MLES & Multivariate Normal Theory

STA721 Linear Models Duke University

Merlise Clyde

September 1, 2015

Outline

- Expectations of Quadratic Forms
- Multivariate Normal Distribution
- Linear Transformations

Properties of MLE's Recap

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

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

- 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 this an unbiased estimate of σ^2?

Need expectations of quadratic forms $\mathbf{Y}^T\mathbf{A}\mathbf{Y}$ for $\mathbf{A}$ an $n \times n$ matrix $\mathbf{Y}$ a random vector in $\mathbb{R}^n$

Quadratic Forms

Without loss of generality we can assume that $\mathbf{A} = \mathbf{A}^T$

- $\mathbf{Y}^T\mathbf{A}\mathbf{Y}$ is a scalar
- $\mathbf{Y}^T\mathbf{A}\mathbf{Y} = (\mathbf{Y}^T\mathbf{A}\mathbf{Y})^T = \mathbf{Y}^T\mathbf{A}^T\mathbf{Y}$

$$\begin{aligned}
\frac{\mathbf{Y}^T\mathbf{A}\mathbf{Y} + \mathbf{Y}^T\mathbf{A}^T\mathbf{Y}}{2} &= \mathbf{Y}^T\mathbf{A}\mathbf{Y} \\
\mathbf{Y}^T\frac{(\mathbf{A} + \mathbf{A}^T)}{2}\mathbf{Y} &= \mathbf{Y}^T\mathbf{A}\mathbf{Y}
\end{aligned}$$

- may take $\mathbf{A} = \mathbf{A}^T$

Expectations of Quadratic Forms

Theorem

Let $\mathbf{Y}$ *be a random vector in* $\mathbb{R}^n$ *with* $E[\mathbf{Y}] = \boldsymbol{\mu}$ *and* $Cov(\mathbf{Y}) = \boldsymbol{\Sigma}$*. Then* $E[\mathbf{Y}^T\mathbf{A}\mathbf{Y}] = tr\mathbf{A}\Sigma + \boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}$*.*

Result useful for finding expected values of Mean Squares; no normality required!

Proof

Start with $(\mathbf{Y}-\boldsymbol{\mu})^T\mathbf{A}(\mathbf{Y}-\boldsymbol{\mu})$, expand and take expectations

$$
\begin{aligned}
\mathsf{E}[(\mathbf{Y}-\boldsymbol{\mu})^T\mathbf{A}(\mathbf{Y}-\boldsymbol{\mu})] &= \mathsf{E}[\mathbf{Y}^T\mathbf{A}\mathbf{Y}+\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}-\mu^T\mathbf{A}\mathbf{Y}-\mathbf{Y}^T\mathbf{A}\mu] \\
&= \mathsf{E}[\mathbf{Y}^T\mathbf{A}\mathbf{Y}]+\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}-\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}-\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu} \\
&= \mathsf{E}[\mathbf{Y}^T\mathbf{A}\mathbf{Y}]-\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}
\end{aligned}
$$

Rearrange

$$
\begin{aligned}
\mathsf{E}[\mathbf{Y}^T\mathbf{A}\mathbf{Y}] &= \mathsf{E}[(\mathbf{Y}-\boldsymbol{\mu})^T\mathbf{A}(\mathbf{Y}-\boldsymbol{\mu})]+\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu} \\
&= \mathsf{E}[\mathrm{tr}(\mathbf{Y}-\boldsymbol{\mu})^T\mathbf{A}(\mathbf{Y}-\boldsymbol{\mu})]+\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu} \\
&= \mathsf{E}[\mathrm{tr}\mathbf{A}(\mathbf{Y}-\boldsymbol{\mu})(\mathbf{Y}-\boldsymbol{\mu})^T]+\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu} \\
&= \mathrm{tr}\mathsf{E}[\mathbf{A}(\mathbf{Y}-\boldsymbol{\mu})(\mathbf{Y}-\boldsymbol{\mu})^T]+\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu} \\
&= \mathrm{tr}\mathbf{A}\mathsf{E}([(\mathbf{Y}-\boldsymbol{\mu})(\mathbf{Y}-\boldsymbol{\mu})^T]+\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu} \\
&= \mathrm{tr}\mathbf{A}\boldsymbol{\Sigma}+\boldsymbol{\mu}^T\mathbf{A}\boldsymbol{\mu}
\end{aligned}
$$

