Eigenvalues and Eigenvectors

Suppose $A$ is an $n \times n$ symmetric matrix with real entries. The function from $R^n$ to $R$ defined by

$$x \mapsto x^t A x$$

is called a quadratic form.

We can maximize $x^T A x$ subject to $x^T x = ||x||^2 = 1$ by Lagrange multipliers:

$$x^T A x - \lambda (x^T x - 1)$$

Take derivatives and get

$$x^T x = 1$$

and

$$2A x - 2\lambda x = 0$$
We say that \( v \) is an eigenvector of \( A \) with eigenvalue \( \lambda \) if \( v \neq 0 \) and
\[
Av = \lambda v
\]
For such a \( v \) and \( \lambda \) with \( v^T v = 1 \) we find
\[
v^T Av = \lambda v^T v = \lambda.
\]
So the quadratic form is maximized over vectors of length one by the eigenvector with the largest eigenvalue.
Call that eigenvector \( v_1 \), eigenvalue \( \lambda_1 \).
Maximize \( x^T Ax \) subject to \( x^T x = 1 \) and \( v_1^T x = 0 \).
Get new eigenvector and eigenvalue.
Summary of Linear Algebra Results

Theorem
Suppose $A$ is a real symmetric $n \times n$ matrix.

1. There are $n$ orthonormal eigenvectors $v_1, \ldots, v_n$ with corresponding eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$.

2. If $P$ is the $n \times n$ matrix whose columns are $v_1, \ldots, v_n$ and $\Lambda$ is the diagonal matrix with $\lambda_1, \ldots, \lambda_n$ on the diagonal then

   \[ A P = P \Lambda \quad \text{or} \quad P^T \Lambda P = A \quad \text{and} \quad P^T A P = \Lambda \quad \text{and} \quad P^T P = I \]

3. If $A$ is non-negative definite (that is, $A$ is a variance covariance matrix) then each $\lambda_i \geq 0$.

4. $A$ is singular if and only if at least one eigenvalue is 0.

5. The determinant of $A$ is $\prod \lambda_i$. 

Richard Lockhart
STAT 350: General Theory
The trace of a matrix

**Definition:** If $A$ is square then the trace of $A$ is the sum of its diagonal elements:

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

**Theorem**

*If $A$ and $B$ are any two matrices such that $AB$ is square then*

$$\text{tr}(AB) = \text{tr}(BA)$$

*If $A_1,\ldots,A_r$ are matrices such that $\prod_{j=1}^r A_j$ is square then*

$$\text{tr}(A_1 \cdots A_r) = \text{tr}(A_2 \cdots A_r A_1) = \cdots = \text{tr}(A_s \cdots A_r A_1 \cdots A_{s-1})$$

*If $A$ is symmetric then*

$$\text{tr}(A) = \sum_i \lambda_i$$

Richard Lockhart
STAT 350: General Theory
Idempotent Matrices

**Definition:** A symmetric matrix $A$ is idempotent if $A^2 = AA = A$.

**Theorem**

A matrix $A$ is idempotent if and only if all its eigenvalues are either 0 or 1. The number of eigenvalues equal to 1 is then $\text{tr}(A)$.

**Proof:** If $A$ is idempotent, $\lambda$ is an eigenvalue and $v$ a corresponding eigenvector then

$$\lambda v = Av = AAv = \lambda Av = \lambda^2 v$$

Since $v \neq 0$ we find $\lambda - \lambda^2 = \lambda(1 - \lambda) = 0$ so either $\lambda = 0$ or $\lambda = 1$. 
Conversely

- Write
  \[ A = P \Lambda P^T \]
  so
  \[ A^2 = P \Lambda P^T P \Lambda P^T = P \Lambda^2 P^T \]
- Have used the fact that \( P \) is orthogonal.
- Each entry in the diagonal of \( \Lambda \) is either 0 or 1
- So \( \Lambda^2 = \Lambda \)
- So
  \[ A^2 = A.\]
Finally

$$\text{tr}(A) = \text{tr}(P \Lambda P^T)$$
$$= \text{tr}(\Lambda P^T P)$$
$$= \text{tr}(\Lambda)$$

Since all the diagonal entries in $\Lambda$ are 0 or 1 we are done the proof.
Independence

Definition: If $U_1, U_2, \ldots, U_k$ are random variables then we call $U_1, \ldots, U_k$ independent if

$$P(U_1 \in A_1, \ldots, U_k \in A_k) = P(U_1 \in A_1) \times \cdots \times P(U_k \in A_k)$$

for any sets $A_1, \ldots, A_k$.

