Problems Sheet 8

Statistics 2

Dom Hutchinson

```r
set.seed(16111998)
xtab<-c(109,65,22,3,1,0)
xobs<-rep(0:5,xtab)
n<-length(xobs)
xobs.bar<-mean(xobs)
```

Question 1

Let $\mathbf{X} := (X_1, \dots, X_{200}) \overset{\text{iid}}{\sim} \text{Poisson}(\lambda)$ be a random vector representing the number of deaths each year & $\mathbf{x} := (x_1, \dots, x_{200})$ be a realisation of $\mathbf{X}$.
Here we shall test $H_0 : \lambda \leq 0.5$ against $H_1 : \lambda > 0.5$.
Define test statistic $T(\mathbf{X}) = \frac{1}{n}\sum_{i=1}^n X_i$, from the data we have

$$T(\mathbf{x}) = -\frac{1}{200}\sum_{i=1}^{200} x_i = \frac{122}{200} = -0.61$$

Since $\sum_{i=1}^n X_i$ is an equivalent test statistic to the Neyamn-Pearson Test statistic then $T(\mathbf{X})$ is an equivalent test statistic too.
Thus $T(\mathbf{X})$ is the uniformly most powerful test statistic for H_0 & H_1 and has associated p-value

$$\begin{aligned}
p(\mathbf{X}) &= \mathbb{P}(T(\mathbf{X}) \geq T(\mathbf{x}); 0.5) \\
&= \mathbb{P}\left(\tfrac{1}{200}\textstyle\sum_{i=1}^{200} X_i \geq 0.61; 0.5\right) \\
&= \mathbb{P}\left(\textstyle\sum_{i=1}^{200} X_i \geq 122; 0.5\right) \\
&= \mathbb{P}\left(\textstyle\sum_{i=1}^{200} X_i \geq 122; 0.5\right) \\
&= \mathbb{P}(Y \geq 122) \text{ by independence where } Y \sim \text{Poisson}(200 \times 0.5) = \text{Poisson}(100) \\
&= 1 - \mathbb{P}(Y < 122) \\
&= 0.018
\end{aligned}$$

This is a very low p-value suggesting that the observered data is very extreme if H_0 is true.

Question 2

a)

We can approximate $Y \sim \text{Poisson}(100)$ as $\tilde{Y} \approx \text{Normal}(100, 100)$. Then

$$p(\mathbf{X}) = 1 - \mathbb{P}(Y < 122) \approx 1 - \mathbb{P}(\tilde{Y} < 122) = 0.4129356$$

b)

```r
lambda<-.5
trials<-1000

samples.raw<-sapply(1:trials, function(i) sum(rpois(200,lambda)))
```

```
count<-sum(samples.raw<=122)
p<-1-count/trials
p
```

```
## [1] 0.011
```

Question 3

Let $\mathbf{X} := (X_1, \dots, X_{200}) \overset{\text{iid}}{\sim} \text{Poisson}(\lambda)$ be a random vector representing the number of deaths each year & $\mathbf{x} := (x_1, \dots, x_{200})$ be a realisation of $\mathbf{X}$.
I shall test $H_0 : \lambda = 0.55$ against $H_1 : \lambda \neq 0.55$ at a significance level of 5% where $\Theta := \mathbb{R}^{>0}$.
Thus $\Theta_0 := \{0.55\}$ and is nested in Θ. Note that $\dim(\Theta) = 1$ and $\dim(\Theta_0) = 0$.
Since Θ_0 is nested within Θ we have

$$T_n(\mathbf{X}) := -2\ln \Lambda_n(\mathbf{X}) \to_{\mathcal{D}(\cdot;\theta)} \chi^2_{\dim(\Theta)-\dim(\Theta_0)} = \chi^2_{1-0} = \chi^2_1 \text{ where } \Lambda_n(\mathbf{x}) := \frac{\sup_{\lambda\in\Theta_0} f_n(\mathbf{x};\lambda)}{\sup_{\lambda\in\Theta} f_n(\mathbf{x};\lambda)}$$

