A discrete truncated Pareto distribution.

Tomasz J. Kozubowski and Anna K. Panorska¹

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

Matthew L. Forister

Department of Biology, University of Nevada, Reno, NV 89557, USA

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.

AMS 2000 classification: 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

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 [34], Reed and Hughes [35], Sandland [36], Stegen et al. [38], Sornette [37], Wedner et al. [40] 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,

¹Corresponding Author: Tomasz J. Kozubowski, Department of Mathematics and Statistics, MS 084, University of Nevada Reno, RENO, NV 89557-0084, USA; email: tkozubow@unr.edu, Phone: +1-775 784 6548.

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)

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

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

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

where $0 < \gamma \leq x \leq \nu + 1 < \infty$ (with integer 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

$$(3) \quad X = \lfloor Y \rfloor,$$

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 $X \stackrel{d}{=} Y|\gamma \leq Y \leq \nu$, where Y has a *discrete Pareto* (DP) distribution given by the probability mass function (PMF)

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

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 Zipf's law (Zipf, [41]), 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

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

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

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

where $\beta = -\alpha > 0$. It is not hard to see that this model arises by discretizing a doubly-truncated power function random variable Y via (3) as follows. First, recall that the power function random variable W , which belongs to the family of beta distributions (see, e.g., Chapter 25 of Johnson et al. [21]), has the cumulative distribution function (CDF) of the form

$$(7) \quad F_W(x) = P(W \leq x) = (x/b)^\beta \text{ for } 0 < x < b.$$

Next, if W is transformed into Y by truncating its distribution below at γ and above at $\nu+1$ (so that $Y \stackrel{d}{=} W | \gamma \leq W \leq \nu+1$), where γ and ν are positive integers satisfying $0 < \gamma < \nu+1 \leq b$, and subsequently Y is discretized via (3), we obtain a distribution with the PMF (6). When $\alpha = -1$, this model reduces to a discrete uniform distribution on the set $\gamma, \gamma+1, \dots, \nu$.

Note that when $\alpha = \beta = 0$, PMFs (5) and (6) are understood in the limiting sense, where we have

$$(8) \quad f(x) = P(X = x) = \frac{\log\left(1 + \frac{1}{x}\right)}{\log\left(\frac{\nu+1}{\gamma}\right)} \text{ for } x = \gamma, \gamma+1, \dots, \nu.$$

In particular, if $\gamma = 1$ and $\nu = 9$, we obtain Benford’s 1st digit law (see, e.g., Benford [5], Hill [19, 20]), given by the PMF

$$(9) \quad f(x) = P(X = x) = \log_{10} \left(1 + \frac{1}{x}\right), \text{ for } x = 1, 2, \dots, 9.$$

Thus, the discrete truncated Pareto law provides a generalization of this important model as well.

This paper is organized as follows. Section 2 contains a definition and properties of the discrete Pareto law. In Section 3 we consider parameter estimation. Section 4 is devoted to an illustration of modeling using DTP. We collect technical lemmas and proofs of the main results in the last section.

2. DEFINITION AND BASIC PROPERTIES

We start with a formal definition and notation. Throughout this paper, the quantities γ and ν are positive integers satisfying the relation $0 < \gamma \leq \nu < \infty$, while $\alpha \in (-\infty, \infty)$.

Definition 2.1. *A random variable X is said to have a discrete truncated Pareto (DTP) distribution with parameters α , γ , and ν , denoted by $DTP_\alpha(\gamma, \nu)$, if its PMF is given by (5). We express this by writing $X \sim DTP_\alpha(\gamma, \nu)$.*

Remark 2.1. As noted in the introduction, this distribution arises by discretizing a truncated Pareto distribution ($\alpha > 0$) or truncated power function distribution ($\alpha < 0$). For $\alpha = 0$, the PMF (5) is understood in the limiting sense. Since the limits of the numerator and the denominator of (5) are both zero as $\alpha \rightarrow 0$, we invoke L'Hospital's rule to obtain

$$\lim_{\alpha \rightarrow 0} \frac{\frac{1}{x^\alpha} - \frac{1}{(x+1)^\alpha}}{\frac{1}{\gamma^\alpha} - \frac{1}{(\nu+1)^\alpha}} = \lim_{\alpha \rightarrow 0} \frac{\frac{1}{x^\alpha} \log \frac{1}{x} - \frac{1}{(x+1)^\alpha} \log \frac{1}{x+1}}{\frac{1}{\gamma^\alpha} \log \frac{1}{\gamma} - \frac{1}{(\nu+1)^\alpha} \log \frac{1}{\nu+1}} = \frac{\log(1 + \frac{1}{x})}{\log(\frac{\nu+1}{\gamma})}.$$

Thus, the PMF of the $DTP_0(\gamma, \nu)$ distribution is indeed given by (8), as stated in the introduction. Moreover, with this convention, the PMF (5) is continuous in α over the entire real line.

Remark 2.2. In the case that $\alpha > 0$, the DTP distribution can also be obtained by truncating a discrete non-truncated Pareto (DP) distribution below at γ and above at ν . The discrete non-truncated Pareto random variable Y , which appeared in Krishna and Pundir [25] and was studied by Buddana [7] and Buddana and Kozubowski [8], arises by discretizing a Pareto Type II (or Lomax) distribution given by the survival function

$$(10) \quad S(x) = \left(\frac{1}{1 + \frac{x}{\sigma}} \right)^\alpha,$$

and has the PMF of the form

$$(11) \quad f_Y(k) = P(Y = k) = \left(\frac{1}{1 + \frac{k-1}{\sigma}} \right)^\alpha - \left(\frac{1}{1 + \frac{k}{\sigma}} \right)^\alpha, \quad k = 1, 2, \dots$$

This distribution is an attractive alternative to the Zipf's law (see, e.g., Zipf [41], which is perhaps the most widely used discrete power-law distribution in applications (see, e.g., Consul [14] or Newman [30] and the references therein). However, the PMF of the latter,

$$(12) \quad P(X = k) = \frac{1}{\xi(\alpha)} \frac{1}{k^\alpha}, \quad k = 1, 2, \dots,$$

involves the Riemann-Zeta special function

$$(13) \quad \xi(\alpha) = \sum_{k=1}^{\infty} \frac{1}{k^\alpha}, \quad \alpha > 0,$$

which is inconvenient to work with in an applied setting. It is not hard to see that the truncated variable, $X \stackrel{d}{=} Y | \gamma \leq Y \leq \nu$, has the PMF

$$(14) \quad f_X(x) = P(X = x) = \frac{\frac{1}{(\sigma+x-1)^\alpha} - \frac{1}{(\sigma+x)^\alpha}}{\frac{1}{(\sigma+\gamma-1)^\alpha} - \frac{1}{(\sigma+\nu)^\alpha}} \quad \text{for } x = \gamma, \gamma+1, \dots, \nu,$$

which coincides with the DTP mass function (5) when $\sigma = 1$. For other (positive) values of σ the PMF (14) is a more general distribution, particularly when the parameter $\sigma > 0$ is not an integer (for integer values of σ , this is just a shifted DTP distribution).

Straightforward calculations lead to the CDF and the SF of the DTP distribution, given below.

