physical sciences forum

MDPI

Proceeding Paper

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

Sony Computer Science Laboratories Inc., Tokyo 141-0022, Japan; frank.nielsen@acm.org

† Presented at the 41st International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, Paris, France, 18–22 July 2022.

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 a 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

check for updates

Citation: Nielsen, F. 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. *Phys. Sci. Forum* **2022**, *5*, 2. https://doi.org/10.3390/psf2022005002

Academic Editors: Frédéric Barbaresco, Ali Mohammad-Djafari and Martino Trassinelli

Published: 31 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

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 $\mathbb{N}$:

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

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]:

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

The set of zeta distributions $\mathcal{Z} = \{p_s(x) \ : \ s \in (1, \infty)\}$ forms a *discrete exponential family* [6,7] 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 [7] 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:

$$p_s(x) = \exp(\theta(s)t(x) - F(\theta(s))). \tag{3}$$

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).

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 $\mathcal{X}$	$\mathbb{N} = \{1, 2, \ldots\}$	$(1, \infty)$
Natural parameter θ	s	s
Cumulant $F(\theta)$	$\log \zeta(\theta)$	$-\log(\theta - 1)$
Sufficient statistic $t(x)$	$-\log x$	$-\log x$
Moment parameter η	$\frac{\zeta'(\theta)}{\zeta(\theta)}$	$-\frac{1}{s-1}$
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_α	$\frac{\zeta(\alpha s_1 + (1-\alpha) s_2)}{\zeta(s_1)^\alpha \zeta(s_2)^{1-\alpha}}$	$\frac{\alpha s_1 + (1-\alpha) s_2}{s_1^\alpha s_2^{1-\alpha}}$
Kullback-Leibler divergence	$\log(\zeta(s_2)) - \sum_{i=1}^{\infty} \frac{1}{i^s \zeta(s)} \log(i^s \zeta(s)) - s_2 \frac{\zeta'(s_1)}{\zeta(s_1)}$	$\log\left(\frac{s_1 - 1}{s_2 - 1}\right) + \frac{s_2 - s_1}{s_1 - 1}$

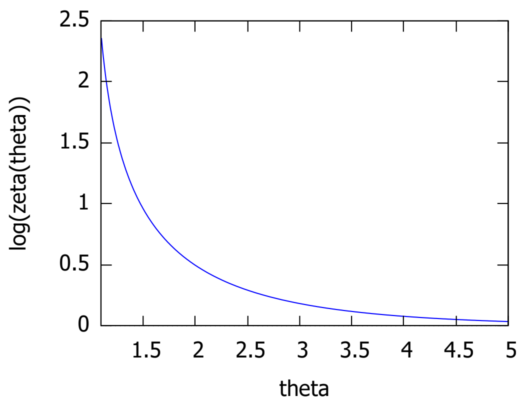

Figure 1. Plot of $F(\theta) = \log \zeta(\theta)$, a strictly convex and analytic function.

The zeta function is known to be irrational at many positive odd integers [8–10] and can be calculated using Bernoulli numbers [11] 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 [12] and precisely [13]. The derivatives of the zeta function have also been studied [12,14].

The zeta distributions are related to the Zipf distributions [15] $p_{s,N}(x) \propto \frac{1}{x^s}$ for $x \in \{1, \ldots, N\}$ and the Zipf–Mandelbrot distributions [16,17] $p_{s,q,N}(x) \propto \frac{1}{(x+q)^s}$ for $x \in \{1, \ldots, N\}$, which play an important role in quantitative linguistics. See [6] 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 [18]. Left-only truncations of the Zeta distributions are called Hurwitz zeta distributions [19]. Similarly, truncated Pareto distributions are used in applications [20]. Notice that truncated distributions of an exponential family with fixed truncation support form another exponential family [21]. The zeta distributions are infinite divisible [19,22]: 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 [23,24] or [25]). Mixtures of zeta distributions have also been used to model social networks [26]. 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 [22].

In this paper, we first study various information-theoretic measures between zeta distributions by considering them as a discrete exponential family [7]: That is, we consider the α-divergences [27] between zeta distributions in Section 2, and study their limit Kullback–Leibler-oriented divergences when $\alpha \to 1$ and $\alpha \to 0$ in Section 3. We then compare these results with the counterpart results obtained for the continuous exponential family of Pareto distributions in Section 4. Finally, we conclude this work in Section 5.

