

Article

# Comparing the Zeta distributions with the Pareto distributions from the viewpoint of information theory and information geometry: Discrete versus continuous exponential families of power laws

Frank Nielsen<sup>1</sup> orcid number:0000-0001-5728-0726

<sup>1</sup> Sony Computer Science Laboratories Inc., Tokyo, Japan; Frank.Nielsen@acm.org

\* Frank Nielsen

† Proceedings of the International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, IHP, Paris, July 18–22, 2022.

Version October 18, 2022 submitted to Entropy

**Abstract:** We consider the zeta distributions which are discrete power law distributions that can be interpreted as the counterparts of the continuous Pareto distributions with unit scale. The family of zeta distributions forms a discrete exponential family with normalizing constants expressed using the Riemann zeta function. We present several information-theoretic measures between zeta distributions, study their underlying information geometry, and compare the results with their continuous counterparts, the Pareto distributions.

**Keywords:** Maximum entropy; exponential family; convex duality; zeta sum; truncated exponential family; Von Mangoldt function

## 1. Introduction

The zeta distributions [1,2] are parametric discrete distributions with probability mass functions indexed by a scalar parameter  $s \in (1, \infty)$  whose support is the set of positive integers  $\mathbb{N}$ :

$$p_s(x) = \Pr[X = x] \propto \frac{1}{x^s}, \quad x \in \mathcal{X} = \mathbb{N} = \{1, 2, \dots\}. \quad (1)$$

The normalization function  $\zeta(s)$  of the zeta distributions  $p_s(x) = \frac{1}{\zeta(s)} \frac{1}{x^s}$  such that  $\sum_{x \in \mathbb{N}} p_s(x) = 1$  is the real Riemann zeta function [3–5]:

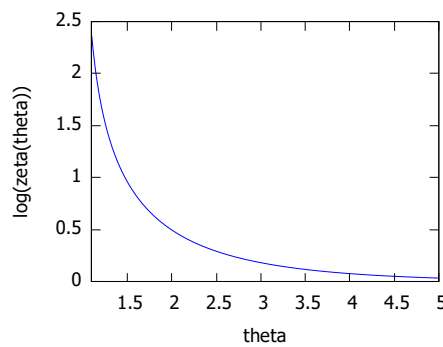
$$\zeta(s) = \sum_{i=1}^{\infty} \frac{1}{i^s} = 1 + \frac{1}{2^s} + \frac{1}{3^s} + \dots, \quad s > 1. \quad (2)$$

The set of zeta distributions  $\mathcal{Z} = \{p_s(x) : s \in (1, \infty)\}$  forms a *discrete exponential family* [6] with natural parameter  $\theta(s) = s$  lying in the natural parameter space  $\Theta = (1, \infty)$ , the sufficient statistic  $t(x) = -\log x$ , and the cumulant function or log-normalizer  $F(\theta) = \log \zeta(\theta)$ . Therefore it follows from the theory of exponential families [6] that  $\log \zeta(\theta)$  is a strictly convex and real analytic function (see Figure 1). Thus the pmf of zeta distributions can be rewritten in the canonical form of exponential families as:

$$p_s(x) = \exp(\theta(s)t(x) - F(\theta(s))). \quad (3)$$

The characteristic function is thus  $\phi_s(t) = \frac{\zeta(s+it)}{\zeta(s)}$ .

Thus a zeta distribution  $p_s(x)$  can be interpreted as the discrete equivalent of a Pareto distribution  $q_s(x)$  of scale 1 and shape  $s - 1$  with probability density function  $q_s(x) = \frac{s-1}{x^s}$  for  $x > 1$  (see Table 1).



**Figure 1.** Plot of  $F(\theta) = \log \zeta(\theta)$ , a strictly convex and analytic function.

The zeta function is known to be irrational at many positive odd integers [7–9] and can be calculated using Bernoulli numbers [10] for positive even integers:  $\zeta(2n) = \frac{(-1)^{n+1} B_{2n} (2\pi)^{2n}}{2(2n)!}$ ,  $n \in \mathbb{N}$ . The zeta function can be calculated fast [11] and precisely [12]. The derivatives of the zeta function have also been studied [11,13].

