Homework 5

Yuhao Mao

2020/7/13

1. The `percentile_ratio_discrepancies()` function.

```
percentile_ratio_discrepancies <- function(P99, P99.5, P99.9, a){
  term1 <- ((P99/P99.9)^(-a+1)-10)^2
  term2 <- ((P99.5/P99.9)^(-a+1)-5)^2
  term3 <- ((P99/P99.5)^(-a+1)-2)^2
  return(term1+term2+term3)
}
percentile_ratio_discrepancies(1e6,2e6,1e7,2)
```

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

2. Estimation function.

```
exponent.multi_ratios_est <- function(P99, P99.5, P99.9){
  result <- optimize(percentile_ratio_discrepancies, P99=P99, P99.5=P99.5, P99.9=P99.9, interval = c(0,
  return(result$minimum)
}
P99 <- 1e6
P99.5 <- 2e6
P99.9 <- 1e7
exponent.multi_ratios_est(P99, P99.5, P99.9)
```

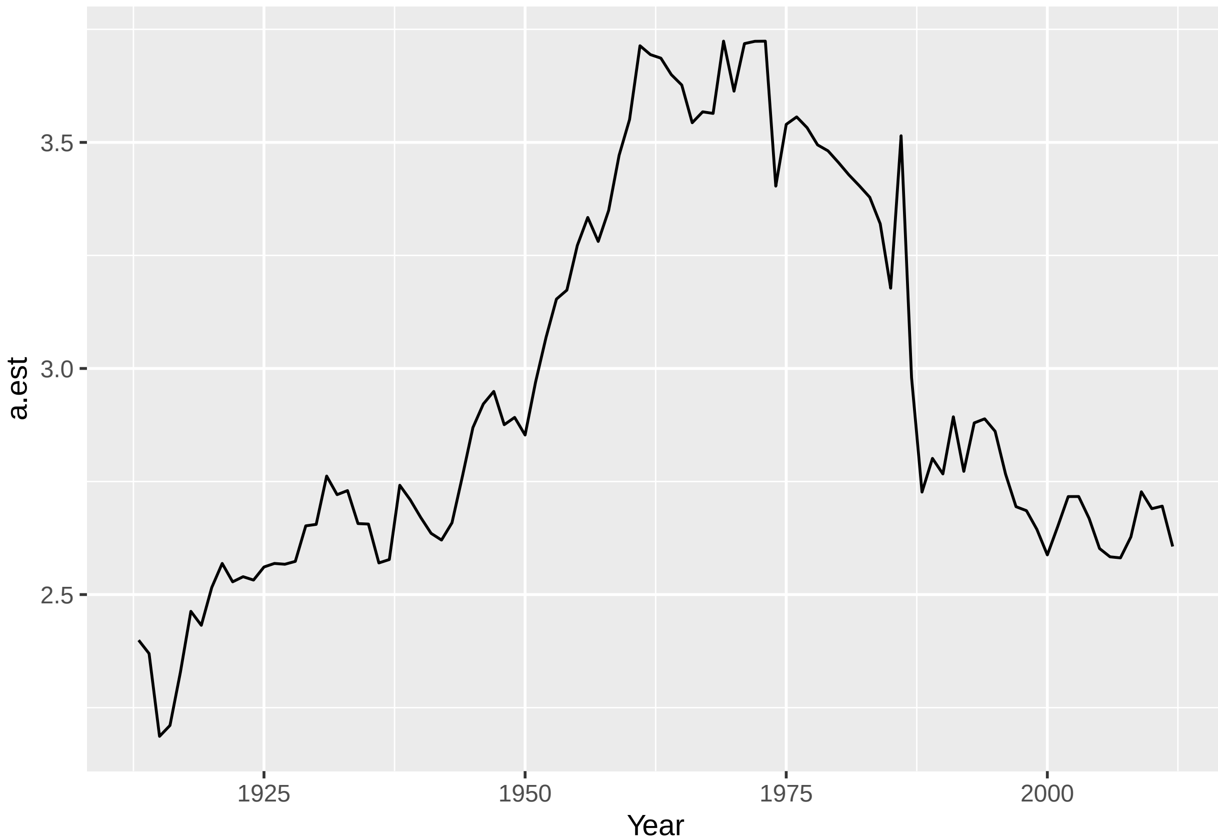
```
## [1] 2.000001
```

3. Estimate a for the US from 1913 to 2012.

```
wt <- read_csv('data/wtid-report.csv') %>% select(Country,
                                                  Year,
                                                  `P99 income threshold`,
                                                  `P99.5 income threshold`,
                                                  `P99.9 income threshold`)
```

