STAT 350: Lecture 17

Reading: There is no truly relevant part of the text except chapter 5.

Summary of Last Time:

Suppose

$$X = \left[\frac{X_1}{X_2}\right] \sim MVN\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}\right)$$

Then

1. The covariance between X_1 and X_2 is

$$\Sigma_{12} = \mathrm{E}[(X_1 - \mu_1)(X_2 - \mu_2)^T] \equiv \mathrm{Cov}(X_1, X_2)$$

2. $\Sigma_{12} = 0 \Leftrightarrow X_1$ independent of X_2

3. In the regression model $Y = X\beta + \epsilon$ with $\epsilon \sim MVN(0, \sigma^2 I)$

$$\left[\frac{\hat\mu}{\hat\epsilon}\right] \sim MVN\left(\left[\frac{\mu}{0}\right]; \sigma^2 \begin{bmatrix} H & 0 \\ 0 & I - H \end{bmatrix}\right)$$

so that $\hat\mu$ and $\hat\epsilon$ are independent.

4. It follows that the Regression Sum of Squares (unadjusted) (=$\hat\mu\hat\mu$) and the Error Sum of Squares (=$\hat\epsilon^T\hat\epsilon$) are independent.

5. Similarly

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

so that $\hat\beta$ and $\hat\epsilon$ are independent.

Conclusion:

$$a^T\hat\beta - a^T\beta \sim N(0, \sigma^2 a^t (X^TX)^{-1} a)$$

is independent of

$$\hat\sigma^2 = \frac{\hat\epsilon^T\hat\epsilon}{n-p}$$

If we know that

$$\frac{\hat\epsilon^T\hat\epsilon}{n-p} \sim \chi^2_{n-p}$$

then it would follow that

$$\frac{\frac{a^T\hat\beta - a^T\beta}{\sigma\sqrt{a^t(X^TX)^{-1}a}}}{\sqrt{\frac{\hat\epsilon^T\hat\epsilon}{n-p}/(n-p)}} = \frac{a^T(\hat\beta-\beta)}{\sqrt{\text{MSE}a^t(X^TX)^{-1}a}} \sim t_{n-p}$$

This leaves only the question: how do I know that

$$\frac{\hat\epsilon^T\hat\epsilon}{n-p} \sim \chi^2_{n-p}$$

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 try to rewrite $\hat\epsilon^T\hat\epsilon/(n-p)$ as $Z_1^2 + \cdots + Z_n^2$ for some $Z_1, \ldots, Z_{n-p}$ which are iid $N(0,1)$. Here is how:
Put:

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

Then

$$\begin{aligned}\frac{\hat\epsilon^T\hat\epsilon}{\sigma^2} &= Z^{*T}(I-H)(I-H)Z^* \\ & Z^{*T}(I-H)Z^*\end{aligned}$$

We are now going to define a new vector Z from Z^* in such a way that

1. $Z \sim MVN(0, I)$
2. $Z^{*T}(I-H)^Z * = \sum_{i=1}^{n-p} Z_i^2$

We use Eigenvalues and Eigenvectors to do so.

Eigenvalues, Eigenvectors, Diagonalization and Quadratic Forms

Linear Algebra theorem: If Q is an $n \times n$ symmetric (real) matrix then there are scalars $\lambda_1, \ldots, \lambda_n$ and vectors $v_1, \ldots, v_n$ such that

1. $Qv_i = \lambda_i v_i$ for each i. We call λ_i an **eigenvalue** and v_i a corresponding **eigenvector**.

2. $v_i^T v_j = 0$ for $i \neq j$. We say that the vectors v_i and v_j are **orthogonal**.

3. $v_i^T v_1 = 1$. We say that v_i is **normalized**.

