STT 843: Multivariate Analysis

5. The Multivariate Normal Distribution (Chapter 4.2-4.3)

Guanqun Cao

Department of Statistics and Probability
Michigan State University

Spring 2026

Outline

1. Additional Properties of the Multivariate Normal Distribution
2. Sampling from a Multivariate Normal Distribution and Maximum Likelihood Estimation

Additional Properties of the MND

The following are true for a normal vector $\boldsymbol{X}$ having a multivariate normal distribution:

1. Linear combination of the components of $\boldsymbol{X}$ are normally distributed.
2. All subsets of the components of $\boldsymbol{X}$ have a (multivariate) normal distribution.
3. Zero covariance implies that the corresponding components are independently distributed.
4. The conditional distributions of the components are normal.

- If $\boldsymbol{X}$ is distributed as $N_p(\boldsymbol{\mu}, \Sigma)$, then any linear combination of variables $\mathbf{a}^\top \boldsymbol{X} = a_1 X_1 + a_2 X_2 + \cdots + a_p X_p$ is distributed as $N\left(\mathbf{a}^\top \boldsymbol{\mu}, \mathbf{a}^\top \Sigma \mathbf{a}\right)$.
- Also if $\mathbf{a}^\top \boldsymbol{X}$ is distributed as $N\left(\mathbf{a}^\top \boldsymbol{\mu}, \mathbf{a}^\top \Sigma \mathbf{a}\right)$ for every $\mathbf{a}$, then $\boldsymbol{X}$ must be $N_p(\boldsymbol{\mu}, \Sigma)$.

Example 3.3 (The distribution of a linear combination of the component of a normal random vector)
Consider the linear combination $\mathbf{a}^\top \boldsymbol{X}$ of a multivariate normal random vector determined by the choice $\mathbf{a}^\top = [1, 0, \ldots, 0]$.
If $\boldsymbol{X}$ is distributed as $N_p(\boldsymbol{\mu}, \Sigma)$, the q linear combinations

$$\mathbf{A}_{(q\times p)}\mathbf{X}_{p\times 1} = \begin{bmatrix} a_{11}X_1 + \cdots + a_{1p}X_p \\ a_{21}X_1 + \cdots + a_{2p}X_p \\ \vdots \\ a_{q1}X_1 + \cdots + a_{qp}X_p \end{bmatrix}$$

are distributed as $N_q\left(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\Sigma\mathbf{A}^\top\right)$. Also $\boldsymbol{X}_{p\times 1} + \mathbf{d}_{p\times 1}$, where $\mathbf{d}$ is a vector of constants, is distributed as $N_p(\boldsymbol{\mu} + \mathbf{d}, \Sigma)$.

Example 3.4 (The distribution of two linear combinations of the components of a normal random vector) For $\boldsymbol{X}$ distributed as $N_3(\boldsymbol{\mu}, \Sigma)$, find the distribution of

$$\begin{bmatrix} X_1 - X_2 \\ X_2 - X_3 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = \mathbf{A}\boldsymbol{X}$$

All subsets of $\boldsymbol{X}$ are normally distributed. If we respectively partition $\boldsymbol{X}$, its mean vector $\boldsymbol{\mu}$, and its covariance matrix Σ as

$$\boldsymbol{X}_{(p\times 1)} = \begin{bmatrix} X_1 \\ (q \times 1) \\ \cdots\cdots \\ X_2 \\ (p-q)\times 1 \end{bmatrix} \qquad \boldsymbol{\mu}_{(p\times 1)} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ (q \times 1) \\ \cdots\cdots \\ \boldsymbol{\mu}_2 \\ (p-q)\times 1 \end{bmatrix}$$

and

$$\Sigma_{(p\times p)} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ (q \times 1) & (q \times (p-q)) \\ \cdots\cdots\cdots & \cdots\cdots \\ \Sigma_{21} & \Sigma_{22} \\ ((p-q)\times q) & ((p-q)\times(p-q)) \end{bmatrix}$$

then $\boldsymbol{X}_1$ is distributed as $N_q(\boldsymbol{\mu}_1, \Sigma_{11})$.

Example 3.5 (The distribution of a subset of a normal random vector)
If $\boldsymbol{X}$ is distributed as $N_5(\boldsymbol{\mu}, \Sigma)$, find the distribution of $[X_2, X_4]^{\mathsf{T}}$.

- If $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are independent, then $\mathrm{Cov}\,(\boldsymbol{X}_1, \boldsymbol{X}_2) = 0$, a $q_1 \times q_2$ matrix of zeros, where $\boldsymbol{X}_1$ is $q_1 \times 1$ random vector and $\boldsymbol{X}_2$ is $q_2 \times 1$ random vector
- If $\begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{bmatrix}$ is $N_{q_1+q_2}\left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}\right)$, then $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are independent if and only if $\Sigma_{12} = \Sigma_{21} = 0$.

