Solution HW 10

December 12, 2019

Section 9.11 of Rice; Exercises 14, 18, 20, 24, 40-43, 49(a,b), 55, 58

9.14

(a) We look at the ratio of posterior probabilities

$$\frac{P(H_0|X)}{P(H_A|X)} = \frac{P(X|H_0)2P(H_A)}{P(X|H_A)P(H_A)} = 2\frac{P(X|H_0)}{P(X|H_A)}$$

$$= 2\frac{e^{\frac{-1}{2\sigma^2}x^2}}{e^{\frac{-1}{2\sigma^2}(x-1)^2}} = 2e^{\frac{1}{2\sigma^2}((x-1)^2-x^2)}$$

We want to see which values this is greater than 1 for:

$$2e^{\frac{1}{2\sigma^2}((x-1)^2-x^2)} > 1$$

$$\implies \frac{1}{2\sigma^2}(1-2x) > ln(1/2)$$

$$\implies 1-2x > ln(1/2)\sigma^2 \implies x < \frac{1-ln(1/2)2\sigma^2}{2}$$

For $\sigma^2 = 0.1, 0.5, 1, 5$ we get $x < 0.5693, 0.8466, 1.1931, 3.9657$, respectively.

(b) We can use a law of total probability idea (let $\sigma^2 = 1$),

$$P(X < 1.1931) = P(X < 1.1931|H_0)P(H_0) + P(X < 1.1931|H_A)P(H_A)$$

$$= pnorm(1.1931)(2/3) + pnorm(1.1931-1)(1/3) = 0.7812$$

9.18 -

$$X_i \overset{iid}{\sim} \frac{1}{2}\lambda e^{-\lambda|x|}$$

for $x > 0$ and $\lambda > 0$. Therefore,

$$L(\lambda) = \left(\frac{1}{2}\lambda\right)^n e^{-\lambda\sum_{i=1}^n |x_i|}$$

We want to test $H_0 : \lambda = \lambda_0$ vs $H_A : \lambda = \lambda_1$ where $\lambda_1 > \lambda_0$. The LRT says reject for small values of the likelihood ration.

$$\Lambda = \frac{L(H_0)}{L(H_A)} = \frac{\left(\frac{1}{2}\lambda_0\right)^n e^{-\lambda_0 \sum_{i=1}^n |x_i|}}{\left(\frac{1}{2}\lambda_1\right)^n e^{-\lambda_1 \sum_{i=1}^n |x_i|}}$$

$$= \left(\frac{\lambda_0}{\lambda_1}\right)^n e^{\sum_{i=1}^n |x_i|(\lambda_1 - \lambda_0)}$$

The rejection region is of the form

$$\left\{x_1, ..., x_n : \left(\frac{\lambda_0}{\lambda_1}\right)^n e^{\sum_{i=1}^n |x_i|(\lambda_1 - \lambda_0)} < c\right\}$$

To find the value of c we would need to know the null distribution of Λ. Alternatively, we could use an equivalent rejection region:

$$\iff e^{\sum_{i=1}^n |x_i|(\lambda_1 - \lambda_0)} < c_1$$

$$\iff \sum_{i=1}^n |x_i|(\lambda_1 - \lambda_0) < c_2$$

$$\iff \sum_{i=1}^n |x_i| < c_3 \text{ since } \lambda_1 > \lambda_0$$

Now our RR is

$$\left\{x_1, ..., x_n : \sum_{i=1}^n |x_i| < c_3\right\}$$

We would now need to know the distribution of $\sum_{i=1}^n |x_i|$ under H_0. In lieu of this, we could simulate many values of $\sum_{i=1}^n |x_i|$ and find an approximate cut off using the empirical α quantile of the observed values.
Since this rejection region didn't depend on the particular value of λ_1 (just that it was larger than λ_0), we can conclude that this is the UMP test for any $\lambda_1 > \lambda_0$.

9.20 -
By NP, the most powerful test is Reject H_0 for $\Lambda = \frac{L(H_0)}{L(H_A)} = \frac{1}{2x} < c$ or Reject H_0 for $x > c_1$. To have level $\alpha = 0.1$ we need

$$0.1 = P(\text{Reject } H_0|H_0) = P(X > c_1|f_0(x) = 1) = 1 - c_1 \implies c_1 = 0.9$$

Thus, we should reject for $x > 0.9$. This is the MP test for this alternative. Therefore, the power of an $\alpha = 0.1$ level test cannot exceed

$$PWR(H_A) = P(\text{Reject } H_0|H_A) = P(X > 0.9|f_0(x) = 2x)$$

$$= \int_{0.9}^{1} 2x dx = x^2|_{0.9}^{1} = 1 - 0.9^2 = 0.19$$

9.24 -

(a) $X \sim Bin(n,p)$ then $L(p) = \binom{n}{x} p^x (1-p)^{n-x}$. We have $H_0 : p = 0.5, H_A : p \neq 0.5$. The null space is $\omega_0 = \{p : p = 0.5\}$. The entire parameter space is $\Omega = \{p : 0 \leq p \leq 1\}$. The LR is then

$$\Lambda = \frac{L(0.5)}{L(\hat{p}_{MLE})} = \frac{\binom{n}{x} 0.5^x (1-0.5)^{n-x}}{\binom{n}{x} \hat{p}_{MLE}^x (1-\hat{p}_{MLE})^{n-x}}$$

$$= \left(\frac{n}{2x}\right)^x \left(\frac{1}{2(1-\hat{p}_{MLE})}\right)^{n-x}$$

We should reject if this is less than some value c that controls the type I error rate.