We usually either:

- Assume independence because there is no physical way for the value of any of the random variables to influence any of the others.
- OR
- We prove independence.
Joint Densities

- How do we prove independence?
- We use the notion of a **joint density**.
- $U_1, \ldots, U_k$ have joint density function $f = f(u_1, \ldots, u_k)$ if

$$P((U_1, \ldots, U_k) \in A) = \int_A \cdots \int f(u_1, \ldots, u_k) du_1 \cdots du_k$$

- Independence of $U_1, \ldots, U_k$ is equivalent to

$$f(u_1, \ldots, u_k) = f_1(u_1) \times \cdots \times f_k(u_k)$$

for some densities $f_1, \ldots, f_k$.
- In this case $f_i$ is the density of $U_i$.
- **ASIDE:** notice that for an independent sample the joint density is the likelihood function!
Application to Normals: Standard Case

If

\[
Z = \begin{bmatrix} Z_1 \\ \vdots \\ Z_n \end{bmatrix} \sim MVN(0, I_{n \times n})
\]

then the joint density of \( Z \), denoted \( f_Z(z_1, \ldots, z_n) \) is

\[
f_Z(z_1, \ldots, z_n) = \phi(z_1) \times \cdots \times \phi(z_n)
\]

where

\[
\phi(z_i) = \frac{1}{\sqrt{2\pi}} e^{-z_i^2/2}
\]
So

\[ f_Z = (2\pi)^{-n/2} \exp \left\{ -\frac{1}{2} \sum_{i=1}^{n} z_i^2 \right\} \]

\[ = (2\pi)^{-n/2} \exp \left\{ -\frac{1}{2} z^T z \right\} \]

where

\[ z = \begin{bmatrix}
    z_1 \\
    \vdots \\
    z_n
\end{bmatrix} \]
Application to Normals: General Case

If $X = AZ + \mu$ and $A$ is invertible then for any set $B \in \mathbb{R}^n$ we have

$$P(X \in B) = P(AZ + \mu \in B)$$
$$= P(Z \in A^{-1}(B - \mu))$$
$$= \int_{A^{-1}(B-\mu)} \cdots \int (2\pi)^{-n/2} \exp \left\{ -\frac{1}{2} z^T z \right\} dz_1 \cdots dz_n$$

Make the change of variables $x = Az + \mu$ in this integral to get

$$P(X \in B) = \int_{B} \cdots \int (2\pi)^{-n/2}$$
$$\times \exp \left\{ -\frac{1}{2} (A^{-1}(x - \mu))^T (A^{-1}(x - \mu)) \right\} J(x) dx_1 \cdots dx_n$$
Here $J(x)$ denotes the Jacobian of the transformation

$$J(x) = J(x_1, \ldots, x_n) = \left| \det \left( \frac{\partial z_i}{\partial x_j} \right) \right| = |\det (A^{-1})|$$

Algebraic manipulation of the integral then gives

$$P(X \in B) = \int \cdots \int_B (2\pi)^{-n/2}$$

$$\times \exp \left\{ -\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right\} |\det A^{-1}| dx_1 \cdots dx_n$$

where

$$\Sigma = AA^T$$

$$\Sigma^{-1} = (A^{-1})^T (A^{-1})$$

$$\det \Sigma^{-1} = (\det A^{-1})^2$$

$$= \frac{1}{\det \Sigma}$$
Multivariate Normal Density

- Conclusion: the $MVN(\mu, \Sigma)$ density is

\[
(2\pi)^{-n/2} \exp \left\{ -\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right\} (\det \Sigma)^{-1/2}
\]

- What if $A$ is not invertible? Ans: there is no density.
- How do we apply this density?
- Suppose

\[
X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}
\]

and

\[
\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}
\]

- Now suppose $\Sigma_{12} = 0$
Assuming $\Sigma_{12} = 0$

1. $\Sigma_{21} = 0$

2. In homework you checked that

$$\Sigma^{-1} = \begin{bmatrix} \Sigma_{11}^{-1} & 0 \\ 0 & \Sigma_{22}^{-1} \end{bmatrix}$$

