1. Suppose Y is a single observation from a Pareto-type distribution with population level pdf

$$f_Y(y|\theta) = \begin{cases} \dfrac{\theta}{y^{\theta+1}}, & y > 1 \\ 0, & \text{otherwise.} \end{cases}$$

The parameter $\theta > 0$ is unknown.

(a) Plot $f_Y(y|\theta)$ versus y under both H_0 and H_a. Find the most powerful level $\alpha = 0.05$ rejection region to test

$$H_0 : \theta = 2$$
$$\text{versus}$$
$$H_a : \theta = 4.$$

(b) Calculate the power of the test when H_a is true.

(c) Redo part (a) using $H_a : \theta = 5$ instead. Does your most powerful level $\alpha = 0.05$ rejection region change? What does this suggest?

2. There are approximately 540 coronavirus testing locations in South Carolina. At the beginning of the day, officials at each location record

$$Y = \text{number of specimens tested to find the first positive case}$$

and assume Y follows a geometric distribution with probability p. The probability p satisfies $0 < p < 1$ and is unknown. In this application, we might also call p the "population prevalence" of the disease. Of course, careful thought should go into defining exactly what the "population" is here.

Suppose $Y_1, Y_2, ..., Y_{540}$ are iid geometric(p) random variables (one for each site) observed on a given day. Epidemiologists at SC-DHEC would like to test

$$H_0 : p = 0.02$$
$$\text{versus}$$
$$H_a : p < 0.02.$$

(a) Show the likelihood function of p on the basis of observing $\mathbf{y} = (y_1, y_2, ..., y_{540})$ is given by

$$L(p|\mathbf{y}) = (1-p)^{\sum_{i=1}^{540} y_i - 540}\, p^{540}.$$

(b) Show the uniformly most powerful (UMP) level α test of H_0 versus H_a has a rejection region of the form

$$\text{RR} = \left\{ t = \sum_{i=1}^{540} y_i \geq k^* \right\}.$$

How would you choose k^* to ensure the test is level $\alpha = 0.05$? *Hint:* What is the sampling distribution of $T = \sum_{i=1}^{540} Y_i$ when H_0 is true?

3. Suppose $Y_1, Y_2, ..., Y_n$ is an iid sample from a beta$(\theta, 1)$ population with pdf

$$f_Y(y|\theta) = \begin{cases} \theta y^{\theta-1}, & 0 < y < 1 \\ 0, & \text{otherwise.} \end{cases}$$

The goal is to test

$$\begin{gathered} H_0 : \theta = 1 \\ \text{versus} \\ H'_a : \theta = \theta_a. \end{gathered}$$

where $\theta_a > 1$.
(a) What is another name for the population-level pdf when H_0 is true?
(b) Show the most powerful test of H_0 versus H'_a has a rejection region of the form

$$\text{RR} = \left\{ t = \prod_{i=1}^{n} y_i > k^* \right\}.$$

(c) Show that choosing k^* to satisfy

$$\alpha = P_{H_0}\left(\prod_{i=1}^{n} Y_i > k^* \right)$$

in part (b) is the same as choosing k^{**} to satisfy

$$\alpha = P_{H_0}\left(\sum_{i=1}^{n} -\ln Y_i < k^{**} \right).$$

Now, derive the sampling distribution of $\sum_{i=1}^{n} -\ln Y_i$ (under H_0) and specify how to choose k^{**} to ensure

$$\text{RR} = \left\{ \sum_{i=1}^{n} -\ln Y_i < k^{**} \right\}$$

is a level α rejection region.
(d) Is the rejection region in part (c) the *uniformly* most powerful level α rejection region for $H_0 : \theta = 1$ versus $H_a : \theta > 1$? Explain.

4. Suppose $Y_1, Y_2, ..., Y_n$ is an iid sample from a $\mathcal{N}(0, \sigma^2)$ population distribution, where $\sigma^2 > 0$ is unknown (note the population mean here is $\mu = 0$). The goal is to test

$$\begin{gathered} H_0 : \sigma^2 = \sigma_0^2 \\ \text{versus} \\ H_a : \sigma^2 \neq \sigma_0^2, \end{gathered}$$

where σ_0^2 is known. In HW3 (Problem 2), you considered the rejection region of the form

$$\text{RR} = \{t < k_1 \ \text{ or } \ t > k_2\},$$

where $t = \sum_{i=1}^n y_i^2$ and $k_2 > k_1$. Show this is the same rejection region you would get by performing a likelihood ratio test (LRT) for H_0 versus H_a. Then remind yourself how to choose k_1 and k_2 to ensure a level α rejection region.

5. We would like to compare the population mean referral waiting times for patients in Lexington, SC and Augusta, GA seeking care from a gastrointestinal specialist. By "referral waiting time," I mean the time it takes to see a gastrointestinal specialist once a referral has been made by another health professional (e.g., a primary care physician, etc.). We have independent random samples of patients from the two locations. Here are the corresponding waiting times and population-level models for them:

- Lexington: $X_1, X_2, ..., X_m \sim$ iid exponential(θ_1)
- Augusta: $Y_1, Y_2, ..., Y_n \sim$ iid exponential(θ_2).

The population parameters satisfy $\theta_1 > 0$ and $\theta_2 > 0$ and are unknown. The goal is to test

$$\begin{gathered} H_0 : \theta_1 = \theta_2 \\ \text{versus} \\ H_a : \theta_1 \neq \theta_2. \end{gathered}$$

(a) Preparing for a LRT derivation below, carefully describe the null parameter space Θ_0 and the entire parameter space Θ. Draw a picture of what both spaces look like.
(b) Show the likelihood function is given by

$$L(\boldsymbol{\theta}|\mathbf{x},\mathbf{y}) = L(\theta_1,\theta_2|\mathbf{x},\mathbf{y}) = \frac{1}{\theta_1^m} e^{-\sum_{i=1}^m x_i/\theta_1} \times \frac{1}{\theta_2^n} e^{-\sum_{j=1}^n y_j/\theta_2},$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2)$. This is just the likelihoods from each sample multiplied together (because the two samples are independent).
(c) Show the (restricted) MLE of $\boldsymbol{\theta}$ over the null parameter space Θ_0 is

$$\widehat{\boldsymbol{\theta}}_0 = \begin{pmatrix} \dfrac{m\overline{X} + n\overline{Y}}{m+n} \\ \dfrac{m\overline{X} + n\overline{Y}}{m+n} \end{pmatrix}.$$

(d) Show the (unrestricted) MLE of $\boldsymbol{\theta}$ over the entire parameter space Θ is

$$\widehat{\boldsymbol{\theta}} = \begin{pmatrix} \overline{X} \\ \overline{Y} \end{pmatrix}.$$

(e) Show the likelihood ratio test (LRT) statistic

$$\lambda = \frac{L(\widehat{\boldsymbol{\theta}}_0|\mathbf{x},\mathbf{y})}{L(\widehat{\boldsymbol{\theta}}|\mathbf{x},\mathbf{y})} = \frac{\overline{x}^m\overline{y}^n}{\left(\dfrac{m\overline{x}+n\overline{y}}{m+n}\right)^{m+n}}.$$

(f) Here are the observed data on the referral waiting times for both groups of patients:

Lexington, SC		Augusta, GA	
36	11	1	49
47	52	16	43
52	9	22	8
1	39	11	2
9	20	8	8
72	32	26	53
		12	24
		39	

Calculate $-2\ln\lambda$ using the data above and implement a large-sample LRT to test H_0 versus H_a. What is your conclusion at $\alpha = 0.05$?