STA380 Term Test 2 Practice Version B

Solutions to the practice tests will not be posted. Students are encouraged to ask questions on Piazza or during office hours; however, they should include evidence of their own attempts when doing so. The primary purpose of providing the practice test is to help students become familiar with the format, length, and style of the questions, as well as to offer an opportunity for additional practice. Several questions are directly from the lecture slides, practice problems, and quizzes. The difficulty of the practice test and the actual test may differ.

Question 1

Consider the Pareto distribution:

$$f_X(x) = \frac{\alpha\beta^\alpha}{x^{\alpha+1}}, \quad x > \beta.$$

Let X be a random variable from a Pareto distribution where $\alpha = 1, \beta = 2$. Note that α is supposed to represent the shape parameter, and β is supposed to represent the scale parameter.

(a) **[2 points]** Use the hit or miss method to compute $\mathbb{P}(X < 3)$. Use $n = 10^5$. To generate a random sample from the Pareto distribution, one could use the `rpareto(n, scale = 1, shape)` function from `VGAM` package, where `n` represents the number of observations, `scale` represents the scale parameter, and `shape` represents the shape parameter.

You may assume that the `VGAM` package is installed but has not been loaded.

(b) **[2 points]** The Monte Carlo estimate for the variance of an estimator is:

$$\widehat{\mathbb{V}ar[\hat{\theta}]} = \frac{\hat{\sigma}^2}{n} = \frac{\sum_{i=1}^n \left[g(x_i) - \overline{g(x)}\right]^2}{n^2}$$

Use this to construct a 95% confidence interval of this estimate in `R`. You can return the answer as a vector, where the first value represents the lower confidence bound, and the second represents the upper confidence bound. **You do not need to write the algorithm.**

Question 2

Estimate

$$\theta = \int_{\mathbb{R}} \cos(x) \frac{2}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\} dx$$

using importance sampling, where the importance function is the pdf of the Normal distribution. Assume that you need $n = 10^4$ replicates. You may select any choice of μ and σ.

To generate a random sample from the Normal distribution, one could use the `rnorm(n, mean = 0, sd = 1)` function, where `n` represents the number of observations, `mean` represents the mean, and `sd` represents the standard deviation.

(a) **[4 points]** Write the algorithm using importance sampling. **Show your work.**

(b) **[2 points]** Write out the `R` code for the algorithm using importance sampling.

Question 3

Consider the doubly truncated exponential distribution that is between T and T^* (for $T, T^* > 0$):

$$f_X(x) = \frac{\theta^{-1}e^{-x/\theta}}{e^{-T/\theta} - e^{-T^*/\theta}}, \quad T < x < T^*.$$

For now, consider the case where $T = 1$ and $T^* = 3$.

(a) **[2 points]** What is the true value of $\int_1^2 f_X(x)dx$?

(b) **[1 point]** Is it possible to use the antithetic approach to estimate $\int_1^2 f_X(x)dx$? Why or why not?

(c) **[2 points]** Assuming it is possible, write the algorithm to estimate $\int_1^2 f_X(x)dx$ using the antithetic variable approach. Let $n = 10^4$. Use any value of $\theta > 0$ that you desire.

(d) **[2 points]** Write out the `R` code for the algorithm using the antithetic variable approach.

Question 4

[6 points] Suppose we want to compare two estimators, $\hat{\theta}_1$ and $\hat{\theta}_2$, compared to a common parameter θ. Recall the MSE:

$$MSE(\hat{\theta}) = \mathbb{E}[(\hat{\theta} - \theta)^2] = \mathbb{V}[\hat{\theta}] + Bias(\hat{\theta})^2$$

Suppose $X_1, X_2, \ldots, X_n$ are i.i.d. $N(\theta, \theta^2)$. We want to estimate θ. We have two possible candidates:

$$\hat{\theta}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad \hat{\theta}_2 = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

Write the appropriate `R` code to compute the MSE of these two estimators. Use $\theta = 5$, sample size of $n = 50$, and $m = 10^4$ number of replicates.

Question 5

Let $X_1, X_2, \ldots, X_{30}$ be a random sample from the *Exponential*(θ) distribution. The goal of this question is to simulate a hypothesis test. Consider the hypothesis test:

$$H_0 : \theta = 3 \quad \text{v.s.} \quad H_1 : \theta < 3$$

Someone begins simulating values using the following code:

```
n <- 30
alpha <- 0.05
theta <- 2
m <- 10^4
x <- matrix(rexp(n*m, rate = 1/theta), nrow = m)
```

The same person found the following **rejection region** (RR):

$$RR := \left\{ \sum_{i=1}^{30} X_i < k \right\}$$

(a) **[1 point]** Using `R`, find what the critical value k should be given that the significance level is 0.05.
Hint: the formula sheet includes `qexp()`, `qgamma()`, and `qbeta()`. Potentially one or more could be useful.

(b) **[1 point]** Using `R` and the code present in the question, compute the test statistic. *Hint:* any one of these could be useful: `colSums()`, `rowSums()`, `colMeans()`, `rowMeans()`.

(c) **[1 point]** Using `R`, compute the Monte Carlo type-I error rate.

Question 6

[1 point] Which one of the following statements are correct?

(a) A p-value below 0.05 guarantees a practically important result.

(b) The p-value measures the probability that the researcher made a mistake.

(c) If the null hypothesis is exactly true, statisticians are morally obligated to reject it eventually.

(d) The primary purpose of hypothesis testing is to confirm results we already believe.

(e) None of the above.