WIKIPEDIA

Pareto distribution

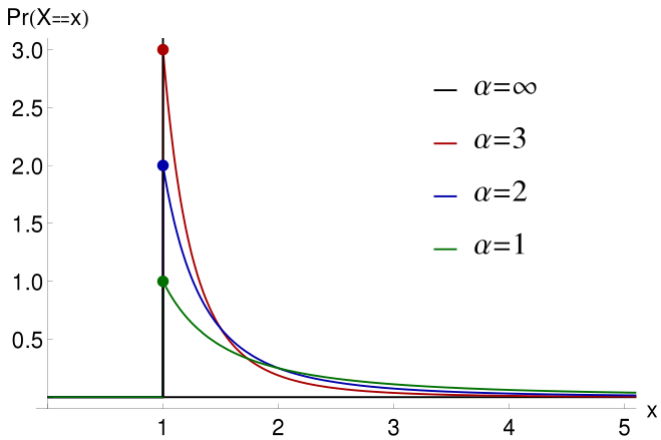
The **Pareto distribution**, named after the Italian civil engineer, economist, and sociologist Vilfredo Pareto,[1] (Italian: [paˈreːto] US: /pəˈreɪtoʊ/ *pə-RAY-toh*),[2] is a power-law probability distribution that is used in description of social, quality control, scientific, geophysical, actuarial, and many other types of observable phenomena. Originally applied to describing the distribution of wealth in a society, fitting the trend that a large portion of wealth is held by a small fraction of the population.[3][4] The Pareto principle or "80-20 rule" stating that 80% of outcomes are due to 20% of causes was named in honour of Pareto, but the concepts are distinct, and only Pareto distributions with shape value (α) of $\log_4 5 \approx 1.16$ precisely reflect it. Empirical observation has shown that this 80-20 distribution fits a wide range of cases, including natural phenomena[5] and human activities.[6]

Contents

- **Definitions**
  - Cumulative distribution function
  - Probability density function
- **Properties**
  - Moments and characteristic function
  - Conditional distributions
  - A characterization theorem
  - Geometric mean
  - Harmonic mean
  - Graphical representation
- **Related distributions**
  - Generalized Pareto distributions
    - Pareto types I–IV
    - Feller–Pareto distribution
  - Relation to the exponential distribution
  - Relation to the log-normal distribution
  - Relation to the generalized Pareto distribution
  - Bounded Pareto distribution
    - Generating bounded Pareto random variables
  - Symmetric Pareto distribution

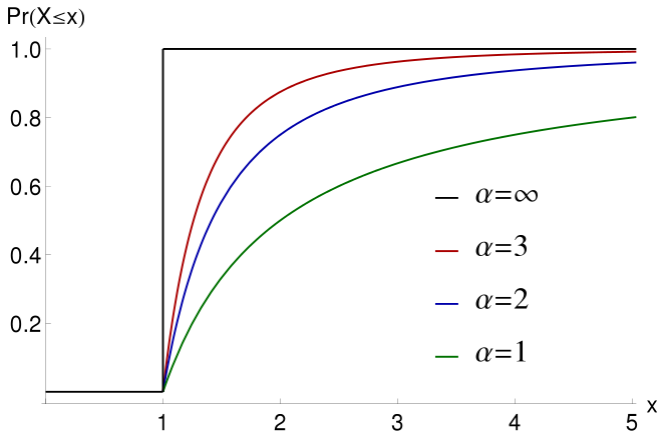
Pareto Type I

Probability density function



Pareto Type I probability density functions for various α with $x_\mathrm{m} = 1$. As $\alpha \to \infty$, the distribution approaches $\delta(x - x_\mathrm{m})$, where δ is the Dirac delta function.

Cumulative distribution function



Pareto Type I cumulative distribution functions for various α with $x_\mathrm{m} = 1$.

Parameters	$x_\mathrm{m} > 0$ scale (real) $\alpha > 0$ shape (real)
Support	$x \in [x_\mathrm{m}, \infty)$
PDF	$\frac{\alpha x_\mathrm{m}^\alpha}{x^{\alpha+1}}$
CDF	$1 - \left(\frac{x_\mathrm{m}}{x}\right)^\alpha$
Mean	$\begin{cases} \infty & \text{for } \alpha \le 1 \\ \frac{\alpha x_\mathrm{m}}{\alpha - 1} & \text{for } \alpha > 1 \end{cases}$
Median	$x_\mathrm{m} \sqrt[\alpha]{2}$
Mode	x_m

