Bernoulli **24**(2), 2018, 873–894
DOI: 10.3150/15-BEJ786

Domains of attraction on countable alphabets

ZHIYI ZHANG

Department of Mathematics and Statistics, University of North Carolina at Charlotte, Charlotte, NC 28223, USA. E-mail: zzhang@uncc.edu

For each probability distribution on a countable alphabet, a sequence of positive functionals are developed as tail indices. By and only by the asymptotic behavior of these indices, domains of attraction for all probability distributions on the alphabet are defined. The three main domains of attraction are shown to contain distributions with thick tails, thin tails and no tails respectively, resembling in parallel the three main domains of attraction, Fréchet, Gumbel and Weibull families, for continuous random variables on the real line. In addition to the probabilistic merits associated with the domains, the tail indices are partially motivated by the fact that there exists an unbiased estimator for every index in the sequence, which is therefore statistically observable, provided that the sample is sufficiently large.

Keywords: distributions on alphabets; domains of attraction; tail index; Turing's formula

1. Introduction and summary

Consider an alphabet with countably many letters $\mathscr{X} = \{\ell_k; k \ge 1\}$ and an associated probability distribution $P = \{p_k; k \ge 1\} \in \mathscr{P}$ where $\mathscr{P}$ is the class of all probability distributions on $\mathscr{X}$. Let $X_1, \dots, X_n$ be a sample of independently and identically distributed (i.i.d.) random elements from $\mathscr{X}$ under P. Let $\{\hat{p}_k; k \ge 1\}$ be the relative letter frequencies in the sample.

Before proceeding further, let us first give a little thought to possible notions of an "extreme value" and a "tail" of a distribution in the current setting, as domains of attraction are commonly discussed in association with such notions. While such notions are not required in the mathematics of this paper, it is nevertheless comforting to have them at least on an intuitive level. Unlike a sample of i.i.d. random variables on the real line where the values are numerically ordered and therefore an extreme value is naturally defined, the letters in an alphabet do not assume numerical values nor do they necessarily admit natural ordering. If we insist on having a notion of an extreme value associated with a sample, then perhaps such a value should be based on its rarity or unusualness with respect to the observed values in the sample. The rarest values in the sample are those with frequency one and there are most commonly many more than one such observed value in a sample. If we entertain a rarer value, it has to be one of those with frequency zero, that is, the letters in the alphabet that are not represented in the sample, which, though not in the sample, are nevertheless associated with and specified by the sample. If we anticipate that another independent observation, say X_{n+1}, is to be taken from $\mathscr{X}$, it would be reasonable then to consider the value of X_{n+1} to be extreme if X_{n+1} takes a letter that is not observed in the original sample of size n. To fix the idea, we will subsequently use the term "an extreme value" to mean a letter of the alphabet that is not represented in the sample of size n. Similarly we can also entertain what a notation of a tail should be on an alphabet. Whenever there is no risk of ambiguity, let us loosely refer to a subset of $\mathscr{X}$ with low probability letters as a "tail" in the

1350-7265

subsequent text. In this sense, a subset of $\mathscr{X}$ with very low probability letters may be referred to as a "distant tail", and a distribution on a finite alphabet has essentially "no tail". Furthermore we note that, though there is no natural ordering among the letters in $\mathscr{X}$, there is one on the index set $\{k; k \geq 1\}$. There therefore exists a natural notion of a distribution $P = \{p_k\}$ having a thinner tail than that of another distribution $Q = \{q_k\}$, in the sense of $p_k \leq q_k$ for all $k \geq k_0$ for some integer $k_0 \geq 1$, when P and Q share a same alphabet and are enumerated by a same index set. In such a case, we will subsequently say that P has a thinner tail than Q *in the usual sense*. Finally, we note that the discussion of domains of attraction for continuous random variables very much hinges on a well-defined extreme value, which is lacking on alphabets, and the differentiability of its cumulative distribution function, which is completely non-existent due to the discrete nature of alphabets. As a result of these characteristics, let us adopt the notation of an out-of-sample extreme value as described above. We may then entertain the probability of X_{n+1} being an extreme value, i.e., $P(\bigcap_{i=1}^{n}\{X_{n+1} \neq X_i\})$, which is, after a few algebraic steps,

$$\zeta_{1,n} = \sum_{k \geq 1} p_k (1 - p_k)^n.$$

Remark 1. $\zeta_{1,n}$ is a member of the family of the generalized Simpson's indices $\zeta_{u,v}$ discussed by Zhang and Zhou [11] which plays an important role in characterizing the underlying distribution $\{p_k\}$ (up to a permutation on the index set) and in giving alternative representations to Shannon's entropy and Rényi's entropy, which are well-known tail indices on an alphabet, as discussed in Zhang [9].

Clearly, $\zeta_{1,n} \to 0$ as $n \to \infty$ for any probability distribution $\{p_k\}$ on $\mathscr{X}$. A multiplicatively adjusted version of $\zeta_{1,n}$ is defined below and will subsequently be referred to as the tail index.

$$\tau_n = n\zeta_{1,n} = \sum_{k \geq 1} n p_k (1 - p_k)^n. \tag{1}$$

Remark 2. Suppose there are two independent samples of i.i.d. random variables of the same size n. The tail index τ_n in (1) may also be interpreted as the expected number of observations in one sample that are not found in the other sample.

On an intuitive level, τ_n is reflective of tail characteristics (or tail-relevant) since $\zeta_{1,n}$ is tail-relevant. To see that $\zeta_{1,n}$ is tail-relevant, let us first consider $\pi_0 = \sum_{k \geq 1} p_k 1[\hat{p}_k = 0]$. $1 - \pi_0$ is often referred to as the sample coverage of a population in the literature. Since the letters not represented in a large sample are likely those with low probabilities, it is reasonable to think that π_0 is a tail-relevant quantity for a large n; and yet $\zeta_{1,n} = \mathrm{E}(\pi_0)$. Intuitively one would expect π_0 to take a smaller (larger) value under a more (less) concentrated probability distribution, and therefore to expect $\zeta_{1,n}$, and hence τ_n, to be a reasonable measure to characterize the tail of a distribution on an alphabet. Also to be noted is that, for any given integer $k_0 \geq 1$, the first k_0 terms in the re-expression of τ_n below converges to zero exponentially fast as $n \to \infty$

$$\tau_n = \sum_{k \leq k_0} n p_k (1 - p_k)^n + \sum_{k > k_0} n p_k (1 - p_k)^n,$$

and therefore the asymptotic behavior of τ_n has essentially nothing to do with how the probabilities are distributed over any fixed and finite subset of $\mathscr{X}$, further noting that τ_n is invariant under any permutation on the index set $\{k\}$.

Domains of attraction for distributions of continuous random variables are a long-standing focal point of the extreme value theory. The large volume of research on this topic in the existing literature goes back to Fréchet [3] and Fisher and Tippett [2], and includes full analyses by Gnedenko [5] and Smirnov [8]. There the three main domains of attraction are defined along the lines of Fréchet family (thick tails), Gumbel family (thin tails) and Weibull family (no tails). For a comprehensive review on the topic, readers may wish to refer to de Haan and Ferreira [1]. The main objective of this paper is to similarly characterize many distributions on alphabets by the indices $\{\tau_n, n \geq 1\}$ into three domains, Domain 0 (no tails), Domain 1 (thin tails), and Domain 2 (thick tails).

Definition 1. *A distribution $P = \{p_k\}$ on $\mathscr{X}$ is said to belong to*:

1. *Domain* 0 *if* $\lim_{n\to\infty} \tau_n = 0$,
2. *Domain* 1 *if* $\limsup_{n\to\infty} \tau_n = c_P$ *for some constant* $c_P > 0$,
3. *Domain* 2 *if* $\lim_{n\to\infty} \tau_n = \infty$, *and*
4. *Domain T, or Domain Transient*, *if it does not belong to Domains* 0, 1, *or* 2.

The four domains so defined above form a partition of $\mathscr{P}$. The primary results established in this paper include:

1. Domain 0 does and only does include probability distributions with positive probabilities on a finite subset of $\mathscr{X}$.
2. Domain 1 includes distributions with thin tails such as $p_k \propto a^{-\lambda k}$, $p_k \propto a^{-\lambda k^2}$, and $p_k \propto k^r a^{-\lambda k}$ where $a > 1$, $\lambda > 0$ and $r \in (-\infty, \infty)$.
3. Domain 2 includes distributions with thick tails such as $p_k \propto k^{-\lambda}$ and $p_k \propto (k \ln^\lambda k)^{-1}$ where $\lambda > 1$.
4. A relative regularity condition between two distributions (one dominates the other) is defined. Under this condition, all distributions on a countably infinite alphabet, that are dominated by a Domain 1 distribution, must also belong to Domain 1.
5. Domain T is not empty.

The secondary results established in this paper include:

1. In Domain 0, $\tau_n \to 0$ exponentially fast for every distribution.
2. The tail index τ_n of a distribution with tail $p_k \propto e^{-\lambda k}$ where $\lambda > 0$ in Domain 1 perpetually oscillates between two positive constants and does not have a limit as $n \to \infty$.
3. There is a uniform positive lower bound for $\limsup_{n\to\infty} \tau_n$ for all distributions with positive probabilities on infinitely many letters of $\mathscr{X}$.

