Lecture note 8: Multivariate normal distribution

Foundation of Probability Theory/STA 203

Zhuosong ZHANG

Department of Statistics and Data Science, SUSTech

Fall, 2023

Bivariate normal distribution

Introduction

The bivariate normal distribution is commonly used to model the joint distribution of two random variables with a linear relationship. Here are some real-world examples where the bivariate normal distribution might be a reasonable model:

- Height and weight of adults
- Father and son's heights
- Test scores in two subjects
- We can assume that these variables both have a marginal normal distribution.
- However, there are some correlation between them.

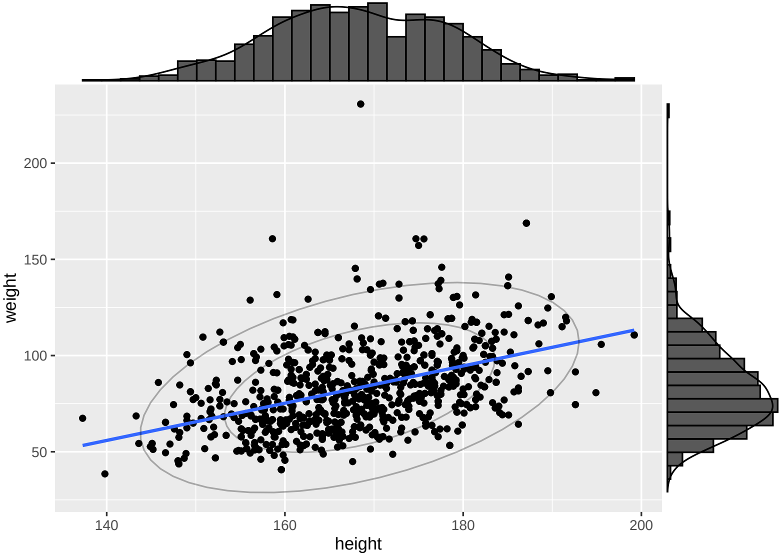


Figure: This is survey data collected by the US National Center for Health Statistics (NCHS)

- Let X be the height (in cm), and Y be the weight (in kg).
- What can you see from the marginal distributions of X and Y?
- Are they independent?

老忠实间歇泉（英语：Old Faithful）是一座位于美国黄石国家公园的间歇泉，为黄石国家公园第一个被命名的间歇泉。现喷发规律是80 分钟左右一次。

Let X be the waiting time (in minutes), and let Y be the duration time of the eruptions (in minutes). History data gives the following graph:

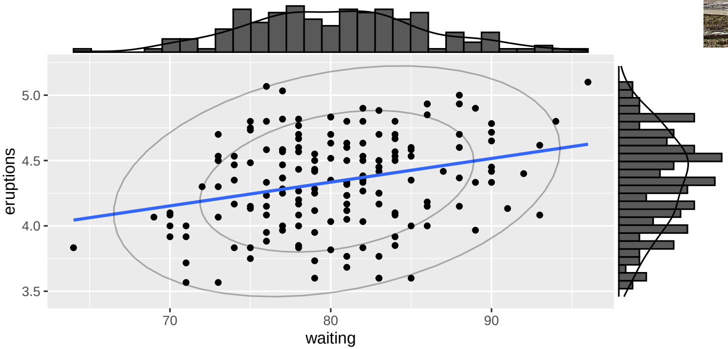


Figure: Old Faithful

Questions

- Whether X and Y are independent? Are they correlated?
- What is the joint pdf of X and Y? How about marginal pdfs?
- What is the conditional distribution of Y given that $X = 80$? How about the conditional expectation?

Joint pdf of independent normal variables

- If $X \sim N(0,1)$ and $Y \sim N(0,1)$ are independent, then

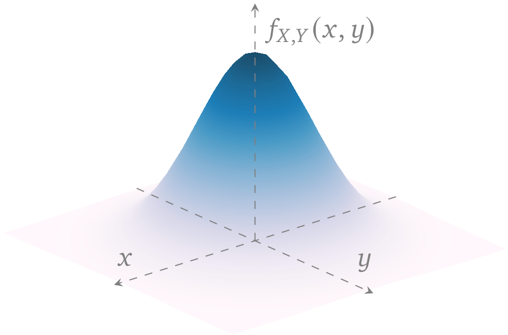
$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2}.$$

- Then the joint pdf is

$$f(x,y) = f_X(x) f_Y(y) = \frac{1}{2\pi} e^{-\frac{x^2}{2} - \frac{y^2}{2}}.$$

- We say X and Y follow a standard bivariate normal distribution.