Now make a matrix $\mathbf{P}$ by putting the n vectors $v_1, \ldots, v_n$ into the columns of $\mathbf{P}$. Then $\mathbf{P}$ is an $n \times n$ matrix. Next we compute $\mathbf{P}^T\mathbf{P}$:

$$\begin{aligned}\mathbf{P}^T\mathbf{P} &= \begin{bmatrix} v_1^T \\ \vdots \\ v_n^T \end{bmatrix} \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix} \\ &= \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix} \\ &= I_{n\times n}\end{aligned}$$

Thus $\mathbf{P}^{-1} = \mathbf{P}^T$; $\mathbf{P}$ is a matrix whose inverse is just its transpose. I remark that this **proves** that $\mathbf{P}\mathbf{P}^T = I$ so that the rows of $\mathbf{P}$ are orthonormal, just like the columns.

Now let Λ be the diagonal matrix whose entries along the diagonal are $\lambda_1, \ldots, \lambda_n$. Then multiplying

$$\mathbf{P}\Lambda = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix} \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix} = \begin{bmatrix} \lambda_1 v_1 & \cdots & \lambda_n v_n \end{bmatrix}$$

On the other hand

$$Q\mathbf{P} = \begin{bmatrix} Qv_1 & \cdots & Qv_n \end{bmatrix} = \begin{bmatrix} \lambda_1 v_1 & \cdots & \lambda_n v_n \end{bmatrix}$$

so

$$Q\mathbf{P} = \mathbf{P}\Lambda$$

Multiply this equation on the right by $\mathbf{P}^T$ to conclude that

$$Q = \mathbf{P}\Lambda\mathbf{P}^T$$

or on the left by $\mathbf{P}^T$ to conclude that

$$\mathbf{P}^T Q\mathbf{P} = \Lambda$$

Rewriting a Quadratic Form as a Sum of Squares

Recall that we are studying $(Z^*)^T Q Z^*$ where Q is the matrix $I - H$ and Z^* is standard multivariate normal. Replace Q by $\mathbf{P}\Lambda\mathbf{P}^T$ in this formula to get

$$\begin{aligned}(Z^*)^T Q Z^* &= (Z^*)^T \mathbf{P}\Lambda\mathbf{P}^T Z^* \\ &= (\mathbf{P}^T Z^*)^T \Lambda (\mathbf{P}^T Z^*) \\ &= Z^T \Lambda Z\end{aligned}$$

where $Z = \mathbf{P}^T Z^*$. Notice that Z has a multivariate normal distribution whose mean is obviously 0 and whose variance is

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

In other words 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 ith entry in Z by the ith diagonal element so

$$\Lambda Z = \begin{bmatrix} \lambda_1 Z_1 \\ \vdots \\ \lambda_n Z_n \end{bmatrix}$$

Taking the dot product of this with Z we see that

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

We have rewritten our original quadratic form as a linear combination of squared independent standard normals, that is, as a linear combination of independent χ_1^2 variables. This is the first big result:

Theorem: If Z has a standard n dimensional multivariate normal distribution and Q is a symmetric $n \times n$ matrix then the distribution of $Z^T Q Z$ is the same as that of

$$\sum \lambda_i Z_i^2$$

where the λ_i are the n eigenvalues of Q.

Now we turn to the conditions under which this linear combination of χ_1^2 variables actually has a χ_ν^2 distribution and how to find ν when it does. The point is that $\sum \lambda_i Z_i^2$ would have a χ_ν^2 distribution if the set of eigenvalues λ_i consisted of ν 1s and all the rest were 0. How can we tell if an eigenvalue is 1 or 0?

Suppose that each eigenvector v_i has an eigenvalue λ_i which is either 0 or 1. Then notice that

$$QQv_i = Q(\lambda_i v_i) = \lambda_i Q v_i = \lambda_i^2 v_i$$

But $0^2 = 0$ and $1^2 = 1$ so $\lambda_i^2 = \lambda_i$. We then learn that

$$Q^2 v_i = \lambda_i v_i = Q v_i$$

or

$$(Q^2 - Q)v_i = 0$$

for all i from 1 to n. Since the $v_1, \ldots, v_n$ are a basis of R_n we have proved that

$$(Q^2 - Q)x = 0$$

for **every** $x \in R^n$. This guarantees that $Q^2 = Q$. Conversely suppose that Q is a symmetric matrix such that $Q^2 = Q$, i.e. Q is idempotent. Then the algebra above shows that

$$\lambda_i v_i = Q v_i = Q^2 v_i = \lambda_i^2 v_i$$

so that

$$\lambda_i (1 - \lambda_i) v_i = 0$$

for all i. The eigenvectors v_i are not 0 so either $\lambda_i = 0$ or $1 - \lambda_i = 0$ and $\lambda_i = 1$.

