LECTURE 13

Example 93. Under what circumstances would you obtain the same coefficients from the linear regression lines of Y on X and of X on Y? Let us recall the formulas derived:

$$\hat{\alpha} = \overline{y} - \hat{\beta}\overline{x}$$

and

$$\hat{\beta} = \frac{S_{XY}}{S_{XX}}$$

From the formula for the slope of the regression line, $\hat{\beta}$, it follows that the slopes are identical if the variances of X and Y are equal, $s_{XX} = s_{YY}$, or if the covariance between X and Y is equal to zero, $s_{XY} = 0$. If the slopes are equal, then it is obvious from the formula for the intercept of the regression line $\hat{\alpha}$ that the intercepts are equal if and only if the means of X and Y are the same.

Example 94. Suppose that X has mean zero and covariance $\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$. Let $Y = X_1 + X_2$. Write Y as a linear transformation, i.e., find the transformation matrix A. Then compute $Var(Y)$. Clearly,

$$Y = X_1 + X_2 = AX = (1,1)\begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$$

and

$$Var(AX) = E(AX - EAX)(AX - EAX)^T = AE(X - EX)(X - EX)^T A^T = AVar(X)A^T$$

Hence,

$$Var(Y) = A\Sigma A^T = (1,1)\Sigma\begin{bmatrix} 1 \\ 1 \end{bmatrix} = (1,1)\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = 3$$

Another possibility is to write

$$Var(Y) = Var(X_1 + X_2) = Var(X_1) + 2Cov(X_1, X_2) + Var(X_2) = 3$$

Example 95. Consider the functions

$$f_1(x_1, x_2) = 4x_1\, x_2 \exp(-x_1^2), \quad x_1, x_2 > 0,$$

$$f_2(x_1, x_2) = 2, \quad 0 < x_1, x_2 < 1 \quad \text{and} \quad x_1 + x_2 < 1,$$

$$f_3(x_1, x_2) = \frac{1}{2}\exp(-x_1), \quad x_1 > |x_2|.$$

Check whether they are joint density functions and then compute $E(X)$, $Var(X)$, $E(X_1|X_2)$, $E(X_2|X_1)$, $Var(X_1|X_2)$ and $Var(X_2|X_1)$.

It is easy to see that the first function,

$$f_1(x_1, x_2) = 4x_1\, x_2 \exp(-x_1^2), \quad x_1, x_2 > 0,$$

is not a joint density function. For any value of x_1, we can choose x_2 such that $f_1(x_1, x_2)$ is arbitrarily large on an infinite interval. Hence, it is clear that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_1(x_1, x_2)\, dx_2\, dx_1 = \infty$$

and therefore the function $f_1(x_1, x_2)$ cannot be a joint density function.

The second function,

$$f_2(x_1, x_2) = 2, \quad 0 < x_1, x_2 < 1 \quad \text{and} \quad x_1 + x_2 < 1,$$

is nonnegative and it obviously integrates to one. Hence, it is a joint density function. Note that the function is symmetric in x_1 and x_2 and it follows that $EX_1 = EX_2$ and $VarX_1 = VarX_2$.

For the expected value, we have

$$EX_1 = \int_0^1\int_0^{1-x_1} x_1 2\, dx_2\, dx_1 = \int_0^1 2x_1(1-x_1)\, dx_2\, dx_1 = x_1^2\Big|_0^1 - \frac{2}{3}x_1^3\Big|_0^1 = \frac{1}{3}.$$

We have already observed that $EX_1 = EX_2$ and, thus,

$$EX = \left(\frac{1}{3}, \frac{1}{3}\right)^T.$$

The variances, $VarX_1 = VarX_2$, can be calculated as follows

$$VarX_1 = EX_1^2 - (EX_1)^2 = \int_0^1\int_0^{1-x_1} x_1^2 2\, dx_2\, dx_1 - \frac{1}{9} = \int_0^1 2x_1^2(1-x_1)\, dx_1 - \frac{1}{9} =$$

$$\frac{2}{3}x_1^3\Big|_0^1 - \frac{2}{4}x_1^4\Big|_0^1 - \frac{1}{9} = \frac{1}{6} - \frac{1}{9} = \frac{1}{18}.$$

