. Show that

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(x_i - \frac{1}{n} \sum_{i=1}^{n} x_i \right)^2$$

is an unbiased estimate for the variance of the distribution we are assuming.

Why are we looking only at the case of $m = 1$ here instead of considering the x_is to be vectors in $\mathbb{R}^m$ in the above equation?

Recap:

An estimate $\hat{\theta}$ for a fixed, "true" value θ_0 is called *unbiased*, if the following holds:

$$\mathbb{E}\left[\hat{\theta}\right] = \theta_0 .$$

Solution:

(a) We assume that there are n i.i.d. copies of the random vector $\boldsymbol{X} = \begin{pmatrix} M_1 \\ M_2 \\ \vdots \\ M_m \end{pmatrix}$ and write

$$\boldsymbol{X}_1, \boldsymbol{X}_2, \dots, \boldsymbol{X}_n \overset{i.i.d.}{\sim} \mathcal{D},$$

where $\mathcal{D}$ is some probability distribution.
Then, we view the row $[x_{i1}, x_{i2}, \dots, x_{im}]$ as a realization of $\boldsymbol{X}_i$.

Next we can make different choices about the assumption we make about $\mathcal{D}$, for example that it belongs to a known distribution family and we just need to estimate the parameters.

Either way, it is usually sensible to assume that $\mathbb{E}[\boldsymbol{Z}], \mathrm{Cov}(\boldsymbol{Z})_{ij} < \infty$ for $\boldsymbol{Z} \sim \mathcal{D}$.

(b) Consider the random variable $\boldsymbol{Z}$ with $\mathbb{E}[\boldsymbol{Z}] = \mu$ and n random variables $X_1, \dots, X_n$ with $\mathbb{E}[X_i] = \mu \; \forall i \in \{1, \dots, n\}$.

$$\begin{aligned} \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] &= \frac{1}{n}\mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[X_i\right] \\ &= \frac{n}{n}\mathbb{E}\left[X_i\right] \\ &= \mu = \mathbb{E}[\boldsymbol{Z}] \quad \checkmark \end{aligned}$$

No, we do not have to make the i.i.d. assumption here. In fact, by the linearity of the expected value, it is completely sufficient to assume that are random variables have the same expectation, even if they follow different distributions!

(c) Consider the random variable $\boldsymbol{Z}$ with **realizations in** $\mathbb{R}$ and $\boldsymbol{Z} \sim \mathcal{D}$ and n random variables $X_1, \dots, X_n$ with **realizations in** $\mathbb{R}$ and $X_1, \dots, X_n \overset{i.i.d.}{\sim} \mathcal{D}$.

$$
\begin{aligned}
\mathbb{E}\left[\hat{\sigma}^2\right] =& \mathbb{E}\left[\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \frac{1}{n}\sum_{j=1}^{n} X_j\right)^2\right] = \frac{n}{n-1}\mathbb{E}\left[\left(X_i - \frac{1}{n}\sum_{j=1}^{n} X_j\right)^2\right] \\
=& \frac{n}{n-1}\mathbb{E}\left[X_i^2 - 2X_i \cdot \frac{1}{n}\sum_{j=1}^{n} X_j + \left(\frac{1}{n}\sum_{j=1}^{n} X_j\right)^2\right] \\
=& \frac{n}{n-1}\mathbb{E}\left[X_i^2 - 2X_i \cdot \frac{1}{n}\sum_{j=1}^{n} X_j + \frac{1}{n^2}\left(\sum_{j=1}^{n} X_j^2 + \sum_{k=1}^{n}\sum_{l=1}^{n} X_k X_l \mathbb{1}_{\{j \neq k\}}\right)\right] \\
=& \frac{n}{n-1}\left(\mathbb{E}\left[X_i^2\right] - \mathbb{E}\left[2X_i \cdot \frac{1}{n}\sum_{j=1}^{n} X_j\right] + \mathbb{E}\left[\frac{1}{n^2}\left(\sum_{j=1}^{n} X_j^2 + \sum_{k=1}^{n}\sum_{l=1}^{n} X_k X_l \mathbb{1}_{\{j \neq k\}}\right)\right]\right) \\
=& \frac{n}{n-1}\left(\mathbb{E}\left[X_i^2\right] - \frac{2}{n}\mathbb{E}\left[X_i^2\right] - \frac{2(n-1)}{n}\mathbb{E}\left[X_i\right]^2 + \frac{1}{n}\mathbb{E}\left[X_i^2\right] + \frac{n(n-1)}{n}\mathbb{E}\left[X_i\right]^2\right) \\
=& \frac{n}{n-1}\left(\frac{(n-1)}{n}\mathbb{E}\left[X_i^2\right] - \frac{(n-1)}{n}\mathbb{E}\left[X_i\right]^2\right) \\
=& \mathbb{E}\left[X_i^2\right] - \mathbb{E}\left[X_i\right]^2 \\
=& \operatorname{Var}(X_i) = \operatorname{Var}(\boldsymbol{Z}) \quad \checkmark
\end{aligned}
$$

