Multivariate Statistics

Lecture 03

Fudan University

Outline

Outline

Normally Distributed Variables

Some properties of normally distributed variables:

1. The marginal distributions derived from multivariate normal distributions are also normal distributions.
2. The conditional distributions derived from multivariate normal distributions are also normal distributions.
3. The linear combinations of multivariate normal variates are normally distributed.

Multivariate Normal Distribution (Linear Combination)

Theorem 1

Let $\mathbf{x} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with $\boldsymbol{\Sigma} \succ \mathbf{0}$. Then

$$\mathbf{y} = \mathbf{C}\mathbf{x}$$

is distributed according to $\mathcal{N}_p(\mathbf{C}\boldsymbol{\mu}, \mathbf{C}\boldsymbol{\Sigma}\mathbf{C}^\top)$ for non-singular $\mathbf{C} \in \mathbb{R}^{p\times p}$.

Sketch of the proof:

1. Let $f(\mathbf{x})$ be the density function of $\mathbf{x}$.
2. Let $g(\mathbf{y})$ be the density function of $\mathbf{y}$.
3. The relation $\mathbf{x} = \mathbf{C}^{-1}\mathbf{y}$ implies

$$g(\mathbf{y}) = f(\mathbf{u}^{-1}(\mathbf{y}))\,|\det(\mathbf{J}^{-1}(\mathbf{y}))|$$

   with $\mathbf{u}(\mathbf{x}) = \mathbf{C}\mathbf{x}$, $\mathbf{u}^{-1}(\mathbf{y}) = \mathbf{C}^{-1}\mathbf{y}$ and $\mathbf{J}^{-1}(\mathbf{y}) = \mathbf{C}^{-1}$.

Outline

Multivariate Normal Distribution (Independence)

Theorem 2

If $\mathbf{x} = [x_1, \ldots, x_p]^\top \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma} \succ \mathbf{0}$. Let

$$\mathbf{x}^{(1)} = [x_1, \ldots, x_q]^\top \quad \text{and} \quad \mathbf{x}^{(2)} = [x_{q+1}, \ldots, x_p]^\top$$

for $q < p$. A necessary and sufficient condition for $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$ to be independent is that each covariance of a variable from $\mathbf{x}^{(1)}$ and a variable from $\mathbf{x}^{(2)}$ is 0.

1. The random vectors $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$ in can be replaced by any subset of $\mathbf{x}$ the subset consisting of the remaining variables respectively.
2. The necessity does not depend on the assumption of normality.

Outline

Multivariate Normal Distribution (Marginal Distribution)

Corollary 2.1

We use the notation in the proof as follows

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}^{(1)} \\ \mathbf{x}^{(2)} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu}^{(1)} \\ \boldsymbol{\mu}^{(2)} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} \right)$$

It shows that if $\mathbf{x}^{(1)}$ is uncorrelated with $\mathbf{x}^{(2)}$, the marginal distribution of $\mathbf{x}^{(1)}$ is $\mathcal{N}(\boldsymbol{\mu}^{(1)}, \boldsymbol{\Sigma}_{11})$ and the marginal distribution of $\mathbf{x}^{(2)}$ is $\mathcal{N}(\boldsymbol{\mu}^{(2)}, \boldsymbol{\Sigma}_{22})$.

In fact, this result holds even if the two sets are NOT uncorrelated.

Multivariate Normal Distribution (Marginal Distribution)

We make a non-singular linear transformation $\mathbf{B} = -\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}$ to subvectors

$$\mathbf{y}^{(1)} = \mathbf{x}^{(1)} + \mathbf{B}\mathbf{x}^{(2)} \quad \text{and} \quad \mathbf{y}^{(2)} = \mathbf{x}^{(2)}$$

leading to the components of $\mathbf{y}^{(1)}$ are uncorrelated with the ones of $\mathbf{y}^{(2)}$.

The vector

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}^{(1)} \\ \mathbf{y}^{(2)} \end{bmatrix} = \begin{bmatrix} \mathbf{I} & -\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{x}^{(1)} \\ \mathbf{x}^{(2)} \end{bmatrix} = \begin{bmatrix} \mathbf{I} & -\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \mathbf{x}$$

is a non-singular transform of $\mathbf{x}$, and therefore it is normally distributed

$$\mathbf{y} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu}^{(1)} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\mu}^{(2)} \\ \boldsymbol{\mu}^{(2)} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{22} \end{bmatrix} \right)$$

Thus $\mathbf{y}^{(1)}$ and $\mathbf{y}^{(2)}$ are independent, which implies the marginal distribution of $\mathbf{x}^{(2)}$ is $\mathcal{N}(\boldsymbol{\mu}^{(2)}, \boldsymbol{\Sigma}_{22})$.

