Distribution of $\sum_{j=1}^n \ln\left(\frac{X_{(j)}}{X_{(1)}}\right)$ when X_i's are i.i.d Pareto variables

Asked 4 years, 3 months ago Modified 2 years, 8 months ago Viewed 765 times

6

1

Let $X_1, X_2, \ldots, X_n$ be i.i.d variables having a Pareto distribution with density

$$f(x) = \frac{a\theta^a}{x^{a+1}} 1_{x>\theta}\,,$$

where $a, \theta > 0$. What is the distribution of $\sum\limits_{j=1}^n \ln\left(\frac{X_{(j)}}{X_{(1)}}\right)$?

Suppose $\operatorname{Gamma}(p, \alpha)$ denotes the density $g(t) \propto e^{-\alpha t} t^{p-1} 1_{t>0}$.

We have

$$T = \sum_{j=1}^n \ln\left(\frac{X_{(j)}}{X_{(1)}}\right) = \sum_{j=1}^n \ln(X_{(j)}) - n\ln(X_{(1)}) = \sum_{j=1}^n \ln X_j - n\ln X_{(1)}$$

Now,

$$\begin{aligned} \ln(X_j/\theta) &\stackrel{\text{i.i.d}}{\sim} \text{Exp with mean } 1/a \qquad , j = 1, \ldots, n \\ \implies \sum_{j=1}^n \ln(X_j/\theta) &= \sum_{j=1}^n \ln X_j - n\ln\theta \sim \operatorname{Gamma}(n, a) \end{aligned}$$

I could show that $X_{(1)}$ has another Pareto density, so that

$$\ln\left(\frac{X_{(1)}}{\theta}\right) = \ln X_{(1)} - \ln\theta \sim \text{Exp with mean } 1/(na)$$

Not sure if the last two facts help me get the exact distribution of T.

Edit:

Turns out this was rather simple had I simply rewritten T as

$$T = \sum_{j=1}^n \ln\left(\frac{X_j}{X_{(1)}}\right) = \sum_{j=1}^n (\ln X_j - \ln X_{(1)}) = \sum_{j=1}^n (Y_j - Y_{(1)})\,,$$

where $Y_j = \ln(X_j/\theta)$. Since $aY_j \sim \operatorname{Exp}(1)$, using this result I have $aT \sim \operatorname{Gamma}(n-1, 1)$.

This is equivalent to $T \sim \text{Gamma}(n-1, a)$ or $2aT \sim \chi^2_{2n-2}$.

distributions | self-study | order-statistics | pareto-distribution | exponential-distribution

Share Cite Improve this question Follow

edited Nov 24, 2019 at 19:20

asked May 12, 2018 at 20:05
StubbornAtom
9,437 1 23 70

1 The first approach ignores the correlation between $\sum_{j=1}^n \ln X_j$ and $\ln X_{(1)}$ – Xi'an May 12, 2018 at 20:51

Yes, I noticed that they are not independent. – StubbornAtom May 12, 2018 at 20:55

1 Answer

Sorted by: Highest score (default)

7

A simpler approach might be to use the fact that if $x \sim \text{Pareto}(\theta, a)$, then conditioning upon $x \geq b$ results in $x \sim \text{Pareto}(b, a)$. Consequently, $x|x_{(1)} \sim \text{Pareto}(x_{(1)}, a)$, except for the single observation corresponding to $x_{(1)}$. When we then take the ratio $x/x_{(1)}$, we are rescaling x by its minimum value, and the resulting variate has a $\text{Pareto}(1, a)$ distribution, independent of $x_{(1)}$.
Therefore, if we don't pay attention to the rank of the x_i in the sample, the ratios $x_i/x_{(1)} \sim \text{Pareto}(1, a)$ and are independent (except for the observation corresponding to $x_{(1)}$, which is equal to 1.)

This, combined with the fact that the log of a $\text{Pareto}(1, a)$ variate is distributed $\text{Exponential}(a)$, and the sum of $n-1$ i.i.d. variates $\sim \text{Exponential}(a)$ is $\sim \text{Gamma}(n-1, a)$, leads directly to the result that the sum

$$\sum_{j=1}^n \ln\left(\frac{X_{(j)}}{X_{(1)}}\right) \sim \text{Gamma}(n-1, a)$$

where the $n-1$ comes from the fact that exactly one of the ratios will have value 1, hence $\log(\cdot) = 0$, leaving $n-1$ nonzero terms in the sum.

Share Cite Improve this answer Follow

edited Nov 24, 2019 at 15:58
StubbornAtom
9,437 1 23 70

answered May 13, 2018 at 3:28
jbowman
33.7k 8 60 114

I think you missed saying $x_{(1)}$ is independent of $x_i/x_{(1)}$ in the last sentence of the first paragraph. – StubbornAtom May 13, 2018 at 6:05

2 That's implied by the point of the next to last sentence, admittedly not nearly as clearly stated as it might have been - the ratio and $x_{(1)}$ are independent. – jbowman May 13, 2018 at 13:57