We are only looking at the case of $m = 1$, because there is no one variance value for random vectors. Instead, they have covariance matrices. The diagonal entries of this covariance matrices will be the variances of the elements of the random vector.

Question 2: Eigenvalue decomposition

Consider the random vector $\boldsymbol{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2)^T$ with covariance

$$
\boldsymbol{\Sigma}_{\boldsymbol{x}} = \begin{pmatrix} 2 & 2 \\ 2 & 5 \end{pmatrix}
$$

a) Determine the eigenvalues λ_1 and λ_2 and the (normalized) eigenvectors of the matrix $\boldsymbol{\Sigma}_{\boldsymbol{x}}$.

b) Use the result of (a) to determine a random vector $\boldsymbol{y} = (\boldsymbol{y}_1, \boldsymbol{y}_2)^T$, whose components $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ are linear combinations of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ and for which additionally holds that

$$
\operatorname{Cov}(\boldsymbol{y}) = \boldsymbol{\Lambda} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}.
$$

Solution:

a)

$$
\boldsymbol{\Sigma}_{\boldsymbol{x}} = \begin{pmatrix} 2 & 2 \\ 2 & 5 \end{pmatrix}
$$

The characteristic polynomial is given by:

$$det(\boldsymbol{\Sigma_x} - \lambda \boldsymbol{I}) = det \begin{pmatrix} 2-\lambda & 2 \\ 2 & 5-\lambda \end{pmatrix} = (2-\lambda)(5-\lambda) - 4 \overset{!}{=} 0$$

$$\Rightarrow \lambda^2 - 7\lambda + 6 = (\lambda - 1)(\lambda - 6) \overset{!}{=} 0$$

$$\Rightarrow \lambda_1 = 1, \lambda_2 = 6$$

As covariance matrix of $\boldsymbol{x} = (x_1, x_2)^T$, $\boldsymbol{\Sigma_x}$ is covariance matrix, since $\boldsymbol{\Sigma_x}$ is

(i) symmetric

(ii) positive definite.

Calculate the eigenvalues:

$$(\boldsymbol{\Sigma_x} - \lambda_1 \boldsymbol{I}) \cdot \boldsymbol{v} = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} \cdot \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} v_1 + 2v_2 \\ 2v_1 + 4v_2 \end{pmatrix} \overset{!}{=} \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\Rightarrow \quad \text{set } v_2 = 1 \Rightarrow v_1 = -2$$

$$\text{length of } \boldsymbol{v}: \; \sqrt{(-2)^2 + 1^2} = \sqrt{5}$$

$$\Rightarrow \quad \widetilde{\boldsymbol{v}} = \frac{1}{\sqrt{5}} \begin{pmatrix} -2 \\ 1 \end{pmatrix}$$

$$(\boldsymbol{\Sigma_x} - \lambda_2 \boldsymbol{I}) \cdot \boldsymbol{w} = \begin{pmatrix} -4 & 2 \\ 2 & -1 \end{pmatrix} \cdot \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} -4w_1 + 2w_2 \\ 2w_1 - 1w_2 \end{pmatrix} \overset{!}{=} \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\Rightarrow \quad \text{set } w_1 = 1 \Rightarrow w_2 = 2$$

$$\text{length of } \boldsymbol{w}: \; \sqrt{1^2 + 2^2} = \sqrt{5}$$

$$\Rightarrow \quad \widetilde{\boldsymbol{w}} = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

