

Linear Models: Random vectors

The theory of linear algebra provides us with a good grounding to analyse our linear models. However we must still do some more groundwork. Once we have done this, the theoretical results come out quite easily!

Previously, we were thinking of matrices and vectors simply as a bunch of numbers. However, there is no reason why we can't think of them as a bunch of random variables!

We can then extend the traditional concepts of expectation, variance, etc. to random vectors.

Expectation

Although traditionally random variables are denoted with capital letters, in keeping with our linear algebra notation, we will denote them by lowercase.

Let

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{bmatrix}.$$

We define the expectation of \mathbf{y} to be the vector of expectations of its components:

$$E[\mathbf{y}] = \begin{bmatrix} E[y_1] \\ E[y_2] \\ \vdots \\ E[y_k] \end{bmatrix} .$$

Expectation properties

- If \mathbf{a} is a vector of constants, then $E[\mathbf{a}] = \mathbf{a}$.
- If \mathbf{a} is a vector of constants, then $E[\mathbf{a}^T \mathbf{y}] = \mathbf{a}^T E[\mathbf{y}]$.
- If A is a matrix of constants, then $E[A\mathbf{y}] = AE[\mathbf{y}]$.

None of these should be surprising as they parallel the scalar case more-or-less exactly.

Example. Let

$$A = \begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

and assume that $E[y_1] = 10$ and $E[y_2] = 20$. Then

$$AE[\mathbf{y}] = \begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} 10 \\ 20 \end{bmatrix} = \begin{bmatrix} 80 \\ 90 \end{bmatrix}.$$

On the other hand,

$$\begin{aligned} E[A\mathbf{y}] &= E \begin{bmatrix} 2y_1 + 3y_2 \\ y_1 + 4y_2 \end{bmatrix} \\ &= \begin{bmatrix} E[2y_1 + 3y_2] \\ E[y_1 + 4y_2] \end{bmatrix} \\ &= \begin{bmatrix} 2E[y_1] + 3E[y_2] \\ E[y_1] + 4E[y_2] \end{bmatrix} \\ &= \begin{bmatrix} 80 \\ 90 \end{bmatrix} = A E[\mathbf{y}]. \end{aligned}$$

Variance

Defining the variance of a random vector is slightly trickier. We want to not just include the variance of the variables themselves, but also how the variables affect each other.

Recall that the variance of a random variable Y with mean μ is defined to be $E[(Y - \mu)^2]$. Now let \mathbf{y} be as before, a $k \times 1$ vector of random variables. We define the variance of \mathbf{y} (sometimes known as the *covariance matrix*) to be

$$\text{var } \mathbf{y} = E[(\mathbf{y} - \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu})^T]$$

where $\boldsymbol{\mu} = E[\mathbf{y}]$.

The diagonal elements of the covariance matrix are just the variances of the individual elements of \mathbf{y} :

$$[\text{var } \mathbf{y}]_{ii} = \text{var } y_i, i = 1, 2, \dots, k.$$

The off-diagonal elements of the covariance matrix are the covariances of the individual elements:

$$[\text{var } \mathbf{y}]_{ij} = \text{cov}(y_i, y_j) = E[(y_i - \mu_i)(y_j - \mu_j)].$$

This means that all covariance matrices are symmetric.

Variance properties

Suppose that \mathbf{y} is a random vector with $\text{var } \mathbf{y} = V$. Then

- If \mathbf{a} is a vector of constants, then $\text{var } \mathbf{a}^T \mathbf{y} = \mathbf{a}^T V \mathbf{a}$.
- If A is a matrix of constants, then $\text{var } A\mathbf{y} = AVA^T$.

These can be derived from first principles quite easily.

