Journal of Modern Mathematics and Statistics 5 (1): 17-18, 2011
ISSN: 1994-5388

Modified Pareto Distribution

Khurshid Ahmad Mir
Department of Statistics, Government Degree College Bemina, Srinagar, Kashmir, India

Abstract: A type of Pareto distribution with its range spread over the non negative part of the real line is considered as a possible model for a given live data wherein the nature of the data imposes some non zero probability at the origin of the random variable. Accordingly the original Pareto model is to be modified to get a new model called modified Pareto distribution. The proposed model fits well for rainfall data leading to the discussion of point and interval estimation of its parameters useful in predictions.

Key words: Pareto distribution, point estimation, interval estimation, probability, random veriable, imposes

INTRODUCTION

There are many situations in practice like life testing experiments where an item fails instantaneously leading to the observed life time of the item to be zero. In the case of rainfall data, there may be days reported with zero rainfall in a given year resulting in non zero probability of zero rainfall on any day in a year. In a similar manner as a model for wages and incomes Pareto distribution can be considered as a continuous model with a peculiar specification that probability of zero daily wages is quite commonly non zero.

In all this description, we will have a random variable X such that P (X = 0) ≠ 0 and P (X>0) is given by a continuous model. The distribution of X given in the above specification is named as modified version of the original distribution. Aitchison (1955), Kleyle and Dahiya (1975) and Muralidharan and Kale (2002) are some researches dealing with such types of mixture models called modified versions.

MODIFIED PARETO DISTRIBUTION

Consider a model given by Pareto distribution as:

$$f(x;\sigma,\alpha)=\frac{\alpha}{\sigma}\left[1+(\frac{x}{\sigma})\right]^{-(\alpha+1)},\ X>0 \quad (1)$$

$$F(X)=1-\left[1+(\frac{x}{\sigma})\right]^{\alpha} \quad (2)$$

The modified version of the above Pareto distribution is given by:

$$G(x;\alpha;\sigma;P)=1-P;\ x=0=1-P+PF(X);\ x>0$$

where, F(x) is given by Eq. 2. The above distribution can be seen to be discrete at x = 0 and continuous for all X>0. The above model was extensively used by Rust *et al*. (2004) and Reinartz and Kumar (2003). If $X_1, X_2, ... X_n$ is an ordered sample of size n from the Pareto distribution, we can have the likelihood equation:

$$L=\prod_{i=1}^{n}(1-P)^{Z(x_i)}\left[P\frac{\alpha}{\sigma}\left(1+\frac{x_i}{\sigma}\right)^{-(\alpha+1)}\right]^{[1-Z(x_i)]} \quad (3)$$

where, $Z(x_i)$ is an indicator function defined as:

$$Z(x_i)=1 \text{ if } X_i=0,\quad =0 \text{ if } X_i>0$$

For M. L. Es of P, α and σ we should have:

$$\frac{\partial \mathrm{Log}\,L}{\partial P}=0\Rightarrow \hat{P}=\sum_{i=1}^{n}\frac{(1-Z(x_i))}{n} \quad (4)$$

$$\frac{\partial \mathrm{Log}\,L}{\partial \alpha}=0$$

$$\Rightarrow \hat{\alpha}=\sum_{i=1}^{n}\frac{(1-Z(x_i))}{(1-Z(x_i))\mathrm{Log}(1+\frac{x_i}{\sigma})} \quad (5)$$

$$\frac{\partial \mathrm{Log}\,L}{\partial \alpha}=0$$

$$\Rightarrow(\alpha+1)\sum_{i=1}^{n}[1-Z(x_i)]\frac{z_i}{1+z_i} \quad (6)$$

If:

$$\sum_{i=1}^{n}Z(x_i)$$

is denoted by k, the above log likelihood equations and their solutions simplify to:

$$\hat{P} = \frac{n-k}{n} \quad (7)$$

$$\hat{\alpha} = \frac{n-k}{\sum_{x_i>0} \log\left(1+\frac{x_i}{\sigma}\right)} \quad (8)$$

