Mahalanobis' distance

Consider a population with $E\underset{\sim}{x} = \underset{\sim}{\mu}$ and $\operatorname{Var}\underset{\sim}{x} = \Sigma$

$$\Delta(\underset{\sim}{x}, \underset{\sim}{\mu}) = \sqrt{(\underset{\sim}{x}-\underset{\sim}{\mu})^T \Sigma^{-1} (\underset{\sim}{x}-\underset{\sim}{\mu})}$$

Compare with: (Euclidian dist.)
$$d(\underset{\sim}{x}, \underset{\sim}{\mu}) = \sqrt{(\underset{\sim}{x}-\underset{\sim}{\mu})^T (\underset{\sim}{x}-\underset{\sim}{\mu})} = \|\underset{\sim}{x} - \underset{\sim}{\mu}\|$$

Let $\underset{\sim}{x}_1, \dots, \underset{\sim}{x}_n$ be a sample

$$D(\underset{\sim}{x}, \underset{\sim}{\bar{x}}) = \sqrt{(\underset{\sim}{x}-\underset{\sim}{\bar{x}})^T S^{-1} (\underset{\sim}{x}-\underset{\sim}{\bar{x}})}$$

$$D(\underset{\sim}{x}_r, \underset{\sim}{x}_s) = \sqrt{(\underset{\sim}{x}_r-\underset{\sim}{x}_s)^T S^{-1} (\underset{\sim}{x}_r-\underset{\sim}{x}_s)}$$

Angle between $\underset{\sim}{x}_r$ and $\underset{\sim}{x}_s$

$$\cos\Theta = \frac{\underset{\sim}{x}_r^T S^{-1} \underset{\sim}{x}_s}{D(\underset{\sim}{x}_r, \underset{\sim}{0})\, D(\underset{\sim}{x}_s, \underset{\sim}{0})}$$

Compare with $\cos\Theta = \dfrac{\underset{\sim}{x}_r^T \underset{\sim}{x}_s}{\|\underset{\sim}{x}_r\| \, \|\underset{\sim}{x}_s\|}$

Consider two populations with equal Σ

$$\Delta(\underset{\sim}{\mu}_1, \underset{\sim}{\mu}_2) = \sqrt{(\underset{\sim}{\mu}_2-\underset{\sim}{\mu}_1)^T \Sigma^{-1} (\underset{\sim}{\mu}_2-\underset{\sim}{\mu}_1)}$$

$$D(\underset{\sim}{\bar{x}}_1, \underset{\sim}{\bar{x}}_2) = \sqrt{(\underset{\sim}{\bar{x}}_1-\underset{\sim}{x}_2)^T S_p^{-1} (\underset{\sim}{\bar{x}}_1-\underset{\sim}{\bar{x}}_2)},$$

where $S_p = \dfrac{(n_1-1)S_1 + (n_2-1)S_2}{n_1+n_2-2}$

Simultaneous test

Define E_i : i'th hypothesis is accepted

$i = 1, \ldots, r$

$P(E_i) = 1 - \alpha_i$

$P(\bigcap_i E_i) = 1 - \alpha$

How to choose α_i ?

Bonferroni's method (conservative)

Note that $P(\bigcap_i E_i) = 1 - P(\bigcup_i \overline{E}_i)$

$\geq 1 - \sum_i P(\overline{E}_i)$

Choose $\alpha_i = \frac{\alpha}{r}$

$$P(\bigcap_i E_i) \geq 1 - \sum_i \frac{\alpha}{r} = 1 - r\frac{\alpha}{r} = 1 - \alpha$$

$\cup$ - $\cap$ - principle

$$\left.\begin{array}{ll} \text{Hypothesis } H_0, \text{ assume } & H_0 = \bigcap_a H_{0a} \\ \text{Alternative } H_1, \text{ assume } & H_1 = \bigcup_a H_{1a} \end{array}\right\}$$

$a \in A$

Ex. $H_0 : \underset{\sim}{\theta} = \underset{\sim}{0}$ $\quad H_1 : \underset{\sim}{\theta} \neq \underset{\sim}{0}$

$$\left.\begin{array}{ll} \text{Choose for } H_{0a} : & \underset{\sim}{a}^T \underset{\sim}{\theta} = 0 \\ H_{1a} : & \underset{\sim}{a}^T \underset{\sim}{\theta} \neq 0 \end{array}\right\} \quad \underset{\sim}{a} \in R^d$$

