Multivariate Normal Theory

STA721 Linear Models Duke University

Merlise Clyde

September 5, 2017

Outline

- Multivariate Normal Distribution
- Linear Transformations
- Distribution of estimates under normality

Properties of MLE's Recap

- $\hat{\mathbf{Y}} = \hat{\boldsymbol{\mu}} = \mathbf{P_X Y}$ is an unbiased estimate of $\boldsymbol{\mu} = \mathbf{X}\beta$
- $\mathrm{E}[\mathbf{e}] = \mathbf{0}$ if $\boldsymbol{\mu} \in C(\mathbf{X})$

$$\mathrm{E}[\mathbf{e}] = \mathrm{E}[(\mathbf{I} - \mathbf{P_X})\mathbf{Y}]$$

- MLE of σ^2:

$$\hat{\sigma}^2 = \frac{\mathbf{e}^T\mathbf{e}}{n} = \frac{\mathbf{Y}^T(\mathbf{I} - \mathbf{P_X})\mathbf{Y}}{n}$$

  Is not an unbiased estimate of σ^2, but

$$\hat{\sigma}^2 \equiv \frac{\mathbf{e}^T\mathbf{e}}{n-p} = \frac{\mathbf{Y}^T(\mathbf{I} - \mathbf{P_X})\mathbf{Y}}{n-p}$$

  where p equals the rank of $\mathbf{X}$ is an unbiased estimate.

Sampling Distributions

- Distribution of $\hat{\beta}$
- Distribution of $\mathbf{P_X Y}$
- Distribution of $\mathbf{e}$

Univariate Normal

Definition

We say that Z has a standard Normal distribution

$$Z \sim \mathrm{N}(0, 1)$$

with mean 0 and variance 1 if it has density

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} z^2}$$

If $Y = \mu + \sigma Z$ then $Y \sim N(\mu, \sigma^2)$ with mean μ and variance σ^2

$$f_Y(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2}$$

Standard Multivariate Normal

Let $z_i \stackrel{\text{iid}}{\sim} \mathrm{N}(0,1)$ for $i = 1, \ldots, d$ and define

$$\mathbf{Z} \equiv \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_d \end{bmatrix}$$

- Density of Z:

$$\begin{aligned} f_{\mathbf{Z}}(\mathbf{z}) &= \prod_{j=1}^{d} \frac{1}{\sqrt{2\pi}} e^{-z_i^2/2} \\ &= (2\pi)^{-d/2} e^{-\frac{1}{2}(\mathbf{Z}^T\mathbf{Z})} \end{aligned}$$

- $\mathrm{E}[\mathbf{Z}] = \mathbf{0}$ and $\mathrm{Cov}[\mathbf{Z}] = \mathbf{I}_d$
- $\mathbf{Z} \sim \mathrm{N}(\mathbf{0}_d, \mathbf{I}_d)$

Multivariate Normal

For a d dimensional multivariate normal random vector, we write $\mathbf{Y} \sim N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

- $\mathrm{E}[\mathbf{Y}] = \boldsymbol{\mu}$: d dimensional vector with means $E[Y_j]$
- $\mathrm{Cov}[\mathbf{Y}] = \boldsymbol{\Sigma}$: $d \times d$ matrix with diagonal elements that are the variances of Y_j and off diagonal elements that are the covariances $\mathrm{E}[(Y_j - \mu_j)(Y_k - \mu_k)]$

Density

If $\boldsymbol{\Sigma}$ is positive definite ($\mathbf{x}'\boldsymbol{\Sigma}\mathbf{x} > 0$ for any $\mathbf{x} \neq 0$ in $\mathbb{R}^d$) then $\mathbf{Y}$ has a density [1]

$$p(\mathbf{Y}) = (2\pi)^{-d/2}|\boldsymbol{\Sigma}|^{-1/2} \exp(-\frac{1}{2}(\mathbf{Y} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \boldsymbol{\mu}))$$

[1] with respect to Lebesgue measure on $\mathbb{R}^d$

Spectral Theorem

Theorem

If $\mathbf{A}$ *$(n \times n)$ is a symmetric real matrix then there exists a* $\mathbf{U}$ *$(n \times n)$ such that* $\mathbf{U}^T\mathbf{U} = \mathbf{U}\mathbf{U}^T = \mathbf{I}_n$ *and a diagonal matrix* $\mathbf{\Lambda}$ *with elements* λ_i *such that* $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$

- $\mathbf{U}$ is an orthogonal matrix; $\mathbf{U}^{-1} = \mathbf{U}^T$
- The columns of $\mathbf{U}$ from an Orthonormal Basis for $\mathbb{R}^n$
- rank of $\mathbf{A}$ equals the number of non-zero eigenvalues λ_i
- Columns of $\mathbf{U}$ associated with non-zero eigenvalues form an ONB for $C(\mathbf{A})$ (eigenvectors of $\mathbf{A}$)
- $\mathbf{A}^p = \mathbf{U}\mathbf{\Lambda}^p\mathbf{U}^T$ (matrix powers)
- a (symmetric) square root of $\mathbf{A} > 0$ is $\mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{U}^T$