Let's check:

(1) Orthogonality:

$$< \widetilde{\boldsymbol{v}}, \widetilde{\boldsymbol{w}} > = -\frac{2}{\sqrt{5}} \frac{1}{\sqrt{5}} + \frac{1}{\sqrt{5}} \frac{2}{\sqrt{5}} = 0 \quad \surd$$

(2) Eigendecomposition:

$$\boldsymbol{P \Lambda P}^T = \frac{1}{\sqrt{5}} \begin{pmatrix} -2 & 1 \\ 1 & 2 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 \\ 0 & 6 \end{pmatrix} \cdot \frac{1}{\sqrt{5}} \begin{pmatrix} -2 & 1 \\ 1 & 2 \end{pmatrix} \overset{(NR)}{=} \begin{pmatrix} 2 & 2 \\ 2 & 5 \end{pmatrix} = \boldsymbol{\Sigma_x} \quad \surd$$

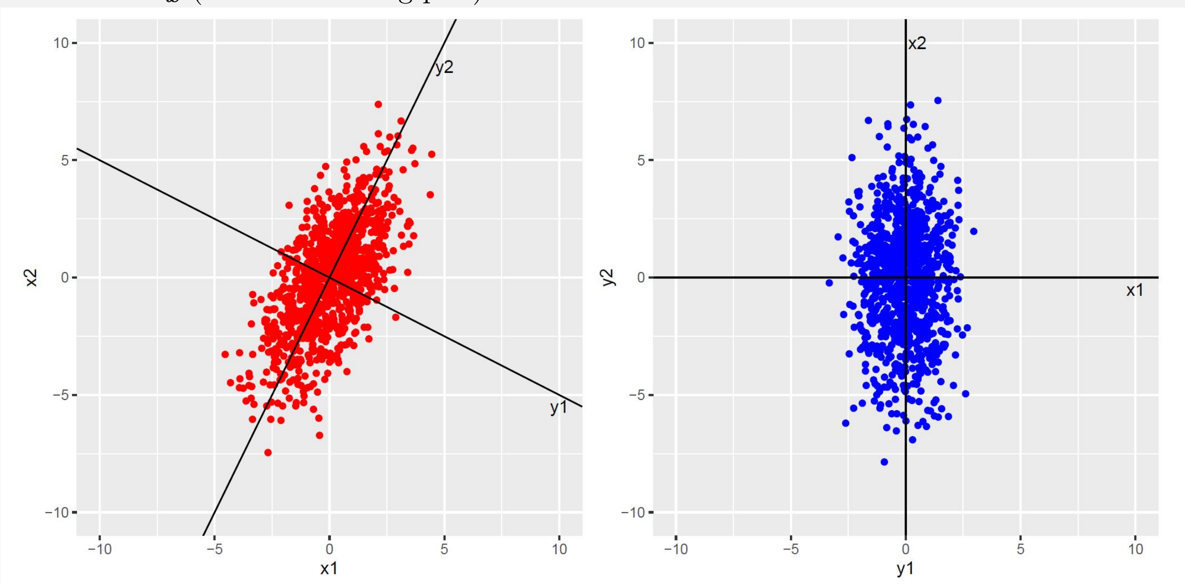
b)

$$\boldsymbol{\Sigma_y} \stackrel{!}{=} \begin{pmatrix} 1 & 0 \\ 0 & 6 \end{pmatrix} = \boldsymbol{\Lambda} = \boldsymbol{P}^T \boldsymbol{\Sigma_x} \boldsymbol{P}$$

