

## Sample Questions: Maximum Likelihood Part 1

STA312 Fall 2023. Copyright information is at the end of the last page.

Let  $X_1, \dots, X_n$  be independent Pareto random variables with density  $f(x|\theta) = \begin{cases} \frac{\theta}{x^{\theta+1}} & \text{for } x \geq 1 \\ 0 & \text{for } x < 1 \end{cases}$  where  $\theta > 0$ . The Pareto has a decreasing density with a heavy right tail, sometimes used as a model for the unequal distribution of wealth.

1. Derive a formula for the maximum likelihood estimate of  $\theta$ . Include the second derivative test. Show your work and **circle your final answer**.

$$\begin{aligned} l(\theta) &= \log \prod_{i=1}^n \frac{\theta}{x_i^{\theta+1}} = \log \left( \theta^n \frac{1}{\left( \prod_{i=1}^n x_i \right)^{\theta+1}} \right) \\ &= n \log \theta - \log \left( \left( \prod_{i=1}^n x_i \right)^{\theta+1} \right) \\ &= n \log \theta - (\theta+1) \sum_{i=1}^n \log x_i \end{aligned}$$

$$\begin{aligned} l'(\theta) &= \frac{n}{\theta} - \frac{d}{d\theta} \left[ \theta \sum_{i=1}^n \log x_i + \sum \log x_i \right] \\ &= \frac{n}{\theta} - \sum_{i=1}^n \log x_i \stackrel{\text{set}}{=} 0 \end{aligned}$$

$$\Rightarrow \frac{n}{\theta} = \sum_{i=1}^n \log x_i \Rightarrow \hat{\theta} = \frac{n}{\sum_{i=1}^n \log x_i}$$

2nd der test:

$$l''(\theta) = \frac{d}{d\theta} \left( n\theta^{-1} - \sum_{i=1}^n \log x_i \right)$$

$$= n(-1)\theta^{-2} - 0$$

$$= \frac{-n}{\theta^2} < 0 \quad \text{NEG} \quad \text{KCD} \quad \text{MAX}$$

$$\text{So } \hat{\theta} = \frac{n}{\sum \log x_i}$$

2. Give a formula for  $\hat{v}_n$ , the estimated asymptotic variance of  $\hat{\theta}_n$ . Show a little work.

$$\hat{v}_n = \frac{-1}{l''(\hat{\theta})} = \frac{-1}{-n/\hat{\theta}^2} = \frac{\hat{\theta}^2}{n} = \frac{1}{n} \left( \frac{n}{\sum \log x_i} \right)^2 = \frac{1}{\left( \sum \log x_i \right)^2}$$

3. The file <http://www.utstat.toronto.edu/~brunner/data/legal/pareto.data.txt> has a set of raw data. Calculate

- The maximum likelihood estimate  $\hat{\theta}_n$ .
- A 95% confidence interval for  $\theta$ .

Your answers are numbers. Circle and label them. Bring your printout to the quiz.

```
> rm(list=ls())
> x = scan("http://www.utstat.toronto.edu/~brunner/data/legal/pareto.data.txt")
Read 150 items
> x
 [1] 5.47 2.54 4.01 1.22 2.74 4.99 1.24 4.35 227.65
[10] 3.20 4.35 1.02 1.17 3.49 1.61 10.43 9.04 1.07
[19] 4.80 1.14 1.41 36.62 5.38 1.98 1.43 1.54 1.42
[28] 1.06 1.68 1.44 1.52 2.25 1.62 1.01 53.79 1.11
[37] 1.52 28.39 15.55 3.96 2.73 6.43 4.35 1.29 2.04
[46] 1.04 1.68 1.89 1.78 2.57 1.39 5.49 1.07 1.74
[55] 5.68 1.43 1.58 42.42 2.11 1.07 1.27 1.03 1.02
[64] 10.92 1.43 2.18 6.28 1.81 4.42 1.93 2.39 3.75
[73] 1.65 1.01 1.31 2.66 1.08 1.36 1.22 2.20 2.79
[82] 1.11 2.01 3.11 2.02 2.21 1.05 5.69 1.16 5.47
[91] 2.19 1.22 1.37 1.37 1.63 3.55 1.13 1.26 1.21
[100] 1.35 4.36 4.59 1.47 2.22 4.34 2.19 1.80 1.68
[109] 31.22 3.63 1.01 1.60 2.39 2.21 1.22 1.54 2.39
[118] 1008.06 3.37 1.30 4.46 1.01 7.51 1.14 1.02 1.12
[127] 2.83 1.79 1.67 1.45 1.74 1.66 1.01 1.34 1.12
[136] 1.30 1.18 10.96 2.28 2.33 1.51 1.41 1.19 1.40
[145] 1.83 1.33 31.98 5.86 3.28 1.53
> thetahat = 1/mean(log(x)); thetahat
[1] 1.100816
> n = length(x); vhat = thetahat^2/n; se = sqrt(vhat); se
[1] 0.08988125
> low95 = thetahat - 1.96*se; up95 = thetahat + 1.96*se
> c(low95,up95) # 95% CI
[1] 0.9246488 1.2769833
```

confidence interval

4. For the Pareto distribution, the well-known 80-20 rule (80% of the wealth is held by 20% of the population) corresponds to a value of  $\theta = 1.16$ . Using a two-sided large-sample  $Z$ -test and the usual  $\alpha = 0.05$  significance level, test whether these data are compatible with  $\theta = 1.16$ .
- (a) There are two critical values, one for the lower tail and one for the upper tail. What are they? The answers are numbers.
- (b) What is the value of the test statistic? The answer is a number. Circle it.
- (c) Use R to calculate the 2-sided  $p$ -value. The answer is a number.

```

> # Critical value(s). Just say plus and minus 1.96, or ...
> c(qnorm(0.025),qnorm(0.975))
a [1] -1.959964  1.959964
>
> Z = (thetahat-1.16)/se; Z
b [1] -0.6584683
>
> pvalue = 2 * (1-pnorm(abs(Z))); pvalue
c [1] 0.5102373

```

- (d) Do you reject the null hypothesis? Answer Yes or **No.**
- (e) Are the results statistically significant? Answer Yes or **No.**
- (f) Do these data contradict claim that  $\theta = 1.16$ ? Answer Yes or **No.**

---

This assignment was prepared by Jerry Brunner, Department of Mathematical and Computational Sciences, University of Toronto. It is licensed under a Creative Commons Attribution - ShareAlike 3.0 Unported License. Use any part of it as you like and share the result freely. The L<sup>A</sup>T<sub>E</sub>X source code is available from the course website:

<http://www.utstat.toronto.edu/brunner/oldclass/312f23>