All above mentioned results are given in Section 2. Section 3 includes several constructed examples, each of which illustrates a point of interest. Section 4 includes a brief discussion on the statistical implication of the established results, along with several other remarks. The paper ends with the Appendix where several lengthy proofs are found.

2. Main results

Let K be the effective cardinality, or simply the cardinality when there is no ambiguity, of $\mathscr{X}$, i.e., $K = \sum_k 1[p_k > 0]$.

Lemma 1. *If $K = \infty$, then there exist a constant $c > 0$ and a subsequence $\{n_k; k \geq 1\}$ in $\mathbb{N}$, satisfying $n_k \to \infty$ as $k \to \infty$, such that $\tau_{n_k} > c$ for all sufficiently large k.*

A proof of Lemma 1 is given in the Appendix.

Theorem 1. *$K < \infty$ if and only if*

$$\lim_{n\to\infty} \tau_n = 0. \tag{2}$$

Proof. Assuming that $P = \{p_k; 1 \leq k \leq K\}$ where K is finite and $p_k > 0$ for all k, $1 \leq k \leq K$, and denoting $p_0 = \min\{p_k; 1 \leq k \leq K\} > 0$, the necessity of (2) follows the fact that as $n \to \infty$

$$\tau_n = n \sum_k^K p_k (1 - p_k)^n \leq n \sum_k^K p_k (1 - p_0)^n = n(1 - p_0)^n \to 0.$$

The sufficiency of (2) follows the fact that, if $K = \infty$, then Lemma 1 would provide a contradiction to (2). □

In fact the proof of Theorem 1 also establishes the following corollary.

Corollary 1. *$K < \infty$ if and only if $\tau_n \leq \mathcal{O}(nq_0^n)$ where q_0 is a constant in $(0, 1)$.*

Theorem 1 and Corollary 1 firmly characterize Domain 0 as a family of distributions on finite alphabets. All distributions outside of Domain 0 must have positive probabilities on infinitely many letters of $\mathscr{X}$. The entire class of such distributions is denoted as $\mathscr{P}_+$. In fact in the subsequent text when there is no ambiguity $\mathscr{P}_+$ will denote the entire class of distributions with a positive probability on every ℓ_k in $\mathscr{X}$. For all distributions in $\mathscr{P}_+$, a natural group would be those for which $\lim_n \tau_n = \infty$ and so Domain 2 is defined.

The following three lemmas are useful in the proof of Theorem 2 below which puts distributions with a power decaying or a slower tail in Domain 2. Lemma 2 is a version of the well-known Euler–Maclaurin formula and therefore is referred to as the Euler–Maclaurin lemma subsequently.

Lemma 2 (Euler–Maclaurin). *Let $f_n(x)$ be a continuous function of x on $[x_0, \infty)$ where x_0 is a positive integer. Suppose $f_n(x)$ is increasing on $[x_0, x(n)]$ and decreasing on $[x(n), \infty)$. If $f_n(x_0) \to 0$ and $f_n(x(n)) \to 0$ as $n \to \infty$, then*

$$\lim_{n\to\infty} \sum_{k\geq x_0} f_n(k) = \lim_{n\to\infty} \int_{x_0}^{\infty} f_n(x)\, dx.$$

A proof of Lemma 2 is given in the Appendix.

The next lemma includes three trivial but useful facts.

Lemma 3. 1. *For any real number* $p \in [0, 1)$, $1 - p \geq \exp(-\frac{p}{1-p})$.

2. *For any real number* $p \in (0, 1/2)$, $\frac{1}{1-p} < 1 + 2p$.

3. *For any real number* $p \in [0, 1]$, *the function* $f_n(p) = np(1-p)^n$ *increases in* p *over the interval* $[0, 1/(n+1)]$, *attains its maximum value of* $[1 - 1/(n+1)]^{n+1}$ *at* $p = 1/(n+1)$, *and decreases over the interval* $[1/(n+1), 1]$.

A proof of Lemma 3 is given in the Appendix.

Lemma 4. *For any given probability distribution* $P = \{p_k; k \geq 1\}$, *as* $n \to \infty$,

$$n^{1-\delta} \sum_{k \geq 1} p_k (1 - p_k)^n \to c > 0$$

for some constants $c > 0$ *and* $\delta \in (0, 1)$, *if and only if*

$$n^{1-\delta} \sum_{k \geq 1} p_k e^{-np_k} \to c > 0.$$

A proof of Lemma 4 is given in the Appendix.

Theorem 2. *For any given probability distribution* $P = \{p_k; k \geq 1\}$, *if there exists constants* $\lambda > 1$, $c > 0$ *and integer* $k_0 \geq 1$ *such that for all* $k \geq k_0$

$$p_k \geq ck^{-\lambda}, \tag{3}$$

then $\lim_{n\to\infty} \tau_n = \infty$.

A proof of Theorem 2 is given in the Appendix.

Theorem 2 puts distributions with power decaying tails, for example $p_k = c_\lambda k^{-\lambda}$, and those with slower decaying tails, for example $p_k = c_\lambda (k \ln^\lambda k)^{-1}$, where $\lambda > 1$ and $c_\lambda > 0$ is a constant which may depend on λ, in Domain 2.

In view of Lemma 1, and Theorems 1 and 2, Domain 1 has a more intuitive definition as given in the following lemma, the proof of which is trivial.

Lemma 5. *A distribution* P *on* $\mathscr{X}$ *belongs to Domain* 1 *if and only if* (1) *the effective cardinality of* $\mathscr{X}$ *is* $K = \infty$, *and* (2) $\tau_n \leq u_P$ *for all* n *and some constant* $u_P > 0$ *which may depend on* P.

The next lemma identifies an important member of Domain 1. An outline of its proof was provided by Professor Stanislav A. Molchanov of the University of North Carolina at Charlotte.

Lemma 6. *For any* $P = \{p_k\} \in \mathscr{P}_+$, *if there exists an integer* $k_0 \geq 1$ *such that* $p_k = c_0 e^{-k}$ *for all* $k \geq k_0$ *where* $c_0 > 0$ *is a constant, then*:

1. *$\tau_n \le u$ for some upper bound $u > 0$; and*
2. *$\lim_{n\to\infty} \tau_n$ does not exist.*

A proof of Lemma 6 is given in the Appendix.

A similar proof to that of Lemma 6 immediately gives Theorem 3 below with a slightly more general statement.

Theorem 3. *For any given probability distribution $P = \{p_k; k \ge 1\}$, if there exists constants $a > 1$ and integer $k_0 \ge 1$ such that for all $k \ge k_0$*

$$p_k = ca^{-k}, \tag{4}$$

then

1. *$\tau_n \le u_a$ for some upper bound $u_a > 0$ which may depend on a; and*
2. *$\lim_{n\to\infty} \tau_n$ does not exist.*

Theorem 3 puts distributions with tails of geometric progression, for examples $p_k = c_\lambda e^{-\lambda k}$ where $\lambda > 0$ and $c_\lambda > 0$ are constants and $p_k = 2^{-k}$, in Domain 1.

Next, we develop a notion of relative dominance of one probability distribution over another on a countable alphabet within $\mathscr{P}_+$. Let $\#A$ denote the cardinality of a set A.

Definition 2. *Let $Q^* \in \mathscr{P}_+$ and $P \in \mathscr{P}_+$ be two distributions on $\mathscr{X}$, and let $Q = \{q_k\}$ be a non-increasingly ordered version of Q^*. Q^* is said to dominate P if*

$$\#\left\{i;\, p_i \in (q_{k+1}, q_k], i \ge 1\right\} \le M < \infty$$

for every $k \ge 1$, where M is a finite positive integer.

It is easy to see that the notion of dominance by Definition 2 is a tail property, and that it is transitive, i.e., if P_1 dominates P_2 and P_2 dominates P_3 then P_1 dominates P_3. It says in essence that, if P is dominated by Q, then the p_is do not get overly congregated locally into some intervals defined by the q_ks.

The following examples shed a bit of intuitive light on the notion of dominance by Definition 2.

Example 1. Let $p_k = c_1 e^{-k^2}$ and $q_k = c_2 e^{-k}$ for all $k \ge k_0$ for some integer $k_0 \ge 1$ and other two constants $c_1 > 0$ and $c_2 > 0$. For every sufficiently large k, suppose $p_j = c_1 e^{-j^2} \le q_k = c_2 e^{-k}$, then $-j^2 \le \ln(c_2/c_1) - k$ and $j + 1 \ge [k + \ln(c_1/c_2)]^{1/2} + 1$. It follows that

$$\begin{aligned} p_{j+1} &= c_1 e^{-(j+1)^2} \le c_1 e^{-(\sqrt{k+\ln(c_1/c_2)}+1)^2} = c_1 e^{-(k+\ln(c_1/c_2)+1)-2\sqrt{k+\ln(c_1/c_2)}} \\ &= c_2 e^{-(k+1)-2\sqrt{k+\ln(c_1/c_2)}} = c_2 e^{-(k+1)} e^{-2\sqrt{k+\ln(c_1/c_2)}} \le c_2 e^{-(k+1)} = q_{k+1}. \end{aligned}$$

This means that if $p_j \in (q_{k+1}, q_k]$ then necessarily $p_{j+1} \notin (q_{k+1}, q_k]$, which implies that each interval $(q_{k+1}, q_k]$ can contain only one p_j at most for a sufficiently large k, that is, $k \ge k_{00} := \max\{k_0, \ln(c_2/c_1)\}$. Since there are only finite p_js covered by $\bigcup_{1\le k<k_{00}} (q_k, q_{k+1}]$, $Q = \{q_k\}$ dominates $P = \{p_i\}$.