Example. Let

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

be a random vector, such that $\text{var } y_i = \sigma^2$ for all i , and that the elements of \mathbf{y} are independent. This means that $\text{cov}(y_i, y_j) = 0$ for $i \neq j$. Then the covariance matrix of \mathbf{y} is

$$\text{var } \mathbf{y} = V = \begin{bmatrix} \sigma^2 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{bmatrix} = \sigma^2 I.$$

Assume that X is a matrix of full rank, which implies that $X^T X$ is nonsingular. Let

$$\mathbf{z} = (X^T X)^{-1} X^T \mathbf{y}.$$

Then

$$\begin{aligned} \text{var } \mathbf{z} &= A V A^T = [(X^T X)^{-1} X^T] \sigma^2 I [(X^T X)^{-1} X^T]^T \\ &= (X^T X)^{-1} X^T (X^T)^T [(X^T X)^{-1}]^T \sigma^2 \\ &= (X^T X)^{-1} X^T X [(X^T X)^T]^{-1} \sigma^2 \\ &= (X^T X)^{-1} \sigma^2. \end{aligned}$$

We will be using this quite a bit later on!

Random quadratic forms

Just as we can consider vectors and matrices to be composed of random variables, we can see what happens when these random vectors are combined into quadratic forms. The result is a function of random variables which is scalar (not vector), and so it is itself a random variable.

We now present some theorems on random quadratic forms.

Theorem

Let \mathbf{y} be a random vector with $E[\mathbf{y}] = \boldsymbol{\mu}$ and $\text{var } \mathbf{y} = V$, and let A be a matrix of constants. Then

$$E[\mathbf{y}^T A \mathbf{y}] = \text{tr}(AV) + \boldsymbol{\mu}^T A \boldsymbol{\mu}.$$

Proof. We denote the (i, j) th element of the matrix V by σ_{ij} . For $i \neq j$,

$$\sigma_{ij} = E[y_i y_j] - \mu_i \mu_j.$$

On the other hand,

$$\sigma_{ii} = E[y_i^2] - \mu_i^2.$$

We write the quadratic form out in summation form:

$$\begin{aligned}E[\mathbf{y}^T A \mathbf{y}] &= E \left[\sum_{i=1}^k \sum_{j=1}^k a_{ij} y_i y_j \right] \\&= \sum_{i=1}^k \sum_{j=1}^k a_{ij} E[y_i y_j] \\&= \sum_{i=1}^k \sum_{j=1}^k a_{ij} (\sigma_{ij} + \mu_i \mu_j) \\&= \sum_{i=1}^k \sum_{j=1}^k a_{ij} \sigma_{ji} + \sum_{i=1}^k \sum_{j=1}^k a_{ij} \mu_i \mu_j \\&= \sum_{i=1}^k [AV]_{ii} + \boldsymbol{\mu}^T A \boldsymbol{\mu} \\&= \text{tr}(AV) + \boldsymbol{\mu}^T A \boldsymbol{\mu}.\end{aligned}$$

Example. Let \mathbf{y} be a 2×1 random vector with

$$\boldsymbol{\mu} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}, V = \begin{bmatrix} 2 & 1 \\ 1 & 5 \end{bmatrix}.$$

Let

$$A = \begin{bmatrix} 4 & 1 \\ 1 & 2 \end{bmatrix}.$$

Consider the quadratic form

$$\mathbf{y}^T A \mathbf{y} = 4y_1^2 + 2y_1y_2 + 2y_2^2.$$

The expectation of this form is

$$E[\mathbf{y}^T \mathbf{A} \mathbf{y}] = 4E[y_1^2] + 2E[y_1 y_2] + E[y_2^2].$$

From the definition of variance and the given covariance matrix,

$$2 = \text{var } y_1 = E[y_1^2] - E[y_1]^2 = E[y_1^2] - 1$$

$$5 = \text{var } y_2 = E[y_2^2] - E[y_2]^2 = E[y_2^2] - 9$$

so $E[y_1^2] = 3$ and $E[y_2^2] = 14$.

From the definition of covariance and the given covariance matrix,

$$1 = \text{cov}(y_1, y_2) = E[y_1 y_2] - E[y_1]E[y_2] = E[y_1 y_2] - 3$$

so $E[y_1 y_2] = 4$. This gives

$$E[\mathbf{y}^T \mathbf{A} \mathbf{y}] = 4 \times 3 + 2 \times 4 + 14 = 48.$$

From the theorem,

$$\begin{aligned} E[\mathbf{y}^T A \mathbf{y}] &= \text{tr}(AV) + \boldsymbol{\mu}^T A \boldsymbol{\mu} \\ &= \text{tr} \left(\begin{bmatrix} 4 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 5 \end{bmatrix} \right) + [1 \ 3] \begin{bmatrix} 4 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \end{bmatrix} \\ &= \text{tr} \left(\begin{bmatrix} 9 & 9 \\ 4 & 11 \end{bmatrix} \right) + [1 \ 3] \begin{bmatrix} 7 \\ 7 \end{bmatrix} \\ &= 9 + 11 + 7 + 21 = 48. \end{aligned}$$

