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 $X_1, \ldots, X_n$ 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^*.

   ¶ Under H_0, μ_0 is a fixed number so the numerator of $\lambda(x)$ is $L(\mu_0)$. Under the overall parameter space, the MLE is $\bar{X}$ so the denominator of $\lambda(x)$ is $L(\bar{x})$. The LRT statistic is

   $$\begin{aligned}\lambda(x) &= \frac{(2\pi)^{-n/2}\exp\{-\sum_{i=1}^n (x_i-\mu_0)^2/(2\sigma^2)\}}{(2\pi)^{-n/2}\exp\{-\sum_{i=1}^n (x_i-\bar{x})^2/(2\sigma^2)\}} \\ &= \exp\left[\left\{-\sum_{i=1}^n (x_i-\mu_0)^2 + \sum_{i=1}^n (x_i-\bar{x})^2\right\}/(2\sigma^2)\right].\end{aligned}$$

   Since $\sum_{i=1}^n (x_i-\mu_0)^2 = \sum_{i=1}^n (x_i-\bar{x})^2 + n(\bar{x}-\mu_0)^2$, the LRT statistic can be simplified to

   $$\lambda(x) = \exp\{-n(\bar{x}-\mu_0)^2/(2\sigma^2)\}.$$

   The rejection region of LRT, $\{x : \lambda(x) \leq c\}$, hence can be written as

   $$\{x : |\bar{x} - \mu_0| \geq c^* = \sqrt{-2\sigma^2(\log c)/n}\}.$$

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

   ¶ Under the null hypothesis,

   $$\alpha = P\left(\left|\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}\right| \geq z_{1-\alpha/2} | \mu = \mu_0\right),$$

   one can have $c^* = z_{1-\alpha/2}\sigma/\sqrt{n}$.

(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.

¶ The power function is a symmetric convex curve with the lowest value α at $\mu = \mu_0$ and the highest value 1 at $+\infty$ and $-\infty$. This test is not the UMP test since the power of the test is lower than a test with a critical region $\{x : \bar{x} - \mu_0 \geq z_{1-\alpha}\sigma/\sqrt{n}\}$ when $\mu > \mu_0$.

(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 $X_1, \ldots, X_n$ 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

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

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

¶ The likelihood function is

$$L(\theta) = \prod_{i=1}^{n} f(x_i|\theta) I(0 < x_i < \theta) = \theta^{-n} I(0 < x_{(n)} < \theta),$$

which is a decreasing function of θ with positive values only when $\theta > x_{(n)}$ and 0 otherwise. Hence the function is maximized at $\theta = x_{(n)}$. We can claim that the MLE of θ is $X_{(n)}$. Moreover, since the CDF of $X_{(n)}$ is

$$F_{X_{(n)}}(y) = P(X_{(n)} \leq y) = \{P(X \leq y)\}^n = \frac{y^n}{\theta^n},$$

the the expectation of $X_{(n)}$ is

$$E(X_{(n)}) = \int_0^\theta y dF_{X_{(n)}}(y) = \frac{n}{\theta^n} \frac{1}{n+1} \left(y^{n+1}\right)_0^\theta = \frac{n}{n+1}\theta,$$

and the variance of $X_{(n)}$ is

$$\mathrm{Var}(X_{(n)}) = E(X_{(n)}^2) - E(X_{(n)})^2 = \frac{n}{(n+2)(n+1)^2}\theta^2.$$

Hence, one can claim that $X_{(n)}$ is a biased estimator under finite n. However, when $n \to \infty$, one can easily see $\lim_{n\to\infty} E(X_{(n)}) = \theta$ and $\lim_{n\to\infty} \mathrm{Var}(X_{(n)}) = 0$. One can claim $X_{(n)}$ is a consistent estimator of θ using 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.

¶ According to the Neyman-Pearson lemma, one can find the UMP test by having a critical region

$$\left\{ x : \frac{f(x|\theta_1)}{f(x|\theta_0)} > c \right\},$$

with some constant c, where $\theta_1 > \theta_0$. Since the ratio can be written as

$$\frac{f(x|\theta_1)}{f(x|\theta_0)} = \frac{\theta_1^{-n} I(0 < x_{(n)} < \theta_1)}{\theta_0^{-n} I(0 < x_{(n)} < \theta_0)},$$

one can have

$$\frac{f(x|\theta_1)}{f(x|\theta_0)} = \begin{cases} \left(\frac{\theta_0}{\theta_1}\right)^n & 0 < x_{(n)} \le \theta_0 \\ \infty & \theta_0 < x_{(n)} < \theta_1, \end{cases}$$

which is a non-decreasing function of $x_{(n)}$. That means, the critical region of the UMP test is equivalent to $\{x : x_{(n)} > c^*\}$. To make the test as a size α test, we can have

$$\alpha = P(X_{(n)} > c^* | \theta = \theta_0) = 1 - F_{X_{(n)}}(c^* | \theta = \theta_0) = 1 - \frac{c^{*n}}{\theta_0^n}.$$

One shall choose $c^* = \theta_0(1-\alpha)^{1/n}$ and the critical region of the UMP size α test is $\{x : x_{(n)} > \theta_0(1-\alpha)^{1/n}\}$.

