General Notes

- You will submit a minimum of two files, the core files must conform to the following naming conventions (including capitalization and underscores). 123456789 is a placeholder, please replace these nine digits with your nine-digit Bruin ID. The files you must submit are:

  1. *123456789_stats102c_hw1.Rmd*: Your markdown file which generates the output file of your submission.
  2. *123456789_stats102c_hw1.html/pdf*: Your output file, either a PDF or an HTML file depending on the output you choose to generate.

  If you fail to submit any of the required core files you will receive ZERO points for the assignment. If you submit any files which do not conform to the specified naming convention, you will receive (at most) half credit for the assignment.

- **Your .Rmd file must knit**. If your .Rmd file does not knit you will receive (at most) half credit for the assignment.
  The two most common reason files fail to knit are because of workspace/directory structure issues and missing include files. To remedy the first, ensure all of the file paths in your document are relative paths pointing at the current working directory. To remedy the second, simply make sure you upload any and all files you source or include in your .Rmd file.

- Your coding should adhere to the tidyverse style guide: https://style.tidyverse.org/.

- Any functions/classes you write should have the corresponding comments as the following format.

```
my_function <- function(x, y, ...){
#A short description of the function
#Args:
#x: Variable type and dimension
#y: Variable type and dimension
#Return:
#Variable type and dimension
Your codes begin here
}
```

NOTE: *Everything* you need to do this assignment is here, in your class notes, or was covered in discussion or lecture.

- Please **DO NOT** look for solutions online.
- Please **DO NOT** collaborate with anyone inside (or outside) of this class.
- Please work **INDEPENDENTLY** on this assignment.
- **EVERYTHING** you submit **MUST** be 100% your, original, work. Any student suspected of plagiarizing, in whole or in part, any portion of this assignment, will be **immediately** referred to the Dean of Student's office without warning.

Problem 1: Suppose $\theta = \int_0^1 2e^{-2x}dx$. Please write corresponding algorithms and functions to answer (a) - (d).

(a) Compute the Monte Carlo estimate of θ without any means of variance reduction.

(b) Compute the Monte Carlo estimate of θ using the antithetic variate approach.

(c) Compute the Monte Carlo estimate of θ using the control variate approach.

(d) Compute the Monte Carlo estimate of θ using Stratified sampling.

(e) Compare your four estimates with the theoretical value, and compute their variances.

Problem 2: Suppose X $\sim f(x)$. Let $\theta = \int g(x)f(x)dx = E_f[g(X)]$. We draw m iid copies $X_1, ..., X_m$ from $h(x)$, which is different from $f(x)$, and define $W(x) = \frac{f(x)}{\phi(x)}$

(a) Prove $E_f[g(X)] = E_\phi[g(X)W(X)]$

(b) Prove $E_\phi[W(X)] = 1$

(c) Let $\hat{\theta} = \sum_{i=1}^m g(X_i)W(X_i)/m$. Find $E[\hat{\theta}]$ and $Var[\hat{\theta}]$.

Problem 3: Given $X \sim N(0,1)$, we want to compute $\theta = P(X > C)$ where C is a positive constant.

(a) Find three importance functions that are supported on $(0, \infty)$, and explain which of your importance functions should produce the smaller $Var[\hat{\theta}]$. Please graph plots to support your answer.

(b) Write a function to compute a Monte Carlo estimate of θ using importance functions you proposed in (a).
Note: You can use the built-in functions in R to generate random variables.

(c) Compare your estimates with the theoretical values for $C = 0.5, 1, 2, 3$.

Problem 4: The density function $f(x)$ is proportional to $q(x) = e^{-x^2\sqrt{x}}[\sin(x)]^2$, for $x \in \mathbb{R}_+$. Consider a trial distribution $h(x) = \frac{r(x)}{Z_r}$. We want to use importance sampling to estimate $E_f(X^2)$. Consider the following choices of un-normalized densities for trial distributions:

(i) $r_1(x) = e^{-2x}$,

(ii) $r_2(x) = x^{-1/2}e^{-x/2}$,

(iii) $r_3(x) = (2\pi)^{-1}(1 + x^2/4)^{-1}$,

for $x \in \mathbb{R}_+$.

(a) Write three functions (E1, E2, and E3) to estimate $E_f(X^2)$ using importance sampling with the above un-normalized densities.

(b) Compare your estimates with the theoretical value.

(c) Which trial distribution above is more efficient? Explain.

(d) Estimate the normalizing constant of $q(x)$.