# Multivariate methods

*Rebecca C. Steorts*

*September 11, 2015*

## Topics we will cover

- Vector, Mean, Covariance, and Cross-covariance
- Eigenvalue, eigenvector
- Trace
- Orthogonal matrix
- Symmetric matrix
- Univariate and multivariate normal distributions
- Taking linear functions of the normal
- Standardizing the normal
- Quadratic forms
- Theorems that are useful

## Vector, Mean, and Covariance Matrix

We first review what a vector is in terms a random variable. Define $\boldsymbol{X}_p$ to be a $p$-variate random vector,

$$\boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix}$$

where its entries $X_1, \ldots, X_p$ are random variables.

Just as in the univariate case, we can take the expectation of a vector. Let $A$ be a $m \times p$ matrix and $\boldsymbol{Y}$ be an $m$-variate random vector. Then

$$E(A\boldsymbol{X} + \boldsymbol{Y}) = AE(\boldsymbol{X}) + E(\boldsymbol{Y}).$$

Let $\boldsymbol{b}$ be a constant vector. Then

$$E(\boldsymbol{b}^T\boldsymbol{X}) = \boldsymbol{b}^T E(\boldsymbol{X}).$$

We now define what the covariance and cross-covariance of a vector $\boldsymbol{X}$ would be. Let $\boldsymbol{X}$ be a $p$-variate random vector. The covariance matrix of $\boldsymbol{X}$ is

$$\begin{aligned}\Sigma_{XX} = \text{Var}(\boldsymbol{X}) &= E\{(\boldsymbol{X} - \mu)^T(\boldsymbol{X} - \mu)\} \\ &= \begin{pmatrix} \text{Var}(X_1) & \text{Cov}(X_1, X_2) & \ldots & \text{Cov}(X_1, X_p) \\ \text{Cov}(X_2, X_1) & \text{Var}(X_2) & \ldots & \text{Cov}(X_2, X_p) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(X_p, X_1) & \text{Cov}(X_p, X_2) & \ldots & \text{Var}(X_p) \end{pmatrix}.\end{aligned}$$

We define the covariance matrix (cross covariance) between $\boldsymbol{X}$ and $\boldsymbol{Y}$ to be

$$\begin{aligned}\Sigma_{XY} &= \text{Cov}(X,Y) = E\{(\boldsymbol{X}-\mu_X)^T(\boldsymbol{Y}-\mu_Y)\} \\ &= \begin{pmatrix} \text{Cov}(X_1,Y_1) & \text{Cov}(X_1,Y_2) & \dots & \text{Cov}(X_1,Y_m) \\ \text{Cov}(X_2,Y_1) & \text{Cov}(X_2,Y_2) & \dots & \text{Cov}(X_2,X_m) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(X_p,Y_1) & \text{Cov}(X_p,Y_2) & \dots & \text{Cov}(X_p,Y_m) \end{pmatrix}.\end{aligned}$$

## Eigenvalues and Eigenvectors

Let $\boldsymbol{v} > 0$ and let $A$ be $d \times d$. Note: $\boldsymbol{v}$ is an eigenvector with eigenvalue $\lambda$ when

$$A\boldsymbol{v} = \lambda\boldsymbol{v}.$$

A few good facts to know (or have handy): Let $\boldsymbol{v} > 0$ and let $A$ be $d \times d$.

1. It's typical to normalize the eigenvector to have length 1 (or have its entries sum to 1).
2. $A$ has at most $p$ distinct eigenvalues (think about why).
3. Eigenvectors with distinct eigenvalues are orthogonal. (we will soon define orthogonal).

## Trace

Let $A = (a_{ij})$ be a square matrices of dimension $d \times d$.

1. The trace of $A$ is the sum of its diagonal elements:

$$\text{tr}(A) = \sum_i a_{ii}.$$

## Orthogonal matrix

A matrix $P$ is orthogonal if and only if $P^TP = 1$ and $PP^T = 1$.

## Symmetric matrix

A matrix $A$ is symmetric if and only if $A = A^T$.

## Univariate and multivariate normal distributions

Just as the probability density of a univariate normal is

$$p(x) = \left(2\pi\sigma^2\right)^{-1/2} \exp\left\{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right\}, \tag{1}$$

the probability density of the multivariate normal is

$$p(\vec{x}) = (2\pi)^{-p/2} \det\Sigma^{-1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}. \tag{2}$$

Univariate normal is special case of the multivariate normal with a one-dimensional mean "vector" and a one-by-one variance "matrix."

## Taking linear functions of the normal

A fact that is true of the normal (both univariate and multivariate), is that any linear function results in a normal distribution.