Proposition 2.1. *The CDF and the SF of a $DTP_\alpha(\gamma, \nu)$ random variable X with $\alpha \neq 0$ are given by*

$$(15) \quad F(x) = P(X \leq x) = \frac{\frac{1}{\gamma^\alpha} - \frac{1}{(\lfloor x \rfloor + 1)^\alpha}}{\frac{1}{\gamma^\alpha} - \frac{1}{(\nu + 1)^\alpha}} \quad \text{for } \gamma \leq x < \nu$$

and

$$(16) \quad S(x) = P(X > x) = \frac{\frac{1}{(\lfloor x \rfloor + 1)^\alpha} - \frac{1}{(\nu + 1)^\alpha}}{\frac{1}{\gamma^\alpha} - \frac{1}{(\nu + 1)^\alpha}} \quad \text{for } \gamma \leq x < \nu,$$

respectively. For $x < \gamma$ the CDF (respectively, the SF) is 0 (respectively, 1) and for $x \geq \nu$ the CDF (respectively, the SF) is 1 (respectively, 0).

When the value of α is equal to zero, the expressions (15)-(16) are to be understood in the limiting sense, as stated below.

Proposition 2.2. *The CDF and the SF of a $DTP_0(\gamma, \nu)$ random variable X ($\alpha = 0$) are given by*

$$(17) \quad F(x) = P(X \leq x) = \frac{\log(\lfloor x \rfloor + 1) - \log(\gamma)}{\log(\nu + 1) - \log(\gamma)}, \quad \text{for } \gamma \leq x < \nu$$

and

$$(18) \quad S(x) = P(X > x) = \frac{\log(\nu + 1) - \log(\lfloor x \rfloor + 1)}{\log(\nu + 1) - \log(\gamma)}, \quad \text{for } \gamma \leq x < \nu,$$

respectively. For $x < \gamma$ the CDF (respectively, the SF) is 0 (respectively, 1) and for $x \geq \nu$ the CDF (respectively, the SF) is 1 (respectively, 0).

2.1. Properties of the DTP mass function. Here we study the behavior of the DTP mass function (5) with respect to x as well as the three parameters. We note that if $\gamma = \nu$, the PMF has only one non-zero value (of one) for $x = \gamma$, so we shall omit this trivial case and assume $\gamma < \nu$ for the remainder of this section. We start with the behavior of the DTP mass function with respect to x .

Proposition 2.3. *The PMF of a $DTP_\alpha(\gamma, \nu)$ random variable X is monotonically increasing in x when $\alpha < -1$, is constant in x when $\alpha = -1$, and is monotonically decreasing in x when $\alpha > -1$.*

Next, we consider the behavior of the DTP PMF with respect to γ and ν . The following result is rather straightforward to prove, and so we shall omit its derivation.

Proposition 2.4. *For any fixed $x > 0$ and $\alpha \in (-\infty, \infty)$, the PMF of a $DTP_\alpha(\gamma, \nu)$ random variable X is monotonically increasing in γ on the interval $(0, x)$ and is monotonically decreasing in ν on the interval (x, ∞) .*

We now proceed with investigating the behavior of the DTP PMF with respect to the parameter α . It turns out that the behavior of this function depends on relationships among γ , ν , and x . We start with the case where x is at the boundary of its range.

Proposition 2.5. *If $x = \gamma$, then the PMF of a $DTP_\alpha(\gamma, \nu)$ random variable X is monotonically increasing in α on the entire real line, with the limits of zero and one at minus infinity and infinity, respectively. On the other hand, if $x = \nu$, then the PMF of a $DTP_\alpha(\gamma, \nu)$ random variable X is monotonically decreasing in α on the interval $(-\infty, \infty)$, with the limits of one and zero at minus infinity and infinity, respectively.*

The behavior of the DTP PMF with respect to α is quite different when X is not on the boundary, as seen in the next several results. The first one concerns differentiability of the PMF with respect to α .

Proposition 2.6. *The PMF $f(x)$ of a $DTP_\alpha(\gamma, \nu)$ random variable X is a continuous and differentiable function of α on the interval $(-\infty, \infty)$. Moreover, the derivative is a continuous function of α on the interval $(-\infty, \infty)$, and its value at zero is given by*

$$(19) \quad \left. \frac{\partial f(x)}{\partial \alpha} \right|_{\alpha=0} = \frac{1}{2} \log \frac{\gamma(\nu+1)}{x(x+1)} \cdot \frac{\log(x+1) - \log(x)}{\log(\nu+1) - \log(\gamma)}.$$

Note that the derivative of the DTP PMF at $\alpha = 0$ is positive when $x = \gamma$ and is negative when $x = \nu$, which is consistent with the results of Proposition 2.5. The next result plays an important role in statistical inference connected with the DTP distribution.

Proposition 2.7. *The PMF $f(x)$ of a $DTP_\alpha(\gamma, \nu)$ random variable X is a log-concave function of α on the interval $(-\infty, \infty)$. Moreover, the limits of $\log f(x)$ at $\pm\infty$ are as follows:*

$$(20) \quad \lim_{\alpha \rightarrow \infty} \log f(x) = \begin{cases} 0 & \text{for } x = \gamma \\ -\infty & \text{for } \gamma < x \leq \nu, \end{cases}$$

$$(21) \quad \lim_{\alpha \rightarrow -\infty} \log f(x) = \begin{cases} 0 & \text{for } x = \nu \\ -\infty & \text{for } \gamma \leq x < \nu. \end{cases}$$

By the above result, when $\gamma < x < \nu$, the DTP PMF $f(x)$ is log-concave in α on $(-\infty, \infty)$, with the limits of $-\infty$ at $\pm\infty$. Consequently, there must be an $\alpha^* \in (-\infty, \infty)$, such that the function $\log f(x)$ is monotonically increasing on the interval $(-\infty, \alpha^*)$ and is monotonically decreasing on the interval (α^*, ∞) . In turn, the same must be true about the PMF itself. We summarize this in the following result.

Proposition 2.8. *If $\gamma < x < \nu$, then there exists an $\alpha^* \in (-\infty, \infty)$ such that the PMF of a $DTP_\alpha(\gamma, \nu)$ random variable X is monotonically increasing in α on the interval $(-\infty, \alpha^*)$ and is monotonically decreasing in α on the interval (α^*, ∞) . Moreover, the limits of the PMF at $\pm\infty$ are both zero.*

Remark 2.3. It follows from these results that, as $\alpha \rightarrow \infty$, the $DTP_\alpha(\gamma, \nu)$ distribution converges weakly to a point mass at γ . Similarly, as $\alpha \rightarrow -\infty$, the $DTP_\alpha(\gamma, \nu)$ distribution converges weakly to a point mass at ν .

2.2. Moments. Since the DTP random variable X is discrete with finite (and positive) support, its moments of all orders $\eta \in (-\infty, \infty)$ exist, and are given by

$$(22) \quad E(X^\eta) = \sum_{k=\gamma}^{\nu} \frac{\frac{k^\eta}{k^\alpha} - \frac{k^\eta}{(k+1)^\alpha}}{\frac{1}{\gamma^\alpha} - \frac{1}{(\nu+1)^\alpha}}, \quad \eta \in (-\infty, \infty).$$

When $\alpha = 0$, the above expression is understood as the limit when $\alpha \rightarrow 0$, given by

$$(23) \quad E(X^\eta) = \sum_{k=\gamma}^{\nu} \frac{k^\eta \log \frac{k+1}{k}}{\log \frac{\nu+1}{\gamma}}, \quad \eta \in (-\infty, \infty), \alpha = 0.$$

Remark 2.4. Since the DTP PMF is a continuous and differentiable function of α on $(-\infty, \infty)$, the same is true for all the moments.

