JOE SEIDEL

STAT 244 HOMEWORK 6

Question 1

Suppose X follows a geometric distribution with parameter p.

1. Derive the likelihood ratio for testing the hypothesis $p = p_0$ versus the alternative $p \neq p_0$.

   THE variable X has PMF given

   $$p(X = x) = p(1-p)^{x-1} \text{ , } x = 1, 2, ...$$

   We have a composite hypothesis a generalized likelihood ratio test is in order.

   $$\Lambda = \frac{\max_{p \in w_0}[L(p)]}{\max_{p \in \Omega}[L(p)]}$$

   Where the rejection region consists of small values for Λ. In this case $w_0 = \{p_0\}$ and $\Omega = \{0 < p < 1\}$

   $$\max_{p \in w_0}[L(p)] = p_0(1-p_0)^{x-1}.$$

   For the denomitor we have to maximize the likelihood for $p \in \Omega$. Which is the mle of $f(x \mid p) = p(1-p)^x$, $\hat{p} = \frac{1}{\overline{X}}$. Which makes

   $$\max_{p \in \Omega}[L(p)] = \frac{1}{\overline{X}}\left(1 - \frac{1}{\overline{X}}\right)^{x-1}.$$

   Therefore

   $$\Lambda = \frac{\overline{X} p_0 (1-p_0)^{x-1} \overline{X}^{x-1}}{(\overline{X}-1)^{x-1}}$$

   It's important to note that I've set this up observing only 1 X, otherwise this would look slightly different, in fact I will end up replacing x for $\overline{X}$ soon.

   Which is the generalized likelihood ratio that will test the hypothesis.

2. For $p_0 = 0.01$, by some combination of numerical experimentation and mathematical analysis, find the set of possible valyes of x for X for which the likelihood ratio is less than 0.1.

   EVALUATING Λ for a few values of X yields.

x	1	2	3	4	5
Λ	.01	.0396	.0066	.09	.117

   Table 1: GLR

   Hence for $x \leq 4$ we have likelihood ratio less than 0.1.

3. Find the probability of Type 1 error for the test the rejects $p_0 = 0.01$ when the likelihood ratio is less than 0.1. Find the power of this test when $p = 0.5$. Find the power of ths test when $p = 0.001$.

Question 2 Rice 9.12

Let $X_1, ..., X_n$ be a random sample from an exponential distribution with the density function $f(x \mid \theta) = \theta e^{-\theta x}$. Derive a likelihood ratio test of H_0: $\theta = \theta_0$ versus H_A: $\theta \neq \theta_0$, and show that the rejection region is of the form $\{\overline{X} e^{-\theta_0 \overline{X}} \leq c\}$.

First, recall that the mle of $L(\theta)$ is $\hat{\theta} = \overline{X}$. Set up the likelihood ratio test

I've found this in previous homeworks.

$$\Lambda = \frac{\max_{p \in w_0}[L(p)]}{\max_{p \in \Omega}[L(p)]}.$$

The numerater will be

$$L(\theta_0) = \prod_{i=1}^{n} \theta e^{-\theta_0 x_i}$$

and the denomitaor

$$L(\hat{\theta}) = \prod_{i=1}^{n} \frac{1}{\overline{X}} e^{-\frac{x_i}{\overline{X}}}.$$

Then

$$\begin{aligned}
\Lambda &= \frac{\prod_{i=1}^{n} \theta e^{-\theta_0 x_i}}{\prod_{i=1}^{n} \frac{1}{\overline{X}} e^{-\frac{x_i}{\overline{X}}}} \\
&= \frac{\theta_0^n e^{-\theta_0 n \overline{X}}}{\frac{1}{\overline{X}^n} e^{-\frac{\overline{X} n}{\overline{X}}}} \\
&= \frac{\theta_0^n \overline{X}^n e^{-\theta_0 n \overline{X}}}{e^{-n}} \\
&= \left(e \theta_0 \overline{X} e^{(-\theta_0 \overline{X})}\right)^n
\end{aligned}$$

Where H_0 is rejected when Λ is small. Since e, n, θ are positive, Λ is small when $\overline{X} e^{(-\theta_0 \overline{X})}$ is small.

$$\begin{aligned}
\left(e \theta_0 \overline{X} e^{(-\theta_0 \overline{X})}\right)^n &< c_1 \\
e \theta_0 \overline{X} e^{(-\theta_0 \overline{X})} &< c_1^{\frac{1}{n}} \\
\overline{X} e^{(-\theta_0 \overline{X})} &< \frac{c_1^{\frac{1}{n}}}{\theta_0 e}
\end{aligned}$$

Therefore, we see the rejection region takes the form $\overline{X}e^{(-\theta_0\overline{X})} \leq c = \frac{c_1^{\frac{1}{n}}}{\theta_0 e}$.

Question 3 Rice 9.13

Suppose, to be specific, that in problem 12, $\theta_0 = 1$, $n = 10$, and that $\alpha = .05$. In order to use the test, we must find the appropriate value of c.

1. Show that rejection region is of the form $\{\overline{X} \leq x_0\} \cup \{\overline{X} \geq x_1\}$, where x_0 and x_1 are deterimined by c.