Theorem: The eigenvalues of a symmetric matrix Q are all either 0 or 1 if and only if Q is idempotent.

We have thus learned that $Z^T Q Z$ has a χ^2 distribution provided that Q is idempotent. How can we count the degrees of freedom? The degrees of freedom ν is just the number of eigenvalues equal to 1. For a list of zeros and ones the number of ones is just the sum of the list. That is

$$\nu = \sum \lambda_i = \text{trace}(\Lambda)$$

Finally, remember the properties of the trace and get

$$\begin{aligned}
\operatorname{trace}(\Lambda) &= \operatorname{trace}(\mathbf{P}^T Q \mathbf{P}) \\
&= \operatorname{trace}(Q\mathbf{P}\mathbf{P}^T) \\
&= \operatorname{trace}(QI) \\
&= \operatorname{trace}(Q)
\end{aligned}$$

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 ESS/σ^2 has a χ^2 distribution with degrees of freedom ν equal to $\operatorname{trace}(I-H)$:

$$\begin{aligned}
\nu &= \operatorname{trace}(I-H) \\
&= \operatorname{trace}(I) - \operatorname{trace}(H) \\
&= n - \operatorname{trace}(X(X^TX)^{-1}X^T) \\
&= n - \operatorname{trace}((X^TX)^{-1}X^TX) \\
&= n - \operatorname{trace}(I_{p\times p}) \\
&= n - p
\end{aligned}$$

Quadratic forms, Diagonalization and Eigenvalues

The function

$$f(x_1,\ldots,x_n) = f(x) = x^TQx = \sum_{i,j} Q_{i,j}x_ix_j$$

is a quadratic form. The coefficient of a cross product term like x_1x_2 is $Q_{1,2}+Q_{2,1}$ so the function is unchanged if each of $Q_{1,2}$ and $Q_{2,1}$ is replaced by their average. In other words we might as well assume that the matrix Q is symmetric. Consider for example the function $f(x_1,x_2) = 6x_1^2+3x_2^2-4x_1x_2$. The matrix Q is

$$\begin{bmatrix} 6 & -2 \\ -2 & 3 \end{bmatrix}$$

What I did in class is the n-dimensional version of the following: Find new variables $y_1 = a_{1,1}x_1 + a_{1,2}x_2$ and $y_2 = a_{2,1}x_1 + a_{2,2}x_2$ and constants λ_1 and λ_2 such that $f(x_1, x_2) = \lambda_1 y_1^2 + \lambda_2 y_2^2$. Put in the expressions for y_i in terms of the x_i and you get

$$f(x_1,x_2) = (\lambda_1 a_{1,1}^2+\lambda_2 a_{2,1}^2)x_1^2+(\lambda_1 a_{2,1}^2+\lambda_2 a_{2,2}^2)x_2^2+2(\lambda_1 a_{1,1}a_{1,2}+\lambda_2 a_{2,1}a_{2,2})x_1x_2\,.$$

Comparing coefficients we can check that

$$Q = A^T \Lambda A$$

where A is the matrix with entries $a_{i,j}$ and Λ is a diagonal matrix with λ_1 and λ_2 on the diagonal. In other words we have to diagonalize Q.