A figure



Bivariate Normal Distribution

The bivariate normal distribution is a probability distribution that describes the joint distribution of two normally distributed variables.

Definition 1

X and Y are said to be bivariate normally distributed with means μ_X and μ_Y and variances σ_X^2 and σ_Y^2 respectively, and with correlation coefficient ρ, if the joint pdf of X and Y is given by:

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_X)^2}{\sigma_X^2} - 2\rho\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \frac{(y-\mu_Y)^2}{\sigma_Y^2}\right]\right)$$

Specially, if $\mu_X = \mu_Y = \rho = 0$ and $\sigma_X = \sigma_Y = 1$, then it is said to be a standard bivariate normal distribution.

Z-scores

Definition 2

The z-score of a random variable X is defined as

$$X^* = \frac{X - \mu}{\sigma},$$

where $\mu = \mathbb{E}[X]$ and $\sigma^2 = \mathrm{Var}(X)$.

Remark

It can be shown that if $X \sim N(\mu, \sigma^2)$, then

$$X^* = \frac{X - \mu}{\sigma} \sim N(0, 1).$$

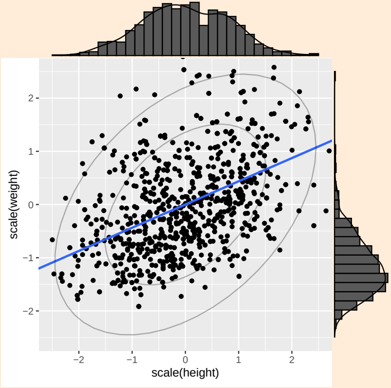
Distribution of the Z scores

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Proposition 3

If X and Y follow the bivariate normal distribution with parameters $(\mu_X, \mu_Y; \sigma_X^2, \sigma_Y^2, \rho)$, then X^* and Y^* follow the bivariate normal distribution with parameters $(0, 0; 1, 1, \rho)$, where

$$\rho = \mathrm{Cor}(X, Y) = \mathrm{Cor}(X^*, Y^*) = \mathrm{Cov}(X^*, Y^*).$$



Proof.

Let

$$x^* = g(x, y) = \frac{x - \mu_X}{\sigma_X}, \quad y^* = h(x, y) = \frac{y - \mu_Y}{\sigma_Y},$$

then

$$|J| = \begin{vmatrix} \frac{1}{\sigma_X} & 0 \\ 0 & \frac{1}{\sigma_Y} \end{vmatrix} = \frac{1}{\sigma_X \sigma_Y}.$$

Then, it follows that

$$f_{X^*,Y^*}(x^*, y^*) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left[(x^*)^2 - 2\rho x^* y^* + (y^*)^2\right]\right). \quad \blacksquare$$

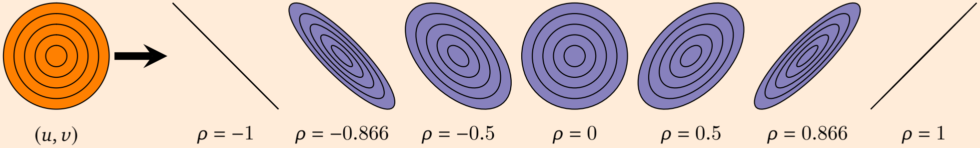
Linear transformations

Proposition 4

If $U \sim N(0,1), V \sim N(0,1)$ are independent random variables and $\rho \in [-1,1]$, and let

$$X = U, \quad Y = \rho U + \sqrt{1-\rho^2}V,$$

then (X,Y) follows the bivariate normal distribution with parameters $(0,0;1,1,\rho)$.



Proof.

Note that the joint pdf of U and V is

$$f_{U,V}(u,v) = \frac{1}{2\pi} e^{-\frac{u^2}{2} - \frac{v^2}{2}},$$

and

$$u = x, \quad v = \frac{1}{\sqrt{1-\rho^2}}(y - \rho x),$$

The Jacobian determinant is given by

$$|J| = \begin{vmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{vmatrix} = \sqrt{1-\rho^2},$$

and thus, the joint pdf of X and Y is

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{x^2}{2} - \frac{(y-\rho x)^2}{2(1-\rho^2)}\right) \quad \blacksquare$$

Marginal distributions

Proposition 5

If X and Y follow the bivariate normal distribution with parameters $(\mu_X, \mu_Y; \sigma_X^2, \sigma_Y^2, \rho)$, then the marginal distributions of X and Y are given by

$$X \sim N(\mu_X, \sigma_X^2), \quad \text{and} \quad Y \sim N(\mu_Y, \sigma_Y^2),$$

respectively.

