STAT 2593 - Problem Set 2
Due: Wednesday, November 22 on Crowdmark
Resubmit: Wednesday, December 6 on Crowdmark

Reminder that you are permitted to discuss to these problems with classmates, but every student must submit their own solutions which are their own work. **Please indicate any students that you discussed solutions with on your submission**. Please ensure that your submissions on Crowdmark are legible, and separated based on the problems included at the submission link. Submissions can be handwritten or typeset.

Part 1: True or False (20 Marks; 2 Marks Each)

For each of the following problems indicate whether the statement is true or false, and give a short justification for your answer. Correct answers without justification will receive only partial credit.

Problem 1: True or False: it is possible to have a 100% confidence interval for p, the probability of success from a binomial random variable.

Problem 2: Two estimators are being compared for the same parameter. The first estimator has a bias of 0 with a variance of 25. The second has a bias of -2 and a variance of 16. **True or False:** the second estimator is always preferred since it has a lower MSE.

Problem 3: Two estimators are being compared for the same parameter. The first estimator has a bias of 0 with a variance of 25. The second has a bias of -2 and a variance of 16. **True or False:** it is possible that the first estimator is the MVUE.

Problem 4: A sample, $y_1, \dots, y_n$ is recorded from a population with a density function, $f(x;\theta)$, where θ is an unknown parameter. In order to better understand the population, researchers consider the quantities $A = \sum_{i=1}^n \exp\left(-y_i^2\right)$ and $B = \sum_{i=1}^n \theta \cdot \frac{y_i}{\log(y_i)}$. **True or False:** both A and B are statistics.

Problem 5: Suppose that two light bulbs have a lifespan which follows an exponential distribution with rate $\lambda = \frac{1}{500}$, independently of one another. Out of curiosity for which one will last longer it is decided that an experiment will be run where each light bulb will be turned on and watched until one of them has gone out. When the experiment was set-up, the first light bulb (bulb A) was turned on successfully, but the second light bulb (bulb B) was not, unbeknownst to the investigators. After 500 hours, the investigators returned to the room, realized that bulb B had never been turned on, and turn it on then; light bulb A is still running at this point. **True or False:** Bulb B is more likely than bulb A to remain on for an additional 500 hours, since bulb A has already ran for 500 hours.

Problem 6: A sample of size 1000 is drawn from a Bernoulli distribution with success probability p, and the sample mean is used to estimate the probability of success, using the normal approximation. **True or False:** It is possible that a correctly calculated 95% confidence interval for p is given by $[0.649, 0.7209]$.

Problem 7: The following data points were observed from a sample of 10.

```
-0.842 -0.829 -0.652 -0.543 -0.342 -0.295 -0.284  0.518  0.764  0.925
```

A probability plot is constructed to compare these data to a $N(0,1)$ distribution. **True or False:** The point $(1.645, 0.925)$ will appear on the probability plot.

Problem 8: **True or False:** To find the value of $Z_{0.0375}$, we can take

$$\frac{1}{2}\left(Z_{0.05} + Z_{0.025}\right).$$

Problem 9: Suppose that a manufacturer measures the sizes of each item in a batch of products that they have produced. With these measurements they form a 95% confidence interval for the mean length, denoted $[a, b]$. **True or False:** The probability that the true mean falls in the interval $[a, b]$ is 0.95.

Problem 10: Suppose that X follows a Gamma distribution with parameters α and β. **True or False:** The quantity

$$Z = \frac{X - \alpha\beta}{\sqrt{\alpha\beta^2}},$$

follows a $N(0,1)$ distribution.

Part 2: Conceptual Question (35 Marks)

For the following questions, provide your answers with justification and clear communication. The answers do not need to be long, but correct responses without complete justification will receive only partial credit.

Problem 11: (**5 Marks**) A continuous random variable has a probability density function given by

$$f(x;\alpha) = (\alpha+1)x^{\alpha}, \quad 0 < x < 1.$$

(a) For what values of α is this a valid PDF? Explain.

(b) Find $P(X \leq 1/2)$, the CDF of X, and $E[X]$.

