Random Vectors[1]

STA442/2101 Fall 2018

[1] See last slide for copyright information.

Background Reading: Renscher and Schaalje's *Linear models in statistics*

- Chapter 3 on Random Vectors and Matrices
- Chapter 4 on the Multivariate Normal Distribution

Overview

Random Vectors and Matrices

A *random matrix* is just a matrix of random variables. Their joint probability distribution is the distribution of the random matrix. Random matrices with just one column (say, $p \times 1$) may be called *random vectors.*

Expected Value

The expected value of a matrix is defined as the matrix of expected values. Denoting the $p \times c$ random matrix $\mathbf{X}$ by $[X_{i,j}]$,

$$E(\mathbf{X}) = [E(X_{i,j})].$$

Immediately we have natural properties like

$$
\begin{aligned}
E(\mathbf{X}+\mathbf{Y}) &= E([X_{i,j}]+[Y_{i,j}]) \\
&= [E(X_{i,j}+Y_{i,j})] \\
&= [E(X_{i,j})+E(Y_{i,j})] \\
&= [E(X_{i,j})]+[E(Y_{i,j})] \\
&= E(\mathbf{X})+E(\mathbf{Y}).
\end{aligned}
$$

Moving a constant through the expected value sign

Let $\mathbf{A} = [a_{i,j}]$ be an $r \times p$ matrix of constants, while $\mathbf{X}$ is still a $p \times c$ random matrix. Then

$$
\begin{aligned}
E(\mathbf{AX}) &= E\left(\left[\sum_{k=1}^{p} a_{i,k} X_{k,j}\right]\right) \\
&= \left[E\left(\sum_{k=1}^{p} a_{i,k} X_{k,j}\right)\right] \\
&= \left[\sum_{k=1}^{p} a_{i,k} E(X_{k,j})\right] \\
&= \mathbf{A}E(\mathbf{X}).
\end{aligned}
$$

Similar calculations yield $E(\mathbf{AXB}) = \mathbf{A}E(\mathbf{X})\mathbf{B}$.

Variance-Covariance Matrices

Let $\mathbf{X}$ be a $p \times 1$ random vector with $E(\mathbf{X}) = \boldsymbol{\mu}$. The *variance-covariance matrix* of $\mathbf{X}$ (sometimes just called the *covariance matrix*), denoted by $cov(\mathbf{X})$, is defined as

$$cov(\mathbf{X}) = E\left\{(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^\top\right\}.$$

$cov(\mathbf{X}) = E\left\{(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})^\top\right\}$

$$
\begin{aligned}
cov(\mathbf{X}) &= E\left\{\begin{pmatrix} X_1-\mu_1 \\ X_2-\mu_2 \\ X_3-\mu_3 \end{pmatrix}\begin{pmatrix} X_1-\mu_1 & X_2-\mu_2 & X_3-\mu_3 \end{pmatrix}\right\} \\
&= E\left\{\left(\begin{array}{lll} (X_1-\mu_1)^2 & (X_1-\mu_1)(X_2-\mu_2) & (X_1-\mu_1)(X_3-\mu_3) \\ (X_2-\mu_2)(X_1-\mu_1) & (X_2-\mu_2)^2 & (X_2-\mu_2)(X_3-\mu_3) \\ (X_3-\mu_3)(X_1-\mu_1) & (X_3-\mu_3)(X_2-\mu_2) & (X_3-\mu_3)^2 \end{array}\right)\right. \\
&= \left(\begin{array}{lll} E\{(X_1-\mu_1)^2\} & E\{(X_1-\mu_1)(X_2-\mu_2)\} & E\{(X_1-\mu_1)(X_3 - \\ E\{(X_2-\mu_2)(X_1-\mu_1)\} & E\{(X_2-\mu_2)^2\} & E\{(X_2-\mu_2)(X_3 - \\ E\{(X_3-\mu_3)(X_1-\mu_1)\} & E\{(X_3-\mu_3)(X_2-\mu_2)\} & E\{(X_3-\mu_3)^2\} \end{array}\right. \\
&= \begin{pmatrix} Var(X_1) & Cov(X_1,X_2) & Cov(X_1,X_3) \\ Cov(X_1,X_2) & Var(X_2) & Cov(X_2,X_3) \\ Cov(X_1,X_3) & Cov(X_2,X_3) & Var(X_3) \end{pmatrix}.
\end{aligned}
$$

So, the covariance matrix $cov(\mathbf{X})$ is a $p \times p$ symmetric matrix with variances on the main diagonal and covariances on the off-diagonals.