Prerequisites

Quadratic forms will pop up regularly in our analysis of linear models. To fully analyse our models, we will want to know the distribution of these forms, under the assumptions that we make on the distribution of the variables in the model.

I am assuming you have some familiarity with the:

- normal distribution;
- Student's- t distribution;
- χ^2 distribution;
- F distribution.

If not, don't worry too much (except for the normal distribution!).

Noncentral χ^2 distribution

The noncentral χ^2 distribution is derived from a normally distributed random vector — in other words, a random vector whose elements all have normal distributions. Furthermore, it has covariance matrix I . This means that the covariance of any two elements of \mathbf{y} is 0. Since they are normally distributed, that means that they are independent.

Definition

Let \mathbf{y} be a $k \times 1$ normally distributed random vector with mean $\boldsymbol{\mu}$ and variance I . Then $\mathbf{y}^T \mathbf{y}$ follows a *noncentral χ^2 distribution* with k degrees of freedom and noncentrality parameter $\lambda = \frac{1}{2} \boldsymbol{\mu}^T \boldsymbol{\mu}$. We denote such a random variable by $X_{k,\lambda}^2$.

Note that $\mathbf{y}^T \mathbf{y} = \sum_{i=1}^k y_i^2$ is the sum of squares of the elements of \mathbf{y} .

Example. Let y_1, y_2 and y_3 be independent, normally distributed random variables with means 4, 2, and -2 respectively, all with variance 1. Then \mathbf{y} is a normally distributed random vector with

$$\boldsymbol{\mu} = \begin{bmatrix} 4 \\ 2 \\ -2 \end{bmatrix}, \text{var } \mathbf{y} = I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Then $y_1^2 + y_2^2 + y_3^2$ follows a noncentral χ^2 distribution with 3 degrees of freedom and noncentrality parameter

$$\lambda = \frac{1}{2} \boldsymbol{\mu}^T \boldsymbol{\mu} = \frac{1}{2} [4 \quad 2 \quad -2] \begin{bmatrix} 4 \\ 2 \\ -2 \end{bmatrix} = 12.$$

Recall that an ordinary χ^2 distribution with k degrees of freedom can also be defined as the sum of the squares of k independent standard normal variables, z_1, z_2, \dots, z_k . Since standard normal variables have mean 0, and variance 1, this means that

$$\mathbf{z} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_k \end{bmatrix}$$

is a normally distributed random vector with mean $\mathbf{0}$ and variance I .

Therefore the sum of squares of its elements has a noncentral χ^2 distribution with k degrees of freedom and noncentrality parameter

$$\lambda = \frac{1}{2} \mathbf{0}^T \mathbf{0} = 0.$$

In other words, taking the noncentrality parameter to be 0 simply gives us the ordinary χ^2 distribution — hence the name.

Theorem

Let $X_{k_1, \lambda_1}^2, X_{k_2, \lambda_2}^2, \dots, X_{k_n, \lambda_n}^2$ be a collection of n independent noncentral χ^2 random variables, with k_1, k_2, \dots, k_n degrees of freedom respectively and noncentrality parameters $\lambda_1, \lambda_2, \dots, \lambda_n$ respectively. Then

$$\sum_{i=1}^n X_{k_i, \lambda_i}^2$$

has a noncentral χ^2 distribution with $\sum_{i=1}^n k_i$ degrees of freedom and noncentrality parameter $\sum_{i=1}^n \lambda_i$.

If we set $\lambda_i = 0$ for all i , we get the result that the sum of independent χ^2 variables is another χ^2 variable.

Distribution of quadratic forms

Now we can find out how certain quadratic forms are distributed, given certain assumptions.