Problem 12: (6 Marks) A manufacturer makes wax cylinders for use in an industrial process. In order to be usable, the produced cylinders must have a length which is at least 13.27 and no more than 13.47. Similarly, the diameter must fall between 4.23 and 4.30. The process runs such that the length of the cylinders are normally distributed with a mean of 13.36 and a standard deviation of 0.05, while the diameters follow a $N(4.26, 0.000225)$ distribution. The lengths and diameters are independent of one another.

(a) What is the probability that a cylinder has an acceptable length?

(b) What is the probability that a cylinder has an acceptable diameter?

(c) What is the probability that a cylinder cannot be used?

(d) What is the probability that, in a sample of 5 cylinders, four or fewer are usable?

Problem 13: (4 Marks) Suppose that we have a sample of size n drawn from a normal population, $X_1, \ldots, X_n$ with mean μ and variance σ^2. Suppose that we wish to estimate σ^2 using an estimator of the form:

$$\widehat{\theta}(c) = c\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

Note, the first variance estimator we looked at took $c = 1/n$ and the standard unbiased estimator took $c = 1/(n-1)$. Find the c which produces the best estimator in terms of MSE. **Hint:** The variance of S_{XX} is given by $2(n-1)\sigma^4$.

Problem 14: (5 Marks) Suppose that the magnitude of earthquakes (measured on the Richter scale) follow an exponential distribution, with a mean of 2.5.

1. What is the probability that an earthquake exceeds 5 on the Richter scale?
2. Suppose that 3 earthquakes occur in a month, independently of one another. What is the probability that at least 1 exceeds 5 on the Richter scale?

3. If an earthquake exceeds 4 on the Richter scale, what is the probability that it exceeds 5?

Problem 15: (7 Marks) Suppose that a sample of values, $x_1, \ldots, x_k$ is drawn from a $\text{Bin}(n, p)$ distribution, with unknown parameters n and p.

(a) Find the method of moments estimator for n.

(b) Suppose that $k = 8$, that the sample mean was 7 and the sample variance was 6. What is an estimate for n?

(c) Does this seem like a reasonable estimator in all scenarios? Why or why not?

Problem 16: (8 Marks) Suppose that individuals are being interviewed to understand Canadian's opinions regarding sensitive, political topics. Each individual interviewed is asked a series of questions, and they respond to each with either "agree" or "disagree". The proportion of Canadians who agree with the question, in actual fact, is called the "support rate".

(a) Suppose that $n = 1000$ people are interviewed. If the first question asked has a true support rate of 0.55, what is the approximate probability that between 500 and 570 or more than 598 people interviewed report agreement? Report your answer in terms of Φ.

(b) Suppose that $n = 1000$ people are interviewed. If the second question asked has a true support rate of 0.75, what is the approximate probability that the sample mean (of the Bernoulli responses) exceeds 0.7768?

(c) Suppose that the interviewers want to be able to provide an approximate 95% confidence interval that has a width not exceeding 0.02 around the proportion of agreement with these questions. How many individuals need to be interviewed to ensure that this is possible?

Part 3: Knowledge Extension (20 Marks)

Please note, the following question is designed to expand on the concepts covered in the course. While you have all of the tools required to solve this question, it requires deeper thinking than the previous conceptual questions.

Problem 17: (20 Marks) Exponential random variables are often used to model the lifetimes of different products or components.

(a) Suppose that we have two different components, A and B, each with a lifetime that follows an exponential distribution with rates λ_A and λ_B, respectively. If our system correctly operates supposing that either A or B is still operating, what the is the probability that the system remains operational at time T?

(b) Suppose that the system now consists of k components operating in parallel, so that as long as any one of the components is operational, the whole system is. What is the CDF and PDF for the length of time that the system operates, if each component has the same rate, λ?

(c) Suppose now that the k components from part (b) are connected in series so that in order for the system to operate, **all** components need to remain operational. What is the distribution for the lifetime of the system? How long do you expect the system to run?

(d) How does your answer change in part (c) if the components are now different, with the j-th component having a rate of λ_j?

(e) Using the formulation in part (d), what is the probability that it is the j-th component (with rate λ_j) which causes the system to fail?