Examination Computational Statistics

Linköpings Universitet, IDA, Statistik

Course:	732A90 Computational Statistics
Date:	2021/03/23, 8–13
Teacher:	Krzysztof Bartoszek
Provided aids:	material in the zip file **exam_material_732A90.zip**
Grades:	A= $[18-20]$ points B= $[16-18)$ points C= $[14-16)$ points D= $[12-14)$ points E= $[10-12)$ points F= $[0-10)$ points
Instructions:	Provide a detailed report that includes plots, conclusions and interpretations. If you are unable to include a plot in your solution file clearly indicate the section of R code that generates it. Give motivated answers to the questions. If an answer is not motivated, the points are reduced. Provide all necessary codes in an appendix. In a number of questions you are asked to do plots. Make sure that they are informative, have correctly labelled axes, informative axes limits and are correctly described. Points may be deducted for poorly done graphs. Name your solution files as: **[your id]_[own file description].[format]** If you have problems with creating a pdf you may submit your solutions in text files with unambiguous references to graphics and code that are saved in separate files. There are **TWO** assignments (with sub–questions) to solve. Provide a separate solution file for each assignment. Include all R code that was used to obtain your answers in your solution files. Make sure it is clear which code section corresponds to which question. If you also need to provide some hand–written derivations please number each page according to the pattern: Question number . page in question number i.e. Q1.1, Q1.2, Q1.3,..., Q2.1, Q2.2, ..., Q3.1, ... Scan/take photos of such derivations preferably into a single pdf file but if this is not possible multiple pdf or .bmp/.jpg/.png files are fine. Please do not use other formats for scanned/photographed solutions. Please submit all your solutions via LISAM or e–mail. If emailing, please email them to **BOTH** krzysztof.bartoszek@liu.se and KB_LiU_exam@protonmail.ch . During the exam you may ask the examiner questions by emailing them to KB_LiU_exam@protonmail.ch **ONLY**. Other exam procedures in LISAM.

NOTE: If you fail to do a part on which subsequent question(s) depend on describe (maybe using dummy data, partial code e.t.c.) how you would do them given you had done that part. You *might* be eligible for partial points.

Assignment 1 (10p)

You are provided with the following summation results

$$\sum_{k=0}^{n} \frac{2}{(k+2)(k+3)} = \frac{n+1}{n+3}, \quad \sum_{k=0}^{\infty} \frac{2}{(k+2)(k+3)} = 1.$$

Let X be a discrete random variable with support on $\{0, 1, 2, \ldots\}$ and probability law

$$P(X = k) = \frac{2}{(k+2)(k+3)}, \quad k = 0, 1, 2, \ldots.$$

Question 1.1 (3p)

Implement a procedure that simulates X by directly using the probability mass function $P(X = k)$. Is there a potential problem with this procedure? On what part of the support can it have issues?

Question 1.2 (4p)

Implement a procedure that simulates X using its cumulative distribution function. Does this approach have any issues or is it exact?

Question 1.3 (3p)

Simulate a large (determine yourself how large will be appropriate) number of values of X using both methods implemented in a) and b). Compare the empirical probabilities with their true values. Which method seems better? If one is better, is there a particular region of the support of X where this is so?

Assignment 2 (10p)

Consider the function

$$\eta_n(s) = \sum_{k=1}^{n} \frac{(-1)^{k-1}}{k^s}$$

for $n \in \mathbb{N}$ and $s > 0$. It is known that

$$\eta_n(s) \xrightarrow{n\to\infty} a \cdot (b - c^{d-s})\zeta(s),$$

where $a, b, c, d \in \mathbb{R}$ and the values of the function $\zeta(s)$ can be obtain using `pracma::zeta()`. Furthermore, the `pracma` package provides `pracma::eta(s)`$= \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k^s}$.

IMPORTANT: Assume that $s \neq 1$ (and consider s that are not close to 1). While the formulæ will all hold, some special numerical treatment could be required to perform the below tasks for $s = 1$.

IMPORTANT: In this question you may **NOT** use `pracma::eta()` to solve any of the problems. You are permitted and required to use it only for a specific task in Question 2.3.

Question 2.1 (3p)

Implement a function that calculates $\eta_n(s)$. Make sure it is as effective numerically as possible.

Question 2.2 (3p)

Plot $\eta_n(s)$ for some useful values of n and s in order to have some informed guesses concerning a, b, c and d. Based on these plots discuss what could be plausible values for a, b, c and d.

Question 2.3 (4p)

Use `R`'s `optim()` function to numerically find the values of a, b, c and d, using your implementation of $\eta_n(s)$. You have to define the objective function, and starting points for the optimization yourself. Evaluate how well your optimization performed by comparing with both `pracma::eta()` and $\eta_n(s)$.