Multivariate Pareto distribution

Statistical inference
- Estimation of parameters
- Confidence intervals

Occurrence and applications
- General
- Relation to Zipf's law
- Relation to the "Pareto principle"
- Relation to Price's law
- Lorenz curve and Gini coefficient

Computational methods
- Random sample generation

See also

References

Notes

External links

Variance	$\begin{cases} \infty & \text{for } \alpha \le 2 \\ \frac{x_{\mathrm{m}}^2 \alpha}{(\alpha-1)^2(\alpha-2)} & \text{for } \alpha > 2 \end{cases}$
Skewness	$\frac{2(1+\alpha)}{\alpha-3}\sqrt{\frac{\alpha-2}{\alpha}}$ for $\alpha > 3$
Ex. kurtosis	$\frac{6(\alpha^3+\alpha^2-6\alpha-2)}{\alpha(\alpha-3)(\alpha-4)}$ for $\alpha > 4$
Entropy	$\log\left(\left(\frac{x_{\mathrm{m}}}{\alpha}\right) e^{1+\frac{1}{\alpha}}\right)$
MGF	does not exist
CF	$\alpha(-ix_{\mathrm{m}}t)^{\alpha}\Gamma(-\alpha,-ix_{\mathrm{m}}t)$
Fisher information	$\mathcal{I}(x_{\mathrm{m}},\alpha) = \begin{bmatrix} \frac{\alpha}{x_{\mathrm{m}}^2} & -\frac{1}{x_{\mathrm{m}}} \\ -\frac{1}{x_{\mathrm{m}}} & \frac{1}{\alpha^2} \end{bmatrix}$ Right: $\mathcal{I}(x_{\mathrm{m}},\alpha) = \begin{bmatrix} \frac{\alpha^2}{x_{\mathrm{m}}^2} & 0 \\ 0 & \frac{1}{\alpha^2} \end{bmatrix}$

Definitions

If X is a random variable with a Pareto (Type I) distribution,[7] then the probability that X is greater than some number x, i.e. the survival function (also called tail function), is given by

$$\overline{F}(x) = \Pr(X > x) = \begin{cases} \left(\frac{x_{\mathrm{m}}}{x}\right)^{\alpha} & x \ge x_{\mathrm{m}}, \\ 1 & x < x_{\mathrm{m}}, \end{cases}$$

where x_{m} is the (necessarily positive) minimum possible value of X, and α is a positive parameter. The Pareto Type I distribution is characterized by a scale parameter x_{m} and a shape parameter α, which is known as the *tail index*. When this distribution is used to model the distribution of wealth, then the parameter α is called the Pareto index.

Cumulative distribution function

From the definition, the cumulative distribution function of a Pareto random variable with parameters α and x_{m} is

$$F_X(x) = \begin{cases} 1 - \left(\frac{x_{\mathrm{m}}}{x}\right)^{\alpha} & x \ge x_{\mathrm{m}}, \\ 0 & x < x_{\mathrm{m}}. \end{cases}$$

Probability density function

It follows (by differentiation) that the probability density function is

$$f_X(x) = \begin{cases} \frac{\alpha x_{\mathrm{m}}^{\alpha}}{x^{\alpha+1}} & x \ge x_{\mathrm{m}}, \\ 0 & x < x_{\mathrm{m}}. \end{cases}$$

When plotted on linear axes, the distribution assumes the familiar J-shaped curve which approaches each of the orthogonal axes asymptotically. All segments of the curve are self-similar (subject to appropriate scaling factors). When plotted in a log-log plot, the distribution is represented by a straight line.

Properties

Moments and characteristic function

- The expected value of a random variable following a Pareto distribution is

$$\mathrm{E}(X) = \begin{cases} \infty & \alpha \leq 1, \\ \frac{\alpha x_\mathrm{m}}{\alpha - 1} & \alpha > 1. \end{cases}$$

- The variance of a random variable following a Pareto distribution is

$$\mathrm{Var}(X) = \begin{cases} \infty & \alpha \in (1, 2], \\ \left(\frac{x_\mathrm{m}}{\alpha - 1}\right)^2 \frac{\alpha}{\alpha - 2} & \alpha > 2. \end{cases}$$

  (If $\alpha \leq 1$, the variance does not exist.)

- The raw moments are

$$\mu'_n = \begin{cases} \infty & \alpha \leq n, \\ \frac{\alpha x_\mathrm{m}^n}{\alpha - n} & \alpha > n. \end{cases}$$

- The moment generating function is only defined for non-positive values $t \leq 0$ as

$$M\left(t; \alpha, x_\mathrm{m}\right) = \mathrm{E}\left[e^{tX}\right] = \alpha(-x_\mathrm{m}t)^\alpha \Gamma(-\alpha, -x_\mathrm{m}t)$$
$$M\left(0, \alpha, x_\mathrm{m}\right) = 1.$$

Thus, since the expectation does not converge on an open interval containing $t = 0$ we say that the moment generating function does not exist.