Multivariate Normal Distribution (Marginal Distribution)

Because the numbering of the components of $\mathbf{x}$ is arbitrary, we can state the following theorem:

Theorem 3

If $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma} \succ \mathbf{0}$, the marginal distribution of any set of components of $\mathbf{x}$ is multivariate normal with means, variances, and covariances obtained by taking the corresponding components of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, respectively.

Outline

Singular Normal Distributions

In previous section, we focus on non-singular normal normally distributed variate $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma} \succ \mathbf{0}$ whose density function is

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

What about the case of singular $\boldsymbol{\Sigma}$?

General Linear Transformation

We can extend Theorem 1 to Theorem 4

Theorem 1

Let $\mathbf{x} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with $\boldsymbol{\Sigma} \succ \mathbf{0}$. Then

$$\mathbf{y} = \mathbf{C}\mathbf{x}$$

is distributed according to $\mathcal{N}_p(\mathbf{C}\boldsymbol{\mu}, \mathbf{C}\boldsymbol{\Sigma}\mathbf{C}^\top)$ for non-singular $\mathbf{C} \in \mathbb{R}^{p\times p}$.

Theorem 4

Let $\mathbf{x} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with $\boldsymbol{\Sigma} \succ \mathbf{0}$. Then

$$\mathbf{z} = \mathbf{D}\mathbf{x}$$

is distributed according to $\mathcal{N}_q(\mathbf{D}\boldsymbol{\mu}, \mathbf{D}\boldsymbol{\Sigma}\mathbf{D}^\top)$ for $\mathbf{D} \in \mathbb{R}^{q\times p}$ of rank $q \leq p$.

General Linear Transformation

For any transformation $\mathbf{x} = \mathbf{D}\mathbf{x}$, for $\mathbf{D} \in \mathbb{R}^{q \times p}$ and p-dimensional random vector $\mathbf{x}$, we have

$$\mathbb{E}[\mathbf{z}] = \mathbf{D}\boldsymbol{\mu} \qquad \text{and} \qquad \mathrm{Cov}[\mathbf{z}] = \mathbf{D}\boldsymbol{\Sigma}\mathbf{D}.$$

If $q \leq p$ and $\mathbf{D}$ is of rank q, we can find a $(p-q) \times p$ matrix $\mathbf{E}$ such that

$$\begin{bmatrix} \mathbf{z} \\ \mathbf{w} \end{bmatrix} = \begin{bmatrix} \mathbf{D} \\ \mathbf{E} \end{bmatrix} \mathbf{x}$$

is a non-singular transformation.

Then $\mathbf{z}$ and $\mathbf{w}$ have a joint normal distribution, and $\mathbf{z}$ has a marginal normal distribution by Theorem 3.

General Linear Transformation

Theorem 4

Let $\mathbf{x} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with $\boldsymbol{\Sigma} \succ \mathbf{0}$. Then

$$\mathbf{z} = \mathbf{D}\mathbf{x}$$

is distributed according to $\mathcal{N}_q(\mathbf{D}\boldsymbol{\mu}, \mathbf{D}\boldsymbol{\Sigma}\mathbf{D}^\top)$ for $\mathbf{D} \in \mathbb{R}^{q \times p}$ of rank $q \leq p$.

Can we extend $\mathbf{D}$ to any real matrix?

General Linear Transformation

In the case of the singular normal, distribution the mass is concentrated on a given linear set. The probability associated with any set not intersecting the given linear set is 0.

For example, consider that

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \right)$$

Such $\mathbf{x}$ cannot have a density at all, because the probability of any set not intersecting the x_2-axis is 0 would imply that the density is 0 almost everywhere.

Singular Normal Distributions

Suppose that $\mathbf{y} \sim \mathcal{N}_q(\boldsymbol{\nu}, \mathbf{T})$, $\mathbf{A} \in \mathbb{R}^{p \times q}$ with $p > q$ and $\boldsymbol{\lambda} \in \mathbb{R}^p$; then we say that

$$\mathbf{x} = \mathbf{A}\mathbf{y} + \boldsymbol{\lambda}$$

has a singular (degenerate) normal distribution in p-space.

We have $\boldsymbol{\mu} = \mathbb{E}[\mathbf{x}] = \mathbf{A}\boldsymbol{\nu} + \boldsymbol{\lambda}$ and

$$\boldsymbol{\mu} = \mathbb{E}[\mathbf{x}] = \mathbf{A}\boldsymbol{\nu} + \boldsymbol{\lambda} \quad \text{and} \quad \boldsymbol{\Sigma} = \mathrm{Cov}(\mathbf{x}) = \mathbf{A}\mathbf{T}\mathbf{A}^\top.$$

The matrix $\boldsymbol{\Sigma}$ is singular and we cannot write the normal density for $\mathbf{x}$.

