Multivariate Statistical Analysis

Lecture 02

Fudan University

luoluo@fudan.edu.cn

Outline

Outline

Random Vectors and Matrices

1. A random matrix (vector) is a matrix (vector) whose elements are random variables.
2. The expected value of a random matrix (or vector) is the matrix (vector) consisting of the expected values of each of its elements.
3. Let $\mathbf{X}$ be an $m \times n$ random matrix, then its expected value, denoted by $\mathbb{E}[\mathbf{X}]$, is the $m \times n$ matrix of numbers (if they exist)

$$\mathbb{E}[\mathbf{X}] = \begin{bmatrix} \mathbb{E}[x_{11}] & \mathbb{E}[x_{12}] & \dots & \mathbb{E}[x_{1n}] \\ \mathbb{E}[x_{21}] & \mathbb{E}[x_{22}] & \dots & \mathbb{E}[x_{2n}] \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{E}[x_{m1}] & \mathbb{E}[x_{m2}] & \dots & \mathbb{E}[x_{mn}] \end{bmatrix}.$$

Expectation of Random Matrices

Let $\mathbf{X}$ and $\mathbf{Y}$ be random matrices of the same dimension, and let $\mathbf{A}$ and $\mathbf{B}$ be conformable matrices of constants. Then we have

$$\mathbb{E}[\mathbf{X}+\mathbf{Y}] = \mathbb{E}[\mathbf{X}] + \mathbb{E}[\mathbf{Y}]$$

and

$$\mathbb{E}[\mathbf{AXB}] = \mathbf{A}\mathbb{E}[\mathbf{X}]\mathbf{B}.$$

Random Vector and Covariance Matrix

For random vector $\mathbf{x} = \left[x_1, \ldots, x_p\right]^\top$, we denote $\boldsymbol{\mu} = \mathbb{E}[\mathbf{x}]$.

The expected value of the random matrix $(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^\top$ is

$$\mathrm{Cov}[\mathbf{x}] = \mathbb{E}\left[(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^\top\right],$$

the covariance or covariance matrix of $\mathbf{x}$.

1. The i-th diagonal element of this matrix, $\mathbb{E}\left[(x_i - \mu_i)^2\right]$, is the variance of x_i.
2. The i,j-th off-diagonal element $(i \neq j)$, $\mathbb{E}[(x_i - \mu_i)(x_j - \mu_j)]$ is the covariance of x_i and x_j.
3. We have $\mathrm{Cov}[\mathbf{x}] = \mathbb{E}\left[\mathbf{x}\mathbf{x}^\top\right] - \boldsymbol{\mu}\boldsymbol{\mu}^\top$.

Outline

Random Vector and Covariance Matrix

Theorem

Let $\mathbf{y} = \mathbf{D}\mathbf{x} + \mathbf{f}$, *where*

1. $\mathbf{D}$ *is an* $n \times p$ *constant matrix,*
2. $\mathbf{x}$ *is a* p*-dimensional random vector,*
3. *and* $\mathbf{f}$ *is a* n*-dimensional constant vector.*

Then we have

$$\mathbb{E}[\mathbf{y}] = \mathbf{D}\mathbb{E}[\mathbf{x}] + \mathbf{f} \qquad \textit{and} \qquad \mathrm{Cov}[\mathbf{y}] = \mathrm{Cov}[\mathbf{D}\mathbf{y}] = \mathbf{D}\mathrm{Cov}[\mathbf{x}]\mathbf{D}^\top.$$

Example

Let $\mathbf{x} = [x_1, x_2]^\top$ be a random vector with

$$\mathbb{E}[\mathbf{x}] = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \qquad \text{and} \qquad \mathrm{Cov}[\mathbf{x}] = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}.$$

Let $\mathbf{z} = [z_1, z_2]$ such that $z_1 = x_1 - x_2$ and $z_2 = x_1 + x_2$.

1. Find the $\mathbb{E}[\mathbf{z}]$ and $\mathrm{Cov}[\mathbf{z}]$.
2. Find the condition that leads to z_1 and z_2 be uncorrelated.