- The characteristic function is given by

$$\varphi(t; \alpha, x_\mathrm{m}) = \alpha(-ix_\mathrm{m}t)^\alpha \Gamma(-\alpha, -ix_\mathrm{m}t),$$

  where $\Gamma(a, x)$ is the incomplete gamma function.

The parameters may be solved for using the method of moments.[8]

Conditional distributions

The conditional probability distribution of a Pareto-distributed random variable, given the event that it is greater than or equal to a particular number x_1 exceeding x_m, is a Pareto distribution with the same Pareto index α but with minimum x_1 instead of x_m. This implies that the conditional expected value (if it is finite, i.e. $\alpha > 1$) is proportional to x_1. In case of random variables that describe the lifetime of an object, this means that life expectancy is proportional to age, and is called the Lindy effect or Lindy's Law.[9]

A characterization theorem

Suppose $X_1, X_2, X_3, \ldots$ are independent identically distributed random variables whose probability distribution is supported on the interval $[x_\mathrm{m}, \infty)$ for some $x_\mathrm{m} > 0$. Suppose that for all n, the two random variables $\min\{X_1, \ldots, X_n\}$ and $(X_1 + \cdots + X_n)/\min\{X_1, \ldots, X_n\}$ are independent. Then the common distribution is a Pareto distribution.

Geometric mean

The geometric mean (G) is[10]

$$G = x_\mathrm{m} \exp\left(\frac{1}{\alpha}\right).$$

Harmonic mean

The harmonic mean (H) is[10]

$$H = x_\mathrm{m} \left(1 + \frac{1}{\alpha}\right).$$

Graphical representation

The characteristic curved 'long tail' distribution when plotted on a linear scale, masks the underlying simplicity of the function when plotted on a log-log graph, which then takes the form of a straight line with negative gradient: It follows from the formula for the probability density function that for $x \geq x_\mathrm{m}$,

$$\log f_X(x) = \log\left(\alpha \frac{x_\mathrm{m}^\alpha}{x^{\alpha+1}}\right) = \log(\alpha x_\mathrm{m}^\alpha) - (\alpha + 1)\log x.$$

Since α is positive, the gradient $-(\alpha + 1)$ is negative.

Related distributions

Generalized Pareto distributions

There is a hierarchy [7][11] of Pareto distributions known as Pareto Type I, II, III, IV, and Feller–Pareto distributions.[7][11][12] Pareto Type IV contains Pareto Type I–III as special cases. The Feller–Pareto[11][13] distribution generalizes Pareto Type IV.

Pareto types I–IV

The Pareto distribution hierarchy is summarized in the next table comparing the survival functions (complementary CDF).

When $\mu = 0$, the Pareto distribution Type II is also known as the Lomax distribution.[14]

In this section, the symbol x_m, used before to indicate the minimum value of x, is replaced by σ.

Pareto distributions

	$\overline{F}(x) = 1 - F(x)$	Support	Parameters
Type I	$\left[\frac{x}{\sigma}\right]^{-\alpha}$	$x \geq \sigma$	$\sigma > 0, \alpha$
Type II	$\left[1 + \frac{x-\mu}{\sigma}\right]^{-\alpha}$	$x \geq \mu$	$\mu \in \mathbb{R}, \sigma > 0, \alpha$
Lomax	$\left[1 + \frac{x}{\sigma}\right]^{-\alpha}$	$x \geq 0$	$\sigma > 0, \alpha$
Type III	$\left[1 + \left(\frac{x-\mu}{\sigma}\right)^{1/\gamma}\right]^{-1}$	$x \geq \mu$	$\mu \in \mathbb{R}, \sigma, \gamma > 0$
Type IV	$\left[1 + \left(\frac{x-\mu}{\sigma}\right)^{1/\gamma}\right]^{-\alpha}$	$x \geq \mu$	$\mu \in \mathbb{R}, \sigma, \gamma > 0, \alpha$

The shape parameter α is the tail index, μ is location, σ is scale, γ is an inequality parameter. Some special cases of Pareto Type (IV) are

$$P(IV)(\sigma, \sigma, 1, \alpha) = P(I)(\sigma, \alpha),$$
$$P(IV)(\mu, \sigma, 1, \alpha) = P(II)(\mu, \sigma, \alpha),$$
$$P(IV)(\mu, \sigma, \gamma, 1) = P(III)(\mu, \sigma, \gamma).$$

The finiteness of the mean, and the existence and the finiteness of the variance depend on the tail index α (inequality index γ). In particular, fractional δ-moments are finite for some $\delta > 0$, as shown in the table below, where δ is not necessarily an integer.