In fact, $\mathbf{x}$ cannot have a density at all.

Singular Normal Distributions

Now we give a formal definition of a normal distribution that includes the singular distribution.

Definition

A p-dimensional random vector $\mathbf{x}$ with $\mathbb{E}[\mathbf{x}] = \boldsymbol{\mu}$ and $\mathrm{Cov}(\mathbf{x}) = \boldsymbol{\Sigma}$ is said to be normally distributed [or is said to be distributed according to $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$] if there is a transformation

$$\mathbf{x} = \mathbf{A}\mathbf{y} + \boldsymbol{\lambda},$$

where $\mathbf{A} \in \mathbb{R}^{p \times r}$, r is the rank of $\boldsymbol{\Sigma}$ and $\mathbf{y} \sim \mathcal{N}_r(\boldsymbol{\nu}, \mathbf{T})$ with $\mathbf{T} \succ \mathbf{0}$.

If $\boldsymbol{\Sigma}$ has rank p, then we can take $\mathbf{A} = \mathbf{I}$ and $\boldsymbol{\lambda} = \mathbf{0}$.

More General Linear Transformation

Theorem 5

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

$$\mathbf{z} = \mathbf{D}\mathbf{x}$$

is distributed according to $\mathcal{N}_q(\mathbf{D}\boldsymbol{\mu}, \mathbf{D}\boldsymbol{\Sigma}\mathbf{D}^\top)$ for any $\mathbf{D} \in \mathbb{R}^{q \times p}$.

We do not require additional assumptions on $\mathbf{D}$ or $\boldsymbol{\Sigma}$.

Outline

Multivariate Normal Distribution (Conditional Distribution)

Let $\mathbf{x}$ be distributed according to $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma} \succ \mathbf{0}$. Let us partition

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}^{(1)} \\ \mathbf{x}^{(2)} \end{bmatrix} \quad \text{with } \mathbf{x}^{(1)} \in \mathbb{R}^q \text{ and } \mathbf{x}^{(2)} \in \mathbb{R}^{p-q}.$$

The joint density of $\mathbf{y}^{(1)} = \mathbf{x}^{(1)} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\mathbf{x}^{(2)}$ and $\mathbf{y}^{(2)} = \mathbf{x}^{(2)}$ is

$$g(\mathbf{y}) = n(\mathbf{y}^{(1)} \mid \boldsymbol{\mu}^{(1)} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\mu}^{(2)}, \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}) n(\mathbf{y}^{(2)} \mid \boldsymbol{\mu}^{(2)}, \boldsymbol{\Sigma}_{22}).$$

Consider that

$$\begin{bmatrix} \mathbf{y}^{(1)} \\ \mathbf{y}^{(2)} \end{bmatrix} = \begin{bmatrix} \mathbf{I} & -\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{x}^{(1)} \\ \mathbf{x}^{(2)} \end{bmatrix} \quad \text{with} \quad \det\left(\begin{bmatrix} \mathbf{I} & -\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \right) = 1.$$

Then the density of $\mathbf{x}$ (joint density of $\mathbf{x}_1$ and $\mathbf{x}_2$) is

$$f(\mathbf{x}) = g(\mathbf{u}(\mathbf{x}))|\det(\mathbf{J}(\mathbf{x}))| = g(\mathbf{u}(\mathbf{x})).$$

Multivariate Normal Distribution (Conditional Distribution)

The resulting joint density of $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$ is

$$\begin{aligned}
& f(\mathbf{x}) = f(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}) \\
=& n(\mathbf{y}^{(1)} \mid \boldsymbol{\mu}^{(1)} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\mu}^{(2)}, \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}) n(\mathbf{y}^{(2)} \mid \boldsymbol{\mu}^{(2)}, \boldsymbol{\Sigma}_{22}) \\
=& \frac{1}{\sqrt{(2\pi)^q \det(\boldsymbol{\Sigma}_{11.2})}} \exp\left(-\frac{1}{2}\left(\mathbf{x}^{(11.2)} - \boldsymbol{\mu}^{(11.2)}\right)^\top \boldsymbol{\Sigma}_{11.2}^{-1}\left(\mathbf{x}^{(11.2)} - \boldsymbol{\mu}^{(11.2)}\right)\right) \\
& \cdot \frac{1}{\sqrt{(2\pi)^{p-q} \det(\boldsymbol{\Sigma}_{22})}} \exp\left(-\frac{1}{2}\left(\mathbf{x}^{(2)} - \boldsymbol{\mu}^{(2)}\right)^\top \boldsymbol{\Sigma}_{22}^{-1}\left(\mathbf{x}^{(2)} - \boldsymbol{\mu}^{(2)}\right)\right)
\end{aligned}$$

