NTNU
Norwegian University of
Science and Technology

TMA4267 Linear Statistical Models V2014 (10)

Properties of the covariance matrix
The multivariate normal distribution [4.3-4.4]

Mette Langaas

To be lectured: February 4, 2014
wiki.math.ntnu.no/emner/tma4267/2014v/start/

The Cork deposit data

- Classical data set from Rao (1948).
- Weigth of bark deposits of $n = 28$ cork trees in $p = 4$ directions (N, E, S, W).

Tree	N	E	S	W
1	72	66	76	77
2	60	53	66	63
3	56	57	64	58
⋮	⋮	⋮	⋮	⋮
28	48	54	57	43

The covariance matrix

Random vector $\boldsymbol{X}_{(p\times 1)}$ with mean vector $\boldsymbol{\mu}_{(p\times 1)}$ and covariance matrix

$$\boldsymbol{\Sigma} = \mathrm{Cov}(\boldsymbol{X}) = \mathrm{E}[(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})^T] = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{12} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1p} & \sigma_{2p} & \cdots & \sigma_{pp} \end{bmatrix}$$

The covariance matrix is by construction symmetric, and we would only consider covariance matrices that are positive definite (PD). Why would we only consider PD matrices?

TMA4267: Lecture 10

Properties of Σ (cont.)

$\underset{p\times p}{\Sigma} = \text{Cov}(X)$, $\underset{p\times 1}{X}$ random vector

We require Σ to be symmetric positive definite (SPD)

$\Sigma^T = \Sigma$ $\quad$ $x^T \Sigma x > 0 \;\; \forall \; x \neq 0$

↳ all eigenvalues are positive

(λ_i, e_i) eigenvalue/eigenvector pair

Spectral theorem $\quad \Sigma = P \Lambda P^T$

$P = [e_1 \; e_2 \cdots e_p]$, $\quad \Lambda = \text{diag}(\lambda_1, .., \lambda_p)$

$\Sigma^{1/2}$ is defined as $\Sigma^{1/2} = P \Lambda^{1/2} P^T$

$\Lambda^{1/2} = \text{diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2}, .., \sqrt{\lambda_p})$

$$\Sigma^{1/2} \Sigma^{1/2} = P \Lambda^{1/2} \underbrace{P^T P}_{I} \Lambda^{1/2} P^T = P \Lambda^{1/2} \Lambda^{1/2} P^T = P \Lambda P^T = \Sigma$$

Since Σ is SPD it has an inverse Σ^{-1}.

$$\Sigma^{-1/2} = P \Lambda^{-\frac{1}{2}} P^T$$

$\Lambda^{-\frac{1}{2}} = \text{diag}\left(\frac{1}{\sqrt{\lambda_1}}, .., \frac{1}{\sqrt{\lambda_p}}\right)$

Cramer-Wold [T4.7]

$\boldsymbol{X}_{(p\times 1)}$ is a random vector. The distribution of $\boldsymbol{X}$ is completely determined by the set of all one-dimensional linear combinations $Y = \boldsymbol{a}^t\boldsymbol{X} = \sum_{i=1}^{p} a_i X_i$ where $\boldsymbol{a}$ ranges over all fixed n-vectors.

Proof:

- $Y = \boldsymbol{a}^T\boldsymbol{X}$ has MGF $M_Y(s) = \mathrm{E}(\exp(sY)) = \mathrm{E}(\exp(s\boldsymbol{a}^T\boldsymbol{X}))$.
- If we choose $s = 1$ $M_Y(1) = \mathrm{E}(\exp(\boldsymbol{a}^T\boldsymbol{X})) = M_{\boldsymbol{X}}(\boldsymbol{a})$, which is the MGF of $\boldsymbol{X}$ and thus determines the distribution of $\boldsymbol{X}$.

The multivariate normal distribution (mvN) [4.3]

B&F **Def 4.8** of multivariate normal:

X is a p-variate normal (N_p) distribution iff $a^T X$ has a univariate (N_1) normal distribution for all constant vectors a.

This implies that: [Let $\underset{p\times 1}{X} \sim N_p(\mu, \Sigma)$]

1) Linear transformation of mvN are also mvN.

$$b^T X \sim N_1(b^T\mu, b^T\Sigma b)$$

$$\underset{m\times p}{B} X \sim N(B\mu, B\Sigma B^T)$$

2) Any vector of elements of a mvN is also mvN, in particular all component are univariate N.

Proof:

1) $\underset{m\times 1}{Y} = \underset{m\times p}{A}X + \underset{m\times 1}{c}$ and b is any m-vector

$$\underset{1\times 1}{b^T Y} = \underbrace{b^T A}_{a^T} X + b^T c =$$

Let $\underset{1\times p}{a^T} = b^T A$, then $a^T X$ is N_1 because X is mvN, and adding $b^T c$ then $b^T Y$ is also N_1. This holds for b, so $Y \sim N_m$

2) Make A a suitable matrix of 0s and 1s.

