



The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search

<http://ageconsearch.umn.edu>

aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

No endorsement of AgEcon Search or its fundraising activities by the author(s) of the following work or their employer(s) is intended or implied.

A simple test for power-law behavior

Carlos M. Urzúa
Tecnológico de Monterrey
Mexico City, Mexico
curzua@tec.mx

Abstract. In this article, I propose a new test for power-law behavior. The statistical test, `pwlaw`, is locally optimal if the possible alternative distributions are contained in the Pareto type (IV) family. After deriving the test, I examine four classical datasets: the frequency of unique words in an English text (*Moby Dick*); the human populations of U.S. cities; the frequency of U.S. family names; and the peak gamma-ray intensity of solar flares. I show that in the first case there is no indication of any power-law behavior and that in the second and fourth cases there is evidence in that regard.

Keywords: `st0610`, `pwlaw`, power-law distribution, Pareto law, Zipf’s law, Pareto (IV) distribution

1 Introduction

A continuous random variable is said to have a power-law (PWL) behavior if it has a probability density of the form

$$g(x) = \frac{\alpha}{\mu} \left(\frac{x}{\mu} \right)^{-(\alpha+1)} \quad x > \mu \quad (1)$$

where $\alpha > 0$ is the shape parameter that determines the tail of the distribution and $\mu > 0$ is the location parameter where the “heavy tail” starts.¹ The law given above is due to Pareto (1897), but in the particular case of $\alpha = 1$, it is occasionally named after Zipf (1949).

The surveys by Brock (1999), Mitzenmacher (2004), and Newman (2005) make clear that the assumption of PWL behavior has been routinely made over the years by many authors in a number of disciplines. Also, in most of that literature, the use of PWL distributions is typically defended by simply pointing out the approximate linear nature of a doubly logarithmic plot or by reporting a simple regression. However, as noted by, for example, Clauset, Shalizi, and Newman (2009) and Urzúa (2011), those arguments are flawed.

Fortunately, there are also in the literature better procedures to test for PWL behavior. The most common, which can be traced back at least to Quandt (1964), is to estimate, for a given dataset, not only the Paretian law but also some competing

1. As is commonly assumed when studying this subject, the discrete counterpart to (1) is not considered here. Nevertheless, Newman (2005) and Clauset, Shalizi, and Newman (2009) present arguments against that common practice.

alternative distributions. Once those estimations are made, the resulting fits are ranked according to some statistical criteria. For instance, in their rather influential article on the empirics of PWL behavior, Clauset, Shalizi, and Newman (2009) suggest as alternatives to that distribution four others: the lognormal, the exponential, the stretched exponential, and the PWL with cutoff.

Other testing procedures have been proposed as well. In particular, using a methodology that is quite common in the econometrics literature, Urzúa (2000) presents a test for Zipf's law that is derived after considering a more general distribution that nests the density given in (1). Afterward, a locally optimal test for Zipf's law is derived by means of Rao's (1948) score test, also known in econometrics as the Lagrange multiplier test.²

The more general density posited in that article is of the form

$$f(x) = \frac{\alpha}{\sigma} \left(1 + \frac{x - \mu}{\sigma} \right)^{-(\alpha+1)} \quad x > \mu \quad (2)$$

where, aside from the two parameters introduced earlier, $\sigma > 0$ is the scale parameter. This distribution is known in the literature by several names, but here it will be simply called Pareto (II), following the terminology of Arnold (see the reference in Arnold [2015]); likewise, Pareto's basic law, given in (1), will be called from now on Pareto (I).

Using the Pareto (II) distribution, Urzúa (2000) derives a test for Zipf's law by testing the statistical hypothesis $H_0: \sigma = \mu, \alpha = 1$. Rao's score test is used because it requires estimation only under the null hypothesis. Note that if the null is reduced to $H_0: \sigma = \mu$, then, as Goerlich (2013) shows, one can even produce a test for the Pareto (I) distribution. Nevertheless, that test would not be appropriate in the case of testing for PWL behavior, because, as I will show in the next section, the densities (1) and (2.1) have a similar heavy-tail behavior.