$\varphi_1 = \theta_1$

$\varphi_i = \theta_i - \theta_{i-1}$

$i = 2, \ldots, d$

$$\left.\begin{array}{ll} \text{Alt. choice for } H_{0a} : & \varphi_a = 0 \\ H_{1a} : & \varphi_a \neq 0 \end{array}\right\} \quad a = 1, \ldots, d$$

Likelihood ratio test

$\underset{\sim}{y}$ stochastic vector with joint density function

$$f(y_1, \dots, y_n ; \theta_1, \dots \theta_k) \quad , \quad (\theta_1, \dots, \theta_k) \in \Omega$$

the parametres

short notation : $f(\underset{\sim}{y} ; \underset{\sim}{\theta})$, $\underset{\sim}{\theta} \in \Omega$

The likelihood function

$$L(\underset{\sim}{\theta}) = f(\underset{\sim}{\theta} ; \underset{\sim}{y})$$

Hypothesis $H_0 : \underset{\sim}{\theta} \in \omega \subset \Omega$

The likelihood ratio

$$l = \frac{\sup_{\underset{\sim}{\theta} \in \omega} L(\underset{\sim}{\theta})}{\sup_{\theta \in \Omega} L(\underset{\sim}{\theta})}$$

Critical area : $l_{obs} < l_\alpha$ (α is the significance level)

Approximative distribution of l :

$$-2 \ln l \sim \chi^2(\nu) \quad , \quad \nu = \text{"}\dim \Omega\text{"} - \text{"}\dim \omega\text{"}$$

The multivariate normal distribution

See lecture note MA 2 (ii)

Definition b : $\underset{\sim}{y}$ is multivariate normal

$$\Leftrightarrow \forall \underset{\sim}{l} : \underset{\sim}{l}^T \underset{\sim}{y} \sim N(\underset{\sim}{l}^T \underset{\sim}{\theta} , \underset{\sim}{l}^T \Sigma \underset{\sim}{l})$$

singular when $\exists \underset{\sim}{l} \neq \underset{\sim}{0} : \underset{\sim}{l}^T \Sigma \underset{\sim}{l} = \underset{\sim}{0}$

The Multidimensional Normal Distribution

Definition

The stochastic vector $\boldsymbol{y} = (y_1, y_2, \cdots, y_d)$ is said to be d-dimensional normal distributed, when $y_1, y_2, \cdots, y_d$ have the simultaneous density function

$$f_{\boldsymbol{y}}(y_1, y_2, \cdots, y_d) = (2\pi)^{-\frac{d}{2}} (\det \Sigma)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\theta})^\top \Sigma^{-1} (\boldsymbol{y}-\boldsymbol{\theta})\right),$$

where $\boldsymbol{\theta} \in \mathbb{R}^d$ and $\Sigma \in \mathbb{R}^{d\times d}, \Sigma > 0$. A short notation is $\boldsymbol{y} \sim \mathrm{N}_d(\boldsymbol{\theta}, \Sigma)$.

By use of the linear transformation

$$\boldsymbol{y} = \Sigma^{\frac{1}{2}}\boldsymbol{x} + \boldsymbol{\theta}$$

with Jacobian $\frac{\partial(y_1, y_2, \cdots, y_d)}{\partial(x_1, x_2, \cdots, x_d)} = \det \Sigma^{\frac{1}{2}}$, the simultaneous density function for $x_1, x_2, \cdots, x_d$ becomes

$$\begin{aligned} f_{\boldsymbol{x}}(x_1, x_2, \cdots, x_d) &= (2\pi)^{-\frac{d}{2}} (\det \Sigma)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\left(\boldsymbol{x}^\top \Sigma^{\frac{1}{2}}\right)\Sigma^{-1}\left(\Sigma^{\frac{1}{2}}\boldsymbol{x}\right)\right) |\det \Sigma^{\frac{1}{2}}| \\ &= (2\pi)^{-\frac{d}{2}} \exp\left(-\frac{1}{2}\boldsymbol{x}^\top \boldsymbol{x}\right) \quad (\text{that is } \boldsymbol{x} \sim \mathrm{N}_d(\boldsymbol{0}, I_d)) \\ &= (2\pi)^{-\frac{d}{2}} \exp\left(-\frac{1}{2}\sum_{i=1}^{d} x_i^2\right) \\ &= \prod_{i=1}^{d} \frac{1}{\sqrt{2\pi}} e^{-\frac{x_i^2}{2}}. \end{aligned}$$

