Preliminaries

Hans Karlsen

UiB: 2021-01-27 12:39:25

Contents

Stat211 Spring 2021

i) $2x2$ lectures per week : Hans Karlsen
Tuesday: 10-12 and Wednesday 10-12.

ii) $1x2$ exercises per week: Robert Clay Glastad
Wednesday 8-10 .

iii) Mandatory homeworks (9?),

iv) Oral exam (I think).

v) Use of the program R in the homworks plus some theory.

vi) You can download the main textbook. (Introduction to time series, ed. 3, 2016, Brockwell & Davis)

Covariance and correlation in the bivariate case

1. Covariance and correlation

For dependent variables covariance and correlation are important concepts that contain information about the dependence.

1.1. Covariance and correlation of two variables

Definition 1. Let X and Y be stochastic variables. Then

$$\sigma_{XY} \stackrel{\text{def}}{=} \operatorname{Cov}(X,Y) \stackrel{\text{def}}{=} \mathbb{E}\,(X-\mu_X)(Y-\mu_Y) \tag{1.1}$$

The expectation of a sum is equal to the sum of the expectations and therefore

$$\sigma_{XY} = \mathbb{E}\,XY - \mu_X\mu_Y$$

The bivariate distribution function

Remark 1.

$$\mathrm{Cov}(aX + b, cY + d) = ac\,\sigma_{X,Y}$$

Remark 2. Note that we in general need the bivariate distribution function in order to compute the covariance between to variables. However, in many cases we have a model or assumptions so that this is not needed and we can relay on rules for the expectation and the covariance. In the general case,

$$\mathbb{E}\,XY = \iint xy\,dF_{X,Y}(x, y)$$

A stochastic vector

1.2. Covariance and correlation of stochastic vectors

Let $\mathbf{X}$ be a p-dimensional stochastic vector.

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}$$

Then

$$\boldsymbol{\mu} \overset{\text{def}}{=} \mathbb{E}\,\mathbf{X} \overset{\text{def}}{=} \begin{bmatrix} \mathbb{E}\,X_1 \\ \mathbb{E}\,X_2 \\ \vdots \\ \mathbb{E}\,X_p \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix}$$

Stochastic matrix

If $\mathbb{X}$ is a stochastic matrix of dimension $p \times m$ then

$$\mathbb{E}\,\mathbb{X} \stackrel{\text{def}}{=} \mathbb{E}\begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1m} \\ X_{21} & X_{22} & \cdots & X_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ X_{p1} & X_{p2} & \cdots & X_{pm} \end{pmatrix} = \begin{pmatrix} \mathbb{E}\,X_{11} & \mathbb{E}\,X_{12} & \cdots & \mathbb{E}\,X_{1m} \\ \mathbb{E}\,X_{21} & \mathbb{E}\,X_{22} & \cdots & \mathbb{E}\,X_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{E}\,X_{p1} & \mathbb{E}\,X_{p2} & \cdots & \mathbb{E}\,X_{pm} \end{pmatrix}$$

This is a definition.

Covariance matrix

DEFINITION 2. The covariance matrix for a stochastic vector of dimension $p \times p$ is

$$\Sigma = \mathrm{Var}(\mathbf{X}) \stackrel{\text{def}}{=} \mathrm{Cov}(\mathbf{X}, \mathbf{X})$$

$$\stackrel{\text{def}}{=} \begin{bmatrix} \mathrm{Cov}(X_1, X_1) & \mathrm{Cov}(X_1, X_2) & \cdots & \mathrm{Cov}(X_1, X_p) \\ \mathrm{Cov}(X_2, X_1) & \mathrm{Cov}(X_2, X_2) & \cdots & \mathrm{Cov}(X_2, X_p) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(X_p, X_1) & \mathrm{Cov}(X_p, X_2) & \cdots & \mathrm{Cov}(X_p, X_p) \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_{11} & \sigma_{11} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{21} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix}, \qquad \sigma_{ij} = \mathrm{Cov}(X_i, X_j)$$

REMARK 3. Note that Σ is a $p \times p$ symmetric matrix.

Vector calculations

We can calculate the covariance matrix using vector notation

PROPOSITION 1.

$$\Sigma = \mathrm{Cov}(\mathbf{X}, \mathbf{X}) = \mathbb{E}\,\mathbf{X}\mathbf{X}^T - \mathbb{E}\,\mathbf{X}\;\mathbb{E}^T\mathbf{X} \tag{1.2}$$

Matrix formula

PROOF OF PROPOSITION 1.
By definition of the expectation of a stochastic matrix and a stochastic vector

$$\mathbb{E}\,\mathbf{X}\mathbf{X}^T = \mathbb{E}\left\{X_i X_j\right\} = \left\{\mathbb{E}\,X_i X_j\right\}$$
$$\mathbb{E}\,\mathbf{X}\;\mathbb{E}^T\mathbf{X} = \left\{\mathbb{E}\,X_i\right\}\left\{\mathbb{E}\,X_j\right\}^T = \left\{\mathbb{E}\,X_i\,\mathbb{E}\,X_j\right\}$$

Hence

$$\begin{aligned}
&\mathbb{E}\,\mathbf{X}\mathbf{X}^T - \mathbb{E}\,\mathbf{X}\;\mathbb{E}^T\mathbf{X} \\
&\quad = \left\{\mathbb{E}\,X_i X_j\right\} - \left\{\mathbb{E}\,X_i\,\mathbb{E}\,X_j\right\} \\
&\quad = \left\{\mathbb{E}\,X_i X_j - \mathbb{E}\,X_i\,\mathbb{E}\,X_j\right\} \\
&\quad = \left\{\mathrm{Cov}(X_i, X_j)\right\} \\
&\quad = \mathbb{\Sigma}
\end{aligned}$$

□

Two vectors - cross covariance matrix

Suppose that $\mathbf{X}$ and $\mathbf{Y}$ are two stochastic vectors then the long vector

$$\left(\mathbf{X}^T, \mathbf{Y}^T\right)^T = \begin{bmatrix} X_1 \\ \vdots \\ X_p \\ Y_1 \\ \vdots \\ Y_m \end{bmatrix}$$

has covariance matrix

$$\Sigma_{(\mathbf{X},\mathbf{Y})} = \begin{bmatrix} \Sigma_{\mathbf{X}} & \Sigma_{\mathbf{XY}} \\ \Sigma_{\mathbf{YX}} & \Sigma_{\mathbf{Y}} \end{bmatrix}$$

(concatenate, coercion, merging)

Cross covariance matrix

The matrix $\Sigma_{\mathbf{XY}}$ is the cross covariance matrix between the vectors,

$$
\begin{aligned}
\Sigma_{\mathbf{XY}} &\stackrel{\text{def}}{=} \mathrm{Cov}(\mathbf{X},\ \mathbf{Y}) \\
&= \begin{bmatrix}
\mathrm{Cov}(X_1, Y_1) & \mathrm{Cov}(X_1, Y_2) & \cdots & \mathrm{Cov}(X_1, Y_m) \\
\mathrm{Cov}(X_2, Y_1) & \mathrm{Cov}(X_2, Y_2) & \cdots & \mathrm{Cov}(X_2, Y_m) \\
\vdots & \vdots & \ddots & \vdots \\
\mathrm{Cov}(X_p, Y_1) & \mathrm{Cov}(X_p, Y_2) & \cdots & \mathrm{Cov}(X_p, Y_m)
\end{bmatrix}_{p\times m}
\end{aligned}
$$

Definition 3. Two stochastic vectors are uncorrelated if their cross covariance matrix is a zero matrix.