2 A test for PWL distributions

Following again the terminology in Arnold (2015), this article assumes that the alternatives to the Pareto (I) distribution belong more generally to the Pareto (IV) family, whose density is given by

$$f(x) = \frac{\alpha}{\gamma\sigma} \left\{ 1 + \left(\frac{x - \mu}{\sigma} \right)^{1/\gamma} \right\}^{-(\alpha+1)} \left(\frac{x - \mu}{\sigma} \right)^{1/\gamma-1} \quad x > \mu \quad (3)$$

2. That other name, which originated from the article by Silvey (1959), did not seem to bother Rao: "[M]y score test for composite hypotheses has become entrenched in the econometrics literature under a fancier name, the Lagrange multiplier test" (Bera and Bilias 2001, 10).

where $\gamma > 0$ is the inequality parameter. Note that (2.1) is obtained from (2.1) when $\sigma = \mu$ and $\gamma = 1$ and that the Pareto (IV) distribution function is given by

$$F(x) = 1 - \left\{ 1 + \left(\frac{x - \mu}{\sigma} \right)^{1/\gamma} \right\}^{-\alpha} \quad x > \mu$$

For purposes of this article, the parameter μ , which is present in the three densities given above, is assumed to be fixed by statistical design. This assumption is quite reasonable from a practical point of view, aside from the fact that it is theoretically required.³ Thus, after defining $w = x - \mu$, we can rewrite the Pareto (IV) density in a more compact form as

$$f(w) = \frac{\alpha}{\gamma\sigma} \left\{ 1 + \left(\frac{w}{\sigma} \right)^{1/\gamma} \right\}^{-(\alpha+1)} \left(\frac{w}{\sigma} \right)^{1/\gamma-1} \quad w > 0 \quad (4)$$

Note that the tail behavior of a Pareto (IV) distribution is determined not only by α , as is the case for the Pareto (I) and (II) distributions, but also by the new parameter γ . This is so because, using L'Hôpital's rule and (2),

$$\begin{aligned} \lim_{w \rightarrow \infty} \frac{1 - F(w)}{w^{-\alpha/\gamma}} &= \lim_{w \rightarrow \infty} \frac{\gamma f(w)}{\alpha w^{-\alpha/\gamma-1}} \\ &= \lim_{w \rightarrow \infty} \left\{ 1 + \left(\frac{w}{\sigma} \right)^{1/\gamma} \right\}^{-(\alpha+1)} \left\{ \left(\frac{\sigma}{w} \right)^{1/\gamma} \right\}^{-(\alpha+1)} \sigma^{\alpha/\gamma} \\ &= \sigma^{\alpha/\gamma} \end{aligned} \quad (5)$$

Thus, the Pareto (IV) family can be approximated, on the right tail, by $w^{-\alpha/\gamma}$. This implies that, for a given PWL parameter α , the tail of the distribution can become as short as desired by reducing the value of γ from one to some small but positive value. Likewise, for a given α , the tail can be made even heavier by setting the value of γ greater than one.

Let $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3) = (\sigma, \gamma, \alpha)$ be the vector of parameters, and let $\{w_i\}_{i=1}^n$ be a random sample. Then, the corresponding log-likelihood function is

$$\ell(\boldsymbol{\theta}) = n \log \left(\frac{\alpha}{\gamma\sigma} \right) - (\alpha + 1) \sum_{i=1}^n \log \left\{ 1 + \left(\frac{w_i}{\sigma} \right)^{1/\gamma} \right\} + (1/\gamma - 1) \sum_{i=1}^n \log \left(\frac{w_i}{\sigma} \right)$$

3. For the purpose of maximum likelihood estimation, that assumption is needed to make the Pareto (IV) distribution a regular family (see, for example, Rao [1973, 364]).

For later use, it is necessary to find its gradient and the expected value of its Hessian. The elements of the gradient are given by

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \sigma} = -\frac{n}{\gamma \sigma} + \frac{\alpha + 1}{\gamma \sigma} \sum_{i=1}^n \frac{(w_i/\sigma)^{1/\gamma}}{1 + (w_i/\sigma)^{1/\gamma}} \quad (6)$$

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \gamma} = -\frac{n}{\gamma} - \frac{1}{\gamma^2} \sum_{i=1}^n \log(w_i/\sigma) + \frac{\alpha + 1}{\gamma^2} \sum_{i=1}^n \frac{(w_i/\sigma)^{1/\gamma} \log(w_i/\sigma)}{1 + (w_i/\sigma)^{1/\gamma}} \quad (7)$$

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \alpha} = \frac{n}{\alpha} - \sum_{i=1}^n \log \left\{ 1 + (w_i/\sigma)^{1/\gamma} \right\} \quad (8)$$

The second-order derivatives are also easy to find, but they are omitted here for lack of space. A less obvious task is to calculate their expected values. Fortunately, they have already been derived in the meticulous article by Brazauskas (2002). Thus, the elements of the information matrix

$$I_{ij}(\boldsymbol{\theta}) = -E_{\boldsymbol{\theta}} \left\{ \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \right\} \quad i, j = 1, 2, 3$$

are also known.

