UNIVERSITY OF OSLO

Faculty of Mathematics and Natural Sciences

Exam in GEO4300/9300 – Geophysical Data Science

Day of exam: 23 November, 2020
Exam hours: 09:00 – 12:00 (3 hours)

This examination paper consists of 6 pages including this page.

Note:
1. This exam is an open book exam. All materials and tools are permitted.
2. There are in total 50 points in this exam.

1 Random variable parameter estimation

A discrete random variable X is defined by

$$X = \begin{cases} -1, & prob. = 1/3 \\ 3, & prob. = 1/2 \\ 4, & prob. = 1/6 \end{cases} \tag{1}$$

(a) find the expected value

(b) find the variance

(c) find the mode

(d) find the coefficient of variation

2 Frequency analysis and linear regression

(a) What is the probability to observe at least one 100-years flood or larger within a period of 10 years?

(b) Figure 1A shows a simple linear regression between average runoff and median annual flood. Figure 1B shows the QQ-plot of the residual where the theoretical quantiles were calculated using the normal distribution. Describe which assumption of a simple linear regression is violated in this analysis, and discuss strategies that can be used to improve the analysis.

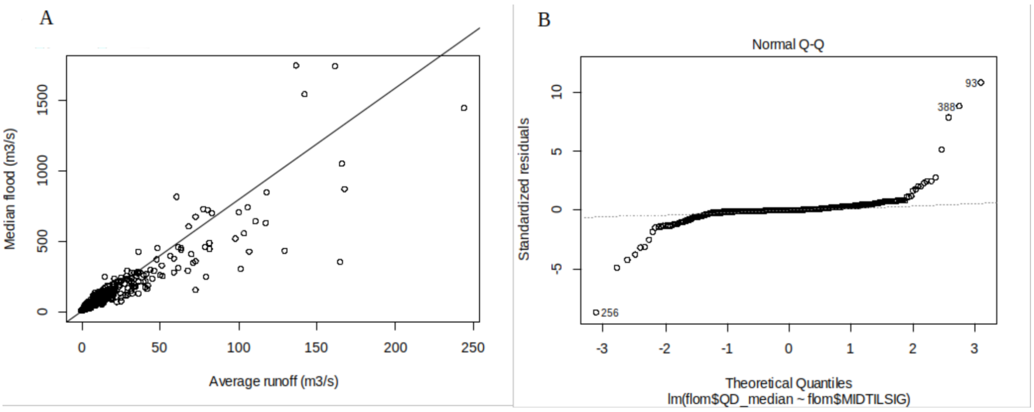


Figure 1: A) linear regression, B) Q-Q plot.

3 Confidence intervals

A sample of 30 random observations produced a mean of 145 and variance of 20.

(a) What is the 95% confidence interval on the mean assuming a normal distribution if

(i) the true variance is unknown and estimated as 20

(ii) the true variance is 20

(b) What is the reason for the difference of results in part (i) and part (ii)?

(c) What is the 95% confidence interval on the variance?

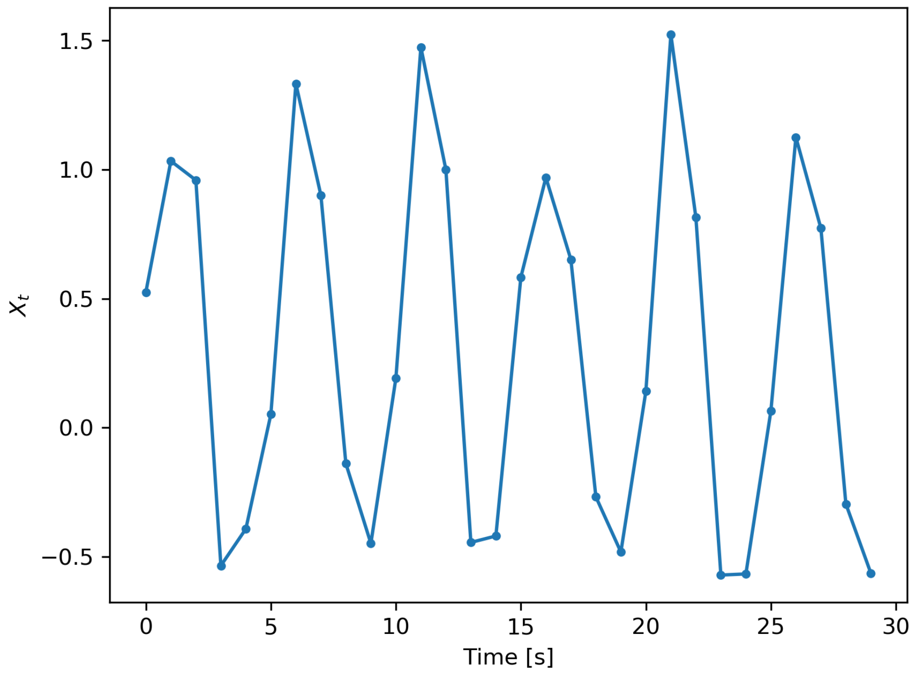
4 Machine learning

(a) Why is it common to split the dataset into a training set and a test set when doing machine learning? In your answer, include in a relevant way the terms training error and test error

(b) In many machine learning algorithms you have a parameter that controls the complexity of the model. Why do we want to control this complexity?

5 Time series analysis and Fourier transformation

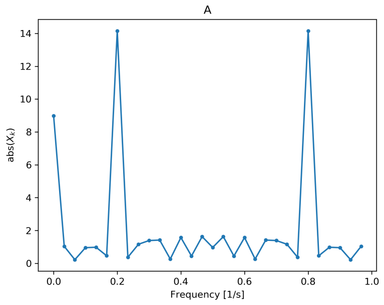
Consider the following time series X_t sampled once per second

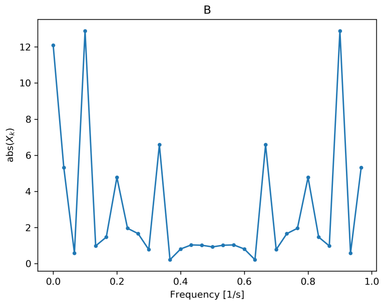


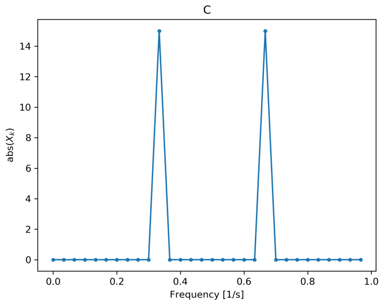
(a) How could you test if there is a significant trend in X_t? Explain a suitable test.

(b) The following three graphs show the absolute values for Fourier coefficients, defined as:

$$X_k = \sum_{n=0}^{N-1} x_n \cdot e^{-i\,2\pi\,k\,n/N}$$







Which one of them (A, B or C) shows the Fourier transform of X_t? Explain your answer.