3. Writing

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

and

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$$

we find

$$(x - \mu)^T \Sigma^{-1} (x - \mu) = (x_1 - \mu_1)^T \Sigma_{11}^{-1} (x_1 - \mu_1)$$

$$+ (x_2 - \mu_2)^T \Sigma_{22}^{-1} (x_2 - \mu_2)$$
4. So, if \( n_1 = \dim(X_1) \) and \( n_2 = \dim(X_2) \) we see that

\[
f_X(x_1, x_2) = (2\pi)^{-n_1/2} \exp \left\{ -\frac{1}{2} (x_1 - \mu_1)^T \Sigma_{11}^{-1} (x_1 - \mu_1) \right\} \times (2\pi)^{-n_2/2} \exp \left\{ -\frac{1}{2} (x_2 - \mu_2)^T \Sigma_{22}^{-1} (x_2 - \mu_2) \right\}
\]

5. So \( X_1 \) and \( X_2 \) are independent.
Summary

- If \( \text{Cov}(X_1, X_2) = \mathbb{E}[(X_1 - \mu_1)(X_2 - \mu_2)^T] = 0 \) then \( X_1 \) is independent of \( X_2 \).

- **Warning**: This only works provided

  \[
  X = \begin{bmatrix}
  X_1 \\
  X_2
  \end{bmatrix} \sim \text{MVN}(\mu, \Sigma)
  \]

- **Fact**: However, it works even if \( \Sigma \) is singular, but you can't prove it as easily using densities.
Application: independence in linear models

\[ \hat{\mu} = X\hat{\beta} = X(X^TX)^{-1}X^TY \]
\[ = X\beta + H\epsilon \]
\[ \hat{\epsilon} = Y - X\hat{\beta} \]
\[ = \epsilon - H\epsilon \]
\[ = (I - H)\epsilon \]

So

\[
\begin{bmatrix}
\hat{\mu} \\
\hat{\epsilon}
\end{bmatrix} = \sigma \begin{bmatrix}
H \\
I - H
\end{bmatrix} \frac{\epsilon}{\sigma} + \begin{bmatrix}
\mu \\
0
\end{bmatrix} \]

Hence

\[
\begin{bmatrix}
\hat{\mu} \\
\hat{\epsilon}
\end{bmatrix} \sim MVN \left( \begin{bmatrix}
\mu \\
0
\end{bmatrix}; AA^T \right)
\]
Now

\[ A = \sigma \left[ \frac{H}{I - H} \right] \]

so

\[ AA^T = \sigma^2 \left[ \frac{H}{I - H} \right] \left[ H^T (I - H)^T \right] \]

\[ = \sigma^2 \left[ \begin{array}{cc} HH & H(I - H) \\ (I - H)H & (I - H)(I - H) \end{array} \right] \]

\[ = \sigma^2 \left[ \begin{array}{cc} H & H - H \\ H - H & I - H - H + HH \end{array} \right] \]

\[ = \sigma^2 \left[ \begin{array}{cc} H & 0 \\ 0 & I - H \end{array} \right] \]

The 0s prove that \( \hat{\epsilon} \) and \( \hat{\mu} \) are independent.

It follows that \( \hat{\mu}^T \hat{\mu} \), the regression sum of squares (not adjusted) is independent of \( \hat{\epsilon}^T \hat{\epsilon} \), the Error sum of squares.
Joint Densities: some manipulations

- Suppose $Z_1$ and $Z_2$ are independent standard normals.
- Their joint density is
  \[ f(z_1, z_2) = \frac{1}{2\pi} \exp\left(-\left(z_1^2 + z_2^2\right)/2\right). \]

- Show meaning of joint density by computing density of a $\chi^2_2$ random variable.
- Let $U = Z_1^2 + Z_2^2$.
- By definition $U$ has a $\chi^2$ distribution with 2 degrees of freedom.
Computing $\chi^2_2$ density

- Cumulative distribution function of $U$ is
  \[ F(u) = P(U \leq u). \]
- For $u \leq 0$ this is 0 so take $u \geq 0$.
- Event $U \leq u$ is same as event that point $(Z_1, Z_2)$ is in the circle centered at the origin and having radius $u^{1/2}$.
- That is, if $A$ is the circle of this radius then
  \[ F(u) = P((Z_1, Z_2) \in A). \]
- By definition of density this is a double integral
  \[ \int \int_A f(z_1, z_2) \, dz_1 \, dz_2. \]
- You do this integral in polar co-ordinates.
Integral in Polar co-ordinates