Example 2. Let $p_k = c_1 a^{-k}$ and $q_k = c_2 b^{-k}$ for all $k \geq k_0$ for some integer $k_0 \geq 1$ and other two constants $a > b > 1$. For every sufficiently large k, suppose $p_j = c_1 a^{-j} \leq q_k = c_2 b^{-k}$, then $-j \ln a \leq \ln(c_2/c_1) - k \ln b$ and $j + 1 \geq k(\ln b/\ln a) + 1 + \ln(c_1/c_2)/\ln a$. It follows that

$$\begin{aligned} p_{j+1} &= c_1 a^{-(k(\ln b/\ln a)+1+(\ln(c_1/c_2)/\ln a))} = c_1 a^{-(k \log_a b+1+(\ln(c_1/c_2)/\ln a))} \\ &= c_1 b^{-k} a^{-1} a^{-(\ln(c_1/c_2)/\ln a)} \leq c_1 b^{-(k+1)} a^{-\log_a(c_1/c_2)} = c_2 b^{-(k+1)} = q_{k+1}. \end{aligned}$$

By a similar argument as that in Example 1, $Q = \{q_k\}$ dominates $P = \{p_i\}$.

Example 3. Let $p_k = c_1 k^{-r} e^{-\lambda k}$ for some integer $k_0 \geq 1$ and constants $\lambda > 0$ and $r > 0$, and $q_k = c_2 e^{-\lambda k}$ for all $k \geq k_0$. Suppose for a $k \geq k_0$ there is a j such that $p_j = c_1 j^{-r} e^{-\lambda j} \in (q_{k+1} = c_2 e^{-\lambda(k+1)}, q_k = c_2 e^{-\lambda k}]$, then

$$\begin{aligned} p_{j+1} &= c_1 (j+1)^{-r} e^{-\lambda(j+1)} = c_1 (j+1)^{-r} e^{-\lambda j} e^{-\lambda} \leq c_1 j^{-r} e^{-\lambda j} e^{-\lambda} \\ &\leq c_2 e^{-\lambda k} e^{-\lambda} = q_{k+1}, \end{aligned}$$

which implies that there is at most one p_j in $(q_{k+1}, q_k]$ for every sufficiently large k. Therefore $Q = \{q_k\}$ dominates $P = \{p_i\}$.

Example 4. Let $p_k = c_1 k^r e^{-\lambda k}$ for some integer $k_0 \geq 1$ and constants $\lambda > 0$ and $r > 0$, and $q_k = c_2 e^{-(\lambda/2)k}$ for all $k \geq k_0$. Suppose for any sufficiently large j, $j \geq j_0 := [e^{\lambda/(2r)} - 1]^{-1}$, we have $p_j = c_1 j^r e^{-\lambda j} \in (q_{k+1} = c_2 e^{-(\lambda/2)(k+1)}, q_k = c_2 e^{-(\lambda/2)k}]$ for some sufficiently large $k \geq k_0$, then

$$\begin{aligned} p_{j+1} &= c_1 (j+1)^r e^{-\lambda(j+1)} = c_1 (j+1)^r e^{-\lambda j} e^{-\lambda} = c_1 j^r e^{-\lambda j} e^{-\lambda} \frac{(j+1)^r}{j^r} \\ &\leq c_2 e^{-(\lambda/2)k} e^{-\lambda} \left(\frac{j+1}{j}\right)^r = c_2 e^{-(\lambda/2)(k+1)} e^{-\lambda/2} \left(\frac{j+1}{j}\right)^r \\ &\leq q_{k+1} e^{-\lambda/2} \left(\frac{j_0+1}{j_0}\right)^r = q_{k+1}, \end{aligned}$$

which implies that there is at most one p_j in $(q_{k+1}, q_k]$ for every sufficiently large k. Therefore, $Q = \{q_k\}$ dominates $P = \{p_i\}$.

Example 5. Let $p_k = q_k$ for all $k \geq 1$. $Q = \{q_k\}$ and $P = \{p_k\}$ dominate each other.

While in each of Examples 1 through 4, the dominating distribution Q has a thicker tail than P in the usual sense, the dominance of Definition 2 in general is not implied by such a thinner/thicker tail relationship. This is so because a distribution $P \in \mathscr{P}_+$, satisfying $p_k \leq q_k$ for all sufficiently large k, could exist yet congregate irregularly to have an unbounded $\sup_{k \geq 1} \#\{p_i; p_i \in (q_{k+1}, q_k], i \geq 1\}$. One such example is given in Section 3 below. In this regard, the dominance of Definition 2 is more appropriately considered as a regularity condition. However, it may be interesting to note that the said regularity is a relative one in the sense that the behavior of P is regulated by a reference distribution Q. This relative regularity gives an umbrella structure in Domain 1 as demonstrated by the theorem below.

Theorem 4. *If two distributions P and Q in $\mathscr{P}_+$ on a same countably infinite alphabet $\mathscr{X}$ are such that Q is in Domain 1 and P is dominated by Q, then P belongs to Domain 1.*

Proof. Without loss of generality, it may be assumed that Q is non-increasingly ordered. For every n, there exists a k_n such that $\frac{1}{n+1} \in (q_{k_n+1}, q_{k_n}]$. Noting part 3 of Lemma 3, consider

$$\begin{aligned}
\tau_n(P) &= \sum_{k\geq 1} np_k(1-p_k)^n \\
&= \sum_{k;\, p_k\leq q_{k_n+1}} np_k(1-p_k)^n + \sum_{k; q_{k_n+1}<p_k\leq q_{k_n}} np_k(1-p_k)^n + \sum_{k;\, p_k>q_{k_n}} np_k(1-p_k)^n \\
&\leq M \sum_{k\geq k_n+1} nq_k(1-q_k)^n + \sum_{k; q_{k_n+1}<p_k\leq q_{k_n}} e^{-1} + M \sum_{1\leq k\leq k_n} nq_k(1-q_k)^n \\
&= M \sum_{k\geq 1} nq_k(1-q_k)^n + \sum_{k; q_{k_n+1}<p_k\leq q_{k_n}} e^{-1} \\
&\leq M\tau_n(Q) + Me^{-1} < \infty.
\end{aligned}$$

The desired result immediately follows. □

Corollary 2. *Any distribution P on a countably infinite alphabet $\mathscr{X}$ satisfying $p_k = ae^{-\lambda k}$, $p_k = be^{-\lambda k^2}$, or $p_k = ck^r e^{-\lambda k}$ for all $k \geq k_0$, where $k_0 \geq 1$, $\lambda > 0$, $r \in (-\infty, +\infty)$, $a > 0$, $b > 0$ and $c > 0$ are constants, is in Domain 1.*

Proof. The result is immediate following Theorem 4 and Examples 1 through 4. □

3. Constructed examples

The first constructed example shows that the notion of thinner tail, in the sense of $p_k \leq q_k$ for $k \geq k_0$ where $k_0 \geq 1$ is some fixed integer and $P = \{p_k\}$ and $Q = \{q_k\}$ are two distributions, does not imply a dominance of Q over P.

Example 6. Consider any strictly decreasing distribution $Q = \{q_k; k \geq 1\} \in \mathscr{P}_+$ and the following grouping of the index set $\{k; k \geq 1\}$.

$$G_1 = \{1\}, \qquad G_2 = \{2, 3\}, \qquad \dots, \qquad G_m = \{m(m-1)/2 + 1, \dots, m(m-1)/2 + m\}, \dots.$$

$\{G_m; m \geq 1\}$ is a partition of the index set $\{k; k \geq 1\}$ and each group G_m contains m consecutive indices. A new distribution $P = \{p_k\}$ is constructed according to the following steps:

1. For each $m \geq 2$, let $p_k = q_{m(m-1)/2+m}$ for all $k \in G_m$.
2. $p_1 = 1 - \sum_{k\geq 2} p_k$.

In the first step, $m(m-1)/2 + m = m(m+1)/2$ is the largest index in G_m and therefore $q_{m(m+1)/2}$ is the smallest q_k with index $k \in G_m$. Since

$$0 \leq \sum_{k\geq 2} p_k = \sum_{m\geq 2} m q_{m(m+1)/2} < \sum_{k\geq 2} q_k \leq 1,$$

p_1 so assigned is a probability. The distribution $P = \{p_k\}$ satisfies $p_k \leq q_k$ for every $k \geq 2 = k_0$. However the number of terms of p_i in the interval $(q_{m(m+1)/2+1}, q_{m(m+1)/2}]$ is at least m and it increases indefinitely as $m \to \infty$; and hence Q does not dominate P.

