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.

   For one observation, the log-likelihood function is

   $$\ell_i(\theta|x_i) = -\frac{x_i^2}{2\theta} - \frac{\theta}{2} - \frac{1}{2}\log\theta + x_i - \frac{1}{2}\log(2\pi).$$

   Taking the first derivative, the score function is

   $$U_i(\theta|x_i) = \frac{\partial}{\partial\theta}\ell(\theta|x_i) = \frac{1}{2\theta^2}\{x_i^2 - (\theta^2 + \theta)\}.$$

   Taking the second derivative, the observed information number is

   $$J_i(\theta|x_i) = -\frac{\partial^2}{\partial\theta^2}\ell(\theta|x_i) = \frac{1}{\theta^2}\left(\frac{x_i^2}{\theta} - \frac{1}{2}\right).$$

   For n observations, the log-likelihood function, score function and observed information are $\ell(\theta|\boldsymbol{x}) = \sum_{i=1}^n \ell_i(\theta|x_i)$, $U(\theta|\boldsymbol{x}) = \sum_{i=1}^n U_i(\theta|x_i)$, and $J(\theta|\boldsymbol{x}) = \sum_{i=1}^n J_i(\theta|x_i)$, respectively

   (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 $v_1(\theta)$.

   By the large sample property of MLE, one can have $v_1(\theta) = I_1^{-1}(\theta)$, where

   $$\begin{aligned} I_1(\theta) = E\{J_i(\theta|x_i)\} &= \frac{1}{\theta^2}\left(\frac{\theta^2+\theta}{\theta} - \frac{1}{2}\right) \\ &= \frac{2\theta+1}{2\theta^2}. \end{aligned}$$

   Hence, $v_1(\theta) = I_1^{-1}(\theta) = 2\theta^2/(2\theta+1)$. Note that $E(X_i^2) = Var(X_i) + \{E(X_i)\}^2 = \theta + \theta^2$.

   (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

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

By the central limit theorem, one can conclude $\sqrt{n}(\bar{X} - \theta) \to_d n(0, \theta)$. Decomposing S^2 as

$$S^2 = \frac{n}{n-1}\frac{1}{n}\sum_{i=1}^{n}(X_i - \theta)^2 + \frac{1}{n-1}\left\{\sqrt{n}(\bar{X} - \theta)\right\}^2,$$

one can conclude that S^2 has the same limiting distribution as

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \theta)^2,$$

since $n/(n-1) \to 1$, $1/(n-1) \to 0$, and $\sqrt{n}(\bar{X} - \theta) \to_d N(0, \theta)$ when $n \to \infty$. Again, by the central limit theorem,

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

since $E(X_i - \theta)^2 = \theta$ and $\text{Var}(X_i - \theta)^2 = 2\theta^2$ (knowing that $(X_i - \theta)^2/\theta \sim \chi^2_1$). One hence can conclude $\sqrt{n}(S^2 - \theta) \to_d N(0, 2\theta^2)$.

(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.

For 1) and 3), the asymptotic relative efficiency (ARE) between $\hat{\theta}$ and $\bar{X}$ is $v_1(\theta)/\theta = 2\theta/(2\theta + 1)$, which is less than 1 if $\theta < \infty$ and converges to 1 if $\theta \to \infty$. That shows $\hat{\theta}$ has a smaller asymptotic variance than $\bar{X}$ for every finite θ. Similarly, the ARE between $\hat{\theta}$ and S^2 is $v_1(\theta)/2\theta^2 = 1/(2\theta + 1)$, which is also less than 1 if $\theta < \infty$ and converges to 0 if $\theta \to \infty$. That shows $\hat{\theta}$ has the smallest asymptotic variance for every $\theta > 0$ among three estimators. For 2), S^2 is asymptotically better than $\bar{X}$ if $\theta^2 < \theta$, i.e., if $\theta < 1/2$.

It follows what we have thought of the MLE, which is asymptotically efficient (having the smallest asymptotic variance among competing ones). When θ (variance in the distribution) getting larger, one can imagine that the observation varies more (more unstable). Hence, using deviation measurement such as S^2 may be unstable.

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 $r = \bar{y}/\bar{x}$.