Remark

It follows that

$$\mathbb{E}[X] = \mu_X, \quad \mathrm{Var}(X) = \sigma_X^2, \quad \mathbb{E}[Y] = \mu_Y, \quad \mathrm{Var}(Y) = \sigma_Y^2.$$

As $\mathrm{Cor}(X, Y) = \rho$, we have

$$\mathrm{Cov}(X, Y) = \rho\sigma_X\sigma_Y.$$

Proof.

We only show for the case where $\mu_X = \mu_Y = 0$ and $\sigma_X = \sigma_Y = 1$. In this case,

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{x^2}{2} - \frac{(y-\rho x)^2}{2(1-\rho^2)}\right) dy \\ &= \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp\left(-\frac{(y-\rho x)^2}{2(1-\rho^2)}\right) dy \\ &= \frac{1}{\sqrt{2\pi}} e^{-x^2/2}. \end{aligned}$$

■

Conditional distribution

Proposition 6

If X and Y follow the bivariate normal distribution with parameters $(\mu_X, \mu_Y; \sigma_X^2, \sigma_Y^2, \rho)$, then

$$Y|X = x \sim N\left(\mu_Y + \frac{\rho\sigma_Y}{\sigma_X}(x - \mu_X), (1-\rho^2)\sigma_Y^2\right),$$

or, equivalently,

$$Y^*|X^* = x^* \sim N(\rho x^*, 1-\rho^2),$$

where (X^*, Y^*) is the z-score of (X, Y).
Moreover,

$$\mathbb{E}[Y|X = x] = \mu_Y + \frac{\rho\sigma_Y}{\sigma_X}(x - \mu_X), \quad \mathrm{Var}(Y|X = x) = (1-\rho^2)\sigma_Y^2.$$

Linear regression

- Usually, the joint distribution of (X, Y) is unknown.
- The regression function

$$h(x) = \mathbb{E}[Y|X = x]$$

  is also unknown.
- However, if we assume that (X, Y) is a bivariate normal random vector, then

$$h(x) = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X) = b_0 + b_1 x,$$

  where

$$b_0 = \mu_Y - b_1\mu_X, \quad b_1 = \rho\frac{\sigma_Y}{\sigma_X}.$$

Examples

Example 7

Assume that the height and weight of a randomly chosen adult, X and Y, follow a bivariate normal distribution with parameters

$$\mu_X = 168.84, \mu_Y = 82.05; \sigma_X^2 = 101.74, \sigma_Y^2 = 448.84, \rho = 0.45.$$

Find (a) $\mathbb{P}\{160 < X < 180\}$. (b) $\mathbb{E}[Y|X = 170]$. (c) $\mathrm{Var}(Y|X = 180)$.

Examples

Solution.

(a) As $X \sim N(168.84, 101.74)$, then $X^* = \frac{X-168.84}{\sqrt{101.74}} \sim N(0,1)$, and hence

$$\mathbb{P}\{160 < X < 180\} = \mathbb{P}\left\{\frac{160-168.84}{\sqrt{101.74}} < X^* < \frac{180-168.84}{\sqrt{101.74}}\right\}$$
$$= \mathbb{P}\{-0.876 < X^* < 1.106\} \approx 0.675.$$

Examples

(b) We have

$$\mathbb{E}[Y|X = 170] = 82.05 + \frac{(0.45)(\sqrt{448.84})}{\sqrt{101.74}}(170 - 168.84) = 83.147.$$

(c) We have

$$\mathrm{Var}[Y|X = 180] = (1 - 0.45^2)(448.84) = 358.28.$$

Actually, we can also obtain

$$Y|X = 170 \sim N(83.147, 358.28).$$

■

Properties

Proposition 8 (Independence)

If X and Y are bivariate normal and uncorrelated, then they are independent.