$$\begin{aligned}
&= \boldsymbol{P}^\top \operatorname{Cov}(\boldsymbol{x}) \boldsymbol{P} = \boldsymbol{P}^\top \mathbb{E}\big[(\boldsymbol{x} - \mathbb{E}[\boldsymbol{x}])(\boldsymbol{x} - \mathbb{E}[\boldsymbol{x}])^\top\big] \boldsymbol{P} \\
&= \mathbb{E}\big[\boldsymbol{P}^\top (\boldsymbol{x} - \mathbb{E}[\boldsymbol{x}])(\boldsymbol{x} - \mathbb{E}[\boldsymbol{x}])^\top \boldsymbol{P}\big] \\
&= \mathbb{E}\big[\boldsymbol{P}^\top (\boldsymbol{x} - \mathbb{E}[\boldsymbol{x}])(\boldsymbol{x} - \mathbb{E}[\boldsymbol{x}])^\top (\boldsymbol{P}^\top)^\top\big] \\
&= \mathbb{E}\big[\boldsymbol{P}^\top (\boldsymbol{x} - \mathbb{E}[\boldsymbol{x}])(\boldsymbol{P}^\top (\boldsymbol{x} - \mathbb{E}[\boldsymbol{x}]))^\top\big] \\
&= \mathbb{E}\big[(\boldsymbol{P}^\top \boldsymbol{x} - \mathbb{E}[\boldsymbol{P}^\top \boldsymbol{x}])(\boldsymbol{P}^\top \boldsymbol{x} - \mathbb{E}[\boldsymbol{P}^\top \boldsymbol{x}])^\top\big] \\
&= \operatorname{Cov}(\boldsymbol{P}^\top \boldsymbol{x})
\end{aligned}$$

$$\Rightarrow \text{define } \boldsymbol{y} = \boldsymbol{P}^T \boldsymbol{x} = \frac{1}{\sqrt{5}} \begin{pmatrix} -2 & 1 \\ 1 & 2 \end{pmatrix} \cdot \begin{pmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{pmatrix} = \frac{1}{\sqrt{5}} \begin{pmatrix} -2\boldsymbol{x}_1 + \boldsymbol{x}_2 \\ \boldsymbol{x}_1 + 2\boldsymbol{x}_2 \end{pmatrix}$$

Note: $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ are coordinates of $\boldsymbol{x}$ with respect to the basis of the eigenvectors $\widetilde{\boldsymbol{v}}$ and $\widetilde{\boldsymbol{w}}$ of $\boldsymbol{\Sigma_x}$ (see the following plot).



Question 3: Multivariate normal distribution

Let $\boldsymbol{x} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p)^T$ be a p-dimensional multivariate-normal distributed random vector. The corresponding density is given by

$$f(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^p |\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu})\right),$$

where $\boldsymbol{\mu}$ denotes the expected value $\mathbb{E}[\boldsymbol{x}]$ and $\boldsymbol{\Sigma}$ the covariance $\operatorname{Cov}(\boldsymbol{x})$.

a) Write out the form of this density for the the case $p = 2$, using the parameters $\sigma_i^2 = var(\boldsymbol{x}_i)$, $i = 1, 2$, and $\rho = \frac{cov(\boldsymbol{x}_1, \boldsymbol{x}_2)}{\sigma_1 \sigma_2}$. Conclude from this that $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are independent if they are uncorrelated.

b) Plot the density for $\boldsymbol{\mu} = \mathbf{0}$, $\sigma_1 = 1$, $\sigma_2 = 3$ and different values of ρ using `R`. (Tip: The function `persp` in combination with the function `manipulate` from the package of the same name is well suited for this).

