Multivariate Normal Theory

STA721 Linear Models Duke University

Merlise Clyde

September 4, 2019

duke.eps

Outline

- Multivariate Normal Distribution Singular Case
- Equal in Distribution
- Conditional Normal Distributions

Singular Case

$\mathbf{Y} = \boldsymbol{\mu} + \mathbf{AZ}$ with $\mathbf{Z} \in \mathbb{R}^d$ and $\mathbf{A}$ is $n \times d$

- $\mathsf{E}[\mathbf{Y}] = \mu$
- $\mathsf{Cov}(\mathbf{Y}) = \mathbf{AA}^T \geq 0$
- $\mathbf{Y} \sim \mathsf{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where $\boldsymbol{\Sigma} = \mathbf{AA}^T$

If $\boldsymbol{\Sigma}$ is singular then there is no density (on $\mathbb{R}^n$), but claim that $\mathbf{Y}$ still has a multivariate normal distribution!

Definition

$\mathbf{Y} \in \mathbb{R}^n$ has a multivariate normal distribution $\mathsf{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ if for any $\mathbf{v} \in \mathbb{R}^n$ $\mathbf{v}^T\mathbf{Y}$ has a univariate normal distribution with mean $\mathbf{v}^T\boldsymbol{\mu}$ and variance $\mathbf{v}^T\boldsymbol{\Sigma}\mathbf{v}$

Proof: need momemt generating or characteristic functions which uniquely characterize distribution.

Moment Generating Functions and Characteristics Functions

Univariate $Y \sim \mathsf{N}(\mu, \sigma^2)$

- MGF or Laplace Transform

$$m_Y(t) = \mathsf{E}[e^{t^T Y}] = \int e^{t^T y} f(y) dy = e^{t^T \mu + \frac{1}{2} t^2 \sigma^2}$$

- Characteristic function or Fourier transform

$$\varphi_Y(t) = \mathsf{E}[e^{-it^T Y}] = \int e^{-it^T y} f(y) dy = e^{-it^T \mu - \frac{1}{2} t^2 \sigma^2}$$

To show that $\mathbf{Y} = \boldsymbol{\mu} + \mathbf{AZ}$ has a $\mathsf{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution, we need to show that $\mathbf{v}^T\mathbf{Y}$ has a MGF or Characteristic function of a univariate normal with mean $\mathbf{v}^T\boldsymbol{\mu}$ and variance $\mathbf{v}^T\Sigma^b\mathbf{v}$.

Proof

$$
\begin{aligned}
\mathsf{E}[e^{-it\mathbf{v}^T\mathbf{Y}}] &= \mathsf{E}[e^{-it\mathbf{v}^T(\boldsymbol{\mu}+\mathbf{A}\mathbf{Z})}] \\
&= e^{-it\mathbf{v}^T\boldsymbol{\mu}}\mathsf{E}[e^{-it\mathbf{v}^T\mathbf{A}\mathbf{Z}}] \\
&= e^{-it\mathbf{v}^T\boldsymbol{\mu}}\mathsf{E}[e^{-it\mathbf{u}^T\mathbf{Z}}] \text{ for } \mathbf{u} = \mathbf{A}^T\mathbf{v} \in \mathbb{R}^n \\
&= e^{-it\mathbf{v}^T\boldsymbol{\mu}} \prod_{j=i}^{n} \mathsf{E}[e^{-itu_jZ_j}] \\
&= e^{-it\mathbf{v}^T\boldsymbol{\mu}} \prod \varphi(tu_j) \\
&= e^{-it\mathbf{v}^T\boldsymbol{\mu}} \prod e^{-\frac{1}{2}t^2u_j^2} \\
& e^{-it\mathbf{v}^T\boldsymbol{\mu}-\frac{1}{2}\sum_j t^2u_j^2} \\
& e^{-it\mathbf{v}^T\boldsymbol{\mu}-\frac{1}{2}t^2\mathbf{u}^T\mathbf{u}} \text{ note: } \mathbf{u}^T\mathbf{u} = \mathbf{v}^T\mathbf{A}\mathbf{A}^T\mathbf{v} = \mathbf{v}^T\boldsymbol{\Sigma}\mathbf{v} \\
& e^{-it\mathbf{v}^T\boldsymbol{\mu}-\frac{1}{2}t^2\mathbf{v}^T\boldsymbol{\Sigma}\mathbf{v}}
\end{aligned}
$$