The product structure shows that $x_1, x_2, \cdots, x_d$ are independent, and that $x_i \sim \mathrm{N}(0,1)$, $i = 1, \cdots, d$. Hence $\mathrm{E}\boldsymbol{x} = \boldsymbol{0}$ and $\mathrm{Var}\,\boldsymbol{x} = I_d$.

Now it is easily seen that

$$\mathrm{E}\boldsymbol{y} \;=\; \Sigma^{\frac{1}{2}}\mathrm{E}\boldsymbol{x} + \boldsymbol{\theta} \;=\; \Sigma^{\frac{1}{2}}\boldsymbol{0} + \boldsymbol{\theta} \;=\; \boldsymbol{\theta}$$

and

$$\operatorname{Var}\boldsymbol{y} = \Sigma^{\frac{1}{2}}\operatorname{Var}\boldsymbol{x}\left(\Sigma^{\frac{1}{2}}\right)^{\top} = \Sigma^{\frac{1}{2}}I_d\Sigma^{\frac{1}{2}} = \Sigma^{\frac{1}{2}}\Sigma^{\frac{1}{2}} = \Sigma.$$

□

After these preliminaries we will look into theorem 2.1 in the book starting by deducing the moment generating function (part (iii)). Next we will prove part (i), (ii), (iv), (v), (vi) and (viii). For the proof of part (vii) the reader is refered to exercise 2.1.

Proof of theorem 2.1 in the book

Ad (iii):

By definition $M_{\boldsymbol{y}}(t_1, \cdots, t_d)$, also written as $M_{\boldsymbol{y}}(\boldsymbol{t})$, becomes

$$\begin{aligned} M_{\boldsymbol{y}}(\boldsymbol{t}) &= \mathrm{E}\left(\exp\left(\boldsymbol{t}^{\top}\boldsymbol{y}\right)\right) \\ &= \int_{\mathbb{R}^d} \exp\left(\boldsymbol{t}^{\top}\boldsymbol{y}\right)(2\pi)^{-\frac{d}{2}}(\det\Sigma)^{-\frac{1}{2}}\exp\left(-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\theta})^{\top}\Sigma^{-1}(\boldsymbol{y}-\boldsymbol{\theta})\right)d\Omega. \end{aligned}$$

We will here use the same change of variables as on page 1. The calculations are straight forward.

$$\begin{aligned} M_{\boldsymbol{y}}(\boldsymbol{t}) &= \int_{\mathbb{R}^d} \exp\left(\boldsymbol{t}^{\top}\left(\Sigma^{\frac{1}{2}}\boldsymbol{x}+\boldsymbol{\theta}\right)\right)(2\pi)^{-\frac{d}{2}}(\det\Sigma)^{-\frac{1}{2}}\exp\left(-\frac{1}{2}\boldsymbol{x}^{\top}\boldsymbol{x}\right)|\det\Sigma^{\frac{1}{2}}|d\Omega_1 \\ &= \exp\left(\boldsymbol{t}^{\top}\boldsymbol{\theta}\right)\int_{\mathbb{R}^d}(2\pi)^{-\frac{d}{2}}\exp\left(\boldsymbol{t}^{\top}\Sigma^{\frac{1}{2}}\boldsymbol{x}-\frac{1}{2}\boldsymbol{x}^{\top}\boldsymbol{x}\right)d\Omega_1 \\ &= \exp\left(\boldsymbol{t}^{\top}\boldsymbol{\theta}+\frac{1}{2}\boldsymbol{t}^{\top}\Sigma\boldsymbol{t}\right)\int_{\mathbb{R}^d}(2\pi)^{-\frac{d}{2}}\exp\left(-\frac{1}{2}\left(\boldsymbol{x}^{\top}\boldsymbol{x}-2\boldsymbol{t}^{\top}\Sigma^{\frac{1}{2}}\boldsymbol{x}+\boldsymbol{t}^{\top}\Sigma\boldsymbol{t}\right)\right)d\Omega_1 \\ &= \exp\left(\boldsymbol{t}^{\top}\boldsymbol{\theta}+\frac{1}{2}\boldsymbol{t}^{\top}\Sigma\boldsymbol{t}\right)\int_{\mathbb{R}^d}(2\pi)^{-\frac{d}{2}}\exp\left(-\frac{1}{2}\left(\boldsymbol{x}-\Sigma^{\frac{1}{2}}\boldsymbol{t}\right)^{\top}\left(\boldsymbol{x}-\Sigma^{\frac{1}{2}}\boldsymbol{t}\right)\right)d\Omega_1 \\ &= \exp\left(\boldsymbol{t}^{\top}\boldsymbol{\theta}+\frac{1}{2}\boldsymbol{t}^{\top}\Sigma\boldsymbol{t}\right), \end{aligned}$$