Correlation

For random vector $\mathbf{x} = [x_1, \ldots, x_p]^\top$, we write its covariance as

$$\mathrm{Cov}[\mathbf{x}] = \mathbf{\Sigma} = \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1p} \\ \vdots & \ddots & \vdots \\ \sigma_{p1} & \cdots & \sigma_{pp} \end{bmatrix}.$$

The correlation coefficient ρ_{ij} is defined as

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

which measures linear association between x_i and x_j.

Correlation

The population correlation matrix of **x** is defined as

$$\boldsymbol{\rho} = \begin{bmatrix} \frac{\sigma_{11}}{\sqrt{\sigma_{11}\sigma_{11}}} & \dots & \frac{\sigma_{1p}}{\sqrt{\sigma_{11}\sigma_{pp}}} \\ \vdots & \ddots & \vdots \\ \frac{\sigma_{p1}}{\sqrt{\sigma_{pp}\sigma_{11}}} & \dots & \frac{\sigma_{pp}}{\sqrt{\sigma_{pp}\sigma_{pp}}} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & \dots & \rho_{1p} \\ \vdots & \ddots & \vdots \\ \rho_{p1} & \dots & 1 \end{bmatrix}.$$

Transformation of Variables

Let the density of p-dimensional random vector $\mathbf{x} = [x_1, \dots, x_p]^\top$ be $f(\mathbf{x})$.

Consider the p-dimensional random vector $\mathbf{y} = [y_1, \dots, y_p]^\top$ such that $y_i = u_i(\mathbf{x})$ for $i = 1, \dots, p$. Let the density function of $\mathbf{y}$ be $g(\mathbf{y})$.

Assume the transformation $\mathbf{u}(\mathbf{x}) = [u_1(\mathbf{x}), \dots, u_p(\mathbf{x})]^\top : \mathbb{R}^p \to \mathbb{R}^p$ from the space of $\mathbf{x}$ to the space of $\mathbf{y}$ is smooth and one-to-one.

Then we have $f(\mathbf{x}) = g(\mathbf{u}(\mathbf{x}))|\det(\mathbf{J}(\mathbf{x}))|$ where

$$\mathbf{J}(\mathbf{x}) = \begin{bmatrix} \dfrac{\partial u_1(\mathbf{x})}{\partial x_1} & \dfrac{\partial u_1(\mathbf{x})}{x_2} & \cdots & \dfrac{\partial u_1(\mathbf{x})}{\partial x_p} \\ \dfrac{\partial u_2(\mathbf{x})}{\partial x_1} & \dfrac{\partial u_2(\mathbf{x})}{\partial x_2} & \cdots & \dfrac{\partial u_2(\mathbf{x})}{\partial x_p} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial u_p(\mathbf{x})}{\partial x_1} & \dfrac{\partial u_p(\mathbf{x})}{\partial x_2} & \cdots & \dfrac{\partial u_p(\mathbf{x})}{\partial x_p} \end{bmatrix}.$$

Transformation of Variables

Similarly, we also have $g(\mathbf{y}) = f(\mathbf{u}^{-1}(\mathbf{y}))|\det(\mathbf{J}^{-1}(\mathbf{y}))|$ where

$$\mathbf{J}^{-1}(\mathbf{y}) = \begin{bmatrix} \dfrac{\partial u_1^{-1}(\mathbf{y})}{\partial y_1} & \dfrac{\partial u_1^{-1}(\mathbf{y})}{\partial y_2} & \cdots & \dfrac{\partial u_1^{-1}(\mathbf{y})}{\partial y_p} \\ \dfrac{\partial u_2^{-1}(\mathbf{y})}{\partial y_1} & \dfrac{\partial u_2^{-1}(\mathbf{y})}{\partial y_2} & \cdots & \dfrac{\partial u_2^{-1}(\mathbf{y})}{\partial y_p} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial u_p^{-1}(\mathbf{y})}{\partial y_1} & \dfrac{\partial u_p^{-1}(\mathbf{y})}{\partial y_2} & \cdots & \dfrac{\partial u_p^{-1}(\mathbf{y})}{\partial y_p} \end{bmatrix}.$$

Random Samples

We use the notation $x_{\alpha j}$ to indicate the value of the α-th variable that is observed on the j-th item, or trial.

