Pareto Distribution:
Estimation of Parameters

M Pal

Estimation of Parameters in Pareto Distribution

- There is a basic problem of estimation of parameters in Pareto distribution because Pareto law holds only for upper tail of the distribution, and one does not know which values in the data set representing the entire population are to be retained for estimation purpose.
- In other words, one has to determine the cut-off point first and then start estimation process using the data with values larger than the cut-off point. But the determination of cut-off point, if possible, implies that this is nothing but the threshold parameter (c). Thus, it remains to estimate the shape parameter α only. However, it is not known how to determine the cut-off point objectively.
- To make the concept clearer, let us assume that we have observations $x_1, x_2, \ldots x_m$. Without loss of generality, we can assume that these are arranged in non-decreasing order, i.e., $x_1 \leq x_2 \leq \cdots \leq x_m$. Since Pareto law holds for upper tail of the distribution, our task is to find a point x_0 above which Pareto law holds.

Logarithmic Transformation

- Pareto law implies

$$N_x = Ax^{-\alpha},$$

$$Or, \qquad Ln(N_x) = Ln(A) - \alpha Ln(x).$$

- If Pareto law holds then $Ln(N_x)$ must be a linear function of $Ln(x)$. Objective way of finding a value of x_0 (above which $Ln(N_x)$ will be a linear function of Ln x) is very difficult if not impossible.
- One may take subjective or semi-objective judgement. One way of making a semi-objective judgement is to plot different ($Ln(x)$, $Ln(N_x)$) values join the consecutive points by smooth curve and see for which points the smooth curve is approximately a straight line. In actual practice, if Pareto law holds, the curve behaves haphazardly in the beginning and then stabilizes to a straight line till the end.

The choice of cut-off point

- The choice of cut-off point thus remains subjective.
- To see how difficult is to find an objective criterion of estimating the cut-off point, let us take an objective criteria like minimizing

$$F = \frac{1}{n_c}\sum_{x>c}\left(Ln(N_x) - a - bLn(x)\right)^2,$$

- where n_c is the number of x values such that x > c with respect to the parameters c, a and b. The procedure is to choose different values of c and select the optimum c value for which it gives minimum of the minimum values. This fails because it may so happen that the last few $\left(Ln(N_x), Ln(x)\right)$ values behave almost linearly forcing the cut-off point to be much above than what it should be.
- Since the estimate of x_0 value is subjective, we assume that the threshold parameter c is situated somewhere near the value of x_0. We then take the sample of x values which are above the value of x_0. Suppose there are n such values $\geq x_0$ out of m values. We may rename these as $x_1, x_2, \dots x_n$. Without loss of generality, we assume that $x_1 \leq x_2 \leq \cdots \leq x_n$.

Estimation Using Moments (The Case of Raw Data)

- **Estimation Using Moments (The Case of Raw Data):**
- We assume that $x_1, x_2, \ldots x_n$ are i.i.d. random samples from Pareto (c, α), and α > 2.
- If X ~ Pareto(α, c) then it can be proved that

$$\mathrm{E(X)} = \frac{\alpha c}{\alpha - 1}$$

$$and\ \mathrm{E}(X^2) = \frac{\alpha c^2}{\alpha - 2}.$$

$$\text{Thus,} \mathrm{Var(X)} = \mathrm{E}(X^2) - \left(E(X)\right)^2 = \frac{\alpha c^2}{\{(\alpha-1)^2\,(\alpha-2)\}},$$

- and Coefficient of Variation is

$$C(X) = \sqrt{\frac{Var(X)}{\left(E(X)\right)^2}}\ ,$$

Estimation … (Continued)

$$Or, C^2(X) = \frac{Var(X)}{\left(E(X)\right)^2} = \frac{1}{\alpha(\alpha-2)}.$$