(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 $X_1, \ldots, X_n$ be a random sample from $N(i\theta, \sigma^2)$, where θ and $\sigma^2 > 0$ are unknown parameters. Let

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

(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 $(n-1)$.

¶ First, one has

$$E(Q) = \frac{\sum_{i=1}^n iE(X_i)}{\sum_{i=1}^n i^2} = \theta \text{ and } \mathrm{Var}(Q) = \frac{\sum_{i=1}^n i^2\mathrm{Var}(X_i)}{(\sum_{i=1}^n i^2)^2} = \frac{\sigma^2}{\sum_{i=1}^n i^2}.$$

Since Q is a linear combination of normal random variables, Q follows a normal distribution with mean θ and variance $\sigma^2/\sum_{i=1}^n i^2$. Hence,

$$\frac{(Q-\theta)^2}{\sigma^2/\sum_{i=1}^n i^2} \sim \chi_1^2.$$

Secondly, one can write

$$\frac{\sum_{i=1}^n (X_i - i\theta)^2}{\sigma^2} = \frac{(n-1)S^2}{\sigma^2} + \frac{(Q-\theta)^2}{\sigma^2/\sum_{i=1}^n i^2}.$$

Since $\sum_{i=1}^n (X_i - i\theta)^2/\sigma^2$ follows χ_n^2, $(Q-\theta)^2/(\sigma^2/\sum_{i=1}^n i^2)$ follows χ_1^2, and Q is independent of S^2, one can claim that $(n-1)S^2/\sigma^2$ follows χ_{n-1}^2 by Cochran's theorem. Since

$$\frac{\sqrt{\sum_{i=1}^n i^2}(Q-\theta)}{S} = \frac{(Q-\theta)/(\sigma/\sqrt{\sum_{i=1}^n i^2})}{\sqrt{S^2/\sigma^2}},$$

one can claim $\sqrt{\sum_{i=1}^n i^2}(Q-\theta)/S$ is a pivotal quantity since it follows a t distribution with degree of freedom $n-1$, which is independent of θ.

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

¶ Since $\sqrt{\sum_{i=1}^n i^2}(Q-\theta)/S$ is a pivotal quantity and follows a t_{n-1} distribution, one can have

$$\begin{aligned} 1-\alpha &= P(t_{n-1,\alpha/2} \le a_n(Q-\theta)/S \le t_{n-1,1-\alpha/2}) \\ &= P(Q - t_{n-1,1-\alpha/2}S/a_n \le \theta \le Q - t_{n-1,\alpha/2}S/a_n), \end{aligned}$$

where $a_n = \sqrt{\sum_{i=1}^n i^2}$. One can claim that the $(1-\alpha)$ confidence interval is

$$\left(Q - t_{n-1,1-\alpha/2}S/a_n, Q - t_{n-1,\alpha/2}S/a_n\right).$$

(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 $x_1, \ldots, x_n$ are the observed values (or realizations) of the n random variables.

(a) To test the hypothesis $H_0 : \theta = \theta_0$ versus $H_0 : \theta \neq \theta_0$, 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.

¶ The log-likelihood function $\ell(\theta)$ can be written as

$$\ell(\theta) = n \log \theta - \theta \sum_{i=1}^{n} x_i.$$

Taking the first derivative with respect to θ, one can get the score function as

$$U(\theta) = n\theta^{-1} - \sum_{i=1}^{n} x_i.$$

Setting $U(\theta) = 0$, one can easily see that the maximum likelihood estimator of θ is $\hat{\theta} = \bar{X}^{-1}$, where $\bar{X} = n^{-1} \sum_{i=1}^{n} X_i$. This is indeed the global maximizer since the second derivative, which is the negative of the observed information, is

$$-J(\theta) = -n\theta^{-2} < 0,$$

with expected information $I_n(\theta) = E\{J(\theta)\} = n\theta^{-2}$ and $I_1(\theta) = n^{-1} I_n(\theta) = \theta^{-2}$. Under the null hypothesis, we know that the large sample likelihood ratio test statistic is

$$-2 \log \lambda(x) = -2\{\ell(\theta_0) - \ell(\hat{\theta})\} = 2 \left\{ n \log \left(\frac{\hat{\theta}}{\theta_0} \right) - (\hat{\theta} - \theta_0) \sum_{i=1}^{n} x_i \right\},$$

for which one will reject the null hypothesis if $-2 \log \lambda(x) \geq \chi^2_{1,1-\alpha}$. The critical region for the score test is

$$\left\{ x : \left| \frac{n^{-1/2} U(\theta_0)}{\sqrt{I_1(\theta_0)}} \right| = \left| \frac{\sqrt{n}(\hat{\theta}^{-1} - \theta_0^{-1})}{\theta_0^{-1}} \right| \geq z_{1-\alpha/2} \right\}.$$

The critical region for the Wald test is otherwise

$$\left\{ x : \left| \frac{\sqrt{n}(\hat{\theta} - \theta_0)}{\sqrt{I_1(\hat{\theta})^{-1}}} \right| = \left| \frac{\sqrt{n}(\hat{\theta} - \theta_0)}{\hat{\theta}} \right| \geq z_{1-\alpha/2} \right\}.$$

(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 $\hat{\theta}^*$ based on $Z_1, \ldots, Z_n$ and its large sample distribution in an explicit form.