The zeta distributions are related to the Zipf distributions [14]  $p_{s,N}(x) \propto \frac{1}{x^s}$  for  $x \in \{1, \dots, N\}$  and the Zipf-Mandelbrot distributions [15,16]  $p_{s,q,N}(x) \propto \frac{1}{(x+q)^s}$  for  $x \in \{1, \dots, N\}$  which play an important role in quantitative linguistics. See [?] for more details. The Zipf distributions and the Zipf-Mandelbrot distributions both have finite support and can be interpreted as truncated zeta distributions (right truncation for Zipf distributions and both left & right truncations for the Zipf-Mandelbrot distributions) with normalizing constants which can be calculated approximately using properties of the zeta function [17]. Left-only truncations of the Zeta distributions are called Hurwitz zeta distributions [18]. Similarly, truncated Pareto distributions are used in applications [19]. Notice that truncated distributions of an exponential family with fixed truncation support form another exponential family [20]. The zeta distributions are infinite divisible [18,21]: A random variable following a zeta distribution can be expressed as the probability distribution of the sum of an arbitrary number of independent and identically distributed random variables. In applications, it is important to quantitatively discriminate between zeta distributions (see, for example [22,23] or [24]). Mixtures of zeta distributions have also been used to model social networks [25]. In general, products of exponential families yield other exponential families. The products of  $d$  zeta distributions form an exponential family called the zeta-star distributions [21].

In this paper, we first study various information-theoretic measures between zeta distributions by considering them as a discrete exponential family [26]: Namely, we consider the  $\alpha$ -divergences [27] between zeta distributions in §2, and study their limit Kullback-Leibler oriented divergences when  $\alpha \rightarrow 1$  and  $\alpha \rightarrow 0$  in §3. We then compare these results with the counterpart results obtained for the continuous exponential family of Pareto distributions in §4. Finally, we conclude this work in §5.

## 2. Amari's $\alpha$ -divergences and Sharma-Mittal divergences

To measure the dissimilarity between two zeta distributions  $p_{s_1}$  and  $p_{s_2}$ , one can use the  $\alpha$ -divergences [27] defined for a real  $\alpha \in (0, 1)$  as follows:

$$D_\alpha[p_{s_1} : p_{s_2}] := \frac{1}{\alpha(1-\alpha)} (1 - I_\alpha[p_{s_1} : p_{s_2}]) = D_{1-\alpha}[p_{s_2} : p_{s_1}], \quad (4)$$

where

$$I_\alpha[p_1, p_2] := \sum_{i=1}^{\infty} p_1(x)^\alpha p_2(x)^{1-\alpha} = I_{1-\alpha}[p_2 : p_1], \quad (5)$$

is the  $\alpha$ -Bhattacharyya coefficient (a similarity measure also called an affinity coefficient).

It follows from [28] that the skewed Bhattacharyya coefficient amounts to a skewed Jensen divergence between the natural parameters of the exponential family  $\mathcal{E}$ :

$$I_\alpha[p_{s_1} : p_{s_2}] = \exp(-J_{F,\alpha}(s_1 : s_2)), \quad (6)$$

where  $J_{F,\alpha}$  is the skewed Jensen divergence induced by a strictly convex and smooth convex function  $F(\theta)$ :

$$J_{F,\alpha}(s_1 : s_2) := \alpha F(s_1) + (1 - \alpha)F(s_2) - F(\alpha s_1 + (1 - \alpha)s_2) \geq 0, \quad (7)$$

$$= \log \left( \frac{\zeta(s_1)^\alpha \zeta(s_2)^{1-\alpha}}{\zeta(\alpha s_1 + (1 - \alpha)s_2)} \right). \quad (8)$$

Thus we have the  $\alpha$ -divergences between two zeta distributions  $p_{s_1}$  and  $p_{s_2}$  available in closed-form.

**Theorem 1** ( $\alpha$ -divergences between two zeta distributions). *The  $\alpha$ -divergence for  $\alpha \in (0, 1)$  between two zeta distributions  $p_{s_1}$  and  $p_{s_2}$  is:*

$$D_\alpha[p_{s_1} : p_{s_2}] = \frac{1}{\alpha(1 - \alpha)} \left( 1 - \frac{\zeta(\alpha s_1 + (1 - \alpha)s_2)}{\zeta(s_1)^\alpha \zeta(s_2)^{1-\alpha}} \right).$$

It follows that when  $s_1$ ,  $s_2$ , and  $\alpha s_1 + (1 - \alpha)s_2$  are all positive even integers, we can evaluate exactly the  $\alpha$ -divergences between  $p_{s_1}$  and  $p_{s_2}$ .

