Instructions: You are required to do questions 1(a)(b), 2(a)(b)(c), 3(a)(b)(d)(e)(f). Questions 1(c), 2(d) 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 variables from a uniform distribution on $(-\theta, \theta)$. Let $X_{(1)}$ and $X_{(n)}$ are minimum and maximum order statistics, respectively.

   (a) The range as a random variable can be defined by $R = X_{(n)} - X_{(1)}$. Derive the mean of R, and find an unbiased estimator of θ.

   (b) Show that $T = n^{-1} \sum_{i=1}^{n} |2X_i|$ is also an unbiased estimator. Without deriving the actual variance of R and T, comment on which estimator might have a smaller variance.

   (c) [**TAKE HOME**] Derive the variance of R and T and comment on which estimator you would prefer.

2. In a simple modeling strategy, the time to tumor recurrence after treatment may follow an exponential distribution with probability density function (pdf)

$$f(x|\theta) = \frac{1}{\theta} \exp\left(-\frac{x}{\theta}\right),$$

   and survivor function

$$S(x|\theta) = P(X > x|\theta) = \exp\left(-\frac{x}{\theta}\right)$$

   Let $X_1, \ldots, X_n$ be a random sample of size n from this distribution.

   (a) Show that $Q = 2n\bar{X}/\theta$ is a pivotal quantity and use the quantity to find (1-α) exact confidence interval for θ.

   (b) Find the maximum likelihood estimator (MLE) of the survival probability $p = S(x_0|\theta)$ at some point x_0 (e.g., 3 months), and derive its large sample distribution.

   (c) To test whether the survival probability is higher than half at the given point x_0, one can postulate a null hypothesis $H_0 : p \leq 0.5$ versus $H_1 : p > 0.5$. To find the uniformly most powerful (UMP) test, a biostatistician decides to make a new random variable $Y_i = I(X_i > x_0)$, where $I(\cdot)$ is the indicator function. Find the UMP test using the random sample $Y_1, \ldots, Y_n$ with a significant level α.

(d) [**TAKE HOME**] One can also use the large sample property in (b) to find an approximate test for the null hypothesis in (c). Derive any approximate test and comment on which test you would prefer.

3. In the environmental study, the distributions of the concentrations of two air pollutants X and Y can be modeled as follows: the conditional density of Y, given $X = x$, can be written as

$$f_Y(y|X = x, \alpha, \beta) = \frac{1}{(\alpha + \beta)x} \exp\left\{-\frac{y}{(\alpha + \beta)x}\right\}, \quad y > 0, \quad x > 0, \quad \alpha, \beta > 0,$$

and the marginal density of X can be written as

$$f_X(x|\beta) = \frac{1}{\beta} \exp\left(-\frac{x}{\beta}\right), \quad x > 0, \quad \beta > 0.$$

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be a random sample of X and Y in n monitoring stations.

(a) Using the fact that the joint probability density function (pdf) $f_{X,Y}(x, y|\alpha, \beta)$ can be written as

$$f_{X,Y}(x, y|\alpha, \beta) = f_Y(y|X = x, \alpha, \beta) f_X(x|\beta),$$

show that two statistics $U_1 = \sum_{i=1}^n X_i$ and $U_2 = \sum_{i=1}^n (Y_i/X_i)$ are joint sufficient statistics for (α, β).

(b) Show that U_1 and U_2 are uncorrelated, i.e., $\mathrm{Cov}(U_1, U_2) = 0$.

(c) [**TAKE HOME**] Derive a 95% confidence interval, either approximate or exact, for the parameter $\gamma = \alpha - \beta$ if $n = 30$, $\hat{\alpha} = 2$, and $\hat{\beta} = 1$, where $\hat{\alpha}$ and $\hat{\beta}$ are maximum likelihood estimators (MLE) of α and β, respectively.

(d) Now, assuming β is *known*, derive the explicit expression of the maximum likelihood estimator (MLE) of α.

(e) Show that the MLE is an unbiased estimator and its variance reaches the Cramér-Rao Lower Bound (CRLB).

(f) Again, assuming β is *known*, derive the critical regions of the likelihood ratio, score, and Wald-type test to test the null hypothesis $H_0 : \alpha = \beta$ versus $H_1 : \alpha \neq \beta$, when n is large.