IND522
Advanced Statistical Modelling

Fall 2025

Professor Orhan FEYZİOĞLU

Assignment ♮4: Due November 21

1) Write a function that will generate chi-square random variables with v degrees of freedom by generating v standard normals, squaring them and then adding them up. This uses the fact that

$$X = Z_1^2 + \ldots + Z_v^2$$

is chi-square with v degrees of freedom. Generate some random variables and plot in a histogram. The degrees of freedom should be an input argument set by the user. Generate also chi-square random variables with the same v degrees of freedom using available **chi2rnd** function and contrast your results.

2) Generate four sets of random variables with $v = 2, 5, 15, 20$ using the function **chi2rnd**. Create histograms for each sample. How does the shape of the distribution depend on the degrees of freedom v?

3) To generate random variable X from a beta distribution with parameters α and β, the following method is suggested: generate two variates $Y_1 = U_1^{1/\alpha}$ and $Y_2 = U_2^{1/\beta}$ where the U_i are from the uniform distribution, and if $Y_1 + Y_2 \leq 1$ then

$$X = \frac{Y_1}{Y_1 + Y_2}.$$

Implement this algorithm. Generate also beta random variables with the same parameters using available **betarnd** function and contrast your results.

4) Generate 1000 binomial random variables for $n = 5$ and $p = 0.3, 0.5, 0.8$. In each case, determine the observed relative frequencies and the corresponding theoretical probabilities. How is the agreement between them?

5) Generate random variables from the bivariate t distribution for $v = 5, 15, 25, 35$. Use the correlation matrix $\Sigma = [1 - 0.8; -0.81]$ and $n = 100$. Construct scatterplots and compare them to a scatterplot of random variables that are bivariate normal with the same correlation matrix and sample size.

6) The heights of 32 Tibetan skulls measured in millimetres is given in the file `tibetan`. These data comprise two groups of skulls collected in Tibet. One group of 17 skulls comes from graves in Sikkim and nearby areas of Tibet and the other 15 skulls come from a battlefield in Lhasa. The data contain five measurements: greatest length of skull (X_1), greatest horizontal breadth of skull (X_2), height of skull (X_3), upper face height (X_4) and face breadth between outermost points of cheek bones (X_5). Do a quantile plot of the Tibetan skull data using the standard normal quantiles. Is it reasonable to assume the data follow a normal distribution?

7) In the `counting` data set, we have the number of scintillations in 72-second intervals arising from the radioactive decay of polonium. Construct a Poissonness plot. Does this indicate agreement with the Poisson distribution?

8) The `biology` data set contains the number of research papers for 1534 biologists. The variables `numpaps` and `freqs` are related to the number of papers per author and the observed number of authors, respectively. Construct a binomial plot of these data. Analyze your results.

9) The `thrombos` data set contains measurements of urinary-thromboglobulin excretion in 12 normal and 12 diabetic patients. Put each of these into a column of a matrix and use the **boxplot** function to compare normal versus diabetic patients.

10) The `bank` data contains two matrices comprised of measurements made on genuine money and forged money. Use some of the visualization techniques mentioned in this chapter to explore the data.