*J.Natn.Sci.Foundation Sri Lanka 2021 49 (1): 79 - 87*
DOI: http://dx.doi.org/10.4038/jnsfsr.v49i1.9710

**RESEARCH ARTICLE**

# A study on some geometrical and statistical properties of the continuous power-law distribution with a comparison of similar properties of a particular exponential distribution

**N Yapage**
*Department of Mathematics, Faculty of Science, University of Ruhuna, Matara.*

Submitted: 27 December; 2019 Revised: 02 October 2020; Accepted: 23 October 2020

**Abstract:** In literature, one can find diverse applications of the power-law distribution to model naturally occurring phenomena in the sciences. With the emerging field of complex networks this applicability has been observed and emphasised more and more. In the present paper several interesting as well as important properties of this distribution have been explored. First, it is shown that the totality of power-law distributions form a one-parameter family or a statistical model which turns out to be a Riemannian differentiable manifold. Closed form expressions are obtained for several information theoretically important measures such as differential entropy, information divergence and Fisher information which are interpreted to have geometrical and statistical meanings. Next, it is shown that the statistical manifold of power-law distributions forms an exponential family which is a very important aspect in mathematical statistics and information geometry including a number of other fields in the sciences. The parameter estimation problem is addressed using both maximum likelihood and entropy methods. The close relationship of the two in the sense of Statistics is elucidated. Finally, an example of an exponential distribution having the same information divergence and Fisher information as that of power-law distributions is given, thus having the same lower bound in the Cramer-Rao inequality. In this case, an approximate structural similarity can be expected between the two statistical manifolds.

**Keywords:** Exponential family, information geometry, information theory, mathematical statistics, power-law distribution.

## INTRODUCTION

Power laws can be seen very frequently in physics, biology, earth and planetary sciences, economics and finance, computer science and the social sciences. When the probability of measuring a particular value of some quantity varies inversely as a power of that value, the quantity is said to follow a power law. Preliminary introduction to power laws and power-law distributions is available in Newman (2005).

Any network can be represented mathematically with nodes (vertices) and edges. In a network the most important characteristic of a single node is its *degree*. The degree $k_i$ of a node $i$ is usually defined to be the total number of its connections. Thus, the larger the degree, the more important the node is in a network. The average of $k_i$ over all $i$ is called the *average degree* of the network and is denoted by $< k >$. The spread of node degrees over a network is characterised by a distribution function $p(k)$ which is the probability that a randomly selected node has exactly $k$ edges. In particular, for a number of networks, for instance computer networks, the *degree distribution* can be better described by a power law of the form $p(k) \sim k^{(-\gamma)}$, where $\gamma$ is called the exponent of the power law.

* Corresponding author (nihal@maths.ruh.ac.lk; https://orcid.org/0000-0002-1192-2823)

This article is published under the Creative Commons CC-BY-ND License (http://creativecommons.org/licenses/by-nd/4.0/). This license permits use, distribution and reproduction, commercial and non-commercial, provided that the original work is properly cited and is not changed in anyway.

A continuous real random variable $x$ with a power-law distribution has a probability $p(x)$ of taking a value in the interval from $x$ to $x + \mathrm{d}x$, where,

$$p(x) = p(x;\theta) = \mathcal{Z}x^{-\theta} \quad ...(1)$$

with $\theta > 0$. The real parameter $\theta$ is called the exponent of the power law and $\mathcal{Z}$ is the normalisation constant. There must be some lowest value $x_{\min}$ at which the power law is obeyed, and we consider only the real numbers $x$ in the range $x_{\min} \le x < \infty$. The constant $\mathcal{Z}$ in equation (1) is given by the normalisation requirement that,

$$1 = \int_{x_{\min}}^{\infty} p(x;\theta)\,\mathrm{d}x = \mathcal{Z}\int_{x_{\min}}^{\infty} x^{-\theta}\,\mathrm{d}x = \frac{\mathcal{Z}}{1-\theta}\left[x^{-\theta+1}\right]_{x_{\min}}^{\infty}. \quad ...(2)$$

We see immediately that this only makes sense if $\theta > 1$, since otherwise the right-hand side of the equation would diverge; power laws with exponents less than unity cannot be normalised and do not normally occur in nature. If $\theta > 1$ then equation (2) gives,

$$\mathcal{Z} = (\theta - 1)\, x_{\min}^{\theta-1}$$

and the correct normalised expression for the power-law distribution itself is,

$$p(x;\theta) = \frac{(\theta-1)}{x_{\min}}\left(\frac{x}{x_{\min}}\right)^{-\theta}. \quad ...(3)$$

It should be mentioned that power laws also occur in many situations other than the statistical distributions of quantities. For example, Newton's famous $r^{-2}$ law for gravity has a power-law form with exponent $\theta = 2$. While such laws are certainly interesting in their own way, they are not the topic of this paper. Now we outline some examples obeying power laws. It has been shown by several authors that the cumulative distributions of different quantities measured in physical, biological, technological and social systems follow power laws. Some of them are: word frequency, citations of scientific papers, number of hits received by the websites, copies of books sold, telephone calls, magnitude of earthquakes, intensity of wars, wealth of the richest people, frequencies of family names, population of cities etc.

