Homework 5 Solutions,

1. Question 10.103. Let $Y_1, \ldots, Y_n$ denote a random sample from a uniform distribution over the interval $(0, \theta)$.

a. Find the most powerful $\alpha$-level test for testing $H_0 : \theta = \theta_0$ against $H_a : \theta = \theta_a$, where \( \theta_a < \theta_0 \). [Hint: You will find a large collections of tests that are equally powerful. Choose one of these.]

b. Is the test in part (a) uniformly most powerful for testing $H_0 : \theta = \theta_0$ against $H_a : \theta < \theta_0$?

   a. The Neyman–Pearson Lemma says the most powerful test rejects $H_0$ when
   \[ L(\theta_0)/L(\theta_a) < k. \]
   Note that under either hypothesis, $Y_{(n)} < \theta_0$. The likelihood ratio is
   \[ \begin{cases} \infty & \text{if } Y_{(n)} > \theta_a \\ \theta_a^\theta_0/\theta_0^{\theta_a} & \text{if } Y_{(n)} < \theta_a. \end{cases} \]
   Hence one should reject $H_0$ first if $Y_{(n)} < \theta_a$, and power considerations do not give any preference to which rejection region you choose, so long as it has the right size. Then, add points from $Y_{(n)} \geq \theta_1$. You can’t do better than a rejection region of \{ $Y_{(n)} \leq \theta_1$ \}. Recall that under $H_0$, \[ P_0 \left[ Y_{(n)} \leq \theta_1 \right] = \left( \theta_0/\theta_1 \right)^n, \] and so let $C = \theta_0 \alpha^{1/n}$.

   b. Because the form of the rejection region did not depend on the particular value of $\theta_a$, this is the most powerful test.

2. Question 10.106. A survey of voter sentiment was conducted in four midcity political wards to compare the fraction of voters favoring candidate A. A random sample of 200 voters were polled in each of the four wards, with the results as shown in the accompanying table. The numbers of voters favoring A in the four samples can be regarded as four independent binomial random variables. Construct a likelihood ratio test for the hypothesis that the fractions of voters favoring candidate A are the same in all four wards. Use $\alpha = 0.05$.

<table>
<thead>
<tr>
<th>Ward</th>
<th>Opinion</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Favor A</td>
<td>76</td>
<td>53</td>
<td>59</td>
<td>48</td>
<td>236</td>
<td></td>
</tr>
<tr>
<td>Do not favor A</td>
<td>124</td>
<td>147</td>
<td>141</td>
<td>152</td>
<td>564</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>200</td>
<td>200</td>
<td>200</td>
<td>200</td>
<td>800</td>
<td></td>
</tr>
</tbody>
</table>

Let $Y_i$ be the number favoring A in Ward $i$. Let $\hat{\theta}_i$ be the population proportion favoring A in Ward $i$. Then the likelihood is \[ L(\theta_1, \ldots, \theta_4) = \prod_{i=1}^{4} \left( \frac{\theta_0}{Y_i} \right)^{Y_i} \left( 1 - \frac{\theta_0}{Y_i} \right)^{200-Y_i}. \] Under the alternative hypothesis, \( \hat{\theta}_i = Y_i/200 \), and under the null, \( \hat{\theta}_i = \sum_{j=1}^{4} Y_j/800 \). (Note that under the null hypothesis, the product of binomial coefficients does not collapse to a single binomial coefficient).

Then
\[
\Lambda = \frac{\prod_{i=1}^{4} \left( \frac{\theta_0}{Y_i} \right)^{Y_i} \left( 1 - \frac{\theta_0}{Y_i} \right)^{200-Y_i}}{\prod_{i=1}^{4} \left( \frac{\theta_0}{Y_i} \right)^{\sum_{j=1}^{4} Y_j/800} \left( 1 - \frac{\theta_0}{Y_i} \right)^{200-Y_i}}.
\]

Let $Y_i$ be the number favoring A in Ward $i$. Let $\theta_i$ be the population proportion favoring A in Ward $i$. Then the likelihood is $L(\theta_1, \ldots, \theta_4) = \prod_{i=1}^{4} \left( \frac{\theta_0}{Y_i} \right)^{Y_i} \left( 1 - \frac{\theta_0}{Y_i} \right)^{200-Y_i}$. Under the alternative hypothesis, $\hat{\theta}_i = Y_i/200$, and under the null, $\hat{\theta}_i = \sum_{j=1}^{4} Y_j/800$. (Note that under the null hypothesis, the product of binomial coefficients does not collapse to a single binomial coefficient). Then
\[
\Lambda = \frac{\prod_{i=1}^{4} \left( \frac{\theta_0}{Y_i} \right)^{Y_i} \left( 1 - \frac{\theta_0}{Y_i} \right)^{200-Y_i}}{\prod_{i=1}^{4} \left( \frac{\theta_0}{Y_i} \right)^{\sum_{j=1}^{4} Y_j/800} \left( 1 - \frac{\theta_0}{Y_i} \right)^{200-Y_i}}.
\]