Matrix of covariances between two random vectors

Let $\mathbf{X}$ be a $p \times 1$ random vector with $E(\mathbf{X}) = \boldsymbol{\mu}_x$ and let $\mathbf{Y}$ be a $q \times 1$ random vector with $E(\mathbf{Y}) = \boldsymbol{\mu}_y$. The $p \times q$ matrix of covariances between the elements of $\mathbf{X}$ and the elements of $\mathbf{Y}$ is

$$cov(\mathbf{X}, \mathbf{Y}) = E\left\{(\mathbf{X} - \boldsymbol{\mu}_x)(\mathbf{Y} - \boldsymbol{\mu}_y)^\top\right\}.$$

Adding a constant has no effect

On variances and covariances

- $cov(\mathbf{X}+\mathbf{a}) = cov(\mathbf{X})$
- $cov(\mathbf{X}+\mathbf{a},\mathbf{Y}+\mathbf{b}) = cov(\mathbf{X},\mathbf{Y})$

These results are clear from the definitions:

- $cov(\mathbf{X}) = E\left\{(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})^\top\right\}$
- $cov(\mathbf{X},\mathbf{Y}) = E\left\{(\mathbf{X}-\boldsymbol{\mu}_x)(\mathbf{Y}-\boldsymbol{\mu}_y)^\top\right\}$

Sometimes it is useful to let $\mathbf{a} = -\boldsymbol{\mu}_x$ and $\mathbf{b} = -\boldsymbol{\mu}_y$.

Analogous to $Var(a\,X) = a^2\,Var(X)$

Let $\mathbf{X}$ be a $p \times 1$ random vector with $E(\mathbf{X}) = \boldsymbol{\mu}$ and $cov(\mathbf{X}) = \boldsymbol{\Sigma}$, while $\mathbf{A} = [a_{i,j}]$ is an $r \times p$ matrix of constants. Then

$$\begin{aligned}
cov(\mathbf{AX}) &= E\left\{(\mathbf{AX} - \mathbf{A}\boldsymbol{\mu})(\mathbf{AX} - \mathbf{A}\boldsymbol{\mu})^\top\right\} \\
&= E\left\{\mathbf{A}(\mathbf{X} - \boldsymbol{\mu})\left(\mathbf{A}(\mathbf{X} - \boldsymbol{\mu})\right)^\top\right\} \\
&= E\left\{\mathbf{A}(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^\top\mathbf{A}^\top\right\} \\
&= \mathbf{A}E\{(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^\top\}\mathbf{A}^\top \\
&= \mathbf{A}cov(\mathbf{X})\mathbf{A}^\top \\
&= \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top
\end{aligned}$$

The Multivariate Normal Distribution

The $p \times 1$ random vector $\mathbf{X}$ is said to have a *multivariate normal distribution*, and we write $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, if $\mathbf{X}$ has (joint) density

$$f(\mathbf{x}) = \frac{1}{|\boldsymbol{\Sigma}|^{\frac{1}{2}}(2\pi)^{\frac{p}{2}}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right),$$

where $\boldsymbol{\mu}$ is $p \times 1$ and $\boldsymbol{\Sigma}$ is $p \times p$ symmetric and positive definite.

$\mathbf{\Sigma}$ positive definite

In the multivariate normal definition

- Positive definite means that for any non-zero $p \times 1$ vector $\mathbf{a}$, we have $\mathbf{a}^\top \mathbf{\Sigma} \mathbf{a} > 0$.
- Since the one-dimensional random variable $Y = \sum_{i=1}^p a_i X_i$ may be written as $Y = \mathbf{a}^\top \mathbf{X}$ and $Var(Y) = cov(\mathbf{a}^\top \mathbf{X}) = \mathbf{a}^\top \mathbf{\Sigma} \mathbf{a}$, it is natural to require that $\mathbf{\Sigma}$ be positive definite.
- All it means is that every non-zero linear combination of $\mathbf{X}$ values has a positive variance.
- And recall $\mathbf{\Sigma}$ positive definite is equivalent to $\mathbf{\Sigma}^{-1}$ positive definite.