In this section, for completion, we give the values of fundamental characteristic quantities expectation $\mathrm{E}_\theta[x]$, $r^{th}$, moment, and variance $\mathrm{V}ar[x]$ of a random variable $x$ distributed according to a power-law distribution in the form of equation (3) as follows:

$$\mathrm{E}_\theta[x] = \int_{x_{\min}}^{\infty} xp(x;\theta)\mathrm{d}x = \left(\frac{\theta-1}{\theta-2}\right)x_{\min}, \text{ for } \theta > 2$$

$$\mathrm{E}_\theta[x^2] = \int_{x_{\min}}^{\infty} x^2p(x;\theta)\mathrm{d}x = \left(\frac{\theta-1}{\theta-3}\right)x_{\min}^2, \text{ for } \theta > 3$$

$$\mathrm{E}_\theta[x^3] = \int_{x_{\min}}^{\infty} x^3p(x;\theta)\mathrm{d}x = \left(\frac{\theta-1}{\theta-4}\right)x_{\min}^3, \text{ for } \theta > 4 \quad ...(4)$$

Thus, we see the $r^{th}$ moment of a random variable $x$ distributed according to the power-law distribution (3) is given by

$$\mathrm{E}_\theta[x^r] = \int_{x_{\min}}^{\infty} x^r p(x;\theta)\,\mathrm{d}x = \left(\frac{\theta-1}{\theta-(r+1)}\right)x_{\min}^r, \text{ for } \theta > r+1.$$

Moreover, the variance $\mathrm{V}ar(x)$ is obtained as

$$\begin{aligned}\mathrm{V}ar(x) &= \mathrm{E}_\theta[x^2] - \{\mathrm{E}_\theta[x]\}^2 \\ &= \left(\frac{\theta-1}{\theta-3}\right)x_{\min}^2 - \left[\left(\frac{\theta-1}{\theta-2}\right)x_{\min}\right]^2 \\ &= \left[\frac{\theta-1}{\theta-3} - \left(\frac{\theta-1}{\theta-2}\right)^2\right]x_{\min}^2, \text{ for } \theta > 2, \theta > 3.\end{aligned} \quad ...(5)$$

In the case of parameter estimation in the power-law distribution, several authors have tried different methods including maximum likelihood used by Bauke (2007), Clauset *et al*. (2009), and White *et al.* (2008) and entropy-based method used by Fu and Gao (2006). However, in this paper, we estimate the parameter in the power-law distribution using both maximum likelihood and entropy method independently and elucidate the close relationship between the two methods.

## METHODOLOGY

### Definitions of information measures for probability densities

In this section, we define several information measures of general continuous probability distributions and briefly introduce the notion of a statistical model. A discussion on the statistical model of power-law distributions and the information measures can be found at the end.

### Differential entropy

The differential entropy $h(p)$ of a continuous random variable $x$ with density $p(x)$ is defined as,

$$h(p) = -\int_S p(x)\ln p(x)dx,$$

where $S$ is the support set of the random variable.

**Information divergence**

The information divergence (also known as the Kullback-Leibler distance or relative entropy)

$D(p \parallel q)$ between two probability densities $p(x)$ and $q(x)$ on the same set $\mathcal{X}$ is defined as,

$$D(p \parallel q) = \int p(x) \ln \frac{p(x)}{q(x)} dx.$$

Here, $D(p \parallel q)$ is finite only if the support set of $p(x)$ is contained in the support set of $q(x)$. It is used to measure the distance between two probability distributions but is not a metric. The most important fact is that geometrically it is related to the Fisher information matrix (see next section) and plays an important role in maximum likelihood estimation and the exponential family structure.

**Statistical model of probability densities and Fisher information matrix**

Consider a family $\mathcal{S}$ of probability densities on the set $\mathcal{X}(=\mathbb{R}^n)$. Suppose each element of $\mathcal{S}$, a probability density function, may be parameterised using $n$ real-valued variables $(\theta^1, \theta^2, ..., \theta^n)$ so that

$$\mathcal{S} = \{p_\theta = p(x;\theta) | \theta = (\theta^1, \theta^2, ..., \theta^n) \in \Theta\},$$

where $\Theta$ is a subset of $\mathbb{R}^n$ and the mapping $\theta \mapsto p_\theta$ is injective and infinitely differentiable for each $x \in \mathcal{X}$. We call such $\mathcal{S}$ an $n$-dimensional statistical model or a parametric model on $\mathcal{X}$.

Now, given a point $\theta \in \Theta$, the Fisher information matrix of $\mathcal{S}$ at a point $\theta$ is given by the $n \times n$ matrix,

$$G(\theta) := [g_{ij}(\theta)],$$

where the $(i,j)^{\text{th}}$ element $g_{ij}(\theta)$ is defined by the equation,

$$g_{ij}(\theta) = \mathrm{E}_\theta[\partial_i \ell_\theta \partial_j \ell_\theta] = \int \partial_i \ell_\theta \partial_j \ell_\theta p(x;\theta)\, \mathrm{d}x,\ \partial_i := \frac{\partial}{\partial \theta^i},$$

where $\ell_\theta(x) = \ell(x;\theta) = \ln p(x;\theta)$ is the log-likelihood function and $\mathrm{E}_\theta$ denotes the expectation with respect to the distribution $p_\theta$. It should be mentioned that, according to the general theory, the Fisher information matrix turns out to be the unique Riemannian metric on the statistical model [see Amari & Nagaoka (2000) for more information].

