Random Vectors and Multivariate Normal[1]

STA 431 Spring 2023

[1] See last slide for copyright information.

Overview

Random Vectors and Matrices

See Section A.3 in Appendix A.

- 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 random 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

If the random matrices $\mathbf{X}$ and $\mathbf{Y}$ are the same size,

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

Moving a constant matrix 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

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

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

$$
\begin{aligned}
cov(\mathbf{x}) &= E\left\{\begin{pmatrix} x_1-\mu_1 \\ x_2-\mu_2 \\ x_3-\mu_3 \end{pmatrix}\begin{pmatrix} x_1-\mu_1 & x_2-\mu_2 & x_3-\mu_3 \end{pmatrix}\right\} \\
&= E\left\{\begin{pmatrix} (x_1-\mu_1)^2 & (x_1-\mu_1)(x_2-\mu_2) & (x_1-\mu_1)(x_3-\mu_3) \\ (x_2-\mu_2)(x_1-\mu_1) & (x_2-\mu_2)^2 & (x_2-\mu_2)(x_3-\mu_3) \\ (x_3-\mu_3)(x_1-\mu_1) & (x_3-\mu_3)(x_2-\mu_2) & (x_3-\mu_3)^2 \end{pmatrix}\right\} \\
&= \begin{pmatrix} E\{(x_1-\mu_1)^2\} & E\{(x_1-\mu_1)(x_2-\mu_2)\} & E\{(x_1-\mu_1)(x_3-\mu_3)\} \\ E\{(x_2-\mu_2)(x_1-\mu_1)\} & E\{(x_2-\mu_2)^2\} & E\{(x_2-\mu_2)(x_3-\mu_3)\} \\ E\{(x_3-\mu_3)(x_1-\mu_1)\} & E\{(x_3-\mu_3)(x_2-\mu_2)\} & E\{(x_3-\mu_3)^2\} \end{pmatrix} \\
&= \begin{pmatrix} Var(x_1) & Cov(x_1,x_2) & Cov(x_1,x_3) \\ Cov(x_1,x_2) & Var(x_2) & Cov(x_2,x_3) \\ Cov(x_1,x_3) & Cov(x_2,x_3) & Var(x_3) \end{pmatrix}.
\end{aligned}
$$

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.

Covariance matrix of a 1×1 random vector

That is, a scalar random variable

$$
\begin{aligned}
cov(\mathbf{x}) &= E\left\{(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^\top\right\} \\
&= E\left\{(x-\mu)(x-\mu)\right\} \\
&= E\left\{(x-\mu)^2\right\} \\
&= Var(x)
\end{aligned}
$$

A rule 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 $\mathbf{A}$ is an $r \times p$ matrix of constants. Then

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

Positive definite is a natural assumption

For covariance matrices

- Let $cov(\mathbf{x}) = \boldsymbol{\Sigma}$
- $\boldsymbol{\Sigma}$ positive definite means $\mathbf{a}^\top \boldsymbol{\Sigma} \mathbf{a} > 0$ for all $\mathbf{a} \neq \mathbf{0}$.
- $y = \mathbf{a}^\top \mathbf{x} = a_1 x_1 + \cdots + a_p x_p$ is a scalar random variable.
- $Var(y) = \mathbf{a}^\top cov(\mathbf{x}) \mathbf{a} = \mathbf{a}^\top \boldsymbol{\Sigma} \mathbf{a}$
- $\boldsymbol{\Sigma}$ positive definite just says that the variance of any (non-trivial) linear combination is positive.
- This is usually what you want.

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

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

Note $cov(\mathbf{x}, \mathbf{x}) = cov(\mathbf{x})$.