**Example 1.** Consider  $s_1 = 4$  and  $s_2 = 12$  with  $\alpha = \frac{1}{2}$  so that  $\alpha s_1 + (1 - \alpha)s_2 = 8$ . Using the formula [10]  $\zeta(2n) = \frac{(-1)^{n+1} B_{2n} (2\pi)^{2n}}{2(2n)!}$ ,  $n \in \mathbb{N}$  where  $B_{2n}$  denotes the Bernoulli numbers, the zeta functions can be calculated exactly at 4, 8 and 12:  $\zeta(4) = \frac{\pi^4}{90}$ ,  $\zeta(8) = \frac{\pi^8}{9450}$ , and  $\zeta(12) = \frac{691\pi^{12}}{638512875}$ . The  $\alpha$ -divergence for  $\alpha = \frac{1}{2}$  is the squared Hellinger divergence  $D_{\frac{1}{2}}[p_{s_1}, p_{s_2}] = \sum_{i=1}^{\infty} \left( \sqrt{p_{s_1}(i)} - \sqrt{p_{s_2}(i)} \right)^2$ . Thus we find the exact squared Hellinger divergence:  $D_{\frac{1}{2}}[p_4, p_{12}] = 4 \left( 1 - 3\sqrt{\frac{715}{6910}} \right) \simeq 0.139929 \dots$

Let us report another example where the squared Hellinger divergence is expressed using the zeta function:

**Example 2.** We consider  $s_1 = 3$ ,  $s_2 = 7$  and  $\alpha = \frac{1}{2}$  so that  $\alpha s_1 + (1 - \alpha)s_2 = 5$ . Then we have  $D_{\frac{1}{2}}[p_3, p_7] = 4 \left( 1 - \frac{\zeta(5)}{\sqrt{\zeta(3)\zeta(7)}} \right) \simeq 0.23261 \dots$

Since  $\lim_{\alpha \rightarrow 1} D_\alpha[p_{s_1} : p_{s_2}] = D_{\text{KL}}[p_{s_1} : p_{s_2}]$  is the Kullback-Leibler divergence [27] (KLD)

$$D_{\text{KL}}[p_{s_1} : p_{s_2}] := \sum_{i=1}^{\infty} p_{s_1}(i) \log \frac{p_{s_1}(i)}{p_{s_2}(i)}, \quad (9)$$

we can approximate the KLD by  $D_{1-\epsilon}[s_1 : s_2]$  for a small value of  $\epsilon$  (say,  $\epsilon = 1.0e^{-3}$ ) using fast methods to compute the zeta function [11].

**Corollary 1** (Approximation of the Kullback-Leibler divergence). *The Kullback-Leibler divergence between two zeta distributions  $p_{s_1}$  and  $p_{s_2}$  can be approximated for small values  $\epsilon > 0$  by*

$$D_{\text{KL}}[p_{s_1} : p_{s_2}] \simeq D_{1-\epsilon}[p_{s_1} : p_{s_2}] = \frac{1}{\epsilon(1 - \epsilon)} \left( 1 - \frac{\zeta((1 - \epsilon)s_1 + \epsilon s_2)}{\zeta(s_1)^{1-\epsilon} \zeta(s_2)^\epsilon} \right).$$

**Example 3.** We let  $1 - \epsilon = 0.99$ ,  $1 - \epsilon = 0.999$ ,  $1 - \epsilon = 0.9999$ ,  $1 - \epsilon = 0.99999$ , and find the following numerical approximations:

$$\begin{aligned} D_{\text{KL}}[p_{s_1} : p_{s_2}] &\simeq_{1-\epsilon=0.99} 0.473 \dots \\ D_{\text{KL}}[p_{s_1} : p_{s_2}] &\simeq_{1-\epsilon=0.999} 0.482 \dots \\ D_{\text{KL}}[p_{s_1} : p_{s_2}] &\simeq_{1-\epsilon=0.9999} 0.483 \dots \\ D_{\text{KL}}[p_{s_1} : p_{s_2}] &\simeq_{1-\epsilon=0.99999} 0.483 \dots \end{aligned}$$

We can also calculate the KLD  $D_{\text{KL}}[p_{s_1}^{\mathcal{X}_1} : p_{s_2}^{\mathcal{X}_2}]$  between two truncated zeta distributions with nested supports  $\mathcal{X}_1 \subseteq \mathcal{X}_2$ . See [20]. A truncated zeta distribution on the support  $\{a, a+1, \dots, b\} \subset \mathbb{N}$  (with  $b > a$ ) has pmf  $p_s^{a,b}(x) = \frac{p_s(x)}{\Phi_s(b) - \Phi_s(a)}$  where  $\Phi_s(u)$  is the cumulative distribution function  $\Phi_s(u) = \sum_{x \in \{1, \dots, u\}} p_s(x) = \frac{1}{\zeta(s)} \sum_{x \in \{1, \dots, u\}} \frac{1}{x^s}$ .