Alternative expressions for the moments can be derived for specific values of η . In particular, for $\eta = 1$, we have the following result, which can be derived by routine algebra.

Proposition 2.9. *The mean of a $DTP_\alpha(\gamma, \nu)$ random variable X is a continuous function of α on the entire real line, given by*

$$(24) \quad E(X) = \frac{\gamma(\nu+1)^\alpha - \nu\gamma^\alpha}{(\nu+1)^\alpha - \gamma^\alpha} + \frac{(\nu+1)^\alpha \gamma^\alpha}{(\nu+1)^\alpha - \gamma^\alpha} \sum_{k=\gamma+1}^{\nu} \frac{1}{k^\alpha}$$

for $\alpha \neq 0$ and

$$(25) \quad E(X) = \frac{\nu \log(\nu+1) - \gamma \log \gamma - \sum_{k=\gamma+1}^{\nu} \log k}{\log(\nu+1) - \log \gamma}$$

for $\alpha = 0$.

We note that when $\gamma = \nu$, the sums in the above expressions are understood as zeros, in which case the mean coincides with γ . In the special case $\alpha = -1$, the DTP distribution is uniform on the set $\gamma, \gamma+1, \dots, \nu$, and its mean reduces to

$$(26) \quad E(X) = \frac{\sum_{k=\gamma+1}^{\nu} k}{\nu+1-\gamma} = \frac{1}{2} \frac{\nu(\nu+1) - \gamma(\gamma-1)}{\nu+1-\gamma}.$$

Close examination of (24) reveals that, when $\gamma < \nu$, the mean converges to γ (respectively, to ν) as α approaches ∞ (respectively, $-\infty$). This is consistent with the fact that under these conditions, the DTP distribution converges weakly to a point mass at γ (respectively, ν), as observed earlier. Moreover, as shown in the next result, the mean actually changes monotonically with α across these two extreme values.

Proposition 2.10. *When $\gamma < \nu$, the mean of a $DTP_\alpha(\gamma, \nu)$ random variable is a monotonically decreasing function of α on the entire real line, with the limits of γ and ν at ∞ and $-\infty$, respectively.*

To obtain the variance of X , it is useful to know its second moment. The latter is given in the result below, whose elementary, albeit computationally intensive, proof will be omitted.

Proposition 2.11. *The second moment of a $DTP_\alpha(\gamma, \nu)$ random variable X is a continuous function of α on the entire real line, given by*

$$(27) \quad E(X^2) = \frac{\gamma^2(\nu+1)^\alpha - \nu^2\gamma^\alpha}{(\nu+1)^\alpha - \gamma^\alpha} + \frac{(\nu+1)^\alpha\gamma^\alpha}{(\nu+1)^\alpha - \gamma^\alpha} \sum_{k=\gamma+1}^{\nu} \frac{2k-1}{k^\alpha}$$

for $\alpha \neq 0$ and

$$(28) \quad E(X^2) = \frac{\nu^2 \log(\nu+1) - \gamma^2 \log \gamma - \sum_{k=\gamma+1}^{\nu} (2k-1) \log k}{\log(\nu+1) - \log \gamma}$$

for $\alpha = 0$. Moreover, when $\gamma < \nu$, the second moment of X converges to γ^2 (respectively, to ν^2) as α approaches ∞ (respectively, $-\infty$).

As before, when $\gamma = \nu$, the sums in the expressions for $E(X^2)$ are understood as zeros, in which case we obtain $E(X^2) = \gamma^2$, consistent with the fact that the variance is zero in this case. In the case of the uniform distribution ($\alpha = -1$), we have

$$(29) \quad E(X^2) = \frac{\sum_{k=\gamma+1}^{\nu} k^2}{\nu+1-\gamma} = \frac{1}{6} \frac{\nu(\nu+1)(2\nu+1) - \gamma(\gamma-1)(2\gamma-1)}{\nu+1-\gamma}.$$

2.3. Further properties. Here we present several interesting properties of DTP distributions. First, we note that if a DTP distribution is truncated, either above or below (or both), then the resulting distribution is also a DTP distribution. This is not surprising in view of the fact that this distribution is obtained by truncation in the first place. The following result can be shown in a straightforward way by inspection, so we shall omit its derivation.

Proposition 2.12. *Suppose that X has a $DTP_\alpha(\gamma, \nu)$ distribution, and let $Y \stackrel{d}{=} X | \tilde{\gamma} \leq X \leq \tilde{\nu}$, where $\gamma \leq \tilde{\gamma} \leq \tilde{\nu} \leq \nu$. Then Y has a $DTP_\alpha(\tilde{\gamma}, \tilde{\nu})$ distribution.*

Next, we derive a stochastic representation of a DTP random variable with $\alpha > 0$, involving a *truncated geometric* (TG) variable. The latter is a geometric random variable truncated below and above by γ and ν , respectively, where, as before, γ and ν are two positive integers satisfying $0 < \gamma \leq \nu$, see, e.g., Kapadia and Thomasson [23], Sandland [36], and Thomasson and Kapadia [39]. Thus, the PMF of a TG random variable Y is of the form

$$(30) \quad f(x) = P(Y = x) = \frac{p(1-p)^x}{(1-p)^\gamma - (1-p)^{\nu+1}} \quad \text{for } x = \gamma, \gamma+1, \dots, \nu.$$

Suppose that the parameter p of the above distribution, denoted by $TG_p(\gamma, \nu)$, is also a random variable, with the same distribution as that of $1 - e^{-T}$, where T has a *weighted gamma* distribution $WG_\alpha(\beta_1, \beta_2)$ given by the PDF

$$(31) \quad g(t) = \frac{\beta_1^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\beta_1 t} \left\{ \frac{\beta_2^\alpha}{\beta_2^\alpha - \beta_1^\alpha} \left(1 - e^{-(\beta_2 - \beta_1)t} \right) \right\} \quad \text{for } 0 < t < \infty.$$

It follows that, with a suitable choice of the parameters β_1 and β_2 , the distribution of such a TG variable with a stochastic p has a DTP distribution.

