Computer Lab 3
Computational Statistics

Linköpings Universitet, IDA, Statistik

2019/02/01

Kurskod och namn:	732A90 Computational Statistics
Datum:	2019/01/30–2019/02/13 (lab session 1 February 2019)
Delmomentsansvarig:	Krzysztof Bartoszek, Eric Herwin
Instruktioner:	This computer laboratory is part of the examination for the Computational Statistics course Create a group report, (that is directly presentable, if you are a presenting group), on the solutions to the lab as a **.PDF** file. Be concise and do not include unnecessary printouts and figures produced by the software and not required in the assignments. **All R code should be included as an appendix into your report.** A typical lab report should 2-4 pages of text plus some amount of figures plus appendix with codes. In the report reference **ALL** consulted sources and disclose **ALL** collaborations. The report should be handed in via LISAM (or alternatively in case of problems e–mailed to krzysztof.bartoszek@liu.se or Eric Herwin, erihe068@student.liu.se), by **23:59 13 February 2019** at latest. Notice there is a final deadline of **23:59 1 April 2019** after which no submissions nor corrections will be considered and you will have to redo the missing labs next year. The seminar for this lab will take place **8 March 2019**. The report has to be written in English.

Question 1: Cluster sampling

An opinion pool is assumed to be performed in several locations of Sweden by sending interviewers to this location. Of course, it is unreasonable from the financial point of view to visit each city. Instead, a decision was done to use random sampling without replacement with the probabilities proportional to the number of inhabitants of the city to select 20 cities. Explore the file `population.xls`. Note that names in bold are counties, not cities.

1. Import necessary information to R.
2. Use a uniform random number generator to create a function that selects 1 city from the whole list by the probability scheme offered above (do not use standard sampling functions present in `R`).
3. . Use the function you have created in step 2 as follows:
   (a) Apply it to the list of all cities and select one city
   (b) Remove this city from the list
   (c) Apply this function again to the updated list of the cities
   (d) Remove this city from the list
   (e) ... and so on until you get exactly 20 cities.
4. Run the program. Which cities were selected? What can you say about the size of the selected cities?
5. Plot one histogram showing the size of all cities of the country. Plot another histogram showing the size of the 20 selected cities. Conclusions?

Question 2: Different distributions

The double exponential (Laplace) distribution is given by formula:

$$DE(\mu, \alpha) = \frac{\alpha}{2} \exp(-\alpha|x - \mu|)$$

1. Write a code generating double exponential distribution $DE(0, 1)$ from Unif$(0, 1)$ by using the inverse CDF method. Explain how you obtained that code step by step. Generate 10000 random numbers from this distribution, plot the histogram and comment whether the result looks reasonable.
2. Use the Acceptance/rejection method with $DE(0, 1)$ as a majorizing density to generate $\mathcal{N}(0, 1)$ variables. Explain step by step how this was done. How did you choose constant c in this method? Generate 2000 random numbers $\mathcal{N}(0, 1)$ using your code and plot the histogram. Compute the average rejection rate R in the acceptance/rejection procedure. What is the expected rejection rate ER and how close is it to R? Generate 2000 numbers from $\mathcal{N}(0, 1)$ using standard `rnorm()` procedure, plot the histogram and compare the obtained two histograms.