- We estimate α and c by solving the following two equations

$$\bar{x} = \frac{1}{n}\sum x_i = \frac{\hat{\alpha}\hat{c}}{(\hat{\alpha}-1)},$$

$$\overline{x^2} = \frac{1}{n}\sum x_i^2 = \frac{\hat{\alpha}\hat{c}^2}{(\hat{\alpha}-2)}.$$

- If the sample counter part of Coefficient of Variation $C^2(X)$, say, is less than 1, we have

$$\hat{\alpha} = 1 + \sqrt{\frac{1+C^2(X)}{C^2(X)}}, \text{ and } \hat{c} = \frac{\bar{x}(\hat{\alpha}-1)}{\hat{\alpha}}.$$

- The method of moment fails if α < 2 and this may happen in most cases as Pareto law states that 1 < α < 2.

Modified Method of Moments

- **Modified Method of Moments (Johnson and Kotz)**
- Here we take the first order moment and the first order statistic as the set of estimators.

$$\bar{x} = \left(\frac{1}{n}\right)\sum x_i = \frac{\hat{\alpha}\hat{c}}{\hat{\alpha}-1} ,$$

$$x_{(1)} = \mathrm{Min}(x_1, x_2, \ldots x_n) = \frac{n\hat{\alpha}\hat{c}}{n\hat{\alpha}-1} .$$

- Thus,

$$\hat{\alpha} = \frac{\left(n\bar{x} - x_{(1)}\right)}{\left(n\left(\bar{x} - x_{(1)}\right)\right)},$$

$$\text{and } \hat{c} = \frac{(n\hat{\alpha} - 1)x_{(1)}}{(n\hat{\alpha})}.$$

Maximum Likelihood Estimate

- **Maximum Likelihood Estimate**
- The likelihood function is

$$\mathrm{L}(\mathrm{c}, \alpha) = \prod_{i=1}^{n} \frac{\alpha c^{\alpha}}{x_i^{\alpha+1}} = \alpha^{n} c^{n\alpha} \prod \frac{1}{x_i^{\alpha+1}} .$$

- Since α > 0, L increases as c increases. So, we take the maximum possible value of c as the Maximum Likelihood (ML) estimate, which is

$$\hat{c} = \mathrm{x}_{(1)} = Min(\mathrm{x}_1, \mathrm{x}_2, \ldots \mathrm{x}_\mathrm{n}).$$

- Again, $\mathrm{Ln}\ \mathrm{L}(\hat{c}, \alpha) = \mathrm{n}\ \mathrm{Ln}(\alpha) + \mathrm{n}\alpha \mathrm{Ln}(\hat{c}) - (\alpha + 1) \sum \mathrm{Ln}(x_i)$

$$= nLn(\alpha) - \alpha \sum Ln\left(\frac{x_i}{x_{(1)}}\right) - \sum Ln(x_i) .$$

Maximum Likelihood Estimate (Continued)

- Differentiating the loglikelihood function with respect to α and equating it to 0 we get,

$$\frac{\partial Ln(L)}{\partial \alpha} = \frac{n}{\alpha} - \sum Ln\left(\frac{x_i}{x_{(1)}}\right) = 0,$$

- we get,

$$\hat{\alpha} = \frac{n}{\sum Ln\left(\frac{x_i}{x_{(1)}}\right)} = \frac{n}{\sum Ln\left(\frac{x_i}{\hat{c}}\right)}.$$

Estimation Using Grouped Data

- **Estimation Using Grouped Data**
- Estimation using grouped data is more difficult since the cut-off point chosen by graphical method is usually one of the end points. This introduces a bias to the ultimate estimates in the parameters. However, let us assume that the observations are grouped as given in the following frequency distribution.