Proposition 2.13. *Suppose that T has a $WG_\alpha(\beta_1, \beta_2)$ distribution given by the PDF (31) with $\alpha > 0$, $\beta_1 = \gamma$, and $\beta_2 = \nu + 1$, and, given $T = t$, Y has a TG distribution given by the PMF (30) with $p = 1 - e^{-t}$. Then, unconditionally, Y has the $DTP_\alpha(\gamma, \nu)$ distribution.*

3. ESTIMATION

Here, we consider the problem of estimating the three parameters of a $DTP_\alpha(\gamma, \nu)$ distribution, based on a random sample $X_1, \dots, X_n$. Denoting by

$$(32) \quad X_{(1)} \leq \dots \leq X_{(n)}$$

the corresponding order statistics allows us to express the likelihood function as

$$(33) \quad L(\alpha, \gamma, \nu) = \prod_{i=1}^n \frac{\frac{1}{X_{(i)}^\alpha} - \frac{1}{(X_{(i)}+1)^\alpha}}{\frac{1}{\gamma^\alpha} - \frac{1}{(\nu+1)^\alpha}} \quad \text{for } \gamma \leq X_{(1)}, \nu \geq X_{(n)}, -\infty < \alpha < \infty.$$

We can now utilize the properties of the $DTP_\alpha(\gamma, \nu)$ PMF (which appears under the product in the above expression) stated in Propositions 2.4 - 2.7 to establish the existence and uniqueness of maximum likelihood estimators (MLEs) of all three parameters. Indeed, by Proposition 2.4, the likelihood function is monotonically increasing in γ and decreasing in ν (over their respective domains) for any fixed value of $\alpha \in (-\infty, \infty)$. This immediately leads to the conclusion that the MLEs of these two parameters are the extreme order statistics,

$$(34) \quad \hat{\gamma} = X_{(1)} \quad \text{and} \quad \hat{\nu} = X_{(n)}.$$

This is analogous to the continuous truncated Pareto case (2) considered in Aban et al. [1]. To find the MLE of α , we need to maximize the function $L(\alpha, \hat{\gamma}, \hat{\nu})$ with respect to $\alpha \in (-\infty, \infty)$, or, equivalently, maximize its logarithm given by

$$(35) \quad \log L(\alpha, \hat{\gamma}, \hat{\nu}) = \sum_{i=1}^n \log \frac{\frac{1}{X_{(i)}^\alpha} - \frac{1}{(X_{(i)}+1)^\alpha}}{\frac{1}{X_{(1)}^\alpha} - \frac{1}{(X_{(n)}+1)^\alpha}} \quad -\infty < \alpha < \infty.$$

In connection with this, we have the following result.

Proposition 3.1. *Let $X_1, \dots, X_n$ be a random sample from a $DTP_\alpha(\gamma, \nu)$ distribution, and let (32) be the corresponding order statistics. If $X_{(1)} < X_{(n)}$, then there exists a unique MLE of α , which satisfies the equation*

$$(36) \quad \frac{1}{n} \sum_{i=1}^n \frac{\log(X_{(i)} + 1)}{\left(\frac{X_{(i)}+1}{X_{(i)}}\right)^\alpha - 1} - \frac{1}{n} \sum_{i=1}^n \frac{\log X_{(i)}}{1 - \left(\frac{X_{(i)}}{X_{(i)}+1}\right)^\alpha} = \frac{\log(X_{(n)} + 1)}{\left(\frac{X_{(n)}+1}{X_{(1)}}\right)^\alpha - 1} - \frac{\log X_{(1)}}{1 - \left(\frac{X_{(1)}}{X_{(n)}+1}\right)^\alpha}.$$

Moreover, the sign of the unique MLE $\hat{\alpha}$ coincides with that of the statistic

$$(37) \quad S(X_1, \dots, X_n) = \sum_{i=1}^n \log \frac{X_{(1)}}{X_{(i)}} - \sum_{i=1}^n \log \frac{X_{(i)} + 1}{X_{(n)} + 1}.$$

In particular, if $S(X_1, \dots, X_n) = 0$ then also $\hat{\alpha} = 0$.

Remark 3.1. Let us note that finding the MLE of α in practice is a standard numerical maximization process, as the proof of Proposition 3.1 shows that the log-likelihood function is differentiable and concave with a unique local maximum. To find the latter, one can utilize any of the standard statistical packages, for example R, to compute and optimize the likelihood function for the given sample.

Remark 3.2. We note, that when $\gamma = 1$ and $\nu \rightarrow \infty$, the $DTP_\alpha(\gamma, \nu)$ converges (in distribution) to the discrete non-truncated Pareto distribution on $\mathbb{N} = \{1, 2, \dots\}$, given by the PMF (4). Thus, the latter may be considered as a special case of the DTP distribution. Further, when we have a sample $X_1, \dots, X_n$ from this discrete non-truncated Pareto, the likelihood function (33) is replaced by its special case obtained when $\gamma = 1$ and $\nu \rightarrow \infty$, which is

$$(38) \quad L(\alpha) = \prod_{i=1}^n \left\{ \frac{1}{X_{(i)}^\alpha} - \frac{1}{(X_{(i)}+1)^\alpha} \right\}.$$

In this case, similar to the result in Proposition 3.1, it can be shown that, when $X_{(1)} < X_{(n)}$, there exists a unique MLE of α , which satisfies equation (36) with the right hand side replaced by zero:

$$(39) \quad \frac{1}{n} \sum_{i=1}^n \frac{\log(X_{(i)} + 1)}{\left(\frac{X_{(i)}+1}{X_{(i)}}\right)^\alpha - 1} - \frac{1}{n} \sum_{i=1}^n \frac{\log X_{(i)}}{1 - \left(\frac{X_{(i)}}{X_{(i)}+1}\right)^\alpha} = 0.$$

Equation (39) can be solved numerically using any of the standard computing packages.

An alternative way of finding an estimate of α is to use the method of moments, utilizing the formulas for the mean of a DTP distribution given in Proposition 2.9. One reasonable strategy is to use the MLEs of γ and ν , given by (34), in the expression on the right-hand-side of (24) and set the latter equal to the sample mean $\bar{X}$. This leads to the following equation for α :

$$(40) \quad \bar{X} = \frac{X_{(1)}(X_{(n)} + 1)^\alpha - X_{(n)}X_{(1)}^\alpha}{(X_{(n)} + 1)^\alpha - X_{(1)}^\alpha} + \frac{(X_{(n)} + 1)^\alpha X_{(1)}^\alpha}{(X_{(n)} + 1)^\alpha - X_{(1)}^\alpha} \sum_{k=X_{(1)}+1}^{X_{(n)}} \frac{1}{k^\alpha}.$$

To solve equation (40) we note that, according to Proposition 2.10 with $\gamma = X_{(1)}$ and $\nu = X_{(n)}$, the function on the right-hand-side of (40) is monotonically decreasing in α . Moreover, by the same result, the limits of this function at $-\infty$ and ∞ are $X_{(n)}$ and $X_{(1)}$, respectively. Since $X_{(1)} < \bar{X} < X_{(n)}$ whenever $X_{(1)} < X_{(n)}$, under the latter condition the solution of (40) exists and is unique. Again, the solution of (40) for a given sample can be found using standard numerical solvers available in computing packages such as R. We can also evaluate the expression on the right-hand-side of (40) for $\alpha = 0$, which coincides with the right-hand-side of (25) with $\gamma = X_{(1)}$ and $\nu = X_{(n)}$. When we denote this quantity by

$$(41) \quad T = T(X_1, \dots, X_n) = \frac{X_{(n)} \log(X_{(n)} + 1) - X_{(1)} \log X_{(1)} - \sum_{k=X_{(1)}+1}^{X_{(n)}} \log k}{\log(X_{(n)} + 1) - \log X_{(1)}},$$

we can state the following useful practical rule: the sign of $\hat{\alpha}$ is the same as that of $T - \bar{X}$. In particular, if $T = \bar{X}$ then $\hat{\alpha} = 0$.