We shall us $T_n(\mathbf{X})$ as our test statistic.
We have

$$\begin{array}{rcl} T(\mathbf{X}) & := & -2\ln \Lambda_n(\mathbf{X}) \\ & = & -2\ln \frac{\sup_{\lambda=0.55} f_n(\mathbf{x};\lambda)}{\sup_{\lambda\in\mathbb{R}^{>0}} f_n(\mathbf{x};\lambda)} \\ & = & -2\left[\ell_n(\mathbf{x};\lambda) - \ell_n(\mathbf{x};\hat{\lambda}_{MLE})\right] \end{array}$$

Since I am constructing $\alpha = 0.95$ confidence interval we wish to retain λ if

$$T_n(\mathbf{x}) := -2\left[\ell_n(\lambda;\mathbf{x}) - \ell_n(\hat{\lambda}_{MLE};\mathbf{x})\right] < \chi^2_{1,\alpha}$$

This is the confidence set

$$\begin{array}{rcl} C(\mathbf{x}) & = & \left\{\lambda \in \mathbb{R}^{>0} : T_n(\mathbf{x}) < \chi^2_{1,\alpha}\right\} \\ & = & \left\{\lambda \in \mathbb{R}^{>0} : -2\left[\ell_n(\lambda;\mathbf{x}) - \ell_n(\hat{\lambda}_{MLE};\mathbf{x})\right] < \chi^2_{1,\alpha}\right\} \\ & = & \left\{\lambda \in \mathbb{R}^{>0} : \ell_n(\lambda;\mathbf{x}) > \ell_n(\hat{\lambda}_{MLE};\mathbf{x}) - \frac{1}{2}\chi^2_{1,\alpha}\right\} \end{array}$$

There is a rule of thumb that at the 5 signifcance level the confidence set can be approximated as

$$C(\mathbf{x}) = \left\{\lambda \in \mathbb{R}^{>0} : \ell_n(\lambda;\mathbf{x}) \geq \ell(\hat{\theta}_n;\mathbf{x}) - 2\right\}$$

We can estimate the bounds of this set using r to calculate $\ell_n(\lambda;\mathbf{x})$ at regular intervals.

```
ell <- function(lambda) {
  stopifnot(all(lambda > 0))
  n <- length(xobs)
  n * (-lambda + mean(xobs) * log(lambda))
}

ell.mle<-optimise(ell,interval=c(0,1),maximum=TRUE)$objective
cat("ell.mle=",ell.mle,sep="")
```

```
## ell.mle=-182.3042
```

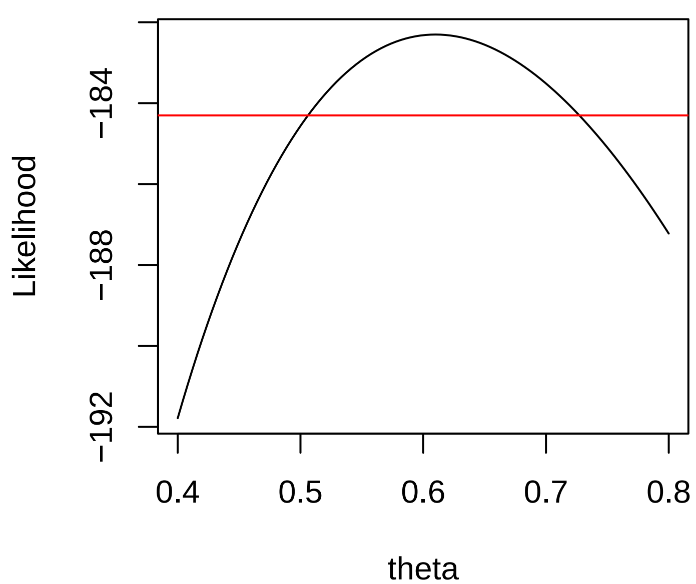
Giving

$$\begin{array}{rcl} C(\mathbf{x}) & = & \left\{\ell_n(\lambda;\mathbf{x}) > \ell_n(\hat{\lambda}_{MLE};\mathbf{x}) - 2\right\} \\ & = & \{\ell_n(\lambda;\mathbf{x}) > -182.3 - 2\} \\ & = & \{\ell_n(\lambda;\mathbf{x}) > -184.3\} \end{array}$$

```
x<-seq(0.4,.8,length.out=100)
y<-sapply(x,function(p) ell(p))
z<-x[which(y>ell.mle-2)] # The values where the inequality holds
cat("lower:",min(z),"\nupper:",max(z),sep="")
```

```
## lower:0.5090909
## upper:0.7232323
```

```
plot(x,y,type="l",xlab="theta",ylab="Likelihood")
abline(h=ell.mle-2,col="red")
```



Thus

$$C(\mathbf{x}) = \{\theta \in [0.509, 0.723]\}$$

We accept H_0 in this case.