Suppose that $X \sim N(\mu, \sigma^2)$. Let $Y = aX + b$. Then $Y \sim N(a\mu + b, a^2\sigma^2)$.

Why? $E(Y) = a\mu + b$ and $Var(Y) = a^2\sigma^2$.

Can you think of what this would be for the multivariate normal? How does this change with matrices?

If $X$ were a multi-variate normal random vector, with mean vector $\vec{\mu}$ and a covariance matrix $\Sigma$ and $Y = AX + B$ where $A, B$ are $p \times 1$ and $p \times p$ respectively, then the $Y \sim Normal(B + A\vec{\mu}, A\Sigma A^T)$.

Why:

$E(Y) = A\mu + B$

$Var(Y) = V(AX + B) = V(AX) = A\Sigma A^T$

## Standardizing the normal

It's useful to "standardize a normal." This means either putting it in N(0,1) or N(0,I) form since it's easier to work with.

Suppose we have

$$X \sim N(\mu, \sigma^2).$$

Then we can do a simple transformation.

Let

$$Z = (X - \mu)/\sigma \sim N(0, 1).$$

This is easier to work with since - $Z^2 \sim \chi^2_1$ distribution (The form of the chi-squared density we won't work with a lot, but we will plot it some in R). - The important thing to know is that the distribution is called a chi-squared distribution and here it has one-degree of freedom (it could have more).

Now let

$$X \sim N_p(\mu, \Sigma)$$

(here we have a p-variate normal). Again, we do a transformation. Assume $\Sigma$ is non-singular. Let

$$Z = \Sigma^{-1/2}(X - \mu).$$

Then $Z \sim N(0, I)$. - An interesting fact is that $Z^2 \sim \chi^2_p$. (Think about why this might be true. Hint: You really have p-univaraite normals.)

## Quadratic forms

Let $A$ be a symmetric matrix and $\boldsymbol{x}$ a vector. A quadratic form is written as

$$\boldsymbol{x}^T A \boldsymbol{x} = \sum_i \sum_j a_{ij} x_i x_j.$$

Note: it's a quadratic function of $\boldsymbol{x}$.

We cover an example: Suppose $Z_1 \dots Z_p \overset{iid}{\sim} N(0, 1)$.

Then

$$||\boldsymbol{Z}||^2 = \sum_i Z_i^2 \sim \chi^2_p.$$

Now

$$Z_i \overset{ind}{\sim} N(\mu_i, 1), \ \forall i.$$

Then

$$\boldsymbol{Y}_{p\times 1} \sim N_p(\boldsymbol{\mu}, I) \implies \boldsymbol{Y}^T\boldsymbol{Y} \sim \chi_p^2(\frac{1}{2}\boldsymbol{\mu}^T\boldsymbol{\mu}).$$

This is called a non-central Chi-squared distribution.

### Useful theorems

Spectral Decomposition Theorem: Let $A_{p\times p}$ be symmetric with orthonormal eigenvectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p$ and corresponding eigenvalues $\lambda_1, \ldots, \lambda_p$. Then $A = P\Lambda P^T$, where $P = (\boldsymbol{v}_1 \ \cdots \ \boldsymbol{v}_p)$ and $\Lambda = \text{Diag}(\lambda_1, \ldots, \lambda_p)$.

The spectral decomposition theorem allows some operations with positive definite matrices to be computed more easily:

- $A^{-1} = P\Lambda^{-1}P^T$.
- $A^{1/2} = P\Lambda^{1/2}P^T$.

Alternative Spectral Decomposition Theorem: Let $A$ be symmetric $n \times n$. Then we can write

$$A = PDP^T,$$

where $D = diag(\lambda_1, \ldots, \lambda_n)$ and $P$ is orthogonal. The $\lambda$s are the eigenvalues of $A$ and ith column of $P$ is an eigenvector corresponding to $\lambda_i$.

Remark: In practice, it's often easier to work with the alternative formulation, however, both are useful.

Singular Value Decomposition

- Suppose that $X$ is data of size $d \times n$.
- The singular value decomposition (SVD) of $X$ is $X = UDV^T$, where $D$ is an $r \times r$ diagonal matrix.
- The matrices $U$ and $V$ have sizes $d \times r$ and $n \times r$, respectively.
- Their columns are the left and right eigenvectors of $X$. The left eigenvectors $\boldsymbol{u}_j$ and the right eigenvectors $\boldsymbol{v}_j$ of $X$ are unit vectors such that for all $j$,

$$X^T\boldsymbol{u}_j = (\boldsymbol{u}_j^T X)^T = d_j\boldsymbol{v}_j \qquad \text{and} \qquad X\boldsymbol{v}_j = d_j\boldsymbol{u}_j.$$

Used in PCA and dimension reduction methods (You will see these in machine learing if taking this course).