Recap:

<u>Calculation of the inverse of a matrix:</u>

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \Rightarrow \boldsymbol{A}^{-1} = \frac{1}{|\boldsymbol{A}|} \underbrace{\begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix}}_{Adj(\boldsymbol{A})}$$

<u>Calculation of the quadratic form:</u>

$$\boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x} = (x_1 \;\; x_2) \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = x_1^2 a_{11} + x_1 x_2 (a_{12} + a_{21}) + x_2^2 a_{22}$$

Solution:

Here, we have: $p = 2$, $\sigma_1^2 = \boldsymbol{Var}\,(x_1)$, $\sigma_2^2 = \boldsymbol{Var}\,(x_2)$, $\boldsymbol{Cov}(x_1, x_2) = \rho\sigma_1\sigma_2$

$$\Rightarrow \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

(1) The determinant is given by:

$$det(\boldsymbol{\Sigma}) = |\boldsymbol{\Sigma}| = \sigma_1^2\sigma_2^2 - \rho^2\sigma_1^2\sigma_2^2 = \sigma_1^2\sigma_2^2(1 - \rho^2)$$

(2) The inverse is given by:

$$\boldsymbol{\Sigma}^{-1} = \frac{1}{\sigma_1^2\sigma_2^2(1 - \rho^2)} \begin{pmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix} = \frac{1}{1 - \rho^2} \begin{pmatrix} \frac{1}{\sigma_1^2} & -\frac{\rho}{\sigma_1\sigma_2} \\ -\frac{\rho}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2} \end{pmatrix}$$

<u>It, therefore, follows that:</u>

$$f(\boldsymbol{x}) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1 - \rho^2}} \exp\left\{ -\frac{1}{2(1 - \rho^2)} \left[\frac{(x_1 - \mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x_1 - \mu_1)(x_2 - \mu_2)}{\sigma_1\sigma_2} + \frac{(x_2 - \mu_2)^2}{\sigma_2^2} \right] \right\}$$

Let us now assume that x_1 and x_2 are uncorrelated $\Rightarrow \;\rho = 0$

$$\begin{aligned} \overset{\rho=0}{\Longrightarrow} f(\boldsymbol{x}) &= \frac{1}{2\pi\sigma_1\sigma_2} \exp\left\{ -\frac{1}{2} \left[\left(\frac{x_1 - \mu_1}{\sigma_1} \right)^2 + \left(\frac{x_2 - \mu_2}{\sigma_2} \right)^2 \right] \right\} \\ &= \underbrace{\frac{1}{\sqrt{2\pi}\sigma_1} \exp\left\{ -\frac{1}{2} \left(\frac{x_1 - \mu_1}{\sigma_1} \right)^2 \right\}}_{=f_{\boldsymbol{x}_1}(\boldsymbol{x}_1)} \cdot \underbrace{\frac{1}{\sqrt{2\pi}\sigma_2} \exp\left\{ -\frac{1}{2} \left(\frac{x_2 - \mu_2}{\sigma_2} \right)^2 \right\}}_{=f_{\boldsymbol{x}_2}(\boldsymbol{x}_2)} \end{aligned}$$

b) *See R-Code.*

Question 4: Determining marginal distributions

Consider the random variable $\boldsymbol{z} = (\boldsymbol{y}, \boldsymbol{x})^T \sim \mathcal{N}_{q+p}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu_y} \\ \boldsymbol{\mu_x} \end{pmatrix} \quad , \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma_y} & \boldsymbol{\Sigma_{yx}} \\ \boldsymbol{\Sigma_{xy}} & \boldsymbol{\Sigma_x} \end{pmatrix}.$$

Derive the marginal distributions for the component vectors $\boldsymbol{y}$ and $\boldsymbol{x}$.