- Let $z_1 = r \cos \theta$ and $z_2 = r \sin \theta$.
- we see that
  \[
  f(r \cos \theta, r \sin \theta) = \frac{1}{2\pi} \exp(-r^2/2).
  \]
- The Jacobian of the transformation is $r$ so that $dz_1 \, dz_2$ becomes $r \, dr \, d\theta$.
- Finally the region of integration is simply $0 \leq \theta \leq 2\pi$ and $0 \leq r \leq u^{1/2}$ so that
  \[
  P(U \leq u) = \int_0^{u^{1/2}} \int_0^{2\pi} \frac{1}{2\pi} \exp(-r^2/2) r \, dr \, d\theta
  \]
  \[
  = \int_0^{u^{1/2}} r \exp(-r^2/2) dr
  \]
  \[
  = - \exp(-r^2/2) \bigg|_0^{u^{1/2}}
  \]
  \[
  = 1 - \exp(-u/2).
  \]
Density of $U$ found by differentiating to get

$$f(u) = \frac{1}{2} \exp(-u/2)$$

which is the exponential density with mean 2.

This means that the $\chi^2_2$ density is really an exponential density.
We have shown that \( \hat{\mu} \) and \( \hat{\epsilon} \) are independent.

So the Regression Sum of Squares (unadjusted) \( (=\hat{\mu}^T \hat{\mu}) \) and the Error Sum of Squares \( (=\hat{\epsilon}^T \hat{\epsilon}) \) are independent.

Similarly

\[
\left[ \begin{array}{c} \hat{\beta} \\ \hat{\epsilon} \end{array} \right] \sim MVN \left( \left[ \begin{array}{c} \beta \\ 0 \end{array} \right] ; \sigma^2 \left[ \begin{array}{cc} (X^TX)^{-1} & 0 \\ 0 & I - H \end{array} \right] \right)
\]

so that \( \hat{\beta} \) and \( \hat{\epsilon} \) are independent.
Conclusions

- We see
  \[ a^T \hat{\beta} - a^T \beta \sim N \left( 0, \sigma^2 a^t(X^TX)^{-1}a \right) \]
  is independent of
  \[ \hat{\sigma}^2 = \frac{\hat{\epsilon}^T \hat{\epsilon}}{n - p} \]

- If we know that
  \[ \frac{\hat{\epsilon}^T \hat{\epsilon}}{\sigma^2} \sim \chi^2_{n-p} \]
  then it would follow that
  \[ \frac{a^T \hat{\beta} - a^T \beta}{\sigma \sqrt{a^t(X^TX)^{-1}a}} \frac{\sqrt{\hat{\epsilon}^T \hat{\epsilon}/\{(n - p)\sigma^2\}}}{\sqrt{\text{MSE}a^t(X^TX)^{-1}a}} \sim t_{n-p} \]

- This leaves only the question: how do I know that
  \[ \hat{\epsilon}^T \hat{\epsilon}/\{\sigma^2\} \sim \chi^2_{n-p} \]
Distribution of the Error Sum of Squares

Recall: if $Z_1, \ldots, Z_n$ are iid $N(0, 1)$ then

$$U = Z_1^2 + \cdots + Z_n^2 \sim \chi_n^2$$

So we rewrite $\hat{\epsilon}^T \hat{\epsilon} / \{\sigma^2\}$ as $Z_1^2 + \cdots + Z_{n-p}^2$ for some $Z_1, \ldots, Z_{n-p}$ which are iid $N(0, 1)$.

Put

$$Z^* = \frac{\epsilon}{\sigma} \sim \text{MVN}_n(0, I_{n \times n})$$

Then

$$\frac{\hat{\epsilon}^T \hat{\epsilon}}{\sigma^2} = Z^*^T (I - H)(I - H)Z^* = Z^*^T (I - H)Z^*.$$ 

Now define new vector $Z$ from $Z^*$ so that

1. $Z \sim \text{MVN}(0, I)$
2. $Z^*^T (I - H)Z^* = \sum_{i=1}^{n-p} Z_i^2$
Theorem
If $Z$ has a standard $n$ dimensional multivariate normal distribution and $A$ is a symmetric $n \times n$ matrix then the distribution of $Z^T AZ$ is the same as that of
\[\sum \lambda_i Z_i^2\]
where the $\lambda_i$ are the $n$ eigenvalues of $Q$.