The second constructed example shows that the notion of the dominance of $Q = \{q_k\}$ over $P = \{p_k\}$, as defined in Definition 2, does not imply that P has thinner tail than Q, in the sense of $p_k \leq q_k$ for $k \geq k_0$ where $k_0 \geq 1$ is some fixed integer.

Example 7. Consider any strictly decreasing distribution $Q = \{q_k; k \geq 1\} \in \mathscr{P}_+$ and the following grouping of the index set $\{k; k \geq 1\}$

$$G_1 = \{1, 2\}, \qquad G_2 = \{3, 4\}, \qquad \ldots, \qquad G_m = \{2m-1, 2m\}, \ldots.$$

$\{G_m; m \geq 1\}$ is a partition of the index set $\{k; k \geq 1\}$ and each group G_m contains 2 consecutive indices, the first one odd and the second one even. The construction of a new distribution $P = \{p_k\}$ is as follows: for each group G_m with its two indices $k = 2m-1$ and $k+1 = 2m$, let $p_k = p_{k+1} = (q_k + q_{k+1})/2$. With the new distribution $P = \{p_k\}$ so defined, we have $p_{2m} < q_{2m}$ and $p_{2m-1} > q_{2m-1}$ for all $m \geq 1$. Clearly Q dominates P (P dominates Q as well), but P does not have a thinner tail in the usual sense.

At this point, it becomes clear that the notation of dominance of Definition 2 and the notation of thinner/thicker tail in the usual sense are two independent notions.

The next constructed example below shows that there exists a distribution such that the associated τ_n approaches infinity along one subsequence of n and is bounded above along another subsequence of n, hence belonging to Domain T. Domain T is not empty.

Example 8. Consider the probability sequence $q_j = 2^{-j}$, for $j = 1, 2, \ldots$, along with a diffusion sequence $d_i = 2^i$, for $i = 1, 2, \ldots$. A probability sequence $\{p_k\}$, for $k = 1, 2, \ldots$, is constructed by the following steps:

1st: (a) Take the first value of d_i, $d_1 = 2^1$, and assign the first $2d_1 = 2^2 = 4$ terms of q_j, $q_1 = 2^{-1}, q_2 = 2^{-2}, q_3 = 2^{-3}, q_4 = 2^{-4}$, to the first 4 terms of p_k, $p_1 = 2^{-1}$, $p_2 = 2^{-2}$, $p_3 = 2^{-3}$, $p_4 = 2^{-4}$.

(b) Take the next unassigned term in q_j, $q_5 = 2^{-5}$, and diffuse it into $d_1 = 2$ equal terms, 2^{-6} and 2^{-6}.

(i) Starting at q_5 in the sequence $\{q_j\}$, look forwardly ($j > 5$) for terms greater or equal to 2^{-6}, if any, continue to assign them to p_k. In this case, there is only one such term $q_6 = 2^{-6}$ and it is assigned to $p_5 = 2^{-6}$.

(ii) Take the $d_1 = 2$ diffused terms and assign them to $p_6 = 2^{-6}$ and $p_7 = 2^{-6}$. At this point, the first few terms of the partially assigned sequence $\{p_k\}$ are

$$\begin{aligned} &p_1 = 2^{-1}, \quad &&p_2 = 2^{-2}, \quad &&p_3 = 2^{-3}, \\ &p_4 = 2^{-4}, \quad &&p_5 = 2^{-6}, \quad &&p_6 = 2^{-6}, \quad &&p_7 = 2^{-6}. \end{aligned}$$

2nd: (a) Take the next value of d_i, $d_2 = 2^2$, and assign the next $2d_2 = 2^3 = 8$ unused terms of q_j, $q_7 = 2^{-7}, \ldots, q_{14} = 2^{-14}$, to the next 8 terms of p_k, $p_8 = 2^{-7}, \ldots, p_{15} = 2^{-14}$.

(b) Take the next unassigned term in q_j, $q_{15} = 2^{-15}$, and diffuse it into $d_2 = 4$ equal terms of 2^{-17} each.

(i) Starting at q_{15} in the sequence of $\{q_j\}$, look forwardly ($j > 15$) for terms greater or equal to 2^{-17}, if any, continue to assign them to p_k. In this case, there are 2 such terms $q_{16} = 2^{-16}$ and $q_{17} = 2^{-17}$, and they are assigned to $p_{16} = 2^{-16}$ and $p_{17} = 2^{-17}$.

(ii) Take the $d_2 = 2^2 = 4$ diffused terms and assign them to $p_{18} = 2^{-17}, \ldots, p_{21} = 2^{-17}$. At this point, the first few terms of the partially assigned sequence $\{p_k\}$ are

$$\begin{aligned} &p_1 = 2^{-1}, \quad &&p_2 = 2^{-2}, \quad &&p_3 = 2^{-3}, \quad &&p_4 = 2^{-4}, \\ &p_5 = 2^{-6}, \quad &&p_6 = 2^{-6}, \quad &&p_7 = 2^{-6}, \\ &p_8 = 2^{-7}, \quad &&p_9 = 2^{-8}, \quad &&\ldots, \quad &&p_{15} = 2^{-14}, \quad &&p_{16} = 2^{-16}, \\ &p_{17} = 2^{-17}, \quad &&p_{18} = 2^{-17}, \quad &&\ldots, \quad &&p_{21} = 2^{-17}. \end{aligned}$$

ith: (a) In general, take the next value of d_i, say $d_i = 2^i$, and assign the next $2d_i = 2^{i+1}$ unused terms of q_j, say $q_{j_0} = 2^{-j_0}, \ldots, q_{j_0+2^{i+1}-1} = 2^{-(j_0+2^{i+1}-1)}$, to the next $2d_i = 2^{i+1}$ terms of p_k, say $p_{k_0} = 2^{-j_0}, \ldots, p_{k_0+2^{i+1}-1} = 2^{-(j_0+2^{i+1}-1)}$.

(b) Take the next unassigned term in q_j, $q_{j_0+2^{i+1}} = 2^{-(j_0+2^{i+1})}$, and diffuse it into $d_i = 2^i$ equal terms, $2^{-(j_0+i+2^{i+1})}$ each.

(i) Starting at $q_{j_0+2^{i+1}}$ in the sequence of $\{q_j\}$, look forwardly ($j > j_0 + 2^{i+1}$) for terms greater or equal to $2^{-(j_0+i+2^{i+1})}$, if any, continue to assign them to p_k. Denote the last assigned p_k as p_{k_0}.

(ii) Take the $d_i = 2^i$ diffused terms and assign them to $p_{k_0+1} = 2^{-(j_0+i+2^{i+1})}, \ldots, p_{k_0+2^i} = 2^{-(j_0+i+2^{i+1})}$.

In essence, the sequence $\{p_k\}$ is generated based on the sequence $\{q_j\}$ with infinitely many selected j's at each of which q_j is diffused into increasingly many equal probability terms according a diffusion sequence $\{d_i\}$. The diffused sequence is then rearranged in a non-increasing order. By construction, it is clear that the sequence $\{p_k; k \geq 1\}$, satisfies the following properties:

$\mathcal{A}_1$: $\{p_k\}$ is a probability sequence in a non-increasing order.

$\mathcal{A}_2$: As k increases, $\{p_k\}$ is a string of segments alternating between two different types: (1) a strictly decreasing segment and (2) a segment (a run) of equal probabilities.

$\mathcal{A}_3$: As k increases, the length of the last run increases and approaches infinity.

$\mathcal{A}_4$: In each run, there are exactly $d_i + 1$ equal terms, d_i of which are diffused terms and 1 of which belongs to the original sequence q_j.

$\mathcal{A}_5$: Between two consecutive runs (with lengths $d_i + 1$ and $d_{i+1} + 1$ respectively), the strictly decreasing segment in the middle has at least $2d_{i+1} = 4d_i = d_i + 3d_i > d_i + d_{i+1}$ terms.

$\mathcal{A}_6$: For any k, $1/p_k$ is a positive integer.

