homework5

zza

2025-07-01

1

```
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(sum(term1, term2, term3))
}


test_case <- percentile_ratio_discrepancies(P99 = 1e6, P99.5 = 2e6, P99.9 = 1e7, a = 2)
cat(" 验证案例误差值: ", test_case, "（应返回 0）")
```

```
## 验证案例误差值:  0 （应返回0）
```

```
exponent.multi_ratios_est <- function(P99, P99.5, P99.9) {

  initial_a <- 1 - log(10) / log(P99 / P99.9)


  result <- optim(
    par = initial_a,
    fn = percentile_ratio_discrepancies,
    P99 = P99,
    P99.5 = P99.5,
    P99.9 = P99.9,
```

```
    method = "Brent",
    lower = 1,
    upper = 10
  )


  return(result$par)
}


test_est <- exponent.multi_ratios_est(P99 = 1e6, P99.5 = 2e6, P99.9 = 1e7)
cat(" 验证案例估计 a 值: ", test_est, "（应返回 2）")
```

```
## 验证案例估计a值:  2 （应返回2）
```

3

```
data <- read.csv('wtid-report.csv')


colnames(data)
```

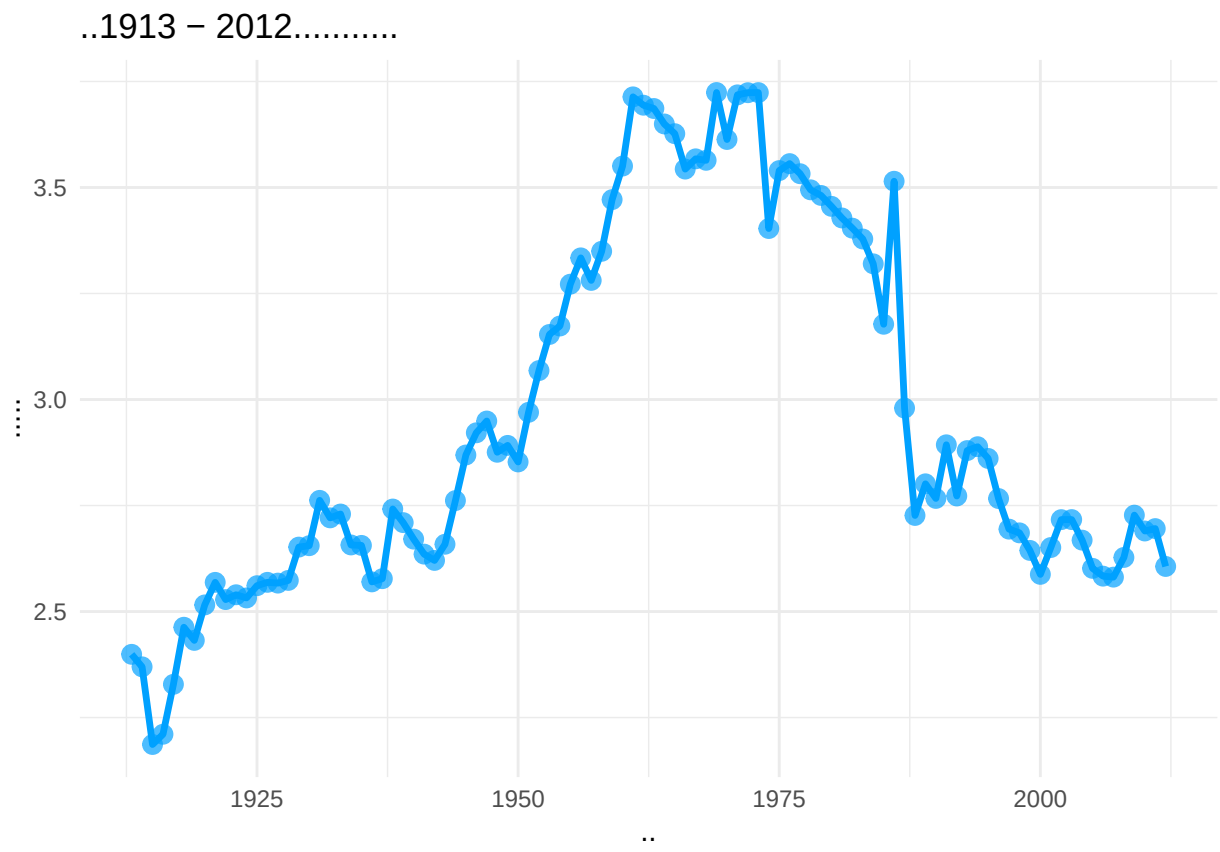
```
## [1] "Country"                 "Year"
## [3] "P90.income.threshold"    "P95.income.threshold"
## [5] "P99.income.threshold"    "P99.5.income.threshold"
## [7] "P99.9.income.threshold"  "P99.99.income.threshold"
```

```
us_data <- subset(data, Country == 'United States' & Year >= 1913 & Year <= 2012)


us_data$estimated_a <- apply(us_data[, c('P99.income.threshold', 'P99.5.income.threshold', 'P99.9.
  exponent.multi_ratios_est(x[1], x[2], x[3])
})


ggplot(us_data, aes(x = Year, y = estimated_a)) +
  geom_line(color = "#00A1FF", size = 1.2) +
  geom_point(color = "#00A1FF", size = 3, alpha = 0.7) +
```

```
  labs(x = '年份', y = '帕累托指数', title = '美国 1913 - 2012 年帕累托指数时间序列图') +
  theme_minimal()
```