We display the N measurements on p variables as the $N \times p$ matrix

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1j} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2j} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{\alpha 1} & x_{\alpha 2} & \dots & x_{\alpha j} & \dots & x_{\alpha p} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \dots & x_{Nj} & \dots & x_{Np} \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1^\top \\ \mathbf{x}_2^\top \\ \vdots \\ \mathbf{x}_i^\top \\ \vdots \\ \mathbf{x}_N^\top \end{bmatrix}.$$

We mainly focus on the following case.

1. The independence of measurements from trial to trial may not hold when the variables are likely to drift over time.

Sample Mean and Covariance

Let $\mathbf{x}_1, \ldots, \mathbf{x}_N$ be a random sample from a joint distribution that has mean vector $\boldsymbol{\mu}$, and covariance matrix $\boldsymbol{\Sigma}$. Then the sample means

$$\hat{\boldsymbol{\mu}} = \bar{\mathbf{x}} = \frac{1}{N}\sum_{\alpha=1}^{N}\mathbf{x}_\alpha$$

is an unbiased estimator of $\boldsymbol{\mu}$, and its covariance matrix is

$$\mathrm{Cov}[\bar{\mathbf{x}}] = \frac{1}{N}\boldsymbol{\Sigma}.$$

However, the matrix

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{N}\sum_{\alpha=1}^{N}(\mathbf{x}_\alpha - \bar{\mathbf{x}})(\mathbf{x}_\alpha - \bar{\mathbf{x}})^\top$$

is a biased estimator of $\boldsymbol{\Sigma}$.

Sample Covariance

We define the sample (variance-covariance) covariance matrix as

$$\mathbf{S} = \frac{N}{N-1}\hat{\mathbf{\Sigma}} = \frac{1}{N-1}\sum_{\alpha=1}^{N}(\mathbf{x}_\alpha - \bar{\mathbf{x}})(\mathbf{x}_\alpha - \bar{\mathbf{x}})^\top, \tag{1}$$

which is an unbiased estimator of $\mathbf{\Sigma}$.

Let $\mathbf{1}_N = [1, \dots, 1]^\top \in \mathbb{R}^N$, then we have

$$\begin{aligned}\mathbf{S} =& \frac{1}{N-1}\mathbf{X}^\top\left(\mathbf{I}_N - \frac{1}{N}\mathbf{1}_N\mathbf{1}_N^\top\right)\mathbf{X} & (2)\\ =& \frac{1}{N-1}\left(\mathbf{X}^\top\mathbf{X} - \frac{1}{N}\mathbf{X}^\top\mathbf{1}_N\mathbf{1}_N^\top\mathbf{X}\right). & (3)\end{aligned}$$

It provides a more efficient implementation.

Sample Correlation

Given sample covariance matrix

$$\mathbf{S} = \begin{bmatrix} s_{11} & \dots & s_{1p} \\ \vdots & \ddots & \vdots \\ s_{p1} & \dots & s_{pp} \end{bmatrix} \in \mathbb{R}^{p \times p},$$

we define the sample correlation matrix as

$$\mathbf{R} = \begin{bmatrix} r_{11} & \dots & r_{1p} \\ \vdots & \ddots & \vdots \\ r_{p1} & \dots & r_{pp} \end{bmatrix} \in \mathbb{R}^{p \times p},$$

where $r_{ij} = \dfrac{s_{ij}}{\sqrt{s_{ii}}\sqrt{s_{jj}}}$.