Analogies

(Multivariate normal reduces to the univariate normal when $p = 1$)

- Univariate Normal
  - $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right\}$
  - $E(X) = \mu, Var(X) = \sigma^2$
  - $\frac{(X-\mu)^2}{\sigma^2} \sim \chi^2(1)$
- Multivariate Normal
  - $f(\mathbf{x}) = \frac{1}{|\mathbf{\Sigma}|^{\frac{1}{2}}(2\pi)^{\frac{p}{2}}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}$
  - $E(\mathbf{X}) = \boldsymbol{\mu}$, $cov(\mathbf{X}) = \mathbf{\Sigma}$
  - $(\mathbf{X}-\boldsymbol{\mu})^\top\mathbf{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu}) \sim \chi^2(p)$

More properties of the multivariate normal

- If $\mathbf{c}$ is a vector of constants, $\mathbf{X} + \mathbf{c} \sim N(\mathbf{c} + \boldsymbol{\mu}, \boldsymbol{\Sigma})$
- If $\mathbf{A}$ is a matrix of constants, $\mathbf{AX} \sim N(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top)$
- Linear combinations of multivariate normals are multivariate normal.
- All the marginals (dimension less than p) of $\mathbf{X}$ are (multivariate) normal, but it is possible in theory to have a collection of univariate normals whose joint distribution is not multivariate normal.
- For the multivariate normal, zero covariance implies independence. The multivariate normal is the only continuous distribution with this property.

An easy example

If you do it the easy way

Let $\mathbf{X} = (X_1, X_2, X_3)^\top$ be multivariate normal with

$$\boldsymbol{\mu} = \left(\begin{array}{c} 1 \\ 0 \\ 6 \end{array} \right) \text{ and } \boldsymbol{\Sigma} = \left(\begin{array}{ccc} 2 & 1 & 0 \\ 1 & 4 & 0 \\ 0 & 0 & 2 \end{array} \right).$$

Let $Y_1 = X_1 + X_2$ and $Y_2 = X_2 + X_3$. Find the joint distribution of Y_1 and Y_2.

In matrix terms

$Y_1 = X_1 + X_2$ and $Y_2 = X_2 + X_3$ means $\mathbf{Y} = \mathbf{AX}$

$$\left(\begin{array}{c} Y_1 \\ Y_2 \end{array} \right) = \left(\begin{array}{ccc} 1 & 1 & 0 \\ 0 & 1 & 1 \end{array} \right) \left(\begin{array}{c} X_1 \\ X_2 \\ X_3 \end{array} \right)$$

$\mathbf{Y} = \mathbf{AX} \sim N(\mathbf{A}\boldsymbol{\mu}, \mathbf{A\Sigma A}^\top)$

You could do it by hand, but

```
> mu = cbind(c(1,0,6))
> Sigma = rbind( c(2,1,0),
+                c(1,4,0),
+                c(0,0,2) )
> A = rbind( c(1,1,0),
+            c(0,1,1) ); A
> A %*% mu                 # E(Y)
     [,1]
[1,]    1
[2,]    6
> A %*% Sigma %*% t(A)   # cov(Y)
     [,1] [,2]
[1,]    8    5
[2,]    5    6
```

Regression

$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, with $\boldsymbol{\epsilon} \sim N_n(\mathbf{0}, \sigma^2\mathbf{I}_n)$.

So $\mathbf{y} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}_n)$.

$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y} = \mathbf{A}\mathbf{y}$.

So $\widehat{\boldsymbol{\beta}}$ is multivariate normal.

Just calculate the mean and covariance matrix.

$$
\begin{aligned}
E(\widehat{\boldsymbol{\beta}}) &= E\left((\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y}\right) \\
&= (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top E(\mathbf{y}) \\
&= (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{X}\boldsymbol{\beta} \\
&= \boldsymbol{\beta}
\end{aligned}
$$

Covariance matrix of $\widehat{\boldsymbol{\beta}}$

Using $cov(\mathbf{Aw}) = \mathbf{A}cov(\mathbf{w})\mathbf{A}^\top$

$$
\begin{aligned}
cov(\widehat{\boldsymbol{\beta}}) &= cov\left((\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y}\right) \\
&= (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top cov(\mathbf{y})\left((\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\right)^\top \\
&= (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\sigma^2\mathbf{I}_n\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1\top} \\
&= \sigma^2(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1} \\
&= \sigma^2(\mathbf{X}^\top\mathbf{X})^{-1}
\end{aligned}
$$