Next, we want to show that there is a subsequence $\{n_i\} \in \mathbb{N}$ such that τ_{n_i} defined with $\{p_k\}$ approaches infinity. Toward that end, consider the subsequence $\{p_{k_i}; i \geq 1\}$ of $\{p_k\}$ where the index k_i is such that p_{k_i} is first term in the ith run segment. Let $\{n_i\} = \{1/p_{k_i}\}$ which by $\mathcal{A}_6$ is a subsequence of $\mathbb{N}$. By $\mathcal{A}_3$ and $\mathcal{A}_4$,

$$\begin{aligned}\tau_{n_i} &= n_i \sum_{k \geq 1} p_k (1 - p_k)^{n_i} \\ &> n_i (d_i + 1) p_{k_i} (1 - p_{k_i})^{n_i} \\ &= (d_i + 1)\left(1 - \frac{1}{n_i}\right)^{n_i} \to \infty.\end{aligned}$$

Consider next the subsequence $\{p_{k_i-(d_i+1)}; i \geq 1\}$ of $\{p_k\}$ where the index k_i is such that p_{k_i} is first term in the ith run segment, and therefore $p_{k_i-(d_i+1)}$ is the $(d_i + 1)$th term counting backwards from p_{k_i-1}, into the preceding segment of at least $2d_i$ strictly decreasing terms. Let $\{m_i\} = \{1/p_{k_i-(d_i+1)} - 1\}$ (so $p_{k_i-(d_i+1)} = (m_i + 1)^{-1}$) which by $\mathcal{A}_6$ is a subsequence of $\mathbb{N}$

$$\begin{aligned}\tau_{m_i} &= m_i \sum_{k \geq 1} p_k (1 - p_k)^{m_i} \\ &= m_i \sum_{k \leq k_i-(d_i+1)} p_k (1 - p_k)^{m_i} + m_i \sum_{k \geq k_i - d_i} p_k (1 - p_k)^{m_i} := t_{m_i,1} + t_{m_i,2}.\end{aligned}$$

Before proceeding further, let us note several detailed facts. First, part 3 of Lemma 3. Second, since $p_{k_i-(d_i+1)} = (m_i + 1)^{-1}$, by $\mathcal{A}_1$ each summand in $\tau_{m_i,1}$ is bounded above by $m_i p_{k_i-(d_i+1)} (1 - p_{k_i-(d_i+1)})^{m_i}$ and each summand in $\tau_{m_i,2}$ is bounded above by $m_i p_{k_i-d_i} (1 - p_{k_i-d_i})^{m_i}$. Third, by $\mathcal{A}_4$ and $\mathcal{A}_5$, for each diffused term of $p_{k'}$ with $k' \leq k_i - (d_i + 1)$ in a run there is a different non-diffused term $p_{k''}$ with $k'' \leq k_i - (d_i + 1)$ such that $p_{k'} > p_{k''}$ and therefore $m_i p_{k'} (1 - p_{k'})^{m_i} \leq m_i p_{k''} (1 - p_{k''})^{m_i}$; and similarly, for each diffused term of $p_{k'}$ with $k' \geq k_i - d_i$ in a run there is a different non-diffused term $p_{k''}$ with $k'' \geq k_i - d_i$ such that $p_{k'} < p_{k''}$ and therefore $m_i p_{k'} (1 - p_{k'})^{m_i} \leq m_i p_{k''} (1 - p_{k''})^{m_i}$. These facts imply that

$$\begin{aligned}\tau_{m_i} &= \tau_{m_i,1} + \tau_{m_i,2} \\ &= m_i \sum_{k \leq k_i-(d_i+1)} p_k (1 - p_k)^{m_i} + m_i \sum_{k \geq k_i - d_i} p_k (1 - p_k)^{m_i} \\ &\leq 2m_i \sum_{j \geq 1} q_j (1 - q_j)^{m_i} < \infty\end{aligned}$$

and the last inequality above is due to Corollary 2.

4. Concluding remarks

While the domains of attraction on alphabets have probabilistic merits, the statistical implication is also quite significant. Zhang and Zhou [11] showed that $\zeta_{1,v}$ is estimable (there exists at least one unbiased estimator of $\zeta_{1,v}$), and established an unbiased estimator of $\zeta_{1,v}$ for every $v \leq n-1$. Their estimator is

$$Z_{1,v} = \frac{n^{1+v}[n-(1+v)]!}{n!} \sum_{k\geq 1} \left[\hat{p}_k \prod_{j=0}^{v-1} \left(1 - \hat{p}_k - \frac{j}{n} \right) \right]. \tag{5}$$

Therefore there readily exists an unbiased estimator of τ_v for every $v \leq n-1$ namely

$$t_v = v Z_{1,v}. \tag{6}$$

Zhang and Zhou [11] also established several useful statistical properties of t_v, including the asymptotic normality and that t_v is the uniformly minimum variance unbiased estimator (umvue) when $K < \infty$.

The availability of t_v gives much added value to the discussion of the domains of attraction on alphabets as presented in this paper. Specifically the fact that the asymptotic behavior of τ_n characterizes the tail probability of the underlying P and the fact that the trajectory of τ_v up to $v = n-1$ is estimable suggest that much could be revealed by a sufficiently large sample.

The three main domains of attraction identified in this paper bear a strong parallelism to Fréchet, Gumbel and Weibull families, for continuous random variables. However the theory in this paper is about general random elements on alphabets. While the theory can be applied to sets of discrete and countable points on the real line as special cases, where the parallelism to Fréchet, Gumbel and Weibull families materializes, they are only special cases. In this broader perspective, it would not seem entirely suitable to identify the three main domains (D_2, D_1, D_0) as Fréchet, Gumbel and Weibull families.

The work reported in this paper is inspired by a long line of researchers going back to as far as Gini [4] and Simpson [7]. The well-known Gini–Simpson diversity index,

$$\zeta_{1,1} = \sum_{k\geq 1} p_k(1-p_k),$$

clearly bears resemblance to the tail index τ_n. In fact, when $n = 1$, $\tau_n = \zeta_{1,1}$. Inspired by Turing's formula introduced by Good [6] (therefore also known as the Good–Turing formula), a weighted version of $\zeta_{1,1}$, in the form of $\zeta_{u,v} = \sum_{k\geq 1} p_k^u (1-p_k)^v$ for any positive integer u and non-negative integer v was introduced by Zhang and Zhou [11]. Zhang and Grabchak [10] showed that $\{\zeta_{1,v}; v \geq 1\}$ and $\{p_k; k \geq 1\}$ uniquely determine each other up to a permutation on the index set, $\{k; k \geq 1\}$. This paper shows that, along the sequence $\{\tau_v = v\zeta_{1,v}; v \geq 1\}$, the distributions of $\mathscr{X}$ splinter into domains of attraction. The term "a domain of attraction on alphabet" was first used by Professor Stanislav A. Molchanov in a private conversation with the author after a seminar on asymptotic normal laws for Turing's formula given by the author in 2008. Professor Molchanov pointed out that the normal law condition on $\{p_k\}$ for Turing's formula constituted a domain of attraction. In retrospect, that was the origin of the idea that grew into this paper.

To honor the great minds of mathematics whose works marked the trail leading to this paper, the author wishes to have the three domains of attraction defined in this paper to be identified as the Gini–Simpson family (for Domain 0), the Molchanov family (for Domain 1), and the Turing–Good family (for Domain 2).

Appendix

Proof of Lemma 1. Let us assume without loss of generality that $p_k > 0$ for all $k \geq 1$. Since $\zeta_{1,n}$ is invariant with respect to any permutation on the index set $\{k; k \geq 1\}$, it can be assumed without loss of generality that $\{p_k\}$ is non-increasing in k. For every k, let $n_k = \lfloor 1/p_k \rfloor$. With n_k so defined, we have $1/(n_k+1) < p_k \leq 1/n_k$ for every k and $\lim_{k\to\infty} n_k = \infty$ though $\{n_k\}$ may not necessarily be strictly increasing. By construction, the following are true about the n_k, $k \geq 1$:

1. $\{n_k; k \geq 1\}$ is an infinite subset of $\mathbb{N}$.
2. Every p_k is covered by the interval $(1/(n_k+1), 1/n_k]$.
3. Every interval $(1/(n_k+1), 1/n_k]$ covers at least one p_k and at most finitely many p_ks.

Let $f_n(x) = nx(1-x)^n$ for $x \in [0,1]$. $f_n(x)$ attains its maximum at $x = (n+1)^{-1}$ with value

$$f_n\left(\frac{1}{n+1}\right) = \frac{n}{n+1}\left(1 - \frac{1}{n+1}\right)^n = \left(\frac{n}{n+1}\right)^{n+1} \to e^{-1}.$$

Also we have

$$f_n\left(\frac{1}{n}\right) = \left(1 - \frac{1}{n}\right)^n \to e^{-1}.$$

Furthermore since $f_n'(x) < 0$ for $(n+1)^{-1} < x < 1$, we have

$$f_n\left(\frac{1}{n}\right) < f_n(x) < f_n\left(\frac{1}{n+1}\right) \qquad \text{for } \frac{1}{n+1} < x < \frac{1}{n}.$$

For a small but fixed $\varepsilon > 0$, let $c = e^{-1} - \varepsilon$. Since $f_n(1/n) \to e^{-1}$ and $f_n(1/(n+1)) \to e^{-1}$, there exists a positive N_ε such that for any $n > N_\varepsilon$, $f_n(1/(n+1)) > f_n(1/n) > c$.