2. Amari's α-Divergences and Sharma–Mittal Divergences

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

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

where

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

is the α-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 $\mathcal{E}$:

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

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

$$\begin{aligned} J_{F,\alpha}(s_1 : s_2) &:= \alpha F(s_1) + (1-\alpha)F(s_2) - F(\alpha s_1 + (1-\alpha)s_2) \geq 0, &(7)\\ &= \log\left(\frac{\zeta(s_1)^\alpha \zeta(s_2)^{1-\alpha}}{\zeta(\alpha s_1 + (1-\alpha)s_2)}\right). &(8)\end{aligned}$$

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

Theorem 1 (α-divergences between two zeta distributions)**.** *The α-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 α-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 [11] $\zeta(2n) = \frac{(-1)^{n+1}B_{2n}(2\pi)^{2n}}{2(2n)!}$, $\quad n \in \mathbb{N}$ 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 α-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: $D_{\frac{1}{2}}[p_4, p_{12}] = 4\left(1 - 3\sqrt{\frac{715}{6910}}\right) \simeq 0.139929\ldots$.*

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* $D_{\frac{1}{2}}[p_3, p_7] = 4\left(1 - \frac{\zeta(5)}{\sqrt{\zeta(3)\zeta(7)}}\right) \simeq 0.23261\ldots$

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

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

we can approximate the KLD by $D_{1-\epsilon}[s_1 : s_2]$ for a small value of ϵ (say, $\epsilon = 10^{-3}$) using fast methods to compute the zeta function [12].

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*

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

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:*

$$\begin{array}{lll} D_{\mathrm{KL}}[p_{s_1} : p_{s_2}] & \simeq_{1-\epsilon=0.99} & 0.473\ldots \\ D_{\mathrm{KL}}[p_{s_1} : p_{s_2}] & \simeq_{1-\epsilon=0.999} & 0.482\ldots \\ D_{\mathrm{KL}}[p_{s_1} : p_{s_2}] & \simeq_{1-\epsilon=0.9999} & 0.483\ldots \\ D_{\mathrm{KL}}[p_{s_1} : p_{s_2}] & \simeq_{1-\epsilon=0.99999} & 0.483\ldots \end{array}$$

We can also calculate the KLD $D_{\mathrm{KL}}[p_{s_1}^{\mathcal{X}_1} : p_{s_2}^{\mathcal{X}_2}]$ between two truncated zeta distributions with nested supports $\mathcal{X}_1 \subseteq \mathcal{X}_2$. See [21]. A truncated zeta distribution on the support $\{a, a+1, \ldots, b\} \subset \mathbb{N}$ (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 $\Phi_s(u) = \sum_{x \in \{1,\ldots,u\}} p_s(x) = \frac{1}{\zeta(s)} \sum_{x \in \{1,\ldots,u\}} \frac{1}{x^s}$.

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 α^* maximizing the Chernoff α-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]

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

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 a uniorder exponential family such as the zeta distributions, a closed-form formula for the optimal Chernoff exponent α^* 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 defined by

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

The Sharma–Mittal divergence is induced from the Sharma–Mittal entropies, which unify the extensive Rényi entropies with the non-extensive Tsallis entropies [30]. The

Sharma–Mittal divergences include the Rényi divergences ($\beta \to 1$) and the Tsallis divergences ($\beta \to \alpha$), and in the limit case of $\alpha, \beta \to 1$, 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]:

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

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]: $D_{\mathrm{KL}}[p_{s_1} : p_{s_2}] = B_F^*(\theta_1 : \theta_2) := B_F(\theta_2 : \theta_1)$ (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

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

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 [7]. 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:

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

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 [7] of n independently and identically distributed observations $x_1, \ldots, x_n$ is $\hat{\eta} = \frac{1}{n}\sum_{i=1}^{n} t(x_i)$. Thus we have:

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

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

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

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

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

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

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 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 ζ and the von Mangoldt function Λ as:*