PROPOSITION 2.

$$\Sigma_{\mathbf{XY}} = \mathrm{Cov}(\mathbf{X}, \mathbf{X}) = \mathbb{E}\,\mathbf{X}\mathbf{Y}^T - \mathbb{E}\,\mathbf{X}\;\mathbb{E}^T\mathbf{Y} \tag{1.3}$$

It follows directly from Proposition 1.

A useful formula

PROPOSITION 3. Let $\mathbf{X}$ and $\mathbf{Y}$ be a p and m-dimensional stochastic vectors with finite covariance matrices, respectively. Let $\mathbb{A}$ and $\mathbb{B}$ be $r \times p$ and $m \times s$ dimensional matrices, respectively. Then

$$\mathrm{Cov}(\mathbb{A}\,\mathbf{X}, \mathbb{B}\,\mathbf{Y}) = \mathbb{A}\,\Sigma_{\mathbf{XY}}\mathbb{B}^T$$

PROOF OF PROPOSITION 3.
Assume without loss of generality that both mean vectors are zero. By (1.3),

$$\begin{aligned}
\mathrm{Cov}(\mathbb{A}\,\mathbf{X}, \mathbb{B}\,\mathbf{Y}) &= \mathbb{E}\,(\mathbb{A}\,\mathbf{X})(\mathbb{B}\,\mathbf{Y})^T \\
&= \mathbb{E}\,\mathbb{A}\mathbf{X}\,\mathbf{Y}^T\mathbb{B}^T \\
&= \mathbb{A}\,\mathbb{E}\left(\mathbf{X}\,\mathbf{Y}^T\right)\mathbb{B}^T \\
&= \mathbb{A}\,\Sigma_{\mathbf{XY}}\,\mathbb{B}^{\mathrm{T}}
\end{aligned}$$

□

Details

$$\mathbb{C} \stackrel{\text{def}}{=} \mathbb{A}\,\mathbf{X}\,\mathbf{Y}^T\mathbb{B}^T$$

The expectation of a random matrix is the matrix of expectations. At the component level

$$\begin{aligned} C_{ij} &= \left[\mathbb{A}\mathbf{X}\,\mathbf{Y}^T\mathbb{B}^T\right]_{ij} \\ &= \sum_k \sum_\ell a_{ik}\,(\mathbf{X}\mathbf{Y})_{k\ell}\,b^T_{\ell j} \\ &= \sum_k \sum_\ell a_{ik} X_k Y_\ell\, b_{j\ell} \end{aligned}$$

Hence

$$\begin{aligned} \mathbb{E}\,C_{ij} &= \sum_k \sum_\ell a_{ik}\mathrm{Cov}(X_k, Y_\ell)\, b_{j\ell} \\ &= \left(\mathbb{A}\,\Sigma_{\mathbf{XY}}\,\mathbb{B}^{\mathrm{T}}\right)_{ij} \end{aligned}$$

Covariance and Correlation

Let Σ be a covariance matrix and define

$$\mathbb{D} = \operatorname{diag} \sigma_{ii},\ i = 1, \ldots, p = \begin{pmatrix} \sigma_{11} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_{pp} \end{pmatrix}$$

The corresponding correlation matrix is

$$\mathcal{R} = \{\rho_{ij}\} = \mathbb{D}^{-1/2}\, \Sigma\, \mathbb{D}^{-1/2}$$

with

$$\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\,\sigma_{jj}}}$$

The regression matrix

PROPOSITION 4. Let $\mathbf{X}$ and $\mathbf{Y}$ are two stochastic vectors with $\Sigma_{\mathbf{X}}$ nonsingular. Then $\mathbf{Y} - \mathbb{M}\mathbf{X}$ and $\mathbf{X}$ are uncorrelated when $\mathbb{M} = \Sigma_{\mathbf{YX}}\Sigma_{\mathbf{X}}^{-1}$.

Choose $\mathbb{M}$

PROOF OF PROPOSITION 4.
Let $\mathbb{B}$ be any $m \times p$ matrix and $\mathbf{Z} \stackrel{\text{def}}{=} \mathbf{Y} - \mathbb{B}\,\mathbf{X}$. Then

$$\begin{aligned} \Sigma_{\mathbf{ZX}} &= \mathrm{Cov}(\mathbf{Z}, \mathbf{X}) \\ &= \mathrm{Cov}(\mathbf{Y} - \mathbb{B}\,\mathbf{X}, \mathbf{X}) \\ &= \mathrm{Cov}(\mathbf{Y}, \mathbf{X}) \; - \mathrm{Cov}(\mathbb{B}\,\mathbf{X}, \mathbf{X}) \\ &= \Sigma_{\mathbf{YX}} - \mathbb{B}\,\Sigma_{\mathbf{X}} \end{aligned}$$

Then $\mathbf{Z}$ and $\mathbf{X}$ are uncorrelated if

$$\Sigma_{\mathbf{YX}} - \mathbb{B}\,\Sigma_{\mathbf{X}} = \mathbb{0}$$

which has solution $\mathbb{B} = \Sigma_{\mathbf{YX}}\Sigma_{\mathbf{X}}^{-1} = \mathbb{M}$. $\square$

Covariance between to sums

COROLLARY 1.

$$\mathrm{Cov}\Big(\sum_{i=1}^{p} a_i X_i,\ \sum_{j=1}^{p} b_j X_j\Big) = \mathbf{a}^T\,\Sigma\,\mathbf{b}$$

PROOF.

$$\mathrm{Cov}\Big(\sum_{i=1}^{p} a_i X_i,\ \sum_{j=1}^{p} b_j X_j\Big) = \mathrm{Cov}\big(\mathbf{a}^T\mathbf{X},\ \mathbf{b}^T\mathbf{X}\big) = \mathbf{a}^T\,\Sigma\,\mathbf{b}$$

$\square$

REMARK 4. Note that $\mathbf{a}^T\,\Sigma\,\mathbf{b} = \sum_{i=1}^{p}\sum_{j=1}^{p} a_i b_j \sigma_{ij}$

Variance of a sum of variables

COROLLARY 2.

$$\mathrm{Var}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_{ij} = \sum_{i=1}^{n}\sigma_{ii} + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\sigma_{ij}$$

PROOF.
Let $p = n$, $i = j$, and $a_i \equiv 1$. $\square$

REMARK 5. For $n = 2$,

$$\mathrm{Var}(X_1 + X_2) = \sigma_{11} + \sigma_{22} + 2\sigma_{12}$$

Eigenvalues - the spectral theorem

THEOREM. [The spectral theorem]
Let $\mathbb{A}$ be a $p \times p$ symmetric matrix. Then we can find an orthogonal matrix $\mathbb{P}$ such that

$$\mathbb{P}^T \mathbb{A} \mathbb{P} = \Lambda,$$

with

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 & 0 \\ 0 & \lambda_2 & \cdots & \vdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \lambda_{p-1} & 0 \\ 0 & 0 & \cdots & 0 & \lambda_p \end{pmatrix}$$

where $\{\lambda_i,\, i = 1, \ldots, p\}$ are the eigenvalues.

Nonnegative definite [nnd] and positive definite [pd] matrices

1.2.1 Nonnegative definite and positive definite [pd] matrices

Definition 4. A quadratic symmetric $n \times n$ matrix, $\mathbb{B}$, is nonnegative definite [nnd] if $\mathbf{a}^T \mathbb{B} \mathbf{a} \geq 0$ for all n-dimensional vectors $\mathbf{a}$. The matrix is positive definite [pd] if the previous inequality is strict for all non trivial $\mathbf{a}$.

Proposition 5. A symmetric matrix $\mathbb{B}$ is nonnegative definite iff all eigenvalues are nonnegative. It is positive definite iff all eigenvalues are strictly positive.

Proof of Proposition 5.
It follows directly from the spectral theorem. $\square$

A covariance matrix is nnd

PROPOSITION 6. A $p \times p$ Σ matrix is covariance matrix if and only if it is nonnegative definite.