Theorem

Let \mathbf{y} be a $n \times 1$ normally distributed random vector with mean $\boldsymbol{\mu}$ and variance I and let A be a $n \times n$ symmetric matrix. Then $\mathbf{y}^T A \mathbf{y}$ has a noncentral χ^2 distribution with k degrees of freedom and noncentrality parameter $\lambda = \frac{1}{2} \boldsymbol{\mu}^T A \boldsymbol{\mu}$ if and only if A is idempotent and has rank k .

Proof. (\Leftarrow) Assume that A is idempotent and has rank k . Because it is symmetric, it can be diagonalized. Let the (orthogonal) diagonalizing matrix be P , so that

$$P^T A P = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix},$$

where the λ 's are the eigenvalues of A . Since A is symmetric and idempotent, all the eigenvalues are either 0 or 1. However, we know that the sum of the eigenvalues is

$$\text{tr}(P^T A P) = \text{tr}(P P^T A) = \text{tr}(A) = r(A) = k.$$

Therefore, A must have k 1 eigenvalues and $n - k$ 0 eigenvalues.

Now arrange the columns of P so that all the 1 eigenvalues come first and the 0 eigenvalues come last. Then we can partition the diagonalization of A as

$$P^T A P = \left[\begin{array}{c|c} I & 0 \\ \hline 0 & 0 \end{array} \right],$$

where the I has dimension $k \times k$. Now define the random vector $\mathbf{z} = P^T \mathbf{y}$, so that $\mathbf{y} = P P^T \mathbf{y} = P \mathbf{z}$. Partition the vectors/matrices as

$$\mathbf{z} = \left[\begin{array}{c} \mathbf{z}_1 \\ \mathbf{z}_2 \end{array} \right], P = [P_1 \mid P_2]$$

where \mathbf{z}_1 is $k \times 1$ and P_1 is $n \times k$.

Then

$$\begin{aligned}\mathbf{y}^T A \mathbf{y} &= (P\mathbf{z})^T A (P\mathbf{z}) = \mathbf{z}^T P^T A P \mathbf{z} \\ &= \left[\mathbf{z}_1^T \mid \mathbf{z}_2^T \right] \left[\begin{array}{c|c} I & 0 \\ \hline 0 & 0 \end{array} \right] \begin{bmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \end{bmatrix} \\ &= \mathbf{z}_1^T \mathbf{z}_1.\end{aligned}$$

The elements of \mathbf{y} are independent, because \mathbf{y} has covariance matrix I . Since $\mathbf{z}_1 = P_1^T \mathbf{y}$, each element of \mathbf{z}_1 is a linear combination of independent normal variables — hence it is also normal.

$$E[\mathbf{z}_1] = E[P_1^T \mathbf{y}] = P_1^T E[\mathbf{y}] = P_1^T \boldsymbol{\mu}.$$

$$\text{var } \mathbf{z}_1 = \text{var} (P_1^T \mathbf{y}) = P_1^T I P_1 = P_1^T P_1.$$

But

$$\begin{aligned} P^T P &= \left[\begin{array}{c|c} P_1^T & P_2^T \\ \hline P_1 & P_2 \end{array} \right] \\ &= \left[\begin{array}{c|c} P_1^T P_1 & P_1^T P_2 \\ \hline P_1^T P_1 & P_2^T P_2 \end{array} \right] \\ &= \left[\begin{array}{c|c} I & 0 \\ \hline 0 & I \end{array} \right], \end{aligned}$$

so $P_1^T P_1 = I$.

So \mathbf{z}_1 is a $k \times 1$ normally distributed random vector with mean $P_1^T \boldsymbol{\mu}$ and variance I . Therefore, $\mathbf{y}^T A \mathbf{y} = \mathbf{z}_1^T \mathbf{z}_1$ has a noncentral χ^2 distribution with k degrees of freedom and noncentrality parameter

$$\lambda = \frac{1}{2} \boldsymbol{\mu}^T P_1 P_1^T \boldsymbol{\mu}.$$

Since

$$\begin{aligned} A &= [P_1 \mid P_2] \left[\begin{array}{c|c} I & 0 \\ \hline 0 & 0 \end{array} \right] \left[\begin{array}{c} P_1^T \\ \hline P_2^T \end{array} \right] \\ &= P_1 P_1^T, \end{aligned}$$

we have

$$\lambda = \frac{1}{2} \boldsymbol{\mu}^T A \boldsymbol{\mu}.$$

Corollary

Let \mathbf{y} be a $n \times 1$ normally distributed random vector with mean $\mathbf{0}$ and variance I and let A be a $n \times n$ symmetric matrix. Then $\mathbf{y}^T A \mathbf{y}$ has a (ordinary) χ^2 distribution with k degrees of freedom if and only if A is idempotent and has rank k .

