

STAT 3006 Assignment 2

Due date: 5:00 pm on 16 March

(25%)**Q1:** Please use the inverse method to generate 5000 samples from $\text{Poisson}(\lambda = 10)$ and draw a histogram for the 5000 samples. The probability mass function of $\text{Poisson}(\lambda)$ is

$$P(X = m) = e^{-\lambda} \frac{\lambda^m}{m!}, \quad m = 0, 1, 2, 3, \dots$$

(25%)**Q2:** Please use the accept-reject method to generate 5000 samples from $f(x) = \frac{\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \cdot I(x \geq 3)}{1 - \Phi(3)}$, in which $g(x) = 3e^{-3(x-3)}I(x \geq 3)$ and $M = \frac{e^{-4.5}}{3\sqrt{2\pi}(1-\Phi(3))}$, and prove that $f(x) \leq M \cdot g(x)$.

(25%)**Q3:** Please estimate the integration $\int_3^{+\infty} \sin(x)e^{-x^2/2}dx$:

- (a) using 5000 samples from Q2;
- (b) using importance sampling (drawing 5000 samples from $g(x) = 3e^{-3(x-3)}I(x \geq 3)$ to estimate the integration).

(25%)**Q4:** Salary data set contains all persons' salary in a county. Your goal is to use stratified sampling to estimate this county's salary mean. In the data set, the first column represents salary per month for each person and the second column indicates the age range. "1" denotes the age interval (20, 30], "2" denotes (30, 50], and "3" denotes (50, 70]. There are total 11,000 persons. 2000 persons have age indicator "1", 5000 persons have age indicator "2", 6000 persons have age indicator "3".

- (a) Randomly draw 100 samples from the data set, and estimate the standard deviation for each subpopulation.
- (b) If the total sample number is set as 1000, please determine the sample number for each subpopulation.
- (c) Estimate population salary mean based on these 1000 samples and compare it to 4166.165 (underlying truth of the population salary mean).

Requirements:

-	in the paper report	in the R code file
Q1	algorithm histogram	R code
Q2	algorithm proof	R code
Q3	algorithms integration values for the two approaches	R code
Q4	standard deviation sample number population salary mean estimate	R code