Some Information measures of Power-law Distributions

N Yapage*
Department of Mathematics
University of Ruhuna
Matara
Sri Lanka

Abstract

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. Power laws can be seen very frequently in physics, biology, earth and planetary sciences, economics and finance, computer science and the social sciences.

In this paper, we calculate several important information measures of power-law distributions with continuous random variables such as differential entropy, information divergence and Fisher information.

Keywords: power-law distributions, differential entropy, information divergence and Fisher information

*nihal@maths.ruh.ac.lk

Some Information measures of Power-law Distributions

Abstract

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. Power laws can be seen very frequently in physics, biology, earth and planetary sciences, economics and finance, computer science and the social sciences.

In this paper, we calculate several important information measures of power-law distributions with continuous random variables such as differential entropy, information divergence and Fisher information.

1 Introduction

A continuous real random variable with a power-law distribution [1] has a probability $p(x)dx$ of taking a value in the interval from x to $x + dx$, where

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

with $\theta > 0$. The real parameter θ is called the exponent of the power law and $\mathcal{Z}$ is the normalization 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 $x_{\min} \leq x < \infty$. The constant $\mathcal{Z}$ in (1) is given by the normalization requirement that

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

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 normalized and do not normally occur in nature. If $\theta > 1$ then (2) gives

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

and the correct normalized expression for the power law itself is

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

It should be mentioned that power laws also occur in many situations other than the statistical distributions of quantities. For example, Newton's famous $\frac{1}{r^2}$ 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 web sites, Copies of books sold, Telephone calls, Magnitude of earthquakes, Intensity of wars, Wealth of the richest people, frequencies of family names, population of cities.

2 Information measures of probability distributions

2.1 Diffrential 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.

2.2 Information divergence

The information divergence (a.k.a. Kullback-Leibler distance or relative entropy) $D(p\|q)$ between two probability densities $p(x)$ and $q(x)$ on the same set $\mathcal{X}$ is defined as

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

Here $D(p\|q)$ is finite only if the support set of $p(x)$ is contained in the support set of $q(x)$.

2.3 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 parameterized using n real-valued variables $(\theta^1, \theta^2, \ldots, \theta^n)$ so that

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

where Θ 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 [2] on $\mathcal{X}$.

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

$$g_{ij}(\theta) = E_\theta[\partial_i \ell_\theta \partial_j \ell_\theta] = \int \partial_i \ell_\theta \partial_j \ell_\theta p(x : \theta) dx,$$

where $\partial_i := \frac{\partial}{\partial \theta^i}$, the log-likelihood function $\ell_\theta(x) = \ell(x; \theta) = \ln p(x; \theta)$ and E_θ denotes the expectation with respect to the distribution p_θ.

3 Calculation of information quantities of power-law distributions

3.1 Statistical model of power-law distributions

Power-law distributions are parameterized using one real parameter θ 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 an one dimensional statistical model and geometrically this can be viewed as a Riemannian manifold with the coordinate system θ.

3.2 Calculation of diiffrential entropy, information divergence and Fisher information

The expressions obtained for diiffrential entropy, information divergence and Fisher information are

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

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

and

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

respectively. The detail calculations are given in the appendix.

4 Conclusion & Discussion

We have derived after fairly lengthy calculations the expressions for diiffrential entropy, information divergence and Fisher information of power-law distributions and hope that these reults would be useful in studying other characteristics of power-law distributions from information theoretic or information geometric point of view.

References

[1] M. E. J. Newman. *Power laws, Pareto distributions and Zipf's law.* Contemporary Physics 46, 323-351, 2005.

[2] S. Amari and H. Nagaoka. *Methods of Information Geometry.* American Mathematical Society and Oxford University Press, 2000.

Appendix

Differential entropy

Let

$h(p) = -\int p(x;\theta)\ln p(x;\theta)\,dx$

$= -\int \frac{\theta-1}{x_{min}}\left(\frac{x}{x_{min}}\right)^{-\theta}\ln\frac{\theta-1}{x_{min}}\left(\frac{x}{x_{min}}\right)^{-\theta}\,dx$

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

$= -\frac{\theta-1}{(x_{min})^{(-\theta+1)}}[\ln(\theta-1)-\ln x_{min}]\int x^{-\theta}\,dx$

$+\frac{\theta-1}{(x_{min})^{(-\theta+1)}}\theta\int x^{-\theta}\ln x\,dx-\frac{\theta-1}{(x_{min})^{(-\theta+1)}}\ln x_{min}\int x^{-\theta}\,dx$

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

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

Since

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

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

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

Information divergence

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

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

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

Consider

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

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

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

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

$D(p \parallel q) = \int \frac{\theta - 1}{x_{min}^{-\theta+1}} x^{-\theta} [\ln \left(\frac{\theta - 1}{\beta - 1} \right) - (\theta - \beta) \ln x + (\theta - \beta) \ln x_{min}] \, dx$

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

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

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

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

Fisher information

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

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

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

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

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

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

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

$E_\theta[(\ln X)^2] = \int_{x_{min}}^{\infty} (\ln x)^2 p(x;\theta)\,dx$

$= \int (\ln x)^2 \frac{\theta-1}{x_{min}} \left(\frac{x}{x_{min}}\right)^{-\theta} \,dx$

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

By substituting $v = (\ln x)^2$ and $du = x^{-\theta}dx$ we get,

$\frac{dv}{dx} = 2.\frac{1}{x}\ln x$ and $u = \frac{x^{-\theta+1}}{-\theta+1}$,

Then,

$\int (\ln x)^2 x^{-\theta} dx = \frac{x^{-\theta+1}}{-\theta+1}(\ln x)^2 - \int \frac{x^{-\theta+1}}{-\theta+1}.2\ln x.\frac{1}{x}\,dx$

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

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

Therefore ,

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

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

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

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

We have used following results to derive the above expression for Fisher information. $\langle x\rangle = \int_{x_{min}}^{\infty} xp(x)\,dx$

$\langle \ln x\rangle = \int_{x_{min}}^{\infty} (\ln x)p(x)\,dx$

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

$= \frac{\theta-1}{x_{min}x_{min}^{-\theta}}\int_{x_{min}}^{\infty} x^{-\theta}\ln x\,dx$

$\int v\,du = uv - \int u\,dv$

$v = lnx$ and $du = x^{-\theta}dx$

$\frac{d}{dx}v = \frac{1}{x}$ and $u = \frac{x^{-\theta+1}}{-\theta+1}$

$\int lnx.x^{-\theta}\,dx = \frac{x^{-\theta+1}}{\theta+1}.lnx - \int \frac{x^{-\theta+1}}{-\theta+1}.\frac{1}{x}\,dx - \frac{1}{1-\theta}\int x^{-\theta}\,dx$

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

$= 0 - \frac{x_{min}}{-\theta+1}\ln x_{min} + \frac{1}{-\theta+1}\frac{x_{min}^{-\theta+1}}{-\theta+1}$

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