Statistical Methodology 26 (2015) 135–150

Contents lists available at ScienceDirect

Statistical Methodology

journal homepage: www.elsevier.com/locate/stamet

A discrete truncated Pareto distribution

Tomasz J. Kozubowski [a,*], Anna K. Panorska [a], Matthew L. Forister [b]

[a] Department of Mathematics and Statistics, University of Nevada, Reno, NV 89557, USA

[b] Department of Biology, University of Nevada, Reno, NV 89557, USA

ARTICLE INFO

Article history:
Received 15 October 2014
Received in revised form 19 March 2015
Accepted 7 April 2015
Available online 5 May 2015

MSC:
primary 60E05
secondary 62F10
62H05
62P05

Keywords:
Benford's law
Beta distribution
Discrete Pareto
Ecology
First digit law
Heavy tail
Maximum likelihood estimation
Power law
Truncated geometric
Truncated Pareto

ABSTRACT

We propose a new discrete distribution with finite support, which generalizes truncated Pareto and beta distributions as well as uniform and Benford's laws. Although our focus is on basic properties and stochastic representations, we also consider parameter estimation and include an illustration from ecology showing potential applications of this new stochastic model.

* Correspondence to: Department of Mathematics and Statistics, MS 084, University of Nevada Reno, RENO, NV 89557-0084, USA. Tel.: +1 775 784 6548.

E-mail address: tkozubow@unr.edu (T.J. Kozubowski).

http://dx.doi.org/10.1016/j.stamet.2015.04.002
1572-3127/

1. Introduction

In this paper we introduce a discrete truncated Pareto (DTP) distribution, derive its properties and develop the estimation of parameters. The motivation for this work comes from a biological problem involving the distribution of diet breadth in Lepidoptera (butterflies and moths), where the empirical survival function suggested the need for an upper truncated model with power law tail.

Pareto distributions are popular models for data with power-law tails. The definitions and properties of continuous Pareto distributions are given in Arnold [2], Kotz and Johnson [24], Balakrishnan and Nevzorov [3], or Krishnamoorthy [26]. Numerous applications of power law distributions in stochastic modeling are offered in Clauset and Newman [12], Greiner et al. [18], Muskulus et al. [29], Newman et al. [30], Park et al. [32], Reed [33], Reed and Hughes [34], Sandland [35], Stegen et al. [37], Sornette [36], Weidner et al. [39] and the references within these papers. The need for truncation of the continuous Pareto distribution was extensively discussed in Aban [1], which also included applications of the truncated Pareto model in climate, geology and finance research. Truncated power law models were discussed in Burroughs and Tebbens [10,11] in the context of modeling earthquake magnitudes, fault lengths, oil or gas field sizes, and areas burnt by forest fires. Estimation for the truncated Pareto distribution was presented in Burroughs and Tebbens [9] using the least squares fitting of the probability plot, and in Aban et al. [1] using a maximum likelihood approach.

A Pareto random variable W is given by the survival function (SF)

$$S_W(x) = P(W > x) = (\gamma/x)^\alpha, \quad \text{for } 0 < \gamma \le x \text{ and } \alpha > 0, \tag{1}$$

while its truncated version, $Y = W|W \le \nu + 1$, has the survival function (SF) of the form

$$S_Y(x) = P(Y > x) = \frac{(\gamma/x)^\alpha - (\gamma/(\nu+1))^\alpha}{1 - (\gamma/(\nu+1))^\alpha}, \tag{2}$$

where $0 < \gamma \le x \le \nu + 1 < \infty$ (with integers γ and ν) and $\alpha > 0$. A *discrete truncated Pareto* variable X can be derived by discretizing the continuous truncated Y with SF (2) using the relation

$$X = \lfloor Y \rfloor, \tag{3}$$

where $\lfloor y \rfloor$ is the largest integer that is less than or equal to x (the floor function). Alternatively, in analogy with the continuous truncated Pareto, its discrete counterpart arises by the restriction $X \stackrel{d}{=} Y|\gamma \le Y \le \nu$, where Y has a *discrete Pareto* (DP) distribution given by the probability mass function (PMF)

$$f_Y(k) = P(Y = k) = \left(\frac{1}{k}\right)^\alpha - \left(\frac{1}{1+k}\right)^\alpha, \quad k = 1, 2, \ldots. \tag{4}$$

The discrete Pareto distribution was developed in Buddana [7] and Buddana and Kozubowski [8] (cf. Krishna and Pundir, [25]) as an attractive, heavy-tail alternative to Zip'f law (Zipf, [40]), since the PDF of DP distributions does not involve any special functions and estimation of their parameters is relatively straightforward. Since we discuss both DT and DTP distributions often in this paper, we will refer to DTP as "discrete truncated Pareto" and to DP as "discrete non-truncated Pareto" distribution, for clarity of the exposition.

A routine derivation shows that the random variable X in (3) has a discrete distribution with the PMF

$$f(x) = P(X = x) = \frac{\frac{1}{x^\alpha} - \frac{1}{(x+1)^\alpha}}{\frac{1}{\gamma^\alpha} - \frac{1}{(\nu+1)^\alpha}} \quad \text{for } x = \gamma, \gamma+1, \ldots, \nu. \tag{5}$$

This discretization of the truncated Pareto distribution (DTP) is the main focus of this paper.

An interesting property of this new distribution is that the "tail" parameter α is not restricted to the positive values. Indeed, for negative values of α, the PMF (5) takes the form

$$f(x) = P(X = x) = \frac{(x+1)^\beta - x^\beta}{(\nu+1)^\beta - \gamma^\beta} \quad \text{for } x = \gamma, \gamma+1, \ldots, \nu, \tag{6}$$