Adding a constant has no effect

On variances and covariances

It's clear from the definitions

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

That

- $cov(\mathbf{x}+\mathbf{a}) = cov(\mathbf{x})$
- $cov(\mathbf{x}+\mathbf{a},\mathbf{y}+\mathbf{b}) = cov(\mathbf{x},\mathbf{y})$

For example, $E(\mathbf{x}+\mathbf{a}) = \boldsymbol{\mu}+\mathbf{a}$, so

$$
\begin{aligned}
cov(\mathbf{x}+\mathbf{a}) &= E\left\{(\mathbf{x}+\mathbf{a}-(\boldsymbol{\mu}+\mathbf{a}))(\mathbf{x}+\mathbf{a}-(\boldsymbol{\mu}+\mathbf{a}))^\top\right\} \\
&= E\left\{(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^\top\right\} \\
&= cov(\mathbf{x})
\end{aligned}
$$

Here's a useful formula

Let $E(\mathbf{y}) = \boldsymbol{\mu}$, $cov(\mathbf{y}) = \boldsymbol{\Sigma}$, and let $\mathbf{A}$ and $\mathbf{B}$ be matrices of constants. Then

$$cov(\mathbf{Ay}, \mathbf{By}) = \mathbf{A}\boldsymbol{\Sigma}\mathbf{B}^\top.$$

Centering

Denote the *centered* version of the random vector $\mathbf{x}$ by $\overset{c}{\mathbf{x}} = \mathbf{x} - \boldsymbol{\mu}_x$, so that

- $E(\overset{c}{\mathbf{x}}) = \mathbf{0}$ and
- $E(\overset{c}{\mathbf{x}}\overset{c}{\mathbf{x}}^\top) = E\left\{(\mathbf{x} - \boldsymbol{\mu}_x)(\mathbf{x} - \boldsymbol{\mu}_x)^\top\right\} = cov(\mathbf{x})$ and
- $E(\overset{c}{\mathbf{x}}\overset{c}{\mathbf{y}}^\top) = E\left\{(\mathbf{x} - \boldsymbol{\mu}_x)(\mathbf{y} - \boldsymbol{\mu}_y)^\top\right\} = cov(\mathbf{x}, \mathbf{y})$

Linear combinations of random vectors

$$\begin{aligned}
\mathbf{L} &= \mathbf{A}_1\mathbf{x}_1 + \cdots + \mathbf{A}_m\mathbf{x}_m + \mathbf{b} \\
\overset{c}{\mathbf{L}} &= \mathbf{L} - E(\mathbf{L}) \\
&= \mathbf{A}_1\mathbf{x}_1 + \cdots + \mathbf{A}_m\mathbf{x}_m + \mathbf{b} \\
&\quad -\mathbf{A}_1\boldsymbol{\mu}_1 - \cdots - \mathbf{A}_m\boldsymbol{\mu}_m - \mathbf{b} \\
&= \mathbf{A}_1(\mathbf{x}_1 - \boldsymbol{\mu}_1) + \cdots + \mathbf{A}_m(\mathbf{x}_m - \boldsymbol{\mu}_m) \\
&= \mathbf{A}_1\overset{c}{\mathbf{x}}_1 + \cdots + \mathbf{A}_m\overset{c}{\mathbf{x}}_m
\end{aligned}$$

So that

$$\begin{aligned}
cov(\mathbf{L}) &= E(\overset{c}{\mathbf{L}}\overset{c}{\mathbf{L}}^\top) \\
cov(\mathbf{L}_1, \mathbf{L}_2) &= E(\overset{c}{\mathbf{L}}_1\overset{c}{\mathbf{L}}_2^\top)
\end{aligned}$$

$cov(\mathbf{L}_1, \mathbf{L}_2) = E(\overset{c}{\mathbf{L}}_1 \overset{c}{\mathbf{L}}_2^\top)$

Let

$$\begin{aligned} \overset{c}{\mathbf{L}}_1 &= \mathbf{A}_1 \overset{c}{\mathbf{x}}_1 + \cdots + \mathbf{A}_m \overset{c}{\mathbf{x}}_m \\ \overset{c}{\mathbf{L}}_2 &= \mathbf{B}_1 \overset{c}{\mathbf{y}}_1 + \cdots + \mathbf{B}_n \overset{c}{\mathbf{y}}_n \end{aligned}$$