PROOF OF PROPOSITION 6. Suppose that Σ is a covariance matrix. Let $\mathbf{a} \in \mathbb{R}^p$. Then

$$\mathbf{a}^T \Sigma \mathbf{a} = \mathrm{Var}(\mathbf{a}^T \mathbf{X}) \geq 0.$$

Close to a Cholesky decomposition

The condition is also sufficient. Suppose that $\mathbb{\Sigma}$ is nnd. Then by the spectral theorem $\mathbb{\Sigma} = \mathbb{P}\Lambda\mathbb{P}^T$ for some orthogonal matrix $\mathbb{P}$. Since all eigenvalues are nonnegative we can write

$$\mathbb{\Sigma} = \mathbb{P}\mathbb{\Lambda}^{1/2}\mathbb{\Lambda}^{1/2}\mathbb{P}^T = (\mathbb{P}\mathbb{\Lambda}^{1/2})(\mathbb{P}\mathbb{\Lambda}^{1/2})^T = \mathbb{A}\,\mathbb{A}^T$$

where

$$\mathbb{\Lambda}^{1/2} \stackrel{\text{def}}{=} \begin{pmatrix} \lambda_1^{1/2} & 0 & \cdots & 0 & 0 \\ 0 & \lambda_2^{1/2} & \cdots & \vdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \lambda_{p-1}^{1/2} & 0 \\ 0 & 0 & \cdots & 0 & \lambda_p^{1/2} \end{pmatrix}, \qquad \mathbb{A} \stackrel{\text{def}}{=} \mathbb{P}\mathbb{\Lambda}^{1/2}$$

Let $\mathbf{Z}$ be a p dimesional vector of independent standard normally distributed variables. Define $\mathbf{X} = \mathbb{A}\,\mathbf{Z}$. Then $\mathbb{\Sigma}_{\mathbf{X}} = \mathbb{A}\,\mathbb{\Sigma}_{\mathbf{Z}}\mathbb{A}^T = \mathbb{A}\,\mathbb{A}^T = \mathbb{\Sigma}$.

$\square$

Cholesky decomposition

Problem 1.1. What is the a Cholesky decomposition? Why is it useful? Can you obtain it from $\mathbb{\Sigma} = \mathbb{A}\,\mathbb{A}^T$?

Problem 1.2. Let $\mathbb{\Sigma} > 0$. Define $\mathbb{\Sigma}^{-1/2}$. Can you it in R?

Reduced rank of the covariance matrix iff linear dependence

DEFINITION 5. There is a linear dependence in $\mathbf{X}$ if there exist a nonzero vector $\mathbf{a}$ such that $\mathbf{a}^T\mathbf{X} \equiv c$ for some constant c.

EXAMPLE 1. Suppose that $p = 3$ and $(X_1, X_2.X_3)$ is linearly dependent. Then

$$a_1X_1 + a_2X_2 + a_3X_3 = c$$

If $a_3 \neq 0$,

$$X_3 = \frac{c}{a_3} - \frac{a_1}{a_3}X_1 - \frac{a_2}{A_3}X_2$$

$$= b_0 + b_1X_1 + b_2X_2$$

Linear dependence means a singular covariance matirx

PROPOSITION 7. There is a linear dependence in $\mathbf{X}$ iff the covariance matrix has reduced rank. Lety $\mathbf{a}$ be any vector in $\mathbb{R}^p$.

PROOF OF PROPOSITION 7.
We have by the decomposition of the covariance matrix;

$$\mathrm{Var}(\mathbf{a}^T\mathbf{X}) = \mathbf{a}^T\Sigma\mathbf{a} = (\mathbb{A}^T\mathbf{a})^T(\mathbb{A}^T\mathbf{a}) = \|\mathbb{A}^T\mathbf{a}\|^2$$

$\square$

Properties of a covariance matrix

Let Σ be a $p \times p$ covariance matrix. Then

i) It is symmetric.

ii) It is nnd.

iii) All eigenvalues are nonnegative.

iv) The rank of Σ is equal to the number of strictly positive eigenvalues (counted with multiplisities).

v) It is pd iff it has full rank.

vi) If $\mathbf{X}$ has covariance matrix Σ of rank $r < p$, then we find subvectors such that $\mathbf{X} = (\mathbf{X}_{(1)}^T, \mathbf{X}_{(2)}^T)^T$ where $\mathbf{X}_{(1)}$ has dimension r and a covariance matrix with full rank while $\mathbf{X}_{(2)}$ is linear function of $\mathbf{X}_{(1)}$.

The multivariate Gaussian distribution

2. The multinormal distritbution

2.1. Definition

The p-dimensional stochastic vector $\mathbf{X}$ is multinormal distributed if

$$\mathbf{X} = \mathbf{b} + \mathbb{A}\,\mathbf{Z}$$

where $\mathbb{A}$ is $p \times m$ matrix and $\mathbf{Z}$ is m-dimesional vector of iid standard normal distributed variables. Here $r \leq p \leq m$ where r is the rank of $\mathbb{A}$. If $r = p$, $\mathbf{X}$ has ordinary multinormal distribution, otherwise $\mathbf{X}$ has generalised multinormal distribution. We write $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$ in the nonsingular case.

In this definition the multinormal distribution is a linear transformation of iid standard normal distributed varibles.

It follows immediately that the multinormal distribution is closed under linear transformations.

The density

2.2. Density

We see the $\boldsymbol{\mu} = \mathbf{b}$ and the covariance matrix for $\mathbf{X}$ is

$$\Sigma = \mathbb{A}\,\mathbb{A}^T \tag{2.4}$$

If $r = p$ then Σ is non-singular and $\mathbf{X}$ has density

$$f_{\mathbf{X}}(\mathbf{x}) = (2\pi)^{-p/2}\left|\Sigma\right|^{-p/2}\exp\left(-2^{-1}(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$$

The inverse of the covariance matrix is called the precision matrix. Each element in the covariance matrix itself depends only on a pair of stochastic variables. However, this is not true for the precision matrix.

In standard notation $\Sigma^{-1} = \{\sigma^{ij}\}$ and

$$(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu}) = \sum_i\sum_j(x_i-\mu_i)\sigma^{ij}(x_j-\mu_j)$$

Details

By the transformation formula,

$$f_{\mathbf{X}}(\mathbf{x}) = f_{\mathbf{Z}}\big(z(\mathbf{x})\big)\,\Big|\frac{\partial \mathbf{z}}{\partial \mathbf{x}}\Big| \tag{2.5}$$

We have

$$f_{\mathbf{Z}}(\mathbf{z}) = (2\pi)^{-p/2}\exp\big(-2^{-1}\mathbf{z}^T\mathbf{z}\big)$$

The Jacobi determinant is absolute value of the determinant of the matrix and

$$\frac{\partial \mathbf{z}}{\partial \mathbf{x}} = \Big\{\frac{\partial z_i}{\partial x_j}\Big\},$$

$$\frac{\partial z_i}{\partial x_j} = \frac{\partial}{\partial x_j}\sum_{k=1}^{p} a^{ik}(x_k - b_k) = a^{ij}$$

This means that

$$\Big|\frac{\partial \mathbf{z}}{\partial \mathbf{x}}\Big| = \big|\det(\mathbb{A}^{-1})\big| = \frac{1}{\big|\det(\mathbb{A})\big|} = \det(\mathbb{A}\mathbb{A}^T)^{-1/2} = |\Sigma|^{-1/2}$$

From (2.5) and the calculations above

$$
\begin{aligned}
f_{\mathbf{X}}(\mathbf{x}) &= (2\pi)^{-p/2}|\mathbb{\Sigma}|^{-1/2}\exp\big(-2^{-1}(\mathbf{x}-\boldsymbol{\mu})^T\mathbb{A}^{-T}\mathbb{A}^{-1}(\mathbf{x}-\boldsymbol{\mu})\big) \\
&= (2\pi)^{-p/2}|\mathbb{\Sigma}|^{-1/2}\exp\big(-2^{-1}(\mathbf{x}-\boldsymbol{\mu})^T\mathbb{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\big)
\end{aligned}
$$

Chi-square distribution

PROBLEM 2.1. Let $\mathbf{X}$ be multinormal distributed with mean $\boldsymbol{\mu}$ and covariance matrix Σ. Assume that $\mathbf{X}$ has a density. Explain that

$$(\mathbf{X}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{X}-\boldsymbol{\mu}) \sim \mathcal{X}^2(p)$$

Any nonsingular covariance matrix corresponds to a multinormal denisity

Corollary 3. Any nonsingular covariance matrix defines a multinormal denisity and $\mathcal{N}(\boldsymbol{\mu}, \mathbb{\Sigma})$ is well defined.