The covariance $Cov(X_1, X_2)$ is equal to

$$Cov(X_1, X_2) = E(X_1X_2) - E(X_1)E(X_2) = \int_0^1 \int_0^{1-x_1} x_1\, x_2 2\, dx_2\, dx_1 - \frac{1}{9} = \int_0^1 x_1(1-x_1)^2\, dx_1 - \frac{1}{9} =$$

$$= \int_0^1 (x_1 - 2x_1^2 + x_1^3)\, dx_1 - \frac{1}{9} = \frac{1}{2} - \frac{2}{3} + \frac{1}{4} - \frac{1}{9} = -\frac{1}{36}.$$

The resulting covariance matrix is

$$Var(X) = \begin{pmatrix} \frac{1}{18} & -\frac{1}{36} \\ -\frac{1}{36} & \frac{1}{18} \end{pmatrix}.$$

The conditional expectations could be calculated by evaluating the appropriate integrals. However, in this case, the solution can be seen immediately. Clearly, the conditional distribution of X_2 given $X_1 = x_1$ is uniform on $(0, 1 - x_1)$. The expected value of uniform distribution is its center, i.e., $E(X_2|X_1 = x_1) = (1 - x_1)/2$. Due to the symmetry of the distribution, we have also that $E(X_1|X_2 = x_2) = (1 - x_2)/2$.

The conditional variances are also variances of uniform distributions:

$$Var(X_2|X_1 = x_1) = E(X_2^2|X_1 = x_1) - \{E(X_2|X_1 = x_1)\}^2 = \int_0^{1-x_1} \frac{x_2^2}{1-x_1}\, dx_2 - \left(\frac{1-x_1}{2}\right)^2 =$$

$$= \frac{(1-x_1)^2}{3} - \frac{(1-x_1)^2}{4} = \frac{(1-x_1)^2}{12}.$$

Due to the symmetry, we have also that

$$Var(X_1|X_2 = x_2) = \frac{(1-x_2)^2}{12}.$$

For the third function,

$$f_3(x_1, x_2) = \frac{1}{2}\exp(-x_1), \qquad x_1 > |x_2|$$

we again start by verifying that it is a joint density function. We have

$$\int_0^\infty \int_{-x_1}^{x_1} f_3(x_1, x_2)\, dx_2\, dx_1 = \int_0^\infty \int_{-x_1}^{x_1} \frac{1}{2}\exp\{-x_1\}\, dx_2\, dx_1 = \int_0^\infty x_1\, \exp\{-x_1\}\, dx_1 = 1.$$

Here, it is helpful to notice that the value of $f_3(x_1, x_2)$ is for any value of x_1 symmetric around zero in x_2 and that the value of the joint density function does not depend on x_2. Notice that the conditional expected value of X_2 is finite since X_2 has bounded support for each value of X_1. From the symmetry, it follows that $E(X_2|X_1 = x_1) = 0$, this in turn implies that $EX_2 = E(E(X_2|X_1)) = 0$.

The fact that the value of the joint density function does not depend on x_2 implies that the conditional distribution of X_2 given $X_1 = x_1$ is uniform on the interval $(-x_1, x_1)$. Looking at the above calculations for the variance of the uniform distribution, we can immediately write:

$$Var(X_2|X_1 = x_1) = \frac{(2x_1)^2}{12} = \frac{x_1^2}{3}.$$

In order to calculate the moments of X_1, we have to evaluate some integrals:

$$EX_1 = \int_0^\infty \frac{1}{2} x_1 \exp\{-x_1\}\, dx_2\, dx_1 = \int_0^\infty x_1^2 \exp\{-x_1\}\, dx_1 =$$

$$= 2x_1^2 \exp(-x_1)\Big|_0^\infty + 2\int_0^\infty \exp\{-x_1\}\, dx_1 = 2.$$

Hence, the vector of expected values is $EX = (2, 0)^T$.