## Exploring data

We first look at some simple data that I just made up.

```r
x <- c(1,2,3,4)
A <- matrix(c(1,2,3,4),nrow=2)
B <- matrix(c(1,2,2,1),nrow=2)
length(x)
```

```
## [1] 4
```

```r
dim(A)
```

```
## [1] 2 2
```

```r
dim(B)
```

```
## [1] 2 2
```

```r
eigen(A)
```

```
## $values
## [1]  5.3722813 -0.3722813
##
## $vectors
##             [,1]       [,2]
## [1,] -0.5657675 -0.9093767
## [2,] -0.8245648  0.4159736
```

```r
eigen(B)
```

```
## $values
## [1]  3 -1
##
## $vectors
##           [,1]       [,2]
## [1,] 0.7071068 -0.7071068
## [2,] 0.7071068  0.7071068
```

```r
library(psych)
tr(A)
```

```
## [1] 5
```

```r
tr(B)
```

```
## [1] 2
```

```r
mean(x)
```

```
## [1] 2.5
```

```r
mean(A)
```

```
## [1] 2.5
```

```r
mean(B)
```

```
## [1] 1.5
```

```r
var(x)
```

```
## [1] 1.666667
```

```r
var(A)
```

```
##      [,1] [,2]
## [1,]  0.5  0.5
## [2,]  0.5  0.5
```

```r
mean(A * x)
```

```
## [1] 7.5
```

```r
var(A * x)
```

```
##      [,1] [,2]
## [1,]  4.5 10.5
## [2,] 10.5 24.5
```

```r
# Is A orthogonal?
t(A) %*% A
```

```
##      [,1] [,2]
## [1,]    5   11
## [2,]   11   25
```

```r
A %*% t(A)
```

```
##      [,1] [,2]
## [1,]   10   14
## [2,]   14   20
```

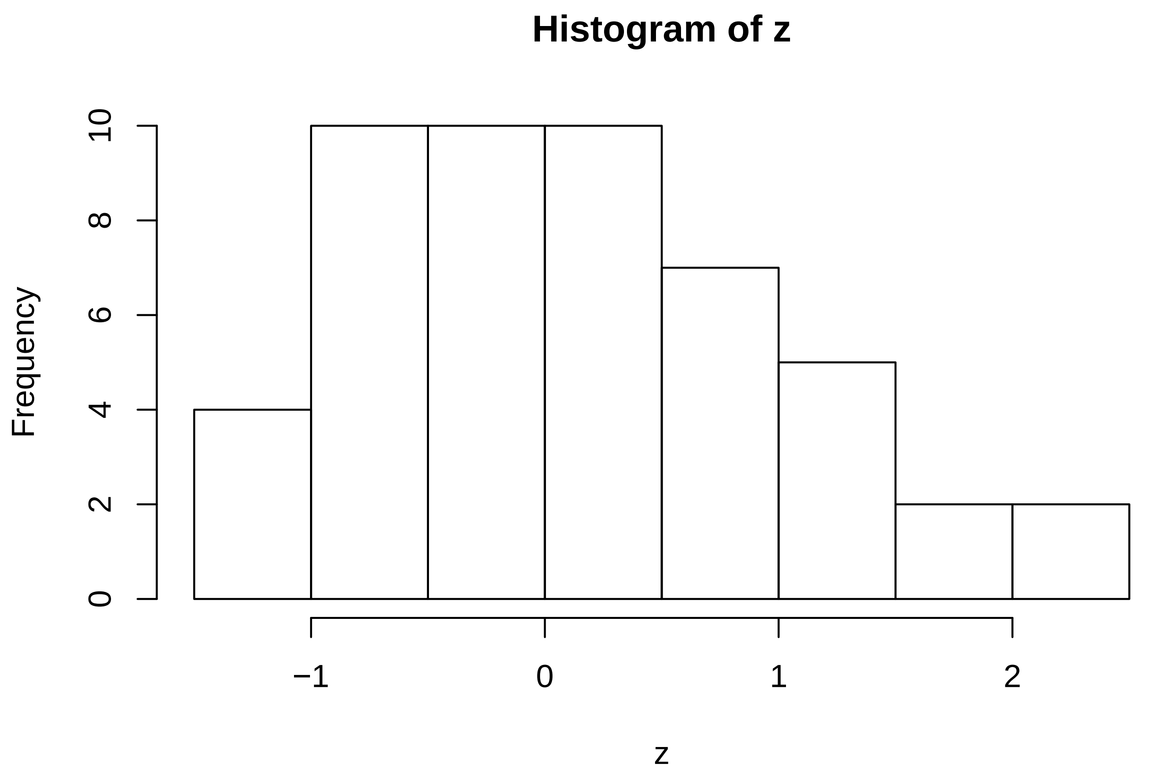
```r
# Is A symmetric? No!
A == t(A)
```

```
##       [,1]  [,2]
## [1,]  TRUE FALSE
## [2,] FALSE  TRUE
```