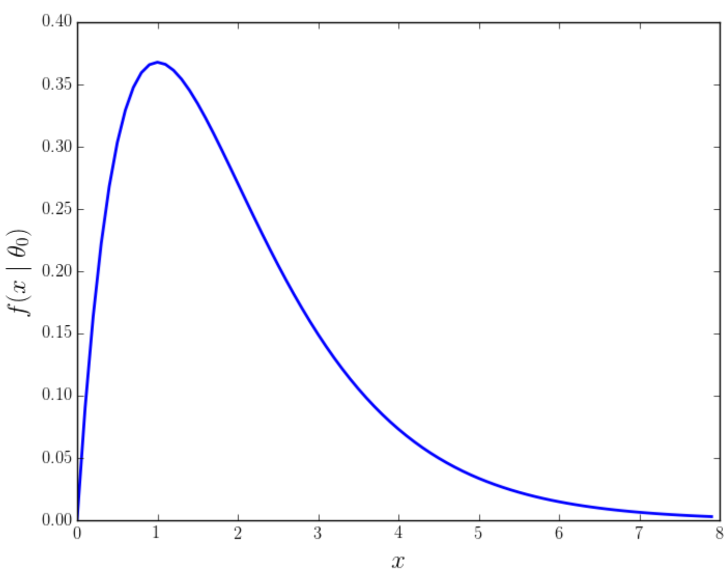
   Now that we are given a value for θ_0 we have

   $$f(x \mid \theta_0) = xe^{-x}.$$

   But we also found in the previous question that our test rejects when

   $$f(\overline{X} \mid \theta_0) = \overline{X}e^{(-\overline{X})}$$

   is small, specifically when less than c. To see why it takes the form $\{\overline{X} \leq x_0\} \cup \{\overline{X} \geq x_1\}$, consider the graph of the function.



   There would be a c chosen that corresponds with the Y axis of the graph and a horizontal line would interect the function at two values of x.

2. Explain why c should be chosen so that $P(\overline{X}e^{(-\overline{X})} \leq c) = .05$.

THIS is just a restatement of Type I error under Neyman-Pearson.

$$Pr(\text{Reject } H_0 \mid H_0) = Pr(x \in [0, c] \mid \theta_0)$$

Which can be worded to say, the probability that x falls in the rejection zone. Furthermore, we found that $P(\overline{X}e^{(-\overline{X})})$ provides a rejection zone for Λ. What remains is to determine how willing we are to make Type 1 error.

If we want $\alpha = .05$ then we should set

$$Pr(x \in [0, c] \mid \theta_0) = Pr(f(\overline{X} \mid \theta)) < c) = .05$$

to determine what c should be.

3. Explain why $\sum_{i=1}^{1} 0X_i$ and hence $\overline{X}$ follow gamma distributions when $\theta_0 = 1$. How could this knowledge be used to choose c?

Under H_0: $\theta_0 = 1$ $X_i \sim Exponetial(1)$ which is a special ase of $\Gamma(1, \lambda)$ or in this particular case $\Gamma(1, 1)$. On the last homework, I found that $\sum_{i=1}^{n} X_i \sim \Gamma(n, 1)$ and $\overline{X} \sim \Gamma(n, n)$. Knowing the exact distribution, means I could compute the exact value of c for which 95 coverage for acceptable values of H_0.

In this day and age, a computer would be the easiest way to do this. The gist would be first solve $f(\overline{X}) = c$ to get $x_0(c)$ and $x_1(c)$. Then solve

$$\alpha(c) = F(x_0(c)) + 1 - F(x_1(c))$$

Where $F(x)$ is cdf of $\Gamma(10, 10)$.

4. Suppose you hadn't thought of the preceding fact. Explain how you could determine a good approximation to c by generating random numbers on a computer (simulation).

GENERATE a bunch of samples from $Exponential(1)$ of size $n = 10$. Then compute $\overline{X}e^{\overline{X}}$ for each sample. Sort the result and set the cuttoff as the value indexed at the 5% of the number of samples generated. For example, if the results are stored in some list of size 10000 then you'd take value indexed at list[500] as the cuttoff value for a close approximation.

Question 4

Call "This, it thus, and" Class I workds; Class II is "everything else". For each of 215 groups of 5 of James Mill's sentences, the number of Class I words was counted.

Test whether a Binomial Distribution $(n = 5, \theta)$ fits these data.

Figure 1: James Mill, the economist

FIRST, knowing the MLE of a binomial will be useful.

$$\begin{aligned} L(\theta) &= \prod_{i=1}^{n} \binom{n}{x_i} \theta^{x_i} (1-\theta)^{n-x_i} \\ &= \prod_{i=1}^{n} \binom{n}{x_i} \theta^{\sum_{i=1}^{n} x_i} (1-\theta)^{n-\sum_{i=1}^{n} x_i} \end{aligned}$$

Taking the log and the derivative we have

$$\frac{d}{d\theta} \log L(\theta) = \frac{\overline{X}n}{\theta} + \frac{n - \overline{X}n}{\theta - 1}$$

Setting the above equal to zero and solving for θ give the mle.

$$\hat{\theta} = \overline{X}$$

Using the table we can compute the mle, $\hat{\theta} = 339/215(5) = .31$.

No. Class I words	0	1	2	3	4	5
No. groups(observed)	87	11	51	42	20	4
No. groups(expected)	34	76	68	30	7	1

Table 2: Class I words