The Chernoff information [29] is defined by  $C[p_1, p_2] = -\log \min_{\alpha \in (0,1)} I_\alpha[p_1, p_2]$ . The unique optimal value  $\alpha^*$  maximizing the Chernoff  $\alpha$ -divergences  $C_\alpha[p_1, p_2] = -\log I_\alpha[p_1, p_2]$  is called the Chernoff exponent [29] due to its role in bounding the probability of error in Bayesian hypothesis testing. When both pdfs or pmfs belong to the same exponential family, we have [29]

$$C[p_{\theta_1}, p_{\theta_2}] = J_F(\theta_1 : (\theta_1 \theta_2)_{\alpha^*}) = B_F(\theta_1 : (\theta_1 \theta_2)_{\alpha^*}) = B_F(\theta_2 : (\theta_1 \theta_2)_{\alpha^*}), \quad (10)$$

where  $B_F$  denotes the Bregman divergence (corresponding to the KLD) and  $(\theta_1 \theta_2)_{\alpha^*} = \alpha^* \theta_1 + (1 - \alpha^*) \theta_2$ . For unioorder exponential family like the zeta distributions, a closed-form formula for the optimal Chernoff exponent  $\alpha^*$  is reported in [29]:  $\alpha^* = \frac{F'^{-1}\left(\frac{F(\theta_2) - F(\theta_1)}{\theta_2 - \theta_1}\right) - \theta_1}{\theta_2 - \theta_1}$ .

The Sharma-Mittal divergences [30] between two densities  $p$  and  $q$  is a biparametric family of relative entropies is defined by

$$D_{\alpha,\beta}[p : q] = \frac{1}{\beta - 1} \left( \left( \int p(x)^\alpha q(x)^{1-\alpha} dx \right)^{\frac{1-\beta}{1-\alpha}} - 1 \right), \forall \alpha > 0, \alpha \neq 1, \beta \neq 1. \quad (11)$$

The Sharma-Mittal divergence is induced from the Sharma-Mittal entropies which unifies the extensive Rényi entropies with the non-extensive Tsallis entropies [30]. The Sharma-Mittal divergences include the Rényi divergences ( $\beta \rightarrow 1$ ) and the Tsallis divergences ( $\beta \rightarrow \alpha$ ), and in the limit case of  $\alpha, \beta \rightarrow 1$  the Kullback-Leibler divergence [31]. When both densities  $p = p_{\theta_1}$  and  $q = p_{\theta_2}$  belong to the same exponential family, we have the following closed-form formula [31]:

$$D_{\alpha,\beta}[p_{\theta_1} : p_{\theta_2}] = \frac{1}{\beta - 1} \left( e^{-\frac{1-\beta}{1-\alpha} J_{F,\alpha}(\theta_1 : \theta_2)} - 1 \right). \quad (12)$$

Thus we get the following theorem:

**Theorem 2.** For  $\alpha > 0$ ,  $\alpha \neq 1$ ,  $\beta \neq 1$ , the Sharma-Mittal divergence between two zeta distributions  $p_{s_1}$  and  $p_{s_2}$  is

$$D_{\alpha,\beta}[p_{s_1} : p_{s_2}] = \frac{1}{\beta - 1} \left( \left( \frac{\zeta(\alpha s_1 + (1-\alpha)s_2)}{\zeta(s_1)^\alpha \zeta(s_2)^{1-\alpha}} \right)^{\frac{1-\beta}{1-\alpha}} - 1 \right).$$

### 3. The Kullback-Leibler divergence between two zeta distributions

It is well-known that the KLD between two probability mass functions of an exponential family amounts to a reverse Bregman divergence induced by the cumulant function [32]:  $D_{\text{KL}}[p_{s_1} : p_{s_2}] =$

$B_F^*(\theta_1 : \theta_2) := B_F(\theta_2 : \theta_1)$  (with  $\theta_1 = s_1$  and  $\theta_2 = s_2$ ). Furthermore, this Bregman divergence amounts to a Fenchel-Young divergence [33] so that we have

$$D_{\text{KL}}[p_{s_1} : p_{s_2}] = B_F(\theta_2 : \theta_1) = F(\theta(s_2)) + F^*(\eta(s_1)) - \theta(s_2)\eta(s_1), \quad (13)$$

where  $F^*(\eta)$  denotes the Legendre convex conjugate of  $F$ ,  $\theta(s) = s$  and  $\eta(s) = F'(\theta(s)) = E_{p_s}[t(x)] = -E_{p_s}[\log x]$ , see [6]. Moreover, the convex conjugate  $F^*(\eta(s))$  corresponds to the negentropy [34]:  $F^*(\eta(s)) = -H[p_s]$ , where the entropy of a zeta distribution  $p_s$  is defined by:

$$H[p_s] := \sum_{i=1}^{\infty} p_s(i) \log \frac{1}{p_s(i)}. \quad (14)$$

68 Using the fact that  $\sum_{i=1}^{\infty} p_s(i) = 1 = \sum_{i=1}^{\infty} \frac{1}{i^s \zeta(s)}$ , we can express the entropy as follows:

$$\begin{aligned} H[p_s] &= \sum_{i=1}^{\infty} \frac{1}{i^s \zeta(s)} \log i^s + \log(\zeta(s)) \sum_{i=1}^{\infty} \frac{1}{i^s \zeta(s)}, \\ &= \sum_{i=1}^{\infty} \frac{1}{i^s \zeta(s)} \log(i^s \zeta(s)). \end{aligned}$$

