Power Laws and Extreme Values

Paul Schuette
Meredith College

Schuette@meredith.edu
What are the Zipf-Pareto Distributions?

**Pareto Distribution.**

Cumulative distribution function. For $\alpha > 0$, 

$$
F(x) = \begin{cases} 
1 - \frac{1}{x^\alpha} & \text{if } x \geq 1 \\
0 & \text{if } x < 1 
\end{cases}
$$

(1)

Density function.

$$
f(x) = \begin{cases} 
\frac{\alpha}{x^{\alpha+1}} & \text{if } x \geq 1 \\
0 & \text{if } x < 1 
\end{cases}
$$

(2)

**Zipf Distribution.**

For $a > 0$, 

$$P(X = x) = \frac{1}{\zeta(\alpha + 1)x^{-(\alpha+1)}} \quad x = 1, 2, 3, \ldots$$

Pareto distributions and Zipf distributions are both known as “Power Laws.”
Origins and Development.

Vilfredo Pareto. 1848-1923.

• trained as an engineer, but worked as an economist.

• influential in mathematical economics.

“La Courbe des revenus” (1896), also in

_Cours d’economie politique_ (1897).

Let \( N \) denote the number of taxpayers whose income exceeds \( x \). Then

\[
\log N = \log A - \alpha \log x,
\]

where \( \alpha \) and \( A \) are constants.
Pareto recognized that chance (le hasard) played a role, but after examining data for England, Ireland, Germany, Italian cities, and Peru concluded that:

“These results are very remarkable. It is absolutely impossible to recognize that these effects are due only to randomness.”

(*Cours d’economie politique* §960, presenter’s translation)
Other Developments.

Willis (1912) genera species relationship for Ceylon (Sri Lanka).

Willis (1922) area-species relationship.

Auerbach (1923) proposed a population-rank relationship for cities. $P \cdot R = C$.

Estoup (1916) interest in stenography leads to investigation of word frequency relationships.

Lotka (1924) refines Auerbach’s conjecture, $P \cdot R^\alpha = C$.

Lotka (1926) scholarly productivity (publication) follows a power law.

Bradford (1934) journal productivity follows a power law.
George Kingsley Zipf (1902-1950)

- trained in Philology (linguistics)

- degrees from Harvard, bachelors 1924, Ph.D 1930, also studied in Bonn and Berlin

- Academic career: Instructor, Assistant Professor, and University Lecturer of German at Harvard.

- Numerous papers, three major works:


  2. *National Unity and Disunity* (1941) city sizes and movement.

- Zipf attempted to explain ubiquity of power laws due to “Principle of Least Effort” and to “Forces of Unification and Diversification.”

- Zipf saw “Principle of Least Effort” as a unifying theme for the social sciences.

- Zipf thanked J. L. Walsh and M.H. Stone of the Harvard Mathematics department for their advice, and also thanked Walsh for reading the book in its entirety.

- J. L. Walsh (AMS president 1949,1950) wrote a favorable review of *Human Behavior and the Principle of Least Effort* for the Scientific American (July 1949) in which he compared Zipf to Kepler.
“At a critical point in my life, I read a wise review of Human Behavior by the mathematician J. L. Walsh. By only mentioning what was good, this review influenced greatly my early scientific work, and its indirect influence continues. Therefore, I owe a great deal to Zipf through Walsh.” Benoit B. Mandelbrot, The Fractal Geometry of Nature. (1983, page 404)

“The body of evidence is impressive, but this work of Zipf's, which I admire even where I disagree, seems to illustrate the dangers of attaching too much importance to analogies with the physical sciences.” M. G. Kendall, Presidential address, Royal Statistical Society, 1960.
Attempts to Explain Zipf’s Law

(1924) Yule obtained

\[ f(n) \approx \frac{\Gamma(1 + \rho^{-1})}{\rho} n^{-1(1+\rho^{-1})} \]

for a mutation model of the data of Willis.


1953 Mandelbrot “An informational theory of the statistical structure of language” Communication Theory

1953 Champernowne, Markov chain model for the distribution of incomes.

- used a stochastic process model with proportional growth assumption to obtain distributions

- proportional to a Beta function.

- Due to resemblance with Yule, Simon called these “Yule” distributions.