Example 9

If X and Y follow the bivariate normal distribution with parameters $(\mu_X, \mu_Y; \sigma_X^2, \sigma_Y^2, \rho)$, find the joint distribution of X and $W = Y - \frac{\rho\sigma_Y}{\sigma_X}X$. Whether they are independent?

Properties

Proposition 10 (Linear combinations of X and Y)

Random variables X and Y follow the bivariate normal distribution with parameters $(\mu_X, \mu_Y; \sigma_X^2, \sigma_Y^2, \rho)$, if and only if for any $a, b \in \mathbb{R}$,

$$aX + bY \sim N(a\mu_X + b\mu_Y, a^2\sigma_X^2 + 2ab\rho\sigma_X\sigma_Y + b^2\sigma_Y^2).$$

Example 11

Let X and Y be jointly normal random variables with parameters $\mu_X = 1, \sigma_X^2 = 1, \mu_Y = 0, \sigma_Y^2 = 4$, and $\rho = 1/2$. Find (a) $\mathbb{P}\{2X + Y \leqslant 3\}$, (b) $\mathrm{Cov}(X + Y, 2X - Y)$, and (c) $\mathbb{P}\{Y > 1 | X = 2\}$.

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Solution.

(a) Since X and Y are jointly normal, then $2X+Y \sim N(2\mu_X+\mu_Y, 4\sigma_X^2+2\rho(2\sigma_X)\sigma_Y+\sigma_Y^2) = N(2,12)$. Therefore,

$$\mathbb{P}\{V \leqslant 3\} = \mathbb{P}\left\{V^* \leqslant \frac{3-2}{\sqrt{12}}\right\} \approx \Phi(0.2887) \approx 0.6136.$$

(b) Note that $\mathrm{Cov}(X,Y) = \rho\sigma_X\sigma_Y = 1$. Therefore,

$$\mathrm{Cov}(X+Y, 2X-Y) = 2\,\mathrm{Var}(X) + 2\,\mathrm{Cov}(X,Y) - \mathrm{Cov}(X,Y) - \mathrm{Var}(Y) = -1.$$

Examples

Solution (Cont'd).

(c) As

$$\mathbb{E}[Y|X=2] = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(2-\mu_X) = 1, \quad \mathrm{Var}(Y|X=2) = (1-\rho^2)\sigma_Y^2 = 3,$$

it follows that $Y|X=2 \sim N(1,3)$, and therefore,

$$\mathbb{P}\{Y > 1|X=2\} = 1 - \Phi\left(\frac{1-1}{\sqrt{3}}\right) = 0.5.$$

Notice!

If X and Y are jointly normal, then each random variable X and Y is normal. However, the converse is not true.

Example 12

Let $X \sim N(0, 1)$ and

$$W = \begin{cases} 1 & \text{with probability } 1/2 \\ -1 & \text{with probability } 1/2 \end{cases}$$

be independent random variables. Let $Y = WX$. Find the pdf of Y. Does (X, Y) bivariate normal distributed? Why? Or why not?

Solution.

By symmetry of $N(0,1)$, we have $-X \sim N(0,1)$. Therefore,

$$\begin{aligned}\mathbb{P}\{Y \leqslant y\} &= \mathbb{P}\{Y \leqslant y | W = -1\}\,\mathbb{P}\{W = -1\} + \mathbb{P}\{Y \leqslant y | W = 1\}\,\mathbb{P}\{W = 1\} \\ &= \frac{1}{2}\mathbb{P}\{X \leqslant y\} + \frac{1}{2}\mathbb{P}\{-X \leqslant y\} \\ &= \frac{1}{2}\Phi(y) + \frac{1}{2}\Phi(y) = \Phi(y).\end{aligned}$$

Hence, $Y \sim N(0,1)$.

However, X and Y are not jointly normal, because $Z = X + Y$ has the following form:

$$Z = \begin{cases} 2X & \text{if } W = 1 \\ 0 & \text{if } W = -1. \end{cases}$$

Therefore, if $z \geqslant 0$,

$$\begin{aligned}\mathbb{P}\{Z \leqslant z\} &= \mathbb{P}\{Z \leqslant z | W = 1\}\,\mathbb{P}\{W = 1\} + \mathbb{P}\{Z \leqslant z | W = -1\}\,\mathbb{P}\{W = -1\} \\ &= \frac{1}{2}\,\mathbb{P}\{X \leqslant \frac{z}{2}\} + \frac{1}{2} = \frac{1}{2}(1 + \Phi(\frac{z}{2})),\end{aligned}$$

while if $z < 0$,

$$\mathbb{P}\{Z \leqslant z\} = \frac{1}{2}\,\mathbb{P}\{X \leqslant \frac{z}{2}\} = \frac{1}{2}\Phi(\frac{z}{2}).$$