## RESULTS AND DISCUSSION

First, we give a short description to the statistical model of power-law distributions and summarise the closed form expressions for differential entropy, information divergence and Fisher information of power-law distribution. The detailed calculations of the information measures with respect to the power-law distribution are given in the Appendix.

Power-law distributions are parameterised using one real parameter $\theta$ as

$$p(x;\theta) = \frac{(\theta-1)}{x_{\min}} \left(\frac{x}{x_{\min}}\right)^{-\theta}$$

and $\mathcal{S} = \{p(x;\theta)\}$ of such distributions becomes a one dimensional statistical model and geometrically this can be viewed as a Riemannian manifold with the natural coordinate system $\theta$ or $-\theta$. The expressions obtained for differential entropy $h(p)$, information divergence $D(p \parallel q)$ and Fisher information matrix $G(\theta)$ of power-law distribution are

$$h(p) = \ln x_{\min} - \ln(\theta - 1) + \frac{\theta}{\theta-1}, \theta > 1 \quad ...(6)$$

$$D(p \parallel q) = \ln\left(\frac{\theta-1}{\beta-1}\right) - \left(\frac{\theta-\beta}{\theta-1}\right),\ \beta, \theta > 1 \text{ and} \quad ...(7)$$

$$G(\theta) = \frac{1}{(\theta-1)^2}, \theta > 1 \quad ...(8)$$

respectively.

**Exponential family structure of the manifold of power-law distributions**

First, we define the notion of an exponential family in mathematical statistics following Amari and Nagaoka (2000). Let $\mathcal{X}$ be a finite set or, more generally, a measurable space with an underlying measure $\mathrm{d}\mu$. We denote the set of positive probability distributions (probability mass functions for a finite $\mathcal{X}$ and probability density functions for a general $(\mathcal{X}, \mathrm{d}\mu)$) on by $\mathcal{P} = \mathcal{P}(\mathcal{X})$. When a family of distributions, say

$$\mathcal{M} = \{p_\theta \mid \theta = (\theta^i);\ i = 1, ..., n\} \subset \mathcal{P}, \quad ...(9)$$

is represented in the form,

$$p_\theta(x) = \exp\{c(x) + \sum_{i=1}^{n} \theta^i f_i(x) - \psi(\theta)\}, \quad x \in \mathcal{X}, \qquad ...(10)$$

$\mathcal{M}$ is called an exponential family. Here, $\theta^i;\ i = 1, ..., n$ are $\mathbb{R}$-valued parameters, $c$ and $f_i$ are functions on $\mathcal{X}$ and $\psi(\theta)$ is a real-valued convex function. Further, we assume that the correspondence $\theta \mapsto p_\theta$ is one-to-one if and only if the $n+1$ functions $\{f_1, ..., f_n, 1\}$ are linearly independent, where 1 denotes the constant function which identically takes the value 1.

These $\theta = (\theta^i)$ are called the natural coordinates of $\mathcal{M}$.

Now, for the exponential family $\mathcal{M}$, if we let

$$\eta_i(\theta) \overset{def}{=} \mathbb{E}_\theta[f_i] = \sum_{\boldsymbol{x}} p_\theta(\boldsymbol{x}) f_i(\boldsymbol{x})$$

then $\eta = (\eta_i)$ and $\theta = (\theta^i)$ are in one-to-one correspondence. That is, we can also use $\eta$ instead of $\theta$ to specify an element of $\mathcal{M}$. These $(\eta_i)$ are called the expectation coordinates of $\mathcal{M}$. The expectation coordinates are, in general, represented as,

$$\eta_i = \partial_i \psi(\theta) \qquad (\partial_i \overset{def}{=} \frac{\partial}{\partial \theta^i}). \qquad ...(11)$$

We now consider the power-law distributions in the form,

$$p(x;\theta) = \frac{\theta-1}{x_{\min}} \left(\frac{x}{x_{\min}}\right)^{-\theta}.$$

Let us next rewrite $p(x;\theta)$ as,

$$p(x;\theta) = \exp[\ln(\theta-1) - \ln x_{\min} - \theta \ln x + \theta \ln x_{\min}] \qquad ...(12)$$

$$= \exp[-\ln x_{\min} - \theta \ln x + \ln(\theta-1) + \theta \ln x_{\min}]. \qquad ...(13)$$

Comparing equations (10) and (13), we obtain

$$f(x) = \ln x,\ \psi(\theta) = -\ln(\theta-1) - \theta \ln x_{\min},\ c(x) = -\ln x_{\min}$$

and the natural parameter is $-\theta$. So we see that the manifold of power-law distributions forms an exponential family. The expectation parameter is obtained by taking the expectation of the sufficient statistic and is given by

$$\eta(\theta) \overset{def}{=} \mathbb{E}_\theta[\ln x] = \frac{\partial}{\partial \theta}\psi(\theta)$$

$$= \frac{\partial}{\partial \theta}(-\ln(\theta-1) - \theta \ln x_{\min})$$

$$= -\frac{1}{\theta-1} - \ln x_{\min}. \qquad ...(14)$$

It should be mentioned that we may also consider

$$\frac{1}{\theta-1} + \ln x_{\min}$$

as an expectation parameter.