- If $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are independent and are distributed as $N_{q_1}(\boldsymbol{\mu}_1, \Sigma_{11})$ and $N_{q_2}(\boldsymbol{\mu}_2, \Sigma_{22})$, respectively, then $\begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{bmatrix}$ has the multivariate normal distribution

$$N_{q_1+q_2}\left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{bmatrix}\right)$$

Example 3.6 (The equivalence of zero covariance and independence for normal variables)

Let $\boldsymbol{X}_{3\times 1}$ be $N_3(\boldsymbol{\mu}, \Sigma)$ with

$$\Sigma = \begin{bmatrix} 4 & 1 & 0 \\ 1 & 3 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

Are X_1 and X_2 independent? What about (X_1, X_2) and X_3 ?

- Let $\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{bmatrix}$ be distributed as $N_p(\boldsymbol{\mu}, \Sigma)$ with $\begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$, and $|\Sigma_{22}| > 0$.
- Then the conditional distribution of $\boldsymbol{X}_1$, given that $\boldsymbol{X}_2 = \mathbf{x}_2$ is normal and has

$$\text{Mean } = \boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2)$$

and

$$\text{Covariance } = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

- Note that the covariance does not depend on the value $\mathbf{x}_2$ of the conditioning variable.

Example 3.7 (The conditional density of a bivariate normal distribution) Obtain the conditional density of X_1, give that $X_2 = x_2$ for any bivariate distribution.

Let $\boldsymbol{X}$ be distributed as $N_p(\boldsymbol{\mu}, \Sigma)$ with $|\Sigma| > 0$. Then

- $(\boldsymbol{X} - \boldsymbol{\mu})^\mathsf{T}\Sigma^{-1}(\boldsymbol{X} - \boldsymbol{\mu})$ is distributed as χ^2_p, where χ^2_p denotes the chi-square distribution with p degrees of freedom.
- The $N_p(\boldsymbol{\mu}, \Sigma)$ distribution assign probability $1 - \alpha$ to the solid ellipsoid $\left\{\mathbf{x} : (\mathbf{x} - \boldsymbol{\mu})^\mathsf{T}\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}) \leq \chi^2_p(\alpha)\right\}$, where $\chi^2_p(\alpha)$ denote the upper (100α) th percentile of the χ^2_p distribution.

Some facts

Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ be mutually independent with $\boldsymbol{X}_j$ distributed as $N_p\left(\boldsymbol{\mu}_j, \Sigma\right)$. (Note that each $\boldsymbol{X}_j$ has the same covariance matrix Σ.) Then

$$\mathbf{V}_1 = c_1\boldsymbol{X}_1 + c_2\boldsymbol{X}_2 + \cdots + c_n\boldsymbol{X}_n$$

is distributed as $N_p\left(\sum_{j=1}^n c_j\boldsymbol{\mu}_j, \left(\sum_{j=1}^n c_j^2\right)\Sigma\right)$. Moreover, $\mathbf{V}_1$ and $\mathbf{V}_2 = b_1\boldsymbol{X}_1 + b_2\boldsymbol{X}_2 + \cdots + b_n\boldsymbol{X}_n$ are jointly multivariate normal with covariance matrix

$$\begin{bmatrix} \left(\sum_{j=1}^n c_j^2\right)\Sigma & \mathbf{b}^\top\mathbf{c}\Sigma \\ \mathbf{b}^\top\mathbf{c}\Sigma_{21} & \left(\sum_{j=1}^n b_j^2\right)\Sigma \end{bmatrix}$$

Consequently, V_1 and V_2 are independent if $\mathbf{b}^\top\mathbf{c} = \sum_{j=1}^n c_j b_j = 0$.

Example 3.8 (Linear combinations of random vectors)

Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3$ and $\boldsymbol{X}_4$ be independent and identically distributed 3×1 random vectors with

$$\boldsymbol{\mu} = \begin{bmatrix} 3 \\ -1 \\ 1 \end{bmatrix} \quad \text{and} \quad \Sigma = \begin{bmatrix} 3 & -1 & 1 \\ -1 & 1 & 0 \\ 1 & 0 & 2 \end{bmatrix}$$

(a) find the mean and variance of the linear combination $\mathbf{a}^\top \boldsymbol{X}_1$ of the three components of $\boldsymbol{X}_1$ where $a = \begin{bmatrix} a_1 & a_2 & a_3 \end{bmatrix}^\top$.
(b) Consider two linear combinations of random vectors

$$\frac{1}{2}\boldsymbol{X}_1 + \frac{1}{2}\boldsymbol{X}_2 + \frac{1}{2}\boldsymbol{X}_3 + \frac{1}{2}\boldsymbol{X}_4$$

and

$$\boldsymbol{X}_1 + \boldsymbol{X}_2 + \boldsymbol{X}_3 - 3\boldsymbol{X}_4$$

Find the mean vector and covariance matrix for each linear combination of vectors and also the covariance between them.