Proof.
Let $\mathbb{\Sigma}$ be given. By the decomposition, $\mathbb{\Sigma} = \mathbb{A}\,\mathbb{A}^T$. Let $\mathbf{X} = \mathbb{A}\mathbf{Z}$. $\square$

Remark 6. You have a technical challenge if you start with the density. In that case you need to prove that it integrates to one and that $\mathbb{\Sigma}$ is the covariance matrix.

The precision matrix

EXAMPLE 2. For $p = 2$,

$$\Sigma^{-1} = \left(\sigma_{11}\sigma_{22} - \sigma_{12}^2\right)^{-1} \begin{bmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{21} & \sigma_{11} \end{bmatrix}$$

Hence σ^{ij} is fraction of two multivariate polynomials of degree $p-1$ and p, respectively.

$$\Sigma^{-1} = \left\{\sigma^{ij}, \qquad 1 \le i, j \le p\right\}$$

$$\sigma^{ij} = \frac{(-1)^{i+j}\det(\Sigma(j|i))}{\det(\Sigma)}$$

where $\Sigma(j|i)$ is equal to Σ with row j and column i removed. Note that

$$\det(\Sigma) = \sum_{j=1}^{p} \sigma_{ij}\,(-1)^{i+j}\det(\Sigma(i|j))$$

The vector consists of two subvectors

2.3. Correlation and dependence

Suppose $\mathbf{X} = (\mathbf{X}_{(1)}^T, \mathbf{X}_{(2)}^T)^T$ where the dimensions are m and k, respectively;

$$\mathbf{X} = \begin{bmatrix} X_1 \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ X_p \end{bmatrix} = \begin{bmatrix} X_1 \\ \vdots \\ X_m \\ X_{m+1} \\ \vdots \\ X_{m+k} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_{(1)} \\ \mathbf{X}_{(2)} \end{bmatrix}$$

2×2 block covariance matrix

Then

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{11} & \cdots & \sigma_{1m} & \sigma_{1,m+1} & \cdots & \sigma_{2,m+k} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{1m} & \sigma_{2,m+1} & \cdots & \sigma_{2,m+k} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{m,1} & \sigma_{m,2} & \cdots & \sigma_{mm} & \sigma_{m,m+1} & \cdots & \sigma_{m,m+k} \\ \sigma_{m+1,1} & \sigma_{m+1,2} & \cdots & \sigma_{m+1,m} & \sigma_{m+1,m+1} & \cdots & \sigma_{m+1,m+k} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{m+k,1} & \sigma_{m+k,2} & \cdots & \sigma_{m+k,m} & \sigma_{m+k,m+k} & \cdots & \sigma_{m+k,m+k} \end{pmatrix}$$

$$= \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}, \qquad \Sigma_{ij} = \mathrm{Cov}(\mathbf{X}_{(i)}, \mathbf{X}_{(j)})$$

Independence when cross covariances are zero

THEOREM 1. Suppose that $\mathbf{X}$ is multinormal distribution. Let $\mathbf{X} = (\mathbf{X}_{(1)}^T, \mathbf{X}_{(2)}^T)^T$. If the cross covariance matrix is zero, the two components are independent and multinormal distributed.

PROOF OF THEOREM 1.
Use the density in nonsingular case. $\square$

Details

Assume first that Σ is nonsingular. Show that

$$\begin{aligned} f_{\mathbf{X}_{(1)},\mathbf{X}_{(2)}}(\mathbf{x}_{(1)},\mathbf{x}_{(2)}) &= g_1(\mathbf{x}_{(1)})g_2(\mathbf{x}_{(2)}) \\ &= f_{\mathbf{X}_{(1)}}(\mathbf{x}_{(1)})\, f_{\mathbf{X}_{(2)}}(\mathbf{x}_{(1)}) \end{aligned}$$

(look at the multinormal density)

General proof

Problem 2.2. Prove Theorem 1 without assuming that $\mathbf{X}$ has a density.

All the $2^p - 1$ subvectors are multinormal

THEOREM 2. Suppose that $\mathbf{X}$ is multinormal distribution. Then any subvector is multinormal distributed.

PROOF OF THEOREM 2.
Assume first that the covariance matrix is nonsingular. Use $\mathbb{M}$ such that $\mathbf{V}_{(1)} \stackrel{\text{def}}{=} \mathbf{X}_{(1)}$ and $\mathbf{V}_{(2)} = \mathbf{X}_{(2)} - \mathbb{M}\mathbf{X}_{(1)}$ are uncorrelated. By definition $\mathbf{V}$ is multinormal distributed and by the previous theorem $\mathbf{X}_{(1)} = \mathbf{V}_{(1)}$ is multinormal. $\square$

Details

$$\begin{bmatrix}\mathbf{V}_{(1)}\\ \mathbf{V}_{(2)}\end{bmatrix} = \begin{bmatrix}\mathbb{I} & \mathbb{0}\\ -\mathbb{M} & \mathbb{I}\end{bmatrix}\begin{bmatrix}\mathbf{X}_{(1)}\\ \mathbf{X}_{(2)}\end{bmatrix}$$

REMARK 7. The trick is not so tricky. You have already the result when the cross covariance matrix zero. It appears to be an obvious idea to start by removing the dependence between the two subvectors with an adjustment of the second one.

The role of the multinormal distribution

i) The multinormal distribution is more important for multivariate distribuitons then ordinary normal distribution among univariate distributions.

ii) The assumption that a multivariate distribution is Gaussian is relatively stronger than the univariate normality assumption.

iii) In a multinormal distribution the entire dependency is determined by the pairwise covariances. No interactions are alllowed.

iv) In a multinormal distribution two components are independent of each other iff they are uncorrelated.

v) The mean in a multivariate sample converge in distribution to a multinormal distribution whenever the covariance matrix is finite. This is the multivariate central limit theorem.

