STAT5703 HW1 Ex5

Chao Huang (ch3474), Wancheng Chen (wc2687), Chengchao Jin (cj2628)

Bonus Question.

Question 1.

The author want to answer the question whether techno-scientific findings are inevitable or not by fitting the findings dataset using a Poisson model. The optimal parameter chosen after experiments tends to show that techno-scientific discoveries are not inevitable and highly depends on luck. I think for me, the choice of Poisson distribution seems reasonable, since techno-scientific findings are odd and can happen at a low probability. And Poisson distribution is quite suitable for modeling the probability of rare events happening.

Question 2.

Since there are no data for singleton and no-findings in the dataset, so using a truncated model will not give weird expected values for $k = 0$ or $k = 1$.

Question 3.

Suppose $X \sim Poisson(\mu)$, then we can derive the expectation and variance of Y using $\mathbf{E}[X]$ and $Var[X]$. We have,

$$\begin{aligned}
\mathbf{E}[X] &= \mu \\
&= \sum_{k=0}^{\infty} k \frac{e^{-\mu}\mu^k}{k!} \\
&= \mu e^{-\mu} + \sum_{k=2}^{\infty} k \frac{e^{-\mu}\mu^k}{k!}
\end{aligned}$$

If we denote $C = \frac{1}{1-e^{-\mu}-\mu e^{-\mu}}$, we have,

$$\begin{aligned}
\mathbf{E}[Y] &= C \sum_{k=2}^{\infty} k \frac{e^{-\mu}\mu^k}{k!} \\
&= C(\mu - \mu e^{-\mu})
\end{aligned}$$

Similarly, we can derive $Var[Y]$ with the help of $\mathbf{E}[X]$ and $\mathbf{E}[X^2]$. We have,

$$Var[Y] = C\mu(2\mu - e^{-\mu}) - C^2\mu^2(1 - e^{-\mu})^2$$

Question 4.

The data can be saved as a dataframe as below,

```r
tbl1 <- data.frame(
  k = seq.int(2, 9),
  count = c(179, 51, 17, 6, 8, 1, 0, 2)
)
tbl1
```

```
##   k count
## 1 2   179
## 2 3    51
```

```
## 3 4     17
## 4 5      6
## 5 6      8
## 6 7      1
## 7 8      0
## 8 9      2
```

Then the log likelihood function can be derived as,

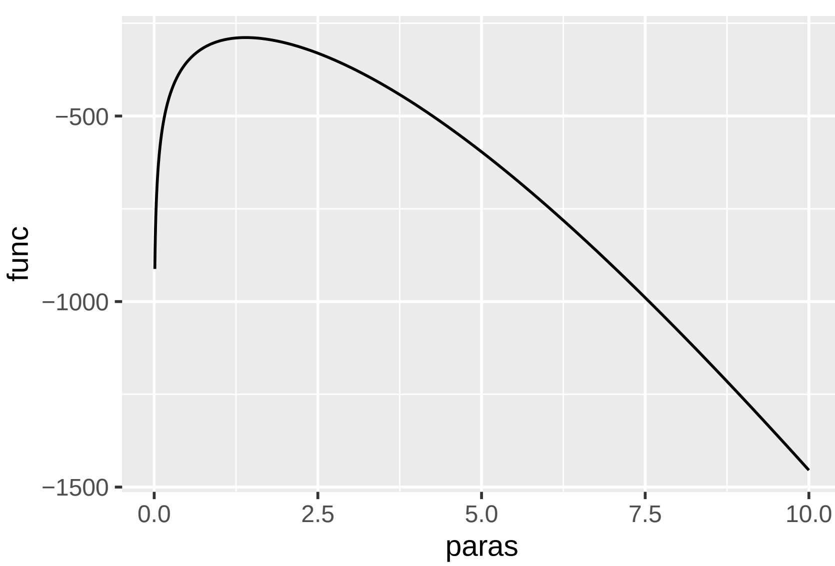
$$\begin{aligned}\log L &= \log(\prod_{k=2}^{9}(C\frac{e^{-\mu}\mu^k}{k!})^{COUNT_k}) \\ &= \sum_{k=2}^{9} COUNT_k \log(\frac{e^{-\mu}\mu^k}{k!})\end{aligned}$$

```
logL <- function (mu) {
  prob_dist <- function(x) {
    exp(-mu) * mu^x / factorial(x) /
      (1 - exp(-mu) - mu * exp(-mu))
  }
  data_ <- tbl1 %>%
    dplyr::mutate(prob=prob_dist(k)) %>% # likelihood
    dplyr::mutate(likelihood=count*log(prob)) # log-likelihood
  sum(data_$likelihood) # sum them up
}
```

And the log likelihood can be plotted as below,

```
plot_curve <- function(pars, f) {
 df_curve <- data.frame(
  paras = pars,
  func = unlist(lapply(pars, f))
 )
 ggplot(df_curve, aes(x=paras, y=func)) + geom_line()
}

plot_curve(10^seq.int(-2, 1, 0.01), logL)
```



Question 5.