$\mathrm{tr}\mathbf{A} \equiv \sum_{i=1}^n a_{ii}$

Expectation of $\hat{\sigma}^2$

Use the theorem:

$$\begin{aligned}
\mathrm{E}[\mathbf{Y}^T(\mathbf{I}-\mathbf{P_X})\mathbf{Y}] &= \mathrm{tr}(\mathbf{I}-\mathbf{P_X})\sigma^2\mathbf{I} + \boldsymbol{\mu}^T(\mathbf{I}-\mathbf{P_X})\boldsymbol{\mu} \\
&= \sigma^2\mathrm{tr}(\mathbf{I}-\mathbf{P_X}) \\
&= \sigma^2 r(\mathbf{I}-\mathbf{P_X}) \\
&= \sigma^2(n - r(\mathbf{X}))
\end{aligned}$$

Therefore an unbiased estimate of σ^2 is

$$\frac{\mathbf{e}^T\mathbf{e}}{n - r(\mathbf{X})}$$

If $\mathbf{X}$ is full rank ($r(\mathbf{X}) = p$) and $\mathbf{P_X} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ then the

$$\begin{aligned}
\mathrm{tr}(\mathbf{P_X}) &= \mathrm{tr}(\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T) \\
&= \mathrm{tr}(\mathbf{X}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}) \\
&= \mathrm{tr}(\mathbf{I}_p) = p
\end{aligned}$$

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 $\boldsymbol{\Lambda}$ with elements λ_i such that $\mathbf{A} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T$

- $\mathbf{U}$ 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 $\mathbf{A}$)
- $\mathbf{A}^p = \mathbf{U}\boldsymbol{\Lambda}^p\mathbf{U}^T$ (matrix powers)
- a square root of $\mathbf{A} > 0$ is $\mathbf{U}\boldsymbol{\Lambda}^{1/2}\mathbf{U}^T$

Projections

Projection Matrix

If $\mathbf{P}$ is an orthogonal projection matrix, then its eigenvalues λ_i are either zero or one with $\text{tr}(\mathbf{P}) = \sum_i(\lambda_i) = r(\mathbf{P})$

- $\mathbf{P} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T$
- $\mathbf{P} = \mathbf{P}^2 \Rightarrow \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T = \mathbf{U}\boldsymbol{\Lambda}^2\mathbf{U}^T$
- $\boldsymbol{\Lambda} = \boldsymbol{\Lambda}^2$ is true only for $\lambda_i = 1$ or $\lambda_i = 0$
- Since $r(\mathbf{P})$ is the number of non-zero eigenvalues, $r(\mathbf{P}) = \sum \lambda_i = \text{tr}(\mathbf{P})$

$$\mathbf{P} = [\mathbf{U}_P \mathbf{U}_{P^\perp}] \begin{bmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0}_{n-r} \end{bmatrix} \begin{bmatrix} \mathbf{U}_P^T \\ \mathbf{U}_{P^\perp}^T \end{bmatrix} = \mathbf{U}_P \mathbf{U}_P^T$$

$$\mathbf{P} = \sum_{i=1}^{r} \mathbf{u}_i \mathbf{u}_i^T$$

sum of r rank 1 projections.

Distributions

- Distribution of $\hat{\beta}$
- Distribution of $\mathbf{P_X Y}$
- Distribution of $\mathbf{e}$
- Distribution ot $\hat{\sigma}^2$

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

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

- Density of Z:

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

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

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 [a]

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

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

Multivariate Normal Density

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

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

- 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

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

Proof.

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

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

- 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

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

$\square$

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 Lessons in Sakai for videos using Characteristic functions

Linear Transformations are Normal

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

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

$\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T$ does not have to be positive definite!

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* $\mathbf{A}\mathbf{A}^T = \mathbf{B}\mathbf{B}^T = \boldsymbol{\Sigma}$

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

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

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

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

Another Useful Result

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 $\mathbf{A}\mathbf{B}^T = \mathbf{0}$

$\square$