Corollary

Let \mathbf{y} be a $n \times 1$ normally distributed random vector with mean $\boldsymbol{\mu}$ and variance $\sigma^2 I$ and let A be a $n \times n$ symmetric matrix. Then $\frac{1}{\sigma^2} \mathbf{y}^T A \mathbf{y}$ has a noncentral χ^2 distribution with k degrees of freedom and noncentrality parameter $\lambda = \frac{1}{2\sigma^2} \boldsymbol{\mu}^T A \boldsymbol{\mu}$ if and only if A is idempotent and has rank k .

Example. Let y_1 and y_2 be independent normal random variables with means 3 and -2 respectively and common variance 1. Let

$$A = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

It is easy to verify that A is symmetric and idempotent, and has rank 1. Therefore

$$\mathbf{y}^T \mathbf{A} \mathbf{y} = \frac{1}{2} \begin{bmatrix} y_1 & y_2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \frac{1}{2} y_1^2 + y_1 y_2 + \frac{1}{2} y_2^2$$

has a noncentral χ^2 distribution with 1 degree of freedom and noncentrality parameter

$$\lambda = \frac{1}{4} \begin{bmatrix} 3 & -2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ -2 \end{bmatrix} = \frac{1}{4}.$$

What happens if \mathbf{y} does not have variance I ? To analyse this case, we remind ourselves of the multivariate normal distribution.

Definition

Let \mathbf{y} be a normally distributed random vector with mean $\boldsymbol{\mu}$ and variance I and let C be a nonsingular matrix. Then

$$\mathbf{z} = C^T \mathbf{y}$$

is said to have a *multivariate normal distribution*, and we call \mathbf{z} a normal random vector.

Note that:

- Each element of \mathbf{z} is a linear combination of independent random variables, and so is normally distributed;
- From results for expectation and variance, $E[\mathbf{z}] = C^T \boldsymbol{\mu}$ and $\text{var } \mathbf{z} = C^T I C = C^T C$;
- \mathbf{y} itself is a multivariate normal random vector;
- The elements of \mathbf{z} are not independent;
- This is not the same as having a bunch of normal random variables!

Example. Let y_1 and y_2 be independent normal variables with means 2 and 5 respectively, and common variance 1. Then the following are normal random vectors:

- $\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$;
- $\begin{bmatrix} 4y_1 - y_2 \\ -2y_1 + 8y_2 \end{bmatrix}$.

However, $\begin{bmatrix} y_1 - 2y_2 \\ 2y_1 - 4y_2 \end{bmatrix}$ is not a normal random vector.

Theorem

Let \mathbf{y} be a $n \times 1$ normal random vector with mean $\boldsymbol{\mu}$ and variance V , and let A be a $n \times n$ symmetric matrix. Then $\mathbf{y}^T A \mathbf{y}$ has a noncentral χ^2 distribution with k degrees of freedom and noncentrality parameter $\lambda = \frac{1}{2} \boldsymbol{\mu}^T A \boldsymbol{\mu}$ if and only if AV is idempotent and has rank k .

Corollary

Let \mathbf{y} be a $n \times 1$ normal random vector with mean $\mathbf{0}$ and variance V and let A be a $n \times n$ symmetric matrix. Then $\mathbf{y}^T A \mathbf{y}$ has a (ordinary) χ^2 distribution with k degrees of freedom if and only if AV is idempotent and has rank k .

Corollary

Let \mathbf{y} be a $n \times 1$ normal random vector with mean $\boldsymbol{\mu}$ and variance V . Then $\mathbf{y}^T V^{-1} \mathbf{y}$ has a noncentral χ^2 distribution with n degrees of freedom and noncentrality parameter $\lambda = \frac{1}{2} \boldsymbol{\mu}^T V^{-1} \boldsymbol{\mu}$.