Since  $F(\theta) = \log \zeta(\theta)$ , we have  $\eta(\theta) = F'(\theta) = \frac{\zeta'(\theta)}{\zeta(\theta)}$ . The function  $\frac{\zeta'(\theta)}{\zeta(\theta)}$  has been tabulated in [35] (page 400). Notice that the maximum likelihood estimator [6] of  $n$  independently and identically distributed observations  $x_1, \dots, x_n$  is  $\hat{\eta} = \frac{1}{n} \sum_{i=1}^n t(x_i)$ . Thus we have:

$$\hat{\eta} = \frac{\zeta'(\hat{\theta})}{\zeta(\hat{\theta})} = -\frac{1}{n} \sum_{i=1}^n \log x_i. \quad (15)$$

69 The inverse of the zeta function  $\zeta^{-1}(\cdot)$  has been studied in [36].

70 **Proposition 1** (KLD between zeta distributions). *The Kullback-Leibler divergence between two zeta*  
71 *distributions can be written as:*

$$\begin{aligned} D_{\text{KL}}[p_{s_1} : p_{s_2}] &= \log(\zeta(s_2)) - H[p_{s_1}] + s_2 E_{p_{s_1}}[\log x], \\ &= \log(\zeta(s_2)) - \sum_{i=1}^{\infty} \frac{1}{i^{s_1} \zeta(s_1)} \log(i^{s_1} \zeta(s_1)) - s_2 \frac{\zeta'(s_1)}{\zeta(s_1)}. \end{aligned}$$

Moreover, the logarithmic derivative of the zeta function can be expressed using the von Mangoldt function [37] (page 1850) for  $\theta > 1$ :

$$\eta(\theta) = \frac{\zeta'(\theta)}{\zeta(\theta)} = -\sum_{i=1}^{\infty} \frac{\Lambda(i)}{i^{\theta}}, \quad (16)$$

72 where  $\Lambda(i) = \log p$  is  $i = p^k$  for some prime  $p$  and integer  $k \geq 1$  and 0 otherwise. Notice that the zeta  
73 function can be calculated using Euler product formula:  $\zeta(\theta) = \prod_{p:\text{prime}} \frac{1}{1-p^{-\theta}}$ .

**Theorem 3.** *The Kullback-Leibler divergence between two zeta distributions can be expressed using the real zeta function  $\zeta$  and the von Mangoldt function  $\Lambda$  as:*

$$D_{\text{KL}}[p_{s_1} : p_{s_2}] = \log(\zeta(s_2)) - \sum_{i=1}^{\infty} \frac{1}{i^s \zeta(s)} \log(i^s \zeta(s)) + s_2 \sum_{i=1}^{\infty} \frac{\Lambda(i)}{i^{s_1}}.$$

**Example 4.** Consider  $s_1 = 4$  and  $s_2 = 12$ . Letting  $1 - \epsilon = 0.9999$  and using Corollary 1, we get

$$D_{\text{KL}}[p_{s_1} : p_{s_2}] \simeq D_{1-\epsilon}[p_{s_1} : p_{s_2}] = 0.430479743738878 \dots$$

Let us now calculate the KLD using Theorem 3, we get  $\log(\zeta(s_2)) = \log \frac{691\pi^{12}}{638512875}$ ,  $H[p_{s_1}] \simeq 0.3337829096182664 \dots$  (using 100 terms), and  $\eta(s_1) = -0.06366938697034288 \dots$  (using 100 terms) so that we have

$$D_{\text{KL}}[p_{s_1} : p_{s_2}] = \log(\zeta(s_2)) - \sum_{i=1}^{\infty} \frac{1}{i^s \zeta(s)} \log(i^s \zeta(s)) + s_2 \sum_{i=1}^{\infty} \frac{\Lambda(i)}{i^{s_1}}, \quad (17)$$

$$\simeq 0.430495790304827 \dots \quad (18)$$

It is well-known that the KLD between two arbitrarily close zeta distributions  $p_s$  and  $p_{s+ds}$  amounts to half of the quadratic distance induced by the Fisher information:

$$D_{\text{KL}}[p_s : p_{s+ds}] \approx \frac{1}{2} I(s) ds^2, \quad (19)$$

where

$$I(s) = E_{p_s}[(\log p_s(x))'^2] = -E_{p_s}[(\log p_s(x))''], \quad (20)$$

where the first-order and second-order derivatives are taken with respect to the parameter  $s$ . Thus for unioorder exponential families, the Fisher information matrix is

$$I(s) = -E_{p_s}[(\log p_s(x))''] = (\log \zeta(s))'' = \frac{\zeta(s)\zeta''(s) - \zeta'(s)^2}{\zeta^2(s)}. \quad (21)$$

This second-order derivative  $(\log \zeta(s))''$  has been studied in [38]. We have

$$I(s) = \sum_{n=1}^{\infty} \Lambda(n) \log(n) n^{-s} \quad (22)$$

where  $\Lambda(n)$  is the Von Mangoldt function.