Geometrical Interpretation

We display p-dimensional random vectors $\mathbf{x}_1, \ldots, \mathbf{x}_N$ as follows

$$\mathbf{X} = \begin{bmatrix} x_{11} & \ldots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{N1} & \ldots & x_{Np} \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1^\top \\ \vdots \\ \mathbf{x}_N^\top \end{bmatrix} = \begin{bmatrix} \mathbf{y}_1 & \ldots & \mathbf{y}_p \end{bmatrix} \in \mathbb{R}^{N \times p}.$$

We denote $\bar{\mathbf{x}} = \begin{bmatrix} \bar{x}_1 & \ldots & \bar{x}_p \end{bmatrix}^\top$ and $\mathbf{d}_i = \mathbf{y}_i - \bar{x}_i \mathbf{1}_N$.

1. The projection of $\mathbf{y}_i$ onto the equal angular vector $\mathbf{1}_N$ is the vector $\bar{x}_i \mathbf{1}_N$.
2. The information comprising $\mathbf{S}$ is obtained from the deviation vectors $\{\mathbf{d}_i\}$.
3. The sample correlation r_{ij} is the cosine of the angle between $\mathbf{d}_i$ and $\mathbf{d}_j$.

Outline

Sample Covariance

When all variables are observed, the variation is described by the sample covariance matrix

$$\mathbf{S} = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{21} & s_{22} & \cdots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \cdots & s_{pp} \end{bmatrix} \in \mathbb{R}^{p\times p},$$

where $s_{ij} = \dfrac{1}{N-1}\sum_{\alpha=1}^{N}(x_{\alpha i} - \bar{x}_i)(x_{\alpha j} - \bar{x}_j)$.

The sample covariance matrix contains p variances and $p(p-1)/2$ potentially different covariances.

Generalized Sample Variance

The value of $\det(\mathbf{S})$ reduces to usual sample variance when $p = 1$.

This determinant is called the generalized sample variance:

$$\text{generalized sample variance} = \det(\mathbf{S}).$$

Geometrical Interpretation: Parallelotope

Theorem

Define $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_p] \in \mathbb{R}^{N\times p}$ *and let*

$$\mathrm{Vol}(\mathbf{v}_1, \ldots, \mathbf{v}_\mathrm{p})$$

be the p-dimensional volume of the parallelotope with $\mathbf{v}_1, \ldots, \mathbf{v}_p \in \mathbb{R}^N$ *as principal edges* ($N \geq p$)*, then*

$$\big(\mathrm{Vol}(\mathbf{v}_1, \ldots, \mathbf{v}_\mathrm{p})\big)^2 = \det(\mathbf{V}^\top \mathbf{V}).$$

For $\mathbf{d}_i = \mathbf{y}_i - \bar{x}_i \mathbf{1}_N$, we have

$$\det(\mathbf{S}) = (N-1)^{-p}\big(\mathrm{Vol}(\mathbf{d}_1, \ldots, \mathbf{d}_\mathrm{p})\big)^2.$$

Geometrical Interpretation: Parallelotope

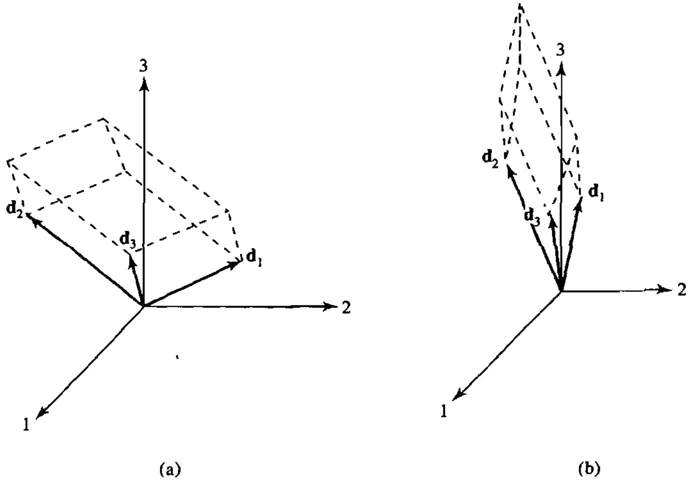


Figure 3.6 (a) "Large" generalized sample variance for $p = 3$.
(b) "Small" generalized sample variance for $p = 3$.