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

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

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

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

$$\begin{aligned} D_{\mathrm{KL}}[p_{s_1} : p_{s_2}] &= \log(\zeta(s_2)) - \sum_{i=1}^{\infty} \frac{1}{i^s \zeta(s)} \log(i^s \zeta(s)) + s_2 \sum_{i=1}^{\infty} \frac{\Lambda(i)}{i^{s_1}}, & (17) \\ &\simeq 0.430495790304827\ldots & (18) \end{aligned}$$

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

$$D_{\mathrm{KL}}[p_s : p_{s+\mathrm{d}s}] \approx \frac{1}{2} I(s) \mathrm{d}s^2, \tag{19}$$

where

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

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

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

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

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

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" in the literature [2].

We can compute the α-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]:

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

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

and we obtain the following closed form for the α-divergences between two Pareto distributions q_{s_1} and q_{s_2}:

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

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

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

The differential entropy of the Pareto distribution q_s is

$$h[q_s] = -\int_1^\infty q_s(x) \log q_s(x) \mathrm{d}x = -F^*(\eta(s)) \tag{27}$$

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

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

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

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

5. Conclusions

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 μ 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).

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Kotz, S.; Balakrishnan, N.; Read, C.; Vidakovic, B. *Encyclopedia of Statistical Sciences*; Wiley: Hoboken, NJ, USA, 2005; Volume 15.
2. Goldstein, M.L.; Morris, S.A.; Yen, G.G. Problems with fitting to the power-law distribution. *Eur. Phys. J. Condens. Matter Complex Syst.* **2004**, *41*, 255–258. [CrossRef]
3. Titchmarsh, E.C.; Heath-Brown, D.R.; Titchmarsh, E.C.T. *The Theory of the Riemann Zeta-Function*; Oxford University Press: Oxford, UK, 1986.
4. Tempesta, P. Group entropies, correlation laws, and zeta functions. *Phys. Rev. E* **2011**, *84*, 021121. [CrossRef] [PubMed]
5. Iwaniec, H. *Lectures on the Riemann Zeta Function*; American Mathematical Society: Providence, RI, USA, 2014; Volume 62.
6. Nielsen, F. A note on some information-theoretic divergences between Zeta distributions. *arXiv* **2021**, arXiv:2104.10548.
7. Barndorff-Nielsen, O. *Information and Exponential Families in Statistical Theory*; John Wiley & Sons: Hoboken, NJ, USA, 2014.
8. Apéry, R. Irrationalité de $\zeta(2)$ et $\zeta(3)$. *Astérisque* **1979**, *61*, 1.
9. Rivoal, T. La fonction zêta de Riemann prend une infinité de valeurs irrationnelles aux entiers impairs. *C. R. de l'Académie des Sci.-Ser. I-Math.* **2000**, *331*, 267–270. [CrossRef]
10. Fischler, S.; Sprang, J.; Zudilin, W. Many odd zeta values are irrational. *Compos. Math.* **2019**, *155*, 938–952. [CrossRef]
11. Graham, R.L.; Knuth, D.E.; Patashnik, O. *Concrete Mathematics: A Foundation for Computer Science*; Addison-Wesley Professional: Boston, MA, USA, 1994.
12. Hiary, G.A. Fast methods to compute the Riemann zeta function. *Ann. Math.* **2011**, *174*, 891–946. [CrossRef]
13. Johansson, F. Rigorous high-precision computation of the Hurwitz zeta function and its derivatives. *Numer. Algorithms* **2015**, *69*, 253–270. [CrossRef]
14. Yildirim, C.Y. A Note on $\zeta''(s)$ and $\zeta'''(s)$. *Proc. Am. Math. Soc.* **1996**, *124*, 2311–2314. [CrossRef]
15. Powers, D.M. Applications and explanations of Zipf's law. In *New Methods in Language Processing and Computational Natural Language Learning*; ACL anthology: Cambridge, MA, USA, 1998 .
16. Mandelbrot, B. *Information Theory and Psycholinguistics: A Theory of Word Frequencies, Readings in Mathematical Social Sciences*; MIT Press: Cambridge, MA, USA, 1966.
17. Lovričević, N.; Pečarić, D.; Pečarić, J. Zipf–Mandelbrot law, f-divergences and the Jensen-type interpolating inequalities. *J. Inequalities Appl.* **2018**, *2018*, 36. [CrossRef]
18. Naldi, M. Approximation of the truncated Zeta distribution and Zipf's law. *arXiv* **2015**, arXiv:1511.01480.
19. Hu, C.Y.; Iksanov, A.M.; Lin, G.D.; Zakusylo, O.K. The Hurwitz zeta distribution. *Aust. N. Z. J. Stat.* **2006**, *48*, 1–6. [CrossRef]
20. Deluca, A.; Corral, Á. Fitting and goodness-of-fit test of non-truncated and truncated power-law distributions. *Acta Geophys.* **2013**, *61*, 1351–1394. [CrossRef]
21. Nielsen, F. Statistical Divergences between Densities of Truncated Exponential Families with Nested Supports: Duo Bregman and Duo Jensen Divergences. *Entropy* **2022**, *24*, 421. [CrossRef] [PubMed]
22. Saito, S.; Tanaka, T. A note on infinite divisibility of zeta distributions. *Appl. Math. Sci.* **2012**, *6*, 1455–1461.
23. 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 Trans. Commun.* **2013**, *62*, 280–292. [CrossRef]
24. Oosawa, T.; Matsuda, T. SQL injection attack detection method using the approximation function of zeta distribution. In Proceedings of the 2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC), San Diego, CA, USA, 5–8 October 2014; pp. 819–824.
25. Doray, L.G.; Luong, A. Quadratic distance estimators for the zeta family. *Insur. Math. Econ.* **1995**, *16*, 255–260. [CrossRef]
26. Jung, H.; Phoa, F.K.H. A Mixture Model of Truncated Zeta Distributions with Applications to Scientific Collaboration Networks. *Entropy* **2021**, *23*, 502. [CrossRef]
27. Cichocki, A.; Amari, S.i. Families of alpha-beta-and gamma-divergences: Flexible and robust measures of similarities. *Entropy* **2010**, *12*, 1532–1568. [CrossRef]
28. Nielsen, F.; Boltz, S. The Burbea-Rao and Bhattacharyya centroids. *IEEE Trans. Inf. Theory* **2011**, *57*, 5455–5466. [CrossRef]
29. Nielsen, F. An information-geometric characterization of Chernoff information. *IEEE Signal Process. Lett.* **2013**, *20*, 269–272. [CrossRef]
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. *J. Phys. Math. Theor.* **2011**, *45*, 032003. [CrossRef]
32. Azoury, K.S.; Warmuth, M.K. Relative loss bounds for on-line density estimation with the exponential family of distributions. *Mach. Learn.* **2001**, *43*, 211–246. [CrossRef]
33. Nielsen, F. On geodesic triangles with right angles in a dually flat space. In *Progress in Information Geometry*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 153–190.

