

STAT4CI3/6CI3 Computational Methods for Inference

Assignment 2

Due at 1:30pm on Monday, February 25, 2019

Instructions:

1. Please indicate **clearly** on your solutions whether you are in STATS 4CI3 or STATS 6CI3.
 2. Non-code parts of the solutions need not be typed but must be readable.
 3. Ensure that all R code is properly commented and attach a print out with your written solution. Also mail your R code as a plain text file to `cantya@mcmaster.ca` using the subject S4CI3 Assignment 1: *<Name> <Student ID>*
 4. Start each question on a new page and submit questions in the same order as given below. **It is important to ensure that all parts of a question are together.**
 5. You are expected to show all details of your solution and any results taken from my notes or elsewhere must be clearly and properly referenced.
 6. No extensions to the due date and time will be given except in extreme circumstances and late assignments without prior approval will not be accepted.
 7. Students are reminded that submitted assignments must be entirely their own work. Submission of all or part of someone else's solution (including solutions from the internet or other sources) under your name is academic misconduct and will be dealt with as such. Penalties for academic misconduct can include a 0 for the assignment, an F for the course with an annotation on your transcript and/or dismissal from your program of study.
 8. Questions 4(c) and 5 marked [**STAT6CI3**] are for the graduate students only.
-

Q. 1 Consider the following rejection algorithm

1. Generate 2 independent $\text{Unif}(0, 1)$ random variables U_1 and U_2 .
 2. Let $X = -\ln U_1$ and $Y = -\ln U_2$.
 3. If $2X \geq (Y - 1)^2$ return Y .
- a) Prove that X and Y are independent $\exp(1)$ random variables.
- b) Use Bayes Theorem to prove that

$$f_Y(y|2X \geq (Y - 1)^2) \propto e^{-y^2/2} \quad \text{for } y > 0.$$

- c) Using a transformation show that

$$\int_0^\infty e^{-y^2/2} dy = \frac{\Gamma(\frac{1}{2})}{\sqrt{2}} = \sqrt{\frac{\pi}{2}}$$

and hence give the density of the accepted observations.

- d) Suppose that Y is generated using this method and U_3 is another $\text{Unif}(0, 1)$ random variable. Define the random variable Z as

$$Z = \begin{cases} Y & \text{if } U_3 > 0.5 \\ -Y & \text{if } U_3 \leq 0.5 \end{cases}$$

Show that Z is a standard normal random variable.

- e) Write an R function which will generate n random standard normal observations using this technique.

Q. 2 a) Using only $\text{Unif}(0, 1)$ random variates, use a Monte Carlo algorithm to approximate the value of the Gamma function

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$$

by considering the function as an expectation of a function of a random variable.

- b) Show that if a Monte Carlo simulation of size N is used then the variance of the Monte Carlo estimator is

$$\text{Var}(\hat{\Gamma}(\alpha)) = \frac{1}{N} [\Gamma(2\alpha - 1) - [\Gamma(\alpha)]^2].$$

provided that $\alpha \geq 0.5$.

- c) Write an R function to implement the method returning the estimated value of the function and the standard error of the estimation. Examine how well the method works for various values of α and various simulation sizes.

Q. 3 Suppose that we wish to use Monte Carlo methods to estimate the integral

$$I = \int_0^1 e^{x^2} dx$$

In the following questions write an R function to estimate this quantity based on uniform(0,1) random variates only. Your functions should return both the estimate and the standard error of the estimate and you should present the estimates and standard errors using $N = 100000$. Please use your student number to set the random seed (using `set.seed`) for each part of the question.

- a) Implement the basic Monte Carlo method by writing I as an expected value.
- b) Implement a control variable estimate using the average of squared uniform(0, 1) random variables as the control covariate. In your solution you should estimate the best β to use based on the output.
- c) Implement an antithetic variable method to estimate I based on a sample of uniform(0, 1) random variables and the fact that if $U \sim \text{uniform}(0, 1)$ then $1 - U \sim \text{uniform}(0, 1)$.
- d) Implement an importance sampling estimator using a observations from a Beta($\alpha, 1$) density. Give the value of α you use and describe how you selected that value.

Q. 4 A *large-sample* $100(1 - \alpha)\%$ confidence interval for the mean is given by

$$\left[\bar{x} - z_{\alpha/2} \frac{s}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \frac{s}{\sqrt{n}} \right]$$

where $\bar{x}$ and s are the sample mean and standard deviation respectively and $z_{\alpha/2}$ is the upper $\alpha/2$ quantile of the standard normal distribution. The coverage of the confidence interval is defined to be the probability that the interval contains the true value μ . For the following questions suppose that $\alpha = 0.05$.

- a) Show (analytically) that, for a fixed n , the coverage probability does not depend on μ and σ if the data come from a normal(μ, σ^2) distribution. Use a simulation study to estimate the coverage of the interval when the data truly comes from a normal distribution and examine how the coverage changes as the sample size changes. Use your simulation study to say how large a sample would you need to get within one percentage point of the nominal coverage?
- b) Suppose now that the data comes from an exponential distribution with mean μ . Show that the coverage probability does not depend on $\mu > 0$. Implement a simulation study to examine how the coverage changes with n . Estimate the probability that the confidence interval includes negative values of μ . How large a sample would you need to get within one percentage point of the nominal coverage?

- c) [STAT6CI3] Now suppose that the underlying data is log-normal. That is $X = \exp\{Y\}$ where Y is a normal(μ, σ^2) random variable. Find the mean of this distribution. Show that the coverage probability does not depend on μ but does depend on σ . Use simulations to estimate the coverage probability and the probability that the interval will contain negative values of the true mean for different values of σ and increasing sample sizes.

Q. 5 [STAT6CI3] Prove Theorem 10 in my notes as follows.

- a) Show that for an arbitrary sampling density g , with the same support as f , the variance of the importance sampling estimator is

$$\text{Var}(\hat{I}_{IS}) = \frac{1}{N} \left\{ \int \frac{h^2(x)f^2(x)}{g(x)} dx - I^2 \right\}$$

- b) Show that when the sampling density g is given by

$$g(x) \propto |h(x)|f(x)$$

this variance reduces to

$$\text{Var}(\hat{I}_{IS}) = \frac{1}{N} \left\{ \left[\int |h(x)|f(x) dx \right]^2 - I^2 \right\}$$

and prove that this variance is less than or equal to the variance in (a).

Hint: For any random variable Y , $\text{Var}(Y) = E(Y^2) - (E(Y))^2 \geq 0$.