Thus, on the manifold of power-law distributions, we consider $\theta$ and $(\theta-1)^{-1} + \ln x_{\min}$ or $(\theta-1)^{-1}$ as natural and expectation coordinates, respectively. It should also be mentioned that the Fisher information defines the unique metric on the manifold of power-law distributions.

## Parameter estimation

### Maximum likelihood method

We apply the maximum likelihood method to estimate the parameter in the power-law distribution given by equation (3). The logarithm of the likelihood function, $L(\theta)$ for a sample $x_1, x_2, ..., x_n$ obtained from a power-law distribution is given by,

$$\ln L(\theta) = n\ln(\theta-1) - n\ln x_{\min} - \theta \sum_{i=1}^{n} \ln x_i + \theta n \ln x_{\min} \qquad ...(15)$$

Now differentiating equation (15) with respect to $\theta$ and equating to zero, we get

$$\hat{\theta} = 1 + \frac{1}{\frac{1}{n}\sum_{i=1}^{n} \ln x_i - \ln x_{\min}},$$

where $\hat{\theta}$ is the maximum likelihood estimator of $\theta$. Since, from the weak law of large numbers in probability theory,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} x_i$$

approaches $\mathrm{E}_\theta[x]$, where $\mathrm{E}_\theta$ denotes expectation with respect to $p(x;\theta)$, it follows that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \ln x_i$$

approaches the expectation of $\ln x$, that is $\mathrm{E}_\theta[\ln x]$, so we rewrite $\hat{\theta}$ as

$$\hat{\theta} = 1 + \frac{1}{\mathrm{E}_\theta[\ln x] - \ln x_{\min}} \qquad ...(16)$$

It is well-known that according to the Cramer-Rao inequality, the variance of any unbiased estimator is at least $(nG(\theta))^{-1}$, where $G(\theta)$ is the Fisher information matrix. Hence, we have

$$Var[\hat{\theta}] \geq \frac{1}{n} G^{-1}(\theta).$$

The maximum likelihood estimator attains this lower bound, that is, equality holds. Thus, in the case of power-law distributions,

$$Var[\hat{\theta}] = \frac{1}{n}(\theta - 1)^2.$$

**Entropy method**

From equation (6) and

$$\mathrm{E}_\theta[\ln x] = \ln x_{\min} + \frac{1}{\theta - 1},$$

we get,

$$h(p) = \mathrm{E}_\theta[\ln x] - \ln(\theta - 1) + 1.$$

So we get,

$$\theta = 1 + \exp\{\mathrm{E}_\theta[\ln x] - h(p) + 1\}$$

and if we know entropy, the parameter $\theta$ in the power-law distribution can be estimated. If we substitute for $\mathrm{E}_\theta[\ln x]$ in equation (16), then the result turns out to be $\hat{\theta} = \theta$.

**A particular exponential distribution having similar properties**

We identify that a particular exponential family having several properties similar to that of power-law distribution. Let us now consider the exponential distribution

$$f(x;\theta) = (\theta - 1)\exp\{-(\theta - 1)x\}, (\theta - 1) > 0. \quad ...(17)$$

The differential entropy of equation (17) and information divergence of and can be calculated as,

$$h(f) = 1 - \ln(\theta - 1),\ \theta > 1 \text{ and} \quad ...(18)$$

$$D(f \parallel f') = \ln\left(\frac{\theta-1}{\beta-1}\right) - \left(\frac{\theta-\beta}{\theta-1}\right), \theta, \beta > 1 \quad ...(19)$$

respectively. We observe that although the differential entropies are different the information divergences of both distributions are same. In particular, the totality of exponential distributions of the form equation (17) turns out to be an exponential family.

For completion, we calculate the Fisher information matrix of $f(x;\theta)$. The log-likelihood function is written as

$$l_\theta(x) = \ln f(x;\theta) = \ln(\theta - 1) - (\theta - 1)x$$

Now differentiating with respect to $\theta$., we get

$$\frac{\mathrm{d}}{\mathrm{d}\theta} l_\theta(x) = \frac{1}{\theta - 1} - x$$

The Fisher information matrix is thus written as,

$$G(\theta) = \mathrm{E}_\theta\left[\left(\frac{1}{\theta-1} - x\right)^2\right]$$

$$= \left(\frac{1}{\theta-1}\right)^2 - \left(\frac{2}{\theta-1}\right)\mathrm{E}_\theta[x] + \mathrm{E}_\theta[x^2], \quad ...(20)$$

where $\mathrm{E}_\theta$ denotes the expectation with respect to the distribution $f(x;\theta)$. We now calculate $\mathrm{E}_\theta[x]$ and $\mathrm{E}_\theta[x^2]$ using integration by parts as,

$$\mathrm{E}_\theta[x] = \int_0^\infty x f(x;\theta)\,\mathrm{d}x = \frac{1}{\theta-1} \quad ...(21)$$

$$\mathrm{E}_\theta[x^2] = \int_0^\infty x f(x;\theta)\,\mathrm{d}x = \frac{2}{(\theta-1)^2}. \quad ...(22)$$

Thus, substituting equations (21) and (22) in equation (20), we get the Fisher information matrix for the given exponential distribution as,

$$\frac{1}{(\theta-1)^2}$$

which is the same as the one we obtained for power-law distribution.

