TMA4267 Linear statistical models
Recommended exercises 2

Problem 1   Principal components analysis

In R, the data set `USArrests` is included. Type `USArrests` to view it, and `?USArrests` to get more information. If you need hints, take look at the R code linked to in Blackboard for the fourth lecture.

a) Find the loadings (rotations) of the principal components. Also do it without using the `prcomp` function.

b) Find the sample variances of the principal components. Also do it without using the `prcomp` function.

c) Find the scores, that is, the values of the principal components for each data point. Check that the scores for Alabama are indeed the linear combinations of the data for Alabama with the loadings as coefficients.

d) Make a biplot of the first two principal components.

e) How many principal components are needed to capture $80\,\%$ of the variability of the data?

f) Investigate what happens if we forget to use centered and scaled data.

Problem 2   Multivariate transformation – the t distribution

Let $U \sim N(0,1)$ and $V \sim \chi_p^2$, and assume that U and V are independent. Let $T = U/\sqrt{V/p}$. The distribution of T is called the *t distribution with p degrees of freedom*, $T \sim t_p$.

a) What is the joint pdf of U and V? Use the multivariate transformation formula for T and $W = V$ to find the joint pdf of T and W

b) Find the marginal pdf of T. Hint: Try to recognize the integral of the pdf of the gamma distibution, given by

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta},$$

with parameters $\alpha = (p+1)/2$ and $\beta = 2/(1 + t^2/p)$.

Problem 3   **The standard normal and chi-squared distributions**

a) Let $U \sim N(0,1)$. Find the pdf of U^2. Also find the MGF of U^2.

b) Let V be the sum of the squares of p independent $N(0,1)$ variables. Then V has the *chi-squared distribution with p degrees of freedom*, and we write $V \sim \chi_p^2$. Show that the pdf of V is given by

$$f(v) = \frac{1}{2^{p/2}\Gamma(p/2)} v^{p/2-1} e^{-v/2}$$

Hint: First use the result from a) to find the MGF for V. Then find the MGF of a variable having pdf f. Show that the two coincide.

Problem 4   **Normal and chi-squared distributions in R**

The pdf, cdf and quantile function (inverse of cdf, e.g. to find critical values) are available in R for many standard distributions. The first letter of the name of the function is always `d` for the pdf, `p` for the cdf and `q` for the quantile function, and the rest of the name indicates the distribution class (for the normal distribution, the functions are named `dnorm`, `pnorm` and `qnorm`). In addition, random data can be generated from a distribution using a function having prefix `r` (`rnorm` for the normal distribution).

a) Find out more about these functions by typing e.g. `rnorm` and `pchisq`. Try some of them.

b) Make a plot of the standard normal pdf. Add vertical lines at the 0.05 and 0.95 quantiles, e.g. by using the `abline` function. Colour the tails, e.g. by using the `polygon` function.

c) Generate 10000 observations from the standard normal distribution and square the data. Plot a histogram of the data using `hist`. Add the pdf of the chi-squared distribution with one degree of freedom to the histogram. Add vertical lines at the 0.1 and 0.9 quantiles.