4. AN ILLUSTRATION

The DTP model has potentially wide applications in the science of ecology and related disciplines, where counts are frequently studied (Fordyce et al. [16]). For example, ecologists study the number of plants in survey plots and the number of lice on birds (Lee and Clayton [27]), and paleontologists study the number of species that survive mass extinction events (Coccioni and Galeotti [13]). A DTP can be expected to arise from some of the processes involved in these phenomena, and similar patterns are evident in certain systems studied by geneticists and evolutionary biologists. For example, Lynch [28] reports a frequency distribution for the age of gene duplications in the human genome: the vast majority are young, while a long, thin tail of the distribution is comprised of relatively old duplicates, thus resembling a DTP. Here we offer a frequency distribution for the number of host plant species and families associated with (attacked by) insect herbivores in a temperate deciduous forest.

The data shown in Figure 1 come from 4 years (2006-2009) of sampling Lepidoptera (moths and butterflies) across thousands of hectares in southwestern Ohio in the midwestern United States. As described in Dyer et al. [15] and Forister et al. [17], associations between Lepidoptera species and plants that they feed on were established by sampling caterpillars directly from vegetation. For each species of butterfly and moth, the data consists of the count of the number of plant species and, subsequently, the number of plant families (a higher taxonomic unit encompassing species) with which they were associated. In what follows, we will refer to that count (the number of plant species and families) as the “diet breadth” of the insect herbivores. It is of interest to consider diet breadth at both the species and family level because they represent different levels of evolutionary resolution, one fine and one more coarse-grained. We fit four models to diet breadth measured by the number of host plant species and families: Poisson, geometric, DTP, and discrete non-truncated Pareto. We chose Poisson and geometric for comparison with DTP, because they are often used when modeling count data in the ecological sciences (Bolker [6]). We also fit the discrete non-truncated Pareto model to show that the truncated model fits the data better.

We estimated the parameters of the Poisson and geometric models using maximum likelihood estimation. For Poisson models, the estimated rate was 0.3168 for the species, and 0.3986 for the families of host plants. For the geometric model, the estimated probability of success was 0.3168 for the species and 0.3986 for the families of host plants. The estimates of the Poisson rate and geometric probability of success are the same, both estimated by the inverse of the sample mean. The parameters of the DTP model were estimated using maximum likelihood estimators as well. This estimation yielded, for species diet breadth, $\hat{\alpha} = 0.897$, $\hat{\nu} = 26$. For family diet breadth we obtained $\hat{\alpha} = 1.068$, $\hat{\nu} = 15$. The lower truncation parameter estimates were $\hat{\gamma} = 1$ for both data sets. Parameters of the discrete non-truncated Pareto model were also estimated using maximum likelihood method², as indicated in Remark 3.2.

We assess the fit of the models graphically using the logarithm of the model and empirical survival functions, see Figure 1. Note, that the species diet breadth is always at least as large as the family diet breadth, because each family contains several species. The distribution of both the number of host species associations and host family associations per Lepidoptera species are well fit by a DTP (Figure 1). In fact, DTP fits the data better than all the other models used. The histograms of diet breadth for both plant species and families show that the majority of Lepidoptera species attack one plant species (or family), while a few species have relatively large diet breadths in the far right tail of

²R code for all estimation is available from A. Panorska upon request.

the distribution. This right skewness and peakedness of the distribution of diet breadth is typical for power tail models such as Pareto. Although both geometric and Poisson models are right skewed and peaked, their tails are shorter than the tails of power law models. However, the main indication of the power-law distribution of the data is linearity of the log-log plot of the empirical survival function. In the case of regular (not truncated) power law models, the entire right tail of the log-log plot of the empirical survival function is following a straight line. Truncation is indicated by the right tail following a line for a while and then falling off faster at the extreme right end of the log-log plot (see also plots in [1]). In Figure 1 we see this change in slope of the empirical survival function suggesting truncation in the far right of the log-log plots for both species and family diet breadths. Figure 1 shows that the best model for the diet breadth data is provided by DTP, which is especially evident for family-level diet breadth. The better fit of the discrete, truncated Pareto relative to the discrete non-truncated Pareto is to be expected given the nature of the data. In any one ecological community, there is a finite number of plant species available to any one herbivore, and the number of hosts eaten by herbivores with even the most generalized feeding habits tends to be far fewer than the possible number of hosts.

Although the factors that determine host plant associations for insect herbivores are an area of active research and much has yet to be learned (e.g. Jorge et al. [22]), we can speculate as to why diet breadth (the number of hosts attacked) is well fit by a DTP. The predominance of extreme specialists (with a diet breadth of 1) potentially reflects a process of speciation or diversification that is biased towards the proliferation of specialists (Nosil [31]), while the relative scarcity of generalists could reflect inherent constraints that make the colonization of new host species a difficult and rare outcome. However, once new hosts have been colonized, it may become easier and easier for an increasingly-generalized herbivore to utilize even more hosts. Thus we suggest that a kind of “snowball effect” could explain the long, thin tail of the diet breadth distribution (Forister et al. [17]). It is interesting to note that Bascompte [4] reports a frequency distribution of the number of interactions for each species of pollinator in a mutualistic network, and the distribution shows the same high peakedness and extreme skew as the diet breadth distributions reported here (and in Forister et al. [17]), which suggests that the DTP might have general utility in ecology for the study of interactions among species.

FIGURE 1. Distribution of diet breadth for Lepidoptera species (top row) and families (bottom row). Left panel histograms: Horizontal axis shows the number of host plant species (top) and families (bottom) attacked by species of Lepidoptera; i.e. bar height is the number of Lepidoptera attacking a given number of hosts. Tick marks under the histograms indicate individual observations. Right panel: Natural logarithm of the empirical survival function (solid dots), and the best fitting DTP (open circles), non-truncated Pareto (triangles), geometric (squares) and Poisson (diamonds) models.

5. PROOFS

We start with several lemmas, which are subsequently used in our proofs.

Lemma 5.1. *For any $\alpha \in (0, \infty)$ and $c \in (0, \infty)$ we have*

$$(42) \quad \frac{c^\alpha - 1}{\alpha} \geq \log c,$$

with a strict inequality in (42) whenever $c \neq 1$.

Proof. The inequality (42) is easily shown using standard calculus methods, and we skip the details.

Lemma 5.2. *For any real α and $0 < \gamma < \delta < \infty$, let*

$$(43) \quad u(x) = \frac{\delta^\alpha \log \frac{x}{\gamma} + \gamma^\alpha \log \frac{\delta}{x}}{x^\alpha}, \quad \gamma \leq x \leq \delta.$$

Then, $u(x) > \log(\delta/\gamma) = u(\gamma) = u(\delta)$ whenever $\gamma < x < \delta$, unless $\alpha = 0$, in which case $u(x) = \log(\delta/\gamma)$ for all $\gamma \leq x \leq \delta$.