(b) Plotting $\left(\frac{(n/2)}{x}\right)^x \left(\frac{(n/2)}{n-x}\right)^{n-x}$ over $x = 0, 1, ..., n$ and $n = 20$ (chosen arbitrarily). We get the plot below (implying the given rejection region)

```
n<-20
x<-seq(from=0,to=n,by=1)
plot(x-n/2,(n/(2*x))^x*(n/(2*(n-x)))^(n-x))
```

(c) Under H_0, $X \sim Bin(n, 0.5)$.

$$\alpha = P\left(|X - n/2| > k | p = 0.5\right) = 1 - P\left(|X - n/2| \leq k | p = 0.5\right)$$

$$= 1 - P\left(-k + n/2 \leq X \leq k + n/2 | p = 0.5\right)$$

Now we use the fact that $X \overset{H_0}{\sim} Bin(n, 0.5)$ to find the value of k.

(d) $n = 10, k = 2$

$$\implies \alpha = 1 - P(-2 + 5 \leq X \leq 2 + 5 | p = 0.5)$$

$$= 1 - P(3 \leq X \leq 7 | p = 0.5) = 0.1094$$

(e) The normal approximation to the binomial says that

$$X \overset{\bullet}{\sim} N(np, np(1-p))$$

if we have a large sample.

$$\alpha = 1 - P(-10 + 100/2 \leq X \leq 10 + 100/2) = 1 - P(40 \leq X \leq 60)$$

$$= 1 - P\left(\frac{40 - 50}{\sqrt{100/4}} \leq Z \leq \frac{60 - 50}{\sqrt{100/4}}\right) = 1 - P(-2 \leq Z \leq 2) = 0.0455$$

9.40

$$X^2 = \frac{(X_1 - np_1)^2}{np_1} + \frac{(X_2 - np_2)^2}{np_2} = \frac{(X_1 - np_1)^2}{np_1} + \frac{(n - X_1 - n(1-p_1))^2}{n(1-p_1)}$$
$$= \frac{(X_1 - np_1)^2(1-p_1) + (np_1 - X_1)^2 p_1}{np_1(1-p_1)} = \frac{(X_1 - np_1)^2}{np_1(1-p_1)}$$

9.41

$$L(p_1, ..., p_m) = \prod_{i=1}^{m} \left[\binom{n_i}{x_i} p_i^{x_i} (1-p_i)^{n_i - x_i} \right]$$

Under ω_0 the max for the common value is at $\hat{p}_0 = \frac{\sum_{i=1}^m x_i}{\sum_{i=1}^m n_i}$.
Under Ω the max is at $\hat{p}_i = \frac{x_i}{n_i}$.

$$\implies \Lambda = \frac{\left(\prod_{i=1}^m \binom{n_i}{x_i}\right) (\hat{p}_0)^{\sum_{i=1}^m x_i} (1-\hat{p}_0)^{\sum_{i=1}^m (n_i - x_i)}}{\prod_{i=1}^m \left[\binom{n_i}{x_i} \hat{p}_i^{x_i} (1-\hat{p}_i)^{n_i - x_i}\right]} = \frac{(\hat{p}_0)^{\sum_{i=1}^m x_i} (1-\hat{p}_0)^{\sum_{i=1}^m (n_i - x_i)}}{\prod_{i=1}^m (\hat{p}_i)^{x_i} (1-\hat{p}_i)^{n_i - x_i}}$$

For a rejection region, we can use the large sample theory that says

$$-2ln\Lambda = 2\left[\sum_{i=1}^m n_i ln\left(\frac{1-\hat{p}_i}{1-\hat{p}_0}\right) + \sum_{i=1}^m x_i ln\left(\frac{\hat{p}_i(1-\hat{p}_0)}{\hat{p}_0(1-\hat{p}_i)}\right)\right] \overset{H_0}{\dot{\sim}} \chi^2_{m-1}$$

9.42

(a) Based on problem 9.41 we have that the MLE is$\hat{p}_0 = \frac{\sum_{i=1}^m x_i}{\sum_{i=1}^m n_i} = (0*157 + 1*69 + ... + 1*5)/(5*280) = 0.142$

(b) Using the Pearson Chi-square GOF test stat we just need the expected counts under the null hypothesis. Here

$$P(Y = y) = \binom{5}{y} \hat{p}_0^y (1-\hat{p}_0)^{5-y}$$

This gives expected counts of 280 times each probability, 130.09, 107.77, 35.72, and 5.92, respectively. Utilizing the Pearson Chi-square test stat we have

$$X^2 = \sum_{i=0}^{3} \frac{(obs_i - exp_i)^2}{exp_i} = \frac{(157 - 130.09)^2}{130.09} + \frac{(69 - 107.77)^2}{107.77} + ... = 44.53$$

Comparing this to the 0.95 χ^2_2 quantile we reject H_0 that the Binomial with a constant p models our data in favor of a more general model.