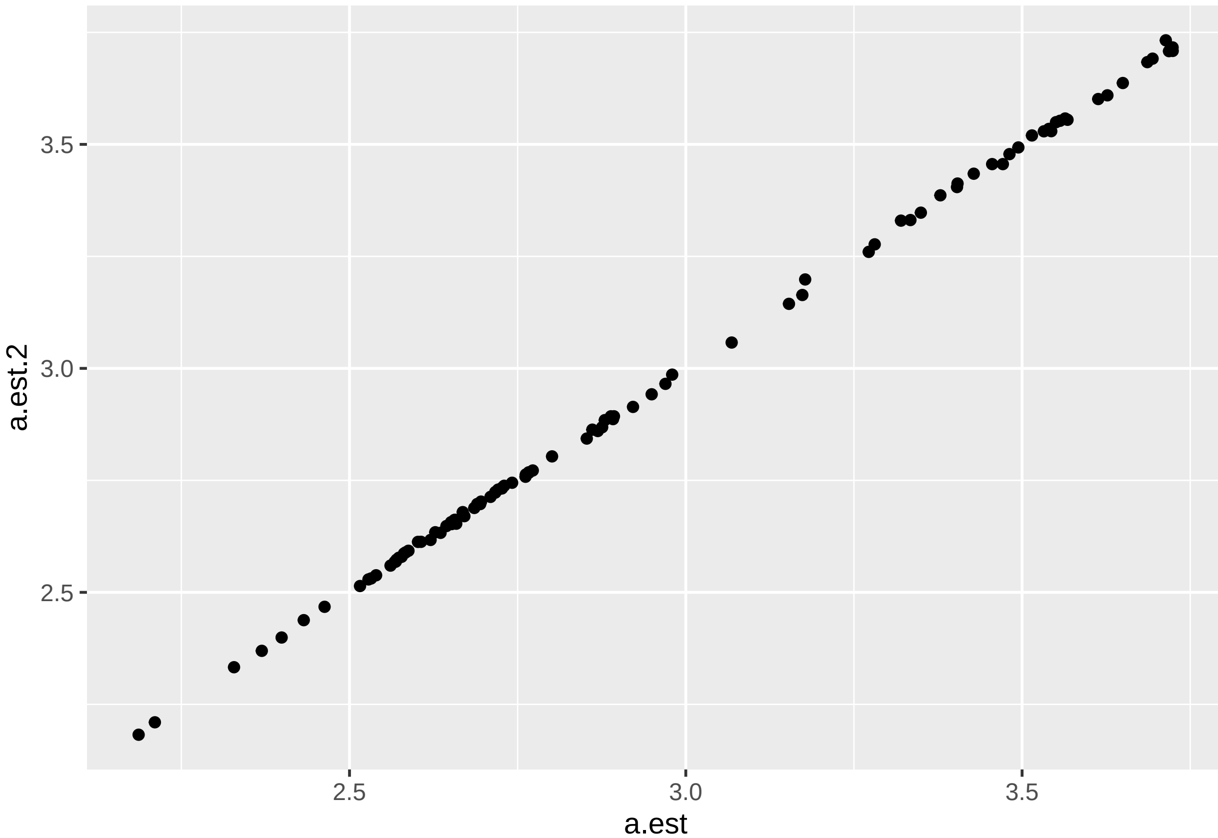
```
est.a <- function(x){
  P99 <- as.numeric(x[[3]])
  P99.5 <- as.numeric(x[[4]])
  P99.9 <- as.numeric(x[[5]])
  return(exponent.multi_ratios_est(P99,P99.5,P99.9))
}
```

```r
wt$a.est <- apply(wt, 1, est.a)
wt %>% ggplot()+geom_line(aes(Year,a.est))
```



4. Estimation from $a = 1 - \frac{\log 10}{\log(P99/P99.9)}$

```r
wt$a.est.2 <- 1-log(10)/log(wt$`P99 income threshold`/wt$`P99.9 income threshold`)
wt %>% ggplot()+geom_point(aes(a.est,a.est.2))
```




```r
all.equal(wt$a.est, wt$a.est.2)
```

```
## [1] "Mean relative difference: 0.001873415"
```

They are almost equal but not identical. That means a derived from equation $a = 1 - \frac{\log 10}{\log(P99/P99.9)}$ is a good answer for the MSE loss.