Geometrical Interpretation: Hyperellipsoid

The coordinates

$$\mathbf{x} = [x_1, x_2, \ldots, x_p]^\top$$

of the points a constant distance $c > 0$ from $\bar{\mathbf{x}}$ satisfy (suppose $\mathbf{S} \succ \mathbf{0}$)

$$(\mathbf{x} - \bar{\mathbf{x}})^\top \mathbf{S}^{-1} (\mathbf{x} - \bar{\mathbf{x}}) = c^2,$$

which defines hyperellipsoid centered at $\bar{\mathbf{x}}$.

The volume of this hyperellipsoid is

$$\frac{2\pi^{p/2}}{p\Gamma(p/2)} \cdot c^p (\det(\mathbf{S}))^{1/2},$$

where

$$\Gamma(p) = \int_0^\infty t^{p-1} \exp(-t) \, \mathrm{d}t.$$

Generalized Sample Variance is Zero

The generalized variance is zero when, and only when, at least one of

$$\{\mathbf{d}_1, \dots, \mathbf{d}_p\}$$

lies in the hyperplane formed by all linear combinations of the others.

That is, the columns of the matrix of deviations

$$\begin{aligned}\mathbf{X} - \mathbf{1}_N\bar{\mathbf{x}}^\top = \begin{bmatrix} (\mathbf{x}_1 - \bar{\mathbf{x}})^\top \\ \vdots \\ (\mathbf{x}_N - \bar{\mathbf{x}})^\top \end{bmatrix} &= \begin{bmatrix} \mathbf{y}_1 - \bar{x}_1\mathbf{1}_N & \dots & \mathbf{y}_p - \bar{x}_p\mathbf{1}_N \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{d}_1 & \dots & \mathbf{d}_p \end{bmatrix} \in \mathbb{R}^{N\times p}\end{aligned}$$

are linearly dependent.

Generalized Sample Variance Determined by Correlation

We can also define generalized variance by

$$\det(\mathbf{R}),$$

where $\mathbf{R}$ is the sample correlation matrix

$$\mathbf{R} = \begin{bmatrix} r_{11} & \cdots & r_{1p} \\ \vdots & \ddots & \vdots \\ r_{p1} & \cdots & r_{pp} \end{bmatrix} \in \mathbb{R}^{p \times p},$$

where $r_{ij} = \dfrac{s_{ij}}{\sqrt{s_{ii}}\sqrt{s_{jj}}}$.

It holds that

$$\det(\mathbf{S}) = \det(\mathbf{R}) \prod_{i=1}^{p} s_{ii}.$$

Total Sample Variance

We define the total sample variance as the sum of the diagonal elements of the sample covariance matrix, that is

$$\text{total sample variance} = \sum_{i=1}^{p} s_{ii}.$$

1. It is the sum of the squared lengths of the p deviation vectors

$$\mathbf{d}_1 = \mathbf{y}_1 - \bar{x}_1 \mathbf{1}_N, \ldots, \mathbf{d}_p = \mathbf{y}_1 - \bar{x}_p \mathbf{1}_N$$

   divided by $n - 1$.
2. It pays no attention to the orientation of $\mathbf{d}_i$.

