STA380 Term Test 2 Practice Version A

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 $X \sim \textit{Exponential}(\theta = 2)$.

(a) [**2 points**] Use the hit or miss method to compute $\mathbb{P}(X < 3)$. You may use `rexp()` and $n = 10^5$. **You do not need to write the algorithm.**

In `rexp(n, rate)`:

- `n` represents the number of observations.
- `rate` represents the rate parameter of the Exponential distribution. Note that rate $= \frac{1}{\text{scale}}$.

(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 99% 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

$$\theta = \int_{\mathbb{R}} \sin(x)\frac{\sqrt{2}}{3\sqrt{\pi}}\exp\left\{-\frac{x^2}{18}\right\}dx$$

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 left-truncated exponential distribution at T for $T > 0$:

$$f_X(x) = \frac{1}{\theta}\exp\left\{\frac{-x+T}{\theta}\right\}, \quad T < x < \infty.$$

(a) [**2 points**] What is the true value of $\int_T^{T+1} f_X(x)dx$?

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

(c) [**2 points**] Assuming it is possible, write the algorithm to estimate $\int_T^{T+1} f_X(x)dx$ 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. Then, the PC of $\hat{\theta}_1$ relative to $\hat{\theta}_2$ is:

$$\pi_{\hat{\theta}_1,\hat{\theta}_2}(\theta) := \mathbb{P}(|\hat{\theta}_1 - \theta| \leq |\hat{\theta}_2 - \theta|).$$

If $\pi_{\hat{\theta}_1,\hat{\theta}_2}(\theta) > \frac{1}{2}$ then $\hat{\theta}_1$ is said to be **Pitman closer** to θ, implying it is a more desirable estimator.

Suppose $X_1, X_2, \dots, X_n$ 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 PC probability of these two estimators. Use $\theta = 1$, sample size of $n = 10$, and $m = 10^4$ number of replicates.

Question 5

Let $X_1, X_2, \dots, X_{20}$ 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 <- 20
alpha <- 0.05
theta <- 5
m <- 10^4
x <- matrix(rexp(n*m, rate = 1/theta), nrow = m)
```

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

$$RR := \left\{ \sum_{i=1}^{20} X_i > k \right\}$$

(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) All statisticians love setting up hypotheses with point nulls and unanimously agree that they reflect realistic scientific questions.

(b) Traditional statistics education heavily emphasises hypothesis testing and p-values.

(c) A small p-value implies that the null hypothesis is unlikely to be true.

(d) Point null hypotheses arise naturally in most applied scientific settings.

(e) If the null hypothesis is false, then the p-value must be close to zero.