The stage is now set to devise a test for the PWL distribution. This is accomplished here by employing Rao's score test in the case of the null hypothesis $H_0: \sigma = \mu, \gamma = 1$. As a first step to that end, we find the restricted maximum-likelihood estimate for the nuisance parameter α by equating (8) to zero and replacing w_i with $x_i - \mu$:

$$\hat{\alpha} = \left\{ \frac{1}{n} \sum_{i=1}^n \log(x_i/\mu) \right\}^{-1} \quad (9)$$

Consequently, the restricted score vector is given by $\hat{\mathbf{d}} = (d_1, d_2, 0)'$, where the first two components can be found using (8) in (4)–(5) and imposing $\sigma = \mu$ and $\gamma = 1$. After some simple algebra, they are found to be

$$d_1 = \frac{n\hat{\alpha}}{\mu} - (\hat{\alpha} + 1) \sum_{i=1}^n \frac{1}{x_i} \quad (10)$$

$$d_2 = -n + \hat{\alpha} \sum_{i=1}^n \log \left(\frac{x_i - \mu}{\mu} \right) - (\hat{\alpha} + 1) \sum_{i=1}^n \frac{\mu}{x_i} \log \left(\frac{x_i - \mu}{\mu} \right) \quad (11)$$

Before presenting the corresponding information matrix, I need to introduce some notation. Let $\Gamma(a) = \int_0^\infty t^{a-1} e^{-t} dt$, $\psi(a) = \Gamma'(a)/\Gamma(a)$, and $\psi'(a) = d\psi(a)/da$ be, respectively, the gamma, digamma, and trigamma functions. Also, to have cleaner expressions in what follows, I define $p(a) \equiv \psi(a) - \psi(1) - 1$ and $q(a) \equiv \psi'(a) + \psi'(1)$. Given that notation, and after imposing $\sigma = \mu$ and $\gamma = 1$ in the information matrix for the Pareto (IV) distribution in Brazauskas (2002, 165), we obtain the following restricted information matrix:

$$\hat{\mathbf{I}} = n \begin{pmatrix} \frac{\hat{\alpha}}{\mu^2(\hat{\alpha} + 2)} & -\frac{\hat{\alpha}p(\hat{\alpha}) + 1}{\mu(\hat{\alpha} + 2)} & -\frac{1}{\mu(\hat{\alpha} + 1)} \\ -\frac{\hat{\alpha}p(\hat{\alpha}) + 1}{\mu(\hat{\alpha} + 2)} & \frac{\hat{\alpha}\{p(\hat{\alpha})^2 + q(\hat{\alpha})\} + 2(p(\hat{\alpha}) + 1)}{\hat{\alpha} + 2} & \frac{p(\hat{\alpha})}{\hat{\alpha} + 1} \\ -\frac{1}{\mu(\hat{\alpha} + 1)} & \frac{p(\hat{\alpha})}{\hat{\alpha} + 1} & \frac{1}{\hat{\alpha}^2} \end{pmatrix} \quad (12)$$

Finally, keeping in mind that $\hat{\mathbf{d}} = (d_1, d_2, 0)'$ is defined by (10) and (9) and that $\hat{\mathbf{I}}$ is given in (10), we define the score test for PWL distributions as

$$\text{PWL} = \hat{\mathbf{d}}' \hat{\mathbf{I}}^{-1} \hat{\mathbf{d}} \stackrel{a}{\sim} \chi_2^2$$

Under the null, PWL is asymptotically distributed as a chi-squared with two degrees of freedom. The statistic can be readily computed using the `pwlaw` command, with the following syntax:

```
pwlaw varname [ in ] [ , mu( # ) ]
```

Note that the program automatically drops any $x_i \leq \mu$.

How well does the test statistic behave in the case of small samples? Table 1 presents evidence that the asymptotic critical values can be safely used when $1 \leq \alpha \leq 3$, the typical range for the shape parameter, and $n \geq 100$, as is the case for most applications. Otherwise, the significance points in table 1 can be used as approximations, or for sharper results, the `pwlaw` command can be used to perform a specific Monte Carlo simulation.