Class interval ($x_{i-1} – x_i$)	**Frequency** (f_i)	**Class mark** (y_i)
$x_0 – x_1$	f_1	y_1
$x_1 – x_2$	f_2	y_2
---	---	---
$x_{i-1} – x_i$	f_i	y_i
---	---	---
$x_{K-1} – x_K$	f_K	y_K
Total	N	---

Table 1: Frequency distribution of income of N persons in a given community

Estimation Using Grouped Data

- **Method of Moments:**
- Using the class marks y_1, y_2, ..., y_K , which are usually the mid values of the class intervals, we can find the first and second order raw moments, equate these to the corresponding population moments, and solve these to get the estimates of the parameters c and α.
- The problem with the moment estimators is that the estimates are not much efficient and hence may sometimes lead to absurd values.

Estimation Using Grouped Data

- **Modified Method of Moments:**
- One may modify, as in the case of raw data, the method of moments by taking $\bar{x}$ and x_P (the Pth quantile). The two equations are then

$$\bar{x} = \frac{\widehat{\alpha}\hat{c}}{\widehat{\alpha} - 1}$$

$$x_P = c(1 - P)^{-1/\alpha}.$$

- We solve the two equations to get $\hat{c}$ and $\widehat{\alpha}$. Usually, x_P is taken close to the minimum value. But as P approaches ‘0’ the estimate of $\widehat{\alpha}$ becomes more and more inefficient. We may take $P \approx 0.05$ if sample size is greater than 200.

Estimation Using Grouped Data

- **Maximum Likelihood Estimate**
- We assume that c is some value in the first interval or is close to x_0. We insert one more interval, which is $(x_k - \infty)$, in the frequency distribution which is given below. f_{k+1} is zero. π_i's in the third column of the frequency distribution are the corresponding probabilities of the intervals. It can be calculated by integrating the probability density function of the Pareto distribution in the given domain as shown in the first column of the frequency distribution if the values of the parameters are known. The mathematical expression of this integral can be found explicitly as the Pareto density allows the same. We assume that the probabilities of the intervals are $\pi_1, \pi_2, \ldots, \pi_K, \pi_{K+1}$. Naturally, $\sum_1^{K+1} \pi_i = 1$. π_i's are functions of c and α.
- $\pi_i = F(x_i) - F(x_{i-1}) = \left\{1 - \left(\frac{c}{x_i}\right)^\alpha\right\} - \left\{1 - \left(\frac{c}{x_{i-1}}\right)^\alpha\right\} = \left(\frac{c}{x_{i-1}}\right)^\alpha - \left(\frac{c}{x_i}\right)^\alpha$.

Estimation Using Grouped Data
(Continued)

Class interval ($x_{i-1} - x_i$)	Frequency (f_i)	Probability of the interval (π_i)
$x_0 - x_1$	f_1	π_1
$x_1 - x_2$	f_2	π_2
---	---	---
$x_{i-1} - x_i$	f_i	π_i
---	---	---
$x_{K-1} - x_K$	f_K	π_K
$x_K - \infty$	f_{K+1}	π_{K+1}
Total	N	1.00

Table 2: Frequency distribution and the interval probabilities of income of N persons in a given community

Estimation Using Grouped Data
(Continued)

- We may assume that the random samples are coming from Pareto(c, α). The joint distribution of the no. of observations falling in the intervals is then the multinomial distribution since the random samples are iid. The likelihood function is

$$L(c,\alpha) = \frac{N!}{f_1!\,f_2!\,\dots f_K!\,f_{K+1}!}(\pi_1)^{f_1}(\pi_2)^{f_2}\dots(\pi_K)^{f_K}(\pi_{K+1})^{f_{K+1}}$$

$$= \frac{N!}{f_1!\,f_2!\,\dots f_K!}(\pi_1)^{f_1}(\pi_2)^{f_2}\dots(\pi_K)^{f_K}, \text{since } f_{K+1} \text{ is zero.}$$

- Thus, it reduces to the likelihood function of the original frequency distribution except that $\sum_1^K \pi_i \neq 1$. We maximize the likelihood function to get the estimates of the parameters. The solution is found by the numerical method using computer programming or using a statistical computer package.

- The problem with the ML estimates is that the estimates may not have the desirable properties since one of the parameters (threshold parameter) depends on the support of the distribution.

Thank You