It seems that the maximum likelihood estimator of the parameter in our exponential distribution has an interesting similarity as that of the power-law distribution. So we apply the maximum likelihood method to estimate the parameter in the given exponential distribution

$$f(x;\theta) = (\theta - 1)\exp\{-(\theta - 1)x\}, (\theta - 1) > 0.$$

The likelihood function $L(\theta)$ for a sample $x_1, x_2, ..., x_n$ obtained from the given exponential distribution is given by,

The log-likelihood function is now written as,

$$\ln L(\theta) = \ln\prod_{i=1}^{n}(\theta - 1)\exp\{-(\theta - 1)x\}$$

$$= \sum_{i=1}^{n}\ln(\theta - 1)\exp\{-(\theta - 1)x\}$$

$$= \sum_{i=1}^{n}[\ln(\theta - 1) - (\theta - 1)x_i]$$

$$= n\ln(\theta - 1) - (\theta - 1)\sum_{i=1}^{n} x_i. \quad ...(23)$$

Now differentiating with respect to $\theta$ and equating to zero, we get,

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\ln L(\theta) = \frac{n}{\hat{\theta}-1} - \sum_{i=1}^{n} x_i = 0,$$

where $\hat{\theta}$ is the maximum likelihood estimator of . Thus we get,

$$\frac{n}{\hat{\theta}-1} = \sum_{i=1}^{n} x_i$$

from which we obtain $\hat{\theta}$ as,

$$\hat{\theta} = 1 + \frac{1}{\frac{1}{n}\sum_{i=1}^{n} x_i}$$

Since, from the weak law of large numbers in probability theory,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} x_i$$

approaches $\mathrm{E}_\theta[x]$, where $\mathrm{E}_\theta$ denotes expectation with respect to $f(x;\theta)$, we can write $\hat{\theta}$ as,

$$\hat{\theta} = 1 + \frac{1}{\mathrm{E}_\theta[x]}.$$

Since $\mathrm{E}_\theta[x] = (\theta-1)^{-1}$, we obtain $\hat{\theta} = \theta$ as in the case of power-law distributions.

Finally, we have the same lower bound in the Cramer-Rao inequality as the Fisher information matrix is same in both cases.

From equation (18), we get

$$\theta = 1 + \exp\{-h(f) + 1\}$$

and if we know entropy, the parameter $\theta$ in the exponential distribution can be estimated.

We derived after fairly lengthy calculations the expressions for differential entropy, information divergence and Fisher information of power-law distribution. It was shown that the totality of power-law distributions form a statistical manifold and it turns out to be an exponential family. We estimated the parameter in the power-law distribution using both maximum likelihood and entropy methods. Moreover, entropy will give a clue to understand the complexity of the network concerned. Divergence will be helpful in comparing degree distributions of two complex networks.

## CONCLUSION

We observed a particular exponential distribution having the same Fisher information matrix and thus the same lower bound in the Cramer-Rao inequality in both cases. On the other hand, Fisher information will be useful in studying the geometrical structure of a network from an information geometrical point of view and also in parameter estimation in complex networks.

We also hope that the same Fisher information for power-law and exponential distributions we have obtained will definitely be helpful to understand the geometry of such networks.

## REFERENCES

Amari S. & Nagaoka H. (2000). *Methods of Information Geometry.* American Mathematical Society, Oxford University Press, UK.

Bauke H. (2007). Parameter estimation for power-law distributions by maximum likelihood methods. *The European Physical Journal B* **58**: 167–173. DOI: https://doi.org/10.1140/epjb/e2007-00219-y

Clauset A., Shalizi C.R. & Newman M.E.J. (2009). Power-law distributions in empirical data. *Society for Industrial and Applied Mathematics Review* **51**(4): 661–703. DOI: https://doi.org/10.1137/070710111

Fu B.B. & Gao Z.Y. (2006). An estimation for the power-law distribution parameter based on entropy. *Chinese Physics Letters* **23**: 520–522. DOI: https://doi.org/10.1088/0256-307X/23/2/068

Newman M.E.J. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary Physics* **46**: 323–351. DOI: https://doi.org/10.1080/00107510500052444

White E.P., Enquist B.J. & Green J.L. (2008). On estimating the exponent of power-law frequency distributions. *Ecology* **89**(4): 905–912. DOI: https://doi.org/10.1890/07-1288.1

Yapage N. (2013). Some Information Measures of Power-law Distributions. *Proceedings of (Ruhuna International Science and Technology Conference) RISTCON 2013*, Faculty of Science, University of Ruhuna, Matara, Sri Lanka, pp. 53–54.

## APPENDIX

In this Appendix, we give detailed calculations of differential entropy, information divergence and Fisher information matrix of the power-law distribution following Yapage (2013).