Table 1. Significance points for PWL

α	Level	$n =$	20	25	30	40	60	100	∞
1	5%		5.64	5.70	5.77	5.81	5.85	5.93	5.99
	10%		4.19	4.28	4.35	4.40	4.46	4.53	4.61
1.5	5%		5.51	5.60	5.71	5.74	5.84	5.88	5.99
	10%		4.06	4.18	4.25	4.32	4.41	4.48	4.61
2	5%		5.44	5.54	5.63	5.70	5.80	5.86	5.99
	10%		3.96	4.08	4.17	4.26	4.36	4.45	4.61
2.5	5%		5.39	5.49	5.60	5.68	5.79	5.83	5.99
	10%		3.89	4.01	4.10	4.21	4.32	4.41	4.61
3	5%		5.35	5.45	5.55	5.65	5.78	5.85	5.99
	10%		3.83	3.95	4.05	4.16	4.28	4.39	4.61

NOTE: Simulations using 100,000 replications.

3 Applications

We now exemplify the use of the PWL test in the case of four interesting datasets studied by Newman (2005) and Clauset, Shalizi, and Newman (2009).⁴ They are listed in table 2, together with seven values associated with each of them. Regarding the columns, note that although x_{\max} is clearly the maximum value in each sample, \hat{x}_{\min} is not the minimum value but rather an estimated lower bound on PWL behavior.⁵ Given the value of that bound, the next smaller number in the sample corresponds to μ (to enforce $x > \mu$), while n_{tail} is the effective sample size.

Table 2. Tests of PWL behavior

	n	x_{\max}	\hat{x}_{\min}	μ	n_{tail}	PWL	p -value
Count of word use	18,855	14,086	7	6	2,958	189	0.00
Population of cities	19,447	8,008,654	52,457	52,360	580	0.16	0.92
Frequency of surnames	2,753	2,502,020	111,919	109,432	239	3.32	0.19
Solar flare intensity	12,773	231,300	323	322	1,711	0.87	0.65

The first dataset in table 2 refers to the number of times that unique words occur in a classical piece of English literature, the novel *Moby Dick* by Herman Melville. Our statistic can be computed with the command `pwlav x, mu(6)`. As shown in the table, the value of the statistic thus obtained is quite large in that particular case, and hence the associated p -value is quite small. Thus, according to the PWL test, the null hypothesis of PWL behavior is strongly rejected. This result is at odds with many

4. Those datasets were directly downloaded from the site maintained by Clauset (2019).

5. The procedure followed by those authors to estimate the lower bounds is described at length in Clauset, Shalizi, and Newman (2009, sec. 3.3).

previous studies on word frequencies, ranging from Zipf (1949) to Clauset, Shalizi, and Newman (2009).

The second case corresponds to the human populations of U.S. cities as recorded by the U.S. Census Bureau in 2000. As shown in the table, and according to the proposed test, there is convincing evidence of PWL behavior. This result is in line with most of the literature on the size of cities. Note also that this finding does not contradict the claim in Urzúa (2000) that U.S. cities do not follow Zipf's law, because the latter is just an extreme case of a PWL distribution (when $\alpha = 1$).

The third sample is made of the frequencies of U.S. family names in the 1990 U.S. Census. As reported in the table, there is moderate evidence of PWL behavior, because the null hypothesis would not be rejected at the typical significance levels. A very similar finding is reported by Clauset, Shalizi, and Newman (2009).

Finally, the fourth sample listed in table 2 is made of the peak gamma-ray intensity of solar flares, measured from orbit between 1980 and 1989 (see Newman [2005] for more details). Given the p -value reported in the table, there is good evidence of PWL behavior in this case. Note that a theoretical model that renders precisely that behavior is due to Lu and Hamilton (1991).

4 Concluding remark

The Pareto (IV) distribution used in this article belongs to the Feller–Pareto family (see Arnold [2015]), whose general density is of the form

$$f(x) = \frac{\Gamma(\gamma_1 + \gamma_2)}{\gamma\sigma\Gamma(\gamma_1)\Gamma(\gamma_2)} \left\{ 1 + \left(\frac{x - \mu}{\sigma} \right)^{1/\gamma} \right\}^{-(\gamma_1 + \gamma_2)} \left(\frac{x - \mu}{\sigma} \right)^{(\gamma_2/\gamma) - 1} \quad x > \mu$$

where $\gamma_2 > 0$ and, to use the notation given earlier, $\gamma_1 = \alpha$. Thus, following the same procedure given before, we can derive a new, more general test statistic. But because of its algebraic intricacy, it is not presented here. In any case, the gain from that generalization is not totally apparent, because, as can be checked in the same way as in (3), the Feller–Pareto and Pareto (IV) distributions have similar tail behavior.