Moments of Pareto I–IV distributions (case $\mu = 0$)

	$\mathrm{E}[X]$	Condition	$\mathrm{E}[X^\delta]$	Condition
Type I	$\frac{\sigma\alpha}{\alpha - 1}$	$\alpha > 1$	$\frac{\sigma^\delta \alpha}{\alpha - \delta}$	$\delta < \alpha$
Type II	$\frac{\sigma}{\alpha - 1} + \mu$	$\alpha > 1$	$\frac{\sigma^\delta \Gamma(\alpha - \delta)\Gamma(1 + \delta)}{\Gamma(\alpha)}$	$0 < \delta < \alpha$
Type III	$\sigma\Gamma(1 - \gamma)\Gamma(1 + \gamma)$	$-1 < \gamma < 1$	$\sigma^\delta \Gamma(1 - \gamma\delta)\Gamma(1 + \gamma\delta)$	$-\gamma^{-1} < \delta < \gamma^{-1}$
Type IV	$\frac{\sigma\Gamma(\alpha - \gamma)\Gamma(1 + \gamma)}{\Gamma(\alpha)}$	$-1 < \gamma < \alpha$	$\frac{\sigma^\delta \Gamma(\alpha - \gamma\delta)\Gamma(1 + \gamma\delta)}{\Gamma(\alpha)}$	$-\gamma^{-1} < \delta < \alpha/\gamma$

Feller–Pareto distribution

Feller[11][13] defines a Pareto variable by transformation $U = Y^{-1} - 1$ of a beta random variable Y, whose probability density function is

$$f(y) = \frac{y^{\gamma_1 - 1}(1 - y)^{\gamma_2 - 1}}{B(\gamma_1, \gamma_2)}, \qquad 0 < y < 1; \gamma_1, \gamma_2 > 0,$$

where $B(\ )$ is the beta function. If

$$W = \mu + \sigma(Y^{-1} - 1)^{\gamma}, \qquad \sigma > 0, \gamma > 0,$$

then W has a Feller–Pareto distribution FP(μ, σ, γ, γ_1, γ_2).[7]

If $U_1 \sim \Gamma(\delta_1, 1)$ and $U_2 \sim \Gamma(\delta_2, 1)$ are independent Gamma variables, another construction of a Feller–Pareto (FP) variable is[15]

$$W = \mu + \sigma\left(\frac{U_1}{U_2}\right)^{\gamma}$$

and we write $W \sim$ FP(μ, σ, γ, δ_1, δ_2). Special cases of the Feller–Pareto distribution are

$$FP(\sigma, \sigma, 1, 1, \alpha) = P(I)(\sigma, \alpha)$$
$$FP(\mu, \sigma, 1, 1, \alpha) = P(II)(\mu, \sigma, \alpha)$$
$$FP(\mu, \sigma, \gamma, 1, 1) = P(III)(\mu, \sigma, \gamma)$$
$$FP(\mu, \sigma, \gamma, 1, \alpha) = P(IV)(\mu, \sigma, \gamma, \alpha).$$

Relation to the exponential distribution

The Pareto distribution is related to the exponential distribution as follows. If X is Pareto-distributed with minimum x_m and index α, then

$$Y = \log\left(\frac{X}{x_\mathrm{m}}\right)$$

is exponentially distributed with rate parameter α. Equivalently, if Y is exponentially distributed with rate α, then

$$x_\mathrm{m} e^{Y}$$

is Pareto-distributed with minimum x_m and index α.

This can be shown using the standard change-of-variable techniques:

$$\begin{aligned}\Pr(Y < y) &= \Pr\left(\log\left(\frac{X}{x_\mathrm{m}}\right) < y\right) \\ &= \Pr(X < x_\mathrm{m} e^{y}) = 1 - \left(\frac{x_\mathrm{m}}{x_\mathrm{m} e^{y}}\right)^{\alpha} = 1 - e^{-\alpha y}.\end{aligned}$$

The last expression is the cumulative distribution function of an exponential distribution with rate α.

Pareto distribution can be constructed by hierarchical exponential distributions.[16] Let $\phi|a \sim \mathbf{Exp}(a)$ and $\eta|\phi \sim \mathbf{Exp}(\phi)$. Then we have $p(\eta|a) = \dfrac{a}{(a+\eta)^2}$ and, as a result, $a + \eta \sim \mathbf{Pareto}(a, 1)$.