4

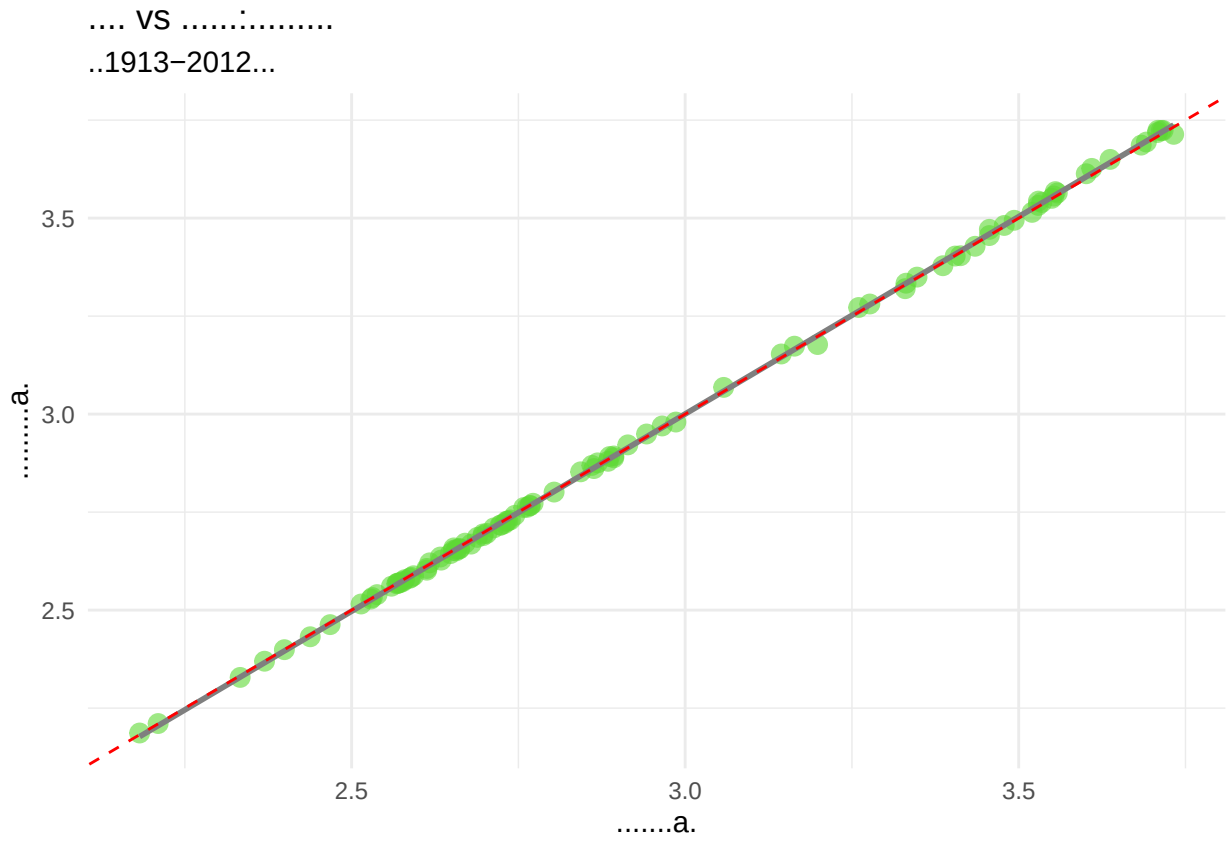
```
us_data$single_ratio_a <- 1 - log(10) / log(us_data$P99.income.threshold / us_data$P99.9.income.th


if(any(is.na(us_data$single_ratio_a))) {
  warning(" 存在 NA 值，可能是由于 P99 或 P99.9 为 0 或负值")
  us_data <- us_data[!is.na(us_data$single_ratio_a), ]
}



ggplot(us_data, aes(x = single_ratio_a, y = estimated_a)) +
```

```
  geom_point(color = "#5ed935", size = 3, alpha = 0.6) +
  geom_smooth(method = "lm", color = "gray50", alpha = 0.2) +
  geom_abline(intercept = 0, slope = 1, linetype = "dashed", color = "red") +
  labs(
    title = " 单比率法 vs 多比率联合法: 帕累托指数估计对比",
    subtitle = " 美国 1913-2012 年数据",
    x = " 单比率法估计的 a 值",
    y = " 多比率联合法估计的 a 值"
  ) +
  theme_minimal()
```




```
correlation <- cor(us_data$single_ratio_a, us_data$estimated_a)
mean_diff <- mean(us_data$estimated_a - us_data$single_ratio_a)
sd_diff <- sd(us_data$estimated_a - us_data$single_ratio_a)

cat(" 两种方法的相关系数:", round(correlation, 3), "\n")
```

```
## 两种方法的相关系数: 1
```

```r
cat(" 多比率法估计平均比单比率法高:", round(mean_diff, 3), "\n")
```

```
## 多比率法估计平均比单比率法高: 0
```

```r
cat(" 两种方法差异的标准差:", round(sd_diff, 3), "\n")
```

```
## 两种方法差异的标准差: 0.007
```