Examination Computational Statistics

Linköpings Universitet, IDA, Statistik

Course code and name:	732A90 Computational Statistics
Date:	2018/05/16, 8–19
Assisting teacher:	Krzysztof Bartoszek
Allowed aids:	Printed books, 100 page computer document, and material in the zip file **732A90_Examination_ExtraMaterial.zip**
Grades:	A= $[17 - 20]$ points B= $[14.5 - 17)$ points C= $[10.5 - 14.5)$ points D= $[8.5 - 10.5)$ points E= $[7 - 8.5)$ points F= $[0 - 7)$ 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 exam account]_[own file description].[format]** 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.

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)

Consider the Pareto distribution with density

$$f_{\alpha,\beta}(x) = \begin{cases} \alpha \frac{\beta^\alpha}{x^{\alpha+1}} & x > \beta,\ \alpha, \beta > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Question 1.1 (5p)

Calculate the cumulative distribution function and then the inverse of the cumulative distribution function. Implement a function that simulates from the Pareto distribution using the inverse cumulative distribution function method. Your function should take α and β as its parameters. Simulate a sample of size 10000 for a couple of pairs of α and β values (take $\alpha \geq 3$ and $\beta \geq 1$), calculate the sample statistics, plot histograms. Compare the sample statistics with the theoretical values (see second page of file `732A90_CS_Examination_GutAppendixB.pdf` in `732A90_Examination_ExtraMaterial.zip`). Compare your simulated histogram with the "true" one, use `actuar::rpareto()`.

Question 1.2 (5p)

Assume $\alpha \geq 3$ and $\beta > 1$. Notice that the density $f_{3,\beta}$ (when $\beta > 1$) can be a dominating density for all other Pareto distributions with $\alpha > 3$ (and same $\beta > 1$). Show this and calculate the majorizing constant. Implement an acceptance–rejection algorithm to simulate from the Pareto distribution with $\alpha > 3$ and $\beta > 1$. Simulate a sample of size 10000 for a couple of pairs of α and β values (take $\alpha > 3$ and $\beta > 1$), calculate the sample statistics, plot histograms. Compare the sample statistics with the theoretical values and compare with the histograms in Part a) and those by `actuar::rpareto()`. Which simulation algorithm is faster (use `R`'s `Sys.time()` and `difftime()` if needed)? Explain.

Assignment 2 (10p)

Finding the minimum or maximum of a function is usually presented as a goal in itself. Here we ill use the function `optim()` to create a procedure to approximate another function, through so–called parabolic interpolation. For this exercise let $f(x)$ be a continuous function on the interval $[0, 1]$ and let $x_0, x_1, x_2 \in [0, 1]$ such that $f(x_1) < f(x_0), f(x_2)$. We will approximate the function $f(x)$ with a function $\tilde{f}(x) = a_0 + a_1 x + a_2 x^2$, i.e. a quadratic function.

Question 2.1 (5p)

Write a function that uses `optim()` and finds values of (a_0, a_1, a_2) for which $\tilde{f}$ interpolates f at user provided points x_0, x_1, x_2. Interpolate means $f(x_0) = \tilde{f}(x_0)$, $f(x_1) = \tilde{f}(x_1)$ and $f(x_2) = \tilde{f}(x_2)$.

Question 2.2 (5p)

Apply your function to $f_1(x) = -x(1 - x)$ for a couple of different triples of x_0, x_1, x_2. Plot $f_1(\cdot)$ and $\tilde{f}_1(\cdot)$. How did `optim()` and your parabolic interpolater fare? Explain what you observe.
Now apply your function to $f_2(x) = \sin(2\pi(x + 1))e^{-(x+1)}$ for a couple of different triples of x_0, x_1, x_2. Plot $f_2(\cdot)$ and $\tilde{f}_2(\cdot)$. How did `optim()` and your parabolic interpolater fare? Explain what you observe. Which function was easier to approximate f_1 or f_2? Explain.