Homework 5 Solutions

3190300985 LUIS LUZERN YUVEN

1

```
percentile_ratio_discrepancies <- function(p99, p99.5, p99.9, a){
  return (((p99/p99.9)^(-a+1) - 10)^2 +
          ((p99.5/p99.9)^(-a+1) - 5)^2 +
          ((p99/p99.5)^(-a+1) - 2)^2)
}
percentile_ratio_discrepancies(1e6, 2e6, 1e7, 2)
```

```
## [1] 0
```

2

```
exponent.multi_ratios_est <- function(p99, p99.5, p99.9){
  a <- 1 - (log(10)/log(p99/p99.9))
  o <- optimize(percentile_ratio_discrepancies, p99=p99,
                p99.5=p99.5, p99.9=p99.9, interval=c(-100,100))
  return (o$minimum)
}
exponent.multi_ratios_est(1e6, 2e6, 1e7)
```

```
## [1] 1.999985
```

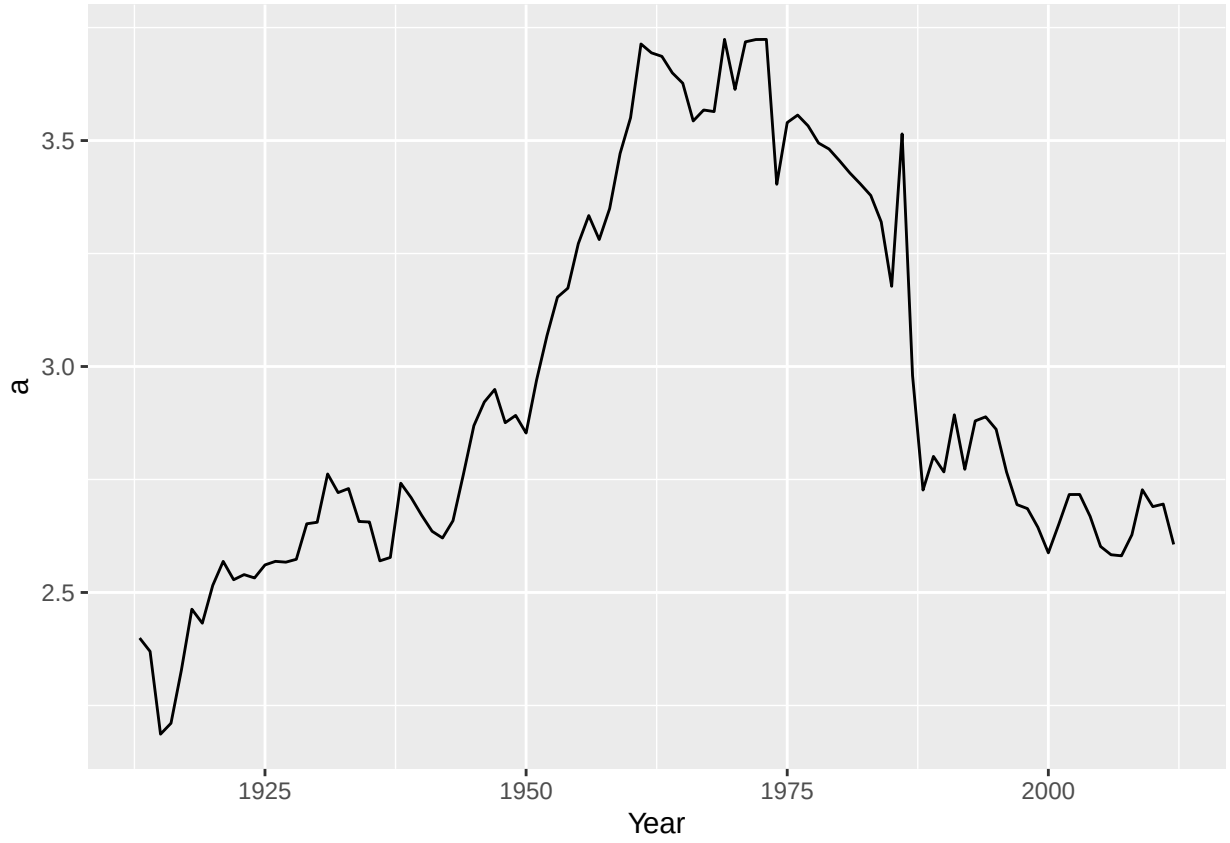
```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.9
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.1
```

