

Random Vectors¹

STA442/2101 Fall 2013

¹See 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

- 1 Definitions and Basic Results
- 2 Multivariate Normal

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

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

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

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 $V(\mathbf{X})$, is defined as

$$V(\mathbf{X}) = E \{ (\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})' \}.$$

$$V(\mathbf{X}) = E \{ (\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})' \}$$

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

So, the covariance matrix $V(\mathbf{X})$ 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

$$C(\mathbf{X}, \mathbf{Y}) = E \{ (\mathbf{X} - \boldsymbol{\mu}_x)(\mathbf{Y} - \boldsymbol{\mu}_y)' \}.$$

Adding a constant has no effect

On variances and covariances

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

It's clear from the definitions:

- $V(\mathbf{X}) = E \{(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})'\}$
- $C(\mathbf{X}, \mathbf{Y}) = E \{(\mathbf{X} - \boldsymbol{\mu}_x)(\mathbf{Y} - \boldsymbol{\mu}_y)'\}$

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

Analogous to $Var(aX) = a^2 Var(X)$

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

$$\begin{aligned}V(\mathbf{AX}) &= E\{(\mathbf{AX} - \mathbf{A}\boldsymbol{\mu})(\mathbf{AX} - \mathbf{A}\boldsymbol{\mu})'\} \\&= E\{\mathbf{A}(\mathbf{X} - \boldsymbol{\mu})(\mathbf{A}(\mathbf{X} - \boldsymbol{\mu}))'\} \\&= E\{\mathbf{A}(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})'\mathbf{A}'\} \\&= \mathbf{A}E\{(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})'\}\mathbf{A}' \\&= \mathbf{A}V(\mathbf{X})\mathbf{A}' \\&= \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}'\end{aligned}$$

The Multivariate Normal Distribution

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

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

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

Σ positive definite

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

Analogies

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

- Univariate Normal

- $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right]$
- $E(X) = \mu, V(X) = \sigma^2$
- $\frac{(X-\mu)^2}{\sigma^2} \sim \chi^2(1)$

- Multivariate Normal

- $f(\mathbf{x}) = \frac{1}{|\Sigma|^{\frac{1}{2}}(2\pi)^{\frac{p}{2}}} \exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})\right]$
- $E(\mathbf{X}) = \boldsymbol{\mu}, V(\mathbf{X}) = \Sigma$
- $(\mathbf{X} - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{X} - \boldsymbol{\mu}) \sim \chi^2(p)$

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, $\mathbf{A}\mathbf{X} \sim N(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$
- 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 $\mathbf{X} = (X_1, X_2, X_3)'$ be multivariate normal with

$$\boldsymbol{\mu} = \begin{bmatrix} 1 \\ 0 \\ 6 \end{bmatrix} \text{ and } \boldsymbol{\Sigma} = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 4 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

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 $\mathbf{Y} = \mathbf{A}\mathbf{X}$

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}$$

$$\mathbf{Y} = \mathbf{A}\mathbf{X} \sim N(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$$

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)   # V(Y)
      [,1] [,2]
[1,]    8    5
[2,]    5    6

```

A couple of things to prove

- $(\mathbf{X} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\mu}) \sim \chi^2(p)$
- \bar{X} and S^2 independent

Recall the square root matrix

Covariance matrix Σ is real and symmetric matrix, so we have the spectral decomposition

$$\begin{aligned}\Sigma &= \mathbf{P}\Lambda\mathbf{P}' \\ &= \mathbf{P}\Lambda^{1/2}\Lambda^{1/2}\mathbf{P}' \\ &= \mathbf{P}\Lambda^{1/2}\mathbf{I}\Lambda^{1/2}\mathbf{P}' \\ &= \mathbf{P}\Lambda^{1/2}\mathbf{P}' \mathbf{P}\Lambda^{1/2}\mathbf{P}' \\ &= \Sigma^{1/2} \quad \Sigma^{1/2}\end{aligned}$$

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

Square root of an inverse

Positive definite \Rightarrow Positive eigenvalues \Rightarrow Inverse exists

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

so

$$(\mathbf{\Sigma}^{-1})^{1/2} = \mathbf{P}\mathbf{\Lambda}^{-1/2}\mathbf{P}'.$$

It's easy to show

- $(\mathbf{\Sigma}^{-1})^{1/2} = (\mathbf{\Sigma}^{1/2})^{-1}$
- Justifying the notation $\mathbf{\Sigma}^{-1/2}$

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

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

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

$$\mathbf{Y}' \boldsymbol{\Sigma}^{-1} \mathbf{Y} = \mathbf{Z}' \mathbf{Z} = \sum_{j=1}^p Z_j^2 \sim \chi^2(p) \quad \blacksquare$$

\bar{X} and S^2 independent

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

$$Y = AX$$

In more detail

$$\begin{pmatrix} 1 - \frac{1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} & -\frac{1}{n} \\ -\frac{1}{n} & 1 - \frac{1}{n} & \cdots & -\frac{1}{n} & -\frac{1}{n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ -\frac{1}{n} & -\frac{1}{n} & \cdots & 1 - \frac{1}{n} & -\frac{1}{n} \\ \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} & \frac{1}{n} \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_{n-1} \\ X_n \end{pmatrix} = \begin{pmatrix} X_1 - \bar{X} \\ X_2 - \bar{X} \\ \vdots \\ X_{n-1} - \bar{X} \\ \bar{X} \end{pmatrix}$$

The argument

$$\mathbf{Y} = \mathbf{A}\mathbf{X} = \begin{pmatrix} X_1 - \bar{X} \\ \vdots \\ X_{n-1} - \bar{X} \\ \bar{X} \end{pmatrix} = \begin{pmatrix} \mathbf{Y}_2 \\ \hline \bar{X} \end{pmatrix}$$

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

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

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

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

Multivariate normal likelihood

For reference

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

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

Copyright Information

This slide show was prepared by **Jerry Brunner**, Department of Statistics, University of Toronto. It is licensed under a **Creative Commons Attribution - ShareAlike 3.0 Unported License**. Use any part of it as you like and share the result freely. The \LaTeX source code is available from the course website:
<http://www.utstat.toronto.edu/~brunner/oldclass/appliedf13>