Outline

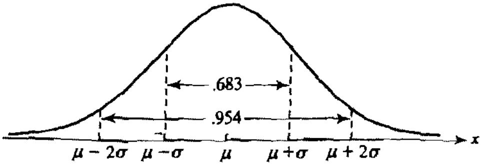
Univariate Normal Distribution

A random variable x is normally distributed with mean μ and standard deviation $\sigma > 0$ can be written in the following notation

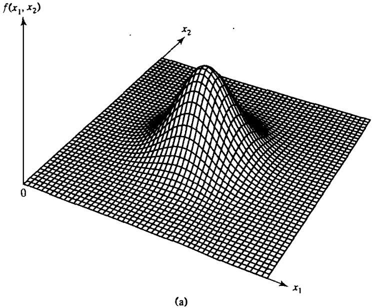
$$x \sim \mathcal{N}(\mu, \sigma).$$

The probability density function of univariate normal distribution is

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



Bivariate Normal Density



Two bivariate normal distributions:

- (a) $\sigma_1 = \sigma_2$ and $\rho_{12} = 0$
- (b) $\sigma_1 = \sigma_2$ and $\rho_{12} = 0.75$

The Central Limit Theorem

Let $x_1, \ldots, x_n$ be independent and identically distributed random variables with the same arbitrary distribution, mean μ, and variance σ^2.

Let $\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$, then the random variable

$$z = \lim_{n\to\infty} \sqrt{n}\left(\frac{\bar{x}_n - \mu}{\sigma}\right)$$

is a standard normal distribution.

The standard normal distribution is a normal distribution with a mean of 0 and standard deviation of 1.

What about multivariate case?

Multivariate Normal Distribution

The multivariate normal distribution of a p-dimensional random vector $\mathbf{x} = [x_1, \ldots, x_p]^\top$ can be written in the following notation:

$$\mathbf{x} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

or to make it explicitly known that $\mathbf{x}$ is p-dimensional.

$$\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

with p-dimensional mean vector

$$\boldsymbol{\mu} = \mathbb{E}[\mathbf{x}] = \begin{bmatrix} \mathbb{E}[x_1] \\ \vdots \\ \mathbb{E}[x_p] \end{bmatrix} \in \mathbb{R}^p$$

and covariance matrix

$$\boldsymbol{\Sigma} = \mathbb{E}\left[(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^\top\right] \in \mathbb{R}^{p \times p}.$$

Multivariate Normal Distribution

The density function of univariate normal distribution is

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

where μ is the mean and σ^2 is the variance with $\sigma > 0$.

The density function of non-singular p-dimensional multivariate normal distribution is

$$f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^p \det(\boldsymbol{\Sigma})}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right),$$

where $\boldsymbol{\mu} \in \mathbb{R}^p$ is the mean and $\boldsymbol{\Sigma}$ is the $p \times p$ (non-singular) covariance matrix.

Density Function of Multivariate Normal Distribution

We generalize the form of pdf for univariate normal distribution

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

to the multivariate case

$$f(\mathbf{x}) = K \exp\left(-\frac{1}{2}(\mathbf{x}-\mathbf{b})^\top \mathbf{A}(\mathbf{x}-\mathbf{b})\right),$$

where $\mathbf{A}$ is symmetric positive definite.

We can verify that if $\mathbb{E}[\mathbf{x}] = \boldsymbol{\mu}$ and $\mathrm{Cov}[\mathbf{x}] = \boldsymbol{\Sigma}$, then

$$K = \frac{1}{\sqrt{(2\pi)^p \det(\boldsymbol{\Sigma})}}, \quad \mathbf{b} = \boldsymbol{\mu} \quad \text{and} \quad \mathbf{A} = \boldsymbol{\Sigma}^{-1}.$$

Multivariate Normal Distribution

If the density of a p-dimensional random vector $\mathbf{x}$ is

$$K \exp\left(-\frac{1}{2}(\mathbf{x}-\mathbf{b})^\top \mathbf{A}(\mathbf{x}-\mathbf{b})\right),$$

where $\mathbf{A} \in \mathbb{R}^{p\times p}$ is symmetric positive definite, then $\mathbb{E}[\mathbf{x}] = \mathbf{b}$ and $\mathrm{Cov}[\mathbf{x}] = \mathbf{A}^{-1}$.

Conversely, given a vector $\boldsymbol{\mu} \in \mathbb{R}^p$ and a positive definite matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{p\times p}$, there is a multivariate normal density

$$n(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^p \det(\boldsymbol{\Sigma})}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right).$$