as the function under the last integral sign is the density of $\boldsymbol{x} \sim \mathrm{N}\left(\Sigma^{\frac{1}{2}}\boldsymbol{t}, I_d\right)$.

□

Ad (i):

We begin by proving

Lemma 1

$$M_{A\boldsymbol{y}+\boldsymbol{b}}(\boldsymbol{t}) = \exp\left(\boldsymbol{t}^\top \boldsymbol{b}\right) M_{\boldsymbol{y}}\left(A^\top \boldsymbol{t}\right).$$

Proof

$$\begin{aligned}
M_{A\boldsymbol{y}+\boldsymbol{b}}(\boldsymbol{t}) &= E\left(\exp\left(\boldsymbol{t}^\top (A\boldsymbol{y}+\boldsymbol{b})\right)\right) \\
&= \exp\left(\boldsymbol{t}^\top \boldsymbol{b}\right) E\left(\exp\left(\boldsymbol{t}^\top A\boldsymbol{y}\right)\right) \\
&= \exp\left(\boldsymbol{t}^\top \boldsymbol{b}\right) E\left(\exp\left(\left(A^\top \boldsymbol{t}\right)^\top \boldsymbol{y}\right)\right) \\
&= \exp\left(\boldsymbol{t}^\top \boldsymbol{b}\right) M_{\boldsymbol{y}}\left(A^\top \boldsymbol{t}\right).
\end{aligned}$$

□

With $\boldsymbol{y} \sim \mathrm{N}_d(\boldsymbol{\theta}, \Sigma)$ and $C \in \mathbb{R}^{q\times d}$, $\operatorname{rank} C = q$, we find

$$\begin{aligned}
M_{C\boldsymbol{y}+\boldsymbol{d}}(\boldsymbol{t}) &= \exp\left(\boldsymbol{t}^\top \boldsymbol{d}\right) M_{\boldsymbol{y}}\left(C^\top \boldsymbol{t}\right), \quad \text{cf. lemma 1} \\
&= \exp\left(\boldsymbol{t}^\top \boldsymbol{d}\right) \exp\left(\left(C^\top \boldsymbol{t}\right)^\top \boldsymbol{\theta} + \frac{1}{2}\left(C^\top \boldsymbol{t}\right)^\top \Sigma \left(C^\top \boldsymbol{t}\right)\right) \\
&= \exp\left(\boldsymbol{t}^\top (C\boldsymbol{\theta}+\boldsymbol{d}) + \frac{1}{2}\boldsymbol{t}^\top C\Sigma C^\top \boldsymbol{t}\right),
\end{aligned}$$

which shows – as $C\Sigma C^\top > 0$ cf. app. A5.7 – that

$$C\boldsymbol{y}+\boldsymbol{d} \sim \mathrm{N}_q\left(C\boldsymbol{\theta}+\boldsymbol{d}, C\Sigma C^\top\right),$$

in particular

$$C\boldsymbol{y} \sim \mathrm{N}_q\left(C\boldsymbol{\theta}, C\Sigma C^\top\right).$$

