Sampling Distributions and Distribution Theory

STA 721: Lecture 7

Merlise Clyde (clyde@duke.edu)
Duke University

Outline

- distributions of $\hat{\boldsymbol{\beta}}$, $\hat{\mathbf{Y}}$, $\hat{\boldsymbol{\epsilon}}$ under normality
- Unbiased Estimation of σ^2
- sampling distribution of $\hat{\sigma^2}$
- independence

Readings:

- Christensen Chapter 1, 2.91 and Appendix C
- Seber & Lee Chapter 3.3 - 3.5

Multivariate Normal

Under the linear model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, $\textsf{E}[\boldsymbol{\epsilon}] = \mathbf{0}_n$ and $\textsf{Cov}[\boldsymbol{\epsilon}] = \sigma^2\mathbf{I}_n$, we had

- $\textsf{E}[\hat{\boldsymbol{\beta}}] = \boldsymbol{\beta}$
- $\textsf{E}[\hat{\mathbf{Y}}] = \mathbf{P_X}\mathbf{Y} = \mathbf{X}\boldsymbol{\beta}$
- $\textsf{E}[\hat{\boldsymbol{\epsilon}}] = (\mathbf{I}_n - \mathbf{P_X})\mathbf{Y} = \mathbf{0}_n$
- distributions if $\epsilon_i \sim \textsf{N}(0, \sigma^2)$?

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

Transformations of Normal Random Variables

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

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

- $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} \sim \mathsf{N}(\boldsymbol{\beta}, \sigma^2(\mathbf{X}^T\mathbf{X})^{-1})$
- $\hat{\mathbf{Y}} = \mathbf{P}_{\mathbf{X}}\mathbf{Y} \sim \mathsf{N}(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{P}_{\mathbf{X}})$
- $\hat{\boldsymbol{\epsilon}} = (\mathbf{I}_n - \mathbf{P}_{\mathbf{X}})\mathbf{Y} \sim \mathsf{N}(\mathbf{0}, \sigma^2(\mathbf{I}_n - \mathbf{P}_{\mathbf{X}}))$

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

Singular Case

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

Definition: Multivariate Normal

$\mathbf{Y} \in \mathbb{R}^n$ 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

Use moment generating or characteristic functions which uniquely characterize distribution to show that $\mathbf{v}^T\mathbf{Y}$ has a univariate normal distribution.

- both $\hat{\mathbf{Y}}$ and $\hat{\boldsymbol{\epsilon}}$ have multivariate normal distributions even though they do not have densities! (singular distributions)

Distribution of MLE of σ^2

Recall we found the MLE of σ^2

$$\hat{\sigma}^2 = \frac{\hat{\boldsymbol{\epsilon}}^T \hat{\boldsymbol{\epsilon}}}{n}$$

Distribution of RSS

Since $\boldsymbol{\epsilon} \sim \mathsf{N}(\mathbf{0}_n, \sigma^2 \mathbf{I}_n)$ and $\mathsf{N} \in \mathbb{R}^{n \times (n-p)}$,

$$\mathsf{N}^T \boldsymbol{\epsilon} = \mathbf{e} \sim \mathsf{N}(\mathbf{0}_{n-p}, \sigma^2 \mathsf{N}^T \mathsf{N}) = \mathsf{N}(\mathbf{0}_{n-p}, \sigma^2 \mathbf{I}_{n-p})$$

$$\begin{aligned}
\mathsf{RSS} &= \sum_{i=1}^{n-p} e_i^2 \\
&\stackrel{\mathrm{D}}{=} \sum_{i=1}^{n-p} (\sigma z_i)^2 \qquad \text{where } \mathbf{Z} \sim \mathsf{N}(\mathbf{0}_{n-p}, \mathbf{I}_{n-p}) \\
&= \sigma^2 \sum_{i=1}^{n-p} z_i^2 \\
&\stackrel{\mathrm{D}}{=} \sigma^2 \chi^2_{n-p}
\end{aligned}$$

Background Theory: If $\mathbf{Z} \sim \mathsf{N}_d(\mathbf{0}_d, \mathbf{I}_d)$, then $\mathbf{Z}^T \mathbf{Z} \sim \chi^2_d$

Unbiased Estimate of σ^2

- Expected value of a χ^2_d random variable is d (the degrees of freedom)
- $\textsf{E}[\textsf{RSS}] = \textsf{E}[\sigma^2 \chi^2_{n-p}] = \sigma^2(n-p)$
- the expected value of the MLE is

$$\hat{\sigma}^2 = \textsf{E}[\textsf{RSS}]/n = \sigma^2 \frac{(n-p)}{n}$$

  so is biased
- an unbiased estimator of σ^2, is $s^2 = \textsf{RSS}/(n-p)$
- note: we can find the expectation of $\hat{\sigma}^2$ or s^2 based on the covariance of $\boldsymbol{\epsilon}$ without assuming normality by exploiting properties of the trace.