Conditioning

2.4. Conditioning in the multinormal distribution

Theorem 3. Suppose that $\mathbf{X}$ is multinormal distribution with a non-singular covariance matrix. Let $\mathbf{X} = (\mathbf{X}_{(1)}^T, \mathbf{X}_{(2)}^T)^T$. Then the conditional distribution of $\mathbf{X}_{(2)}$ given $\mathbf{X}_{(1)}$ is multinormal distributed;

$$\mathbf{X}_{(2)} \,|\, \mathbf{X}_{(1)} = \mathbf{x}_{(1)} \sim \mathcal{N}\big(\boldsymbol{\mu}_{(2)}(\mathbf{x}_{(1)}),\, \Sigma_{22\cdot 1}\big)$$

where

$$\begin{aligned} \boldsymbol{\mu}_{(2)}(\mathbf{x}_{(1)}) &\overset{\text{def}}{=} \boldsymbol{\mu}_{(2)} + \Sigma_{21}\Sigma_{11}^{-1}\big(\mathbf{x}_{(1)} - \boldsymbol{\mu}_{(1)}\big) \\ \Sigma_{22\cdot 1} &\overset{\text{def}}{=} \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \end{aligned} \tag{2.6}$$

Remark 8. Note that

i)

$$\boldsymbol{\mu}_{(2)}(\mathbf{x}_{(1)}) = \mathbb{E}\,[\mathbf{X}_{(2)}\,|\,\mathbf{X}_{(1)} = \mathbf{x}_{(1)}]$$

ii)

$$\Sigma_{22\cdot 1} = \mathrm{Var}[\mathbf{X}_{(2)}\,|\,\mathbf{X}_{(1)} = \mathbf{x}_{(1)}] \tag{2.7}$$

iii)

$$\mathbf{Z} \stackrel{\mathrm{def}}{=} \mathbf{X}_{(2)} - \mathbb{E}\,[\mathbf{X}_{(2)}\,|\,\mathbf{X}_{(1)}] \qquad \text{is independent of } \mathbf{X}_{(1)} \tag{2.8}$$

iv)

$$\mathbf{Z} \sim \mathcal{N}\big(0, \Sigma_{22\cdot 1}\big)$$

v)

$$\mathrm{Var}(\mathbb{E}\,(\mathbf{X}_{(2)}\,|\,\mathbf{X}) = \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \tag{2.9}$$

Proof

PROOF OF THEOREM 3.
The trick (again) is to transform the two subvectors to independent variables with aid of the **V**-variable,

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}_{(1)} \\ \mathbf{V}_{(2)} \end{bmatrix} = \begin{bmatrix} \mathbb{I} & \mathbb{0} \\ -\mathbb{M} & \mathbb{I} \end{bmatrix} \begin{bmatrix} \mathbf{X}_{(1)} \\ \mathbf{X}_{(2)} \end{bmatrix} \tag{2.10}$$

By the transformation formula,

$$\begin{aligned} f_{\mathbf{X}_{(1)},\mathbf{X}_{(2)}}(\mathbf{x}_{(1)}, \mathbf{x}_{(2)}) &= f_{\mathbf{V}_{(1)}}(\mathbf{x}_{(1)}) f_{\mathbf{V}_{(2)}}(\mathbf{x}_{(2)} - \mathbb{M}\mathbf{x}_{(1)}) \\ &= f_{\mathbf{X}_{(1)}}(\mathbf{x}_{(1)}) f_{\mathbf{V}_{(2)}}(\mathbf{x}_{(2)} - \mathbb{M}\mathbf{x}_{(1)}) \end{aligned} \tag{2.11}$$

Therefore,

$$
\begin{aligned}
f_{\mathbf{X}_{(2)}\,|\,\mathbf{X}_{(1)}}(\mathbf{x}_{(2)}\,|\,\mathbf{x}_{(1)}) &= \frac{f_{\mathbf{X}_{(1)},\mathbf{X}_{(2)}}(\mathbf{x}_{(1)},\mathbf{x}_{(2)})}{f_{\mathbf{X}_{(1)}}(\mathbf{x}_{(1)})} \\
&= \frac{f_{\mathbf{X}_{(1)}}(\mathbf{x}_{(1)})f_{\mathbf{V}_{(2)}}(\mathbf{x}_{(2)} - \mathbb{M}\mathbf{x}_{(1)})}{f_{\mathbf{X}_{(1)}}(\mathbf{x}_{(1)})} \\
&= f_{\mathbf{V}_{(2)}}(\mathbf{x}_{(2)} - \mathbb{M}\mathbf{x}_{(1)})
\end{aligned}
$$

Now

$$
\mathbf{V}_{(2)} = \boldsymbol{\mu}_{(2)} - \mathcal{M}\boldsymbol{\mu}_{(1)} + \mathbf{Z} \qquad\qquad \text{with } \mathbf{Z} \text{ defined by (2.8)}
$$

From (2.8),

$$
\begin{aligned}
\mathbf{X}_{(2)} &= \mathbb{E}\left[\mathbf{X}_{(2)} \,|\, \mathbf{X}_{(1)} = \mathbf{x}_{(1)}\right] + \mathbf{Z} \\
&= \boldsymbol{\mu}_{(2)} + \mathbb{M}(\mathbf{X}_{(1)} - \boldsymbol{\mu}_{(1)}) + \mathbf{Z} && \text{by def. of } \mathbb{M} \text{ the variables are independent} \\
\mathbb{\Sigma}_{22} &= \mathbb{M}\,\mathbb{\Sigma}_{11}\mathbb{M}^T + \mathrm{Var}(\mathbf{Z}) \\
&= \mathbb{\Sigma}_{21}\mathbb{\Sigma}_{11}^{-1}\mathbb{\Sigma}_{12} + \mathrm{Var}(\mathbf{Z})
\end{aligned}
$$

Hence

$$
\mathrm{Var}(\mathbf{Z}) = \mathbb{\Sigma}_{22\cdot 1}
$$

$$
\mathbf{V}_{(2)} \sim \mathcal{N}\left(\boldsymbol{\mu}_{(2)} - \mathcal{M}\boldsymbol{\mu}_{(1)}, \mathbb{\Sigma}_{22\cdot 1}\right)3
$$

$\square$

Details

Alternatively we could use

$$\mathbf{X}_{(2)} = \mathbb{E}\,(\mathbf{X}_{(2)}\,|\,\mathbf{X}_{(1)}) + \mathbf{Z}$$
$$\mathbf{Z} \sim \mathcal{N}\big(\mathbf{0}, \Sigma_{22\cdot 1}\big) \qquad \text{indep. of } \mathbf{X}_{(1)}$$

$$\begin{aligned}
\mathbb{P}(\mathbf{X}_{(1)} \le \mathbf{x}_{(1)}\,|\,\mathbf{X}_{(1)} = \mathbf{x}_{(1)}) &= \mathbb{P}\Big(\boldsymbol{\mu}_{(2)}(\mathbf{x}_{(1)}) + \mathbf{Z} \le \mathbf{x}_{(1)}\,|\,\mathbf{X}_{(1)} = \mathbf{x}_{(1)})\Big) \\
&= \mathbb{P}\big(\boldsymbol{\mu}_{(2)}(\mathbf{x}_{(1)}) + \mathbf{Z} \le \mathbf{x}_{(1)}\big) \\
&= \mathbb{P}\big(\mathbf{Z} \le \mathbf{x}_{(1)} - \boldsymbol{\mu}_{(2)}(\mathbf{x}_{(1)})\big)
\end{aligned}$$

It works although is it is multivariate inequality. It is done at the component level.

