# Exercise5

*Chutian Chen cc4515; Congcheng Yan cy2550; Mingrui Liu ml4404*

## (1)

There are many independent and simultaneous discoveries or inventions. The two hypotheses about this are the social deterministic perspective and probabilistic model. The author thinks the later one deserves detailed inspection and he would like to look deeper in it based on three questions: the theoretical basis, the performance and the implication. The model is to apply Poisson distribution to the problem. The assumption is that the product of n and p is a constant $\mu$. I think it is reasonable to use Poisson distribution because firstly the inventions are random events and they happen with a probability. Also, Poisson distribution is more proper than Bernoulli because we cannot get enough information to estimate the parameters of Bernoulli while with the assumption, we only have one parameter in Poisson and the model could be studied better.

## (2)

The frequencies of $k = 0$ and $k = 1$ cannot be observed. So, in the truncated Poisson, we only count the situations when $k > 2$. To keep the total probability being 1, a coefficient is added.

## (3)

Let X be Poisson Distribution. $EY = \frac{1}{1-e^{-\mu}-\mu e^{-\mu}}(EX - P(X=0)*0 - P(X=1)*1) = \frac{\mu - \mu e^{-\mu}}{1-e^{-\mu}-\mu e^{-\mu}}$

## (4)

Log likelihood: $\sum_{i=1}^{n}(-\mu + y_i ln(\mu) - ln(y_i!) - ln(1 - e^{-\mu} - \mu e^{-\mu}))$

```r
library(ggplot2)
freq <- c(179,51,17,6,8,1,0,2)
y <- seq(2,9,1)
data <- as.data.frame(y)
data$freq <- freq
log_mle <- function(miu) {
  likehood <- function(miu,y) {
    return(-miu+y*log(miu)-log(factorial(y))-log(1-exp(-miu)-miu*exp(-miu)))
  }
  return(sum(likehood(miu,data$y)*data$freq))
}
miu <- seq(0.8,1.6,0.1)
mle <- sapply(miu,log_mle)
ggplot() + geom_line(aes(x = miu, y = mle))
```
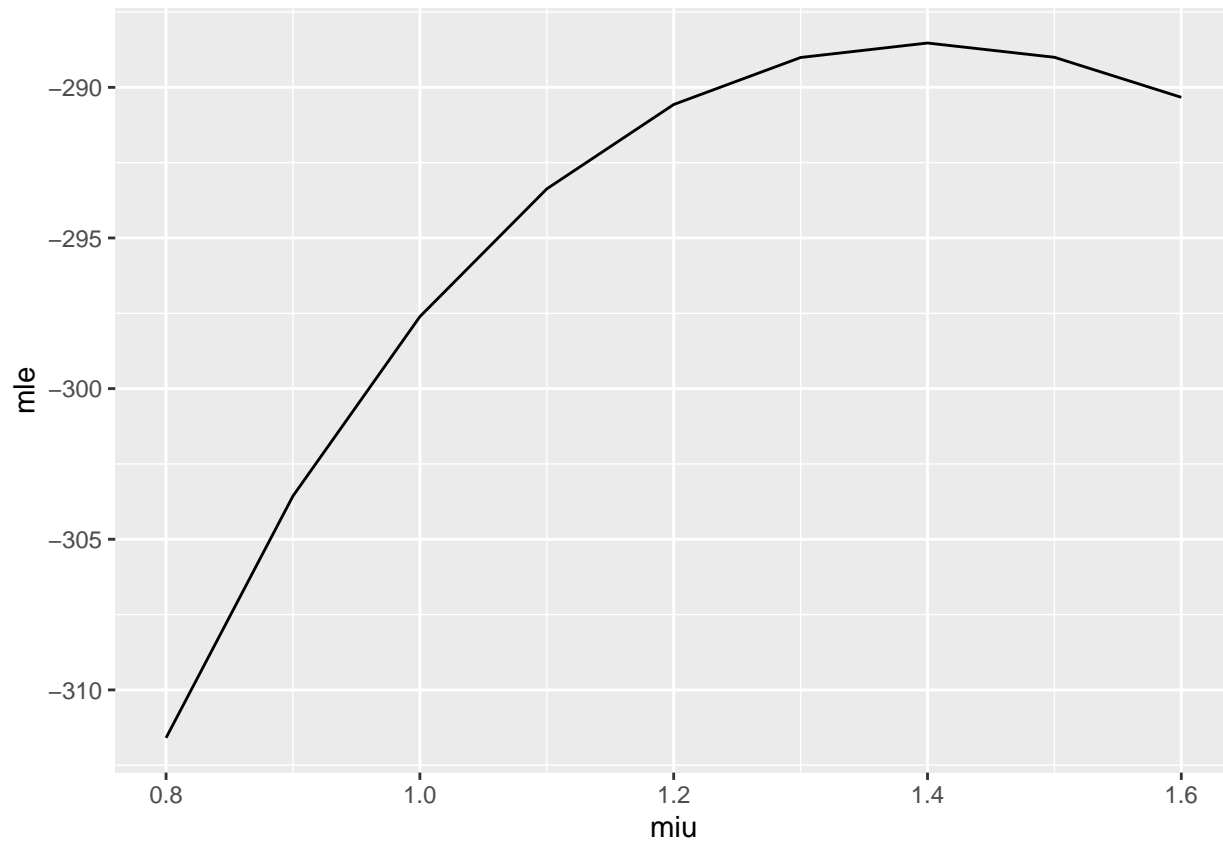
**(5)**

```
library(optim.functions)
initial <- c(1)
optim(initial, a <- function(x){return(-log_mle(x))}, method = "BFGS")
```

```
## $par
## [1] 1.398391
##
## $value
## [1] 288.5289
##
## $counts
## function gradient
##       11        4
##
## $convergence
## [1] 0
##
## $message
## NULL
```

We use BFGS. The MLE of $\mu$ is 1.398.

**(6)**

$$\sqrt{n}(\hat{\mu}_{MLE} - \mu) \to^D N(0, I^{-1}(\mu))$$

$$I(\mu) = -E\left(\frac{d^2(-\mu + y\ln(\mu) - \ln(y!) - \ln(1 - e^{-\mu} - \mu e^{-\mu}))}{d\mu^2}\right)$$

$$= E\left(\frac{y}{\mu^2} + \frac{e^\mu - \mu e^\mu - 1}{(e^\mu - 1 - \mu)^2}\right) = \frac{(e^\mu - 1)^2 - \mu^2 e^\mu}{\mu(e^\mu - 1 - \mu)^2}$$

## (7)

Through the asymptotic distribution of $\mu_{MLE} - \mu$, we can get a 0.95 asymptotic confidence interval.

$$[\hat{\mu}_{MLE} - \frac{Z_{0.975}}{\sqrt{nI(\hat{\mu})}}, \hat{\mu}_{MLE} + \frac{Z_{0.975}}{\sqrt{nI(\hat{\mu})}}]$$

```r
m <- 1.398
i <- ((exp(m)-1)^2-m^2*exp(m))/m/(exp(m)-1-m)^2
interval <- qnorm(0.975)/sqrt(sum(data$freq)*i)
print(c(m-interval,m+interval))
```

```
## [1] 1.197427 1.598573
```

95% CI is [1.20,1.60]

## (8)

It seems like a reasonable approach without using algorithm such as gradient descent.

However, this approach can't get many properties of the estimators like bias, consistency and variance. We can only compared the goodness of fit with the other estimators.

## (9)

In the paper $\mu = 1.4$. Our estimation of $\mu$ is 1.398, almost equal to the answer in paper. Simonton's technique is very accurate.