1. Let $X_1, \ldots, X_n$ be iid random variables from the $n(\theta, \theta)$ distribution, $\theta > 0$.

   (a) Develop expressions for the log-likelihood function, score function, and observed information.

   (b) Conditions are satisfied in this problem for the MLE $\hat{\theta}$. One can show

   $$\sqrt{n}(\hat{\theta} - \theta) \to_d N(0, v_1(\theta)), \quad \text{as} \quad n \to \infty.$$

   Give an expression for $v_1(\theta)$.

   (c) Possible estimators of θ are $\bar{X} = \sum_{i=1}^n X_i/n$ and $S^2 = \sum_{i=1}^n (X_i - \bar{X})^2/(n-1)$. Show that

   $$\sqrt{n}(\bar{X} - \theta) \to_d N(0, \theta),$$

   and

   $$\sqrt{n}(S^2 - \theta) \to_d N(0, 2\theta^2).$$

   (d) Now we have three estimators for θ. Show that 1) $\hat{\theta}$ has the smallest asymptotic variance for every $\theta > 0$; 2) S^2 may have a smaller asymptotic variance than $\bar{X}$ given some θ; 3) the asymptotic relative efficiency (ARE) of $\hat{\theta}$ and $\bar{X}$, which is defined by the ratio of the asymptotic variances, converges to 1 when $\theta \to \infty$.

   Summarize what you found from these three arguments.

2. The exponential distribution is often used to model survival times. This problem develops a simple model for comparing survival times in two groups of patients. Let $X_1, \ldots, X_m$ be a random sample from an exponential distribution with pdf

   $$f(x|\mu_1) = \frac{1}{\mu_1} e^{-x/\mu_1}, \;\; x > 0, \quad \mu_1 > 0,$$

   and let $Y_1, \ldots, Y_n$ be a random sample from an exponential distribution with pdf

   $$f(y|\mu_2) = \frac{1}{\mu_2} e^{-y/\mu_2}, \;\; y > 0, \quad \mu_2 > 0.$$

   Assume that X and Y are independent. Define $\psi = \mu_2/\mu_1$, and let $\bar{X} = m^{-1}\sum_{i=1}^m X_i$ and $\bar{Y} = n^{-1}\sum_{i=1}^n Y_i$ be the sample means.

   (a) Show that the **exact** likelihood ratio test statistic for the hypothesis $H_0 : \mu_1 - \mu_2 = 0$ against $H_1 : \mu_1 - \mu_2 \neq 0$ is

   $$\lambda(\boldsymbol{x}, \boldsymbol{y}) = \frac{(m+n)^{m+n}}{m^m n^n} w^m (1-w)^n,$$

   where $w = m/(m + nr)$ and $r = \bar{y}/\bar{x}$.

(b) Demonstrate that the rejection region $\{(\boldsymbol{x}, \boldsymbol{y}); \lambda(\boldsymbol{x}, \boldsymbol{y}) < c\}$ is equivalent to $\{r; r < c_1^*\} \cup \{r; r > c_2^*\}$. That means one may reject the null hypothesis by observing either $r < c_1^*$ or $r > c_2^*$. Given a type-I error rate α, find c_1^* and c_2^* using the fact that $\mu_1 \bar{Y}/\mu_2 \bar{X}$ follows $F_{2n,2m}$, which is F distribution with degree of freedoms $2n$ and $2m$.

(c) Express the critical region of the Wald test for the hypothesis $H_0 : \mu_1 - \mu_2 = 0$ against $H_1 : \mu_1 - \mu_2 \neq 0$ given that the type-I error probability is α.

(d) Explain why $\psi \bar{X}/\bar{Y}$ is a pivotal quantity. Use that pivot to derive an exact 95% confidence interval for ψ.

3. An epidemiologist gathers data (x_i, Y_i) on each of n randomly chosen noncontiguous and demographically similar cities in the United States, where x_i, $i = 1, \ldots, n$, is the known population size (in millions of people) in city i, and where Y_i is the random variable denoting the number of people in city i with colon cancer. It is reasonable to assume that Y_i, $i = 1, \ldots, n$, has a Poisson distribution with mean $E(Y_i) = \theta x_i$, where $\theta > 0$ is an unknown parameter, and that $Y_1, Y_2, \ldots, Y_n$ are mutually independent random variables.

(a) Use the available data (x_i, Y_i), $i = 1, \ldots, n$, construct a uniformly most powerful (UMP) level α test of $H_0 : \theta = 1$ versus $H_1 : \theta > 1$.

(b) Use the available data (x_i, Y_i), $i = 1, \ldots, n$, construct a uniformly most powerful (UMP) level α test of $H_0 : \theta \leq 1$ versus $H_1 : \theta > 1$. Is this critical region the same as the one used in (a)? Explain.

(c) One can show that $S = \sum_{i=1}^n Y_i$ follows Poisson$(\theta \sum_{i=1}^n x_i)$. If one observes $\sum_{i=1}^n x_i = 0.8$, find c^* in the critical region $\mathcal{R} = \{S : S \geq c^*\}$ with size $\alpha = 0.05$. Explain.

(If $X \sim$ Poisson(0.8), then $P(X = 0) = 0.449$, $P(X \leq 1) = 0.808$, $P(X \leq 2) = 0.952$, $P(X \leq 3) = 0.990$, $P(X \leq 4) = 0.999$).

(d) What is the power when in reality $\theta = 5$, using the critical region in (c) and $\sum_{i=1}^n x_i = 0.8$?

(If $X \sim$ Poisson(4), then $P(X = 0) = 0.018$, $P(X \leq 1) = 0.092$, $P(X \leq 2) = 0.238$, $P(X \leq 3) = 0.433$, $P(X \leq 4) = 0.628$).