Math 575 HW 2

Fall 2021

Washington University in St. Louis, University College

Due date: 10/24/2021

Instruction:

Please type your answers clearly and show your work neatly. You are encouraged to use the Rmarkdown version of this assignment as a template to submit your work. Unless stated otherwise, all programming references in the assignment will be in R. For this assignment, problems roughly covers content from lecture 4-6.

Problem 1

Consider the function $g(x) = ln(1+x)$

a. Plot the function g for x from 0 to 10.
b. Evaluate the $\int_0^{10} g(x)dx$ using trapezoid rule (hint: see `lecture4_code.R` for an example). Experiment with different choices of interval legngth (e.g. choose n between 10, 50, 100), report the results.
c. Solve the integral analytically (hint: use integration by parts), compare with results from *b*. Comment on your findings.

Problem 2

Consider a 4-sided dice with faces numbered 1,2,3, and 4. Conduct an experiment of rolling the dice n times. Let X be the event where we roll a '1'.

a. What is the expected value of X, as well as the sample variance and coefficient of variation of the procedure?
b. Using methods illustrated in `lecture5_code.R`, create a series of simulated samples for X, calculate the coefficient of variation as you increase the number of samples.
c. Define a new variable Y, such that $P(Y = 1) = 1/2$ by introducing a 'biased' dice. Use importance sampling to generate a series of simulated samples for X. Note that, in order to get X, we need to apply a correction factor based on outcome of the biased sampler Y.
d. Discuss your results from b) and c). How many samples did it take, respectively, to achieve a coefficient of variation (CV) of less than 5 %?

Problem 3

Suppose we have observed data sampled i.i.d. from a mixture distribution $\delta N(7, (0.5)^2 + (1-\delta)N(10, (0.5)^2$, the data y, can be obtained from the file `mixture.dat`.

a. Using $Beta(2, 10)$ as proposal density, estimate δ using the Metropolis-Hastings algorithm, repeat the experiment with different values of n (e.g. 100, 500, 1000, 5000, etc), where n is the length of the Markov chain generated. (hint: see `lecture6_code.R`)

b. Plot the trajectory of the δ estimates, as well as its distribution.

c. Compare the results with the example shown in class, where a proposal function of $Beta(1, 1)$ is used instead. Comment on your findings.