Advanced Computational Statistics – Spring 2023
Assignment for Lecture 6 and 7

Frank Miller, frank.miller@liu.se,
Department of Computer and Information Science, Linköpings University

May 17, 2023

Perform the solutions individually and send your report **until June 7** by email to me. Try to keep this deadline. However, if you have problems with it, there will be a final deadline on September 30 for all assignments. Please include your name in the filename(s) of your solution file(s). Please send me one **pdf-file** with your report (alternatively, Word is ok, too), and additionally, please send me your code in one separate **plain-text file** (an R-markdown, .rmd, is possible but not required).

The following four problems are mandatory.

Problem 6.1

There are several significance tests which can be used to test if a sample is normally distributed:

$$H_0: \text{ sample is } N(\mu, \sigma^2)\text{-distributed for some } \mu \text{ and } \sigma.$$

Your task here is to investigate the Shapiro–Wilk test used with a significance level $\alpha = 0.05$ for a sample size of $n = 200$.

a. Simulate the type I error of the test by generating standard normally distributed samples and applying the Shapiro-Wilk test. Use an appropriate number s of repetitions and give a reason why you have chosen it. Determine the size of the Monte Carlo error for your estimated type I error.

b. Simulate the power of the test if data has a distribution with density $f(x)$ for the different densities below. For this, generate random variables following the target density $f(x)$ with an appropriate method (you can use the available functions in base `R` for generation of random variables). Use a number s of repetitions which gives reasonable precision of the result and argue for that.

   - $f(x) = \frac{8}{\pi}\sqrt{x(1-x)}$ if $0 \leq x \leq 1$ and $f(x) = 0$ for other x (Beta(1.5,1.5)-distribution),
   - $f(x) = \frac{3}{8}x^2$ if $0 \leq x \leq 2^*$ and $f(x) = 0$ for other x,
   - $f(x) = const \cdot \exp(-x^2)(2 + \cos(7\pi x)), x \in I\!R$.

*In an earlier version of this assignment, I wrote by mistake $0 \leq x \leq 1$ which leads not to a probability density; but if you already dealt with this in some other way, you need not to revise.

Problem 6.2

We consider the initial iteration of adaptive rejection sampling (without squeezing). The one-dimensional density $f(x)$ which we want to generate is here the standard normal distribution, i.e. $f(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$.

a. We start with $T_2 = \{x_1, x_2\}$, where $x_2 = -x_1 = t > 0$. Compute an envelope $e^*(x)$ (defined via the tangents in x_i) and $e(x) = \exp(e^*(x))$. Plot $f(x)$ together with $e(x)$ for one or two t.

b. Compute the waste generated by this envelope. Optimize t such that the waste is as small as possible.

c. Use now $T_4 = \{-t_2, -t_1, t_1, t_2\}$ with $0 < t_1 < t_2$ as initial grid. Compute the envelope and the waste. Optimize (t_1, t_2) such that the waste is as small as possible. Plot $f(x)$ together with $e(x)$ for the optimal choice of t_1, t_2.

Problem 7.1

We consider a case when the prior distribution for the mean μ in a normal distribution has the following Laplace density:

$$f_{\text{prior}}(\mu) = \frac{1}{2} e^{-|\mu|}.$$

When normally distributed data is observed, the posterior $f_{\text{posterior}}(\mu)$ has a density proportional to:

$$g(\mu|\bar{x}) = f_{\text{prior}}(\mu) \cdot e^{-(\mu-\bar{x})^2/(2\tau^2)},$$

where $\bar{x}$ is the observed mean and $\tau^2 = \sigma^2/n$ is the variance of the observed mean. Assume $\tau^2 = 1$ and that the observed mean was $\bar{x} = 2.04$. It is of interest here to compute the probability $P(\mu \geq 1) = \int_1^\infty f_{\text{posterior}}(\mu)\, d\mu$ based on the posterior distribution using several techniques as required below.

a. Determine the proportionality constant of the posterior density (you can use a build-in function in R for it) and plot it.

b. Write an own program to compute $\int_1^\infty f_{\text{posterior}}(\mu)\, d\mu$ with Simpson's rule where the number of nodes is iteratively increased until a stop criterion is met. For this, you can truncate the integral at a suitably chosen upper bound. Compute the integral.

c. Write an own program using rejection sampling with an envelope of your choice to simulate draws of the posterior distribution. Compute a Monte Carlo estimate for $P(\mu \geq 1)$.

d. Write an own program using a Metropolis algorithm to generate draws of the posterior. Compute a Monte Carlo estimate for $P(\mu \geq 1)$.

Problem 7.2

Independent observations X in the interval $[0, 1]$ are obtained from a technical process. They are distributed according to following probability density:

$$f(x) = \frac{2}{1 + e^{-10x+5}} \cdot \mathbf{1}\{0 \leq x \leq 1\},$$

where $\mathbf{1}\{\dots\}$ is the indicator function (which is 1 if the following condition is fulfilled and is 0 otherwise). It is here of interest to quantify the probability for observations which are smaller than 0.1, $P(X < 0.1)$, using Monte Carlo simulations.

a. Plot the function f.

b. Choose two different sampling functions g for importance sampling which both have the potential to be good in reducing the variance of the Monte Carlo estimator.

c. After motivating and deciding on a number of repetitions, compute the importance sampling estimator (with non-standardized weights) for $P(X < 0.1)$ and quantify the standard error of the estimator for both sampling functions. How much improvement gives this compared with the standard error from the usual Monte Carlo estimator?

d. Repeat part c. with the importance sampling estimator with standardized weights and compare the standard errors. Estimate also the bias of the estimates.