Let $\mathbf{t} = t\mathbf{v}$ then $\varphi(t\mathbf{v}^T\mathbf{Y}) = \varphi(\mathbf{t}^T\mathbf{Y})$ yields multivariate normal mgf or characteristic function.

duke.eps

Multivariate Normal Moment Generating and Characteristic Functions

$\mathbf{Y} \sim \mathsf{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

- MGF or Laplace Transform

$$m_{\mathbf{Y}}(\mathbf{t}) = \mathsf{E}[e^{\mathbf{t}^T\mathbf{Y}}] = \int e^{\mathbf{t}^T\mathbf{y}} f(\mathbf{y}) d\mathbf{y} = e^{\mathbf{t}^T\boldsymbol{\mu} + \frac{1}{2}\mathbf{t}^T\boldsymbol{\Sigma}\mathbf{t}}$$

- Characteristic function or Fourier Transform

$$\phi_{\mathbf{Y}}(\mathbf{t}) = \mathsf{E}[e^{-i\mathbf{t}^T\mathbf{Y}}] = \int e^{-i\mathbf{t}^T\mathbf{y}} f(\mathbf{y}) d\mathbf{y} = e^{-i\mathbf{t}^T\boldsymbol{\mu} - \frac{1}{2}\mathbf{t}^T\boldsymbol{\Sigma}\mathbf{t}}$$

Linear Transformations are Normal

If $\mathbf{Y} \sim \mathsf{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ then for $\mathbf{A}$ $m \times n$

$$\mathbf{A}\mathbf{Y} \sim \mathsf{N}_m(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T)$$

$\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T$ does not have to be positive definite!
(Proof in book or linked video uses characteristic functions or MGFs)

Use to prove that all univariate and multivariate marginal distributions of normals are normal!

Equal in Distribution

Multiple ways to define the same normal:

- $\mathbf{Z}_1 \sim \mathrm{N}(\mathbf{0}, \mathbf{I}_n)$, $\mathbf{Z}_1 \in \mathbb{R}^n$ and take $\mathbf{A}$ $d \times n$
- $\mathbf{Z}_2 \sim \mathrm{N}(\mathbf{0}, \mathbf{I}_p)$, $\mathbf{Z}_2 \in \mathbb{R}^p$ and take $\mathbf{B}$ $d \times p$
- Define $\mathbf{Y} = \boldsymbol{\mu} + \mathbf{A}\mathbf{Z}_1$
- Define $\mathbf{W} = \boldsymbol{\mu} + \mathbf{B}\mathbf{Z}_2$

Theorem

If $\mathbf{Y} = \boldsymbol{\mu} + \mathbf{A}\mathbf{Z}_1$ *and* $\mathbf{W} = \boldsymbol{\mu} + \mathbf{B}\mathbf{Z}_2$ *then* $\mathbf{Y} \stackrel{\mathrm{D}}{=} \mathbf{W}$ *if and only if* $\mathbf{A}\mathbf{A}^T = \mathbf{B}\mathbf{B}^T = \boldsymbol{\Sigma}$

see linked video

Zero Correlation and Independence

Theorem

For a random vector $\mathbf{Y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *partitioned as*

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} \right)$$

then $Cov(\mathbf{Y}_1, \mathbf{Y}_2) = \boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}_{21}^T = \mathbf{0}$ *if and only if* $\mathbf{Y}_1$ *and* $\mathbf{Y}_2$ *are independent.*

Independence Implies Zero Covariance

Proof.

$$\mathrm{Cov}(\mathbf{Y}_1, \mathbf{Y}_2) = \mathrm{E}[(\mathbf{Y}_1 - \boldsymbol{\mu}_1)(\mathbf{Y}_2 - \boldsymbol{\mu}_2)^T]$$

