Instructions: You are required to do questions 1(a), 2(a)(b)(c), 3(a)(b)(c)(e), and 4(a)(b). 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(b), 3(d), and 3(f) 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 of size n from an uniform distribution with pdf

$$f_X(x|\theta) = \frac{1}{\theta}, \quad 0 < x < \theta,$$

where θ is an unknown parameter. Suppose that a researcher wants to test the hypothesis $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$.

   (a) If $\theta_0 = 1/2$, find the cutoff c such that one rejects the null hypothesis with size 0.05 if $X_{(n)} > c$, where $X_{(n)}$ is the maximum order statistic.

   (b) [**TAKE HOME**] If the true value of θ is $3/4$, find the sample size n such that one can detect the difference between $\theta = 1/2$ and $\theta = 3/4$ under 0.8 power.

2. It is of interest to know the utilization of primary care in a city, where the rate of the primary care physician visit is expressed as the number of out-patient visits per person-year of community residence. Suppose that n adult residents are randomly selected and are asked for the total number of out-patient visits, denoted by Y_i, and the length of residency in years in the city, denoted by x_i, $i = 1, \ldots, n$. Assume that Y_i follows a Poisson distribution with mean λx_i with x_i considered as some constants (not a random variable). A hypothesis test for $H_0 : \lambda = \lambda_0$ versus $H_1 : \lambda \neq \lambda_0$ is proposed by a researcher.

   (a) Find the maximum likelihood estimator of λ, denoted by $\hat{\lambda}$.

   (b) Derive the likelihood ratio test (LRT) statistic for the hypothesis, show that the critical region of the likelihood ratio test is equivalent to

$$R = \{\boldsymbol{y} : \ \hat{\lambda} \leq c_1^* \text{ or } \hat{\lambda} \geq c_2^*\},$$

   for some cutoffs c_1^* and c_2^*, and comment on how to find these cutoffs for a *level* α test.

   (c) Assuming the sample size is large, derive the likelihood ratio test (LRT), score test, and Wald test for the hypothesis, by defining the information number as $I_1(\lambda) = n^{-1}I_n(\lambda)$, an average of expected information from n independent (but not identical) patients. Specify the critical region for each test with size α.

3. Let $X_1, \ldots, X_n$ be a random sample of size n from $N(\mu_0, \sigma^2)$, and let $Y_1, \ldots, Y_m$ be a random sample of size m from $N(\mu_1, \sigma^2)$ with *known* $\sigma^2 > 0$. Assume that two samples are mutually independent and that $\mu_1 = \theta\mu_0$.

(a) Show that $(\bar{X}, \bar{Y})$ are (joint) sufficient statistics for (μ_0, θ), where $\bar{X} = n^{-1}\sum_{i=1}^{n} X_i$ and $\bar{Y} = m^{-1}\sum_{i=1}^{m} Y_i$.

(b) Show that the maximum likelihood estimator (MLE) of θ is $\hat{\theta} = \bar{Y}/\bar{X}$.

(c) Letting

$$Q = \frac{\bar{Y} - \theta\bar{X}}{\sigma\sqrt{1/m + \theta^2/n}},$$

show that Q is a pivotal quantity. Use this quantity to find an *approximate* $(1-\alpha)$ confidence interval for θ, and common on whether one can find an *exact* confidence interval for θ.

(d) [**TAKE HOME**] Show that $\hat{\theta}$ is a consistent estimator of θ, i.e., $\hat{\theta} \to_p \theta$, without using the property of MLE directly, i.e., you are not allowed to prove the consistency by saying $\hat{\theta}$ is an MLE so $\hat{\theta}$ is consistent. Is $\hat{\theta}$ an unbiased estimator of θ?

(e) Assuming μ_0 is known, find the uniformly most powerful (UMP) test with size α for the hypothesis $H_0 : \theta = \theta_0$ versus $H_1 : \theta > \theta_0$. Specify the cutoff of the rejection region.

(f) [**TAKE HOME**] Assuming μ_0 is known, find the uniformly most powerful (UMP) test for the hypothesis $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$.

4. Assuming μ_0 is known in Question 3, answer the following questions when n and m are large, i.e., $n, m \to \infty$.

(a) Show that the MLE of θ under this scenario is $\tilde{\theta} = \bar{Y}/\mu_0$, which has large sample normality as

$$\sqrt{m}(\tilde{\theta} - \theta) \to_d N(0, v_1^2),$$

where v_1^2 is the limiting variance. Find v_1^2.

(b) Despite the asymptotic result in (a), one can actually derive the *exact* distribution of $\tilde{\theta}$, i.e., the distribution of $\tilde{\theta}$ when n is finite. Derive the distribution and use it to construct an *exact* $(1-\alpha)$ confidence interval for θ.

(c) [**BONUS**] An investigator thinks that θ should be positive so a confidence interval that covers a non-positive domain may be hard to interpret. To construct a confidence interval that covers only positive domain, one common approach is to first derive the large sample normality of $\log\tilde{\theta}$, such as

$$\sqrt{m}(\log\tilde{\theta} - \log\theta) \to_d N(0, v_2^2),$$

with limiting variance v_2^2. Then, construct a $(1-\alpha)$ confidence interval for $\log\theta$ and exponentiate both ends to obtain the confidence interval for θ. Find v_2^2 and construct the $(1-\alpha)$ confidence interval for θ as desired. Compare this interval to 3(c) and 4(b) and comment on your preference.