□

Ad (ii):

Choosing $C = (I_{d_1}\ O)$ and using (i) we get

$$(I_{d_1}\ O)\boldsymbol{y} \sim \mathrm{N}_{d_1}\left((I_{d_1}\ O)\boldsymbol{\theta}, (I_{d_1}\ O)\Sigma\begin{pmatrix} I_{d_1} \\ O \end{pmatrix}\right),$$

which is

$$\boldsymbol{y}^{(1)} \sim \mathrm{N}_{d_1}\left(\boldsymbol{\theta}^{(1)}, \Sigma_{11}\right).$$

Any choice of components from $\boldsymbol{y}$ will form a normal distributed vector or variable.

$\square$

Ad (iv):

First we prove

Lemma 2
Let $\boldsymbol{x}$ and $\boldsymbol{y}$ have the same dimension. Then we have

$$\boldsymbol{x} \text{ and } \boldsymbol{y} \text{ are independent } \Rightarrow M_{\boldsymbol{x}+\boldsymbol{y}}(\boldsymbol{t}) = M_{\boldsymbol{x}}(\boldsymbol{t})M_{\boldsymbol{y}}(\boldsymbol{t})$$

Proof:

$$\begin{aligned}
M_{\boldsymbol{x}+\boldsymbol{y}}(\boldsymbol{t}) &= \mathrm{E}\left[\exp\left(\boldsymbol{t}^\top(\boldsymbol{x}+\boldsymbol{y})\right)\right] \\
&= \mathrm{E}\left[\exp\left(\boldsymbol{t}^\top\boldsymbol{x}+\boldsymbol{t}^\top\boldsymbol{y}\right)\right] \\
&= \mathrm{E}\left[\exp\left(\boldsymbol{t}^\top\boldsymbol{x}\right)\exp\left(\boldsymbol{t}^\top\boldsymbol{y}\right)\right] \\
&= \mathrm{E}\left[\exp\left(\boldsymbol{t}^\top\boldsymbol{x}\right)\right]\mathrm{E}\left[\exp\left(\boldsymbol{t}^\top\boldsymbol{y}\right)\right] \\
&= M_{\boldsymbol{x}}(\boldsymbol{t})M_{\boldsymbol{y}}(\boldsymbol{t}).
\end{aligned}$$

$\square$

Let $\boldsymbol{y} = \binom{\boldsymbol{y}^{(1)}}{\boldsymbol{y}^{(2)}}$ and $\boldsymbol{t} = \binom{\boldsymbol{t}_1}{\boldsymbol{t}_2}$, where both $\boldsymbol{y}$ and $\boldsymbol{t}$ are partitioned in a d_1- and a d_2-dimensional part.

$$\begin{aligned}
M_{\boldsymbol{y}}(\boldsymbol{t}) &= \exp\left(\boldsymbol{t}^\top\boldsymbol{\theta} + \frac{1}{2}\boldsymbol{t}^\top\Sigma\boldsymbol{t}\right) \\
&= \exp\left(\left(\boldsymbol{t}_1{}^\top\boldsymbol{t}_2{}^\top\right)\begin{pmatrix}\boldsymbol{\theta}^{(1)}\\ \boldsymbol{\theta}^{(2)}\end{pmatrix} + \frac{1}{2}\left(\boldsymbol{t}_1{}^\top\boldsymbol{t}_2{}^\top\right)\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix}\begin{pmatrix}\boldsymbol{t}_1\\ \boldsymbol{t}_2\end{pmatrix}\right) \\
&= \exp\left(\boldsymbol{t}_1{}^\top\boldsymbol{\theta}^{(1)} + \boldsymbol{t}_2{}^\top\boldsymbol{\theta}^{(2)} + \frac{1}{2}\boldsymbol{t}_1{}^\top\Sigma_{11}\boldsymbol{t}_1 + \frac{1}{2}\boldsymbol{t}_2{}^\top\Sigma_{21}\boldsymbol{t}_1 + \frac{1}{2}\boldsymbol{t}_1{}^\top\Sigma_{12}\boldsymbol{t}_2 + \frac{1}{2}\boldsymbol{t}_2{}^\top\Sigma_{22}\boldsymbol{t}_2\right) \\
&= \exp\left(\boldsymbol{t}_1{}^\top\boldsymbol{\theta}^{(1)} + \frac{1}{2}\boldsymbol{t}_1{}^\top\Sigma_{11}\boldsymbol{t}_1\right)\exp\left(\boldsymbol{t}_1{}^\top\Sigma_{12}\boldsymbol{t}_2\right)\exp\left(\boldsymbol{t}_2{}^\top\boldsymbol{\theta}^{(2)} + \frac{1}{2}\boldsymbol{t}_2{}^\top\Sigma_{22}\boldsymbol{t}_2\right) \\
&= M_{\boldsymbol{y}^{(1)}}(\boldsymbol{t}_1)\exp\left(\boldsymbol{t}_1{}^\top\Sigma_{12}\boldsymbol{t}_2\right)M_{\boldsymbol{y}^{(2)}}(\boldsymbol{t}_2) \\
&= M_{\binom{\boldsymbol{y}^{(1)}}{\boldsymbol{0}}}(\boldsymbol{t})\exp\left(\boldsymbol{t}_1{}^\top\Sigma_{12}\boldsymbol{t}_2\right)M_{\binom{\boldsymbol{0}}{\boldsymbol{y}^{(2)}}}(\boldsymbol{t}).
\end{aligned}$$