The Multivariate Normal Likelihood

Joint density function of all $p \times 1$ observed random vectors $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$

$$
\begin{aligned}
& \text{Joint density of } \{\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n\} \\
= & \prod_{j=1}^{n} \left\{ \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} e^{-\left(\mathbf{x}_j - \boldsymbol{\mu}\right)^{\mathsf{T}} \Sigma^{-1} \left(\mathbf{x}_j - \boldsymbol{\mu}\right)/2} \right\} \\
= & \frac{1}{(2\pi)^{np/2} |\Sigma|^{n/2}} e^{-\sum_{j=1}^{n} \left(\mathbf{x}_j - \boldsymbol{\mu}\right)^{\mathsf{T}} \Sigma^{-1} \left(\mathbf{x}_j - \boldsymbol{\mu}\right)/2} \\
= & \frac{1}{(2\pi)^{np/2} |\Sigma|^{n/2}} e^{-\operatorname{tr}\left[\Sigma^{-1} \left(\sum_{j=1}^{n} \left(\mathbf{x}_j - \bar{\mathbf{x}}\right)\left(\mathbf{x}_j - \bar{\mathbf{x}}\right)^{\mathsf{T}} + n(\bar{\mathbf{x}} - \boldsymbol{\mu})(\bar{\mathbf{x}} - \boldsymbol{\mu})^{\mathsf{T}}\right)\right]/2}
\end{aligned}
$$

Likelihood

When the numerical values of the observations become available, they may be substituted for the $\mathbf{x}_j$ in the equation above. The resulting expression, now considered as a function of $\boldsymbol{\mu}$ and Σ for the fixed set of observations $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$, is called the likelihood.

Maximum likelihood estimation

One meaning of best is to select the parameter values that maximize the joint density evaluated at the observations. This technique is called maximum likelihood estimation, and the maximizing parameter values are called maximum likelihood estimates.

Let $\mathbf{A}$ be a $k \times k$ symmetric matrix and $\mathbf{x}$ be a $k \times 1$ vector. Then

- $\mathbf{x}^{\mathsf{T}}\mathbf{A}\mathbf{x} = \mathrm{tr}\left(\mathbf{x}^{\mathsf{T}}\mathbf{A}\mathbf{x}\right) = \mathrm{tr}\left(\mathbf{A}\mathbf{x}\mathbf{x}^{\mathsf{T}}\right)$
- $\mathrm{tr}(\mathbf{A}) = \sum_{i=1}^{n} \lambda_i$, where the λ_i are the eigenvalues of $\mathbf{A}$.

Maximum Likelihood Estimate of μ and Σ

- Given a $p \times p$ symmetric positive definite matrix $\mathbf{B}$ and a scalar $b > 0$, it follows that

$$\frac{1}{|\Sigma|^b} e^{-\operatorname{tr}\left(\Sigma^{-1}\mathbf{B}\right)/2} \leq \frac{1}{|B|^b} (2b)^{pb} e^{-bp}$$

for all positive definite $\Sigma_{p\times p}$, with equality holding only for $\Sigma = (1/2b)\mathbf{B}$.

- Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ be a random sample from a normal population with mean $\boldsymbol{\mu}$ and covariance Σ. Then

$$\hat{\boldsymbol{\mu}} = \overline{\boldsymbol{X}} \quad \text{and} \quad \hat{\Sigma} = \frac{1}{n}\sum_{j=1}^{n}\left(\boldsymbol{X}_j - \bar{X}\right)\left(\boldsymbol{X}_j - \bar{X}\right)^{\mathsf{T}} = \frac{n-1}{n}\mathbf{S}$$

  are the maximum likelihood estimators of μ and Σ, respectively.
- Their observed value $\bar{\mathbf{x}}$ and $(1/n)\sum_{j=1}^{n}\left(\mathbf{x}_j - \bar{\mathbf{x}}\right)\left(\mathbf{x}_j - \bar{\mathbf{x}}\right)^{\mathsf{T}}$, are called the maximum likelihood estimates of μ and Σ.

Invariance Property of Maximum likelihood estimators

Let $\hat{\boldsymbol{\theta}}$ be the maximum likelihood estimator of $\boldsymbol{\theta}$, and consider the parameter $h(\boldsymbol{\theta})$, which is a function of $\boldsymbol{\theta}$. Then the maximum likelihood estimate of $h(\boldsymbol{\theta})$ is given by $h(\hat{\boldsymbol{\theta}})$.

Examples

1. The maximum likelihood estimator of $\boldsymbol{\mu}^{\mathsf{T}}\Sigma^{-1}\boldsymbol{\mu}$ is $\hat{\boldsymbol{\mu}}\hat{\Sigma}^{-1}\hat{\boldsymbol{\mu}}$, where $\hat{\boldsymbol{\mu}} = \overline{\boldsymbol{X}}$ and $\hat{\Sigma} = \frac{n-1}{n}\mathbf{S}$ are the maximum likelihood estimators of $\boldsymbol{\mu}$ and Σ respectively.
2. The maximum likelihood estimator of $\sqrt{\sigma_{ii}}$ is $\sqrt{\hat{\sigma}_{ii}}$, where

$$\hat{\sigma}_{ii} = \frac{1}{n}\sum_{j=1}^{n}\left(X_{ij} - \bar{X}_i\right)^2$$

is the maximum likelihood estimator of $\sigma_{ii} = \mathrm{Var}\left(X_i\right)$.