69

STAT 8003, Homework 4

Group #8
Members: Anastasia Vishnyakova, Nooreen Dabbish, Yinghui Lu

Oct. 3, 2013

Problem 1.

In a large traumatic brain injury experiment, put injured and uninjured rats in a Morris water maze and determine whether each animal reaches the platform in 60 seconds. Repeat the experiment until the animal reaches the platform before 60 seconds and record the number of trials up until the first success. We are interested in estimating the success rate p. Now suppose we have an *i.i.d.* sample of size n.

Background:
The **geometric distribution**, which gives the probability of k failures before the first success.

$$P(X = k|p) = (1 - p)^k p$$

this is not Geo(p)
$X \sim$ Geo(p) if X is the total number of trial including the first success!

−1

a) Find the method of moments estimate of p

Method of moments

In an experiment, we determine n-samples of a real-valued random variable $X = \{x_1, x_2, \dots x_n\}$ and want to estimate the parameters of the distribution of X.

(a) Calculate the "population moments." The moments/expected values associated with the probability distribution.

$$E(X) = rp$$

?

(b) Calculate the "sample moments" associated with the real, observed values.

$$m_1 = \frac{1}{n}\sum_1^n x_n$$

$$m_2 = \frac{1}{n}\sum_1^n x_n^2$$

Wrong start point.

Here,

$$E(x) = \sum_{1}^{\infty} x(1-p)^x p$$
$$= (1-p)\sum_{1}^{\infty} x(1-p)^{x-1} p$$
$$= \frac{1-p}{p}$$

Show how to get the summation.

-3

We set this equal to the first sample momement:

$$E(X) = \frac{1}{n}\sum_{1}^{n} x_n$$
$$\frac{1-p}{p} = \bar{x}$$
$$p = \frac{1}{\bar{x}+1}$$

b) Find the MLE of p.

Maximum likelihood estimation

In this approach, a likelihood estimator is constructed using the sampled data-points. This estimator is a function of the distribution function parameters.

(a) Find the likelihood function. This is the product of the density functions for i.i.d. samples:

$$\mathcal{L}(\theta|x_1,\ldots,x_n) = f(x_1|\theta) \times \ldots \times f(x_n|\theta)$$
$$= \prod_{i=1}^{n} f(x_i|\theta)$$

(b) To make the multiplication of n terms more tractable, use the log-likelihood function:

$$\ell(\theta|x_1,\ldots,x_n) = \ln \mathcal{L}(\theta|x_1,\ldots,x_n) = \sum_{i=1}^{n} \ln f(x_i|\theta)$$

(c) Can find maxima using calculus—solving for points where ℓ' is zero and ℓ'' is negative. This gives the values of theta with the greatest likelihood.

Here we are given the distribution $P(X = k|p) = (1-p)^k p$.

$$\mathcal{L}(p|x_1, \ldots, x_n) = \prod_{i=1}^{n} f(x_i|\theta)$$
$$= \prod_{i=1}^{n} (1-p)^{x_i} p \qquad x_i = \{t_1, \ldots, t_n\}$$

number of trials recorded for each rat

$$\ell = \ln \mathcal{L}(p|x_1, \ldots, x_n) = n\ln(p) + \ln(1-p)n\bar{x}$$
$$\frac{d}{dp}\ell = \frac{n}{p} - \frac{n\bar{x}}{1-p}$$
$$p = \frac{1}{\bar{x}+1}$$
$$\ell'' = \frac{d^2}{d^2p}\ell = -\frac{n}{p^2} - \frac{n\bar{x}}{(1-p)^2} < 0$$

c) After the experiment, we get the following data. Calculate the MOM and MLE with the data.

Note that both MOM and MLE obtained the same value for an estimate of p, $p = \frac{1}{\bar{x}+1} = 0.143$.

Code from R

```
x <- c(19, 2, 2, 12, 2, 1, 1, 20, 0, 1)
n <- length(x)
sum_x <- sum(x)
x_bar <- sum_x/n
p <- 1/(x_bar + 1)
print(paste("The value of p obtained by MOM and MLE estimation was ", p, "."))
## [1] "The value of p obtained by MOM and MLE estimation was
0.142857142857143 ."
> sec_deriv <- -((n/p^2)-((n*x_bar)/(1-p^2)^2))
> print(paste("The second derivative at p=",p," was ",sec_deriv,"."))
[1] "The second derivative at p= 0.142857142857143  was  -427.473958333333 ."
```

(Note that a negative second derivative tells us that we found a value of p that is a maxima.)

Problem 2.

a) Let i be the number of right turns, where $i = 0, 1, 2, 3\cdots$, and X_i be the frequency of i times of right turns, then we have,

$$p(i) = \frac{\lambda^i \exp(-\lambda)}{i!}$$

thus

$$\mathcal{L}_n(\lambda) = \prod_{i=0}^{12} \left(\frac{\lambda^i \exp(-\lambda)}{i!}\right)^{x_i} \prod_{i=13}^{\infty} \left(\frac{\lambda^i \exp(-\lambda)}{i!}\right)^{0}$$
$$= \prod_{i=0}^{12} \left(\frac{\lambda^i \exp(-\lambda)}{i!}\right)^{x_i}$$

$$l_n(\lambda) = \log\lambda \sum_{i=0}^{12} i x_i - \lambda \sum_{i=0}^{12} x_i - \sum_{i=0}^{12} \log i!$$

$$l_n'(\lambda) = \frac{1}{\lambda} \sum_{i=0}^{12} i x_i - \sum_{i=0}^{12} x_i$$