More in general, if $\lambda \sim \mathbf{Gamma}(\alpha, \beta)$ (shape-rate parametrization) and $\eta|\lambda \sim \mathbf{Exp}(\lambda)$, then $\beta + \eta \sim \mathbf{Pareto}(\beta, \alpha)$.

Equivalently, if $Y \sim \mathbf{Gamma}(\alpha, 1)$ and $X \sim \mathbf{Exp}(1)$, then $x_{\mathrm{m}}\left(1+\frac{X}{Y}\right) \sim \mathbf{Pareto}(x_{\mathrm{m}}, \alpha)$.

Relation to the log-normal distribution

The Pareto distribution and log-normal distribution are alternative distributions for describing the same types of quantities. One of the connections between the two is that they are both the distributions of the exponential of random variables distributed according to other common distributions, respectively the exponential distribution and normal distribution. (See the previous section.)

Relation to the generalized Pareto distribution

The Pareto distribution is a special case of the generalized Pareto distribution, which is a family of distributions of similar form, but containing an extra parameter in such a way that the support of the distribution is either bounded below (at a variable point), or bounded both above and below (where both are variable), with the Lomax distribution as a special case. This family also contains both the unshifted and shifted exponential distributions.

The Pareto distribution with scale x_m and shape α is equivalent to the generalized Pareto distribution with location $\mu = x_m$, scale $\sigma = x_m/\alpha$ and shape $\xi = 1/\alpha$. Vice versa one can get the Pareto distribution from the GPD by $x_m = \sigma/\xi$ and $\alpha = 1/\xi$.

Bounded Pareto distribution

The bounded (or truncated) Pareto distribution has three parameters: α, L and H. As in the standard Pareto distribution α determines the shape. L denotes the minimal value, and H denotes the maximal value.

The probability density function is

$$\frac{\alpha L^\alpha x^{-\alpha-1}}{1-\left(\frac{L}{H}\right)^\alpha},$$

where $L \le x \le H$, and $\alpha > 0$.

Bounded Pareto

Parameters	$L > 0$ location (real) $H > L$ location (real) $\alpha > 0$ shape (real)
Support	$L \leqslant x \leqslant H$
PDF	$\frac{\alpha L^\alpha x^{-\alpha-1}}{1-\left(\frac{L}{H}\right)^\alpha}$
CDF	$\frac{1-L^\alpha x^{-\alpha}}{1-\left(\frac{L}{H}\right)^\alpha}$
Mean	$\frac{L^\alpha}{1-\left(\frac{L}{H}\right)^\alpha} \cdot \left(\frac{\alpha}{\alpha-1}\right) \cdot \left(\frac{1}{L^{\alpha-1}} - \frac{1}{H^{\alpha-1}}\right), \alpha \neq 1$ $\frac{HL}{H-L} \ln \frac{H}{L}, \alpha = 1$
Median	$L\left(1-\frac{1}{2}\left(1-\left(\frac{L}{H}\right)^\alpha\right)\right)^{-\frac{1}{\alpha}}$
Variance	

Generating bounded Pareto random variables

If U is uniformly distributed on (0, 1), then applying inverse-transform

method [17]

$$U = \frac{1 - L^{\alpha} x^{-\alpha}}{1 - (\frac{L}{H})^{\alpha}}$$

	$\frac{L^{\alpha}}{1 - \left(\frac{L}{H}\right)^{\alpha}} \cdot \left(\frac{\alpha}{\alpha - 2}\right) \cdot \left(\frac{1}{L^{\alpha-2}} - \frac{1}{H^{\alpha-2}}\right), \alpha \neq 2$ $\frac{2H^2L^2}{H^2 - L^2} \ln \frac{H}{L}, \alpha = 2$ (this is the second raw moment, not the variance)
Skewness	$\frac{L^{\alpha}}{1 - \left(\frac{L}{H}\right)^{\alpha}} \cdot \frac{\alpha * (L^{k-\alpha} - H^{k-\alpha})}{(\alpha - k)}, \alpha \neq j$ (this is the kth raw moment, not the skewness)

$$x = \left(-\frac{UH^{\alpha} - UL^{\alpha} - H^{\alpha}}{H^{\alpha}L^{\alpha}}\right)^{-\frac{1}{\alpha}}$$

is a bounded Pareto-distributed.