#### 4. Comparison of the Zeta family with a Pareto subfamily

The zeta distribution is also called the “pure power-law distribution” is in the literature [2].

We can compute the  $\alpha$ -divergences between two Pareto distributions  $q_{s_1}$  and  $q_{s_2}$  with fixed scale 1 and respective shapes  $s_1 - 1$  and  $s_2 - 1$ . The Pareto density writes  $q_s(x) = \frac{s-1}{x^s}$  for  $x \in \mathcal{X} = (1, \infty)$ . The family of such Pareto distributions forms a continuous exponential family with natural parameter  $\theta = s$ , sufficient statistic  $t(x) = -\log(x)$ , and convex cumulant function  $F(\theta) = -\log(\theta - 1)$  for  $\theta \in \Theta = (1, \infty)$ . Thus we have [28]:

$$I_{\alpha}[q_1 : q_2] = \int q_{s_1}(x)^{\alpha} q_{s_2}(x)^{1-\alpha} dx = \exp(-J_{F,\alpha}(\theta_1 : \theta_2)), \quad (23)$$

$$= \frac{\alpha s_1 + (1 - \alpha) s_2}{s_1^{\alpha} s_2^{1-\alpha}}, \quad (24)$$

and we get the following closed-form for the  $\alpha$ -divergences between two Pareto distributions  $q_{s_1}$  and  $q_{s_2}$ :

$$D_{\alpha}[q_{s_1} : q_{s_2}] = \frac{1}{\alpha(1 - \alpha)} \left( 1 - \frac{\alpha s_1 + (1 - \alpha) s_2}{s_1^{\alpha} s_2^{1-\alpha}} \right). \quad (25)$$

**Table 1.** Comparisons between the Zeta family and the Pareto subfamily. The function  $\zeta(s)$  is the real zeta function.

	Zeta distribution	Pareto distribution
Univariate uni-order exponential family $\exp(\theta t(x) - F(\theta))$		
	Discrete EF	Continuous EF
PMF/PDF	$p_s(x) = \frac{1}{x^s \zeta(s)}$	$q_s(x) = \frac{s-1}{x^s}$
Support $\mathcal{X}$	$\mathbb{N} = \{1, 2, \dots\}$	$(1, \infty)$
Natural parameter $\theta$	$s$	$s$
Cumulant $F(\theta)$	$\log \zeta(\theta)$	$-\log(\theta - 1)$
Sufficient statistic $t(x)$	$-\log x$	$-\log x$
Moment parameter $\eta$	$\frac{\zeta'(\theta)}{\zeta(\theta)}$	$-\frac{1}{s-1}$
Conjugate $F^*(\eta)$	$-H[p_s] = -\sum_{i=1}^{\infty} \frac{1}{i^s \zeta(s)} \log(i^s \zeta(s))$	$\eta - 1 - \log(-\eta)$
Maximum likelihood estimator	$\hat{\eta} = \frac{\zeta'(\hat{\theta})}{\zeta(\hat{\theta})} = -\frac{1}{n} \sum_{i=1}^n \log x_i$	$\hat{s} = \frac{n}{\sum_{i=1}^n \log x_i}$
Fisher information	$\sum_{i=0}^{\infty} \Lambda(i) \log(i) i^{-s}$	$\frac{1}{(s-1)^2}$
Entropy $-F^*(\eta(s))$	$\sum_{i=1}^{\infty} \frac{1}{i^s \zeta(s)} \log(i^s \zeta(s))$	$1 + \frac{1}{s-1} - \log(s-1)$
Bhattacharyya coefficient $I_\alpha$	$\frac{\zeta(\alpha s_1 + (1-\alpha)s_2)}{\zeta(s_1)^\alpha \zeta(s_2)^{1-\alpha}}$	$\frac{\alpha s_1 + (1-\alpha)s_2}{s_1^\alpha s_2^{1-\alpha}}$
Kullback-Leibler divergence	$\log(\zeta(s_2)) - \sum_{i=1}^{\infty} \frac{1}{i^s \zeta(s)} \log(i^s \zeta(s)) - s_2 \frac{\zeta'(s_1)}{\zeta(s_1)}$	$\log\left(\frac{s_1-1}{s_2-1}\right) + \frac{s_2-s_1}{s_1-1}$