Conditional covariances

REMARK 9.

$$\sigma_{ij\cdot 1} = \sigma_{ij} - \boldsymbol{\sigma}_{i1}\Sigma_{11}^{-1}\boldsymbol{\sigma}_{1j}$$

$$\boldsymbol{\sigma}_{i1} = [\sigma_{i1}, \ldots, \sigma_{im}] \text{ for } \mathbf{X}_{(1)} = (X_1, \ldots, X_m)^T$$

REMARK 10.

$$\begin{aligned}\sigma_{ij\cdot 1} &= \rho_{ij}\sigma_{ii}^{1/2}\sigma_{ij}^{1/2} - \boldsymbol{\rho}_{i1}\sigma_{ii}^{1/2}\mathbb{D}_1^{1/2}\Sigma_{11}^{-1}\mathbb{D}_1^{1/2}\boldsymbol{\rho}_{1j}\sigma_{jj}^{1/2} \\ &= \sigma_{ii}^{1/2}\sigma_{ij}^{1/2}\Big(\rho_{ij} - \boldsymbol{\rho}_{i1}\mathbb{D}_1^{1/2}\Sigma_{11}^{-1}\mathbb{D}_1^{1/2}\boldsymbol{\rho}_{1j}\Big) \\ &= \sigma_{ii}^{1/2}\sigma_{ij}^{1/2}\Big(\rho_{ij} - \boldsymbol{\rho}_{i1}\mathcal{R}_{11}^{-1}\boldsymbol{\rho}_{1j}\Big)\end{aligned}$$

with

$$\mathbb{D}_1 = \operatorname{diag}\Sigma_{11}$$

Linear predictor

REMARK 11.

i) The predictor, $\mathbb{E}[\mathbf{X}_{(2)} \,|\, \mathbf{X}_{(1)}]$ of $\mathbf{X}_{(2)}$ is a linear function of the components of $\mathbf{X}_{(1)}$.

ii) The conditional variance of $\mathbf{X}_{(2)}$ given $\mathbf{X}_{(1)}$ is the predictor variance. It does not on the dependent variable, i.e.

$$
\begin{aligned}
&\mathrm{Var}(\mathbf{X}_{(2)} \,|\, \mathbf{X}_{(1)}) \\
&\qquad = \mathbb{E}\,\mathrm{Var}(\mathbf{X}_{(2)} \,|\, \mathbf{X}_{(1)}) \\
&\qquad = \mathbb{E}\left[\left(\mathbf{X}_{(2)} - \mathbb{E}\,[\mathbf{X}_{(2)} \,|\, \mathbf{X}_{(1)}]\right)\left(\mathbf{X}_{(2)} - \mathbb{E}[\mathbf{X}_{(2)} \,|\, \mathbf{X}_{(1)}]\right)^T\right] \qquad (2.12)\\
&\qquad = \mathrm{Var}\left(\mathbf{X}_{(2)} - \mathbb{E}\,[\mathbf{X}_{(2)} \,|\, \mathbf{X}_{(1)}]\right)
\end{aligned}
$$

iii) We can write

$$\mathbf{X}_{(2)} = \boldsymbol{\mu}_{(2)} + \boldsymbol{\beta}^T \left(\mathbf{X}_{(1)} - \boldsymbol{\mu}_{(1)}\right) + \mathbf{Z}$$

and the residual Z is independent of the explanatory variable $\mathbf{X}_{(1)}$. This is the population multivariate multiple regression formula.

Partial correlation

2.5. Partial correlation in the multinormal distribution

i) Partial correlations in the multinormal distribution are the correlations derived from a conditional covariance matrix.

ii) Partial correlation in time series are defined in terms of the covariance structure. Hence the formulas can be derived under a Gaussian assumption.

iii) The conditional covariance matrix in the multinormal distribution is independent of the conditioning variable.

Normalisation

Let $\mathbb{D}_{2\cdot 1} = \operatorname{diag} \Sigma_{22\cdot(1)}$ and

$$\mathbb{R}_{22|(1)} = \mathbb{D}_{2\cdot 1}^{-1/2} \Sigma_{22\cdot 1} \mathbb{D}_{2\cdot 1}^{-1/2} = \left\{ \rho_{ij\cdot(1)} \right\}$$

This is the conditional correlation matrix, i.e. the correlations in the conditional distribution.

Remark 12. Note the normalisation going from conditional variances to partial correlations is not based on the diagonal matrix of ordinary covariances, but on conditional covariances which are somewhat more complicated.

Calculations

From (2.7) we get, for $(i, j) \in (2)$,

$$\rho_{ij\cdot 1} = \frac{\mathrm{Cov}\big\{X_i - \mathbb{E}\,(X_i \,|\, \mathbf{X}_{(1)}),\, X_j - \mathbb{E}\,(X_j \,|\, \mathbf{X}_{(1)})\big\}}{\mathrm{Var}^{1/2}\big(X_i - \mathbb{E}\,[X_i \,|\, \mathbf{X}_{(1)}]\big)\mathrm{Var}^{1/2}\big(X_j - \mathbb{E}\,[X_j \,|\, \mathbf{X}_{(1)}]\big)}$$

where

$$\mathrm{Var}\big[X_i - \mathbb{E}\,(X_i \,|\, \mathbf{X}_{(1)})\big] = \sigma_{ii} - \boldsymbol{\sigma}_{i(1)}\Sigma_{11}^{-1}\boldsymbol{\sigma}_{(1)i}$$

$$\mathrm{Cov}\big\{X_i - \mathbb{E}\,(X_i \,|\, \mathbf{X}_{(1)}),\, X_j - \mathbb{E}\,(X_j \,|\, \mathbf{X}_{(1)})\big\} = \sigma_{ij} - \boldsymbol{\sigma}_{i(1)}\Sigma_{11}^{-1}\boldsymbol{\sigma}_{(1)j}$$

The formula for the partial correlation coefficient

We also want to replace covariances with correlations,

$$\boldsymbol{\sigma}_{i(1)} = \sigma_{ii}^{1/2} \boldsymbol{\rho}_{i(1)} \mathbb{D}_1^{1/2}, \qquad \Sigma_{11} = \mathbb{D}_1^{1/2} \mathcal{R}_{11} \mathbb{D}_1^{1/2}, \qquad \Sigma_{11}^{-1} = \mathbb{D}_1^{-1/2} \mathcal{R}_{11}^{-1} \mathbb{D}_1^{-1/2}$$

giving

$$\boldsymbol{\sigma}_{i(1)} \Sigma_{11}^{-1} \boldsymbol{\sigma}_{(1)i} = \boldsymbol{\rho}_{j(1)} \mathcal{R}_{11}^{-1} \boldsymbol{\rho}_{(1)j}$$

We combine these calculations and find that

$$\rho_{ij\cdot(1)} = \frac{\rho_{ij} - \boldsymbol{\rho}_{i(1)} \mathcal{R}_{11}^{-1} \boldsymbol{\rho}_{(1)j}}{\left(1 - \boldsymbol{\rho}_{i(1)} \mathcal{R}_{11}^{-1} \boldsymbol{\rho}_{(1)i}\right)^{1/2} \left(1 - \boldsymbol{\rho}_{j(1)} \mathcal{R}_{11}^{-1} \boldsymbol{\rho}_{(1)j}\right)^{1/2}}$$

A well known formula

EXAMPLE 3. Suppose that $\mathbf{X} = (X_1, X_2, X_3)$ with $\mathbf{X}_{(1)} = X_1$ and $\mathbf{X}_{(2)} = (X_1, X_2)^T$. Then $\mathcal{R}_{11} = 1$ and

$$\rho_{23\cdot(1)} = \frac{\rho_{23} - \rho_{12}\rho_{13}}{\left(1 - \rho_{12}^2\right)^{1/2}\left(1 - \rho_{13}^2\right)^{1/2}}$$

This is a famous formula.

REMARK 13. Suppose that $\mathbf{X}$ is a stochastic random vector with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ. Then we use all formulas from the multinormal distribution since these formulas depends only on the covariance structure and the mean.