So $\widehat{\boldsymbol{\beta}} \sim N_p\left(\boldsymbol{\beta}, \sigma^2(\mathbf{X}^\top\mathbf{X})^{-1}\right)$

A couple of things to prove

- $(\mathbf{X} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\mu}) \sim \chi^2(p)$
- $\overline{X}$ and S^2 independent under normal random sampling.

Recall the square root matrix

Covariance matrix $\mathbf{\Sigma}$ is real and symmetric matrix, so we have the spectral decomposition

$$\begin{aligned}
\mathbf{\Sigma} &= \mathbf{P}\mathbf{\Lambda}\mathbf{P}^\top \\
&= \mathbf{P}\mathbf{\Lambda}^{1/2}\mathbf{\Lambda}^{1/2}\mathbf{P}^\top \\
&= \mathbf{P}\mathbf{\Lambda}^{1/2}\,\mathbf{I}\,\mathbf{\Lambda}^{1/2}\mathbf{P}^\top \\
&= \mathbf{P}\mathbf{\Lambda}^{1/2}\mathbf{P}^\top \;\; \mathbf{P}\mathbf{\Lambda}^{1/2}\mathbf{P}^\top \\
&= \quad \mathbf{\Sigma}^{1/2} \qquad\quad \mathbf{\Sigma}^{1/2}
\end{aligned}$$

So $\mathbf{\Sigma}^{1/2} = \mathbf{P}\mathbf{\Lambda}^{1/2}\mathbf{P}^\top$

Square root of an inverse
Positive definite ⇒ Positive eigenvalues ⇒ Inverse exists

$$\mathbf{P}\boldsymbol{\Lambda}^{-1/2}\mathbf{P}^\top \cdot \mathbf{P}\boldsymbol{\Lambda}^{-1/2}\mathbf{P}^\top = \mathbf{P}\boldsymbol{\Lambda}^{-1}\mathbf{P}^\top = \boldsymbol{\Sigma}^{-1},$$

so

$$\left(\boldsymbol{\Sigma}^{-1}\right)^{1/2} = \mathbf{P}\boldsymbol{\Lambda}^{-1/2}\mathbf{P}^\top.$$

It's easy to show

- $\left(\boldsymbol{\Sigma}^{-1}\right)^{1/2}$ is the inverse of $\boldsymbol{\Sigma}^{1/2}$
- Justifying the notation $\boldsymbol{\Sigma}^{-1/2}$

Now we can show $(\mathbf{X}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu}) \sim \chi^2(p)$

Where $\mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

$$
\begin{aligned}
\mathbf{Y} = \mathbf{X} - \boldsymbol{\mu} & \sim & N\left(\mathbf{0},\ \boldsymbol{\Sigma}\right) \\
\mathbf{Z} = \boldsymbol{\Sigma}^{-\frac{1}{2}}\mathbf{Y} & \sim & N\left(\mathbf{0}, \boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-\frac{1}{2}}\right) \\
& = & N\left(\mathbf{0}, \boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\Sigma}^{\frac{1}{2}}\ \boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{\Sigma}^{-\frac{1}{2}}\right) \\
& = & N\left(\mathbf{0}, \mathbf{I}\right)
\end{aligned}
$$

So $\mathbf{Z}$ is a vector of p independent standard normals, and

$$
\mathbf{Y}^\top \boldsymbol{\Sigma}^{-1}\mathbf{Y} = \mathbf{Z}^\top \mathbf{Z} = \sum_{j=1}^{p} Z_i^2 \sim \chi^2(p) \qquad \blacksquare
$$

$\overline{X}$ and S^2 independent

Let $X_1, \ldots, X_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$.

$$\mathbf{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \sim N\left(\mu\mathbf{1}, \sigma^2\mathbf{I}\right) \qquad \mathbf{Y} = \begin{pmatrix} X_1 - \overline{X} \\ \vdots \\ X_{n-1} - \overline{X} \\ \\ \overline{X} \end{pmatrix} = \mathbf{AX}$$