The moment parameter is  $\eta(\theta) = F'(\theta) = -\frac{1}{\theta-1}$  so that  $\theta(\eta) = 1 - \frac{1}{\eta}$  and  $F^*(\eta) = \theta(\eta)\eta - F(\theta(\eta)) = \eta - 1 - \log(-\eta)$ . It follows that the KLD is

$$D_{\text{KL}}[q_{s_1} : q_{s_2}] = B_F(\theta_2 : \theta_1) = \log\left(\frac{s_1-1}{s_2-1}\right) + \frac{s_2-s_1}{s_1-1}. \quad (26)$$

The differential entropy of the Pareto distribution  $q_s$  is

$$h[q_s] = -\int_1^\infty q_s(x) \log q_s(x) dx = -F^*(\eta(s)) \quad (27)$$

with  $\eta(s) = -\frac{1}{s-1}$ . We find that

$$h[q_s] = 1 + \frac{1}{s-1} - \log(s-1). \quad (28)$$

**Example 5.** For comparison, we calculate the KLD between two Pareto distributions with parameters  $s_1 = 4$  and  $s_2 = 12$ . We find

$$D_{\text{KL}}[q_{s_1} : q_{s_2}] = \log \frac{3}{11} + \frac{8}{3} \simeq 1.367383682536406 \dots$$

## 5. Conclusion

Table 1 compares the discrete exponential family of zeta distributions with the continuous exponential family of Pareto distributions with fixed scale 1.

In general, it is interesting to consider discrete counterparts of continuous exponential families. For example, the discrete Gaussian distributions or discrete normal distributions defined as maximum entropy distributions have been studied in [39,40]. The log-normalizer or cumulant function of the discrete Gaussian distributions are related to the Riemann theta function [41]. Given a prescribed sufficient statistics  $t(x)$ , we may define the continuous exponential family with respect to the Lebesgue measure  $\mu$  as the probability density functions  $p(x)$  maximizing the differential entropy under the moment constraint  $E_p[t(x)] = \eta$ . The corresponding discrete exponential family is obtained by the distributions with probability mass functions maximizing Shannon entropy under the moment constraint  $E_p[t(x)] = \eta$ .

Additional material is available online at <https://franknielsen.github.io/ZetaParetoExpFam/index.html> (accessed on 18 October 2022).

## References

- Kotz, S.; Balakrishnan, N.; Read, C.; Vidakovic, B. *Encyclopedia of Statistical Sciences (Volume 15)*; Wiley, 2005.
- Goldstein, M.L.; Morris, S.A.; Yen, G.G. Problems with fitting to the power-law distribution. *The European Physical Journal B-Condensed Matter and Complex Systems* **2004**, *41*, 255–258.
- Titchmarsh, E.C.; Heath-Brown, D.R.; Titchmarsh, E.C.T. *The theory of the Riemann zeta-function*; Oxford university press, 1986.
- Tempesta, P. Group entropies, correlation laws, and zeta functions. *Physical Review E* **2011**, *84*, 021121.
- Iwaniec, H. *Lectures on the Riemann zeta function*; Vol. 62, American Mathematical Society, 2014.
- Barndorff-Nielsen, O. *Information and exponential families in statistical theory*; John Wiley & Sons: New Jersey, 2014.
- Apéry, R. Irrationalité de  $\zeta(2)$  et  $\zeta(3)$ . *Astérisque* **1979**, *61*, 1.
- Rivoal, T. La fonction zêta de Riemann prend une infinité de valeurs irrationnelles aux entiers impairs. *Comptes Rendus de l'Académie des Sciences-Series I-Mathematics* **2000**, *331*, 267–270.
- Fischler, S.; Sprang, J.; Zudilin, W. Many odd zeta values are irrational. *Compositio Mathematica* **2019**, *155*, 938–952.
- Graham, R.L.; Knuth, D.E.; Patashnik, O. *Concrete mathematics: a foundation for computer science*; Addison-Wesley Professional, 1994.
- Hiary, G.A. Fast methods to compute the Riemann zeta function. *Annals of mathematics* **2011**, pp. 891–946.
- Johansson, F. Rigorous high-precision computation of the Hurwitz zeta function and its derivatives. *Numerical Algorithms* **2015**, *69*, 253–270.
- Yildirim, C.Y. A Note on  $\zeta''(s)$  and  $\zeta'''(s)$ . *Proceedings of the American Mathematical Society* **1996**, pp. 2311–2314.
- Powers, D.M. Applications and explanations of Zipf's law. *New methods in language processing and computational natural language learning*, 1998.
- Mandelbrot, B. *Information Theory and Psycholinguistics: A Theory of Word Frequencies, Readings in Mathematical Social Sciences*; MIT Press, MA, USA, 1966.
- Lovričević, N.; Pečarić, D.; Pečarić, J. Zipf-Mandelbrot law,  $f$ -divergences and the Jensen-type interpolating inequalities. *Journal of inequalities and applications* **2018**, *2018*, 1–20.
- Naldi, M. Approximation of the truncated Zeta distribution and Zipf's law. *arXiv preprint arXiv:1511.01480* **2015**.
- Hu, C.Y.; Iksanov, A.M.; Lin, G.D.; Zakusylo, O.K. The Hurwitz zeta distribution. *Australian & New Zealand Journal of Statistics* **2006**, *48*, 1–6.
- Deluca, A.; Corral, Á. Fitting and goodness-of-fit test of non-truncated and truncated power-law distributions. *Acta Geophysica* **2013**, *61*, 1351–1394.
- Nielsen, F. Statistical Divergences between Densities of Truncated Exponential Families with Nested Supports: Duo Bregman and Duo Jensen Divergences. *Entropy* **2022**, *24*, 421.
- Saito, S.; Tanaka, T. A note on infinite divisibility of zeta distributions. *Applied Mathematical Sciences* **2012**, *6*, 1455–61.
- Wang, T.; Zhang, W.; Maunder, R.G.; Hanzo, L. Near-capacity joint source and channel coding of symbol values from an infinite source set using Elias gamma error correction codes. *IEEE transactions on communications* **2013**, *62*, 280–292.
- Oosawa, T.; Matsuda, T. SQL injection attack detection method using the approximation function of zeta distribution. 2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC). IEEE, 2014, pp. 819–824.
- Doray, L.G.; Luong, A. Quadratic distance estimators for the zeta family. *Insurance: Mathematics and Economics* **1995**, *16*, 255–260.
- Jung, H.; Phoa, F.K.H. A Mixture Model of Truncated Zeta Distributions with Applications to Scientific Collaboration Networks. *Entropy* **2021**, *23*, 502.