34. Nielsen, F.; Nock, R. Entropies and cross-entropies of exponential families. In Proceedings of the 2010 IEEE International Conference on Image Processing, Washington, DC, USA, 11–12 November 2010; pp. 3621–3624.
35. Walther, A. Anschauliches zur Riemannschen zetafunktion. *Acta Math.* **1926**, *48*, 393–400. [CrossRef]
36. Kawalec, A. The inverse Riemann zeta function. *arXiv* **2021**, arXiv:2106.06915.
37. Weisstein, E.W. *CRC Concise Encyclopedia of Mathematics*; CRC press: Boca Raton, FL, USA, 2002.
38. Stopple, J. Notes on $\log(\zeta(s))''$. *Rocky Mt. J. Math.* **2016**, *46*, 1701–1715.
39. Agostini, D.; Améndola, C. Discrete Gaussian distributions via theta functions. *SIAM J. Appl. Algebra Geom.* **2019**, *3*, 1–30. [CrossRef]
40. Nielsen, F. The Kullback–Leibler Divergence Between Lattice Gaussian Distributions. *J. Indian Inst. Sci.* **2022**, 1–12. [CrossRef]
41. Deconinck, B.; Heil, M.; Bobenko, A.; Van Hoeij, M.; Schmies, M. Computing Riemann theta functions. *Math. Comput.* **2004**, *73*, 1417–1442. [CrossRef]