NTNU
Norwegian University of
Science and Technology

TMA4267 Linear Statistical Models V2014 (9)

E and Cov of random vectors and matrices
The multivariate normal distribution [4.3]

Mette Langaas

To be lectured: February 3, 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

Part 3: From random vectors and matrices to the multivariate normal distribution [L9]

[4.3-4.5]

Random vector $\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix} = \underset{p \times 1}{X}$ ex: cork tree $p=4$

X has joint distribution $f(x) = f(x_1, x_2, \dots, x_p)$ and each X_j has marginal distribution $f_j(x_j)$.

Mean

$$\underset{p \times 1}{E(X)} = \begin{bmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_p) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_p \end{bmatrix} = \underset{p \times 1}{\mu}$$

$$E(X_1) = \int_{-\infty}^{\infty}\int \cdots \int x_1 \cdot f(x_1, \dots, x_p)\, dx_1 \cdots dx_p$$

$$= \int x_1 \cdot f_1(x_1)\, dx_1 = \mu_1$$

$E(X_j)$ is calculated from the marginal distribution of X_j and contains no information about dependency between X_j and X_k $k \neq j$.

Rules for mean:

Let $\underset{p\times 1}{X}$ and $\underset{p\times 1}{Y}$ be random vectors and $\underset{k\times p}{A}$ be a constant matrix.

1) $E(X+Y) = E(X) + E(Y)$

2) $E(\underbrace{AX}_{k\times 1}) = A\,E(X)$

Proof:

1) Look at one element $Z_j = X_j + Y_j$

2) $$AX = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ \vdots & & & \\ a_{k1} & \cdots & & a_{kp} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix} = \left\{ \sum_{j=1}^{p} a_{ij} X_j \right\}$$

element i of my $k\times 1$ vector

$$E\left(\sum_{j=1}^{p} a_{ij} X_j \right) = \underbrace{\sum_{j=1}^{p} a_{ij} E(X_j)}_{i\text{ th element of } A\cdot E(X)}$$

Theorem:

Let $\underset{n\times p}{X}$ and $\underset{n\times p}{Y}$ be random matrices [ex. cork data set]

and $\underset{m\times n}{A}$ and $\underset{p\times q}{B}$ be constant matrices.

1) $\underbrace{E(X+Y)}_{n\times p} = \overbrace{E(X)}^{n\times p} + \overbrace{E(Y)}^{n\times p}$

$$E(X) = \begin{bmatrix} E(X_{11}) & E(X_{12}) & \cdots & E(X_{1p}) \\ \vdots & & & \\ E(X_{n1}) & \cdots & & E(X_{np}) \end{bmatrix}$$

2) $E(\underbrace{A\,X\,B}_{m\times q}) = A\,E(X)\,B$

proof.

1) as for vectors

2) by looking at $AXB = \left\{ \sum_{k=1}^{n} a_{ik} \sum_{l=1}^{p} X_{kl} b_{lj} \right\}_{(i,j)}$

Take $E\big(\text{element}(i,j)\big)$ and then observe

that this is the element (i,j) of $AE(X)B$.

Random vectors and matrices

— Random vector $\boldsymbol{X}_{(p\times 1)}$ with mean vector $\boldsymbol{\mu}_{(p\times 1)}$:

$$\boldsymbol{X}_{(p\times 1)} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}, \qquad \boldsymbol{\mu}_{(p\times 1)} = \mathrm{E}(\boldsymbol{X}) = \begin{bmatrix} \mathrm{E}(X_1) \\ \mathrm{E}(X_2) \\ \vdots \\ \mathrm{E}(X_p) \end{bmatrix}$$

— Random vector $\boldsymbol{Y}_{(p\times 1)}$ and random vector $\boldsymbol{X}_{(p\times 1)}$:

$$\mathrm{E}(\boldsymbol{X}+\boldsymbol{Y}) = \mathrm{E}(\boldsymbol{X}) + \mathrm{E}(\boldsymbol{Y})$$

— Random vector $\boldsymbol{X}_{(p\times 1)}$ and conformable constant matrix $\boldsymbol{A}$:

$$\mathrm{E}(\boldsymbol{A}\boldsymbol{X}) = \boldsymbol{A}\mathrm{E}(\boldsymbol{X})$$

— Random matrix $\boldsymbol{X}_{(n\times p)}$ and conformable constant matrices $\boldsymbol{A}$ and $\boldsymbol{B}$:

$$\mathrm{E}(\boldsymbol{A}\boldsymbol{X}\boldsymbol{B}) = \boldsymbol{A}\mathrm{E}(\boldsymbol{X})\boldsymbol{B}$$

Covarience matrix

1) From TMA4245/ST1101: X and Y scalar RV's with $E(X) = \mu_X$, $E(Y)\ \mu_Y$.

$$\mathrm{Cov}(X,Y) = E\big((X-\mu_X)(Y-\mu_Y)\big) = E(X\cdot Y) - \mu_X \cdot \mu_Y$$

$$\mathrm{Var}(X) = \mathrm{Cov}(X,X).$$

$\mathrm{Cov}(X,Y) = 0$ if X and Y are independent (but not the opposite).