Now

$$y^{(1)} \text{ and } y^{(2)} \text{ are independent} \Leftrightarrow \begin{pmatrix} y^{(1)} \\ 0 \end{pmatrix} \text{ and } \begin{pmatrix} 0 \\ y^{(2)} \end{pmatrix} \text{ are independent}$$
$$\Rightarrow \exp\left(\boldsymbol{t_1}^\top \Sigma_{12} \boldsymbol{t_2}\right) = 1 \text{ for all } \boldsymbol{t}, \text{ cf. lemma 2}$$
$$\Leftrightarrow \Sigma_{12} = O$$
$$\Leftrightarrow \mathrm{Cov}(\boldsymbol{y}^{(1)}, \boldsymbol{y}^{(2)}) = O.$$

Conversely, let $\mathrm{Cov}(y^{(1)}, y^{(2)}) = 0$ ... to be continued on page 7 □

Ad (v):

We will show that the linear transformations

$$\boldsymbol{u}_i = A_i \boldsymbol{y}, \quad i = 1, \cdots, m,$$

are mutually independent if and only if

$$\mathrm{Cov}(\boldsymbol{u}_i, \boldsymbol{u}_j) = O, \quad i \neq j.$$

Let

$$\boldsymbol{u} = \begin{pmatrix} \boldsymbol{u}_i \\ \boldsymbol{u}_j \end{pmatrix} = \begin{pmatrix} A_i \boldsymbol{y} \\ A_j \boldsymbol{y} \end{pmatrix} = \begin{pmatrix} A_i \\ A_j \end{pmatrix} \boldsymbol{y} = A\boldsymbol{y}.$$

Assume that A_i has full rank $(< d), i = 1, \cdots, m$. Notice that $\mathrm{Cov}(\boldsymbol{u}_i, \boldsymbol{u}_j) = O \Leftrightarrow \mathrm{Cov}(A_i\boldsymbol{y}, A_j\boldsymbol{y}) = O \Leftrightarrow A_i \Sigma A_j^\top = O$, which shows that the system of rows in A_i are orthogonal on the system of rows in A_j with respect to the inner product $< \boldsymbol{\alpha}, \boldsymbol{\beta} >= \boldsymbol{\alpha}^\top \Sigma \boldsymbol{\beta}$. Therefore $\mathrm{rank}\, A = \mathrm{rank}\, A_i + \mathrm{rank}\, A_j (= q \leq d)$ and $\boldsymbol{u} \sim \mathrm{N}_q\left(A\boldsymbol{\theta}, A\Sigma A^\top\right)$.

Now $\mathrm{Cov}(\boldsymbol{u}_i, \boldsymbol{u}_j) = O \Leftrightarrow \boldsymbol{u}_i$ and $\boldsymbol{u}_j$ independent, cf. (iv).