To find the eigenvalues of Q we can solve $det(A - \lambda I) = 0$ The characteristic polynomial is $(6-\lambda)(3-\lambda) - 4 = \lambda^2 - 9\lambda + 14$ whose two roots are 2 and 7. To find the corresponding eigenvectors you "solve" $(Q - \lambda_i I)v = 0$. For $\lambda_1 = 7$ you get the equations

$$-v_1 - 2v_2 = 0 \qquad \text{and} \qquad -2v_1 - 4v_2 = 0$$

These equations are linearly dependent (otherwise the only solution would be $v = 0$ and λ would not be an eigenvalue). Solving either one gives $v_1 = -2v_2$ so that $(2,-1)^T$ is an eigenvector as is any non-zero multiple of that vector. To get a normalized eigenvector you divide through by the length of the vector, that is, by $\sqrt{5}$. The second eigenvector may be found similarly. We get the equation $2v_2 = 4v_1$ so that $(1,2)^T$ is an eigenvector for the eigenvalue 2. After normalizing we stick these two eigenvectors in the matrix I called P obtaining

$$P = \begin{bmatrix} \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \\ \frac{-1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{bmatrix}$$

Now check that

$$P\Lambda^T P^T = \begin{bmatrix} \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \\ \frac{-1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} 7 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} \frac{2}{\sqrt{5}} & \frac{-1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{bmatrix} = \begin{bmatrix} 6 & -2 \\ -2 & 3 \end{bmatrix} = Q$$

This makes the matrix A above be P^T and $y_1 = (2x_1 - x_2)/\sqrt{5}$ and $y_2 = (x_1 + 2x_2)/\sqrt{5}$. You can check that $7y_1^2 + 2y_2^2 = 6x_1^2 + 3x_2^2 - 4x_1x_2$ as desired.

As a second example consider a sample of size 3 from the standard normal distribution, say, Z_1, Z_2 and Z_3. Then you know that $(n-1)s_Z^2$ is supposed to have a χ^2 distribution on $n-1$ degrees of freedom where now $n=2$. Expanding out

$$2s_Z^2 = (Z_1 - \bar{Z})^2 + (Z_2 - \bar{Z})^2 + (Z_3 - \bar{Z})^2$$

we get the quadratic form

$$2Z_1^2/3 + 2Z_2^2/3 + 2Z_3^2/3 - 2Z_1Z_2/3 - 2Z_1Z_3/3 - 2Z_2Z_3/3$$

for which the matrix Q is

$$Q = \begin{bmatrix} 2/3 & -1/3 & -1/3 \\ -1/3 & 2/3 & -1/3 \\ -1/3 & -1/3 & 2/3 \end{bmatrix}$$

The determinant of $Q - \lambda I$ may be found to be $-\lambda^3 + 2\lambda^2 - \lambda$. This factors as $-\lambda(\lambda-1)^2$ so that the eigenvalues are 1, 1, and 0. An eigenvector corresponding to 0 is $(1,1,1)^T$. Corresponding to the other two eigenvalues there are actually many possibilities. The equations are $v_1 + v_2 + v_3 = 0$ which is 1 equation in 3 unknowns so has a two dimensional solution space. For instance the vector $(1,-1,0)^T$ is a solution. The third solution would then be perpendicular to this, making the first two entries equal. Thus $(1,1,-2)^T$ is a third eigenvector.

The key point in the , however, is that the distribution of the quadratic form Z^TQZ depends only on the eigenvalues of Q and not on the eigenvectors. We can rewrite $2s_Z^2$ in the form $(Z_1^*)^2 + (Z_2^*)^2$. To find Z_1* and Z_2* we fill up a matrix P with columns which are our eigenvectors, scaled to have length 1. This makes

$$P = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{-2}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{bmatrix}$$

and we find $Z^* = P^TZ$ to have components

$$Z_1^* = Z_1/\sqrt{2} - Z_2/\sqrt{2}$$

$$Z_2^* = Z_1/\sqrt{6} + Z_2/\sqrt{6} - 2Z_3/\sqrt{6}$$

and

$$Z_3^* = (Z_1 + Z_2 + Z_3)/\sqrt{3} = \sqrt{3}\bar{Z}$$

You should check that these new variables all have variance 1 and all covariances equal to 0. In other words they are standard normals. Also check that $(Z_1^*)^2 + (Z_2^*)^2 = 2s_Z^2$. Since we have written $2s_Z^2$ as a sum of square of two of these independent normals we can conclude that $2s_Z^2$ has a χ_2^2 distribution.