

A new approach to the Pareto interpolation

Deko Dekov

Abstract. Pareto interpolation is a method of estimating the median and other properties of a population that follows a Pareto distribution. It is used in economics when analysing the distribution of incomes in a population. In this paper we offer a new simple numerical method for approximation of a sample by the Pareto distribution. From the Pareto probability density function we easily find the mean, median and variance of the Pareto distribution.

Keywords: Pareto distribution, approximation of a sample, least squares method

Pareto interpolation is a method of estimating the median and other properties of a population that follows a Pareto distribution. It is used in economics when analysing the distribution of incomes in a population. There is a symbolic method for finding the median of the Pareto approximation from a sample.

In this paper we offer a new simple numerical method for approximation of a sample by the Pareto distribution. The numerical method is described in the paper [1]. Here we show that the method, described in [1], could be applied to the problem of finding the Pareto approximation. Note that from the Pareto probability density function we can easily find the mean, median, and variance of the Pareto distribution.

The numerical method, described in this paper, has the following advantages. The method uses only the definition of a function, so that the school and college students could understand and use it without studying. The method is fast, because it needs small numbers of iterations. Since each iteration adds one true digit to the answer, we need only 100 iterations to receive an answer with 100 true digits. We receive the answer for less than 1 second, if we use a desktop personal computer. By using the described numerical method, we could approximate a sample by any continuous probability distribution, provided the corresponding objective function has one or two variables. But, if we want to receive the answer immediately, we have to use a computer program.

The method, in more general framework, is as follows. Suppose data consisting of n points $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ are known and the goal is to find a function $y = F(x)$ that fits the data reasonably well. We will use the least squares criterion. We suppose that the reader is familiar with the least squares criterion. Suppose that $f(x)$ is the objective

function, so that we have to find the minimum of $f(x)$. We use the data set, in order to localize the minimum of $f(x)$. Suppose that the minimum of the function $f(x)$ is within the segment $[a,b]$. We divide the segment $[a,b]$ by N equal parts by using the points $x_0 = a, x_1, x_2, \dots, x_N = b$. Then we evaluate $f(x_0), f(x_1), f(x_2), \dots, f(x_N)$, and select the minimal of these values. We use the minimal value as the midpoint of a new segment, whose length is 10 times smaller than the previous segment. The process is repeated until the minimum is found. The method works well if $N \geq 10$.

The method is simple, so that it allows a simple implementation. I have created a simple computer program by using PHP. The program could record the calculations, made by the computer. The file containing calculations for the example given below, is available for download as supplementary material.

Recall that the probability density function of the Pareto distribution is given by

$$P(x; \alpha, k) = \frac{\alpha k^\alpha}{x^{\alpha+1}}$$

where the variable $x > k$ and the parameter $\alpha > 0$ are real numbers. For the formulas of the mean, median and the variance of the Pareto distribution.

Example. A sample is given by the following table:

2	3	4	5
1.25	0.3	0.1	0.05

Find the approximation of the sample by the Pareto distribution, and then find its mean, median and variance.

Solution. Set $k = 2$. By using the computer program, which implements the above described method, we obtain the parameter α . The computer program finds α as the solution to the following problem. Find α which minimizes the function

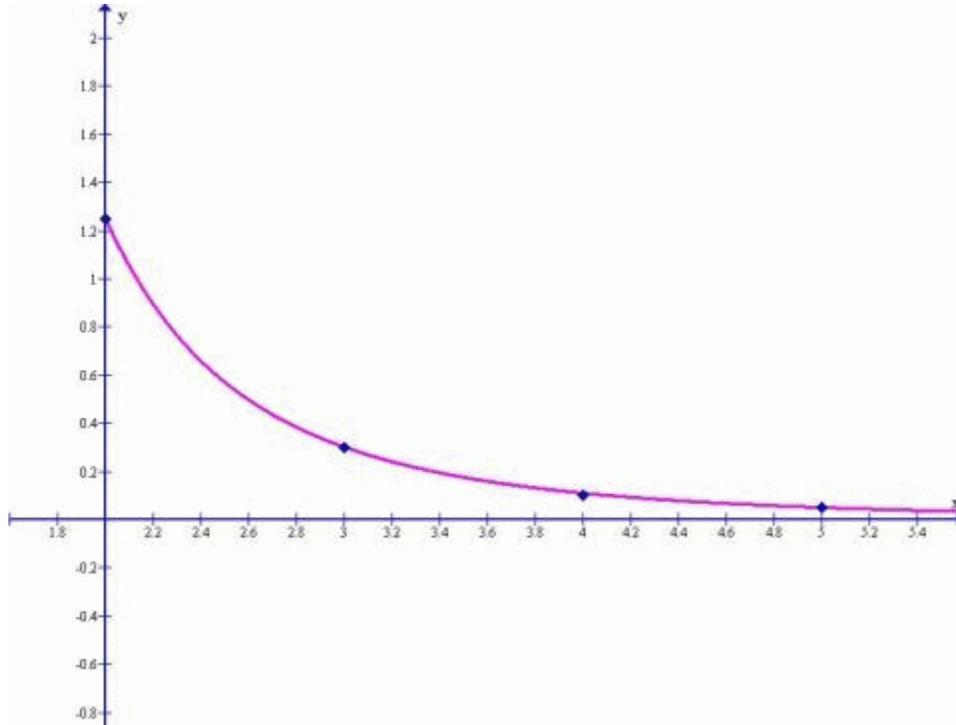
$$f(\alpha) = \left(\frac{\alpha 2^\alpha}{2^{\alpha+1}} - 1.25\right)^2 + \left(\frac{\alpha 2^\alpha}{3^{\alpha+1}} - 0.3\right)^2 + \left(\frac{\alpha 2^\alpha}{4^{\alpha+1}} - 0.1\right)^2 + \left(\frac{\alpha 2^\alpha}{5^{\alpha+1}} - 0.05\right)^2$$

Set $N = 10$. As initial interval for α we take the interval $[2,3]$. If we require 5 digits after the decimal point, we receive the following answer: $\alpha = 2.50143$. Hence, the probability density function of the Pareto approximation is as follows:

$$P(x) = \frac{2.50143(2^{2.50143})}{x^{2.50143+1}}$$

From the probability density function of the Pareto distribution we obtain the mean = 3.33206, median = 2.6386 and variance = 8.85172.

We could use the Ivan Johansen's computer program Graph to draw the graph of the sample - the blue rhombs and the graph of the Pareto approximation to the sample:



References

1. Deko Dekov, A numerical method for solving the horizontal resection problem in Surveying, Journal of Geodetic Science (to appear).

Dr.Deko Dekov
Zahari Knjazeski 81
6000 Stara Zagora

Bulgaria

Submitted on 1 October 2011

Publication date: 1 February 2012