This example illustrates that although X and Y are normally distributed, it is possible that their sum Z is not normally distributed, which further implies that X and Y are not jointly normal. ■

Properties

Some important properties of the bivariate normal distribution include:

- The marginal distributions of X and Y are themselves normally distributed.
- The conditional distribution of X given $Y = y$ and the conditional distribution of Y given $X = x$ are both normally distributed with means and variances that depend on y and x respectively.
- The conditional expectation of X given $Y = y$ and the conditional expectation of Y given $X = x$ are both linear functions of y and x respectively.

Multivariate normal distribution

Multivariate Normal Distribution

The multivariate normal distribution is a probability distribution that describes the joint distribution of p normally distributed variables.

Definition 13

If $\boldsymbol{X} = (X_1, X_2, \dots, X_p)$ is a p-dimensional random vector with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, then the pdf of multivariate normal distribution is given by:

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

and we denote $\boldsymbol{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Here,

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_p \end{pmatrix} \in \mathbb{R}^p, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & \dots & \sigma_{1p} \\ \vdots & \ddots & \vdots \\ \sigma_{p1} & \dots & \sigma_{pp} \end{pmatrix} \in \mathbb{R}^{p\times p},$$

and $\boldsymbol{\Sigma}$ is a positive definite matrix. The symbol $|\boldsymbol{\Sigma}|$ is the determinant of $\boldsymbol{\Sigma}$.

Standard MND

Definition 14

Specially, if $\boldsymbol{\mu} = \mathbf{0}$, and $\boldsymbol{\Sigma} = \boldsymbol{I}_p$, then we say $\boldsymbol{X}$ follows a standard multivariate normal distribution if $\boldsymbol{X} \sim N(\mathbf{0}, \boldsymbol{I}_p)$.

Properties

Some important properties of the multivariate normal distribution include:

- Any linear combination of the components of $\boldsymbol{X}$ is also normally distributed.
- The marginal distributions of any subset of components of $\boldsymbol{X}$ are themselves multivariate normal.
- The conditional distribution of any subset of components of $\boldsymbol{X}$ given the remaining components is also multivariate normal.
- The conditional expectation of any subset of components of $\boldsymbol{X}$ given the remaining components is a linear function of the remaining components.

Properties of $f_{\boldsymbol{X}}(\boldsymbol{x})$

SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Proposition 15

We have

$$\int_{\mathbb{R}^p} f_{\boldsymbol{X}}(\boldsymbol{x}) d\boldsymbol{x} = 1.$$

Proof.

Since $\boldsymbol{\Sigma} > 0$, it follows that a non-singular matrix $\boldsymbol{L}$ such that

$$\boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{L}^T, \quad |\boldsymbol{L}| = |\boldsymbol{\Sigma}|^{1/2}.$$

Consider the transformation

$$\boldsymbol{y} = \boldsymbol{L}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}).$$

Then,

$$\boldsymbol{x} = \boldsymbol{L}\boldsymbol{y} + \mu,$$

Therefore,

$$(\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}) = \boldsymbol{y}^T \boldsymbol{y}.$$

■

Moment generating function of $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Theorem 16

The moment generating function of $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is given by

$$M(\boldsymbol{t}) = \exp\left\{\boldsymbol{\mu}^T\boldsymbol{t} + \frac{1}{2}\boldsymbol{t}^T\boldsymbol{\Sigma}\boldsymbol{t}\right\}, \quad \boldsymbol{t} \in \mathbb{R}^p.$$

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Definition 17

For $\boldsymbol{\mu} \in \mathbb{R}^p$, and $\boldsymbol{\Sigma} \in \mathbb{R}^{p\times p}$ is a non-negative definite matrix. Then $\boldsymbol{X}$ is called to follow a multivariate normal distribution if its moment generating function is

$$M(\boldsymbol{t}) = \exp\left\{\boldsymbol{\mu}^T\boldsymbol{t} + \frac{1}{2}\boldsymbol{t}^T\boldsymbol{\Sigma}\boldsymbol{t}\right\}.$$

Remark

Here, $\boldsymbol{\Sigma}$ may be degenerate, say, rank$(\boldsymbol{\Sigma}) < p$, or $|\boldsymbol{\Sigma}| = 0$. In this case, we say $\boldsymbol{X}$ follows a degenerate normal distribution, or singular normal distribution.

SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY 南方科技大学

Theorem 18

Any subvector of $\boldsymbol{X}$, say,

$$\widetilde{X} = (X_{k_1}, \ldots, X_{k_r})^T, \quad r \leqslant p,$$

also follows a normal distribution $N(\widetilde{\boldsymbol{\mu}}, \widetilde{\boldsymbol{\Sigma}})$, where

$$\widetilde{\boldsymbol{\mu}} = \begin{pmatrix} \mu_{k_1} \\ \vdots \\ \mu_{k_r} \end{pmatrix}, \quad \widetilde{\boldsymbol{\Sigma}} = \begin{pmatrix} \sigma_{k_1,k_1} & \ldots & \sigma_{k_1,k_r} \\ \vdots & \ddots & \vdots \\ \sigma_{k_r,k_1} & \ldots & \sigma_{k_r,k_r} \end{pmatrix}$$

Properties

Remark

The marginal distribution of X_j is $N(\mu_j, \sigma_{jj})$. The marginal distribution of (X_j, X_k) is

$$N\left(\begin{pmatrix}\mu_j \\ \mu_k\end{pmatrix}, \begin{pmatrix}\sigma_{jj} & \sigma_{jk} \\ \sigma_{jk} & \sigma_{kk}\end{pmatrix}\right).$$

Properties

Theorem 19

We have

$$\mu_j = \mathbb{E}[X_j], \quad \sigma_{jj} = \mathrm{Var}(X_j).$$

Moreover,

$$\sigma_{jk} = \mathrm{Cov}(X_j, X_k).$$

Independence

Theorem 20

Random variables $X_1, X_2, \ldots, X_p$ are independent, if and only if $\sigma_{jk} = 0$ for all $j \neq k$. Generally, if $\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2)$, where $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are two subvectors of $\boldsymbol{X}$, and let

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix},$$

where $\boldsymbol{\Sigma}_{11}$ and $\boldsymbol{\Sigma}_{22}$ are the covariance matrices of $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$, respectively, and

$$\boldsymbol{\Sigma}_{12} = \mathbb{E}[(\boldsymbol{X}_1 - \boldsymbol{\mu}_1)(\boldsymbol{X}_2 - \boldsymbol{\mu}_2)^T].$$

Then, $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are independent if and only if $\boldsymbol{\Sigma}_{12} = \mathbf{0}$.

Example 21

Assume that $\boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}$ follows $N\left(\begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 4 & 0 & 2 \\ 0 & 1 & -1 \\ 2 & -1 & 3 \end{pmatrix}\right)$.

Find

(a) The distributions of X_1, X_2 and X_3.

(b) The distribution of $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$.

(c) Whether X_1 and X_2 are independent?

(d) Whether X_1 and $(X_2, X_3)^T$ are independent?

Linear transformation

- Let $\boldsymbol{X} \in \mathbb{R}^p$ be any random vector (not necessarily normal), satisfying

$$\mathbb{E}[\boldsymbol{X}] = \boldsymbol{\mu}, \mathrm{Cov}(\boldsymbol{X}) = \boldsymbol{\Sigma}.$$

- Let $\boldsymbol{a} = (a_1, a_2, \ldots, a_p)^T$. Consider the linear transformation

$$Y = \sum_{j=1}^{p} a_j X_j = \boldsymbol{a}^T \boldsymbol{X}.$$

- It follows that

$$\mathbb{E}[Y] = \sum_{j=1}^{p} a_j \mu_j = \boldsymbol{a}^T \boldsymbol{\mu}.$$

- Moreover,

$$\mathrm{Var}(Y) = \sum_{j=1}^{p} \sum_{k=1}^{p} a_j a_k \sigma_{jk} = \boldsymbol{a}^T \boldsymbol{\Sigma} \boldsymbol{a}.$$

Linear transformation of $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Theorem 22

$\boldsymbol{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ if and only if

$$\boldsymbol{a}^T \boldsymbol{X} \sim N\left(\sum_{j=1}^{p} a_j \mu_j, \sum_{j=1}^{p} \sum_{k=1}^{p} a_j a_k \sigma_{jk}\right) \quad \text{for any } \boldsymbol{a} \in \mathbb{R}^p.$$