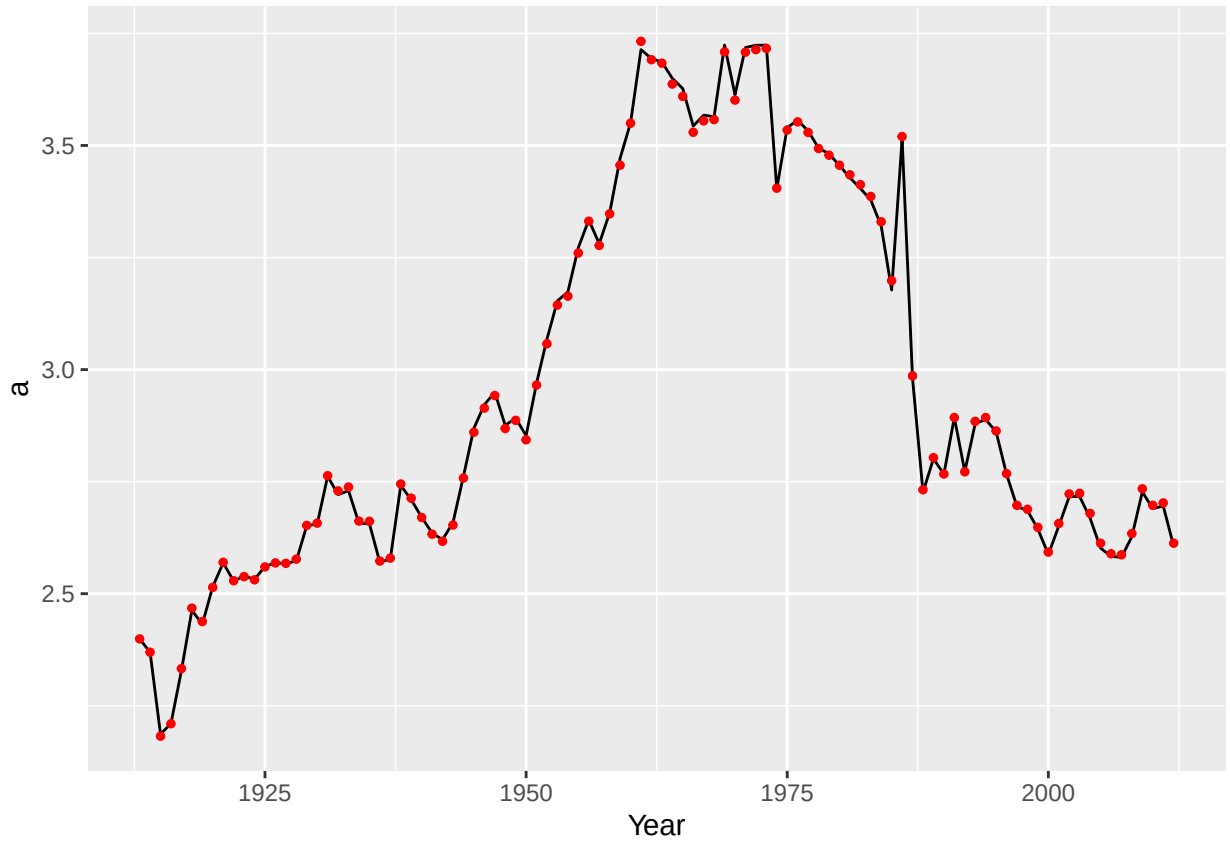
```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
data <- read.csv("data/wtid-report.csv")
row <- dim(data)[1]
est <- vector(length = row)
for(i in 1 : 100){
  est[i] <- exponent.multi_ratios_est(data$P99.income.threshold[i],
                                      data$P99.5.income.threshold[i],
                                      data$P99.9.income.threshold[i])
}
(g <- ggplot() + geom_line(aes(x = data$Year, y = est)) +
  labs(x = "Year", y = "a"))
```



4

```
est1 <- 1 - log(10) / log(data$P99.income.threshold / data$P99.9.income.threshold)
(g <- g + geom_point(aes(x = data$Year, y = est1), col = "red", size = 1))
```



We can see from the above plot, the estimates given by the function `exponent.multi_ratios_est` and by formula (4) are very close, with only slight differences at some points. It can be concluded that both methods are valid to estimate `a`