Proof. Clearly, when $\alpha = 0$, the function u is equal to $\log(\delta/\gamma)$ for every $\gamma \leq x \leq \delta$, so we shall assume in the sequel that $\alpha \neq 0$. Observe that $u(\gamma) = u(\delta) = \log(\delta/\gamma)$. Since u is continuous, the result will follow if we can show that, for some $r \in (\gamma, \delta)$, the function u is increasing on (γ, r) and is decreasing on (r, δ) . To achieve this goal, we examine the derivative of u , which after standard algebra, turns out to be

$$(44) \quad u'(x) = \frac{\delta^\alpha - \gamma^\alpha - [\alpha(\delta^\alpha - \gamma^\alpha) \log x + \alpha\gamma^\alpha \log \delta - \alpha\delta^\alpha \log \gamma]}{x^{\alpha+1}}.$$

Thus, the derivative is positive if and only if

$$(45) \quad v(x) < \delta^\alpha - \gamma^\alpha,$$

where

$$v(x) = \alpha(\delta^\alpha - \gamma^\alpha) \log x + \alpha\gamma^\alpha \log \delta - \alpha\delta^\alpha \log \gamma.$$

Note that for any $\alpha \neq 0$, the quantity $\alpha(\delta^\alpha - \gamma^\alpha)$ is positive, so that the function v given above is monotonically increasing on the interval (γ, δ) . Consequently, if we can show that

$$(46) \quad v(\gamma) < \delta^\alpha - \gamma^\alpha \quad \text{and} \quad v(\delta) > \delta^\alpha - \gamma^\alpha,$$

it follows by continuity of v that there exists a unique $r \in (\gamma, \delta)$ such that (45) holds if and only if $\gamma < x < r$. In turn, this will imply that the function u is increasing on (γ, r) and is decreasing on (r, δ) , as desired.

To establish (46), first note that

$$v(\gamma) = \alpha\gamma^\alpha(\log \delta - \log \gamma) \text{ and } v(\delta) = \alpha\delta^\alpha(\log \delta - \log \gamma).$$

Assuming that $\alpha > 0$, after some algebra we find that the inequalities (46) are equivalent to

$$(47) \quad \frac{1}{\alpha} \left[\left(\frac{\delta}{\gamma} \right)^\alpha - 1 \right] > \log \frac{\delta}{\gamma} \text{ and } \frac{1}{\alpha} \left[\left(\frac{\gamma}{\delta} \right)^\alpha - 1 \right] > \log \frac{\gamma}{\delta},$$

which are true by Lemma 5.1. Similarly, one can show that the inequalities (46) are true for $\alpha < 0$. This concludes the proof.

Proof of Proposition 2.3. Consider the PMF as a function $h(x)$ of a *real* argument x on the interval $[\gamma, \nu]$. The result can be established by checking the sign of the derivative

$$(48) \quad h'(x) = \alpha \frac{\frac{1}{(x+1)^{\alpha+1}} - \frac{1}{x^{\alpha+1}}}{\frac{1}{\gamma^\alpha} - \frac{1}{(\nu+1)^\alpha}}.$$

Proof of Proposition 2.5. First, consider the case $x = \gamma$. Further, assume that $\alpha \in (0, \infty)$ and write the DTP PMF, as a function of α , in the form

$$(49) \quad v(\alpha) = \frac{1 - a^\alpha}{1 - b^\alpha},$$

where

$$(50) \quad 0 < b = \frac{\gamma}{\nu + 1} < \frac{\gamma}{\gamma + 1} = a < 1.$$

Observe that the limit of v at infinity is equal to 1 as stated in the Proposition, since both, a and b , are positive quantities less than 1. Upon taking the derivative of $v(\alpha)$ and setting it to be larger than zero, followed by routine calculations, we arrive at the inequality

$$(51) \quad u(\alpha) = \left[\left(\frac{1}{b} \right)^\alpha - 1 \right] \frac{\log a}{\log b} - \left[\left(\frac{1}{a} \right)^\alpha - 1 \right] > 0.$$

Since

$$(52) \quad u'(\alpha) = \log a \left[\left(\frac{1}{a} \right)^\alpha - \left(\frac{1}{b} \right)^\alpha \right] > 0$$

for $\alpha > 0$ and $u(0) = 0$, we conclude that the inequality (51) holds for $\alpha > 0$, and thus the function $v(\alpha)$ is increasing on $(0, \infty)$. We now assume that $\alpha \in (-\infty, 0)$, in which case the DTP PMF takes on the form

$$(53) \quad w(\beta) = \frac{A^\beta - 1}{B^\beta - 1},$$

where $\beta = -\alpha > 0$ and

$$(54) \quad 1 < A = \frac{\gamma + 1}{\gamma} < \frac{\nu + 1}{\gamma} = B < \infty.$$

Clearly, the function $w(\beta)$ converges to zero at infinity, so that the DTP PMF converges to zero at minus infinity. Using similar argument to the one above, one can show that $w(\beta)$ is decreasing on $(0, \infty)$, so that the DTP PMF is increasing in α on the interval $(-\infty, 0)$. Together with the continuity of the DTP PMF, the above analysis shows the DTP PMF is indeed an increasing function on the entire real line when $x = \gamma$. Using similar arguments, one can show that the result for the case $x = \nu$ stated in Proposition 2.5 holds as well. This concludes the proof.

Proof of Proposition 2.6. Clearly, the DTP PMF is a continuous and differentiable function of α at every $\alpha \neq 0$. The continuity of the PMF at $\alpha = 0$ follows by construction, as discussed in Remark 2.1 following Definition 2.1. It remains to show the PMF $f(x)$ is differentiable at $\alpha = 0$. Rather lengthy algebra involving several applications of L'Hospital's rule, leads to the conclusion that the limit

$$(55) \quad \lim_{\alpha \rightarrow 0} \frac{f(\alpha) - f(0)}{\alpha}$$

exists and is given by (19). This concludes the proof.

Proof of Proposition 2.7. Since the limits of the function $\log f(x)$ as $\alpha \rightarrow \pm\infty$ are straightforward to derive, we focus on proving that the function $\log f(x)$ is concave in α on $(-\infty, \infty)$. We proceed by showing that the second derivative of $\log f(x)$ is negative for any $\alpha \neq 0$. This will prove concavity of $\log f(x)$, since, by Proposition 2.6, this function is differentiable on the entire real line. Routine calculations produce the first and the second derivatives of $\log f(x)$ as follows:

$$(56) \quad \frac{\partial \log f(x)}{\partial \alpha} = -\frac{\log x}{1 - \left(\frac{x}{x+1}\right)^\alpha} + \frac{\log(x+1)}{\left(\frac{x+1}{x}\right)^\alpha - 1} + \frac{\log \gamma}{1 - \left(\frac{\gamma}{\nu+1}\right)^\alpha} - \frac{\log(\nu+1)}{\left(\frac{\nu+1}{\gamma}\right)^\alpha - 1}.$$

$$\frac{\partial^2 \log f(x)}{\partial \alpha^2} = -\frac{(x+1)^\alpha x^\alpha}{[(x+1)^\alpha - x^\alpha]^2} \log^2 \frac{x+1}{x} + \frac{(\nu+1)^\alpha \gamma^\alpha}{[(\nu+1)^\alpha - \gamma^\alpha]^2} \log^2 \frac{\nu+1}{\gamma}.$$

To prove concavity, we need to show that the second derivative of $\log f(x)$ is negative, that is

$$(57) \quad \frac{(\nu+1)^\alpha \gamma^\alpha}{[(\nu+1)^\alpha - \gamma^\alpha]^2} \log^2 \frac{\nu+1}{\gamma} < -\frac{(x+1)^\alpha x^\alpha}{[(x+1)^\alpha - x^\alpha]^2} \log^2 \frac{x+1}{x}$$

for any

$$(58) \quad 0 < \gamma \leq x < x+1 \leq \nu+1$$

and $\alpha \neq 0$ (except for the special case $\gamma = x = \nu$, where the inequality above becomes an equality). Note that we can restrict attention to $\alpha > 0$, since the expressions on both sides of (57) are even functions

of α . It is easily verified that, for $\alpha > 0$, the inequality (57) is equivalent to

$$(59) \quad G\left(\frac{\nu+1}{\gamma}\right) < G\left(\frac{x+1}{x}\right),$$

where

$$(60) \quad G(u) = \frac{\log u}{u^{\alpha/2} - u^{-\alpha/2}}, \quad 1 < u < \infty.$$

Clearly, under the condition (58), we have

$$1 < \frac{x+1}{x} \leq \frac{\nu+1}{x} \leq \frac{\nu+1}{\gamma},$$

with at least one inequality above being strict (unless we have the exceptional case $\gamma = x = \nu$). Thus, the inequality (59) follows if the function G is decreasing. The latter fact is indeed the case, and can be shown by examining the derivative of G . The result follows.