Example. Let y_1 and y_2 follow a multivariate normal distribution with means -1 and 4 respectively, and covariance matrix

$$V = \begin{bmatrix} 3 & 2 \\ 2 & 2 \end{bmatrix}. \text{ Then}$$

$$V^{-1} = \frac{1}{3 \times 2 - 2 \times 2} \begin{bmatrix} 2 & -2 \\ -2 & 3 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ -1 & 3/2 \end{bmatrix},$$

and the quadratic form

$$\mathbf{y}^T V^{-1} \mathbf{y} = y_1^2 - 2y_1y_2 + \frac{3}{2}y_2^2$$

has a noncentral χ^2 distribution with 2 degrees of freedom and noncentrality parameter

$$\lambda = \frac{1}{2} \begin{bmatrix} -1 & 4 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -1 & 3/2 \end{bmatrix} \begin{bmatrix} -1 \\ 4 \end{bmatrix} = \frac{33}{2}.$$

Independence of quadratic forms

Sometimes we will want to know when two quadratic forms are independent. The next theorem tells us when this happens.

Recall from linear algebra that if A_1, A_2, \dots, A_m is a collection of symmetric matrices, then there exists an orthogonal matrix P which diagonalizes all of the A_i if and only if $A_i A_j = A_j A_i$ for all i, j (i.e. the A matrices *commute*).

Theorem

Let \mathbf{y} be a $n \times 1$ normal random vector with mean $\boldsymbol{\mu}$ and variance V , and let A and B be symmetric $n \times n$ matrices. Then $\mathbf{y}^T A \mathbf{y}$ and $\mathbf{y}^T B \mathbf{y}$ are independent if and only if

$$AVB = 0.$$

Proof. (\Leftarrow) Suppose that $AVB = 0$. Since \mathbf{y} is a random normal vector, by definition there exists a nonsingular matrix C so that $V = C^T C$. Then $AC^T CB = 0$. Let

$$R = CAC^T, S = CBC^T.$$

Then

$$RS = CAC^T CBC^T = 0.$$

Because A and B are symmetric, R and S are symmetric as well. Then

$$SR = S^T R^T = (RS)^T = 0 = RS.$$

Therefore we can find an orthogonal matrix P which diagonalizes R and S simultaneously.

Since C is nonsingular, $r(R) = r(CAC^T) = r(A)$. Then

$$P^T R P = \left[\begin{array}{c|c} D_1 & 0 \\ \hline 0 & 0 \end{array} \right]$$

where D_1 has dimension $r(A) \times r(A)$. Because $RS = 0$, it can be shown that this means

$$P^T S P = \left[\begin{array}{c|c} 0 & 0 \\ \hline 0 & D_2 \end{array} \right]$$

where the partition has the same dimensions. Now define

$$\mathbf{z} = P^T (C^T)^{-1} \mathbf{y}.$$

Since P and C are nonsingular, \mathbf{z} is a random normal vector.

It is easy to show that $E[\mathbf{z}] = P^T(C^T)^{-1}\boldsymbol{\mu}$ and $\text{var } \mathbf{z} = I$. In particular, this means that the elements of \mathbf{z} are independent. Now partition \mathbf{z} into

$$\mathbf{z} = \begin{bmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \end{bmatrix}$$

where \mathbf{z}_1 has dimension $r(A) \times 1$. By rewriting our equations, we see that

$$\begin{aligned} \mathbf{y} &= C^T P \mathbf{z} \\ A &= C^{-1} R (C^T)^{-1} \\ B &= C^{-1} S (C^T)^{-1} \end{aligned}$$