(c) Using 9.41 our test stat is

$$-2ln\Lambda \overset{H_0}{\overset{\bullet}{\sim}} \chi^2_{m-1}$$

where

$$\Lambda = \frac{(\hat{p}_0)^{\sum_{i=1}^m x_i}(1-\hat{p}_0)^{\sum_{i=1}^m (n_i - x_i)}}{\prod_{i=1}^m (\hat{p}_i)^{x_i}(1-\hat{p}_i)^{n_i - x_i}}$$

Plugging in we have $\Lambda = 5.24 \times 10^{-119}$ and $-2\log\Gamma = 544.70$. Compared with χ^2_5, we can conclude that binomial distribution is not a good fit here.

9.43

(a) Using Pearson's chi-square test,

$$\chi^2 = \sum_{i=1}^{2} \frac{(N_i - E_i)^2}{E_i} = \frac{(9207-8975)^2}{8975} + \frac{(8743-8975)^2)}{8975} = 11.99$$

Compare it with χ^2_1, we can get $p = 0.00053$. Thus we reject H_0.

(b)

Number of Heads	Frequency	Expected Frequency	$\frac{(N_i-E_i)^2}{E_i}$
0	100	111.24	1.15
1	524	560.04	2.32
2	1080	1123.67	1.70
3	1126	1123.67	0
4	655	560.40	16.80
5	105	111.29	0.36

Thus $\chi^2 = 21.63$ and the corresponding p-value is $p = 0.0006.(\chi^2_5)$

(c) In this problem,

$$H_0 : p_1 = p_2 = \cdots = p_5$$

Under H_0, we can calculate out $\hat{p}_{MLE} = 0.513$. Thus

Number of Heads	Frequency	Expected Frequency	$\frac{(N_i-E_i)^2}{E_i}$
0	100	98.34	0.03
1	524	517.96	0.07
2	1080	1091.23	0.12
3	1126	1149.49	0.48
4	655	605.43	4.06
5	105	127.55	3.99

$$P(\chi^2_4 \geq 8.75) = 0.068$$

. Thus we can not reject H_0.

9.49

```
#Input data
data<-c(rep(18.5,0),rep(19.0,1),rep(19.5,3),rep(20.0,33),rep(20.5,39),
        rep(21.0,156),rep(21.5,152),rep(22.0,392),rep(22.5,288),
        rep(23.0,286),rep(23.5,100),rep(24.0,86),rep(24.5,21),rep(25.0,12),
        rep(25.5,2),rep(26.0,0),rep(26.5,1))
#(a)constructing a histogram and superposing a normal density
hist(data,prob=TRUE)
curve(dnorm(x,mean(data),sd(data)),add=TRUE)

#(b) plotting on normal probability paper
qqnorm(data)
qqline(data)
```

9.55 - Compare to the normal distribution.

```
##########################################
##Problem Session 9 R code
##########################################

##9.55
#a
n<-25
plot(y=sort(rnorm(n)),x=qnorm((1:n-0.5)/n),main=paste("qq-plot of normal
data with n = ",n),
xlab="Normal quantiles",ylab="ordered observed values")
#repeat several times and for n=50,100

#b
plot(y=sort(rchisq(n,df=10)),x=qnorm((1:n-0.5)/n),main=paste("qq-plot of
chisquare data with
df=10 and n = ",n),xlab="Normal quantiles",ylab="ordered observed
values")
#repeat several times and for n=50,100


#d
plot(y=sort(runif(n)),x=qnorm((1:n-0.5)/n),main=paste("qq-plot of
uniform(0,1) data with
n = ",n), xlab="Normal quantiles",ylab="ordered observed values")
#repeat several times and for n=50,100

#e
plot(y=sort(rexp(n)),x=qnorm((1:n-0.5)/n),main=paste("qq-plot of exp(1)
data with n = ",n),
xlab="Normal quantiles",ylab="ordered observed values")
#repeat several times and for n=50,100

#f It is difficult for the chi-square, but reasonably clear for the uniform
and exponential
```

9.58

```
sec<-c(0,60,120,181,243,306,369,432,497,562,628,698,1130,1714,
       2125,2567,3044,3562,4130,4758,5460,6255,7174,8260,9590,
       11304,13719,14347,15049,15845,16763,17849,19179,20893,23309,
       27439)
Freq<-c(115,104,99,106,113,104,101,106,104,96,512,524,468,
        531,461,526,506,509,520,540,542,499,494,500,550,
        465,104,97,101,104,92,102,103,110,112,100)
n<-length(Freq)
cum<-rep(0,n)
cum[1]<-115
for(i in 2:n){
  cum[i]=cum[i-1]+Freq[i]
}
qua_x<-function(x){
  log((sum(Freq)+1)/(sum(Freq)+1-x))
}
plot(qua_x(cum),sec)
```