Recap:

All random vectors that result from linear transformations of normally distributed random vectors are in turn normally distributed. More specifically, the following applies to a p-dimensional random vector:

$$\boldsymbol{w} \sim \mathcal{N}_p(\boldsymbol{\mu},\ \boldsymbol{\Sigma}) \implies \underbrace{\boldsymbol{A}}_{\in\mathbb{R}^{q\times p}} \boldsymbol{w} + \underbrace{\boldsymbol{b}}_{\in\mathbb{R}^{q\times 1}} \sim \mathcal{N}_q(\boldsymbol{A\mu}+\boldsymbol{b},\ \boldsymbol{A\Sigma A}^T) \quad \text{mit } q = rg(\boldsymbol{A}) \leq p$$

Solution:

$$\boldsymbol{z} = \begin{pmatrix} \boldsymbol{y} \\ \boldsymbol{x} \end{pmatrix} \sim \mathcal{N}_{q+p}\left(\begin{pmatrix} \boldsymbol{\mu_y} \\ \boldsymbol{\mu_x} \end{pmatrix} , \begin{pmatrix} \boldsymbol{\Sigma_y} & \boldsymbol{\Sigma_{yx}} \\ \boldsymbol{\Sigma_{xy}} & \boldsymbol{\Sigma_x} \end{pmatrix}\right)$$

Distribution of $\boldsymbol{y}$:

$$\text{Define } \boldsymbol{A_y} = (\mathbf{I}_q\ \mathbf{0}_{q\times p}) = \begin{pmatrix} \underbrace{\begin{matrix} 1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & 1 \end{matrix}}_{q} & \underbrace{\begin{matrix} 0 & & 0 \\ & \ddots & \\ 0 & & 0 \end{matrix}}_{p} \end{pmatrix}$$

$$\Rightarrow \boldsymbol{A_y z} = \mathbf{I}_q \boldsymbol{y} + \mathbf{0}_{q\times p}\boldsymbol{x} = \boldsymbol{y}$$

$$\boldsymbol{y} \sim \mathcal{N}_q\left(\boldsymbol{A_y}\begin{pmatrix} \boldsymbol{\mu_y} \\ \boldsymbol{\mu_x} \end{pmatrix} , \boldsymbol{A_y}\begin{pmatrix} \boldsymbol{\Sigma_y} & \boldsymbol{\Sigma_{yx}} \\ \boldsymbol{\Sigma_{xy}} & \boldsymbol{\Sigma_x} \end{pmatrix}\boldsymbol{A_y}^T\right)$$

$$\Rightarrow \boldsymbol{A_y}\begin{pmatrix} \boldsymbol{\mu_y} \\ \boldsymbol{\mu_x} \end{pmatrix} = \mathbf{I}_q\boldsymbol{\mu_y} + \mathbf{0}_{q\times p}\boldsymbol{\mu_x} = \boldsymbol{\mu_y}$$

$$\begin{aligned} \boldsymbol{A_y \Sigma A_y}^T &= (\mathbf{I}_q\ \mathbf{0}_{q\times p})\begin{pmatrix} \boldsymbol{\Sigma_y} & \boldsymbol{\Sigma_{yx}} \\ \boldsymbol{\Sigma_{xy}} & \boldsymbol{\Sigma_x} \end{pmatrix}\begin{pmatrix} \mathbf{I}_q \\ \mathbf{0}_{p\times q} \end{pmatrix} \\ &= (\boldsymbol{\Sigma_y} + \mathbf{0}_{q\times p}\boldsymbol{\Sigma_{xy}}\ ,\ \ \boldsymbol{\Sigma_{yx}} + \mathbf{0}_{q\times p}\boldsymbol{\Sigma_x})\begin{pmatrix} \mathbf{I}_q \\ \mathbf{0}_{p\times q} \end{pmatrix} \\ &= \boldsymbol{\Sigma_y} + \mathbf{0}_{q\times p}\boldsymbol{\Sigma_{yx}} \end{aligned}$$

$$\Rightarrow \boldsymbol{y} \sim \mathcal{N}_q(\boldsymbol{\mu_y},\ \boldsymbol{\Sigma_y})$$

Distribution of x:

$$\text{Define } \boldsymbol{A_x} = (\mathbf{0}_{p\times q}\ \mathbf{I}_p)$$

And the rest follows analogously:

$$\Rightarrow \boldsymbol{x} \sim \mathcal{N}_p(\boldsymbol{\mu}_{\boldsymbol{x}},\ \boldsymbol{\Sigma}_{\boldsymbol{x}})$$