Property of transformation of $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

Theorem 23

If $\boldsymbol{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then for any $\boldsymbol{C} \in \mathbb{R}^{r\times p}$,

$$\boldsymbol{C}\boldsymbol{X} \sim N(\boldsymbol{C}\boldsymbol{\mu}, \boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}^T).$$

Theorem 24

If $\boldsymbol{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then there exists a orthogonal transformation $\boldsymbol{U}$ such that each component of $\boldsymbol{UX}$ is independent of each other. More specifically,

$$\boldsymbol{UX} \sim N(\boldsymbol{U\mu}, \boldsymbol{\Lambda}),$$

where

$$\boldsymbol{\Lambda} = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_p \end{pmatrix}$$

and λ_j's are the eigenvalues of $\boldsymbol{\Sigma}$.

Example 25

Assume that $\boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}$ follows $N\left[\begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 4 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & -1 & 1 \end{pmatrix}\right]$.

Find

(a) the distribution of $X_1 - 2X_2 + X_3$;

(b) the joint distribution of $X_1 - X_2 + X_3$ and $3X_1 + X_2 - 2X_3$;

(c) an orthogonal matrix $\boldsymbol{U}$ such that $\boldsymbol{U}\boldsymbol{X}$ has independent components.

Solution.

(a) Let $Y = \boldsymbol{a}^T\boldsymbol{X}$, where $\boldsymbol{a} = \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}$, then $Y = X_1 - 2X_2 + X_3$. Note that

$$\boldsymbol{a}^T\boldsymbol{\mu} = \begin{pmatrix} 1 & -2 & 1 \end{pmatrix}\begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix} = 3, \quad \boldsymbol{a}^T\boldsymbol{\Sigma}\boldsymbol{a} = \begin{pmatrix} 1 & -2 & 1 \end{pmatrix}\begin{pmatrix} 4 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & -1 & 1 \end{pmatrix}\begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} = 15.$$

The distribution of Y is $N(3, 15)$.

Solution

(b) Let $\boldsymbol{a}_1 = (1, -1, 1)^T$ and $\boldsymbol{a}_2 = (3, 1, -2)$, and let

$$\boldsymbol{A} = \begin{pmatrix} 1 & -1 & 1 \\ 3 & 1 & -2 \end{pmatrix}$$

Then,

$$\boldsymbol{AX} = \begin{pmatrix} X_1 - 2X_2 + X_3 \\ 3X_1 + X_2 - 2X_3 \end{pmatrix}$$

Note that

$$\boldsymbol{A\mu} = \begin{pmatrix} 1 & -1 & 1 \\ 3 & 1 & -2 \end{pmatrix}\begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 4 \\ 6 \end{pmatrix},$$

$$\boldsymbol{A\Sigma A}^T = \begin{pmatrix} 1 & -1 & 1 \\ 3 & 1 & -2 \end{pmatrix}\begin{pmatrix} 4 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & -1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 3 \\ -1 & 1 \\ 1 & -2 \end{pmatrix} = \begin{pmatrix} 8 & 6 \\ 6 & 45 \end{pmatrix}.$$

Solution.

(c) The eigenvalues of $\boldsymbol{\Sigma}$ are $4, 2$ and 0, and the eigenvectors are

$$\boldsymbol{u}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \boldsymbol{u}_2 = \begin{pmatrix} 0 \\ -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{pmatrix}, \quad \boldsymbol{u}_3 = \begin{pmatrix} 0 \\ \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{pmatrix},$$

Then, with

$$\boldsymbol{U} = (\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3)^T, \boldsymbol{\Lambda} = \text{diag}(4, 2, 0).$$

we have $\boldsymbol{\Sigma} = \boldsymbol{U}^T \boldsymbol{\Lambda} \boldsymbol{U}$. As a consequence,

$$\text{Cov}(\boldsymbol{U}\boldsymbol{X}) = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{U}^T = \boldsymbol{\Lambda}.$$

∎

Chi-squared distribution

Theorem 26

If $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where $|\boldsymbol{\Sigma}| > 0$, then

$$(\boldsymbol{X} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{X} - \boldsymbol{\mu}) \sim \chi_p^2.$$