***Differential entropy:***

From the definition of differential entropy, we get,

$$
\begin{aligned}
h(p) &= -\int_{x_{\min}}^{\infty} p(x;\theta)\ln p(x;\theta)\, dx \\
&= -\int_{x_{\min}}^{\infty} \frac{(\theta-1)}{x_{\min}}\left(\frac{x}{x_{\min}}\right)^{-\theta} \ln \frac{(\theta-1)}{x_{\min}}\left(\frac{x}{x_{\min}}\right)^{-\theta} \mathrm{d}x \\
&= -\int_{x_{\min}}^{\infty} \frac{(\theta-1)}{x_{\min}^{(-\theta+1)}} x^{-\theta}\left[(\ln(\theta-1)-\ln x_{\min}) - \theta \ln\left(\frac{x}{x_{\min}}\right)\right] \mathrm{d}x \\
&= -\int_{x_{\min}}^{\infty} \frac{(\theta-1)}{x_{\min}^{(-\theta+1)}} x^{-\theta}[(\ln(\theta-1)-\ln x_{\min}) - \theta(\ln x - \ln x_{\min})]\, \mathrm{d}x \\
&= -\frac{(\theta-1)}{x_{\min}^{(-\theta+1)}}[\ln(\theta-1)-\ln x_{\min}] \int_{x_{\min}}^{\infty} x^{-\theta}\, \mathrm{d}x \\
&\quad + \frac{\theta(\theta-1)}{x_{\min}^{(-\theta+1)}} \int_{x_{\min}}^{\infty} x^{-\theta} \ln x\, \mathrm{d}x - \frac{\theta(\theta-1)}{x_{\min}^{(-\theta+1)}} \ln x_{\min} \int_{x_{\min}}^{\infty} x^{-\theta}\, \mathrm{d}x \\
&= -\frac{(\theta-1)}{x_{\min}^{(-\theta+1)}}[(\ln(\theta-1)-\ln x_{\min})]\left[\frac{x^{(-\theta+1)}}{(-\theta+1)}\right]_{x_{\min}}^{\infty} \\
&\quad + \frac{\theta(\theta-1)}{x_{\min}^{(-\theta+1)}}\left[\frac{-x_{\min}^{(-\theta+1)}}{(-\theta+1)}\ln x_{\min} + \frac{1}{(-\theta+1)}\frac{x_{\min}^{(-\theta+1)}}{(-\theta+1)}\right] \\
&\quad - \frac{\theta(\theta-1)}{x_{\min}^{(-\theta+1)}} \ln x_{\min}\left[\frac{x^{(-\theta+1)}}{(-\theta+1)}\right]_{x_{\min}}^{\infty} \\
&= -\frac{(\theta-1)}{x_{\min}^{(-\theta+1)}}[(\ln(\theta-1)-\ln x_{\min})]\left[\frac{x^{(-\theta+1)}}{(-\theta+1)}\right] \\
&\quad + \frac{\theta(\theta-1)}{x_{\min}^{(-\theta+1)}}\left[\frac{-x_{\min}^{(-\theta+1)}}{(-\theta+1)}\ln x_{\min} + \frac{1}{(-\theta+1)}\frac{x_{\min}^{(-\theta+1)}}{(-\theta+1)}\right] \\
&\quad - \frac{\theta(\theta-1)}{x_{\min}^{(-\theta+1)}} \ln x_{\min}\left[-\frac{x_{\min}^{(-\theta+1)}}{(-\theta+1)}\right] \\
&= -\frac{(\theta-1)}{x_{\min}^{(-\theta+1)}}[\ln(\theta-1)-\ln x_{\min}]\left[-\frac{x_{\min}^{(-\theta+1)}}{(-\theta+1)}\right] \\
&\quad + \theta\frac{(\theta-1)}{(\theta-1)}\ln x_{\min} + \frac{\theta}{(\theta-1)} + (-\theta \ln x_{\min}) \\
&= \ln x_{\min} - \ln(\theta-1) + \theta \ln x_{\min} + \frac{\theta}{\theta-1} - \theta \ln x_{\min} \\
&= \ln x_{\min} - \ln(\theta-1) + \frac{\theta}{(\theta-1)}.
\end{aligned}
\tag{A.1}
$$

We now evaluate the following integral which was used in the above calculation [see (A.1)]. Using integration by parts formula

$$\int v\,\mathrm{d}u = uv - \int u\,\mathrm{d}v$$

with $v = \ln x$ and $\mathrm{d}u = x^{-\theta}\,\mathrm{d}x$, we get,

$$\int_{x_{\min}}^{\infty} \ln x\, x^{-\theta}\,\mathrm{d}x = \left[\frac{x^{(-\theta+1)}}{(-\theta+1)}\ln x\right]_{x_{\min}}^{\infty} - \int_{x_{\min}}^{\infty} \frac{x^{(-\theta+1)}}{(-\theta+1)}\,\frac{1}{x}\,\mathrm{d}x$$

$$= \left[\frac{x^{(-\theta+1)}}{(-\theta+1)}\ln x\right]_{x_{\min}}^{\infty} - \frac{1}{1-\theta}\int_{x_{\min}}^{\infty} x^{-\theta}\,\mathrm{d}x$$

$$= \left[\frac{x^{(-\theta+1)}}{(-\theta+1)}\ln x\right]_{x_{\min}}^{\infty} - \frac{1}{(-\theta+1)}\left[\frac{x^{(-\theta+1)}}{(-\theta+1)}\right]_{x_{\min}}^{\infty}$$

$$= -\frac{x_{\min}^{(-\theta+1)}}{(-\theta+1)}\ln x_{\min} + \frac{1}{(-\theta+1)}\frac{x_{\min}^{(-\theta+1)}}{(-\theta+1)} \tag{A.2}$$