Projection notation

2.6. Projection notation

Let

$$\langle U, V \rangle = \mathbb{E}\, UV$$
$$\|U - V\|^2 = \mathbb{E}\,(U - V)^2 = \langle U - V, U - V \rangle$$

Projection notation for the partial correlation coefficient

Let $\mathcal{P}(U) \stackrel{\text{def}}{=} \mathbb{E}\,(U \,|\, \mathbf{X}_{(1)})$. Then

$$
\begin{aligned}
\rho_{ij\cdot 1} &= \frac{\operatorname{Cov}\big\{X_i - \mathbb{E}\,(X_i \,|\, \mathbf{X}_{(1)}),\, X_j - \mathbb{E}\,(X_j \,|\, \mathbf{X}_{(1)})\big\}}{\operatorname{Var}^{1/2}\big(X_i - \mathbb{E}\,[X_i \,|\, \mathbf{X}_{(1)}]\big)\, \operatorname{Var}^{1/2}\big(X_j - \mathbb{E}\,[X_j \,|\, \mathbf{X}_{(1)}]\big)} \\
&= \frac{\langle X_i - \mathbb{E}\,(X_i \,|\, \mathbf{X}_{(1)}), X_j - \mathbb{E}\,(X_j \,|\, \mathbf{X}_{(1)})\rangle}{\|X_i - \mathbb{E}\,[X_i \,|\, \mathbf{X}_{(1)}]\| \; \|X_j - \mathbb{E}\,[X_j \,|\, \mathbf{X}_{(1)}]\|} \\
&= \frac{\langle X_i - \mathcal{P}(X_i), X_i - \mathcal{P}(X_j)\rangle}{\|X_i - \mathcal{P}(X_i)\| \|X_j - \mathcal{P}(X_j)\|} \\
&= \frac{\langle X_i, X_j\rangle - \langle \mathcal{P}(X_i), \mathcal{P}(X_j)\rangle}{\|X_i - \mathcal{P}(X_i)\| \|X_j - \mathcal{P}(X_j)\|}
\end{aligned}
$$

The numerator is a version of Pythagoras.

Multiple correlation coeffisient

2.7. Multiple correlation coefficient

Definition 6. The multiple correlation coeffisient for X_i with respect to $\mathbf{X}_{-\mathbf{i}}$ is

$$\rho_{i,-\mathbf{i}} = \underset{\boldsymbol{\beta}}{\operatorname{argmax}} \operatorname{Corr}\big(X_i, \boldsymbol{\beta}^T \mathbf{X}_{-\mathbf{i}}\big)$$

Proposition 8.

$$\rho^2_{i,-\mathbf{i}} = \frac{\sigma_{i,-\mathbf{i}} \Sigma^{-1}_{-\mathbf{i}} \sigma_{i,-\mathbf{i}}}{\sigma_{ii}} = \boldsymbol{\rho}_{i,-\mathbf{i}} \mathcal{R}^{-1}_{-\mathbf{i}} \boldsymbol{\rho}_{i,-\mathbf{i}}$$

Proof

PROOF OF PROPOSITION 8.

$$\mathrm{Corr}^2\left(X_i, \boldsymbol{\beta}^T \mathbf{X}_{-\mathbf{i}}\right) = \frac{\mathrm{Cov}^2\left(X_i, \boldsymbol{\beta}^T \mathbf{X}_{-\mathbf{i}}\right)}{\sigma_{ii}\, \boldsymbol{\beta}^T \Sigma_{-\mathbf{i}} \boldsymbol{\beta}} = \frac{\left(\boldsymbol{\sigma}_{i,-\mathbf{i}} \Sigma_{-\mathbf{i}}^{-1} \boldsymbol{\beta}\right)^2}{\sigma_{ii}\, \boldsymbol{\beta}^T \Sigma_{-\mathbf{i}} \boldsymbol{\beta}}$$

Let $\mathbf{b} = \Sigma_{-\mathbf{i}}^{-1/2} \boldsymbol{\beta}$. Then

$$\mathrm{Corr}^2\left(X_i, \boldsymbol{\beta}^T \mathbf{X}_{-\mathbf{i}}\right) = \frac{1}{\sigma_{ii}} \frac{\left(\boldsymbol{\sigma}_{i,-\mathbf{i}} \Sigma_{-\mathbf{i}}^{-1/2} \mathbf{b}\right)^2}{\mathbf{b}^T \mathbf{b}}$$

The maximum is attained for $\mathbf{b} \propto \Sigma_{-\mathbf{i}}^{-1/2} \boldsymbol{\sigma}_{-\mathbf{i},i}$ by Cauchy-Schwartz. $\square$

Multiple correlation coefficient in projection notation

$$\rho_{i,-\mathbf{i}} = \frac{\langle \mathcal{P}_{-\mathbf{i}}(X_i), \mathcal{P}_{-\mathbf{i}}(X_i) \rangle}{\langle X_i, X_i \rangle}$$

Here we assume that the expectation of $\mathbf{X}$ is zero.

Regression model

In a somewhat abuse notation, let $\mathbf{X} = \mathbf{X}_{(1)}$ and $Y = \mathbf{X}_{(2)}$. Then

$$\Sigma_{(\mathbf{X},Y)} = \begin{bmatrix} \Sigma_{\mathbf{X}} & \Sigma_{\mathbf{X}Y} \\ \Sigma_{Y\mathbf{X}} & \sigma_Y^2 \end{bmatrix}$$

COROLLARY 4. In a multiple regression model, $R^2 = \rho^2_{Y,\mathbf{X}}$ and

$$\rho^2_{Y,\mathbf{X}} = \frac{\langle \mathcal{P}_{\mathbf{X}}(Y), \mathcal{P}_{\mathbf{X}}(Y) \rangle}{\langle Y, Y \rangle}$$

PROOF.
By def. $\square$

REMARK 14. This is without data, otherwise R^2 is the empirical version of $\rho^2_{Y,\mathbf{X}}$.

Empirical case

$$\mathbf{Y} = \mathbf{1}\xi + \mathbb{X}\boldsymbol{\beta} + \boldsymbol{e}$$

with $\mathbb{X}^T\mathbf{1} = \mathbf{0}$.

$$R^2 = \frac{\mathbf{Y}^T\mathbb{X}(\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T\mathbf{Y}}{(\mathbf{Y} - \mathbf{1}\xi)^T(\mathbf{Y} - \mathbf{1}\xi)} \tag{2.13}$$

Let $\mathcal{P}_{\mathbb{X}}$ be the projection of $\mathbf{Y}$ onto the subspace spanned by the columns of $\mathbb{X}$;

$$\mathcal{P}_{\mathbb{X}} = \mathbb{X}(\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T$$

This is a projection matrix since

i) It is symmetric.

ii) It is idempotent: $\mathcal{P}_{\mathbb{X}}^2 = \mathcal{P}_{\mathbb{X}}$.

Geometric projection

The projection is the ordinary propjection in $\mathbb{R}^n$ with inner product $\langle \mathbf{a}, \mathbf{b} \rangle = \mathbf{a}^T \mathbf{b}$. In this notation (2.13) can be written as

$$R^2 = \frac{\langle \mathcal{P}_{\mathbb{X}}(\mathbf{Y}), \mathcal{P}_{\mathbb{X}}(\mathbf{Y}) \rangle}{\langle \mathbf{Y} - \mathbf{1}\xi, \mathbf{Y} - \mathbf{1}\xi \rangle}$$

In the empirical case the projection is purely geometric and does not depend on any distributional assumption. In fact is does not take into account that the data is produced in an random way. If $\mathbb{X}$ has full rank

Total variance formula

2.8. Total variance formula

Definition 7. Let $\mathbb{A}$ and $\mathbb{B}$ be nonnegative definite matrices. Then we define $\mathbb{A} \geq \mathbb{B}$ if $\mathbb{C} \stackrel{\text{def}}{=} \mathbb{A} - \mathbb{B}$ is nonnegative definite. Strict inequality means positive definite.