Distribution of $\hat{\beta}$

$\hat{\boldsymbol{\beta}} \sim \mathsf{N}\left(\boldsymbol{\beta}, \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}\right)$

- do not know σ^2
- Need a distribution that does not depend on unknown parameters for deriving confidence intervals and hypothesis tests for $\boldsymbol{\beta}$.
- what if we plug in $\boldsymbol{s}^2$ or $\hat{\sigma}^2$ for σ^2?
- won't be multivariate normal
- need to reflect uncertainty in estimating σ^2
- first show that $\hat{\boldsymbol{\beta}}$ and $\boldsymbol{s}^2$ are independent

Independence of $\hat{\beta}$ and s^2

If the distribution of $\mathbf{Y}$ is normal, then $\hat{\boldsymbol{\beta}}$ and s^2 are statistically independent.

- The derivation of this result basically has three steps:
  1. $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\epsilon}}$ or $\mathbf{e}$ have zero covariance
  2. $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\epsilon}}$ or $\mathbf{e}$ are independent
  3. Conclude $\hat{\boldsymbol{\beta}}$ and RSS (or s^2) are independent

Step 1:

$$
\begin{aligned}
\textsf{Cov}[\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\epsilon}}] &= \textsf{E}[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\hat{\boldsymbol{\epsilon}}^T] \\
&= \textsf{E}[(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T(\mathbf{I} - \mathbf{P}_{\mathbf{X}})] \\
&= \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\mathbf{I} - \mathbf{P}_{\mathbf{X}}) \\
&= \mathbf{0}
\end{aligned}
$$

Zero Covariance $\Leftrightarrow$ Independence in Multivariate Normals

Step 2: $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\epsilon}}$ are independent

▼ **Theorem:** Zero Correlation and Independence

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

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

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

Proof: Independence implies Zero Covariance

Easy direction

- $\textsf{Cov}[\mathbf{W}_1, \mathbf{W}_2] = \textsf{E}[(\mathbf{W}_1 - \boldsymbol{\mu}_1)(\mathbf{W}_2 - \boldsymbol{\mu}_2)^T]$
- since they are independent

$$\begin{aligned} \textsf{Cov}[\mathbf{W}_1, \mathbf{W}_2] &= \textsf{E}[(\mathbf{W}_1 - \boldsymbol{\mu}_1)]\textsf{E}[(\mathbf{W}_2 - \boldsymbol{\mu}_2)^T] \\ &= \mathbf{0}\mathbf{0}^T \\ &= \mathbf{0} \end{aligned}$$

so $\mathbf{W}_1$ and $\mathbf{W}_2$ are uncorrelated

Zero Covariance Implies Independence

▼ **Proof**

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

- 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

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

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

Proof: continued

$$\mathbf{W} \overset{\mathrm{D}}{=} \mathbf{A}\mathbf{Z} + \boldsymbol{\mu} \sim \mathsf{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

$$\begin{aligned}
\begin{bmatrix} \mathbf{W}_1 \\ \mathbf{W}_2 \end{bmatrix} &\overset{\mathrm{D}}{=} \begin{bmatrix} \mathbf{A}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2 \end{bmatrix} \begin{bmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \end{bmatrix} + \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{A}_1\mathbf{Z}_1 + \boldsymbol{\mu}_1 \\ \mathbf{A}_2\mathbf{Z}_2 + \boldsymbol{\mu}_2 \end{bmatrix}
\end{aligned}$$

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

For Multivariate Normal Zero Covariance implies independence!

▼ Corollary

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

▼ Proof

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

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

- Since $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{Y}$ and $\hat{\boldsymbol{\epsilon}} = (\mathbf{I} - \mathbf{P}_{\mathbf{X}})\mathbf{Y}$ have zero covariance, using the corollary we have that $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\epsilon}}$ are independent to show Step 2.

Step 3:

Show $\hat{\boldsymbol{\beta}}$ and RSS are independent

- $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{Y}$ and $\hat{\boldsymbol{\epsilon}} = (\mathbf{I} - \mathbf{P}_{\mathbf{X}})\mathbf{Y}$ are independent
- functions of independent random variables are independent so $\hat{\boldsymbol{\beta}}$ and $\textsf{RSS} = \hat{\boldsymbol{\epsilon}}^T\hat{\boldsymbol{\epsilon}}$ are independent
- so $\hat{\boldsymbol{\beta}}$ and $s^2 = \textsf{RSS}/(n-p)$ are independent

This result will be critical for creating confidence regions and intervals for $\boldsymbol{\beta}$ and linear combinations of $\boldsymbol{\beta}$, $\lambda^T\boldsymbol{\beta}$ as well as testing hypotheses

Next Class

- shrinkage estimators
- Bayes and penalized loss functions