The algorithm I choose is "BFGS", implemented in "optimx:optimx" function. Since it's a convex and nonlinear optimization problem as plotted above, this algorithm will converge shortly. The results and code are shown below,

```
opt <- optim(as.vector(c(1)), method = "BFGS", fn=function(x) {-logL(x)}, gr = NULL)
opt$par
```

```
## [1] 1.398391
```

Question 6.

Since the given distribution follows the regularity conditions, the asymptotic dsitribution would be a normal distribution,

$$\sqrt{n}(\hat{\mu}^{MLE} - \mu_0) \to N(0, I(\mu_0)^{-1})$$

where the fisher information can be calculated as,

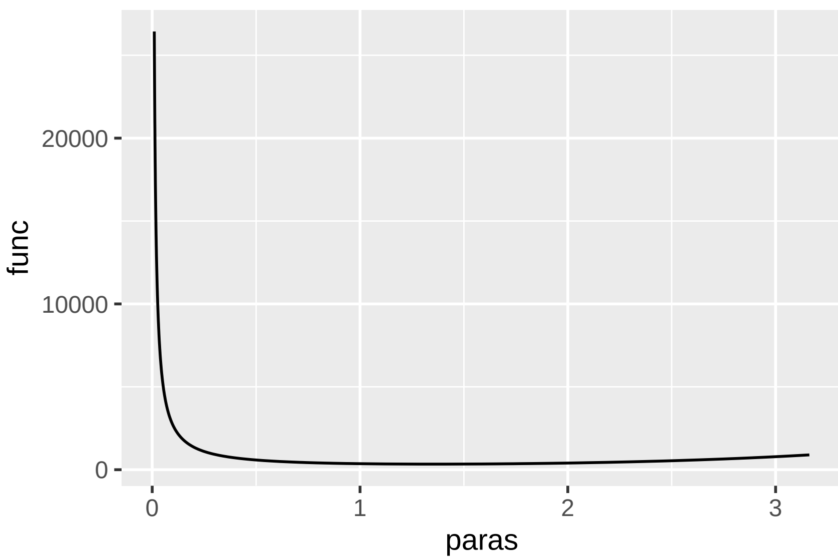
$$I(\mu_0) = -\mathbf{E}[\frac{\partial^2 \log(L)}{\partial \mu^2}] = \frac{n}{\mu} + \frac{n(\mu + e^{-\mu} - 1)}{e^{-\mu}(e^{\mu} - 1 - \mu)^2}$$

```
fisher <- function(mu) {
  n <- 264
  n / mu + n*(mu+exp(-mu)-1)/exp(-mu)/(exp(-mu)-1-mu)^2
}
optimize(fisher, lower=0, upper=10)
```

```
## $minimum
## [1] 1.35379
##
## $objective
## [1] 337.4855
```

Also, from the curve below, we can notice that the curve of fisher information around the MLE or optimal μ is quite flat. Therefore, we use MLE to calculate fisher information, which is 337.8257851.

```
plot_curve(10^seq.int(-2, 0.5, 0.01), fisher)
```



Question 7.

Given the asymptotic distribution given by Q6, we have the confidence interval as,

$$\mu \in [\mu_{ML} - 1.96\frac{I(\mu_{ML})^{-1}}{\sqrt{n}}, \mu_{ML} + 1.96\frac{I(\mu_{ML})^{-1}}{\sqrt{n}}] = [1.39803, 1.39875]$$

Question 8.

It seems like a reasonable choice since different groups of majors have quite different value of μ as mentioned in the paper. But it would be hard to evaluate the mathematical properties of this estimator.

Question 9.

Our ML estimator is 1.3983907, which is quite similar to the result ($\mu = 1.4$) given by the paper. Both of them can show evidence that the techno-scientific findings are not inevitable.