Since $\lim_{k\to\infty} n_k = \infty$ and $\{n_k\}$ is non-decreasing, there exists an integer $K_\varepsilon > 0$ such that $n_k > N_\varepsilon$ for all $k > K_\varepsilon$. Consider the sub-sequence $\{\tau_{n_k}; k \geq 1\}$. For any $k > K_\varepsilon$,

$$\tau_{n_k} = \sum_{i=1}^{\infty} n_k p_i (1-p_i)^{n_k} > f_{n_k}(p_k).$$

Since $p_k \in (1/(n_k+1), 1/n_k]$ and $f_{n_k}(x)$ is decreasing on the interval $(1/(n_k+1), 1/n_k]$, we have

$$f_{n_k}(p_k) > f_{n_k}\left(\frac{1}{n_k}\right) \geq c = e^{-1} - \varepsilon,$$

and hence $\tau_{n_k} > f_{n_k}(p_k) \geq c$ for all $k > K_\varepsilon$. □

Proof of Lemma 2. It can be verified that

$$\sum_{x_0 \le k \le x(n)} f_n(k) - f_n\big(x(n)\big) \le \int_{x_0}^{x(n)} f_n(x)\,dx \le \sum_{x_0+1 \le k < x(n)} f_n(k) + f_n\big(x(n)\big) \quad \text{and}$$

$$\sum_{k > x(n)} f_n(k) - f_n\big(x(n)\big) \le \int_{x(n)}^{\infty} f_n(x)\,dx \le \sum_{k \ge x(n)} f_n(k) + f_n\big(x(n)\big).$$

Adding the corresponding parts of the two expressions above and taking limits give

$$\begin{aligned}\lim_{n\to\infty}\sum_{k=x_0}^{\infty} f_n(k) - 2\lim_{n\to\infty} f_n\big(x(n)\big) &\le \lim_{n\to\infty}\int_{x_0}^{\infty} f_n(x)\,dx \\ &\le \lim_{n\to\infty}\sum_{k=x_0}^{\infty} f_n(k) - \lim_{n\to\infty} f_n(x_0) + 2\lim_{n\to\infty} f_n\big(x(n)\big).\end{aligned}$$

The desired result follows the conditions of the lemma. □

Proof of Lemma 3. For part 1, the function $y = \frac{1}{1+t}e^t$ is strictly increasing over $[0, \infty)$, and has value 1 at $t = 0$. Therefore, $\frac{1}{1+t}e^t \ge 1$ for $t \in [0, \infty)$. The desired inequality follows the change of variable $p = t/(1+t)$. For parts 2 and 3, the proofs are trivial. □

Proof of Lemma 4. Let $\delta^* = \delta/8$. Consider the partition of the index set $\{k; k \ge 1\} = I \cup II$ where

$$I = \big\{k;\ p_k \le 1/n^{1-\delta^*}\big\} \quad \text{and} \quad II = \big\{k;\ p_k > 1/n^{1-\delta^*}\big\}.$$

Since pe^{-np} has a negative derivative with respect to p on interval $(1/n, 1]$ and hence on $(1/n^{1-\delta^*}, 1]$ for large n, $p_k e^{-np_k}$ attains its maximum at $p_k = 1/n^{1-\delta^*}$ for every $k \in II$. Therefore noting that there are at most $n^{1-\delta^*}$ indices in II,

$$\begin{aligned}0 &\le n^{1-\delta}\sum_{II} p_k(1-p_k)^n \le n^{1-\delta}\sum_{II} p_k e^{-np_k} \\ &\le n^{1-\delta}\sum_{II}\left(\frac{1}{n^{1-\delta^*}}e^{-n/n^{1-\delta^*}}\right) \le n^{1-\delta}n^{1-\delta^*}\left(\frac{1}{n^{1-\delta^*}}e^{-n/n^{1-\delta^*}}\right) \\ &= n^{1-\delta}e^{-n^{\delta^*}} \to 0.\end{aligned}$$

Thus

$$\lim_{n\to\infty} n^{1-\delta}\sum_{k} p_k(1-p_k)^n = \lim_{n\to\infty} n^{1-\delta}\sum_{I} p_k(1-p_k)^n \tag{7}$$

and

$$\lim_{n\to\infty} n^{1-\delta}\sum_k p_k e^{-np_k} = \lim_{n\to\infty} n^{1-\delta}\sum_I p_k e^{-np_k}. \tag{8}$$

On the other hand, since $1 - p \le e^{-p}$ for all $p \in [0, 1]$,

$$n^{1-\delta}\sum_I p_k(1-p_k)^n \le n^{1-\delta}\sum_I p_k e^{-np_k}.$$

Furthermore, applying parts 1 and 2 of Lemma 3 in the first and the third steps below respectively leads to

$$\begin{aligned} n^{1-\delta}\sum_I p_k(1-p_k)^n &\ge n^{1-\delta}\sum_I p_k \exp\left(-\frac{np_k}{1-p_k}\right) \\ &\ge n^{1-\delta}\sum_I p_k \exp\left(-\frac{np_k}{1-\sup_I p_k}\right) \\ &\ge n^{1-\delta}\sum_I \exp\left(-2n\left(\sup_I p_k\right)^2\right) p_k e^{-np_k}. \end{aligned}$$

Noting the fact that $\lim_{n\to\infty}\exp(-2n(\sup_I p_k)^2) = 1$ uniformly by the definition of I,

$$\lim_{n\to\infty} n^{1-\delta}\sum_I p_k(1-p_k)^n = \lim_{n\to\infty} n^{1-\delta}\sum_I p_k e^{-np_k},$$

and hence, by (7) and (8), the lemma follows. □

Proof of Theorem 2. For clarity, the proof is given in two cases respectively:

1. $p_k = ck^{-\lambda}$ for all $k \ge k_0$ for some $k_0 > 1$, and
2. $p_k \ge ck^{-\lambda}$ for all $k \ge k_0$ for some $k_0 > 1$.

Case 1: Assuming $p_k = ck^{-\lambda}$ for all $k \ge k_0$, it suffices to consider the partial series $\sum_{k\ge k_0} np_k(1-p_k)^n$. First consider

$$n^{1-1/\lambda}\sum_{k=k_0}^{\infty} p_k e^{-np_k} = n^{1-1/\lambda}\sum_{k=k_0}^{\infty} ck^{-\lambda}e^{-nck^{-\lambda}} = \sum_{k=k_0}^{\infty} f_n(k),$$

where $f_n(x) = n^{1-1/\lambda}cx^{-\lambda}e^{-ncx^{-\lambda}}$. Since it is easily verified that

$$f_n'(x) = -\lambda cn^{1-1/\lambda}x^{-(\lambda+1)}\left(1 - ncx^{-\lambda}\right)e^{-ncx^{-\lambda}},$$

it can be seen that, $f_n(x)$ increases over $[1, (nc)^{1/\lambda}]$ and decreases over $[(nc)^{1/\lambda}, \infty)$. Let $x_0 = k_0$ and $x(n) = (nc)^{1/\lambda}$. It is clear that $f_n(x_0) \to 0$ and

$$f_n\left(x(n)\right) = n^{1-1/\lambda}c(nc)^{-1}e^{-nc(nc)^{-1}} = n^{1-1/\lambda}c(nc)^{-1}e^{-1} = \frac{1}{en^{1/\lambda}} \to 0.$$

Invoking the Euler–Maclaurin lemma, we have, with changes of variable $t = x^{-\lambda}$ and then $s = nct$,

$$\begin{aligned}
n^{1-1/\lambda} \sum_{k=k_0}^{\infty} p_k e^{-np_k} &\sim \int_{x_0}^{\infty} n^{1-1/\lambda} c x^{-\lambda} e^{-ncx^{-\lambda}}\, dx \\
&= \frac{c}{\lambda} \int_0^{x_0^{-\lambda}} n^{1-1/\lambda} t^{-1/\lambda} e^{-nct}\, dt \\
&= \frac{c}{\lambda} n^{1-1/\lambda} \int_0^{x_0^{-\lambda}} (nct)^{-1/\lambda} (nc)^{-1+1/\lambda} e^{-nct}\, d(nct) \\
&= \frac{c}{\lambda} n^{1-1/\lambda} (nc)^{-1+1/\lambda} \int_0^{ncx_0^{-\lambda}} s^{-1/\lambda} e^{-s}\, ds \\
&= \frac{c^{1/\lambda}}{\lambda} n^0 \int_0^{ncx_0^{-\lambda}} s^{-1/\lambda} e^{-s}\, ds = \frac{c^{1/\lambda}}{\lambda} \int_0^{ncx_0^{-\lambda}} s^{(1-1/\lambda)-1} e^{-s}\, ds \\
&= \frac{c^{1/\lambda}}{\lambda} \Gamma\left(1 - \frac{1}{\lambda}\right) \left[\frac{1}{\Gamma(1 - 1/\lambda)} \int_0^{ncx_0^{-\lambda}} s^{(1-1/\lambda)-1} e^{-s}\, ds \right] \\
&\to \frac{c^{1/\lambda}}{\lambda} \Gamma\left(1 - \frac{1}{\lambda}\right) > 0.
\end{aligned}$$

Hence by Lemma 4, $n^{1-1/\lambda} \sum_{k=1}^{\infty} p_k (1 - p_k)^n \to c^{1/\lambda} \lambda^{-1} \Gamma(1 - 1/\lambda) > 0$ and therefore $\tau_n \to \infty$.