If $\mathbf{Y}_1$ and $\mathbf{Y}_2$ are independent

$$\mathrm{E}[(\mathbf{Y}_1 - \boldsymbol{\mu}_1)(\mathbf{Y}_2 - \boldsymbol{\mu}_2)^T] = \mathrm{E}[(\mathbf{Y}_1 - \boldsymbol{\mu}_1)\mathrm{E}(\mathbf{Y}_2 - \boldsymbol{\mu}_2)^T] = \mathbf{0}\mathbf{0}^T = \mathbf{0}$$

therefore $\boldsymbol{\Sigma}_{12} = \mathbf{0}$

$\square$

Zero Covariance Implies Independence

Assume $\boldsymbol{\Sigma}_{12} = \mathbf{0}$

Proof

- Choose an

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2 \end{bmatrix}$$

  such that $\mathbf{A}_1\mathbf{A}_1^T = \boldsymbol{\Sigma}_{11}$, $\mathbf{A}_2\mathbf{A}_2^T = \boldsymbol{\Sigma}_{22}$
- Partition

$$\mathbf{Z} = \begin{bmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \end{bmatrix} \sim \mathsf{N}\left(\begin{bmatrix} \mathbf{0}_1 \\ \mathbf{0}_2 \end{bmatrix}, \begin{bmatrix} \mathbf{I}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_2 \end{bmatrix} \right) \text{ and } \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$$

- then $\mathbf{Y} \stackrel{\mathrm{D}}{=} \mathbf{A}\mathbf{Z} + \boldsymbol{\mu} \sim \mathsf{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

Continued

Proof.

$$\begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{bmatrix} \stackrel{\mathrm{D}}{=} \begin{bmatrix} \mathbf{A}_1\mathbf{Z}_1 + \boldsymbol{\mu}_1 \\ \mathbf{A}_2\mathbf{Z}_2 + \boldsymbol{\mu}_2 \end{bmatrix}$$

- But $\mathbf{Z}_1$ and $\mathbf{Z}_2$ are independent
- Functions of $\mathbf{Z}_1$ and $\mathbf{Z}_2$ are independent
- Therefore $\mathbf{Y}_1$ and $\mathbf{Y}_2$ are independent

$\square$

For Multivariate Normal Zero Covariance implies independence

Corollary

Corollary

If $\mathbf{Y} \sim N(\boldsymbol{\mu}, \sigma^2 \mathbf{I}_n)$ *and* $\mathbf{A}\mathbf{B}^T = \mathbf{0}$ *then* $\mathbf{AY}$ *and* $\mathbf{BY}$ *are independent.*

Proof.

$$\begin{bmatrix} \mathbf{W}_1 \\ \mathbf{W}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{A} \\ \mathbf{B} \end{bmatrix} \mathbf{Y} = \begin{bmatrix} \mathbf{AY} \\ \mathbf{BY} \end{bmatrix}$$

- $\textsf{Cov}(\mathbf{W}_1, \mathbf{W}_2) = \textsf{Cov}(\mathbf{AY}, \mathbf{BY}) = \sigma^2 \mathbf{A}\mathbf{B}^T$
- $\mathbf{AY}$ and $\mathbf{BY}$ are independent if $\mathbf{A}\mathbf{B}^T = \mathbf{0}$

$\square$

Conditional Distributions

Theorem

If joint distribution of $\mathbf{Y}_1$ and $\mathbf{Y}_2$ is

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} \right)$$

and $\Sigma_{22} > 0$ then

$$\mathbf{Y}_1 \mid \mathbf{Y}_2 = \mathbf{y}_2 \sim N\left(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{y}_2 - \boldsymbol{\mu}_2), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\right)$$

- The conditional distribution of $\mathbf{Y}_1$ given $\mathbf{Y}_2$ is also normal!
- Can replace $\boldsymbol{\Sigma}_{22}^{-1}$ by a Generalized inverse if $\boldsymbol{\Sigma}_{22}$ is singular.

Brute Force (full rank case) or Linear Transformations!

Derivation

Proof

- Define

$$\begin{bmatrix} \mathbf{W}_1 \\ \mathbf{W}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{I} & -\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{Y}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\mathbf{Y}_2 \\ \mathbf{Y}_2 \end{bmatrix}$$

- then

$$\mathbf{W}_1 \sim \mathrm{N}\left(\boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\right)$$

$$\mathbf{W}_2 \sim \mathrm{N}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$$

$$\mathrm{Cov}(\mathbf{W}_1, \mathbf{W}_2) = \mathbf{0}$$

Covariance of $\mathbf{W}_1$ and $\mathbf{W}_2$

$$\mathrm{Cov}(\mathbf{W}_1, \mathbf{W}_2) = \left[\mathbf{I} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\right] \begin{bmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{0} \\ \mathbf{I} \end{bmatrix}$$