Linear combinations

- Random vector $\boldsymbol{X}_{(p\times 1)}$ with mean vector $\boldsymbol{\mu}_{\boldsymbol{X}} = \mathrm{E}(\boldsymbol{X})$ and covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{X}} = \mathrm{Cov}(\boldsymbol{X})$.
- The linear combinations $\boldsymbol{Z} = \boldsymbol{C}\boldsymbol{X}$ have

$$\begin{aligned} \boldsymbol{\mu}_{\boldsymbol{Z}} &= \mathrm{E}(\boldsymbol{Z}) = \mathrm{E}(\boldsymbol{C}\boldsymbol{X}) = \boldsymbol{C}\boldsymbol{\mu}_{\boldsymbol{X}} \\ \boldsymbol{\Sigma}_{\boldsymbol{Z}} &= \mathrm{Cov}(\boldsymbol{Z}) = \mathrm{Cov}(\boldsymbol{C}\boldsymbol{X}) = \boldsymbol{C}\boldsymbol{\Sigma}_{\boldsymbol{X}}\boldsymbol{C}^T \end{aligned}$$

Properties of the Multivariate Normal Distribution [4.3]

Let $\boldsymbol{X}_{(p\times 1)}$ be a random vector from $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

1. Linear combinations of components of $\boldsymbol{X}$ are (multivariate) normal.
2. All subsets of the components of $\boldsymbol{X}$ are (multivariate) normal.
3. Zero covariance implies that the corresponding components are independently distributed.

T4.10: $\underset{p\times 1}{X} \sim N_p(\mu, \Sigma)$ has MGF

$$M_X(t) = E\left(e^{t^T X}\right) = \exp\left\{ t^T\mu + \frac{1}{2} t^T \Sigma t \right\}$$

x_1, x_2 $\quad t_1 x_1 + t_2 x_2$

t_1, t_2

Proof: $Y = t^T X$, $E(X) = \mu$, $Cov(X) = \Sigma$

By Def 4.8 $Y \sim N_1(t^T\mu, t^T \Sigma t)$ with MGF (known from Ch 1)

$$M_Y(s) = E(e^{sY}) = \exp\left(s \underbrace{t^T\mu}_{\mu} + \frac{1}{2}s^2 \underbrace{t^T \Sigma t}_{\sigma^2}\right)$$

$$\exp\left(s\mu + \frac{1}{2}s^2\sigma^2\right)$$ ← see lecture 4

$X \sim N_1(\mu, \sigma^2)$ $M_X(t) = \exp(\mu t + \frac{1}{2}\sigma^2 t^2)$

univ.

Choose $s = 1$

$$E(e^{sY}) = E(e^Y) = E\left(e^{t^T X}\right) = M_X(t)$$

is the MGF of X

$$M_X(t) = E\left(e^{t^T X}\right) = \exp\left(t^T\mu + \frac{1}{2}t^T\Sigma t\right)$$

C4.11. The components of X are independent iff Σ is diagonal.

Proof: We looked at the bivariate normal (Ch1, L4) and saw that $M_{X_1,X_2}(t_1,t_2) = M_{X_1}(t_1)\cdot M_{X_2}(t_2)$ when $\rho_{12}=0 \Rightarrow \Sigma$ is diagonal. This means that Σ need to be diagonal to make the X_i's independent, and viceversa.

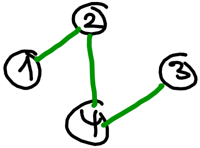
Independent variables Example

$$\Sigma = \begin{bmatrix} 2 & 1 & 0 & 0 \\ 1 & 2 & 0 & 1 \\ 0 & 0 & 2 & 1 \\ 0 & 1 & 1 & 2 \end{bmatrix}$$

X_2 vs X_3

X_1 vs X_3

X_1 vs X_4



Independent variables?

Let $\boldsymbol{X}_{p\times 1} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with

$$\boldsymbol{\Sigma} = \begin{bmatrix} 2 & 1 & 0 & 0 \\ 1 & 2 & 0 & 1 \\ 0 & 0 & 2 & 1 \\ 0 & 1 & 1 & 2 \end{bmatrix}$$

— List the pairs of variables that are independent.

