NMAI061-22-EX2

Matej Nemec

1)Poisson distribution MLE formula derivation

Probability density function for Poisson distribution:

$$P(X = x) = \frac{\lambda^x e^- \lambda}{x!}$$

Likelihood function:

$$L(\lambda; x_1, ..., x_n) = \prod_{i=1}^{n} \frac{\lambda^{x_i} e^- \lambda}{x_i!}$$

We can use natural logarithm to simplify so we are able to replace product with sum, this gives log likelihood function:

$$l(\lambda; x_1, ..., x_n) = ln(\prod_{i=1}^{n} \frac{\lambda^{x_i} e^- \lambda}{x_i!})$$

Which can be written as:

$$l(\lambda; x_1, ..., x_n) = -n\lambda + ln(\lambda) \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} ln(x_i!)$$

Next we calculate the derivative of the log likelihood function w.r.t. λ and put it equal to 0:

$$\frac{d}{d\lambda} l(\lambda; x_1, ..., x_n) = -n + \frac{1}{\lambda} \sum_{i=1}^{n} = 0$$

Finally we solve for λ leaving us with MLE for λ equal to sample mean:

$$\lambda = \frac{1}{n} \sum_{i=1}^{n} x_i$$

2)MLE for simulated data

We keep the λ of 9 from previous excercise and randomly sample from the corresponding Poisson distribution. We sample only 20 times to make it somewhat interesting:

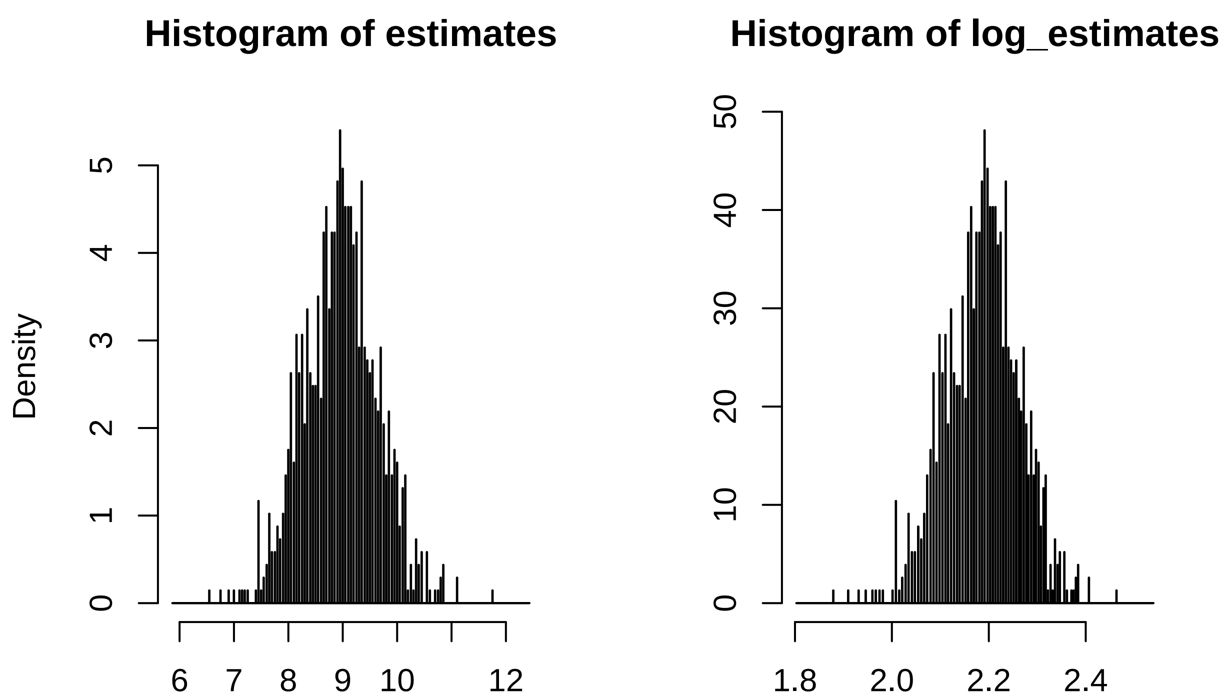
```
n=20
l=9
sample=rpois(n=n,lambda=l)
cat(paste0('Sample mean for ',n,' random samples with lambda ',l,' is ',mean(sample),'.'))
```

```
## Sample mean for 20 random samples with lambda 9 is 8.65.
```

3)Reapeated sampling and estimate distribution

Histogram for λ estimates from repeated random sampling:

```
repeats=1000
simulation=matrix(rpois(n*repeats,lambda=l),nrow=repeats)
estimates=rowSums(simulation)/n
par(mfrow=c(1,2))
hist(estimates,prob=TRUE,breaks = seq(min(estimates)-sd(estimates),
max(estimates)+sd(estimates),by = sd(estimates)/100),xlab='')
log_estimates=log(estimates)
hist(log_estimates,prob=TRUE,breaks = seq(min(log_estimates)-sd(log_estimates),
max(log_estimates)+sd(log_estimates), by = sd(log_estimates)/100),xlab='',ylab = '')
```



3)Confidence interval estimation and testing

We can see that λ of 9 falls within confidence intervals for all λ estimations. (Note: We only visualize 200 randomly picked estimations from 1000 for the graph to be readable):

```
margin=qnorm(0.975)*estimates/sqrt(20)
lower_cis=estimates-margin
upper_cis=estimates+margin
plot(NULL,ylim=c(min(estimates)-8*sd(estimates),max(estimates)+8*sd(estimates)),
xlim=c(0,200),ylab='',xlab='',main='CIs for lambda')
i=1
for(k in sample(1:1000, 200)) {
    lines(x=c(i,i),y=c(lower_cis[k],upper_cis[k]))
    i=i+1
}
abline(h=9)
```