***Information divergence:***

We consider two power-law distributions $p(x;\theta)$ and $q(x;\beta)$ and calculate the information divergence. From the definition of information divergence, we have,

$$D(p \parallel q) = \int_{x_{\min}}^{\infty} p(x)\ln\frac{p(x;\theta)}{q(x;\beta)}\,\mathrm{d}x$$

$$= \int_{x_{\min}}^{\infty} p(x)\{\ln p(x;\theta) - q(x;\beta)\}\,\mathrm{d}x$$

$$= \int_{x_{\min}}^{\infty} \frac{(\theta-1)}{x_{\min}}\left(\frac{x}{x_{\min}}\right)^{-\theta}$$

$$\left(\ln\frac{(\theta-1)}{x_{\min}}\left(\frac{x}{x_{\min}}\right)^{-\theta} - \ln\frac{(\beta-1)}{x_{\min}}\left(\frac{x}{x_{\min}}\right)^{-\beta}\right)\mathrm{d}x \tag{A.3}$$

Consider

$$\ln\frac{(\theta-1)}{x_{\min}}\left(\frac{x}{x_{\min}}\right)^{-\theta} - \ln\frac{(\beta-1)}{x_{\min}}\left(\frac{x}{x_{\min}}\right)^{-\beta}$$

$$= \ln(\theta-1) - \ln x_{\min} - \theta\ln\frac{x}{x_{\min}} - \ln(\beta-1) + \ln x_{\min}$$

$$+ \beta\ln\frac{x}{x_{\min}}$$

$$= \ln\left(\frac{\theta-1}{\beta-1}\right) - \theta\ln x + \theta\ln x_{\min} + \beta\ln x - \beta\ln x_{\min}$$

$$= \ln\left(\frac{\theta-1}{\beta-1}\right) - (\theta-\beta)\ln x + (\theta-\beta)\ln x_{\min} \tag{A.4}$$

Now substituting (A.4) in (A.3), we get

$$D(p \parallel q) = \int_{x_{\min}}^{\infty} \frac{(\theta-1)}{x_{\min}^{(-\theta+1)}} x^{-\theta}\left[\ln\left(\frac{\theta-1}{\beta-1}\right) - (\theta-\beta)\ln x\right.$$

$$\left. + (\theta-\beta)\ln x_{\min}\right]\mathrm{d}x$$

$$= \left[\ln\left(\frac{\theta-1}{\beta-1}\right) + (\theta-\beta)\ln x_{\min}\right]\frac{(\theta-1)}{x_{\min}^{(-\theta+1)}}\int_{x_{\min}}^{\infty} x^{-\theta}\,\mathrm{d}x$$

$$- (\theta-\beta)\frac{(\theta-1)}{x_{\min}^{(-\theta+1)}}\int_{x_{\min}}^{\infty} x^{-\theta}\ln x\,\mathrm{d}x$$

$$= \left[\ln\left(\frac{\theta-1}{\beta-1}\right) + (\theta-\beta)\ln x_{\min}\right]\frac{(\theta-1)}{x_{\min}^{(-\theta+1)}}\left[\frac{x^{(-\theta+1)}}{(-\theta+1)}\right]_{x_{\min}}^{\infty}$$

$$- (\theta-\beta)\frac{(\theta-1)}{x_{\min}^{(-\theta+1)}}\left[\frac{x_{\min}^{(-\theta+1)}}{(\theta-1)}\ln x_{\min} + \frac{1}{(\theta-1)^2}x_{\min}^{(-\theta+1)}\right]$$

$$= \left[\ln\left(\frac{\theta-1}{\beta-1}\right) + (\theta-\beta)\ln x_{\min}\right]\frac{(\theta-1)}{x_{\min}^{(-\theta+1)}}\frac{x_{\min}^{(-\theta+1)}}{(\theta-1)}$$

$$- (\theta-\beta)\ln x_{\min} - \left(\frac{\theta-\beta}{\theta-1}\right)$$

$$= \ln\left(\frac{\theta-1}{\beta-1}\right) - \left(\frac{\theta-\beta}{\theta-1}\right).$$

***Fisher information matrix:***

We consider the power-law distributions

$$p_\theta(x) := p(x;\theta) = \frac{\theta-1}{x_{\min}}\left(\frac{x}{x_{\min}}\right)^{-\theta}.$$

The log-likelihood function is written as,

$$l_\theta(x) = \ln p_\theta(x) = \ln(\theta-1) - \ln(x_{\min}) - \theta\ln\left(\frac{x}{x_{\min}}\right)$$

Now differentiating $l_\theta(x)$ with respect to $\theta$, we get,

$$\frac{\mathrm{d}}{\mathrm{d}\theta}l_\theta(x) = \frac{1}{\theta-1} - \ln\left(\frac{x}{x_{\min}}\right)$$

The Fisher information matrix is thus written as,

$$G(\theta) = \mathrm{E}_\theta\left[\left(\frac{1}{\theta-1} - \ln\left(\frac{x}{x_{\min}}\right)\right)^2\right]$$

$$= \mathrm{E}_\theta\left[\left(\frac{1}{\theta-1}\right)^2\right] - \mathrm{E}_\theta\left[\left(\frac{2}{\theta-1}\right)(\ln X - \ln x_{\min})\right]$$

$$+ \mathrm{E}_\theta[(\ln X - \ln x_{\min})^2]$$

$$= \left(\frac{1}{\theta-1}\right)^2 + \left(\frac{2}{\theta-1}\right)\ln x_{\min} - \left(\frac{2}{\theta-1}\right)\mathrm{E}_\theta[\ln x] + \mathrm{E}_\theta[(\ln x)^2]$$

$$- 2\ln x_{\min}\mathrm{E}_\theta[\ln x] + (\ln x_{\min})^2, \tag{A.5}$$

where $\mathrm{E}_\theta$ denotes the expectation with respect to the distribution $p_\theta$.

