

Use of a basic calculator is allowed. Graphical calculators and mobile phones are not allowed. This exam consists of 4 questions (27 points).

Please write all answers in English. Grade = $\frac{total+3}{3}$.

GOOD LUCK!

Question 1 [7 points]

Are the following statements correct/sensible? Motivate your answer by a short argument or a sketch.

- a. [2 points] In the context of bootstrapping: the parametric bootstrap is always better than the empirical bootstrap.
- b. [1 point] The influence function of the 10%-trimmed mean is bounded.
- c. [2 points] Consider a bivariate sample $(X_1, Y_1), \dots, (X_n, Y_n)$. The two stem-and-leaf plots of X -values and Y -values separately contain the same information as the bivariate scatter plot.
- d. [2 points] If the dots in a QQ -plot of a sample (vertical axis) against some distribution F_0 (horizontal axis) show an S-shape, the distribution F_0 has heavier tails than the distribution of the data.

Question 2 [8 points]

In Figure 1 (see page 3) an empirical (two sample) QQ -plot of two data sets x and y is shown.

- a. [2 points] Do you think that the underlying distributions of the two data sets belong to the same location scale family? Motivate your answer.
- b. [1 point] What can you say about the normality of the underlying distribution of data set x ?
- c. [3 points] Suppose that we would like to test whether or not the underlying distribution of data set x is the normal distribution with expectation 0.5 and variance 1. Evaluate for each of the following tests for goodness of fit how suitable they are for testing this, and motivate your answer:
 - i) Kolmogorov-Smirnov test;
 - ii) chi-square test for goodness of fit;
 - iii) Shapiro-Wilk test.

- d. [2 points] Consider the following goodness-of-fit test situation:
 $H_0 : F \in \mathcal{F}_0$ where F is the unknown underlying distribution of a given sample and \mathcal{F}_0 is some class of distributions. Explain the following statement: “The test statistic T for goodness of fit is nonparametric (distribution free) under the null hypothesis”.

Question 3 [5 points]

Consider the data presented in Figure 2 (see page 3).

- a. [2 points] Empirical bootstrap values for the sample mean and sample median of this data set were computed and some quantiles of these bootstrap values of both location estimators are:

quantile	0.025	0.05	0.5	0.95	0.975
estimator 1	0.68	0.75	1.08	1.39	1.50
estimator 2	1.49	1.55	2.07	2.54	2.78

Indicate which of the two estimators is the mean: estimator 1 or estimator 2? Motivate your answer.

- b. [2 points] Determine the length of the 95% bootstrap confidence intervals both for the mean and the median of the underlying distribution. (You are not asked to determine the intervals, only their lengths.)
- c. [1 point] Which estimator for location do you prefer for this data set? Motivate your answer.

Question 4 [7 points]

Let X_1, \dots, X_n be independent and identically distributed random variables with unknown distribution P . Suppose that the sample variance $T_n(X_1, \dots, X_n) = S_X^2$ is used to estimate the variance of P . To determine the accuracy of this estimator, its standard deviation is estimated by means of the empirical bootstrap.

- a. [3 points] Describe the steps of the empirical bootstrap scheme that you would use to find the bootstrap estimate of the standard deviation of T_n .
- b. [2 points] Describe shortly which two errors are (necessarily) made in this bootstrap procedure.
- c. [2 points] Now consider a bootstrap test for testing $H_0 : P \in \mathcal{P}_0$ using some sensible test statistic which has an unknown distribution under H_0 . Indicate of each error that you mentioned in part (b) whether such an error is also present in the context of this bootstrap test. Motivate your answer.

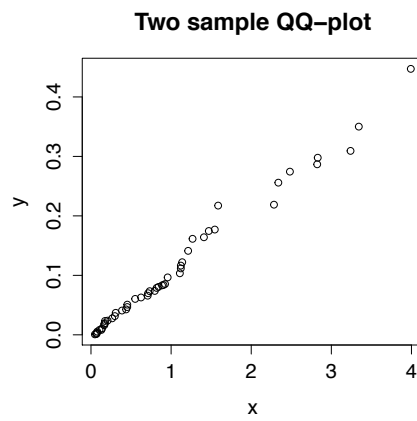


Figure 1: Two sample QQ-plot of two data sets x and y .

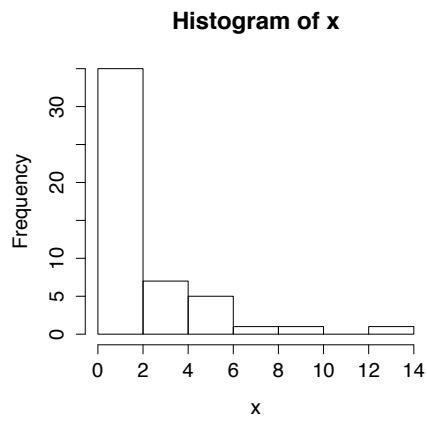


Figure 2: Histogram of a data set x .

THE END