where

$$\begin{aligned}
\mathbf{x}^{(11.2)} =& \mathbf{x}^{(1)} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\mathbf{x}^{(2)}, \\
\boldsymbol{\mu}^{(11.2)} =& \boldsymbol{\mu}^{(1)} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\mu}^{(2)}, \\
\boldsymbol{\Sigma}_{11.2} =& \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}.
\end{aligned}$$

The marginal density of $\mathbf{x}^{(2)}$ is

$$\begin{aligned} f(\mathbf{x}^{(2)}) &= n(\mathbf{y}^{(2)} \mid \boldsymbol{\mu}^{(2)}, \boldsymbol{\Sigma}_{22}) \\ &= \frac{1}{\sqrt{(2\pi)^{p-q}\det(\boldsymbol{\Sigma}_{22})}} \exp\left(-\frac{1}{2}\left(\mathbf{x}^{(2)} - \boldsymbol{\mu}^{(2)}\right)^\top \boldsymbol{\Sigma}_{22}^{-1}\left(\mathbf{x}^{(2)} - \boldsymbol{\mu}^{(2)}\right)\right). \end{aligned}$$

Hence, the conditional density of $\mathbf{x}^{(1)}$ given that $\mathbf{x}^{(2)}$ is

$$\begin{aligned} f(\mathbf{x}^{(1)} \mid \mathbf{x}^{(2)}) &= \frac{f(\mathbf{x}^{(1)}, \mathbf{x}^{(2)})}{f(\mathbf{x}^{(2)})} \\ &= \frac{1}{\sqrt{(2\pi)^{q}\det(\boldsymbol{\Sigma}_{11.2})}} \exp\left(-\frac{1}{2}\left(\mathbf{x}^{(11.2)} - \boldsymbol{\mu}^{(11.2)}\right)^\top \boldsymbol{\Sigma}_{11.2}^{-1}\left(\mathbf{x}^{(11.2)} - \boldsymbol{\mu}^{(11.2)}\right)\right) \end{aligned}$$

Multivariate Normal Distribution (Conditional Distribution)

The conditional density of $\mathbf{x}^{(1)}$ given that $\mathbf{x}^{(2)}$ is

$$
\begin{aligned}
f(\mathbf{x}^{(1)} \mid \mathbf{x}^{(2)}) &= \frac{f(\mathbf{x}^{(1)}, \mathbf{x}^{(2)})}{f(\mathbf{x}^{(2)})} \\
&= \frac{1}{\sqrt{(2\pi)^q \det(\boldsymbol{\Sigma}_{11.2})}} \exp\left(-\frac{1}{2}\left(\mathbf{x}^{(11.2)} - \boldsymbol{\mu}^{(11.2)}\right)^\top \boldsymbol{\Sigma}_{11.2}^{-1}\left(\mathbf{x}^{(11.2)} - \boldsymbol{\mu}^{(11.2)}\right)\right)
\end{aligned}
$$

where $\mathbf{x}^{(11.2)} = \mathbf{x}^{(1)} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\mathbf{x}^{(2)}$, $\boldsymbol{\mu}^{(11.2)} = \boldsymbol{\mu}^{(1)} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\mu}^{(2)}$ and $\boldsymbol{\Sigma}_{11.2} = \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$.

The density $f(\mathbf{x}^{(1)} \mid \mathbf{x}^{(2)})$ is a q-variate normal density with mean

$$
\mathbb{E}\left[\mathbf{x}^{(1)} \mid \mathbf{x}^{(2)}\right] = \boldsymbol{\mu}^{(1)} + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}^{(2)} - \boldsymbol{\mu}^{(2)}) = \boldsymbol{\nu}(\mathbf{x}^{(2)})
$$

and covariance matrix

$$
\begin{aligned}
\mathrm{Cov}\left[\mathbf{x}^{(1)} \mid \mathbf{x}^{(2)}\right] &= \mathbb{E}\left[(\mathbf{x}^{(1)} - \boldsymbol{\nu}(\mathbf{x}^{(2)}))(\mathbf{x}^{(1)} - \boldsymbol{\nu}(\mathbf{x}^{(2)}))^\top \mid \mathbf{x}^{(2)}\right] \\
&= \boldsymbol{\Sigma}_{11.2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}
\end{aligned}
$$