Case 2: Assuming $p_k \geq ck^{-\lambda} =: q_k$ for all $k \geq k_0$ for some $k_0 \geq 1$, we observe

$$\begin{aligned}
n^{1-1/\lambda} \sum_{k \geq [(n+1)c]^{1/\lambda}} ck^{-\lambda} e^{-nck^{-\lambda}} &= \sum_{k \geq 1} \left\{ n^{1-1/\lambda} ck^{-\lambda} e^{-nck^{-\lambda}} 1\left[k \geq \left[(n+1)c\right]^{1/\lambda}\right] \right\} \\
&=: \sum_{k \geq 1} f_n(k).
\end{aligned}$$

Further let us note the following three facts: first the function $f_n^*(x) = n^{1-1/\lambda} cx^{-\lambda} e^{-ncx^{-\lambda}}$ for $x \in (1, \infty)$ increases as x increases from 1 to reach its maximized at $x^*(n) = (nc)^{1/\lambda}$ and then decreases as x increases to ∞; second $x^*(n) = (nc)^{\frac{1}{\lambda}} < [(n + 1)c]^{1/\lambda} = x(n)$; and third $f_n^*(x) \geq f_n(x)$ where $f_n(x) = f_n^*(x) 1[x \geq [(n + 1)c]^{1/\lambda}]$ is a truncated version of $f_n^*(x)$. The fact $f_n^*(x^*(n)) \to 0$ implies $f_n^*(x(n)) \to 0$, which in turn implies $f_n(x(n)) \to 0$. Since $f_n(x)$ maximizes at $x(n)$, the condition of Euler–MacLaurin lemma is satisfied. We then have

$$\begin{aligned}
&n^{1-1/\lambda} \sum_{k \geq [(n+1)c]^{1/\lambda}} ck^{-\lambda} e^{-nck^{-\lambda}} \\
&= c \int_1^{\infty} n^{1-1/\lambda} x^{-\lambda} e^{-ncx^{-\lambda}} 1\left[x \geq \left[(n+1)c\right]^{1/\lambda}\right] dx
\end{aligned}$$

$$= c\int_{[(n+1)c]^{1/\lambda}}^{\infty} n^{1-1/\lambda} x^{-\lambda} e^{-ncx^{-\lambda}}\,dx \tag{9}$$

$$= c^{1/\lambda}\lambda^{-1}\Gamma\left(1-\frac{1}{\lambda}\right)\int_0^{n(n+1)c^2}\frac{1}{\Gamma(1-1/\lambda)}s^{(1-1/\lambda)-1}e^{-s}\,ds$$

$$\to c^{1/\lambda}\lambda^{-1}\Gamma\left(1-\frac{1}{\lambda}\right) > 0.$$

On the other hand, for sufficiently large n, let $I^* = \{k;\, p_k \le \frac{1}{n+1}\} \subseteq \{k;\, k \ge k_0\}$. By parts 1 and 2 of Lemma 3 at steps 2 and 4 below and (9) at step 7, we have

$$\begin{aligned}
n^{1-1/\lambda}\sum_{k\in I^*} p_k(1-p_k)^n &\ge n^{1-1/\lambda}\sum_{k\in I^*} q_k(1-q_k)^n \\
&\ge n^{1-1/\lambda}\sum_{k\in I^*} q_k \exp\left(-\frac{nq_k}{1-q_k}\right) \\
&\ge n^{1-1/\lambda}\sum_{k\in I^*} q_k \exp\left(-\frac{nq_k}{1-\sup_{I^*} q_k}\right) \\
&\ge n^{1-1/\lambda}\sum_{k\in I^*} \exp\left(-2n\left(\sup_{I^*} q_k\right)^2\right) q_k e^{-nq_k} \\
&\ge n^{1-1/\lambda}\sum_{k\in I^*} \exp(-2/n) q_k e^{-nq_k} \\
&= \exp(-2/n) n^{1-1/\lambda}\sum_{k\in I^*} ck^{-\lambda} e^{-nck^{-\lambda}} \\
&\to c^{1/\lambda}\lambda^{-1}\Gamma\left(1-\frac{1}{\lambda}\right) > 0.
\end{aligned}$$

Finally, $\tau_n = n\sum_k p_k(1-p_k)^n \ge n^{1/\lambda} n^{1-1/\lambda}\sum_{k\in I^*} p_k(1-p_k)^n \to \infty$ as $n \to \infty$. □

Proof of Lemma 6. For clarity, the proof of Lemma 6 is given in three parts: Part 1: Preliminaries; Part 2: Part 1 of Lemma 6; and Part 3: Part 2 of Lemma 6.

Part 1: Preliminaries. Noting that the first finite terms of τ_n vanishes exponentially fast for any distribution, we may assume, without loss of generality, that $k_0 = 1$. For any given n, define $k^* = k^*(n)$ by

$$p_{k^*+1} < \frac{1}{n+1} \le p_{k^*}. \tag{10}$$

Noting

$$c_0 e^{-(k^*+1)} < \frac{1}{n+1} \le c_0 e^{-k^*},$$

$$e^{-(k^*+1)} < \frac{1}{c_0(n+1)} \le e^{-k^*},$$
$$-\big(k^*+1\big) < -\ln\big(c_0(n+1)\big) \le -k^* \quad \text{and}$$
$$k^*+1 > \ln\big(c_0(n+1)\big) \ge k^*,$$

we may write

$$k^* = \big\lfloor \ln\big(c_0(n+1)\big)\big\rfloor$$

for each n. That is to say that, although k^* is uniquely defined by any given n, each k^* may correspond to several consecutive integer values of n. For a given integer value k^*, let the said consecutive integer values of n be denoted by

$$\{n_{k^*}, n_{k^*}+1, \dots, n_{k^*+1}-1\}, \tag{11}$$

specifically noting (1) that n_{k^*} is the smallest integer value of n corresponding to k^* by (10), that is, $k^* = \lfloor \ln(c_0(n+1))\rfloor$; (2) n_{k^*+1} is the smallest integer value of n that satisfies $k^*+1 = \lfloor \ln(c_0(n+1))\rfloor$; and (3) $n_{k^*+1}-1$ is the greatest integer value of n that shares the same value of k^* with n_{k^*}.

Since $k^* = k^*(n)$ depends on n, we may express p_{k^*} as, and define $c(n)$ by,

$$p_{k^*} = \frac{c(n)}{n}. \tag{12}$$

At this point, let us observe the following fact: for each given k^*,

$$p_{k^*} = \frac{c(n_{k^*})}{n_{k^*}} = \frac{c(n_{k^*}+1)}{n_{k^*}+1} = \dots = \frac{c(n_{k^*+1}-1)}{n_{k^*+1}-1}. \tag{13}$$

There are two main consequences of the expression in (12). The first is that τ_n defined in (1) may be expressed by (15) below; and the second is that the sequence $c(n)$ perpetually oscillates between 1 and e. Both facts are demonstrated below.

Noting part 3 of Lemma 3, we have for any n

$$\begin{aligned} f_n(p_k) &\le f_n(p_{k^*}), \qquad k \le k^*, \\ f_n(p_k) &< f_n(p_{k^*}), \qquad k \ge k^*+1. \end{aligned} \tag{14}$$

For a given n, let us rewrite each p_k in terms of p_{k^*}, and therefore in terms of n and $c(n)$

$$p_{k^*+i} = e^{-i}\frac{c(n)}{n} \quad \text{and} \quad p_{k^*-j} = e^{j}\frac{c(n)}{n}$$

for all appropriate positive integers i and j. Therefore,

$$f_n(p_{k^*+i}) = ne^{-i}\frac{c(n)}{n}\left(1 - e^{-i}\frac{c(n)}{n}\right)^n = \frac{c(n)}{e^i}\left(1 - \frac{c(n)}{ne^i}\right)^n,$$
$$f_n(p_{k^*-j}) = ne^{j}\frac{c(n)}{n}\left(1 - e^{j}\frac{c(n)}{n}\right)^n = c(n)e^j\left(1 - \frac{c(n)e^j}{n}\right)^n$$

and

$$\tau_n = \sum_{k\le k^*-1} f_n(p_k) + f_n(p_{k^*}) + \sum_{k\ge k^*+1} f_n(p_k)$$
$$= c(n)\sum_{j=1}^{k^*-1} e^j\left(1-\frac{c(n)e^j}{n}\right)^n + c(n)\left(1-\frac{c(n)}{n}\right)^n + c(n)\sum_{i=1}^{\infty} e^{-i}\left(1-\frac{c(n)}{ne^i}\right)^n. \tag{15}$$

Next, we want to show that $c(n)$ oscillates perpetually over the interval $(n/(n+1), e)$ which approaches $[1, e)$ as n increases indefinitely. This is so because, since k^* is defined by (10), we have

$$\frac{c(n)}{n}e^{-1} \le \frac{1}{n+1} \le \frac{c(n)}{n}$$

or

$$e^{-1} < \frac{n}{n+1} \le c(n) \le \frac{n}{n+1}e < e. \tag{16}$$

At this point, the fact $c(n) \in [1, e)$ is established. What remains to be shown is that $c(n)$ oscillates perpetually in n. Toward that end, let us consider $k^*(n)$ as a mapping, which maps every positive integer value of $n \in \mathbb{N}$ to a positive integer value of $k^* \in \mathbb{N}$. The inverse of $k^*(n)$ maps every integer value $k^* \in \mathbb{N}$ to a set as in (11). Let

$$\mathbb{N} = \cup\{n_{k^*}, n_{k^*}+1, \ldots, n_{k^*+1}-1\}, \tag{17}$$

where the union is over all possible integer values of k^*. (In fact, the smallest k^* possible is $k^* = 1$ for $n_{k^*} = 1$. For this case, $p_1 = 1 - e^{-1} \approx 0.6321$, $p_2 = p_1 e^{-1} \approx 0.2325$, and $(1+1)^{-1} = 0.5$, therefore $k^* = 1$ and $n_1 = 1$.)