2) Now back to $\underset{p\times 1}{X}$ randomvector

$$\underset{p\times p}{\Sigma} = \underset{p\times p}{\mathrm{Cov}(X)} = \begin{bmatrix} \overbrace{\mathrm{Cov}(X_1,X_1)}^{\sigma_{11}} & \mathrm{Cov}(X_1,X_2) & \cdots & \overbrace{\mathrm{Cov}(X_1,X_p)}^{\sigma_{1p}} \\ \mathrm{Cov}(X_2,X_1) & \mathrm{Cov}(X_2,X_2) & & \\ & & & \\ \underbrace{\mathrm{Cov}(X_p,X_1)}_{\sigma_{p1}} & \cdots & & \mathrm{Cov}(X_p,X_p) \end{bmatrix}$$

a symmetric real matrix

$$\sigma_{jk} = \mathrm{Cov}(X_j, X_k) = E\big((X_j-\mu_j)(X_k-\mu_k)\big)$$

where $\underset{p\times 1}{E(X)} = \underset{p\times 1}{\mu} = \begin{bmatrix}\mu_1\\ \vdots \\ \mu_p\end{bmatrix}$

$$\underset{p\times p}{\Sigma} = E\Big(\underbrace{(X-\mu)}_{p\times 1}\underbrace{(X-\mu)^T}_{1\times p}\Big)$$

E3

$$= E(XX^T) - \mu\cdot\mu^T$$

Can the information in Σ be condensed?

$$tr(\Sigma) = \text{sum of diagonal elements}$$
$$= \text{sum of variances} \overset{\text{DEF}}{=} \text{total variance}$$

$$\det(\Sigma) \overset{\text{DEF}}{=} \text{generalized variance}$$

DEF: The covariance between two vectors.

$\underset{p\times 1}{X}$ and $\underset{p\times 1}{Y}$ two random vectors $E(X) = \mu_X$, $E(Y) = \mu_Y$

$$Cov(X,Y) = E\Big(\underset{p\times 1}{(X-\mu_X)}\underset{1\times p}{(Y-\mu_Y)^T}\Big)$$

and $Cov(X,Y) = 0$ if X and Y are independent.

Covariance matrix

— Random vector $\boldsymbol{X}_{(p\times 1)}$ with mean vector $\boldsymbol{\mu}_{(p\times 1)}$:

$$\boldsymbol{X}_{(p\times 1)} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}, \quad \boldsymbol{\mu}_{(p\times 1)} = \begin{bmatrix} \mathrm{E}(X_1) \\ \mathrm{E}(X_2) \\ \vdots \\ \mathrm{E}(X_p) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix}$$

— Covariance matrix $\boldsymbol{\Sigma}$ (real and symmetric)

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

— $\sigma_{ij} = \mathrm{E}[(X_i - \mu_i)(X_j - \mu_j)]$

Correlation matrix

Correlation matrix $\boldsymbol{\rho}$ (real and symmetric)

$$\boldsymbol{\rho} = \begin{bmatrix} \frac{\sigma_{11}}{\sqrt{\sigma_{11}\sigma_{11}}} & \frac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}}} & \cdots & \frac{\sigma_{1p}}{\sqrt{\sigma_{11}\sigma_{pp}}} \\ \frac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}}} & \frac{\sigma_{22}}{\sqrt{\sigma_{22}\sigma_{22}}} & \cdots & \frac{\sigma_{2p}}{\sqrt{\sigma_{22}\sigma_{pp}}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\sigma_{1p}}{\sqrt{\sigma_{11}\sigma_{pp}}} & \frac{\sigma_{2p}}{\sqrt{\sigma_{22}\sigma_{pp}}} & \cdots & \frac{\sigma_{pp}}{\sqrt{\sigma_{pp}\sigma_{pp}}} \end{bmatrix} = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{12} & 1 & \cdots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1p} & \rho_{2p} & \cdots & 1 \end{bmatrix}$$

$$\boldsymbol{\rho} = (\boldsymbol{V}^{\frac{1}{2}})^{-1}\boldsymbol{\Sigma}(\boldsymbol{V}^{\frac{1}{2}})^{-1}, \text{ where } \boldsymbol{V}^{\frac{1}{2}} = \begin{bmatrix} \sqrt{\sigma_{11}} & 0 & \cdots & 0 \\ 0 & \sqrt{\sigma_{22}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\sigma_{pp}} \end{bmatrix}$$

E and Cov of linear combinations

$\underset{p\times 1}{X}$ random vector with $\mu_X = E(X)$

$\Sigma_X = Cov(X)$

$\underset{k\times p}{C}$ constants.