Independent variables?

```
sigma <- matrix(c(2,1,0,0,1,2,0,1,0,0,2,1,0,1,1,2),ncol=4)
     [,1] [,2] [,3] [,4]
[1,]    2    1    0    0
[2,]    1    2    0    1
[3,]    0    0    2    1
[4,]    0    1    1    2

eigen(sigma)
$values
[1] 3.618034 2.618034 1.381966 0.381966
$vectors
         [,1]      [,2]      [,3]      [,4]
[1,] 0.371748  0.601501  0.601501  0.371748
[2,] 0.601501  0.371748 -0.371748 -0.601501
[3,] 0.371748 -0.601501  0.601501 -0.371748
[4,] 0.601501 -0.371748 -0.371748  0.601501

library(MASS)
ds <- mvrnorm(1000,c(0,0,0,0),sigma)
pairs(ds)
```

Finally, Teo 4.15

Let $\underset{m\times p}{A}$ and $\underset{r\times p}{B}$ be conformable constant matrices

and $\underset{p\times 1}{X} \sim N_p(\mu, \Sigma)$.

Then $\underbrace{AX}_{m\times 1}$ and $\underbrace{BX}_{r\times 1}$ are independent iff $\underset{m\times p}{A}\,\underset{p\times p}{\Sigma}\,\underset{p\times r}{B^T} = 0$

Seen from MGF, proof p 111.

The mvN density [4.4]

$\underset{1\times 1}{Z} \sim N_1(0,1)$ $\qquad f(z) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}z^2}$

$\downarrow$ indepent

$\underset{p\times 1}{Z} \sim N_p(0, \underset{p\times p}{I})$ $\qquad f(\underset{p\times 1}{z}) = \prod_{i=1}^{p} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z_i^2}$

$$= \left(\frac{1}{2\pi}\right)^{p/2} e^{-\frac{1}{2}\sum z_i^2} = \left(\frac{1}{2\pi}\right)^{p/2} e^{-\frac{1}{2}z^T z}$$

$\downarrow$ $X = \mu + \Sigma^{1/2} Z$

$\underset{p\times 1}{X} \sim N_p(\mu, \underset{p\times p}{\Sigma})$

$$E(X) = \Sigma^{1/2} \underbrace{E(Z)}_{=0} + \mu = \mu$$

$$Cov(X) = \Sigma^{1/2} \underbrace{Cov(Z)}_{I} \Sigma^{1/2} = \Sigma^{1/2}\Sigma^{1/2} = \Sigma$$

The mv transformation formula:

$$f_X(x) = f_Z(\underbrace{z(x)}) \cdot \text{abs}(\underset{\uparrow}{J})$$

$$X = \Sigma^{1/2} Z + \mu \iff Z = \Sigma^{-1/2}(X - \mu) = z(x)$$

$$J = \det\left\{\frac{\partial z_i}{\partial x_j}\right\}$$

$$\underset{p \times p}{\left\{\frac{\partial z_i}{\partial x_j}\right\}} = \Sigma^{-\frac{1}{2}}$$

$$f_X(x) = \left(\frac{1}{2\pi}\right)^{p/2} \cdot \exp\left\{-\frac{1}{2}\left(\Sigma^{-1/2}(X - \mu)\right)^T \Sigma^{-1/2}(X - \mu)\right\}$$

$$\cdot \text{abs}\left(\det\left(\Sigma^{-1/2}\right)\right)$$

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

$$\det(\Sigma^{-1/2}) = \frac{1}{\det(\Sigma^{1/2})} = \left(\frac{1}{\det(\Sigma)}\right)^{1/2}$$

$$\det(\Sigma^{1/2}) = \sqrt{\det(\Sigma)}$$

$$\leftarrow \det(\Sigma) = \det(\Sigma^{1/2} \cdot \Sigma^{1/2})$$
$$= \det(\Sigma^{1/2}) \cdot \det(\Sigma^{1/2})$$

Finally:

$$f_x(x) = 2\pi^{-p/2} \det(\Sigma)^{-1/2}$$

$$\exp\left\{-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right\}$$

The Multivariate Normal Density [4.4]

— Univariate normal, $X \sim N(\mu, \sigma^2)$

$$f(x) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} \exp\{-\frac{1}{2\sigma^2}(x-\mu)^2\}$$

— Multivariate normal:
Let $\boldsymbol{X}_{(p \times 1)}$ be a random vector from $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then

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

Contours of multivariate normal distribution

- Contours of constant density for the p-dimensional normal distribution are ellipsoids defined by $\boldsymbol{x}$ such that

$$(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) = c^2$$

  These ellipsoids are centered at $\boldsymbol{\mu}$ and have axes $\pm c\sqrt{\lambda_i}\boldsymbol{e}_i$, where $\boldsymbol{\Sigma}\boldsymbol{e}_i = \lambda_i\boldsymbol{e}_i$, for $i = 1, ..., p$.
- $(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$ is distributed as χ^2_p.
- The solid ellipsoid of $\boldsymbol{x}$ values satisfying

$$(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) \leq \chi^2_p(\alpha)$$

  has probability $1-\alpha$