Instructions: You are required to do questions 1(a)(b), 2(a)(b), 3(a)(b) and 4(a)(b)(c)(d). Questions 1(c), 3(c) and 4(e) 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 $N(0, \sigma^2)$. To test $H_0 : \sigma^2 \leq 2$ versus $H_1 : \sigma^2 > 2$, answer the following questions in order to find the uniformly most powerful (UMP) test.

   (a) Show that $\sum_{i=1}^n X_i^2$ is a sufficient statistic for σ^2 and that the probability density function of X has the monotone likelihood ratio (MLR) property in $\sum_{i=1}^n X_i^2$.

   (b) Based on the proved conditions in (a), show that the critical region of the UMP test can be written as $R = \{\boldsymbol{x} : \sum_{i=1}^n x_i^2 > c\}$. Find c explicitly given type-I error α, using the fact $\sum_{i=1}^n X_i^2/\sigma^2$ follows a χ^2 distribution with degree of freedom n.

   (c) **[TAKE HOME]** Derive the likelihood ratio test (LRT) for $H_0 : \sigma^2 \leq 2$ versus $H_1 : \sigma^2 > 2$, and comment on whether this critical region is different from the UMP test.

2. Let $X_1, \ldots, X_n$ be a random sample from from a density function

$$f_X(x|\theta) = \theta x^{\theta-1}, \quad 0 < x < 1, \quad \theta > 0.$$

   If one propose a confidence interval $(X_{(1)}, X_{(n)})$ for the population median ξ, where the population median satisfies $P(X_i > \xi) = 1/2$ and $P(X_i < \xi) = 1/2$,

   (a) Derive the expected length of the confidence interval, i.e., $E(X_{(n)} - X_{(1)})$.

   (b) Derive the confidence level $(1-\alpha)$, where $1 - \alpha = P(X_{(1)} < \xi < X_{(n)})$.

3. Let $X_1, \ldots, X_n$ be a random sample from a Poisson distribution with mean λ.

   (a) Show that $\sqrt{n}(\bar{X} - \lambda)$ converges in distribution to $N(0, \lambda)$ and that $\sqrt{n}(\bar{X} - \lambda)/\sqrt{\lambda}$ is a pivotal quantity when n is large.

   (b) Using the result in (a), show that

$$\left(\bar{x} - z_{1-\alpha/2}\sqrt{\frac{\bar{x}}{n}}, \bar{x} + z_{1-\alpha/2}\sqrt{\frac{\bar{x}}{n}}\right)$$

   is a $(1-\alpha)$ confidence interval for λ, where $z_{1-\alpha/2}$ is the $(1-\alpha/2)$ quantile of the standard normal distribution. Comment on whether this interval is an *exact* or *approximate* confidence interval.

(c) **[TAKE HOME]** Comment on how one can construct a better confidence interval using the fact that

$$P(|\sqrt{n}(\bar{X}-\lambda)/\sqrt{\lambda}| \le z_{1-\alpha/2}) = P(\lambda^2 - (2\bar{X} + z^2_{1-\alpha/2}/n)\lambda + \bar{X}^2 \le 0).$$

4. Let Y_i be the random variable that follows a geometric distribution with success probability θ_i with

$$f_{Y_i}(y_i) = (1-\theta_i)^{y_i-1}\theta_i, \quad y_i = 1, 2, \ldots, \quad 0 < \theta_i < 1,$$

for $i = 1, \ldots, n$. This distribution is useful when describing the discrete time to the first event in biostatistics. For example, researchers may want to know how many "weeks" it takes for *P. vivax* malaria to relapse after a certain treatment. A common approach to model the heterogenous θ_i is assumed

$$\theta_i = \frac{\beta x_i}{1+\beta x_i},$$

where x_i is a covariate, e.g., patient's age in the malaria relapse. Given the n pairs (Y_i, x_i), $i = 1, \ldots, n$, of data points, the goal of the analysis is to obtain the maximum likelihood estimator (MLE) of β and use the estimator to make statistical inferences.

(a) Given that the likelihood function is

$$L(\beta|\boldsymbol{y}) = \prod_{i=1}^{n} f_{Y_i}(y_i|\beta) = \prod_{i=1}^{n} \left(1 - \frac{\beta x_i}{1+\beta x_i}\right)^{y_i-1} \frac{\beta x_i}{1+\beta x_i},$$

write down the log-likelihood function, score function and observed information.

(b) Prove that the MLE, denoted by $\hat{\beta}$, satisfies the equation

$$\hat{\beta}^{-1} = n^{-1} \sum_{i=1}^{n} x_i y_i (1+\hat{\beta}x_i)^{-1}.$$

(c) Show that $\sqrt{n}(\hat{\beta}-\beta) \to_d N(0, v(\beta))$, where the asymptotic variance $v(\beta)$ can be consistently estimated by

$$\hat{v}(\hat{\beta}) = \frac{n\hat{\beta}^2}{\sum_{i=1}^{n}(1+\hat{\beta}x_i)^{-1}}.$$

(d) To test the null hypothesis $H_0 : \beta = 1$ versus $H_1 : \beta \neq 1$, derive the critical regions of the likelihood ratio, score, and Wald-type test when n is large.

(e) **[TAKE HOME]** If the research has no interest to consider $\beta < 1$, she re-writes the hypothesis as $H_0 : \beta = 1$ versus $H_0 : \beta > 1$, comment on how the test regions in (d) should be adjusted.