¶ The random variables $Z_1, \ldots, Z_n$ follow a Bernoulli distribution with mean $P(X_i > x^*) = e^{-\theta x^*}$. Let $\gamma = e^{-\theta x^*}$. With $\bar{Z} = n^{-1} \sum_{i=1}^n Z_i$ as the MLE of γ based on $Z_1, \ldots, Z_n$, one can claim that the MLE of θ is $\hat{\theta}^* = -x^{*-1} \log \bar{Z}$ by the invariance property. One can derive the large sample property of $\hat{\theta}^*$ using the delta method. Let $g(y) = -x^{*-1} \log(y)$ and $g'(y) = -x^{*-1} y^{-1}$. By the central limit theorem, one has

$$\sqrt{n}(\bar{Z} - \gamma) \to_d N(0, \gamma(1 - \gamma)).$$

According to the delta method, one has

$$\sqrt{n}\{g(\bar{Z}) - g(\gamma)) \to_d N(0, \{g'(\gamma)\}^2 \gamma(1 - \gamma)).$$

Hence,

$$\sqrt{n}(\hat{\theta}^* - \theta) \to_d N(0, v(\theta)),$$

where $v(\theta) = (e^{\theta x^*} - 1)/x^{*2}$.

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

¶ According to (a), the asymptotic variance of MLE $\hat{\theta}$ based on $X_1, \ldots, X_n$ is $I_1(\theta)^{-1} = \theta^2$. If $x^* = E(X_1) = \theta^{-1}$, the asymptotic variance of $\hat{\theta}^*$ based on $Z_1, \ldots, Z_n$ is $v(\theta) = (e^{\theta\theta^{-1}} - 1)/\theta^{-2} = (e - 1)\theta^2 > \theta^2$. One can claim the asymptotic variance based on $Z_1, \ldots, Z_n$ is larger. That does make sense since $Z_1, \ldots, Z_n$ are dichotomized random variables of $X_1, \ldots, X_n$, respectively, and information will be lost after dichotomization. However, if $X_1, \ldots, X_n$ is measured with error and the error can be avoided by dichotomization, then the estimator $\hat{\theta}$ based on error prone $X_1, \ldots, X_n$ is *inconsistent* and the variance of the estimator is not necessarily smaller than $\hat{\theta}^*$.

(d) Suppose there is a new type of chemotherapy treatment 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 based on random samples $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$ for the hypothesis $H_0 : \theta \leq \beta$ versus $H_1 : \theta > \beta$, using a test statistic $(\hat{\theta} - \hat{\beta})$, where $\hat{\beta} = \bar{Y}^{-1}$ and $\bar{Y} = m^{-1} \sum_{i=1}^{m} Y_i$.

¶ When the sample size is large, we knew that $\sqrt{n}(\hat{\theta} - \theta) \rightarrow_d N(0, \theta^2)$ and $\sqrt{m}(\hat{\beta} - \beta) \rightarrow_d N(0, \beta^2)$. Since two sample are independent, one can derive the statistic $\hat{\theta} - \hat{\beta}$ will behave similarly as a normal distribution with mean $(\theta - \beta)$ and variance $\theta^2/n + \beta^2/m$. Under the null hypothesis, the critical region of the large sample Wald test with size α could be

$$\left\{ (x, y) : \frac{\hat{\theta} - \hat{\beta}}{\sqrt{\hat{\theta}^2/n + \hat{\beta}^2/m}} \geq z_{1-\alpha} \right\}.$$

(e) [**TAKE HOME**] The definition of the *hazard* function is $h(t|\theta) = f_T(t|\theta)/S_T(t|\theta)$, where $S(t|\theta) = \int_t^\infty f_T(s|\theta)ds$ is the survivor function. In the [negative] exponential distribution, one can get $h(t|\theta) = \theta$. Since the hazard function is the reciprocal of the expected length of time, one biostatistician suggests that we should report *hazard ratio*, which is defined by $\psi = \theta/\beta$, and test the hypothesis $H_0 : \psi = 1$ versus $H_1 : \psi \neq 1$ to conclude if the new treatment has a different effect than the current one. Provide a large sample test (either likelihood ratio, score, or Wald test) to test the hypothesis with size α.