A better rule for covariances of linear combinations

$$
\begin{aligned}
cov(\mathbf{L}_1, \mathbf{L}_2) &= E\left\{ \overset{c}{\mathbf{L}}_1 \overset{c}{\mathbf{L}}{}_2^\top \right\} \\
&= E\left\{ \left(\mathbf{A}_1 \overset{c}{\mathbf{x}}_1 + \cdots + \mathbf{A}_m \overset{c}{\mathbf{x}}_1 \right) \left(\mathbf{B}_1 \overset{c}{\mathbf{y}}_1 + \cdots + \mathbf{B}_n \overset{c}{\mathbf{y}}_n \right)^\top \right\} \\
&= E\left\{ \left(\mathbf{A}_1 \overset{c}{\mathbf{x}}_1 + \cdots + \mathbf{A}_m \overset{c}{\mathbf{x}}_1 \right) \left(\overset{c}{\mathbf{y}}{}_1^\top \mathbf{B}_1^\top + \cdots + \overset{c}{\mathbf{y}}{}_n^\top \mathbf{B}_n^\top \right) \right\} \\
&= E\left\{ \mathbf{A}_1 \overset{c}{\mathbf{x}}_1 \overset{c}{\mathbf{y}}{}_1^\top \mathbf{B}_1^\top + \mathbf{A}_1 \overset{c}{\mathbf{x}}_1 \overset{c}{\mathbf{y}}{}_2^\top \mathbf{B}_2^\top + \cdots + \mathbf{A}_m \overset{c}{\mathbf{x}}_m \overset{c}{\mathbf{y}}{}_n^\top \mathbf{B}_n^\top \right\} \\
&= \mathbf{A}_1 E\left\{ \overset{c}{\mathbf{x}}_1 \overset{c}{\mathbf{y}}{}_1^\top \right\} \mathbf{B}_1^\top + \mathbf{A}_1 E\left\{ \overset{c}{\mathbf{x}}_1 \overset{c}{\mathbf{y}}{}_2^\top \right\} \mathbf{B}_2^\top + \cdots + \mathbf{A}_m E\left\{ \overset{c}{\mathbf{x}}_m \overset{c}{\mathbf{y}}{}_n^\top \right\} \mathbf{B}_n^\top \\
&= \mathbf{A}_1\, cov(\mathbf{x}_1, \mathbf{y}_1)\, \mathbf{B}_1^\top + \mathbf{A}_1\, cov(\mathbf{x}_1, \mathbf{y}_2)\, \mathbf{B}_2^\top + \cdots + \mathbf{A}_m\, cov(\mathbf{x}_m, \mathbf{y}_n)\, \mathbf{B}_n^\top \\
&= \sum_{i=1}^{m} \sum_{j=1}^{n} \mathbf{A}_i\, cov(\mathbf{x}_i, \mathbf{y}_j)\, \mathbf{B}_j^\top
\end{aligned}
$$

That is, calculate the covariance of each term in $\mathbf{L}_1$ with each term in $\mathbf{L}_2$, and add them up.

Example: $cov(\mathbf{x}+\mathbf{y})$

$$\begin{aligned} cov(\mathbf{x}+\mathbf{y}) &= cov(\mathbf{x}+\mathbf{y}, \mathbf{x}+\mathbf{y}) \\ &= cov(\mathbf{x},\mathbf{x}) + cov(\mathbf{x},\mathbf{y}) + cov(\mathbf{y},\mathbf{x}) + cov(\mathbf{y},\mathbf{y}) \\ &= cov(\mathbf{x}) + cov(\mathbf{y}) + cov(\mathbf{x},\mathbf{y}) + cov(\mathbf{y},\mathbf{x}) \end{aligned}$$

- $cov(\mathbf{y},\mathbf{x}) \neq cov(\mathbf{x},\mathbf{y})$
- $cov(\mathbf{y},\mathbf{x}) = cov(\mathbf{x},\mathbf{y})^\top$