$\mathbf{Y} = \mathbf{AX}$

In more detail

$$\left(\begin{array}{ccccc} 1-\frac{1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} & -\frac{1}{n} \\ -\frac{1}{n} & 1-\frac{1}{n} & \cdots & -\frac{1}{n} & -\frac{1}{n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ -\frac{1}{n} & -\frac{1}{n} & \cdots & 1-\frac{1}{n} & -\frac{1}{n} \\ \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} & \frac{1}{n} \end{array}\right) \left(\begin{array}{c} X_1 \\ X_2 \\ \vdots \\ X_{n-1} \\ X_n \end{array}\right) = \left(\begin{array}{c} X_1 - \overline{X} \\ X_2 - \overline{X} \\ \vdots \\ X_{n-1} - \overline{X} \\ \overline{X} \end{array}\right)$$

The argument

$$\mathbf{Y} = \mathbf{A}\mathbf{X} = \left(\begin{array}{c} X_1 - \overline{X} \\ \vdots \\ X_{n-1} - \overline{X} \\ \\ \overline{X} \end{array} \right) = \left(\begin{array}{c} \mathbf{Y}_2 \\ \hline \\ \overline{X} \end{array} \right)$$

- $\mathbf{Y}$ is multivariate normal.
- $Cov\left(\overline{X}, (X_j - \overline{X})\right) = 0$ (Exercise)
- So $\overline{X}$ and $\mathbf{Y}_2$ are independent.
- So $\overline{X}$ and $S^2 = g(\mathbf{Y}_2)$ are independent. $\blacksquare$

Leads to the t distribution

If

- $Z \sim N(0,1)$ and
- $Y \sim \chi^2(\nu)$ and
- Z and Y are independent, then

$$T = \frac{Z}{\sqrt{Y/\nu}} \sim t(\nu)$$

Random sample from a normal distribution

Let $X_1, \ldots, X_n \stackrel{i.i.d.}{\sim} N(\mu, \sigma^2)$. Then

- $\frac{\sqrt{n}(\overline{X}-\mu)}{\sigma} = \frac{(\overline{X}-\mu)}{\sigma/\sqrt{n}} \sim N(0,1)$ and
- $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$ and
- These quantities are independent, so

$$
\begin{aligned}
T &= \frac{\sqrt{n}(\overline{X}-\mu)/\sigma}{\sqrt{\frac{(n-1)S^2}{\sigma^2}/(n-1)}} \\
&= \frac{\sqrt{n}(\overline{X}-\mu)}{S} \sim t(n-1)
\end{aligned}
$$

Multivariate normal likelihood

For reference

$$
\begin{aligned}
L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \prod_{i=1}^{n} \frac{1}{|\boldsymbol{\Sigma}|^{\frac{1}{2}} (2\pi)^{\frac{p}{2}}} \exp\left\{ -\frac{1}{2} (\mathbf{x}_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x}_i - \boldsymbol{\mu}) \right\} \\
&= |\boldsymbol{\Sigma}|^{-n/2} (2\pi)^{-np/2} \exp -\frac{n}{2} \left\{ tr(\widehat{\boldsymbol{\Sigma}} \boldsymbol{\Sigma}^{-1}) + (\overline{\mathbf{x}} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\overline{\mathbf{x}} - \boldsymbol{\mu}) \right\},
\end{aligned}
$$

where $\widehat{\boldsymbol{\Sigma}} = \frac{1}{n} \sum_{i=1}^{n} (\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})^\top$ is the sample variance-covariance matrix.

The Multivarite Delta Method

An application

The univariate delta method says that if $\sqrt{n}\,(T_n - \theta) \stackrel{d}{\rightarrow} T$, then $\sqrt{n}\,(g(T_n) - g(\theta)) \stackrel{d}{\rightarrow} g'(\theta)\,T$.

In the multivariate delta method, $\mathbf{T}_n$ and $\mathbf{T}$ are d-dimensional random vectors.

