Norwegian University of Science and Technology
Department of Mathematical Sciences

TMA4267 Linear statistical models
Recommended exercises 4

Problem 1 Symmetric, idempotent matrices

A square matrix A is *idempotent* if $A^2 = A$. The *trace*, $\operatorname{tr} A$, of a square matrix A is the sum of its diagonal entries, which is in general equal to the sum of the eigenvalues (counted with multiplicities as roots of the characteristic polynomial). The *rank*, $\operatorname{rank} A$, of a matrix A is the dimension of the column space, which is equal to the dimension of the row space. A square $n \times n$ matrix A is invertible if and only if $\operatorname{rank} A = n$, and if and only if 0 is not an eigenvalue of A.

a) Find a 2×2 matrix that is idempotent but not symmetric.

We have seen that the eigenvalues of an idempotent matrix A are 0 and 1, and that $\operatorname{rank} A = \operatorname{tr} A$. The latter can be seen more easily in the case that A is in addition symmetric, hence diagonalizable, using the fact that similar matrices have the same rank, i.e., if A and B are two square matrices such that $B = P^{-1}AP$ for an invertible matrix P, then $\operatorname{rank} A = \operatorname{rank} B$.

b) Assume that A is idempotent and symmetric. Show that $\operatorname{rank} A = \operatorname{tr} A$ by considering a diagonalization $\Lambda = P^{-1}AP$ of A.

c) Let $\mathbf{1}$ be a vector of 1s, $\mathbf{1} = (1\ 1\ \cdots\ 1)^{\mathrm{T}}$, so that

$$\mathbf{1}\mathbf{1}^{\mathrm{T}} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix}$$

Show that $\frac{1}{n}\mathbf{1}\mathbf{1}^{\mathrm{T}}$ and $I - \frac{1}{n}\mathbf{1}\mathbf{1}^{\mathrm{T}}$ are both symmetric and idempotent, and find their ranks.

Problem 2 Quadratic form

(From Exam TMA4267, spring 2014, Problem 1. See also Recommended Exercises 1, Problem 2, for first part of exam problem.)

Let $\boldsymbol{X}$ be a trivariate random vector with mean $(1\ 1\ 1)^{\mathrm{T}}$ and covariance matrix I, a 3×3 identity matrix. Let

$$A = \begin{pmatrix} \frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{3} & -\frac{1}{3} & \frac{2}{3} \end{pmatrix}.$$

a) Find the mean of $\boldsymbol{X}^{\mathrm{T}} A \boldsymbol{X}$. (Hint: formula involving the trace)

Now assume that $\boldsymbol{X}$ is trivariate normal.

b) Show that A is a symmetric and idempotent matrix. Find the rank of A. Derive the distribution of $\boldsymbol{X}^{\mathrm{T}} A \boldsymbol{X}$. Find the probability that $\boldsymbol{X}^{\mathrm{T}} A \boldsymbol{X}$ is less than 6.

Problem 3 The F-distribution

This problem is similar to Problems 2–3 of Recommended Exercises 2, which were about the t- and the chi-squared distributions, respectively. Another distribution of great importance in statistical inference is the F-distribution.

The F-distribution with (p, q) degrees of freedom is the distribution of $F = \frac{V/p}{W/q}$, where $V \sim \chi^2_p$ and $W \sim \chi^2_q$ and V and W are independent. We write $F \sim F_{p,q}$.

a) Use the multivariate transformation formula to find the pdf of the F-distribution.

Hint: Let $G = W$ and use the multivariate transformation formula to find the joint pdf of F and G. Find the marginal distribution of F from this joint distribution. For the last part it will help you to recognize the integral of a χ^2 pdf.

b) Let $F \sim F_{p,q}$ (F-distribution with (p, q) degrees of freedom). Show that $1/F \sim F_{q,p}$.

Hint: Use definition of F-distribution in terms of chi-squared distributed variables rather than a transformation formula.

c) Let $T \sim t_q$ (t-distribution with q degrees of freedom). Show that $T^2 \sim F_{1,q}$ (F-distribution with $(1, q)$ degrees of freedom).

Hint: Use definition of t-distribution in terms of normally and chi-squared distributed variables rather than a transformation formula.

Problem 4 T- and F-distributions in R

This problem is similar to Problem 4 of Recommended Exercises 2, which was about the normal and chi-squared distributions, respectively.

Let $B = 10000$ and $p = 9$.

a) Generate a random sample of B data points $U_i \sim N(0, 1)$, and independently a random sample of B data points $V_i \sim \chi^2_p$. Plot a histogram of the t-ratios, $U_i/\sqrt{V_i/p}$. Add the pdf of the t_p distribution to the histogram. Then add vertical lines at the 0.15 and 0.85 quantiles. Repeat this for other values of p.

b) Plot a histogram of the squares of the t-ratios. Add the pdf of the $F_{1,p}$ distribution to the histogram. Then add vertical lines at the 0.05 and 0.95 quantiles.

c) Let $n_1 = 5$ and $n_2 = 40$. Generate a random sample of B data points from the $\chi^2_{n_1}$ distribution and independently a random sample of B data points from the $\chi^2_{n_2}$ distribution. Plot a histogram of the F-ratios from the definition of F given in Problem 3. Add the pdf of the F_{n_1,n_2} distribution to the histogram. Then add vertical lines at the 0.05 and 0.95 quantiles.