- In his autobiography, Simon attributed his interest to Lotka’s work, and found Yule’s and Champernowne’s work during a literature search after obtaining his results.
The Mandelbrot-Simon Debate. appeared in *Information and Control*


1960 Simon, “Some Further Notes on a Class of Skew Distribution Functions” “. . . there is a whole host of stochastic processes that yield equilibrium distributions quite similar to the observed ones, and that, therefore, we should be wary about concluding much more than that some law of large numbers is at work.”

1961 Mandelbrot, “Final Note on a Class of Skew Distribution Functions: Analysis and Critique of a Model Due to H. A. Simon” “The failure of Simon’s and other attempts to derive the Pareto-Yule-Zipf distributions clearly shows that ‘much more than some law of large numbers is at work’ here.” “. . . this family of distributions has challenged many statisticians and is likely to continue to do so.”
1961 Simon, “Reply to ‘Final Note’ by Benoît Mandelbrot”

1961 Mandelbrot, “Post Scriptum to ‘Final Note’” “My criticism has not changed since I first had the privilege of commenting upon a draft of Simon (1955)”

1961 Simon, “Reply to Dr. Mandelbrot’s Post Scriptum” “Dr. Mandelbrot has proposed a new set of objects to my 1955 models of the Yule distribution. Like his earlier objections, these are invalid.”
Bruce Hill’s work


(1975a) Hill and Woodroofe, “Stronger Forms of Zipf’s Law.” JASA


With the exception of 1975b, an occupancy problem based on Bose-Einstein statistics was considered, and the number of cells increased.
Approach after 1970 used order statistics: If 

\[ X_1, X_2, \ldots X_n \] 

are RVs, denote the order statistics by

\[ X_{(1)} \geq X_{(2)} \geq \cdots X_{(n)} \]

In (1975b) Hill, considered iid random variables, whose distribution function is eventually Pareto, introduced the Hill estimator

\[
\hat{\alpha} = \left( \frac{1}{k} \sum_{j=1}^{k} \ln Z_j - \ln Z_{k+1} \right)^{-1}
\]

Over 300 hundred papers written on the Hill estimator.


\[ G_\gamma(x) = \exp \left\{ -(1 + \gamma x)^{-1/\gamma} \right\}, \quad \gamma \in \mathbb{R}, \ 1 + \gamma x > 0 \]

If \( \gamma > 0 \), then \( \gamma = \frac{1}{\hat{\alpha}} \).

Pickands estimator

\[
\hat{\gamma}_{k,n} = \frac{1}{\ln 2} \ln \left( \frac{X_{(k)} - X_{(2k)}}{X_{(2k)} - X_{(4k)}} \right)
\]
(1982) Mason “Laws of Large Numbers for Sums of Extreme Values” *Annals of Prob.* proved and characterized LLNs for $\hat{\alpha}_k(n), n$ with random variables in the domain of attraction of a Frechet distribution.
Power Laws Observed

- Earthquakes
- Floods
- Hurricanes
- Stock returns
- Firm Sizes
- Linguistics
- Demography
- Insurance
- Biodiversity
• Management/Quality Control (80-20 principle)

• Book/Journal usage in libraries (Bradford’s Law)

• Web usage

• Distribution of “junk DNA”

• Blog Popularity

Some hold that “power law” behavior indicates a phenomenon of human origin!
Tail Fit and the Zipf-Pareto Law

Joint work with M.C. Spruill.

A limit theorem with bounds on the rate of convergence is proven. The joint distribution of a fixed number of relative decrements of the top order statistics from a random sample converges to the limit as the sample size increases if and only if the underlying distribution is in essence a Pareto. In conjunction with a chi-square test of fit, it provides an asymptotically distribution-free test of fit to the family of distributions with regularly varying tails at infinity. When the limit distribution holds, rank-size plots obey Zipf’s law. The test can be used to detect departures from this Zipf-Pareto law.
Let $X_1, \ldots, X_n$ be iid RVs with distribution $F$.

Denote the order statistics $X_{1:n} < \cdots < X_{n:n}$.

