

Group: Zining Fan, Mutian Wang, Siyuan Wang

UNI: zf2234, mw3386, sw3418

Graded Homework 1 - Exercise 5

1. The author adopts a probabilistic approach, instead of a traditional social deterministic model, to explain the phenomenon that many independent techno-scientific discoveries/inventions occurred simultaneously in the history.

The main statistical model used by the author is Poisson distribution, an approximation to binomial distribution. Suppose there are n people working on a techno-scientific problem, and each person has the probability p to discover/invent something, then the number of repeats, X , is approximately a Poisson random variable.

Generally speaking, this model is reasonable. Intuitively, in this scenario, n should be large and p should be small, so Poisson distribution is a reasonable approximation. Also the CDF of Poisson distribution is much simpler than that of binomial distribution. More importantly, such approximation reduces the number of parameters from 2 to 1. However, the probability for each person to succeed is certainly not the same, and one person might have influence on others. In addition, the historical data of $X = 0$ and 1 are missing/inadequate, as mentioned in the article. Despite all these problems, this model is fairly good.

2. This is 1-truncated Poisson distribution, also known as conditional Poisson distribution. It is the Poisson distribution conditioned on the event $Y \geq 2$. The advantage is that 1-truncated Poisson distribution fits the data better, because it does not depend on $Y = 0$ and 1 and the historical data do not have such information.

3.

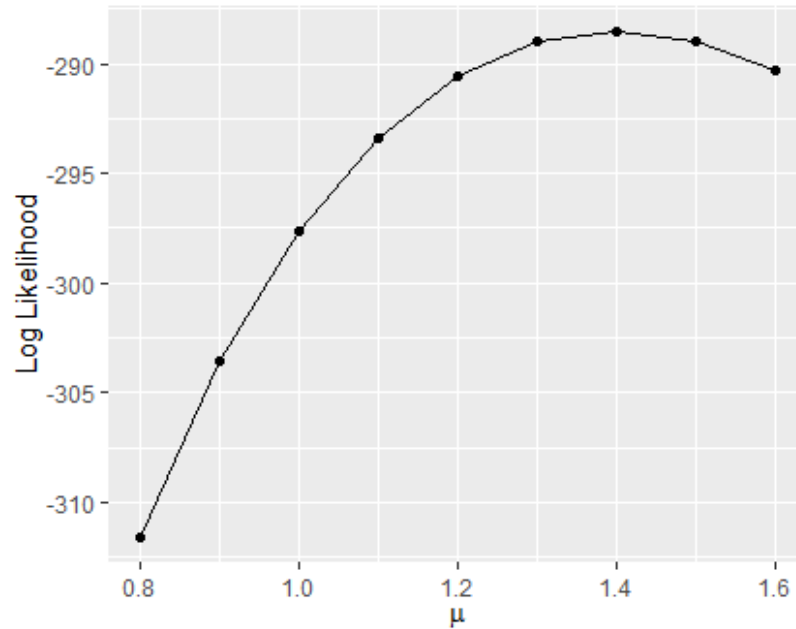
$$\begin{aligned} E[Y] &= \sum_{k=2}^{\infty} \left(\frac{k e^{-\mu} \mu^k}{k!} \cdot \frac{1}{1 - e^{-\mu} - \mu e^{-\mu}} \right) \\ &= \frac{1}{1 - e^{-\mu} - \mu e^{-\mu}} \cdot \left(\sum_{k=0}^{\infty} \frac{k e^{-\mu} \mu^k}{k!} - 0 \cdot e^{-\mu} - \mu e^{-\mu} \right) \\ &= \frac{\mu - \mu e^{-\mu}}{1 - e^{-\mu} - \mu e^{-\mu}} \end{aligned}$$

$$\begin{aligned} E[Y^2] &= \sum_{k=2}^{\infty} \left(\frac{k^2 e^{-\mu} \mu^k}{k!} \cdot \frac{1}{1 - e^{-\mu} - \mu e^{-\mu}} \right) \\ &= \frac{1}{1 - e^{-\mu} - \mu e^{-\mu}} \cdot \left(\sum_{k=0}^{\infty} \frac{k^2 e^{-\mu} \mu^k}{k!} - 0 \cdot e^{-\mu} - \mu e^{-\mu} \right) \\ &= \frac{\mu + \mu^2 - \mu e^{-\mu}}{1 - e^{-\mu} - \mu e^{-\mu}} \end{aligned}$$

$$\begin{aligned}
\text{var}(Y) &= E[Y^2] - E[Y]^2 \\
&= \frac{\mu + \mu^2 - \mu e^{-\mu}}{1 - e^{-\mu} - \mu e^{-\mu}} - \left(\frac{\mu - \mu e^{-\mu}}{1 - e^{-\mu} - \mu e^{-\mu}} \right)^2 \\
&= \frac{\mu - 2\mu e^{-\mu} + \mu e^{-2\mu} - \mu^3 e^{-\mu}}{(1 - e^{-\mu} - \mu e^{-\mu})^2}
\end{aligned}$$