Symmetric Pareto distribution

The purpose of Symmetric Pareto distribution and Zero Symmetric Pareto distribution is to capture some special statistical distribution with a sharp probability peak and symmetric long probability tails. These two distributions are derived from Pareto distribution. Long probability tail normally means that probability decays slowly. Pareto distribution performs fitting job in many cases. But if the distribution has symmetric structure with two slow decaying tails, Pareto could not do it. Then Symmetric Pareto or Zero Symmetric Pareto distribution is applied instead.[18]

The Cumulative distribution function (CDF) of Symmetric Pareto distribution is defined as following:[18]

$$F(X) = P(x < X) = \begin{cases} \frac{1}{2}(\frac{b}{2b - X})^{a} & X < b \\ 1 - \frac{1}{2}(\frac{b}{X})^{a} & X \geq b \end{cases}$$

The corresponding probability density function (PDF) is:[18]

$$p(x) = \frac{ab^{a}}{2(b + |x - b|)^{a+1}}, X \in R$$

This distribution has two parameters: a and b. It is symmetric by b. Then the mathematic expectation is b. When, it has variance as following:

$$E((x - b)^2) = \int_{-\infty}^{\infty} (x - b)^2 p(x) dx = \frac{2b^2}{(a - 2)(a - 1)}$$

The CDF of Zero Symmetric Pareto (ZSP) distribution is defined as following:

$$F(X) = P(x < X) = \begin{cases} \frac{1}{2}(\frac{b}{b - X})^{a} & X < 0 \\ 1 - \frac{1}{2}(\frac{b}{b + X})^{a} & X \geq 0 \end{cases}$$

The corresponding PDF is:

$$p(x) = \frac{ab^a}{2(b+|x|)^{a+1}}, X \in R$$

This distribution is symmetric by zero. Parameter a is related to the decay rate of probability and (a/2b) represents peak magnitude of probability.[18]

Multivariate Pareto distribution

The univariate Pareto distribution has been extended to a multivariate Pareto distribution.[19]

Statistical inference

Estimation of parameters

The likelihood function for the Pareto distribution parameters α and x_m, given an independent sample $x = (x_1, x_2, \ldots, x_n)$, is

$$L(\alpha, x_\mathrm{m}) = \prod_{i=1}^{n} \alpha \frac{x_\mathrm{m}^\alpha}{x_i^{\alpha+1}} = \alpha^n x_\mathrm{m}^{n\alpha} \prod_{i=1}^{n} \frac{1}{x_i^{\alpha+1}}.$$

Therefore, the logarithmic likelihood function is

$$\ell(\alpha, x_\mathrm{m}) = n \ln \alpha + n\alpha \ln x_\mathrm{m} - (\alpha+1) \sum_{i=1}^{n} \ln x_i.$$

It can be seen that $\ell(\alpha, x_\mathrm{m})$ is monotonically increasing with x_m, that is, the greater the value of x_m, the greater the value of the likelihood function. Hence, since $x \geq x_\mathrm{m}$, we conclude that

$$\widehat{x}_\mathrm{m} = \min_i x_i.$$

To find the estimator for α, we compute the corresponding partial derivative and determine where it is zero:

$$\frac{\partial \ell}{\partial \alpha} = \frac{n}{\alpha} + n \ln x_\mathrm{m} - \sum_{i=1}^{n} \ln x_i = 0.$$

Thus the maximum likelihood estimator for α is:

$$\widehat{\alpha} = \frac{n}{\sum_i \ln(x_i / \widehat{x}_\mathrm{m})}.$$

The expected statistical error is:[20]

$$\sigma = \frac{\widehat{\alpha}}{\sqrt{n}}.$$

Malik (1970)[21] gives the exact joint distribution of $(\hat{x}_{\mathrm{m}}, \hat{\alpha})$. In particular, $\hat{x}_{\mathrm{m}}$ and $\hat{\alpha}$ are independent and $\hat{x}_{\mathrm{m}}$ is Pareto with scale parameter x_{m} and shape parameter $n\alpha$, whereas $\hat{\alpha}$ has an inverse-gamma distribution with shape and scale parameters $n - 1$ and $n\alpha$, respectively.

Confidence intervals

Treating $\hat{\alpha}$ as known (i.e. $\alpha = \hat{\alpha}$) and using the result above by Malik (1970), the CDF of a Pareto(x_m, $n\hat{\alpha}$) distribution can be inverted to form a confidence interval for x_m, i.e. CDF($\hat{x}_m$) = $1 - \left(\frac{x_m}{\hat{x}_m}\right)^{n\hat{\alpha}}$. Additionally, α could be profiled. That is, for each hypothesis $H_0 : x_m \leq x_{m_0}$ re-estimate α under the restricted null space producing $\hat{\alpha}_0$and treat this as known. The lower one-sided 97.5% confidence limit is the hypothesis for x_m such that the upper-tailed probability of $\hat{x}_m$ or something more extreme is 2.5%. A similar approach can be used to obtain a confidence interval for α by inverting the CDF of an inverse-gamma distribution with shape and scale parameters $n - 1$ and $n\alpha$, respectively, i.e. CDF($\hat{\alpha}$) = $\frac{\Gamma\left(n-1, \frac{n\alpha}{\hat{\alpha}}\right)}{\Gamma(n-1)}$. The two-sided 95% confidence limits are the hypotheses for α such that the one-sided probability of $\hat{\alpha}$ or something more extreme is 2.5%.