26. Barndorff-Nielsen, O. *Information and exponential families: in statistical theory*; John Wiley & Sons, 2014.
27. Cichocki, A.; Amari, S.i. Families of alpha-beta-and gamma-divergences: Flexible and robust measures of similarities. *Entropy* **2010**, *12*, 1532–1568.
28. Nielsen, F.; Boltz, S. The Burbea-Rao and Bhattacharyya centroids. *IEEE Transactions on Information Theory* **2011**, *57*, 5455–5466.
29. Nielsen, F. An information-geometric characterization of Chernoff information. *IEEE Signal Processing Letters* **2013**, *20*, 269–272.
30. Sharma, B.D.; Mittal, D.P. New non-additive measures of entropy for discrete probability distributions. *J. Math. Sci* **1975**, *10*, 28–40.
31. Nielsen, F.; Nock, R. A closed-form expression for the Sharma–Mittal entropy of exponential families. *Journal of Physics A: Mathematical and Theoretical* **2011**, *45*, 032003.
32. Azoury, K.S.; Warmuth, M.K. Relative loss bounds for on-line density estimation with the exponential family of distributions. *Machine Learning* **2001**, *43*, 211–246.
33. Nielsen, F. On geodesic triangles with right angles in a dually flat space. In *Progress in Information Geometry*; Springer, 2021; pp. 153–190.
34. Nielsen, F.; Nock, R. Entropies and cross-entropies of exponential families. 2010 IEEE International Conference on Image Processing. IEEE, 2010, pp. 3621–3624.
35. Walther, A. Anschauliches zur Riemannschen zetafunktion. *Acta Mathematica* **1926**, *48*, 393–400.
36. Kawalec, A. The inverse Riemann zeta function. *arXiv preprint arXiv:2106.06915* **2021**.
37. Weisstein, E.W. *CRC concise encyclopedia of mathematics*; CRC press, 2002.
38. Stopple, J. Notes on  $\log(\zeta(s))''$ . *The Rocky Mountain Journal of Mathematics* **2016**, *46*, 1701–1715.
39. Agostini, D.; Améndola, C. Discrete Gaussian distributions via theta functions. *SIAM Journal on Applied Algebra and Geometry* **2019**, *3*, 1–30.
40. Nielsen, F. The Kullback–Leibler Divergence Between Lattice Gaussian Distributions. *Journal of the Indian Institute of Science* **2022**, pp. 1–12.
41. Deconinck, B.; Heil, M.; Bobenko, A.; Van Hoeij, M.; Schmies, M. Computing Riemann theta functions. *Mathematics of Computation* **2004**, *73*, 1417–1442.