Instruction: You are required to do questions 1(a)(b)(c), 2(a)(b), 3(a)(b), and 4(a)(b)(d). The question 4(c) is a bonus question worth of 10 points. However, your total score will not be over 100 points if you did really well in other questions. Questions 1(d), 2(c), and 3(c) are take-home questions for those who want to get extra credits. However, doing these questions will not move your grade from P to H.

1. Let $X_1, \dots, X_n$ be a random sample from a normal distribution with unknown mean μ and *known* variance σ^2. To test $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$, where μ_0 is a given number:

   (a) Derive the likelihood ratio test statistic $\lambda(x)$ and show that the critical region can be written as $\{x : |\bar{x} - \mu_0| \geq c^*\}$ with some constant c^*.

   (b) Find the c^* in (a) such that the test is a size α test.

   (c) Sketch the power function $\beta(\mu)$ as a function of μ, where $-\infty < \mu < \infty$. Prove or disprove that the test with the critical region in (b) is the uniformly most powerful (UMP) test.

   (d) [**TAKE HOME**] Provided that σ^2 is *unknown*, derive the likelihood ratio test statistic and find the equivalent critical region with size α using a test statistic with a well-known distribution.

2. Let $X_1, \dots, X_n$ be a random sample from a distribution with pdf $f(x|\theta) = 1/\theta$ for $0 < x < \theta$, and zero otherwise.

   (a) Show that the maximum likelihood estimator is the maximum order statistic $X_{(n)}$, and prove that it is a *biased* estimator of θ under finite n but a *consistent* estimator when $n \to \infty$ by showing that

   $$\lim_{n\to\infty} E(X_{(n)}) = \theta, \quad \text{and} \quad \lim_{n\to\infty} \mathrm{Var}(X_{(n)}) = 0.$$

   [By showing the limiting properties above, one can claim an estimator is consistent by Theorem 10.1.3 in C&B].

   (b) Find the uniformly most powerful (UMP) size α test when testing $H_0 : \theta = \theta_0$ against $H_1 : \theta > \theta_0$ with specification for the cutoff in the critical region.

   (c) [**TAKE HOME**] Based on the answer in (a), find an unbiased estimator of θ and show that its variance is smaller than the Crámer-Rao lower bound (CRLB). Comment on why CRLB fails in this situation.

3. Let $X_1, \dots, X_n$ be an independent sample that were drawn from $N(i\theta, \sigma^2)$ for X_i with unknown θ and $\sigma^2 > 0$. Define

   $$Q = \frac{\sum_{i=1}^n iX_i}{\sum_{i=1}^n i^2} \quad \text{and} \quad S^2 = \frac{\sum_{i=1}^n (X_i - iQ)^2}{(n-1)}.$$

(a) Show that $T = \sqrt{\sum_{i=1}^n i^2}(Q-\theta)/S$ is a pivotal quantity for θ by showing that Q is a normally distributed random variable with mean θ and variance $\sigma^2/\sum_{i=1}^n i^2$ and $(n-1)S^2/\sigma^2$ follows a chi-square distribution with degree of freedom $(n-1)$.

(b) Construct a $(1-\alpha)$ confidence interval for θ using the pivotal quantity T.

(c) [**TAKE HOME**] Find a pivotal quantity for σ^2 and construct a $(1-\alpha)$ confidence interval based on the quantity.

4. The time X (in months) in remission for leukemia patients who have completed a certain type of chemotherapy treatment is assumed to have the [negative] exponential distribution

$$f_X(x|\theta) = \theta e^{-\theta x}, \quad x > 0, \quad \theta > 0.$$

Let $X_1, \ldots, X_n$ represent a random sample of size n from $f_X(x|\theta)$, and let $\hat{\theta}$ denote the MLE of θ based on $X_1, \ldots, X_n$.

(a) To test the hypothesis $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$, derive the explicit expression of the large-sample likelihood ratio test, score test, and Wald test statistics. Specify the critical region for each test with size α.

(b) A biostatistician realizes that it is not possible to know the *exact* number of months that each patient is in remission after completing the chemotherapy treatment. She suggests that one should set a specific time period (in months) of length x^* (a known positive constant) and analyze the data based on dichotomized random variables $Z_1, \ldots, Z_n$, where $P(Z_i = 1) = P(X_i > x^*)$, $i = 1, \ldots, n$. Find the alternative maximum likelihood estimator $\tilde{\theta}$ of θ based on $Z_1, \ldots, Z_n$ and its large sample distribution in an explicit form.

(c) [**BONUS**] Assuming x^* is the expected length of time in remission, compare the asymptotic variance of $\hat{\theta}$ and $\tilde{\theta}$. If one is smaller than the other, comment on why this should be the anticipated finding.

(d) Suppose there is a new type of chemotherapy that is claimed to have a better mean length of time in remission for leukemia patients. To test this hypothesis, investigators intend to collect another independent sample of size m applying this new treatment and assume that the length of time in remission $Y_1, \ldots, Y_m$ follows another [negative] exponential distribution

$$f_Y(y|\beta) = \beta e^{-\beta y}, \quad y > 0, \quad \beta > 0.$$

Derive a large sample Wald test with size α for the hypothesis $H_0 : \theta \leq \beta$ versus $H_1 : \theta > \beta$, using the test statistic $(\hat{\theta} - \hat{\beta})$, where $\hat{\beta} = \bar{Y}^{-1}$ and $\bar{Y} = m^{-1}\sum_{i=1}^m Y_i$.