Proposition 9. Assume that $\mathbb{\Sigma} > 0$. Then

i) $\mathbb{\Sigma}_{22} > \mathbb{\Sigma}_{22\cdot 1} > 0$

ii)

$$\mathbb{\Sigma}_{22} = \underbrace{\mathrm{Var}\big(\mathbb{E}\,[\mathbf{X}_{(2)}\,|\,\mathbf{X}_{(1)}]\big)}_{\mathbb{\Sigma}_{12}\mathbb{\Sigma}_{11}^{-1}\mathbb{\Sigma}_{21}} + \underbrace{\mathbb{E}\,\mathrm{Var}[\mathbf{X}_{(2)}\,|\,\mathbf{X}_{(1)}]}_{\mathbb{\Sigma}_{22\cdot 1}}$$

Remark 15. This is true for any multivariate distribution with covariance matrix $\mathbb{\Sigma}$.

Proof of total variance formula

PROOF OF PROPOSITION 9.
We can write

$$\mathbf{X}_{(2)} = \mathbb{E}\left[\mathbf{X}_{(2)} \,|\, \mathbf{X}_{(1)}\right] + \underbrace{\mathbf{X}_{(2)} - \mathbb{E}\left[\mathbf{X}_{(2)} \,|\, \mathbf{X}_{(1)}\right]}_{\mathbf{Z}}$$

with the right hand side is a sum of two uncorrelated variables and

$$\mathrm{Var}(\mathbf{Z}) = \mathbb{E}\,\mathrm{Var}(\mathbf{X}_{(2)} \,|\, \mathbf{X}_{(1)})$$

$\square$

Other result for the multinormal distribution

2.9. Other result

Remember that

$$\Sigma^{-1} = \{\sigma_{ij}\}^{-1} = \{\sigma^{ij}\}$$

$$\Sigma^{-1} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}, \qquad , \qquad \Sigma^{-1} = \begin{bmatrix} \Sigma^{11} & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} \end{bmatrix}$$

where, for instance,

$$\Sigma^{22} = \{\sigma^{ij}\} \neq \{\sigma_{ij}\}^{-1} = \Sigma_{22}^{-1}$$

The inverse of the dot matrix

PROPOSITION 10. Assume that $\Sigma > 0$. Then

$$\Sigma^{22} = \Sigma_{22\cdot 1}^{-1}$$

PROOF OF PROPOSITION 10.
See the Appendix. $\square$

COROLLARY 5.

$$\sigma^{ii} = \frac{1}{\sigma_{ii}\left(1 - \rho_{i,-\mathbf{i}}^{2}\right)}$$

Proof.

$$\sigma^{ii} = \frac{1}{\sigma_{ii} - \sigma_{i,-\mathbf{i}}\Sigma_{-\mathbf{i}}^{-1}\sigma_{i,-\mathbf{i}}}$$

and

$$\rho_{i,-\mathbf{i}}^{2} = \frac{\sigma_{i,-\mathbf{i}}\Sigma_{-\mathbf{i}}^{-1}\sigma_{i,-\mathbf{i}}}{\sigma_{ii}}$$

so that

$$\sigma_{i,-\mathbf{i}}\Sigma_{-\mathbf{i}}^{-1}\sigma_{i,-\mathbf{i}} = \sigma_{ii}\rho_{i,-\mathbf{i}}^{2}$$

$\square$

Corollary 6.

$$\sigma^{ii} \geq \frac{1}{\sigma_{ii}}$$

with equality iff X_i is independent of $\mathbf{X_{-i}}$.

Proof

Proof.
In this case $(2) = i$ and

$$\mathrm{Var}(X_i \,|\, \mathbf{X}_{-\mathbf{i}}) = \sigma_{ii} - \boldsymbol{\sigma}_{i\cdot-\mathbf{i}}\Sigma_{-\mathbf{i}}^{-1}\boldsymbol{\sigma}_{-\mathbf{i}\cdot i} \leq \sigma_{ii}$$

and

$$\sigma^{ii} = \frac{1}{\mathrm{Var}(X_i \,|\, \mathbf{X}_{-\mathbf{i}})} \geq \frac{1}{\sigma_{ii}}$$

$\square$

Determinant formula

COROLLARY 7.

$$\det(\Sigma) = \det(\Sigma_{11}) \det(\Sigma_{22.1})$$

PROOF.
Use the **V**- transformation defined by (2.10) and (2.11) . It has Jacobi determinant 1.

$$\mathrm{Var}(\mathbf{V}) = \begin{bmatrix} \Sigma_{11} & \mathbb{0} \\ \mathbb{0} & \Sigma_{22\cdot(1)} \end{bmatrix}$$

$\square$

By combining the two corollaries we get

Corollary 8.

$$\det(\Sigma) = \frac{\det(\Sigma_{11})}{\det(\Sigma^{22})}$$

The block inverse covariance matrix

2.10. 2×2 BLOCK SYMMETRIC MATRIX AND ITS 2×2 BLOCK INVERSE

The matrix $\mathbb{A}$ is symmetric and the diagonal blocks are quadratic.

$$\mathbb{A} = \begin{bmatrix} \mathbb{A}_{11} & \mathbb{A}_{12} \\ \mathbb{A}_{21} & \mathbb{A}_{22} \end{bmatrix}, \quad \mathbb{B} = \mathbb{A}^{-1} = \begin{bmatrix} \mathbb{B}_{11} & \mathbb{B}_{12} \\ \mathbb{B}_{21} & \mathbb{B}_{22} \end{bmatrix}$$

We want to express $\mathbb{B}$ in terms of $\mathbb{A}$. From $\mathbb{B}\,\mathbb{A} = \mathbb{I}$ we get two equations for first the row in $\mathbb{B}$:

$$\mathbb{B}_{11}\mathbb{A}_{11} + \mathbb{B}_{12}\mathbb{A}_{21} = \mathbb{I}$$
$$\mathbb{B}_{11}\mathbb{A}_{12} + \mathbb{B}_{12}\mathbb{A}_{22} = \mathbb{0}$$

The formula

Let

$$\mathbb{A}_{11\cdot 2} = \mathbb{A}_{11} - \mathbb{A}_{12}\mathbb{A}_{22}^{-1}\mathbb{A}_{21}, \;\; \mathbb{A}_{1\cdot 2} = \mathbb{A}_{12}\mathbb{A}_{22}^{-1}$$

From the second equation we get $\mathbb{B}_{12} = -\mathbb{B}_{11}\mathbb{A}_{12}\mathbb{A}_{22}^{-1}$ and inserting in the first equation, this gives

$$\mathbb{B}_{11}\left\{\mathbb{A}_{11} - \mathbb{A}_{12}\mathbb{A}_{22}^{-1}\mathbb{A}_{21}\right\} = \mathbb{I} \quad \Rightarrow \quad \mathbb{B}_{11}\mathbb{A}_{11\cdot 2} = I \quad \Rightarrow \quad \mathbb{B}_{11} = \mathbb{A}_{11\cdot 2}^{-1}$$

The off diagonal block in $\mathbb{B}$

We have

$$\mathbb{B}_{12} = -\mathbb{B}_{11}\mathbb{A}_{12}\mathbb{A}_{22}^{-1} = -\mathbb{A}_{11\cdot 2}^{-1}\mathbb{A}_{1\cdot 2}$$

Due to symmetry

$$\mathbb{B}_{21} = -\mathbb{A}_{22\cdot 1}^{-1}\mathbb{A}_{2\cdot 1}$$

and

$$\mathbb{A}^{-1} = \begin{bmatrix} \mathbb{A}^{11} & \mathbb{A}^{12} \\ \\ \mathbb{A}^{21} & \mathbb{A}^{22} \end{bmatrix} = \begin{bmatrix} \mathbb{A}_{11\cdot 2}^{-1} & -\mathbb{A}_{11\cdot 2}^{-1}\mathbb{A}_{1\cdot 2} \\ \\ -\mathbb{A}_{22\cdot 1}^{-1}\mathbb{A}_{2\cdot 1} & \mathbb{A}_{22\cdot 1}^{-1} \end{bmatrix}$$