The function $g : \mathbb{R}^d \rightarrow \mathbb{R}^k$ is a vector of functions:

$$g(x_1, \ldots, x_d) = \left(\begin{array}{c} g_1(x_1, \ldots, x_d) \\ \vdots \\ g_k(x_1, \ldots, x_d) \end{array} \right)$$

$g'(\theta)$ is replaced by a matrix of partial derivatives (a Jacobian):

$$\dot{\mathrm{g}}(x_1, \ldots, x_d) = \left[\frac{\partial g_i}{\partial x_j} \right]_{k \times d} \text{ like } \left(\begin{array}{ccc} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} & \frac{\partial g_1}{\partial x_3} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} & \frac{\partial g_2}{\partial x_3} \end{array} \right).$$

The Delta Method

Univariate and multivariate

The univariate delta method says that if $\sqrt{n}\,(T_n - \theta) \xrightarrow{d} T$, then $\sqrt{n}\,(g(T_n) - g(\theta)) \xrightarrow{d} g'(\theta)\,T$.

The multivariate delta method says that if $\sqrt{n}(\mathbf{T}_n - \boldsymbol{\theta}) \xrightarrow{d} \mathbf{T}$, then $\sqrt{n}(g(\mathbf{T}_n) - g(\boldsymbol{\theta})) \xrightarrow{d} \dot{\mathrm{g}}(\boldsymbol{\theta})\mathbf{T}$,

where $\dot{\mathrm{g}}(x_1, \ldots, x_d) = \left[\frac{\partial g_i}{\partial x_j}\right]_{k \times d}$

In particular, if $\mathbf{T} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$, then

$$\sqrt{n}(g(\mathbf{T}_n) - g(\boldsymbol{\theta})) \xrightarrow{d} \mathbf{Y} \sim N(\mathbf{0}, \dot{\mathrm{g}}(\boldsymbol{\theta})\boldsymbol{\Sigma}\dot{\mathrm{g}}(\boldsymbol{\theta})^\top).$$

Testing a non-linear hypothesis

Consider the regression model $y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \epsilon_i$.

- There is a standard F-test for $H_0 : \mathbf{L}\boldsymbol{\beta} = \mathbf{h}$.
- So testing whether $\beta_1 = 0$ and $\beta_2 = 0$ is easy.
- But what about testing whether $\beta_1 = 0$ or $\beta_2 = 0$ (or both)?
- If $H_0 : \beta_1\beta_2 = 0$ is rejected, it means that *both* regression coefficients are non-zero.
- Can't test non-linear null hypotheses like this with standard tools.
- But if the sample size is large we can use the delta method.

The asymptotic distribution of $\widehat{\beta}_1\widehat{\beta}_2$

The multivariate delta method says that if $\sqrt{n}(\mathbf{T}_n - \boldsymbol{\theta}) \stackrel{d}{\rightarrow} \mathbf{T}$, then $\sqrt{n}(g(\mathbf{T}_n) - g(\boldsymbol{\theta})) \stackrel{d}{\rightarrow} \dot{\mathrm{g}}(\boldsymbol{\theta})\mathbf{T}$,

Know $\widehat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y} \sim N_p\left(\boldsymbol{\beta}, \sigma^2(\mathbf{X}^\top\mathbf{X})^{-1}\right)$.

So $\sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \stackrel{d}{\rightarrow} \mathbf{T} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = \lim_{n\rightarrow\infty} \sigma^2\left(\frac{1}{n}\mathbf{X}^\top\mathbf{X}\right)^{-1}$.

Let $g(\boldsymbol{\beta}) = \beta_1\beta_2$. Have

$$
\begin{aligned}
&= \sqrt{n}(g(\widehat{\boldsymbol{\beta}}_n) - g(\boldsymbol{\beta})) \\
&= \sqrt{n}(\widehat{\beta}_1\widehat{\beta}_2 - \beta_1\beta_2) \\
&\stackrel{d}{\rightarrow} \dot{\mathrm{g}}(\boldsymbol{\beta})\mathbf{T} \\
&= T \sim N(0, \dot{\mathrm{g}}(\boldsymbol{\beta})\boldsymbol{\Sigma}\dot{\mathrm{g}}(\boldsymbol{\beta})^\top)
\end{aligned}
$$

We will say $\widehat{\beta}_1\widehat{\beta}_2$ is asymptotically $N\left(\beta_1\beta_2, \frac{1}{n}\dot{\mathrm{g}}(\boldsymbol{\beta})\boldsymbol{\Sigma}\dot{\mathrm{g}}(\boldsymbol{\beta})^\top\right)$.