□

Ad (vi):

Considering

$$(\boldsymbol{y} - \boldsymbol{\theta})^\top \Sigma^{-1} (\boldsymbol{y} - \boldsymbol{\theta})$$

and once again using the transformation $\boldsymbol{y} = \Sigma^{\frac{1}{2}} \boldsymbol{x} + \boldsymbol{\theta}$ we find

$$(\boldsymbol{y} - \boldsymbol{\theta})^\top \Sigma^{-1} (\boldsymbol{y} - \boldsymbol{\theta}) = \boldsymbol{x}^\top \boldsymbol{x} = \sum_{i=1}^{d} x_i^2 \sim \chi^2(d),$$

as the x_i's are independent and $\mathrm{N}(0, 1)$, cf. page 1. □

Ad (vii):

See exercise 2.1.

Ad (viii):

Let y be partitioned as before:

$$y = \begin{pmatrix} y^{(1)} \\ y^{(2)} \end{pmatrix} \sim \mathrm{N}_{d_1+d_2}\left(\begin{pmatrix} \theta^{(1)} \\ \theta^{(2)} \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right).$$

We introduce $z = y^{(2)} - \Sigma_{21}\Sigma_{11}^{-1} y^{(1)} \sim \mathrm{N}_d(\cdot,\cdot)$ and consider the vector $\binom{y^{(1)}}{z}$. Notice that

$$\begin{pmatrix} y^{(1)} \\ z \end{pmatrix} = \begin{pmatrix} I_{d_1} & 0 \\ -\Sigma_{21}\Sigma_{11}^{-1} & I_{d_2} \end{pmatrix} \begin{pmatrix} y^{(1)} \\ y^{(2)} \end{pmatrix},$$

from which it follows that

$$\begin{aligned} \mathrm{Var}\begin{pmatrix} y^{(1)} \\ z \end{pmatrix} &= \begin{pmatrix} I_{d_1} & 0 \\ -\Sigma_{21}\Sigma_{11}^{-1} & I_{d_2} \end{pmatrix} \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \begin{pmatrix} I_{d_1} & -\left(\Sigma_{21}\Sigma_{11}^{-1}\right)^{\top} \\ 0 & I_{d_2} \end{pmatrix} \\ &= \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ 0 & -\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} + \Sigma_{22} \end{pmatrix} \begin{pmatrix} I_{d_1} & -\Sigma_{11}^{-1}\Sigma_{12} \\ 0 & I_{d_2} \end{pmatrix} \\ &= \begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \end{pmatrix}. \end{aligned}$$

Thus $y^{(1)}$ and z are independent, cf. (iv).

Calculating

$$\mathrm{E}z = \mathrm{E}y^{(2)} - \Sigma_{21}\Sigma_{11}^{-1}\mathrm{E}y^{(1)} = \theta^{(2)} - \Sigma_{21}\Sigma_{11}^{-1}\theta^{(1)},$$

and noticing

$$\begin{aligned} \mathrm{Var}\, z &= \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \\ &= \Sigma_{22\cdot 1}, \end{aligned}$$

we get

$$z \sim \mathrm{N}_{d_2}\left(\theta^{(2)} - \Sigma_{21}\Sigma_{11}^{-1}\theta^{(1)}, \Sigma_{22\cdot 1}\right).$$

Now from the introduction of z we express

$$y^{(2)} = z + \Sigma_{21}\Sigma_{11}^{-1} y^{(1)},$$

which conditionally on $y^{(1)}$ becomes

$$\begin{aligned} y^{(2)} | y^{(1)} &= z | y^{(1)} + \Sigma_{21}\Sigma_{11}^{-1} y^{(1)} \\ &= z + \Sigma_{21}\Sigma_{11}^{-1} y^{(1)} \end{aligned}$$

with

$$\begin{aligned}
\mathrm{E}[\boldsymbol{y}^{(2)}|\boldsymbol{y}^{(1)}] &= \mathrm{E}\boldsymbol{z} + \Sigma_{21}\Sigma_{11}^{-1}\boldsymbol{y}^{(1)} \\
&= \boldsymbol{\theta}^{(2)} - \Sigma_{21}\Sigma_{11}^{-1}\boldsymbol{\theta}^{(1)} + \Sigma_{21}\Sigma_{11}^{-1}\boldsymbol{y}^{(1)} \\
&= \boldsymbol{\theta}^{(2)} + \Sigma_{21}\Sigma_{11}^{-1}(\boldsymbol{y}^{(1)} - \boldsymbol{\theta}^{(1)})
\end{aligned}$$

and

$$\mathrm{Var}\,(\boldsymbol{y}^{(2)}|\boldsymbol{y}^{(1)}) = \mathrm{Var}\,\boldsymbol{z} = \Sigma_{22\cdot 1}.$$