Multivariate Normal Density

- Density of $Z \sim \mathrm{N}(\mathbf{0}, \mathbf{I}_d)$:

$$\begin{aligned} f_{\mathbf{Z}}(\mathbf{z}) &= \prod_{j=1}^{d} \frac{1}{\sqrt{2\pi}} e^{-z_i^2/2} \\ &= (2\pi)^{-d/2} e^{-\frac{1}{2}(\mathbf{Z}^T\mathbf{Z})} \end{aligned}$$

- Write $\mathbf{Y} = \boldsymbol{\mu} + \mathbf{A}\mathbf{Z}$
- Solve for $\mathbf{Z} = g(\mathbf{Y})$
- Jacobian of the transformation $J(\mathbf{Z} \to \mathbf{Y}) = |\frac{\partial g}{\partial \mathbf{Y}}|$
- substitute $g(\mathbf{Y})$ for $\mathbf{Z}$ in density and multiply by Jacobian

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{Z}}(\mathbf{z}) J(\mathbf{Z} \to \mathbf{Y})$$

Multivariate Normal Density

$$\mathbf{Y} = \boldsymbol{\mu} + \mathbf{A}\mathbf{Z} \quad \text{for } \mathbf{Z} \sim \mathsf{N}(\mathbf{0}, \mathbf{I}_d) \tag{1}$$

Proof.

- since $\boldsymbol{\Sigma} > 0$, $\exists$ by the spectral theorem an $\mathbf{A}$ $(d \times d)$ such that $\mathbf{A} > 0$ and $\mathbf{A}\mathbf{A}^T = \boldsymbol{\Sigma}$
- $\mathbf{A} > 0 \Rightarrow \mathbf{A}^{-1}$ exists
- Multiply both sides (1) by $\mathbf{A}^{-1}$:

$$\mathbf{A}^{-1}\mathbf{Y} = \mathbf{A}^{-1}\boldsymbol{\mu} + \mathbf{A}^{-1}\mathbf{A}\mathbf{Z}$$

- Rearrange $\mathbf{A}^{-1}(\mathbf{Y} - \boldsymbol{\mu}) = \mathbf{Z}$
- Jacobian of transformation $d\mathbf{Z} = |\mathbf{A}^{-1}| d\mathbf{Y}$
- Substitute and simplify algebra

$$f(\mathbf{Y}) = (2\pi)^{-d/2} |\boldsymbol{\Sigma}|^{-1/2} \exp(-\frac{1}{2}(\mathbf{Y} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{Y} - \boldsymbol{\mu}))$$

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 normal distribution with mean $\mathbf{v}^T\boldsymbol{\mu}$ and variance $\mathbf{v}^T\boldsymbol{\Sigma}\mathbf{v}$

see linked videos using characteristic functions:

$$Y \sim \mathsf{N}(\mu, \sigma^2) \Leftrightarrow \varphi_y(t) \equiv \mathsf{E}[e^{itY}] = e^{it\mu - t^2\sigma^2/2}$$

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)

Distribution of $\hat{\mathbf{Y}}$ and $\mathbf{e}$ (marginally)

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* $\mathbf{Y} \overset{\mathrm{D}}{=} \mathbf{W}$ *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.

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

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

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

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

$\square$

Zero Covariance Implies Independence

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

Proof

- Choose an
$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2 \end{bmatrix}$$
such that $\mathbf{A}_1\mathbf{A}_1^T = \mathbf{\Sigma}_{11}$, $\mathbf{A}_2\mathbf{A}_2^T = \mathbf{\Sigma}_{22}$
- Partition
$$\mathbf{Z} = \begin{bmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \end{bmatrix} \sim \mathrm{N}\left(\begin{bmatrix} \mathbf{0}_1 \\ \mathbf{0}_2 \end{bmatrix}, \begin{bmatrix} \mathbf{I}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_2 \end{bmatrix} \right) \text{ and } \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$$
- then $\mathbf{Y} \stackrel{\mathrm{D}}{=} \mathbf{AZ} + \boldsymbol{\mu} \sim \mathrm{N}(\boldsymbol{\mu}, \mathbf{\Sigma})$

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.

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

$\square$

Joint Distribution of $\hat{\mathbf{Y}}$ and $\mathbf{e}$

More Distribution Theory

Distributions unconditional on σ^2

- χ^2 distributions ($\hat{\sigma}^2$)
- t distribution ($\hat{\mathbf{Y}}$, $\mathbf{e}$, $\hat{\boldsymbol{\beta}}$)