Need $\dot{\mathrm{g}}(\boldsymbol{\beta})$.

$\dot{\text{g}}(x_1, \ldots, x_d) = \left[\frac{\partial g_i}{\partial x_j}\right]_{k \times d}$

$g(\beta_0, \beta_1, \beta_2) = \beta_1\beta_2$ so $d = 3$ and $k = 1$.

$$\begin{aligned} \dot{\text{g}}(\beta_0, \beta_1, \beta_2) &= \left(\frac{\partial g}{\partial \beta_0}, \frac{\partial g}{\partial \beta_1}, \frac{\partial g}{\partial \beta_2}\right) \\ &= (0, \beta_2, \beta_1) \end{aligned}$$

So $\widehat{\beta}_1\widehat{\beta}_2 \stackrel{\cdot}{\sim} N\left(\beta_1\beta_2, \frac{1}{n}(0, \beta_2, \beta_1)\boldsymbol{\Sigma}\begin{pmatrix} 0 \\ \beta_2 \\ \beta_1 \end{pmatrix}\right)$.

Need the standard error

We have $\widehat{\beta}_1\widehat{\beta}_2 \stackrel{\cdot}{\sim} N\left(\beta_1\beta_2, \frac{1}{n}(0, \beta_2, \beta_1)\boldsymbol{\Sigma}\left(\begin{array}{c} 0 \\ \beta_2 \\ \beta_1 \end{array}\right)\right)$.

Denote the asymptotic variance by

$\frac{1}{n}(0, \beta_2, \beta_1)\boldsymbol{\Sigma}\left(\begin{array}{c} 0 \\ \beta_2 \\ \beta_1 \end{array}\right) = v.$

If we knew v we could compute $Z = \frac{\widehat{\beta}_1\widehat{\beta}_2 - \beta_1\beta_2}{\sqrt{v}}$

And use it in tests and confidence intervals.

Need to estimate v.

Standard error

Estimated standard deviation of $\widehat{\beta}_1\widehat{\beta}_2$

$$v = \frac{1}{n}(0, \beta_2, \beta_1)\mathbf{\Sigma} \begin{pmatrix} 0 \\ \beta_2 \\ \beta_1 \end{pmatrix}$$

where $\mathbf{\Sigma} = \lim_{n\to\infty} \sigma^2 \left(\frac{1}{n}\mathbf{X}^\top\mathbf{X}\right)^{-1}$.

Estimate β_1 and β_2 with $\widehat{\beta}_1$ and $\widehat{\beta}_2$

Estimate σ^2 with $MSE = \mathbf{e}^\top\mathbf{e}/(n-p)$.

Approximate $\frac{1}{n}\mathbf{\Sigma}$ with

$$\begin{aligned} MSE\frac{1}{n}\left(\frac{1}{n}\mathbf{X}^\top\mathbf{X}\right)^{-1} &= MSE\left(n\frac{1}{n}\mathbf{X}^\top\mathbf{X}\right)^{-1} \\ &= MSE\left(\mathbf{X}^\top\mathbf{X}\right)^{-1} \end{aligned}$$

$\widehat{v}$ approximates v

$$
\begin{aligned}
v &= \frac{1}{n}(0, \beta_2, \beta_1)\boldsymbol{\Sigma}\begin{pmatrix} 0 \\ \beta_2 \\ \beta_1 \end{pmatrix} \\
\widehat{v} &= MSE\,(0, \widehat{\beta}_2, \widehat{\beta}_1)\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\begin{pmatrix} 0 \\ \widehat{\beta}_2 \\ \widehat{\beta}_1 \end{pmatrix}
\end{aligned}
$$

Test statistic for $H_0 : \beta_1\beta_2 = 0$

$$Z = \frac{\widehat{\beta}_1\widehat{\beta}_2 - 0}{\sqrt{\widehat{v}}}$$

where

$$\widehat{v} = (0, \widehat{\beta}_2, \widehat{\beta}_1) MSE \left(\mathbf{X}^\top\mathbf{X}\right)^{-1} \begin{pmatrix} 0 \\ \widehat{\beta}_2 \\ \widehat{\beta}_1 \end{pmatrix}$$

Note $MSE\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}$ is produced by R's `vcov` function.

Copyright Information