Proof of Proposition 2.10. Since the derivation of the limits of the mean at $\pm\infty$ is straightforward, we focus on the monotonicity of the mean. To prove that the function (24) is decreasing in α , we shall show that its derivative is negative whenever $\alpha \neq 0$. Routine (but tedious) calculations lead to the following expression for the derivative of (24):

$$(61) \quad \frac{\partial}{\partial \alpha} E(X) = \frac{\gamma^\alpha (\nu+1)^\alpha (\nu-\gamma)}{[(\nu+1)^\alpha - \gamma^\alpha]^2} \left\{ \log \frac{\nu+1}{\gamma} - \frac{1}{\nu-\gamma} \sum_{k=\gamma+1}^{\nu} \frac{(\nu+1)^\alpha \log \frac{k}{\gamma} + \gamma^\alpha \log \frac{\nu+1}{k}}{k^\alpha} \right\}.$$

Thus, the derivative is negative whenever

$$(62) \quad \log \frac{\nu+1}{\gamma} < \frac{1}{\nu-\gamma} \sum_{k=\gamma+1}^{\nu} \frac{(\nu+1)^\alpha \log \frac{k}{\gamma} + \gamma^\alpha \log \frac{\nu+1}{k}}{k^\alpha}.$$

Note that the right-hand-side in (62) is the arithmetic average of the $\nu - \gamma$ quantities $u(k)$, $k = \gamma+1, \gamma+2, \dots, \nu$, where

$$(63) \quad u(x) = \frac{(\nu+1)^\alpha \log \frac{x}{\gamma} + \gamma^\alpha \log \frac{\nu+1}{x}}{x^\alpha}.$$

Since for every k we have $0 < \gamma < k < \nu+1$, an application of Lemma 5.2 with $\delta = \nu+1$ shows that $u(k) > u(\gamma) = u(\nu+1) = \log[(\nu+1)/\gamma]$. Consequently, the arithmetic average the $\{u(k)\}$ is also larger than $\log[(\nu+1)/\gamma]$. This concludes the proof.

Proof of Proposition 2.13. Using total probability formula, we have

$$P(Y = k) = \int_0^\infty P(Y = k|T = t)g(t)dt = \int_0^\infty \frac{(1 - e^{-t})e^{-tk}}{e^{-t\gamma} - e^{-t(\nu+1)}}g(t)dt,$$

where $g(t)$ is given by (31) with $\beta_1 = \gamma$ and $\beta_2 = \nu + 1$. After some algebra, the above simplifies to

$$P(Y = k) = \left(\frac{1}{\gamma^\alpha} - \frac{1}{(\nu + 1)^\alpha} \right)^{-1} \left(\int_0^\infty \frac{1}{\Gamma(\alpha)} t^{\alpha-1} e^{-kt} dt - \int_0^\infty \frac{1}{\Gamma(\alpha)} t^{\alpha-1} e^{-(k+1)t} dt \right),$$

which produces (5), as the two integrals in the above expression reduce to $k^{-\alpha}$ and $(k+1)^{-\alpha}$, respectively. This concludes the proof.

Proof of Proposition 3.1. First, we note that if $X_{(1)} = X_{(n)}$, so that all sample values coincide, the log-likelihood function $\log L(\alpha, \hat{\gamma}, \hat{\nu})$ given by (35) is identically equal to zero. Thus, in this case, there is no unique MLE of α . Second, we recognize the terms under the log on the right-hand-side in (35) to be the PMFs of the $DTP_\alpha(X_{(1)}, X_{(n)})$ distribution evaluated at the order statistics. Consequently, in view of Propositions 2.6 - 2.7, $\log L(\alpha, \hat{\gamma}, \hat{\nu})$ is a differentiable and concave function of α on the entire real line, with the limits of $-\infty$ at $\pm\infty$. This ensures the uniqueness of $\hat{\alpha}$, the MLE of α . The value of $\hat{\alpha}$ is obtained by setting the derivative of $\log L(\alpha, \hat{\gamma}, \hat{\nu})$ equal to zero. In computing this derivative, we utilize the expression (56), with x , γ , and ν replaced by $X_{(i)}$, $X_{(1)}$, and $X_{(n)}$, respectively, used before in the proof of Proposition 2.7. Following routine calculations, this process leads to equation (36) for the MLE. Finally, to get the condition for the sign of $\hat{\alpha}$, we can evaluate the derivate of $\log L(\alpha, \hat{\gamma}, \hat{\nu})$ at $\alpha = 0$, the computation of which is aided by Proposition 2.6 coupled with equation (8). In connection with this, we have

$$(64) \quad \left. \frac{\partial \log L(\alpha, \hat{\gamma}, \hat{\nu})}{\partial \alpha} \right|_{\alpha=0} = \sum_{i=1}^n \frac{1}{2} \log \frac{X_{(1)} (X_{(n)} + 1)}{X_{(i)} (X_{(i)} + 1)}.$$

Clearly, when this derivative is positive, the function $\log L(\alpha, \hat{\gamma}, \hat{\nu})$ is increasing at $\alpha = 0$, so that the maximum must occur at $\hat{\alpha} > 0$. Similarly, the negative value of this expression implies $\hat{\alpha} < 0$, while the value of zero is a reflection of $\hat{\alpha} = 0$. These conditions can be restated in the form stated in the proposition. This concludes the proof.

6. ACKNOWLEDGMENTS

Thanks to John Stireman, Wright State University, for providing Lepidoptera diet breadth data (collected in association with US National Science Foundation award DEB-1020571). Thanks to the associate editor and an anonymous referee for suggestions that improved this paper. Kozubowski's and Panorska's research was partially funded by the European Union's Seventh Framework Programme for Research, Technological Development and Demonstration under grant agreement no 318984 - RARE. Forister's research was partially supported by National Science Foundation grants DEB-1020509 and DEB-1145609.

