Computational Statistics 732A90 – Fall 2023
Computer Lab 4

Krzysztof Bartoszek, Bayu Brahmantio, Frank Miller, Héctor Rodriguez Déniz
Department of Computer and Information Science (IDA), Linköpings University

November 21, 2023

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 by email), by **23:59 28 November 2023** at latest. Notice there is a deadline for corrections **23:59 21 January 2024** and a final deadline of **23:59 11 February 2024** 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 **29 November 2023**.
The report has to be written in English.

Question 1: Computations with Metropolis–Hastings

Consider a random variabel X with the following probability density function:

$$f(x) \propto x^5 e^{-x}, \quad x > 0.$$

The distribution is known up to some constant of proportionality. If you are interested (**NOT** part of the Lab) this constant can be found by applying integration by parts multiple times and equals 120.

a. Use the Metropolis–Hastings algorithm to generate 10000 samples from this distribution by using a log–normal $LN(X_t, 1)$ proposal distribution; take some starting point. Plot the chain you obtained with iterations on the horizontal axis. What can you guess about the convergence of the chain? If there is a burn–in period, what can be the size of this period? What is the acceptance rate? Plot a histogram of the sample.

b. Perform Part a by using the chi–square distribution $\chi^2(\lfloor X_t + 1 \rfloor)$ as a proposal distribution, where $\lfloor x \rfloor$ is the floor function, meaning the integer part of x for positive x, i.e. $\lfloor 2.95 \rfloor = 2$

c. Suggest another proposal distribution (can be a log normal or chi–square distribution with other parameters or another distribution) with the potential to generate a good sample. Perform part a with this distribution.

d. Compare the results of Parts a, b, and c and make conclusions.

e. Estimate

$$E(X) = \int_0^\infty x f(x) \, dx$$

using the samples from Parts a, b, and c.

f. The distribution generated is in fact a gamma distribution. Look in the literature and define the actual value of the integral. Compare it with the one you obtained.

Question 2: Gibbs sampling

Let $X = (X_1, X_2)$ be a bivariate distribution with density $f(x_1, x_2) \propto \mathbf{1}\{x_1^2 + wx_1x_2 + x_2^2 < 1\}$ for some specific w with $|w| < 2$. X has a uniform distribution on some two-dimensional region. We consider here the case $w = 1.999$ (in Lecture 4, the case $w = 1.8$ was shown).

a. Draw the boundaries of the region where X has a uniform distribution. You can use the code provided on the course homepage and adjust it.

b. What is the conditional distribution of X_1 given X_2 and that of X_2 given X_1?

c. Write your own code for Gibbs sampling the distribution. Run it to generate $n = 1000$ random vectors and plot them into the picture from Part a. Determine $P(X_1 > 0)$ based on the sample and repeat this a few times (you need not to plot the repetitions). What should be the true result for this probability?

d. Discuss, why the Gibbs sampling for this situation seems to be less successful for $w = 1.999$ compared to the case $w = 1.8$ from the lecture.

e. We might transform the variable X and generate $U = (U_1, U_2) = (X_1 - X_2, X_1 + X_2)$ instead. In this case, the density of the transformed variable $U = (U_1, U_2)$ is again a uniform distribution on a transformed region (no proof necessary for this claim). Determine the boundaries of the transformed region where U has a uniform distribution on. You can use that the transformation corresponds to $X_1 = (U_2 + U_1)/2$ and $X_2 = (U_2 - U_1)/2$ and set this into the boundaries in terms of X_i. Plot the boundaries for (U_1, U_2). Generate $n = 1000$ random vectors with Gibbs sampling for U and plot them. Determine $P(X_1 > 0) = P((U_2 + U_1)/2 > 0)$. Compare the results with Part c.