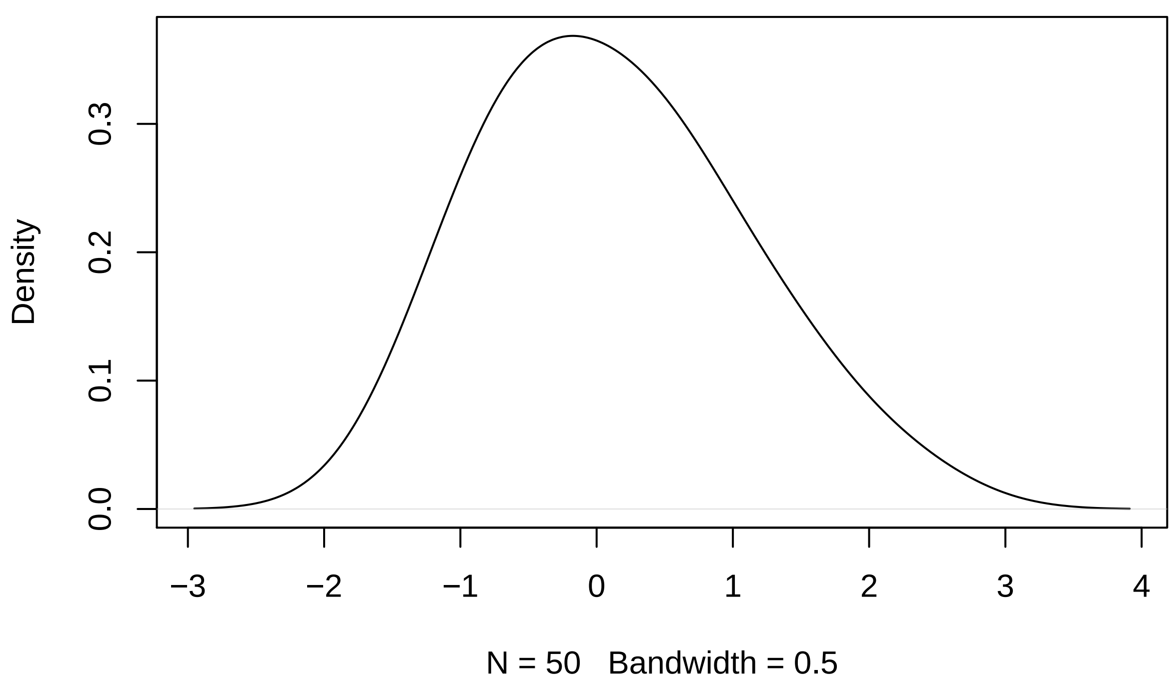
```r
# Is B symmetric? Yes!
B == t(B)
```

```
##      [,1] [,2]
## [1,] TRUE TRUE
## [2,] TRUE TRUE
```