Conditional Characteristic Function

- $\varphi_{\mathbf{Y}_1|\mathbf{Y}_2=\mathbf{y}_2}(t) = \mathsf{E}\left[e^{it^T\mathbf{Y}_1} \mid \mathbf{Y}_2 = \mathbf{y}_2\right]$
- Add zero

$$= \mathsf{E}\left[e^{it^T\mathbf{Y}_1 - it^T\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\mathbf{Y}_2 + it^T\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\mathbf{Y}_2} \mid \mathbf{Y}_2 = \mathbf{y}_2\right]$$

- Factor and exploit conditioning

$$= \mathsf{E}\left[e^{it^T\mathbf{Y}_1 - it^T\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\mathbf{Y}_2}\, e^{it^T\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\mathbf{Y}_2} \mid \mathbf{Y}_2 = \mathbf{y}_2\right]$$

$$= \mathsf{E}\left[e^{it^T\mathbf{W}_1} \mid \mathbf{Y}_2 = \mathbf{y}_2\right] e^{it^T\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\mathbf{y}_2}$$

- Independence of $\mathbf{W}_1 = \mathbf{Y}_1 - \Sigma_{12}\Sigma_{22}^{-1}$ and $\mathbf{Y}_2 = \mathbf{W}_2$

$$= \mathsf{E}\left[e^{it^T\mathbf{W}_1}\right]\, e^{it^T\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\mathbf{y}_2}$$

Combine

- $\mathbf{W}_1 \sim \mathsf{N}(\boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21})$

$$\varphi_{\mathbf{W}_1}(t) = e^{it^T(\boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\mu}_2) - \frac{1}{2}t^T(\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21})t}$$

- Combining

$$\begin{aligned}
\varphi_{\mathbf{Y}_1|\mathbf{Y}_2}(t) &= \varphi_{\mathbf{W}_1}(t)\, e^{it^T\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\mathbf{y}_2} \\
&= e^{it^T(\boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\mu}_2) - \frac{1}{2}t^T(\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21})t}\, e^{it^T\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\mathbf{y}_2} \\
&= e^{it^T(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{y}_2 - \boldsymbol{\mu}_2)) - \frac{1}{2}t^T(\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21})t}
\end{aligned}$$

- Characteristic function implies

$$\mathbf{Y}_1 \mid \mathbf{Y}_2 \sim \mathsf{N}(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{y}_2 - \boldsymbol{\mu}_2), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21})$$

Regression setting

Let $\mathbf{Y}_1 = Y$ and $\mathbf{Y}_2 = \mathbf{x}$
Then

$$Y \mid \mathbf{X} \sim \mathrm{N}(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x} - \boldsymbol{\mu}_2), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21})$$

$$Y \mid \mathbf{X} \sim \mathrm{N}(\boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\mu_2 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\mathbf{x}, \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21})$$

$$Y_i \mid \mathbf{X} \sim \mathrm{N}(\beta_0 + \mathbf{x}_i^T\boldsymbol{\beta}, \sigma^2)$$

Multivariate Normality is not necessary

General Definition

Definition

Let $\mathbf{V}$ be an vector space with inner product $\langle , \rangle$. Then $\mathbf{Y} \in \mathbf{V}$ has a multivariate normal distribution $\mathsf{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ if for any $\mathbf{v} \in \mathbf{V}$, $\langle \mathbf{v}, \mathbf{Y} \rangle$ has a normal distribution with mean $\langle \mathbf{v}, \boldsymbol{\mu} \rangle$ and variance $\langle \mathbf{v}, \boldsymbol{\Sigma}\mathbf{v} \rangle$

For usual Euclidean space inner product $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T\mathbf{y}$

For the energetic Student: Consider space of $n \times m$ matrices, and a random matrix $\mathbf{Y} \sim \mathsf{N}(\boldsymbol{\mu}, \mathbf{I} \otimes \boldsymbol{\Sigma})$ where $(\mathbf{I} \otimes \boldsymbol{\Sigma})M = \mathbf{I}M\boldsymbol{\Sigma}^T$ for M $n \times m$

Under the Inner product $\langle \mathbf{x}, \mathbf{y} \rangle \equiv \operatorname{tr} \mathbf{x}\mathbf{y}$, show that $\mathbf{Y}$ has a multivariate normal distribution on the space of $n \times m$ matrices