Regularly varying functions:

$$\mathcal{R}_\alpha = \{ U : \lim_{t \to \infty} \frac{U(tx)}{U(t)} = x^\alpha \}$$

$F$ absolutely continuous and $F' = f \in \mathcal{R}_{-b}$, $b \geq 1 \Rightarrow \bar{F}(x) = 1 - F(x) \in \mathcal{R}_{-(b-1)}$ and the Von-Mises condition holds:

$$\lim_{x \to \infty} \frac{xf(x)}{1 - F(x)} = b - 1 \quad (3)$$

If $\bar{F} \in \mathcal{R}_{-(b-1)}$ is absolutely continuous and satisfies (3) then $f \in \mathcal{R}_{-b}$. 
Karamata’s Theorem. If $f$ is regularly varying \( \exists \eta : \mathbb{R}^+ \to \mathbb{R} \), constant \( c > 0 \), \( z_0 < \infty \) such that for \( x > z_0 \)

\[
\bar{F}(x) = x^{-(b-1)}L(x)
= cx^{-(b-1)} \exp\left\{ \int_{z_0}^{x} \frac{\eta(s)}{s} ds \right\},
\]

(4)

where \( \lim_{s \to \infty} \eta(s) = 0 \).

For \( x > 0 \) let

\[
\gamma(x) = \sup_{t \geq x} |\eta(t)|.
\]

\( \gamma \) is non-increasing, \( \lim_{x \to \infty} \gamma(x) = 0 \), and \( \eta \) continuous \( \implies \gamma \) is continuous.
If $\bar{F} = 1 - F$ is regularly varying, $k \geq 2$ fixed, as $n \to \infty$ the distribution of the $k-1$ relative decrements

$$V_{n,j} = (X_{n-j+2:n} - X_{n-j+1:n})/X_{n-j+2:n}, \quad (5)$$

$j = 2, \ldots, k$, approaches that of $k-1$ independent random variables $V_j$, where the density of $V_j$ is

$$p_j(v) = (j-1)(b-1)(1-v)^{(j-1)(b-1)-1}I_{(0,1)}(v),$$

and $1 - b$ is the index of variation of $\bar{F}$. (see Leadbetter, Lindgren and Rootzen)

Let $h_n(v)$, be the density of $(V_{n,2}, \ldots, V_{n,k})$. We'll show convergence with rates to the asymptotic density, $h(v) = \prod_{j=2}^{k} p_j(v_j)$. 
Conditions:

(VM) For the density \( f \), \( |x|f(x) \) is bounded, \( f \in \mathcal{R}_b \) for some \( b > 1 \) so (3) holds, and \( \eta \) is continuous.

The following conditions will be invoked only to bound the rate.

(K) The function \( \eta \) is positive for all \( s > 0 \) and for any fixed \( \tau \in (0, 1) \) there are constants \( 1 \leq D_\tau \) and \( 0 < V_\tau \leq 1 \) such that for all \( x \) sufficiently large

\[
D_\tau \cdot \eta(x) \geq \max\{\eta(s) : s \in [x, x/\tau]\}
\]

\[
V_\tau \cdot \eta(x) \leq \min\{\eta(s) : s \in [x, x/\tau]\}.
\]

Furthermore, \( \limsup_{\tau \to 0} D_\tau = D < \infty \). It can be proven that if \( \eta \) is continuous, regularly varying, and tends to zero for large \( x \) then the condition is satisfied.

(C) For \( a > 0 \),

\[
\lim_{n \to \infty} \frac{\gamma(F^{-1}(1 - a/n))}{\gamma^{1+\epsilon}(F^{-1}(1 - t_n))} = \begin{cases} +\infty & \text{if } \epsilon > 0, \\ 0 & \text{if } \epsilon < 0, \end{cases}
\]
**Lemma 1.** For $F$ satisfying (VM) and $T = \infty$, if $1 < \rho < e$ then for $n$ sufficiently large there is a solution $t_n \in (0, 1)$ such that

$$
\frac{1}{t_n \gamma(F^{-1}(1 - t_n))} = \rho^{nt_n}.
$$

(6) Furthermore, $t_n \to 0$ and $nt_n \to \infty$.

**Theorem.** For $F$ satisfying (VM) and any $k \geq 2$ and $v = (v_2, \ldots, v_k) \in (0, 1)^{k-1}$

(a) if $T \in (0, \infty)$ then with $\alpha_n = n^k F^n(T)$ one has (7) for $\epsilon > 0$;

(b) If $T = \infty$ then with $\alpha_n = \gamma(F^{-1}(1 - t_n))$ and $t_n$ given by Lemma 1, for $\epsilon > 0$ one has (7). If $\epsilon < 0$ and, in addition, conditions (K) and (C) obtain, then for some $v$ and $k$, (7) fails.