The joint pdf of $\boldsymbol{X} = (X_1, \ldots, X_m)$ and $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$ are

$$L(\mu_1, \mu_2|\boldsymbol{x}, \boldsymbol{y}) = L(\mu_1|\boldsymbol{x}) L(\mu_2|\boldsymbol{y}) = \mu_1^{-m} e^{-m\bar{x}/\mu_1} \mu_2^{-n} e^{-n\bar{y}/\mu_2}.$$

Under the null hypothesis $\mu_1 = \mu_2 = \mu_0$, the maximization is over the function

$$L(\mu_0|\boldsymbol{x}, \boldsymbol{y}) = \mu_0^{-m} e^{-m\bar{x}/\mu_0} \mu_0^{-n} e^{-n\bar{y}/\mu_0}.$$

One can obtain the MLE of μ_0 as

$$\hat{\mu}_0 = (m\bar{x} + n\bar{y})/(m+n).$$

Under unrestricted space, the MLE of μ_1 and μ_2 can be solved by maximizing individual likelihood functions due to independence between $\boldsymbol{X}$ and $\boldsymbol{Y}$. One obtains $\hat{\mu}_1 = \bar{x}$ and $\hat{\mu}_2 = \bar{y}$. We hence can have the likelihood ratio statistic as

$$\begin{aligned}
\lambda(\boldsymbol{x}, \boldsymbol{y}) &= \frac{\sup_{H_0} L(\mu_1, \mu_2|\boldsymbol{x}, \boldsymbol{y})}{\sup_{H_0 \bigcup H_1} L(\mu_1, \mu_2|\boldsymbol{x}, \boldsymbol{y})} \\
&= \frac{L(\hat{\mu}_0|\boldsymbol{x}, \boldsymbol{y})}{L(\hat{\mu}_1, \hat{\mu}_2|\boldsymbol{x}, \boldsymbol{y})} = \frac{(m+n)^{m+n} (m\bar{x})^m (n\bar{y})^n}{m^m n^n (m\bar{x} + n\bar{y})^{m+n}} \\
&= \frac{(m+n)^{m+n}}{m^m n^n} w^m (1-w)^n,
\end{aligned}$$

where $w = m\bar{x}/(m\bar{x} + n\bar{y}) = m/(m + nr)$ 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 $2m$.

Here $r > 0$, $w \in (0, 1)$, and $\lambda(\boldsymbol{x}, \boldsymbol{y})$ is unimodal and concave in w. That means

$$\lambda(\boldsymbol{x}, \boldsymbol{y}) < c \Leftrightarrow \{w < c_1\} \cup \{w > c_2\}.$$

Further, since w is monotone increasing in r, we can have the critical region written as

$$\{r < c_1^*\} \cup \{r > c_2^*\}.$$

Since $\mu_1 \bar{Y}/\mu_2 \bar{X}$ follows $F_{2n,2m}$, one may choose $c_1^* = \psi F_{2n,2m,\alpha/2}$ and $c_2^* = \psi F_{2n,2m,1-\alpha/2}$, where $F_{2n,2m,\alpha}$ is the $(1-\alpha)$th quantile of $F_{2n,2m}$.

(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.

Under the null, the Wald test statistic by definition is

$$\frac{\bar{X} - \bar{Y}}{\sqrt{\mu_0^2(1/m + 1/n)}},$$

since $E(\bar{X} - \bar{Y}) = 0$ and $Var(\bar{X} - \bar{Y}) = \mu_0^2(1/m + 1/n)$. However, we did not know what μ_0 is. We need a consistent estimator of μ_0 and plug it into the Wald statistic. We knew that the MLE is a consistent estimator so we can use $\hat{\mu}_0 = (m\bar{X} + n\bar{Y})/(m + n)$ to replace μ_0 in the Wald statistic. Therefore, the Wald statistic of practical use is

$$T = \frac{\bar{X} - \bar{Y}}{\hat{\mu}_0 \sqrt{(1/m + 1/n)}},$$

which follows a standard normal distribution. Hence the critical region is $R = \{x, y; \{T \leq z_{\alpha_1}\} \cup \{T \geq z_{1-\alpha_2}\}\}$, where $\alpha_1 + \alpha_2 = \alpha$.

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

Since $\psi \bar{X}/\bar{Y}$ follows $F_{2m,2n}$, an F-distribution with degree of freedoms $2m$ and $2n$ and free of the parameter of interest ψ, one hence can claim that $\psi \bar{X}/\bar{Y}$ is a pivotal quantity. Using the pivotal quantity and its distribution, one can have

$$1-\alpha = P(F_{2m,2n,\alpha_1} < \psi\bar{X}/\bar{Y} < F_{2m,2n,1-\alpha_2}),$$

where $F_{2m,2n,\alpha}$ is the αth quantile of the distribution $F_{2m,2n}$ and $\alpha_1+\alpha_2=\alpha$. Using the equation above, one can easily see

$$1-\alpha = P(F_{2m,2n,\alpha_1}\bar{Y}/\bar{X} < \psi < F_{2m,2n,1-\alpha_2}\bar{Y}/\bar{X}),$$

and the 95% confidence interval for ψ can be

$$\left(\frac{\bar{y}}{\bar{x}}F_{2m,2n,0.25}, \frac{\bar{y}}{\bar{x}}F_{2m,2n,0.975}\right).$$

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, $i=1,\dots,n$, 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, $i=1,\dots,n$, has a Poisson distribution with mean $E(Y_i)=\theta x_i$, where $\theta>0$ is an unknown parameter, and that $Y_1, Y_2, \dots, Y_n$ are mutually independent random variables.

