HW 21 Solutions

The Chi-square approximation for the GLHR test is often used when comparing nested models. For example, we might model $n = 10$ soccer goal time values as $X_1, \ldots, X_n \sim Gamma(\alpha, \lambda)$ and consider the special case with $\alpha = 1$ (i.e. an exponential model). The idea is that if the exponential model is not rejected in favor of the more general Gamma model, this validates the choice of using the simpler (one parameter) distribution.

We will carry out a GLHR test of $H_0 : \alpha = 1, \lambda > 0$ vs $H_1 : \alpha > 0, \lambda > 0$. The sufficient statistics are $\bar{x} = 37.553$ minutes (per goal) and the geometric average $\bar{x}_g = \left(\prod x_i\right)^{1/n} = 19.593$ minutes (and $n = 19$).

Problem 1

(a) Write the likelihood and log-likelihood functions for α and λ in terms of $\bar{x}$ and $\bar{x}_g$ (note $\log(\bar{x}_g) = \sum \log(x_i)/n$).

$$L(\alpha, \lambda) = \prod_{i=1}^{n} \frac{\lambda^{\alpha}}{\Gamma(\alpha)} x_i^{\alpha-1} e^{-\lambda x_i}$$

$$= \left[\frac{\lambda^{\alpha}}{\Gamma(\alpha)}\right]^n \left[\left(\prod_{i=1}^{n} x_i\right)^{1/n}\right]^{n(\alpha-1)} \exp\left\{-\lambda n \bar{x}\right\}$$

$$\ln L(\alpha, \lambda) = n \ln\left(\frac{\lambda^{\alpha}}{\Gamma(\alpha)}\right) + \sum_{i=1}^{n} \ln x_i^{\alpha-1} - \lambda n \bar{x}$$

$$= n \ln\left(\frac{\lambda^{\alpha}}{\Gamma(\alpha)}\right) + (\alpha - 1)n \left(\frac{\sum_{i=1}^{n} \ln x_i}{n}\right) - \lambda n \bar{x}$$

(b) Find the method of moments estimates for α and λ and evaluate the log likelihood at these values. (Hint: You can use the function *lgamma(a)* to evaluate $\log \Gamma(a)$ in R. You may also assume that $\sum_{i=1}^{n} x_i^2 = 2594.1$.)

First take

$$\frac{1}{n} \sum_{i=1}^{n} x_i \stackrel{\text{set}}{=} E(X_1) = \frac{\alpha}{\lambda}$$

$$\frac{1}{n} \sum_{i=1}^{n} x_i^2 \stackrel{\text{set}}{=} E(X_1^2) = Var(X_1) - [E(X_1)]^2 = \frac{\alpha}{\lambda^2}(1 - \alpha)$$

and then solve for $\hat{\alpha}_{MOM}$ and $\hat{\lambda}_{MOM}$:

$$\hat{\lambda}_{MOM} = \frac{\bar{x}}{(1/n)\sum_{i=1}^{n} x_i^2 + \bar{x}^2}, \quad \hat{\alpha}_{MOM} = \frac{\bar{x}^2}{(1/n)\sum_{i=1}^{n} x_i^2 + \bar{x}^2}.$$

```
xbar = 37.553
xg = 19.593
sum_sq_x = 2594.1
```

```
n=19
```

```
(a_obs = xbar/((1/n)*sum_sq_x + xbar^2))
```

```
## [1] 0.0242785
```

```
(l_obs = (xbar^2/((1/n)*sum_sq_x + xbar^2)))
```

```
## [1] 0.9117306
```

```
my_log_lik <- function(a, l, n){
  return(n*log((l^a)/lgamma(a)) + (a-1)*n*log(xg) - l*n*xbar)
}

my_log_lik(a_obs, l_obs, n)
```

```
## [1] -730.6067
```

(c) The MLE (found numerically) are $\hat{\alpha}_{MLE} = 0.9$ and $\hat{\lambda}_{MLE} = 0.024$. Evaluate the log likelihood function at the MLEs and compare this to the value in part (b).

```
a_MLE = 0.9
l_MLE = 0.024
my_log_lik(a_MLE, l_MLE, n)
```

```
## [1] -35.01898
```

Problem 2

(a) Find the MLE for λ under H_0 and evaluate the maximized log-likelihood under this null hypothesis.

Under H_0 the likelihood is

$$L(\alpha = 1, \lambda > 0) = \lambda^n \exp\{\lambda n \bar{x}\} \quad \text{or} \ln L(\alpha = 1, \lambda > 0) = n \ln(\lambda) - n\lambda\bar{x}$$

so the MLE for λ under ω_0 is found by

$$\frac{\partial}{\partial \lambda} \ln L(\alpha = 1, \lambda > 0) = \frac{n}{\lambda} - n\bar{x} \overset{set}{=} 0$$

i.e. the MLE of λ under H_0 is $\hat{\lambda} = \frac{1}{\bar{x}}$. And the likelihood evaluated at $\hat{\lambda}$ is:

```
(1/xbar)^n*exp(-n)
```

```
## [1] 6.76327e-39
```

(b) Recall that the GLHR test considers evidence for H_1 to be smaller values of the test statistic Λ. A small value of Λ corresponds to a large value of $-1 \log \Lambda$ which is double the difference between the overall maximum log likelihood and the maximum under H_0. Evaluate $-2 \log(\Lambda)$ for the data.

Note we need to evaluate the value of the *likelihood* in the numerator and denominator (as opposed to the log likelihood)

```
num =   (1/xbar)^n*exp(-n)
denom =  (l_MLE^a_MLE/lgamma(a_MLE))^n * xg^(n*(a_MLE-1)) * exp(-l_MLE*n*xbar)
-2*log(num/denom)
```

```
## [1] 105.7407
```

Problem 3

For distributions in the exponential family, the null sampling distribution of $-2\log(\Lambda)$ is approximately Chi-square(m), where the degrees of freedom m is the difference in the number of parameters estimated overall vs the number of parameters estimated under H_0. Determine if there is evidence to reject the exponential model by evaluating and interpreting the appropriate degrees of freedom, finding the rejection threshold for an $\alpha = 0.05$ level test, and calculating the approximate p-value for the observed data.

This produces a p-value that is basically zero:

```
TS_obs = -2*log(num/denom)
(pval = 1-pchisq(TS_obs, df=1, lower.tail=TRUE))
```

so we have to answer analytically, based on our knowledge of the chi-square distribution and the p-value is approx zero.