The variance of X_1 can be calculated similarly as the expected value

$$VarX_1 = \int_0^\infty \int_{-x_1}^{x_1} \frac{1}{2} x_1^2 \exp\{-x_1\}\, dx_2\, dx_1 - 4 = \int_0^\infty x_1^3 \exp\{-x_1\}\, dx_1 - 4 = 3EX_1 - 4 = 2.$$

Now it is easy to calculate also the unconditional variance of X_2 since

$$Var(X_2) = E(Var(X_2|X_1)) + Var(E(X_2|X_1)) = E\left(\frac{X_1^2}{3}\right).$$

Note that the symmetry of the joint density function in x_2 implies that also the distribution of the random variable X_1X_2 is symmetric around 0 and, hence, its expected value $E(X_1X_2) = 0$. It follows that

$$Cov(X_1, X_2) = E(X_1X_2) - EX_1\, EX_2 = 0.$$

The variance matrix of the random vector X is $VarX = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$.

It remains to investigate the conditional moments of X_1 given $X_2 = x_2$. The conditional density of X_1 given X_2 is

$$f_{X_1|X_2=x_2}(x_1) = \frac{f_3(x_1, x_2)}{f_{X_2}(x_2)} = \frac{\exp(-x_1)}{\int_{|x_2|}^\infty \exp(-x_1)\, dx_1} = \frac{\exp(-x_1)}{\exp(-|x_2|)},$$

for $x_1 > |x_2|$ and 0 otherwise.

The conditional expectation of X_1 can be calculated as

$$E(X_1|X_2 = x_2) = \int_{|x_2|}^{\infty} x_1\, f_{X_1|X_2=x_2}(x_1)\, dx_1 = \int_{|x_2|}^{\infty} \frac{\exp(-x_1)}{\exp(-|x_2|)}\, dx_1 =$$

$$= \frac{1}{\exp(-|x_2|)}\{|x_2| \exp(-|x_2|) + \exp(-|x_2|)\} = |x_2| + 1.$$

Finally, the conditional variance of X_1 given $X_2 = x_2$ is

$$Var(X_1|X_2 = x_2) = E(X_1^2|X_2 = x_2) - (E(X_1|X_2 = x_2))^2 = \int_{|x_2|}^{\infty} x_1^2 \frac{\exp(-x_1)}{\exp(-|x_2|)}\, dx_1 - (|x_2| + 1)^2 =$$

$$= \frac{1}{\exp(-|x_2|)}\{|x_2|^2 \exp(-|x_2|) + 2(|x_2|\ \exp(-|x_2|) + \exp(-|x_2|))\} - (|x_2| + 1)^2 =$$

$$= |x_2|^2 + 2|x_2| + 2 - (|x_2| + 1)^2 = 1.$$

§56. SOME PROPERTIES OF MATRICES RELATED TO EIGENVALUES AND EIGENVECTORS.

There are many interesting and useful relationships between the eigenvalues and eigenvectors of a matrix and certain characteristics of the latter.

Theorem 1. The sum $\sum_{i=1}^{p} \lambda_i$ of the eigenvalues of a matrix $\mathbf{A} = [a_{ij}]_{i,j=1}^{p}$ is equal to the sum $\sum_{i=1}^{p} a_{ii}$ of the latter's diagonal elements. This sum is called the trace of $\mathbf{A}$, denoted $\mathrm{tr}(\mathbf{A})$.