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

   By Neyman-Pearson Lemma, the uniformly most powerful test has a rejection region as

$$\frac{f(\boldsymbol{y}|\theta_1)}{f(\boldsymbol{y}|\theta=1)} = \frac{\prod_{i=1}^n (y_i!)^{-1}e^{-\theta_1 x_i}(\theta_1 x_i)^{y_i}}{\prod_{i=1}^n (y_i!)^{-1}e^{-\theta_1 x_i}x_i^{y_i}} = e^{(1-\theta_1)\sum_{i=1}^n x_i}\theta_1^{\sum_{i=1}^n y_i} > c,$$

   where $\theta_1>1$. Since $\theta_1>1$, $f(\boldsymbol{y}|\theta_1)/f(\boldsymbol{y}|\theta=1)>c$ is equivalent to $\sum_{i=1}^n y_i > c^*$. One hence can establish the test with a critical region $R=\{\boldsymbol{y}; \sum_{i=1}^n y_i > c^*\}$ is the UMP test.

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

   We intend to use Karlin-Rubin Theorem as it is composite vs. composite hypotheses. Here we need to first prove that $\sum_{i=1}^n Y_i$ is a sufficient statistic, which can be shown by factorizing the pdf as

$$f(\boldsymbol{y}|\theta) = \prod_{i=1}^n (y_i!)^{-1}e^{-\theta x_i}(\theta x_i)^{y_i} = \left(\prod_{i=1}^n (y_i!)^{-1}x_i^{y_i}\right)e^{-\theta\sum_{i=1}^n x_i}\theta^{\sum_{i=1}^n y_i}.$$

We then show that the pdf has a property of maximum likelihood ratio (MLR) in $\sum_{i=1}^n Y_i$. For every $\theta_2 > \theta_1$, one can see the likelihood ratio

$$\frac{f(\boldsymbol{y}|\theta_2)}{f(\boldsymbol{y}|\theta_1)} = e^{(\theta_2-\theta_1)\sum_{i=1}^n x_i}\left(\frac{\theta_2}{\theta_1}\right)^{\sum_{i=1}^n y_i}$$

is monotone increasing in $S = \sum_{i=1}^n Y_i$. Hence the MLR property stands. By the Karlin-Rubin Theorem the UMP test has a critical region $R = \{\boldsymbol{y}; S = \sum_{i=1}^n y_i > s_0\}$. This critical region is the same as in (a) since they are both established via the same test statistic. As suggested in the following question, $\sum_{i=1}^n Y_i$ follows a Poisson distribution and the c^* and s_0 can be found in satisfaction with the type I error probability.

(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 level $\alpha = 0.05$. Explain.

Given type I error $\alpha = 0.05$, one has $P_{\theta=1}(\sum_{i=1}^n Y_i \geq c^*) \leq 0.05$. Under the null hypothesis ($\theta = 1$) and $\sum_{i=1}^n x_i = 0.8$, $\sum_{i=1}^n Y_i$ follows Poisson(0.8). Therefore,

$$\begin{aligned} P(\sum_{i=1}^n Y_i \geq c^*) &= 1 - P(\sum_{i=1}^n Y_i < c^*) \\ &= 1 - P(\sum_{i=1}^n Y_i \leq c^* - 1) \leq 0.05. \end{aligned}$$

By the information provided, one should choose $c^* - 1 = 2$. Hence $c^* = 3$.

(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 $\sum_{i=1}^n x_i = 0.8$?

Given that the critical region is $R = \{\boldsymbol{y} : \sum_{i=1}^n y_i \geq 3\}$, the power at $\theta = 5$ is

$$\begin{aligned} \beta(5) = P_{\theta=5}(\sum_{i=1}^n Y_i \geq 3) &= 1 - P_{\theta=5}(\sum_{i=1}^n Y_i < 3) \\ &= 1 - P_{\theta=5}(\sum_{i=1}^n Y_i \leq 2) \\ &= 1 - 0.238 = 0.762, \end{aligned}$$

since, under $\theta = 5$, $\sum_{i=1}^{n} Y_i$ follows Poisson$(0.8 \times 5) \equiv$ Poisson (4).

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