5 Acknowledgments

I appreciate the very helpful comments made by a referee on an earlier version of the article. I also thank Erick Rosas for his research assistance.

6 Programs and supplemental materials

To install a snapshot of the corresponding software files as they existed at the time of publication of this article, type

```
. net sj 20-3
. net install st0610      (to install program files, if available)
. net get st0610          (to install ancillary files, if available)
```

7 References

- Arnold, B. C. 2015. *Pareto Distributions*. 2nd ed. Boca Raton, FL: CRC Press.
- Bera, A. K., and Y. Biliias. 2001. Rao's score, Neyman's $C(\alpha)$ and Silvey's LM tests: An essay on historical developments and some new results. *Journal of Statistical Planning and Inference* 97: 9–44. [https://doi.org/10.1016/S0378-3758\(00\)00343-8](https://doi.org/10.1016/S0378-3758(00)00343-8).
- Brazauskas, V. 2002. Fisher information matrix for the Feller–Pareto distribution. *Statistics & Probability Letters* 59: 159–167. [https://doi.org/10.1016/S0167-7152\(02\)00143-8](https://doi.org/10.1016/S0167-7152(02)00143-8).
- Brock, W. A. 1999. Scaling in economics: A reader's guide. *Industrial and Corporate Change* 8: 409–446. <https://doi.org/10.1093/icc/8.3.409>.
- Clauset, A. 2019. Power-law distributions. <http://tuvalu.santafe.edu/~aaronc/datacode.htm>.
- Clauset, A., C. R. Shalizi, and M. E. J. Newman. 2009. Power-law distributions in empirical data. *SIAM Review* 51: 661–703. <https://doi.org/10.1137/070710111>.
- Goerlich, F. J. 2013. A simple and efficient test for the Pareto law. *Empirical Economics* 45: 1367–1381. <https://doi.org/10.1007/s00181-012-0654-5>.
- Lu, E. T., and R. J. Hamilton. 1991. Avalanches and the distribution of solar flares. *Astrophysical Journal* 380: L89–L92. <http://doi.org/10.1086/186180>.
- Mitzenmacher, M. 2004. A brief history of generative models for power law and log-normal distributions. *Internet Mathematics* 1: 226–251. <https://doi.org/10.1080/15427951.2004.10129088>.
- Newman, M. E. J. 2005. Power laws, Pareto distributions and Zipf's law. *Contemporary Physics* 46: 323–351. <https://doi.org/10.1080/00107510500052444>.
- Pareto, V. 1897. *Cours d'Économie Politique*. Vol. 2. Lausanne: F. Rouge.
- Quandt, R. E. 1964. Statistical discrimination among alternative hypotheses and some economic regularities. *Journal of Regional Science* 5(2): 1–23. <https://doi.org/10.1111/j.1467-9787.1964.tb01462.x>.

- Rao, C. R. 1948. Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *Mathematical Proceedings of the Cambridge Philosophical Society* 44: 50–57. <https://doi.org/10.1017/S0305004100023987>.
- . 1973. *Linear Statistical Inference and Its Applications*. 2nd ed. New York: Wiley.
- Silvey, S. D. 1959. The Lagrangian multiplier test. *Annals of Mathematical Statistics* 30: 389–407. <https://doi.org/10.1214/aoms/1177706259>.
- Urzúa, C. M. 2000. A simple and efficient test for Zipf's law. *Economics Letters* 66: 257–260. [https://doi.org/10.1016/S0165-1765\(99\)00215-3](https://doi.org/10.1016/S0165-1765(99)00215-3).
- . 2011. Testing for Zipf's law: A common pitfall. *Economics Letters* 112: 254–255. <https://doi.org/10.1016/j.econlet.2011.05.049>.
- Zipf, G. K. 1949. *Human Behavior and the Principle of Least Effort*. Cambridge, MA: Addison–Wesley.

About the author

Carlos M. Urzúa is professor of economics at the Tecnológico de Monterrey. He is also a member of the Mexican Academy of Sciences.