Theorem
The distribution in the last theorem is $\chi^2_\nu$ if and only if all the $\lambda_i$ are 0 or 1 and $\nu$ of them are 1.

Theorem
The distribution is chi-squared if and only if $A$ is idempotent. In this case $\text{tr}(A) = \nu$. 
Consider \((Z^*)^T AZ^*\) where \(A\) is symmetric matrix and \(Z^*\) is standard multivariate normal.

In earlier application \(A = I - H\).

Replace \(A\) by \(P \Lambda P^T\) in this formula

Get

\[
(Z^*)^T QZ^* = (Z^*)^T P \Lambda P^T Z^*
\]

\[
= (P^T Z^*)^T \Lambda (P^T Z^*)
\]

\[
= Z^T \Lambda Z
\]

where \(Z = P^T Z^*\).
Notice that $Z$ has a multivariate normal distribution
mean is 0 and variance is

$$\text{Var}(Z) = P^T P = I_{n \times n}$$

So $Z$ is also standard multivariate normal!

Now look at what happens when you multiply out

$$Z^T \Lambda Z$$

Multiplying a diagonal matrix by $Z$ simply multiplies the $i$th entry in $Z$ by the $i$th diagonal element

So

$$\Lambda Z = \begin{bmatrix} \lambda_1 Z_1 \\ \vdots \\ \lambda_n Z_n \end{bmatrix}$$
Take dot product of this with $Z$:

$$Z^T \Lambda Z = \sum \lambda_i Z_i^2.$$ 

Have rewritten our original quadratic form as a linear combination of squared independent standard normals, that is, as a linear combination of independent $\chi^2_1$ variables.
Application to Error Sum of Squares

- Recall that
  \[ \frac{\text{ESS}}{\sigma^2} = (Z^*)^T (I - H) Z^* \]
  where \( Z^* = \epsilon/\sigma \) is multivariate standard normal.

- The matrix \( I - H \) is idempotent

- So \( \text{ESS}/\sigma^2 \) has a \( \chi^2 \) distribution with degrees of freedom \( \nu \) equal to \( \text{trace}(I - H) \):

  \[
  \nu = \text{trace}(I - H) \\
  = \text{trace}(I) - \text{trace}(H) \\
  = n - \text{trace}(X(X^TX)^{-1}X^T) \\
  = n - \text{trace}((X^TX)^{-1}X^TX) \\
  = n - \text{trace}(I_{p \times p}) \\
  = n - p
  \]
Summary of Distribution theory conclusions

1. \( \epsilon^T A \epsilon / \sigma^2 \) has the same distribution as \( \sum \lambda_1 Z_i^2 \) where the \( Z_i \) are iid \( N(0, 1) \) random variables (so the \( Z_i^2 \) are iid \( \chi_1^2 \)) and the \( \lambda_i \) are the eigenvalues of \( A \).

2. \( A^2 = A \) (\( A \) is idempotent) implies that all the eigenvalues of \( A \) are either 0 or 1.

3. Points 1 and 2 prove that \( A^2 = A \) implies that \( \epsilon^T A \epsilon / \sigma^2 \sim \chi_{\text{trace}(A)}^2 \).

4. A special case is
   \[
   \frac{\hat{\epsilon}^T \hat{\epsilon}}{\sigma^2} \sim \chi_{n-p}^2
   \]

5. \( t \) statistics have \( t \) distributions.

6. If \( H_o : \beta = 0 \) is true then
   \[
   F = \frac{(\hat{\mu}^T \hat{\mu})/p}{\hat{\epsilon}^T \hat{\epsilon}/(n - p)} \sim F_{p, n-p}
   \]
Many Extensions are Possible

The most important of these are:

1. If a “reduced” model is obtained from a “full” model by imposing \( k \) linearly independent linear restrictions on \( \beta \) (like \( \beta_1 = \beta_2, \beta_1 + \beta_2 = 2\beta_3 \)) then

\[
\text{Extra SS} = \frac{\text{ESS}_R - \text{ESS}_F}{\sigma^2} \sim \chi^2_k
\]

assuming that the null hypothesis (the restricted model) is true.

2. So the Extra Sum of Squares \( F \) test has an \( F \)-distribution.

3. In ANOVA tables which add up the various rows (not including the total) are independent.

4. When null \( H_o \) is not true distribution of Regression SS is Non-central \( \chi^2 \).

5. Used in power and sample size calculations.