$$\frac{1}{\sigma}.\sum_{x_i>0}^{n} \frac{z_i}{1+z_i} = 0 \quad (9)$$

MLEs of α and σ are the simultaneous solutions of Eq. 8 and 9. Equation 9 has to be solved iteratively for σ which in turn when substituted in Eq. 8 would give the value of α. Given a sample of size n the exact MLEs of P, α, σ can thus be determined of which MLE of P is analytical, MLE of σ is iterative and MLE of α is analytical depending on σ and hence, indirectly iterative. We now attempt to present the estimates of P, α, σ for a live data on recorded rain fall in a region Jalgaon division, India (Muralidharan and Kale, 2002) which is shown in Table 1. The fittness of modified Pareto model for this data is asserted by the strength of linear relationship measured through the Karl Pearson coeficient of correlation between ordered non zero rainfall data and the corresponding population quantile of a standard Pareto distribution.

The data was available daily for the months of June, July, August and September for the years 1961-1970. The calculated correlation coeficients are presented in the Table 1 which shows that Pareto model is reasonably a good fit for the data. Fitness of Pareto model to the rainfall data. Using the rainfall data, we obtained the MLEs of P, α, σ for each year by solving the Eq. 7-9. It can be seen that the expected value of modified Pareto model is given by:

$$E(X) = \frac{\sigma P}{\alpha - 1}$$

MLE of population mean is given by:

$$\frac{\hat{\alpha}\hat{P}}{\hat{\alpha}-1}$$

Using the well known Cramer's rule, estimated asymptotic variance of MLE of population mean is given by:

Table 1: Fitness of Pareto model to the rainfall data

Years	a = 2	a = 3	a = 4
1961	0.9848	0.9933	0.9941
1962	0.9310	0.9503	0.9580
1963	0.9244	0.9510	0.9619
1964	0.9627	0.9788	0.9835
1965	0.9488	0.9502	0.9497
1966	0.9818	0.9719	0.9653
1967	0.9637	0.9824	0.9889
1968	0.9272	0.9110	0.9007
1969	0.9830	0.9867	0.9852
1970	0.8793	0.9087	0.9216

$$\text{ASVAR}\left(\frac{\hat{\alpha}\hat{P}}{\hat{\alpha}-1}\right) = \left(\frac{\partial h}{\partial \alpha}\right)^2 . V(\hat{\alpha}) + \left(\frac{\partial h}{\partial P}\right)^2 . V(\hat{P}) + \left(\frac{\partial h}{\partial \sigma}\right)^2 . V(\hat{\sigma}) + \left(\frac{\partial h}{\partial \alpha}.\frac{\partial h}{\partial \sigma}\right). 2cov(\hat{\sigma}, \hat{\alpha}) + \left(\frac{\partial h}{\partial \alpha}.\frac{\partial h}{\partial P}\right).2cov(\hat{\sigma}, \hat{\alpha}) + \left(\frac{\partial h}{\partial \sigma}.\frac{\partial h}{\partial P}\right).2cov(\hat{\sigma}, \hat{P})$$

where:

$$h(\alpha, \sigma, P) = \frac{\sigma P}{\alpha - 1}. V(\hat{\alpha}), V(\hat{\sigma}) \ldots\ldots cov(\hat{\sigma}, \hat{P})$$

are elements of estimated asymptotic variance covariance matrix of $(\hat{\sigma}, \hat{\alpha}, \hat{P})$, we can get confidence intervals of population mean at any desired confidence level using the asymptotic normality of MLEs.

REFERENCES

Aitchison, J., 1955. On the distribution of a positive random variable having a discrete probability mass at the origin. J. Am. Stat. Assoc., 50: 901-908.

Kleyle, R.M. and R.C. Dahiya, 1975. Estimation of parameters of mixed failure time distribution from censored data. Commun. Stat. Theory Methods, 4: 873-882.

Muralidharan, K. and B.K. Kale, 2002. Modified gamma distribution with singularity at zero. Commun. Stat. Simulation Comput., 31: 143-158.

Reinartz, W. and V. Kumar, 2003. The impact of customer relationship characteristics on profitable lifetime duration. J. Market., 67: 77-99.

Rust], R.T., K.N. Lemon and V.A. Zeithaml, 2004. Return on marketing: Using customer equity to focus marketing strategy. J. Marketing, 68: 109-127.