So

$$\begin{aligned}\mathbf{y}^T \mathbf{A} \mathbf{y} &= \mathbf{z}^T \mathbf{P}^T \mathbf{C} \mathbf{C}^{-1} \mathbf{R} (\mathbf{C}^T)^{-1} \mathbf{C}^T \mathbf{P} \mathbf{z} \\ &= \mathbf{z}^T \mathbf{P}^T \mathbf{R} \mathbf{P} \mathbf{z} \\ &= \left[\mathbf{z}_1^T \mid \mathbf{z}_2^T \right] \left[\begin{array}{c|c} D_1 & 0 \\ \hline 0 & 0 \end{array} \right] \left[\begin{array}{c} \mathbf{z}_1 \\ \mathbf{z}_2 \end{array} \right] \\ &= \mathbf{z}_1^T D_1 \mathbf{z}_1\end{aligned}$$

and similarly

$$\mathbf{y}^T \mathbf{B} \mathbf{y} = \mathbf{z}_2^T D_2 \mathbf{z}_2.$$

But \mathbf{z}_1 and \mathbf{z}_2 are mutually independent of each other, since all elements of \mathbf{z} are independent. Therefore $\mathbf{y}^T \mathbf{A} \mathbf{y}$ and $\mathbf{y}^T \mathbf{B} \mathbf{y}$ are independent.

Example. Let y_1 and y_2 follow a multivariate normal distribution with covariance matrix

$$V = \begin{bmatrix} 1 & c \\ c & 1 \end{bmatrix}.$$

Consider the symmetric matrices

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, B = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

It is obvious that

$$\mathbf{y}^T A \mathbf{y} = y_1^2, \mathbf{y}^T B \mathbf{y} = y_2^2.$$

Now these quadratic forms will be independent if and only if

$$\begin{aligned} AVB &= \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & c \\ c & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & c \\ 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 0 & c \\ 0 & 0 \end{bmatrix} \end{aligned}$$

is the 0 matrix. But this happens if and only if $c = 0$, i.e. if y_1 and y_2 have zero covariance.

Corollary

Let \mathbf{y} be a random normal vector with mean $\boldsymbol{\mu}$ and variance $\sigma^2 I$, and let A and B be symmetric matrices. Then $\mathbf{y}^T A \mathbf{y}$ and $\mathbf{y}^T B \mathbf{y}$ are independent if and only if $AB = 0$.

Next we consider when a quadratic form is independent of a random vector. Firstly, we define a random variable to be independent of a random vector if and only if it is independent of all elements of that vector.

Theorem

Let \mathbf{y} be a $n \times 1$ normal random vector with mean $\boldsymbol{\mu}$ and variance V , and let A be a $n \times n$ symmetric matrix and B a $m \times n$ matrix. Then $\mathbf{y}^T A \mathbf{y}$ and $B \mathbf{y}$ are independent if and only if $BVA = 0$.

Lastly, we can combine several of the theorems we have seen before to tell when a group of quadratic forms (more than two) are independent.

Theorem

Let \mathbf{y} be a normal random vector with mean $\boldsymbol{\mu}$ and variance I , and let A_1, A_2, \dots, A_m be a collection of m symmetric matrices. If any two of the following statements are true:

- All A_i are idempotent;
- $\sum_{i=1}^m A_i$ is idempotent;
- $A_i A_j = 0$ for all $i \neq j$;

then:

- For all i , $\mathbf{y}^T A_i \mathbf{y}$ has a noncentral χ^2 distribution with $r(A_i)$ degrees of freedom and noncentrality parameter $\lambda_i = \frac{1}{2} \boldsymbol{\mu}^T A_i \boldsymbol{\mu}$;
- $\mathbf{y}^T A_i \mathbf{y}$ and $\mathbf{y}^T A_j \mathbf{y}$ are independent for $i \neq j$; and
- $\sum_{i=1}^m r(A_i) = r(\sum_{i=1}^m A_i)$.

Many of you will be familiar with basic analysis of variance (ANOVA). In this method, we separate the sum of squares of the response variable into a sum of quadratic forms:

$$\mathbf{y}^T \mathbf{y} = \mathbf{y}^T A_1 \mathbf{y} + \mathbf{y}^T A_2 \mathbf{y} + \dots + \mathbf{y}^T A_m \mathbf{y}.$$

Since $\sum_{i=1}^m A_i = I$, it is idempotent. Then if we can establish one of the other two conditions given in the theorem (most commonly that A_i is idempotent), we can use the theorem to determine the distribution of the quadratic forms, and also conclude their independence.