Outline of proof. The eigenvalues λ_i of $\mathbf{A}$ are, by definition, the roots of its characteristic equation,

$$\begin{vmatrix} a_{11} - \lambda & a_{12} & \dots & a_{1p} \\ a_{21} & a_{22} - \lambda & \dots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \dots & a_{pp} - \lambda \end{vmatrix} = 0$$

which is a polynomial equation of degree p in λ. In the theory of equations, it is shown that if $\lambda_1, \lambda_2, \dots, \lambda_p$ are roots of a polynomial equation

$$c_0\lambda^p + c_1\lambda^{p-1} + c_2\lambda^{p-2} + \dots + c_p = 0$$

then

$$\sum \lambda_i = -(c_1/c_0)$$

For our characteristic equation, it can be shown by expanding the determinant that

$$c_0 = (-1)^p$$

and

$$c_1 = (-1)^{p-1}(a_{11} + a_{22} + \ldots + a_{pp})$$

It therefore follows that

$$\sum \lambda_i = \sum a_{ii}$$

as was to be proved.

Theorem 2. The product $\lambda_1, \lambda_2, \ldots, \lambda_p$ (which is abbreviated $\prod_{i=1}^{p} \lambda_i$) of the eigenvalues of a matrix $\mathbf{A}$ is equal to the value of the determinant $|\mathbf{A}|$.

Outline of proof. Borrowing again from the theory of equations, we have the relation

$$\prod_{i=1}^{p} \lambda_i = (-1)^p (c_p/c_0),$$

where c_p is the constant term, and c_0 the coefficient of λ^p in the characteristic equation, as in the proof outline for Theorem 1. But the constant term of $|A - \lambda I|$ is readily found by setting $\lambda = 0$ in this expression: $c_p = |A - 0I| = |A|$ From this and the fact that $c_0 = (-1)^p$, stated before, it immediately follows that

$$\prod_{i=1}^{n} \lambda_i = |\mathbf{A}|.$$

An important corollary to this theorem is that a matrix has one or more eigenvalues equal to zero if and only if the matrix is singular. This is immediately obvious from the result of Theorem 2. Furthermore, the number of zero eigenvalues has an important bearing on the nature of $\mathbf{A}$, especially when the latter is a variance-covariance matrix. Even though the presence of just one eigenvalue equal to zero is sufficient to assure that $\mathbf{A}$ is singular, we somehow feel that it is "more singular" with a greater number of zero eigenvalues.

Theorem 3. Two eigenvectors $\mathbf{v_i}$ and $\mathbf{v_j}$ associated with two distinct eigenvalues λ_i and λ_j of a symmetric matrix are mutually orthogonal; that is, $v_j' v_j = 0$.

Proof. An instance of this theorem was already verified for a numerical example. The general proof is as follows. By hypothesis, we have

$$\mathbf{A}\mathbf{v_i} = \lambda_i \mathbf{v_i} \tag{a}$$

and

$$\mathbf{A}\mathbf{v_j} = \lambda_j \mathbf{v_j} \tag{b}$$

with $\lambda_i \neq \lambda_j$ Taking the transposes of the two members of Equation a, we have:

$$\mathbf{v}_i'\mathbf{A} = \lambda_i \mathbf{v}_i' \tag{a'}$$

(since $\mathbf{A}' = \mathbf{A}$ by hypothesis). Premultiplying both sides of Equation b by $\mathbf{v}_i'$ yields

$$\mathbf{v}_i' A v_j = \lambda_j \mathbf{v}_i' v_j \tag{c}$$

in the left member of which we may, from Equation a', substitute $\lambda_i \mathbf{v}_j{}'$ in place of $\mathbf{v}_i' A$ to obtain

$$\lambda_i \mathbf{A} v_j = \lambda_j \mathbf{v}_i' v_j$$

or

$$(\lambda_i - \lambda_j)(\mathbf{v}_i{}' v_j q) = 0$$

Since $\lambda_i - \lambda_j \neq 0$ by hypothesis, it follows that $\mathbf{v}_i{}'\mathbf{v}_j = 0$. The proof is complete.