Thus

$$\boldsymbol{y}^{(2)}|\boldsymbol{y}^{(1)} \sim \mathrm{N}_{d_2}\left(\boldsymbol{\theta}^{(2)} + \Sigma_{21}\Sigma_{11}^{-1}(\boldsymbol{y}^{(1)} - \boldsymbol{\theta}^{(1)}), \Sigma_{22\cdot 1}\right).$$

$\square$

2.4.2001/BR, LaTeX version by Tina Madsen.

... continued from page 5 :

Set $\underset{\sim}{y} = (\underset{\sim}{y}^{(1)}, \underset{\sim}{y}^{(2)})$ and note that $\Sigma_{12} = 0$. It follows

$$\Sigma = \begin{bmatrix} \Sigma_{11} & 0 \\ 0^T & \Sigma_{22} \end{bmatrix} \Rightarrow \Sigma^{-1} = \begin{bmatrix} \Sigma_{11}^{-1} & 0 \\ 0^T & \Sigma_{22}^{-1} \end{bmatrix}$$

Then

$$\begin{aligned}
(\underset{\sim}{y} - \underset{\sim}{\mu})^T \Sigma^{-1} (\underset{\sim}{y} - \underset{\sim}{\mu}) &= \begin{bmatrix} \underset{\sim}{y}^{(1)} - \underset{\sim}{\mu}^{(1)} \\ \underset{\sim}{y}^{(2)} - \underset{\sim}{\mu}^{(2)} \end{bmatrix}^T \begin{bmatrix} \Sigma_{11}^{-1} & 0 \\ 0^T & \Sigma_{22}^{-1} \end{bmatrix} \begin{bmatrix} \underset{\sim}{y}^{(1)} - \underset{\sim}{\mu}^{(1)} \\ \underset{\sim}{y}^{(2)} - \underset{\sim}{\mu}^{(2)} \end{bmatrix} \\
&= (\underset{\sim}{y}^{(1)} - \underset{\sim}{\mu}^{(1)})^T \Sigma_{11}^{-1} (\underset{\sim}{y}^{(1)} - \underset{\sim}{\mu}^{(1)}) + (\underset{\sim}{y}^{(2)} - \underset{\sim}{\mu}^{(2)})^T \Sigma_{22}^{-1} (\underset{\sim}{y}^{(2)} - \underset{\sim}{\mu}^{(2)}) \\
&= h_1(\underset{\sim}{y}^{(1)}) + h_2(\underset{\sim}{y}^{(2)})
\end{aligned}$$

The density function of $\underset{\sim}{y}$ becomes

$$\begin{aligned}
&f(y_{11}, \dots, y_{1d_1}, y_{21}, \dots, y_{2d_2}) \\
&= (2\pi)^{-\frac{d_1+d_2}{2}} (\det \Sigma_{11} \det \Sigma_{22})^{-\frac{1}{2}} \exp\left(-\tfrac{1}{2}(h_1(\underset{\sim}{y}^{(1)}) + h_2(\underset{\sim}{y}^{(2)}))\right) \\
&= (2\pi)^{-\frac{d_1}{2}} (\det \Sigma_{11})^{-\frac{1}{2}} \exp(-\tfrac{1}{2} h_1(\underset{\sim}{y}^{(1)}))\, (2\pi)^{-\frac{d_2}{2}} (\det \Sigma_{22})^{-\frac{1}{2}} \cdot \\
&\qquad \exp(-\tfrac{1}{2} h_2(\underset{\sim}{y}^{(2)}))
\end{aligned}$$

Thus $\underset{\sim}{y}^{(1)}$ og $\underset{\sim}{y}^{(2)}$ are independent.