$$
\alpha_n^{\epsilon - 1} |h_n(v) - h(v)| \to 0.
$$

(7)
Expression for the joint density of the V’s

Joint density of the order statistics

\[ U_j = X_{n-j+1:n}, j = 1, \ldots, k, \]

is

\[ f_n(u_1, \ldots, u_k) = \frac{n!}{(n-k)!} F^{n-k}(u_k) f(u_k) f(u_{k-1}) \cdots f(u_1). \]

(8)

Making the change of variables to \( V_1 = U_1 \) with the remaining \( V'\)'s defined by (5), and denoting the joint density of the \( V'\)'s by \( g_n(v_1, \ldots, v_k) \), one has the joint density of the relative decrements

\[ h_n(v_2, \ldots, v_k) = \int g_n(v_1, \ldots, v_k) dv_1, \]

Momentarily fixing \( v = (v_2, \ldots, v_k) \), and making the change of variable from \( v_1 \) to \( w \) defined by

\[ w = 1 - F((1-v_k)(1-v_{k-1}) \cdots (1-v_2)v_1), \]

one has

\[ h_n(v) = (1-v_2)^{-2}(1-v_3)^{-3} \cdots (1-v_k)^{-k} \int_0^1 \tilde{K}_n(w) S_v(w) dw, \]

(9)
where

\[ S_v(w) = \frac{1}{w^{k-1}} f\left( \frac{F^{-1}(1 - w)}{1 - v_k} \right) \ldots \]

\[ f\left( \frac{F^{-1}(1 - w)}{\prod_{j=2}^{k}(1 - v_j)} \right)(F^{-1}(1 - w))^{k-1} \] (10)

and \( \tilde{K}_n(w) = \frac{n!}{(n-k)!}(1 - w)^{n-k}w^{k-1}I_{(0,1)}(w). \)
Idea of the proof

Let \( v = (v_2, \ldots, v_k) \in (0, 1)^{k-1} \) be fixed and for \( j = 2, \ldots, k, \)

\[
g_j(w) = (1 - v_j)^{-j} F^{-1}(1 - w) f \left( \frac{F^{-1}(1 - w)}{\prod_{i=j}^k (1 - v_i)} \right)
\]

and

\[
c_j = (b - 1)(1 - v_j)^{-j} \left( \prod_{i=j}^k (1 - v_i) \right)^b.
\]

Let \( W_n \) be the Beta random variable with density

\[
K_n(w) = \frac{\Gamma(n + 1)}{\Gamma(n - k + 1)\Gamma(k)} w^{k-1}(1-w)^{n-k} I_{(0,1)}(w).
\]

Then, the joint density \( h_n(v_2, \ldots, v_k) \) and its asserted limit \( h \) satisfy,

\[
h_n(v_2, \ldots, v_k) - h(v_2, \ldots, v_k) = (k - 1)! E\left[ \prod_{j=2}^k g_j(W_n) - \prod_{j=2}^k c_j \right]. \tag{11}
\]

Provide bounds and rates for the last term.
Example 1. Let $F(x) = 1 - \frac{1}{x \ln x}$. Show that $|h_n - h|$ goes to zero like $\frac{1}{\ln n}$.

Solution: Here $\eta(x) = \frac{1}{\ln x} = \gamma(x)$ so that if

$$\frac{1}{t_n \gamma(F^{-1}(1 - t_n))} = \rho^{nt_n} = \frac{\ln(F^{-1}(1 - t_n))}{t_n}$$

then, setting $w_n = F^{-1}(1 - t_n), t_n = \frac{1}{w_n \ln w_n}$. So

$$w_n \ln^2 w_n = \rho^{\frac{n}{w_n \ln w_n}}$$

or

$$\ln w_n + 2 \ln \ln w_n = \frac{n}{w_n \ln w_n} \ln \rho.$$ 

Therefore,

$$\frac{w_n(\ln w_n)^2}{n} \sim \text{const.}$$

Then $w_n \approx \frac{n}{(\ln n)^2}$ and

$$\gamma(F^{-1}(1 - t_n)) = \gamma(w_n) \sim \frac{1}{\ln(cn/(\ln n)^2)} \sim \frac{1}{\ln n}$$

as was to be proven. □
Example 2. Let $F(x) = 2\tan^{-1}(x)$, $x > 0$.

As in Example 1, $F \in \mathcal{R}_1$. Show that $|h_n - h|$ goes to zero like $\frac{(\ln n)^2}{n^2}$.