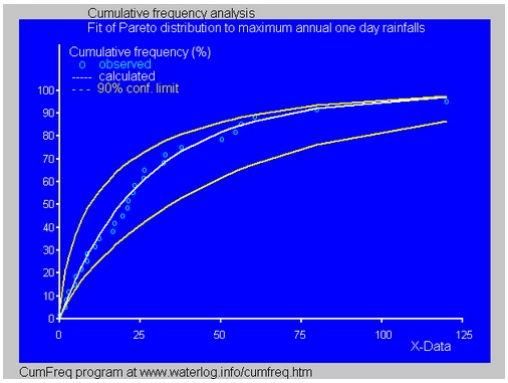
Occurrence and applications

General

Vilfredo Pareto originally used this distribution to describe the allocation of wealth among individuals since it seemed to show rather well the way that a larger portion of the wealth of any society is owned by a smaller percentage of the people in that society. He also used it to describe distribution of income.[4] This idea is sometimes expressed more simply as the Pareto principle or the "80-20 rule" which says that 20% of the population controls 80% of the wealth.[22] However, the 80-20 rule corresponds to a particular value of α, and in fact, Pareto's data on British income taxes in his *Cours d'économie politique* indicates that about 30% of the population had about 70% of the income. The probability density function (PDF) graph at the beginning of this article shows that the "probability" or fraction of the population that owns a small amount of wealth per person is rather high, and then decreases steadily as wealth increases. (The Pareto distribution is not realistic for wealth for the lower end, however. In fact, net worth may even be negative.) This distribution is not limited to describing wealth or income, but to many situations in which an equilibrium is found in the distribution of the "small" to the "large". The following examples are sometimes seen as approximately Pareto-distributed:

- The sizes of human settlements (few cities, many hamlets/villages)[23][24]
- File size distribution of Internet traffic which uses the TCP protocol (many smaller files, few larger ones)[23]
- Hard disk drive error rates[25]
- Clusters of Bose–Einstein condensate near absolute zero[26]
- The values of oil reserves in oil fields (a few large fields, many small fields)[23]

- The length distribution in jobs assigned to supercomputers (a few large ones, many small ones)[27]
- The standardized price returns on individual stocks [23]
- Sizes of sand particles [23]
- The size of meteorites
- Severity of large casualty losses for certain lines of business such as general liability, commercial auto, and workers compensation.[28][29]
- Amount of time a user on Steam will spend playing different games. (Some games get played a lot, but most get played almost never.) [2] (https://docs.google.com/spreadsheets/d/1AjtfgTQc1T84NCyJWGcCPN4jrVsOpX0bp0jgPZJEW6A/edit#gid=0)
- In hydrology the Pareto distribution is applied to extreme events such as annually maximum one-day rainfalls and river discharges.[30] The blue picture illustrates an example of fitting the Pareto distribution to ranked annually maximum one-day rainfalls showing also the 90% confidence belt based on the binomial distribution. The rainfall data are represented by plotting positions as part of the cumulative frequency analysis.
- In Electric Utility Distribution Reliability (80% of the Customer Minutes Interrupted occur on approximately 20% of the days in a given year).



Fitted cumulative Pareto (Lomax) distribution to maximum one-day rainfalls using CumFreq, see also distribution fitting

Relation to Zipf's law

The Pareto distribution is a continuous probability distribution. Zipf's law, also sometimes called the zeta distribution, is a discrete distribution, separating the values into a simple ranking. Both are a simple power law with a negative exponent, scaled so that their cumulative distributions equal 1. Zipf's can be derived from the Pareto distribution if the x values (incomes) are binned into N ranks so that the number of people in each bin follows a 1/rank pattern. The distribution is normalized by defining x_m so that $\alpha x_{\mathrm{m}}^{\alpha} = \dfrac{1}{H(N, \alpha - 1)}$ where $H(N, \alpha - 1)$ is the generalized harmonic number. This makes Zipf's probability density function derivable from Pareto's.

$$f(x) = \frac{\alpha x_{\mathrm{m}}^{\alpha}}{x^{\alpha+1}} = \frac{1}{x^s H(N, s)}$$

where $s = \alpha - 1$ and x is an integer representing rank from 1 to N where N is the highest income bracket. So a randomly selected person (or word, website link, or city) from a population (or language, internet, or country) has $f(x)$ probability of ranking x.