REFERENCES

- [1] I.B. Aban, M.M. Meerschaert, A.K. Panorska, Parameter estimation for the truncated Pareto distribution, *J. Am. Stat. Assoc.* 101(2006) 270-277.
- [2] B.C. Arnold, *Pareto Distributions*, International Co-operative Publishing House, Maryland, 1983.
- [3] N. Balakrishnan, N.D. Nevzorov, *A Primer on Statistical Distributions*, John Wiley & Sons, New Jersey, 2003.
- [4] J. Bascompte, P. Jordano, Plant-animal mutualistic networks: the architecture of biodiversity, *Annual Review of Ecology, Evolution, and Systematics*, (2007) 567-593.
- [5] F. Benford, The law of anomalous numbers, *Proc. Am. Philos. Soc.* 78(4)(1938) 551 - 572.
- [6] B.M. Bolker, *Ecological models and data in R*, Princeton University Press, Princeton, 2008.
- [7] A. Buddana, *A Discrete Pareto Distribution on Positive Integers*, Thesis, University of Nevada, Reno, 2011.
- [8] A. Buddana, T.J. Kozubowski, Discrete Pareto distributions, *Econ. Quality Control*, 29(2)(2014) 143-156.
- [9] S.M. Burroughs, S.F. Tebbens, Upper-truncated power law distributions, *Fractals*, 9(2001a) 209-222.
- [10] S.M. Burroughs, S.F. Tebbens, Upper-truncated power laws in natural systems. *J. of Pure and Appl. Geophysics*, 158(2001b) 741-757.
- [11] S.M. Burroughs, S.F. Tebbens, The upper-truncated power law applied to earthquake cumulative frequency-magnitude distributions. *Bull. of the Seism. Soc. of Am.*, 92(8)(2002) 2983-2993.
- [12] A. Clauset, M.E.J. Newman, Power-law distributions in empirical data, *SIAM Rev.* 2(2009) 661-703.
- [13] R. Coccioni, S. Galeotti, KT boundary extinction: geologically instantaneous or gradual event? Evidence from deep-sea benthic foraminifera. *Geology* 22(9)(1994) 779-782.
- [14] P.C. Consul, Evolution of surnames, *Internat. Stat. Rev.* 59(3)(1991) 271-278.
- [15] L.A. Dyer, M.S. Singer, J.T. Lill, J.O. Stireman, G.L. Gentry, R.J. Marquis, R.E. Ricklefs, H.F. Greeney, D.L. Wagner, H.C. Morais, I.R. Diniz, T.A. Kursar, P.D. Coley, Host specificity of Lepidoptera in tropical and temperate forests, *Nature* 448(2007) 696-699.
- [16] J.A. Fordyce, Z. Gompert, M.L. Forister, C.C. Nice, A hierarchical Bayesian approach to ecological count data: a flexible tool for ecologists, *PLoS one*(2011) e26785.
- [17] M.L. Forister, Y. Basset, A.K. Panorska, P.D. Coley, I.R. Diniz, P. Drozd, M. Fox, A. Glassmire, R. Hazen, J. Hrcek, J.P. Jahner, T.J. Kozubowski, T.A. Kursar, J. Lill, R.J. Marquis, S.E. Miller, H.C. Morais, M. Murakami, V. Novotny, N. Pardikes, R.E. Ricklefs, M.S. Singer, A.M. Smilanich, J.O. Stireman, D.L. Wagner, T. Walla, G.D. Weiblen, L.A. Dyer, The global distribution of diet breadth in insect herbivores, *Proceedings of the National Academy of Sciences of the United States of America*, 112(2)(2015) 442447.
- [18] M. Greiner, M. Jobmann, L. Lipsky, The importance of power-tail distributions for modeling queueing systems, *Oper. Res.* 47(2)(1999) 313-326.
- [19] T.P. Hill, A statistical derivation of the significant-digit law, *Stat. Sci.* 10(1995) 354-363.
- [20] T.P. Hill, The first digit phenomenon, *American Sci.* 86(4)(1998) 358.
- [21] N.L. Johnson, S. Kotz, A.W. Kemp, *Univariate Discrete Distributions*, John Wiley & Sons, New York, 1993.
- [22] L.R. Jorge, P.I. Prado, M. Almeida-Neto, T.M. Lewinsohn, An integrated framework to improve the concept of resource specialization, *Ecology Letters*, 17(11)(2014) 1341-1350.

- [23] C. Kapadia, R. Thomasson, On estimating the parameter of a truncated geometric distribution by the method of moments, *Ann. Inst. Stat. Math.* 27(1)(1975) 269-272.
- [24] S. Kotz, N.L. Johnson, *Continuous Univariate Distributions-1*, John Wiley & Sons, 1976.
- [25] H. Krishna, P.S. Pundir, Discrete Burr and discrete Pareto distributions, *J. Stat. Meth.* 6(2009) 177-188.
- [26] K. Krishnamoorthy, *Handbook of Statistical Distributions with Applications*, Chapman and Hall/CRC Press, Florida, 2006.
- [27] P.L. Lee, D.H. Clayton, Population biology of swift (*Apus apus*) ectoparasites in relation to host reproductive success, *Ecol. Entomol.* 20(1)(1995) 43-50.
- [28] M. Lynch, *The origins of genome architecture*, Sinauer Associates Sunderland, MA, 2007.
- [29] M. Muskulus, A.M. Slats, P.J. Sterk, S. Verduyn-Lunel, Fluctuations and determinism of respiratory impedance in asthma and chronic obstructive pulmonary disease, *J. of Applied Physiology*, 109(6)(2010) 1582-1591.
- [30] M.E.J. Newman, Power laws, Pareto distribution, Zipf's law, *Contemporary Physics* 46(2005) 323-325.
- [31] P. Nosil, *Ecological Speciation*, Oxford University Press, 2012.
- [32] J. Park, T.P. Seager, P.S.C. Rao, M. Convertino, I. Linkov, Integrating risk and resilience approaches to catastrophe management in engineering systems, *Risk Analysis: An International Journal*, 33(3)(2013) 356-367.
- [33] C.R. Rao, *Linear Statistical Inference and its Applications*, Wiley, New York, 1973.
- [34] W.J. Reed, The Pareto, Zipf and other power laws, *Econ. Lett.* 74(2001) 15-19.
- [35] W.J. Reed, B.D. Hughes, From gene families and genera to incomes and internet file sizes: Why power laws are so common in nature, *Phys. Rev. E* 66(2002) 067103.
- [36] R.L. Sandland, A note on some applications of the truncated geometric distribution, *Australian J. Stat.* 16(1974) 57-58.
- [37] D. Sornette, Multiplicative processes and power laws, *Phys. Rev. E* 57(1998) 4811-4813.
- [38] J.C. Stegen, N.G. Swenson, B.J. Enquist, E.P. White, O.L. Phillips, P.M. Jørgensen, M. D. Weiser, A. Monteagudo Mendoza, P. Núñez Vargas, Variation in above-ground forest biomass across broad climatic gradients, *Global Ecology and Biogeography*, 20(5)(2011) 744-754.
- [39] R. Thomasson, C. Kapadia, On estimating the parameter of a truncated geometric distribution, *Ann. Inst. Statist. Math.* 20(1968) 519-523.
- [40] C. Weidner, P. Kroupa, T. Maschberger, The influence of multiple stars on the high-mass stellar initial mass function and age dating of young star clusters, *Monthly Notices of the Royal Astronomical Society*, 393(2)(2009) 663-680.
- [41] G.K. Zipf, *Human Behavior and the Principle of Least Effort*, Cambridge, Massachusetts: Addison-Wesley, 1949.