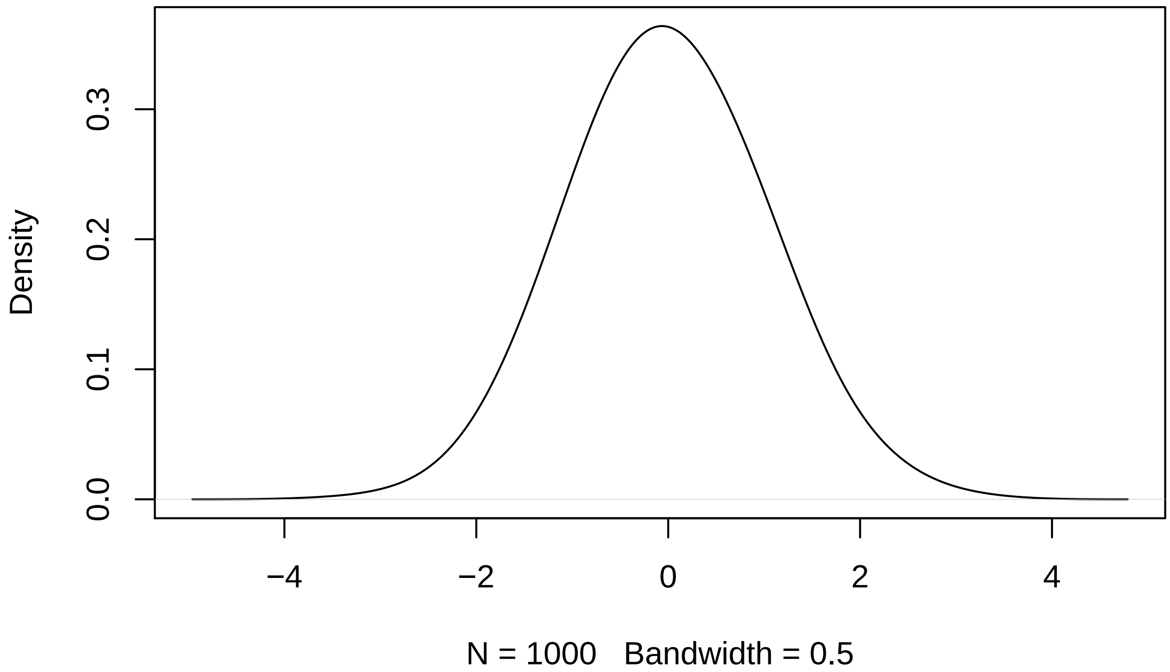
```r
## Let's generate a random univariate normal
## with 50 observations, mean 0 and variance 1
library(MASS)
z <- rnorm(50,0,1)
hist(z)
```




```r
plot(density(z, bw = 0.5), main="")
```



```r
# Let's up the sample size
z <- rnorm(1000,0,1)
plot(density(z, bw = 0.5), main="")
```



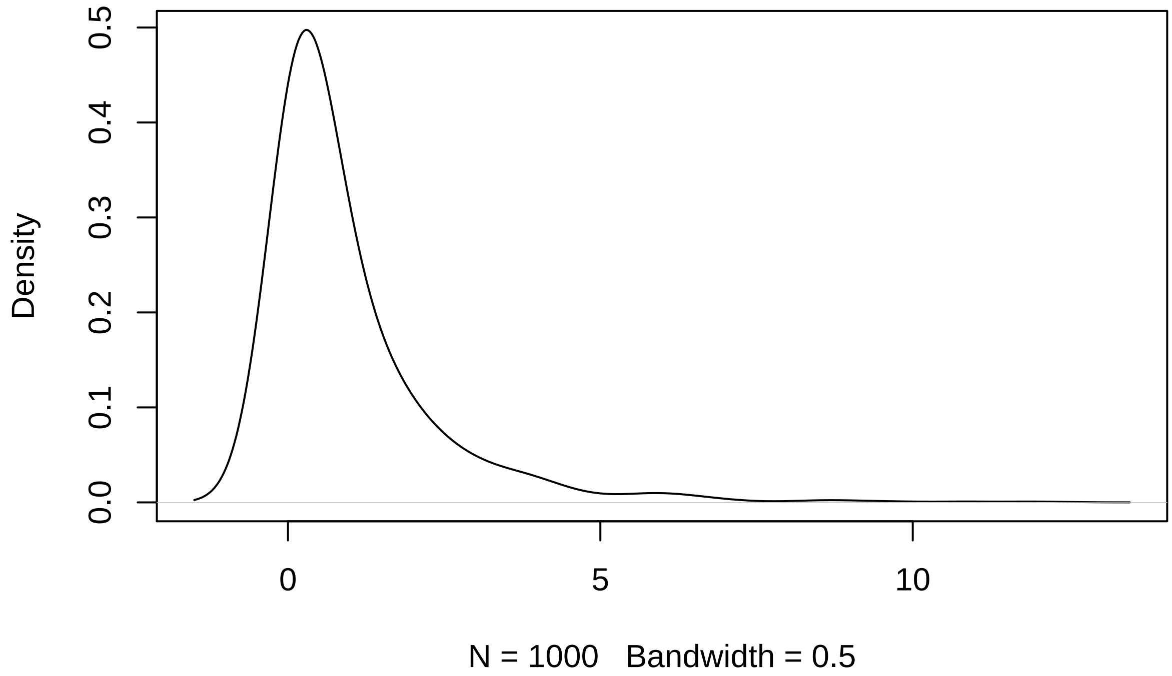

```r
chi_sq <- z^2
mean(chi_sq)
```

```
## [1] 0.9496707
```

```r
var(chi_sq)
```

```
## [1] 1.880129
```

```r
plot(density(chi_sq, bw = 0.5), main="")
```


Density
0.0
0.1
0.2
0.3
0.4
0.5
0
5
10
N = 1000   Bandwidth = 0.5