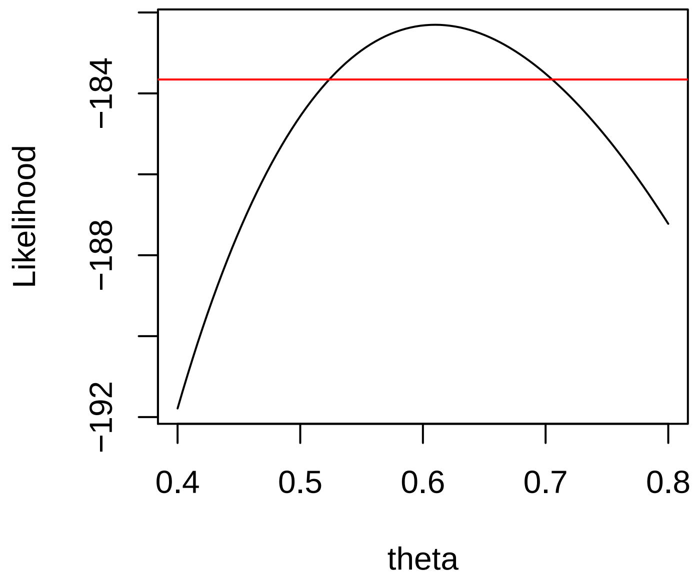
b)

```
alpha=0.9
quantile<-qchisq(alpha,1)

z<-x[which(y>ell.mle-.5*quantile)] # The values where the inequality holds
cat("chisq_alpha=",quantile,"\nlower:",min(z),"\nupper:",max(z))
```

```
## chisq_alpha= 2.705543
## lower: 0.5252525
## upper: 0.7030303
```

```
plot(x,y,type="l",xlab="theta",ylab="Likelihood")
abline(h=ell.mle-.5*quantile,col="red")
```



Thus

$$\begin{aligned} C(\mathbf{x}) &= \left\{\lambda \in \mathbb{R}^{>0} : \ell_n(\lambda;\mathbf{x}) > \ell_n(\hat{\lambda}_{MLE};\mathbf{x}) - \tfrac{1}{2}2.71\right\} \\ &= \left\{\lambda \in \mathbb{R}^{>0} : \ell_n(\lambda;\mathbf{x}) > -182.3 - 1.355\right\} \\ &= \left\{\lambda \in \mathbb{R}^{>0} : \ell_n(\lambda;\mathbf{x}) > -183.655\right\} \\ &= \{\lambda \in [0.525, 0.703] \text{ from R} \end{aligned}$$

We accept H_0 in this case.

Question 4

```
lambda<-round(optimise(ell,interval=c(0,1),maximum=TRUE)$maximum,2) # 2dp MLE
breaks<-c(-Inf,seq(0,4,by=1),Inf)
obs<-table(cut(xobs,breaks))
exp<-ppois(breaks[-1],lambda)-ppois(breaks[-length(breaks)],lambda)
exp<-n*exp
round(cbind(obs,exp),1)
```

```
##           obs   exp
## (-Inf,0] 109 108.7
## (0,1]      65  66.3
## (1,2]      22  20.2
## (2,3]       3   4.1
## (3,4]       1   0.6
## (4, Inf]    0   0.1
```

```
t_obs<-sum((obs-exp)^2/exp)
p_val<-1-pchisq(t_obs,df=length(breaks)-1)
cat("lambda=",lambda,"\nt_obs=",t_obs,"\np-val=",p_val,sep=" ")
```

```
## lambda= 0.61
```

```
## t_obs= 0.7903603
## p-val= 0.9923296
```

Here the p-value is high, suggesting that Poisson(0.61) is a good model for these observations.