The Multivariate Normal Distribution

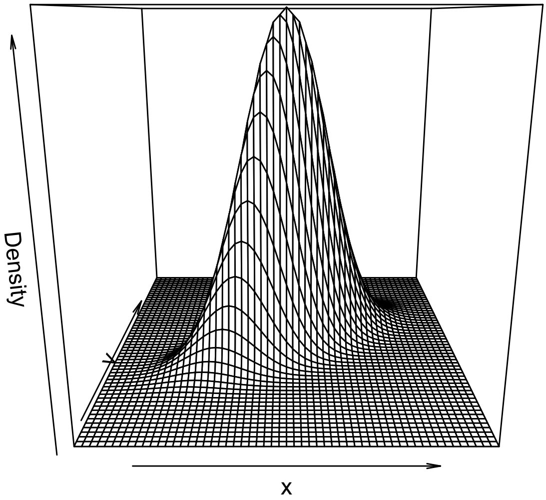
The $p \times 1$ random vector $\mathbf{x}$ is said to have a *multivariate normal distribution*, and we write $\mathbf{x} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, if $\mathbf{x}$ has (joint) density

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

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

The Bivariate Normal Density

Multivariate normal with $p = 2$ variables



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

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.
- For the multivariate normal, zero covariance implies independence. The multivariate normal is the only continuous distribution with this property.

Multivariate Normal Likelihood

$\mathbf{x}_1, \ldots, \mathbf{x}_n \overset{iid}{\sim} N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

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

where $\widehat{\boldsymbol{\Sigma}} = \frac{1}{n} \sum_{i=1}^{n} (\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})^\top$

Simulating from a multivariate norrmal

- Simulation of univariate normals is built-in. Use `rnorm()`.
- Say you want to simulate from $\mathbf{x} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.
- Generate $\mathbf{z} \sim N_p(\mathbf{0}, \mathbf{I})$.
- Calculate $\boldsymbol{\Sigma}^{\frac{1}{2}}$ using spectral decomposition.
- Let $\mathbf{x} = \boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{z} + \boldsymbol{\mu} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

The rmvn Function

```
> source("https://www.utstat.toronto.edu/~brunner/openSEM/fun/rmvn.txt")
> A = rbind(c(1.0,0.5),
+           c(0.5,1.0))
> A
     [,1] [,2]
[1,]  1.0  0.5
[2,]  0.5  1.0
> datta = rmvn(10,mu=c(0,0),sigma=A); datta
              [,1]        [,2]
 [1,] -2.643825316 -0.69926774
 [2,] -1.572814887 -0.21980248
 [3,] -0.387355643 -0.75080547
 [4,] -0.168534571 -1.28075830
 [5,] -0.716922363 -0.06556707
 [6,] -0.272368211 -0.15602646
 [7,] -0.007593983  0.59682941
 [8,]  0.436463462  1.02248006
 [9,] -0.193334362 -1.23877080
[10,] -0.859909183 -0.36091445
```

For the Record

```
#                 rmvn: Simulate from multivariate normal
rmvn <- function(nn,mu,sigma)
# Returns an nn by kk matrix, rows are independent MVN(mu,sigma)
    {
    kk <- length(mu)
    dsig <- dim(sigma)
    if(dsig[1] != dsig[2]) stop("Sigma must be square.")
    if(dsig[1] != kk) stop("Sizes of sigma and mu are inconsistent.")
    ev <- eigen(sigma)
    if(min(eigen(sigma)$values) < 0)
      stop("Sigma must have non-negative eigenvalues.")
    sqrl <- diag(sqrt(ev$values))
    PP <- ev$vectors
    ZZ <- rnorm(nn*kk) ; dim(ZZ) <- c(kk,nn)
    out <- t(PP%*%sqrl%*%ZZ+mu)
    return(out)
    }# End of function rmvn
```

Copyright Information

This slide show was prepared by Jerry Brunner, Department of Statistical Sciences, 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/431s23