Let us make the following three observations:

1. We have $c(n_{k^*}) < c(n_{k^*}+1) < \cdots < c(n_{k^*+1}-1)$. This is so because of (13): all of them sharing the same k^* and therefore the same p_{k^*}. Furthermore, the increments of increase are all identical, namely, p_{k^*}.

2. Consider $\{n_{k*}; k^* \ge 1\}$ where n_{k*} is the smallest integer value in each partitioning set of (17). We have $c(n_{k^*}) = n_{k^*}p_{k^*} \to 1$. This is so because $1/n_{k^*} > p_{k^*} \ge 1/(n_{k^*}+1)$ or

$$1 - p_{k^*} \le n_{k^*}p_{k^*} < 1, \tag{18}$$

which implies that, $n_{k^*}p_{k^*}$ for all sufficiently large k^* (or equivalently sufficiently large n_{k^*} or sufficiently large n),

$$c(n_{k^*}) = n_{k^*}p_{k^*} \in (1-\varepsilon, 1), \tag{19}$$

where $\varepsilon > 0$ is an arbitrarily small real value.

3. Consider $\{n_{k*+1}-1; k^* \geq 1\}$ where $n_{k*+1}-1$ is the greatest integer value in each partitioning set of (17). We have $c(n_{k^*+1}-1) = (n_{k^*+1}-1)p_{k^*} \to e$. This is so because

$$p_{k^*} = p_{k^*+1}e = \frac{n_{k^*+1}-1}{n_{k^*+1}-1} p_{k^*+1}e = \frac{1}{n_{k^*+1}-1}\left(\frac{n_{k^*+1}-1}{n_{k^*+1}}\right)(n_{k^*+1}p_{k^*+1})e$$

and therefore by (18)

$$c(n_{k^*+1}-1) = \left(\frac{n_{k^*+1}-1}{n_{k^*+1}}\right)(n_{k^*+1}p_{k^*+1})e \to e.$$

At this point, it has been established that the range of $c(n)$ for $n \geq n_0$, where n_0 is any positive integer, covers the entire interval $[1, e)$.

Part 2: Part 1 of Lemma 6. Noting that $e^{-1} \leq c(n) \leq e$ (see (16)) and that $1-p \leq e^{-p}$ for all $p \in [0, 1]$, the desired result follows the argument below

$$\begin{aligned}
\tau_n &= c(n)\sum_{j=1}^{k^*-1} e^j\left(1-\frac{c(n)e^j}{n}\right)^n + c(n)\left(1-\frac{c(n)}{n}\right)^n + c(n)\sum_{j=1}^{\infty} e^{-j}\left(1-\frac{c(n)}{ne^j}\right)^n \\
&\leq e\sum_{j=1}^{k^*-1} e^j\left(1-\frac{e^{j-1}}{n}\right)^n + e\left(1-\frac{e^{-1}}{n}\right)^n + e\sum_{j=1}^{\infty} e^{-j}\left(1-\frac{1}{ne^{j+1}}\right)^n \\
&\leq e\sum_{j=1}^{k^*-1} e^j e^{-e^{j-1}} + e\sum_{j=0}^{\infty} e^{-j}e^{-e^{-(j+1)}} \\
&\leq e^2\sum_{j=1}^{k^*-1} e^{j-1}e^{-e^{j-1}} + e^2\sum_{j=0}^{\infty} e^{-(j+1)}e^{-e^{-(j+1)}} \\
&< e^2\sum_{j=0}^{\infty} e^j e^{-e^j} + e^2\sum_{j=1}^{\infty} e^{-j}e^{-e^{-j}} := u.
\end{aligned}$$

Part 3: Part 2 of Lemma 6. Consider, for any fixed $c > 0$,

$$\tau_n^* = c\sum_{j=1}^{k^*-1} e^j\left(1-\frac{ce^j}{n}\right)^n + c\left(1-\frac{c}{n}\right)^n + c\sum_{j=1}^{\infty} e^{-j}\left(1-\frac{c}{ne^j}\right)^n.$$

By Dominated Convergence theorem,

$$\tau(c) := \lim_{n\to\infty} \tau_n^* = c\sum_{j=0}^{\infty} e^j e^{-ce^j} + c\sum_{j=1}^{\infty} e^{-j}e^{-ce^{-j}},$$

and $\tau(c)$ is a non-constant function in c on $[1, e]$.

Noting that, as k^* increases (or equivalently n_{k^*} increases or n increases),

$$1 \leftarrow c(n_{k^*}) < c(n_{k^*} + 1) < \cdots < c(n_{k^*+1} - 1) \rightarrow e$$

and that the increment between two consecutive terms $p_{k^*} \rightarrow 0$, $c(n)$ visits any arbitrarily small closed interval $[a, b] \subset [1, e]$ infinitely often, and therefore there exists for each such interval a subsequence $\{n_l; l \geq 1\}$ of $\mathbb{N}$ such that $c(n_l)$ converges, that is, $c(n_l) \rightarrow \theta$ for some $\theta \in [a, b]$. Since $\tau(c)$ is a non-constant function on $[1, e]$, there exist two non-overlapping closed intervals, $[a_1, b_1]$ and $[a_2, b_2]$ in $[1, e]$, satisfying

$$\max_{a_1 \leq c \leq b_1} \tau(c) < \min_{a_2 \leq c \leq b_2} \tau(c),$$

such that there exist two sub-sequences of $\mathbb{N}$, said $\{n_l; l \geq 1\}$ and $\{n_m; m \geq 1\}$, such that $c(n_l) \rightarrow \theta_1$ for some $\theta_1 \in [a_1, b_1]$ and $c(n_m) \rightarrow \theta_2$ for some $\theta_2 \in [a_2, b_2]$.

Consider the limit of τ_n along $\{n_l; l \geq 1\}$, again by Dominated Convergence theorem,

$$\begin{aligned}
\lim_{n_l \rightarrow \infty} \tau_{n_l} &= \lim_{n_l \rightarrow \infty} \left[c(n_l) \sum_{j=0}^{k^*-1} e^j \left(1 - \frac{c(n_l) e^j}{n}\right)^n + c(n_l) \sum_{j=1}^{\infty} e^{-j} \left(1 - \frac{c(n_l)}{n e^j}\right)^n \right] \\
&= \theta_1 \sum_{j=0}^{\infty} e^j e^{-\theta_1 e^j} + \theta_1 \sum_{j=1}^{\infty} e^{-j} e^{-\theta_1 e^{-j}} = \tau(\theta_1).
\end{aligned}$$

A similar argument gives $\lim_{n_m \rightarrow \infty} \tau_{n_m} = \tau(\theta_2)$, but $\tau(\theta_1) \neq \tau(\theta_2)$ by construction, and hence $\lim_{n \rightarrow \infty} \tau_n$ does not exist. $\square$

Acknowledgement

The author wishes to thank a referee whose comments and suggestions brought significant improvement to an earlier version of this manuscript.

References

[1] de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory*: *An Introduction*. *Springer Series in Operations Research and Financial Engineering*. New York: Springer. MR2234156

[2] Fisher, R.A. and Tippett, L.H.C. (1928). Limiting forms of the frequency-distribution of the largest or smallest member of a sample. *Math*. *Proc*. *Cambridge Philos*. *Soc*. **24** 180.

[3] Fréchet, M. (1927). Sur la loi de probabilité de l'écart maximum. *Annals de la Soc*. *Polonaise*. *de Math*. **6** 92.

[4] Gini, C. (1912). *Variabilità e Mutabilità*. *Contributo Allo Studiodelle Distribuzioni e delle Relazioni Statistiche*. Bologna: C. Cuppini.

[5] Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Ann*. *of Math*. (2) **44** 423–453. MR0008655

[6] Good, I.J. (1953). The population frequencies of species and the estimation of population parameters. *Biometrika* **40** 237–264. MR0061330
[7] Simpsom, E.H. (1949). Measurement of diversity. *Nature* **163** 688.
[8] Smirnov, N.V. (1952). *Limit Distributions for the Terms of a Variational Series*. Am. Math. Soc., Providence, RI.
[9] Zhang, Z. (2012). Entropy estimation in Turing's perspective. *Neural Comput*. **24** 1368–1389. MR2952469
[10] Zhang, Z. and Grabchak, M. (2016). Entropic representation and estimation of Diversity indices. *J. Nonparametr. Stat*. **28** 563–575.
[11] Zhang, Z. and Zhou, J. (2010). Re-parameterization of multinomial distributions and diversity indices. *J. Statist. Plann. Inference* **140** 1731–1738. MR2606713

Received April 2015 and revised August 2015