4.

$$\begin{aligned}
\therefore L(\mu; Y) &= \prod_{i=1}^n \frac{e^{-\mu} \mu^{Y_i}}{Y_i!} \cdot \frac{1}{1 - e^{-\mu} - \mu e^{-\mu}} \\
&= \left(\frac{e^{-\mu}}{1 - e^{-\mu} - \mu e^{-\mu}} \right)^n \cdot \frac{\mu^{\sum Y_i}}{\prod Y_i!} \\
&= \left(\frac{1}{e^{\mu} - 1 - \mu} \right)^n \cdot \frac{\mu^{\sum Y_i}}{\prod Y_i!} \\
\therefore \ln L(\mu; Y) &= -n \cdot \ln(e^{\mu} - 1 - \mu) - \sum_{i=1}^n \ln Y_i! + \ln \mu \cdot \sum_{i=1}^n Y_i
\end{aligned}$$



5. The chosen algorithm is a built-in function in R, `nlm()`. It is the non-linear minimization using a Newton-type algorithm. We use this function to minimize the negative log likelihood.

```

data <- c(rep(2,179), rep(3,51), rep(4,17), rep(5,6),
          rep(6,8), rep(7,1), rep(9,2))
n <- length(data)
fun <- function(mu){
  n*log(exp(mu)-1-mu) + sum(log(factorial(data))) - log(mu)*sum(data)
}
nlm(fun, mu <- 1, hessian = T)

```

\$minimum

[1] 288.5289

\$estimate

[1] 1.398391

\$gradient

[1] -2.398301e-06

\$hessian

[,1]

[1,] 95.45589

\$code

[1] 1

\$iterations

[1] 6

6. Since MLE is always asymptotically normally distributed, we have

$$\sqrt{n}(\hat{\mu}_{\text{MLE}} - \mu) \xrightarrow[n \rightarrow \infty]{D} N(0, I(\mu)^{-1})$$

$$\begin{aligned} I(\mu) &= -E\left[\frac{\partial^2}{(\partial\mu)^2} \ln L(\mu; Y_1)\right] \\ &= \frac{e^\mu - \mu e^\mu - 1}{(e^\mu - \mu - 1)^2} + \frac{E[Y_1]}{\mu^2} \\ &= \frac{e^\mu - \mu e^\mu - 1}{(e^\mu - \mu - 1)^2} + \frac{\mu - \mu e^{-\mu}}{1 - e^{-\mu} - \mu e^{-\mu}} \cdot \frac{1}{\mu^2} \\ &= \frac{-\mu^2 e^\mu + e^{2\mu} - 2e^\mu + 1}{\mu(e^\mu - \mu - 1)^2} \end{aligned}$$

$$\therefore \hat{\mu}_{ML} \xrightarrow[n \rightarrow \infty]{D} N\left(\mu, \frac{1}{nI(\mu)}\right) \text{ where } I(\mu) = \frac{-\mu^2 e^\mu + e^{2\mu} - 2e^\mu + 1}{\mu(e^\mu - \mu - 1)^2}$$

7. Since we know the asymptotic distribution of $\hat{\mu}_{ML}$, we can construct a 0.95 asymptotic confidence interval as follows,

$$\left[\hat{\mu}_{ML} - \widehat{sd}(\hat{\mu}_{ML}) \cdot z_{0.975}, \hat{\mu}_{ML} + \widehat{sd}(\hat{\mu}_{ML}) \cdot z_{0.975}\right]$$

If we plug in values obtained earlier ($n = 264, \hat{\mu}_{ML} = 1.398$), we can get the confidence interval,

$$[1.197427, 1.598573]$$

8. Generally speaking, this is a reasonable way to estimate μ . Since the author picked the value of μ with the minimal χ^2 value, the chosen μ should best fit the data.

Yet we also have some doubts. Manually tuning the value of n_μ for each value of μ sounds strange to us. The optimum μ might be influenced by the author and cannot reflect the truth. Anyway, the result of this technique turns out quite good.

Because of the human intervention, we cannot calculate this estimator's bias, consistency or variance.

9. The results are almost identical. Simonton's estimate of μ is 1.4, while $\hat{\mu}_{\text{MLE}} = 1.398391$.

However, we can have more information about the maximum likelihood estimator, such as its distribution, bias, variance, consistency, etc.

Also, we only need to find a value of μ that maximizes the log likelihood, whereas we have to find the best pair (μ, n_μ) that minimizes the χ^2 value. So it seems the MLE method is somewhat less complex.