Thus, the Fisher information matrix is obtained as,

$$G(\theta) = \frac{1}{(\theta-1)^2} + \frac{2}{\theta-1}\ln x_{\min} - \frac{2}{\theta-1}\left[\ln x_{\min} + \frac{1}{\theta-1}\right]$$

$$+ (\ln x_{\min})^2 + \frac{2}{\theta-1}\ln x_{\min} + \frac{2}{(\theta-1)^2}$$

$$- 2\ln x_{\min}\left[\ln x_{\min} + \frac{1}{\theta-1}\right] + (\ln x_{\min})^2$$

$$= \frac{1}{(\theta-1)^2}$$

We calculate $\mathrm{E}_\theta[\ln x]$ and $\mathrm{E}_\theta[(\ln x)^2]$ below which were used to derive the above expression for Fisher information matrx. We know that the expectation of the random variable $x$ with respect to $p(x;\theta)$ is defined as,

$$\mathrm{E}_\theta[x] = \int_{x_{\min}}^{\infty} x\, p(x;\theta)\, \mathrm{d}x.$$

Corresponding to this definition, we write,

$$\mathrm{E}_\theta[\ln x] = \int_{x_{\min}}^{\infty} (\ln x)\, p(x;\theta)\, \mathrm{d}x$$

$$= \int_{x_{\min}}^{\infty} (\ln x) \frac{(\theta-1)}{x_{\min}} \left(\frac{x}{x_{\min}}\right)^{-\theta} \mathrm{d}x$$

$$= \frac{(\theta-1)}{x_{\min} x_{\min}^{-\theta}} \int_{x_{\min}}^{\infty} x^{-\theta} \ln x\, \mathrm{d}x \tag{A.6}$$

Using the result in equation (25), we get

$$\mathrm{E}_\theta[\ln x] = \frac{\theta-1}{x_{\min}^{-\theta+1}}(-1)\frac{x_{\min}^{-\theta+1}}{-\theta+1}\ln x_{\min} + \frac{\theta-1}{x_{\min}^{-\theta+1}}\frac{x_{\min}^{-\theta+1}}{(-\theta+1)(-\theta+1)}$$

$$= \ln x_{\min} - \frac{(-\theta+1)}{(-\theta+1)(-\theta+1)}$$

$$= \ln x_{\min} + \frac{1}{\theta-1}. \tag{A.7}$$

Next we calculate $\mathrm{E}_\theta[(\ln x)^2]$:

$$\mathrm{E}_\theta[\ln x] = \int_{x_{\min}}^{\infty} (\ln x)\, p(x;\theta)\, \mathrm{d}x$$

$$= \int_{x_{\min}}^{\infty} (\ln x) \frac{(\theta-1)}{x_{\min}} \left(\frac{x}{x_{\min}}\right)^{-\theta} \mathrm{d}x$$

$$= \frac{(\theta-1)}{x_{\min} x_{\min}^{-\theta}} \int_{x_{\min}}^{\infty} x^{-\theta} \ln x\, \mathrm{d}x \tag{A.8}$$

Now using the integration by parts formula

$$\int v\, \mathrm{d}u = uv - \int u\, \mathrm{d}v$$

with $v = (\ln x)^2$ and $\mathrm{d}u = x^{-\theta}\mathrm{d}x$ we get,

$$\int_{x_{\min}}^{\infty} (\ln x)^2 x^{-\theta}\, \mathrm{d}x = \frac{x^{-\theta+1}}{-\theta+1}(\ln x)^2 - \int_{x_{\min}}^{\infty} \frac{x^{-\theta+1}}{-\theta+1}\, 2\ln x\, \frac{1}{x}\, \mathrm{d}x$$

$$= \left[\frac{x^{-\theta+1}}{-\theta+1}(\ln x)^2\right]_{x_{\min}}^{\infty} - \frac{2}{-\theta+1}\int_{x_{\min}}^{\infty} x^{-\theta}\ln x\, \mathrm{d}x$$

$$= \frac{-x_{\min}^{-\theta+1}}{-\theta+1}(\ln x_{\min})^2$$

$$- \frac{2}{-\theta+1}\left[\frac{-x_{\min}^{-\theta+1}}{-\theta+1}\ln x_{\min} + \frac{1}{-\theta+1}\frac{x_{\min}^{-\theta+1}}{-\theta+1}\right] \tag{A.9}$$

Therefore, substituting equation (32) in equation (31) and simplifying, we get

$$\mathrm{E}_\theta[(\ln x)^2] = (\ln x_{\min})^2 + \frac{2\ln x_{\min}}{\theta-1} + \frac{2}{(\theta-1)^2}.$$