Let $Y_i$ be the number favoring A in Ward $i$. Let $\theta_i$ be the population proportion favoring A in Ward $i$. Then the likelihood is $L(\theta_1, \ldots, \theta_4) = \prod_{i=1}^{4} \left( \frac{\theta_0}{Y_i} \right)^{Y_i} \left( 1 - \frac{\theta_0}{Y_i} \right)^{200-Y_i}$. Under the alternative hypothesis, $\hat{\theta}_i = Y_i/200$, and under the null, $\hat{\theta}_i = \sum_{j=1}^{4} Y_j/800$. (Note that under the null hypothesis, the product of binomial coefficients does not collapse to a single binomial coefficient). Then
\[
\Lambda = \frac{\prod_{i=1}^{4} \left( \frac{\theta_0}{Y_i} \right)^{Y_i} \left( 1 - \frac{\theta_0}{Y_i} \right)^{200-Y_i}}{\prod_{i=1}^{4} \left( \frac{\theta_0}{Y_i} \right)^{\sum_{j=1}^{4} Y_j/800} \left( 1 - \frac{\theta_0}{Y_i} \right)^{200-Y_i}}.
\]
Then \( \ln(\Lambda) = \sum_{i=1}^{4} Y_i \log(Y_i/200) + \sum_{i=1}^{4} (200 - Y_i) \log(1 - Y_i/200) - [\left(\sum_{i} Y_i\right) \ln(\left(\sum_{i} Y_i\right)/800) + (800 - \left(\sum_{i} Y_i\right)) \ln(1 - (\sum_{i} Y_i)/800)] = 5.268. \) Using Wilk’s test, compare \( 2 \times 5.268 = 10.536 \) to a \( \chi^2_3 \) distribution. The 0.95 quantile of the \( \chi^2_3 \) distribution is 7.815. Reject the null hypothesis.

3. Question 10.109. Do only part a. Let \( X_1, \ldots, X_m \) denote a random sample from an exponential density with mean \( \theta_1 \) and let \( Y_1, \ldots, Y_n \) denote an independent random sample from an exponential density with mean \( \theta_2 \).

a. Find the likelihood ratio criterion for testing \( H_0 : \theta_1 = \theta_2 \) versus \( H_a : \theta_2 \neq \theta_1 \).

b. Show that the test in part (a) is equivalent to an exact \( F \) test [Hint: transform \( \sum X_i \) and \( \sum Y_j \) to \( \chi^2 \) random variables.

The likelihood ratio condition says to divide the likelihood maximized under the null hypothesis by the likelihood maximized under the alternative. Under the null hypothesis, the log likelihood is

\[
\ell(\theta) = -\left(\sum_{i=1}^{m} X_i + \sum_{i=1}^{n} Y_i\right)/\theta - (m+n) \log(\theta),
\]

and \( \ell'(\theta) = \left(\sum_{i=1}^{m} X_i + \sum_{i=1}^{n} Y_i\right)/\theta^2 \) \( - (m+n)/\theta \). The root of the likelihood equation is \( \hat{\theta} = \left(\sum_{i=1}^{m} X_i + \sum_{i=1}^{n} Y_i\right)/(m+n) \). Furthermore, \( \ell''(\hat{\theta}) = -2\left(\sum_{i=1}^{m} X_i + \sum_{i=1}^{n} Y_i\right)/\theta^3 + (m+n)/\theta^2 \), and \( \ell''(\hat{\theta}) = -2(m+n)/\theta^2 + (m+n)/\hat{\theta}^2 < 0 \), and so \( \hat{\theta} \) is a maximizer. Similarly, maximizers under the alternative hypothesis are \( \hat{\theta}_1 = \sum_{i=1}^{m} X_i/m \) and \( \hat{\theta}_2 = \sum_{i=1}^{n} Y_i/n \). Then reject for small values of

\[
\Lambda = \frac{\prod_{i=1}^{m} \exp(-X_i/\hat{\theta}) \prod_{i=1}^{n} \exp(-Y_i/\hat{\theta}) \hat{\theta}^{-(m+n)}}{\prod_{i=1}^{m} \exp(-X_i/\hat{\theta}_1) \prod_{i=1}^{n} \exp(-Y_i/\hat{\theta}_2) \hat{\theta}_1^{-m} \hat{\theta}_2^{-n}} = \frac{\exp(-(m+n))\hat{\theta}_1^m \hat{\theta}_2^n}{\exp(-(m+n))\hat{\theta}_1^m \hat{\theta}_2^n} = \frac{\hat{\theta}_1^m \hat{\theta}_2^n}{\theta_1^m \theta_2^n}
\]

b Let \( F = \bar{X}/\bar{Y} \). Under the null hypothesis, \( F \) has an \( F \) distribution with \( 2m \) and \( 2n \) degrees of freedom, and \( \Lambda \) can be expressed as an increasing function of \( F \).

4. Question 10.110. Suppose that a probability model involving a parameter \( \theta \) has a sufficient statistic \( T \). Show that a likelihood ratio test of a null hypothesis \( H_0 : \theta \in \Omega_0 \) vs. the alternative hypothesis \( H_A : \theta \in \Omega_A \) depends on the data only through the value of \( T \).

Let \( g(T, \theta)h(X_1, \ldots, X_n) \) be the factorization of the likelihood that the factorization theorem indicates exists. The \( L(\theta) = g(T, \theta)h(X_1, \ldots, X_n) \). The two maximum likelihood estimators \( \hat{\theta}_0 \) and \( \hat{\theta} \) are given by \( \arg\max_{\theta \in \Omega_0} L(\theta) = \arg\max_{\theta \in \Omega_0} g(T, \theta)h(X_1, \ldots, X_n) = \arg\max_{\theta \in \Omega_0} g(T, \theta) \) and \( \arg\max_{\theta \in \Omega_0, \Omega_A} L(\theta) = \arg\max_{\theta \in \Omega_0, \Omega_A} g(T, \theta) \) respectively. Hence \( \hat{\theta}_0 \) and \( \hat{\theta} \) depend on the data only through \( T \), and so \( \Lambda = L(\hat{\theta}_0)/L(\hat{\theta}_A) = g(T, \hat{\theta}_0)/g(T, \hat{\theta}) \).