Set $l_n'(\lambda) = 0$, then we have

$$\lambda = \frac{\sum_{i=0}^{12} i x_i}{\sum_{i=0}^{12} x_i}$$

Use R to get the solution is $\lambda = 3.89$.

```
Code from R:
> x=0:12
> y=c(14,30,36,68,43,43,30,14,10,6,4,1,1)
> z=x%*%y
> s=sum(y)
> lambda=z/s
> lambda
         [,1]
[1,] 3.893333
```

Then check $l_n''(\lambda) = -\frac{1}{\lambda^2} \sum_{i=0}^{12} i x_i < 0$

Hence, the Poisson distribution fit using MLE is

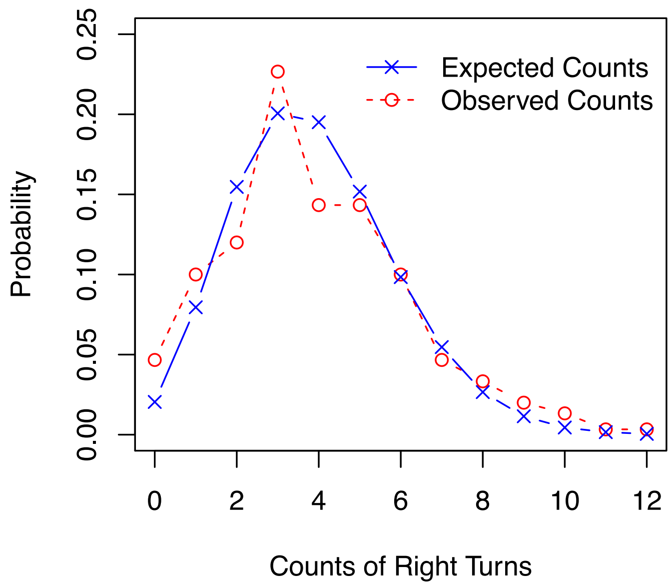
$$p(i) = \frac{3.89^i \exp(-3.89)}{i!}$$

b) Use R to draw the density distribution curve of the observed counts and the expected counts.

```
Code from R
> x=0:12
```

```
> y=c(14,30,36,68,43,43,30,14,10,6,4,1,1)
> pob=y/sum(y)
> pex=dpois(x,3.89)
> plot(x,pob,type="b",col="red",ylim=c(0,0.25),xlab="Counts of Right Turns",
ylab="Probability",lty=2,pch=1)
> lines(x,pex,type="b",col="blue",pch=4,lty=1)
> legend(5,0.25,c("Expected Counts","Observed Counts"),col=c("blue","red"),
pch=c(4,1),lty=c(1,2),bty="n")
```

From the figure we can see that the distribution of expected counts and the observed counts are in very similar shape, which suggest that our fit using MLE is a good fit.




```
Test if the estimate close to the mean value of the value (true for Poisson):

> x=0:12
> y=c(14,30,36,68,43,43,30,14,10,6,4,1,1)
> sample<- rep(t,f)
> mean(sample)
[1] 3.893333
The result is similar to the outcome of the MLE estimation.
```

Because the 300 intervals were distributed over various hours of the day and various days of the week, that is evenly distributed, Poisson distribution assumption returned a close fit.

Problem 3. Suppose X is a discrete random variable with $P(X = 1) = \theta$ and $P(X = 2) = 1 - \theta$. Three independent observations of X are made: $x_1 = 1, x_2 = 2, x_3 = 2$.

a). Find the method of moment estimate of θ .
Find $E(X)$:

$$\begin{aligned} E(X) &= \sum_x xp(X = x) \\ &= 1\theta + 2(1 - \theta) \\ &= 2 - \theta \end{aligned}$$

Find the first sample moment of X:

-2

$$\begin{aligned} E(\hat{X}) &= \frac{1}{n} \sum_{i=1}^{n} X_i \\ &= \frac{x_1 + x_2 + x_3}{n} \\ &= \frac{1 + 2 + 2}{3} \\ &= \frac{5}{3} \end{aligned}$$

Set expectation equal to sample estimates to find θ:

$$\begin{aligned} 2 - \theta &= \frac{5}{3} \\ \theta &= \frac{1}{3} \end{aligned}$$

-2

b). What is the likelihood function?

$$\mathcal{L}_n(\theta) = \prod_{i=1}^{n} p(x_i, \theta)$$

$$\mathcal{L}_n(\theta) = \prod_{i=1}^{n_1} \theta^{n_1} \prod_{i=1}^{n_2} (1 - \theta)^{n_2}$$

$$\begin{aligned} \mathcal{L}_n(\theta | x_1, x_2, x_3) &= \theta \cdot (1 - \theta) \cdot (1 - \theta) \\ &= \theta \cdot (1 - 2\theta + \theta^2) \\ &= \theta^3 - 2\theta^2 + \theta \end{aligned}$$

c). What is the MLE of ?

To optimize the likelihood function, we took the first derivative:

$$\mathcal{L}(\theta|x_1, x_2, x_3) = \theta^3 - 2\theta^2 + \theta \qquad \text{from b) above}$$
$$\mathcal{L}' = \frac{d}{d\theta}\mathcal{L} = 3\theta^2 - 4\theta + 1$$

And set it equal to zero:

$$0 = 3\theta^2 - 4\theta + 1$$
$$= (3\theta - 1)(\theta - 1)$$
$$\theta = \frac{1}{3}, 1$$

-2

We found the second derivative and did the second derivative test.

$$\mathcal{L}'' = 6\theta - 4$$
$$\mathcal{L}''\left(\frac{1}{3}\right) = -2$$
$$\mathcal{L}''(1) = +2 \qquad \text{Does not pass the second derivative test.}$$

So, the MLE of θ is $\frac{1}{3}$.