Solution:

By expanding at $\infty$ one finds

$$\gamma(n) \sim \frac{1}{n^2} \quad \text{and} \quad F^{-1}(1 - t) \sim \frac{1}{t}.$$ 

So

$$\frac{1}{t\gamma(F^{-1}(1 - t))} \sim \frac{1}{t^3}.$$ 

For $t_n - 3\ln t_n \sim nt_n \ln \rho$. Taking $t_n = \frac{\ln n}{n}$ and so

$$\gamma(F^{-1}(1 - t_n)) \sim \frac{(\ln n)^2}{n^2}. \qed$$
Lemma 2. If $X_1, \ldots, X_n$ are i.i.d. from an absolutely continuous distribution $F$ with probability density function $f$ and $X_{n-1:n}/X_{n:n}$ converges in law to the distribution whose density is $(b - 1)x^{b-2}I_{(0,1)}(x)$ then there are random variables $Y_n \xrightarrow{p} \infty$ such that for each $c > 0$

$$E[f(cY_n)/f(Y_n)] \to c^{-b}$$

as $n \to \infty$.

Lemma 3. Under the conditions of our Theorem

$$\lim_{n \to \infty} E[X_{n-j:n}/X_{n:n}] = C_{j}j^{-1/(b-1)}$$

and $j^{1-\epsilon}|C_j - e^\psi/(b-1)| \to 0$ as $j \to \infty$ for any $\epsilon > 0$, where $\psi = 0.57721 \ldots$ is Euler’s constant.
A test of fit

Assume $n$ is large and the underlying distribution is regularly varying at infinity.

Then

$$W_{j,n} = -(j - 1) \ln \left( \frac{X_{n-j+1:n}}{X_{n-j+2:n}} \right),$$

$j = 2, \ldots, k + 1$ will be approximately iid exponential with mean $\theta = 1/a$, where $a = b - 1$.

Test the fit of the distribution of these $W$’s to the exponential family using the Watson-Roy-Kambhampati chi-square test based on the statistic $T_{3k}$ of Moore and Spruill is used.

To execute this random cell chi-square test, the (raw data) maximum likelihood estimator $\hat{\theta}_k = 1/\hat{a}_k$, is employed, where

$$\hat{a}_k = \left[ \frac{1}{k} \sum_{i=0}^{k-1} \ln X_{n-i:n} - \ln X_{n-k:n} \right]^{-1}$$

is Hill’s estimator of the order of regular variation. The random cell chi-square test statistic is formed based on the $m$ random cell boundaries $b_j(\phi) = \phi c_j$, $j = 1, \ldots, m - 1$, where $c_j = - \ln (1 - j/m)$. 

28
Implementation may be found at Carl Spruill's site http://www.opti-stat.com/.

**Example 3.** The top 32 order statistics of \( n = 253 \) observations constituting log returns from 16 June 1999 to 15 June 2000 of the FTSE are plotted in Figure 1 and were analyzed. The statistic \( T = 16.01, \) p-value 0.001, entails a significant lack of fit to the model of i.i.d. regularly varying data.

![Daily Log Returns FTSE](image)

*Figure 1.*
Example 4. A plot of the 28 maximum daily log returns of stock values of BMW is found in Figure 2. A chi square of 4.36, p-value 0.23, is obtained so there is no significant evidence of a lack of fit to an underlying with regularly varying tails. This is in agreement with the results of Embrechts et. al. who analyzed earlier BMW returns using different methods.

Figure 2.
Example 5. The top 28 frequencies were obtained from the access log of requests over the span of 0.87 day, October 15, 2005 for web pages from the departmental web server in a major university. A plot is found in Figure 3. A chi square of 2.83, p-value 0.42, signifies no significant lack of fit to regularly varying tails.

Figure 3.
**Example 6.** A plot of the 28 maximum frequencies of words occurring in the 100,000 words from various sources in Dewey can be found in Figure 4. A chi square of 3.34, p-value 0.34, is obtained so, there is no significant evidence of lack of fit to the model of the word frequencies as iid’s from an asymptotically Pareto probability mass function.

![Figure 4. Dewey Most Frequent Words](image-url)
**Example 7.** A plot of the 28 maximum frequencies of words occurring in Ulysses (Joyce) is found in Figure 5. A chi square of 8.91, p-value 0.03, is obtained so, there is significant evidence of lack of fit to the model of the word frequencies as iid’s from an asymptotically Pareto probability mass function.
Many researchers used log-log plots to make judgments about fit of the model. In Figure 6 such a plot is found. The least squares line accounts for 98% of the variation.

Figure 6.