Conditional distribution

Theorem 27

If $\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{pmatrix}$ follows a p-variate normal distribution $N(\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix})$, then

$$\boldsymbol{X}_2 | \boldsymbol{X}_1 \sim N(\boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{X}_1 - \boldsymbol{\mu}_1), \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}).$$

Example

Example 28

Let

$$\boldsymbol{X} \sim N\left[\begin{pmatrix} 2 \\ 5 \\ -2 \\ 1 \end{pmatrix}, \begin{pmatrix} 9 & 0 & 3 & 3 \\ 0 & 1 & -1 & 2 \\ 3 & -1 & 6 & -3 \\ 3 & 2 & -3 & 7 \end{pmatrix}\right].$$

Let

$$\boldsymbol{Y} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, \quad \boldsymbol{Z} = \begin{pmatrix} X_3 \\ X_4 \end{pmatrix}$$

Find the distribution of $\boldsymbol{Y}|\boldsymbol{Z} = \boldsymbol{z}$.

Solution.

Note that

$$\boldsymbol{\mu}_Y = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \quad \boldsymbol{\mu}_Z = \begin{pmatrix} -2 \\ 1 \end{pmatrix}, \quad \boldsymbol{\Sigma}_{YY} = \begin{pmatrix} 9 & 0 \\ 0 & 1 \end{pmatrix}, \quad \boldsymbol{\Sigma}_{ZZ} = \begin{pmatrix} 6 & -3 \\ -3 & 7 \end{pmatrix}, \quad \boldsymbol{\Sigma}_{YZ} = \begin{pmatrix} 3 & 3 \\ -1 & 2 \end{pmatrix} = \boldsymbol{\Sigma}_{ZY}^T.$$

Then,

$$\begin{aligned}
\mathbb{E}[\boldsymbol{Y}|\boldsymbol{Z} = \boldsymbol{z}] &= \boldsymbol{\mu}_Y + \boldsymbol{\Sigma}_{YZ}\boldsymbol{\Sigma}_{ZZ}^{-1}(\boldsymbol{z} - \boldsymbol{\mu}_X) \\
&= \begin{pmatrix} 2 \\ 5 \end{pmatrix} + \begin{pmatrix} 3 & 3 \\ -1 & 2 \end{pmatrix}\begin{pmatrix} 6 & -3 \\ -3 & 7 \end{pmatrix}^{-1}\begin{pmatrix} z_1 + 2 \\ z_2 - 1 \end{pmatrix} \\
&= \begin{pmatrix} 3 + \frac{10}{11}z_1 + \frac{9}{11}z_2 \\ \frac{14}{3} - \frac{1}{33}z_1 + \frac{3}{11}z_2 \end{pmatrix}.
\end{aligned}$$

Solution

Moreover,

$$\begin{aligned}\mathrm{Cov}(\boldsymbol{Y}|\boldsymbol{Z}=\boldsymbol{z}) &= \boldsymbol{\Sigma}_{YY} - \boldsymbol{\Sigma}_{YZ}\boldsymbol{\Sigma}_{ZZ}^{-1}\boldsymbol{\Sigma}_{ZY} \\ &= \begin{pmatrix} 9 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} 3 & 3 \\ -1 & 2 \end{pmatrix}\begin{pmatrix} 6 & -3 \\ -3 & 7 \end{pmatrix}^{-1}\begin{pmatrix} 3 & -1 \\ 3 & 2 \end{pmatrix} \\ &= \frac{1}{33}\begin{pmatrix} 126 & -24 \\ -24 & 14 \end{pmatrix}.\end{aligned}$$

Fisher's Lemma

Theorem 29

Let $X_1, \ldots, X_n$ be i.i.d. $N(\mu, \sigma^2)$ variables. Let

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad \hat{\sigma}_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

Then,

(i) $\bar{X}$ and $\hat{\sigma}_n^2$ are independent;

(ii) $\bar{X} \sim N(\mu, \sigma^2/n)$;

(iii) $(n-1)\hat{\sigma}_n^2/\sigma^2 \sim \chi_{n-1}^2$.

Further reading

[1] Sheldon M. Ross (谢尔登·M. 罗斯).

A first course in probability (概率论基础教程): Chapter 6.

10th edition (原书第十版), 机械工业出版社

[2] 李贤平.

概率论基础: 第四章第六节.

第三版, 高等教育出版社