Consider $\underset{k\times 1}{Z} = \underset{k\times p}{C}\ \underset{p\times 1}{X} = \begin{bmatrix} \sum_{j=1}^{p} c_{1j} X_j \\ \vdots \\ \sum_{j=1}^{p} c_{kj} \cdot X_j \end{bmatrix}$

1) $\underset{k\times 1}{\mu_Z} = E(Z) = E(CX) = C \cdot E(X) = C\mu_X$

2) $\underset{k\times k}{\Sigma_Z} = Cov(Z) = Cov(CX) = C \cdot Cov(X)\, C^T$

$= C\, \Sigma_X\, C^T$

Proof:

1) already proven.

2) $$\mathrm{Cov}(Z) = E\left[(Z-\mu_Z)(Z-\mu_Z)^T \right]$$

$$= E\left[(CX - C\mu_X)(CX - C\mu_X)^T \right]$$

$$= E\Big[C \underbrace{(X-\mu_X)(X-\mu_X)^T}_{\substack{Y \\ p\times p \text{ matrix}}} C^T \Big]$$

$E(AXB) = AE(X)B$

$$= C \; E(Y) \; C^T$$

$$= C \cdot \underbrace{E\left((X-\mu_X)(X-\mu_X)^T \right)}_{\Sigma_X} C^T$$

$$= C \; \Sigma_X \; C^T$$

Linear combinations

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

$$\begin{aligned} \boldsymbol{\mu_Z} &= \mathrm{E}(\boldsymbol{Z}) = \mathrm{E}(\boldsymbol{CX}) = \boldsymbol{C\mu_X} \\ \boldsymbol{\Sigma_Z} &= \mathrm{Cov}(\boldsymbol{Z}) = \mathrm{Cov}(\boldsymbol{CX}) = \boldsymbol{C\Sigma_X C}^T \end{aligned}$$

Task: The cork example

Let $X = \begin{array}{c} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix} \end{array} \begin{array}{c} N \\ E \\ S \\ W \end{array}$, $\mu = E(X)$, $\Sigma = Cov(X)$

$$= \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{21} & & & \\ \sigma_{31} & & & \\ \sigma_{41} & & & \sigma_{44} \end{bmatrix}$$

Scientists would like to "compare" the following three contrasts

$$N - S$$
$$E - W$$
$$(E+W) - (N+S)$$

X_1 N
X_2 E
X_3 S
X_4 W

1) Write this as $\underset{3\times 1}{Y} = \underset{3\times 4}{C} X$. What is C.

2) What is $E(Y)$?

3) Sketch how to find $Cov(Y)$.

1) $$\underset{3\times 4}{C} = \begin{array}{cccc} N & E & S & W \\ \end{array}$$
$$\begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ -1 & 1 & -1 & 1 \end{bmatrix}$$

2) $$E(CX) = C \cdot \underset{\begin{smallmatrix}\mu_1\\ \mu_2\\ \mu_3\\ \mu_4\end{smallmatrix}}{\mu} = \begin{bmatrix} \mu_1 - \mu_3 \\ \mu_2 - \mu_4 \\ -\mu_1 + \mu_2 - \mu_3 + \mu_4 \end{bmatrix}$$

3) $$Cov(CX) = C \cdot \Sigma C^T = \begin{bmatrix} \sigma_{11} + \sigma_{33} - 2\sigma_{13} & & \\ & & \\ & & \end{bmatrix}$$

$$\underset{3\times 4}{\begin{bmatrix} & \end{bmatrix}} \quad 4\times 4 \quad \underset{4\times 3}{\begin{bmatrix} & \end{bmatrix}}$$

Properties of Σ

Let $\underset{p\times p}{\Sigma}$ be a real symmetric matrix.

A vector e $(e\neq 0)$ and a scalar λ satisfy

$$\Sigma e = \lambda e$$

are called the eigenvectors and eigenvalues of Σ.

Σ has p sets $(\lambda_1, e_1), \dots, (\lambda_p, e_p)$ of eigenvalues and vectors. We assume that we have ortogonal vectors,

$\underset{i\neq j}{e_i^T e_j = 0}$, $e_i^T e_i = 1$.

Observe:

- all eigenvalues are real
- the eigenvectors of distinct eigenvalues are ortogonal
- $\det(\Sigma) = \prod_{i=1}^{p} \lambda_i$
- $tr(\Sigma) = \sum_{i=1}^{p} \lambda_i$

We only consider Σ that are positive definite (PD)

If $\underset{1\times p}{x^T} \underset{p\times p}{\Sigma} \underset{p\times 1}{x} > 0 \quad \forall x \neq 0$

then Σ is PD.

Why do I need PD?

Let $\underset{p\times 1}{X}$ random vector have $Cov(\Sigma)$ and look at

$\underset{1\times 1}{Y} = \underset{1\times p}{c^T} \underset{p\times 1}{X}$. We would like $Var(Y) > 0$ for any $c \neq 0$.

$$Var(Y) = Cov(c^T X) = c^T \underbrace{Cov(X)}_{\Sigma} c = c^T \Sigma c > 0$$

i.e. need Σ to be PD.