Relation to the "Pareto principle"

The "80–20 law", according to which 20% of all people receive 80% of all income, and 20% of the most affluent 20% receive 80% of that 80%, and so on, holds precisely when the Pareto index is $\alpha = \log_4 5 = \dfrac{\log_{10} 5}{\log_{10} 4} \approx 1.161$. This result can be derived from the Lorenz curve formula given below. Moreover, the following have been shown[31] to be mathematically equivalent:

- Income is distributed according to a Pareto distribution with index $\alpha > 1$.

- There is some number $0 \le p \le 1/2$ such that $100p$ % of all people receive $100(1-p)$% of all income, and similarly for every real (not necessarily integer) $n > 0$, $100p^n$ % of all people receive $100(1-p)^n$ percentage of all income. α and p are related by

$$1 - \frac{1}{\alpha} = \frac{\ln(1-p)}{\ln(p)} = \frac{\ln((1-p)^n)}{\ln(p^n)}$$

This does not apply only to income, but also to wealth, or to anything else that can be modeled by this distribution.

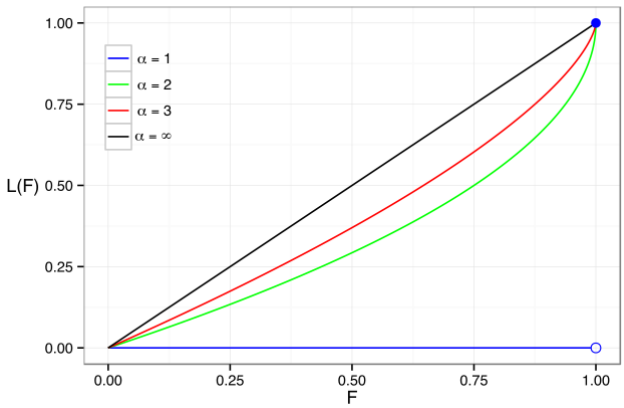
This excludes Pareto distributions in which $0 < \alpha \le 1$, which, as noted above, have an infinite expected value, and so cannot reasonably model income distribution.

Relation to Price's law

Price's square root law is sometimes offered as a property of or as similar to the Pareto distribution. However, the law only holds in the case that $\alpha = 1$. Note that in this case, the total and expected amount of wealth are not defined, and the rule only applies asymptotically to random samples. The extended Pareto Principle mentioned above is a far more general rule.

Lorenz curve and Gini coefficient

The Lorenz curve is often used to characterize income and wealth distributions. For any distribution, the Lorenz curve $L(F)$ is written in terms of the PDF f or the CDF F as



Lorenz curves for a number of Pareto distributions. The case $\alpha = \infty$ corresponds to perfectly equal distribution ($G = 0$) and the $\alpha = 1$ line corresponds to complete inequality ($G = 1$)

$$L(F) = \frac{\int_{x_\mathrm{m}}^{x(F)} x f(x)\,dx}{\int_{x_\mathrm{m}}^{\infty} x f(x)\,dx} = \frac{\int_0^F x(F')\,dF'}{\int_0^1 x(F')\,dF'}$$

where $x(F)$ is the inverse of the CDF. For the Pareto distribution,

$$x(F) = \frac{x_\mathrm{m}}{(1-F)^{\frac{1}{\alpha}}}$$

and the Lorenz curve is calculated to be

$$L(F) = 1 - (1-F)^{1-\frac{1}{\alpha}},$$

For $0 < \alpha \leq 1$ the denominator is infinite, yielding L=0. Examples of the Lorenz curve for a number of Pareto distributions are shown in the graph on the right.

According to Oxfam (2016) the richest 62 people have as much wealth as the poorest half of the world's population.[32] We can estimate the Pareto index that would apply to this situation. Letting ε equal $62/(7 \times 10^9)$ we have:

$$L(1/2) = 1 - L(1 - \varepsilon)$$

or

$$1 - (1/2)^{1-\frac{1}{\alpha}} = \varepsilon^{1-\frac{1}{\alpha}}$$

The solution is that α equals about 1.15, and about 9% of the wealth is owned by each of the two groups. But actually the poorest 69% of the world adult population owns only about 3% of the wealth.[33]

The Gini coefficient is a measure of the deviation of the Lorenz curve from the equidistribution line which is a line connecting [0, 0] and [1, 1], which is shown in black ($\alpha = \infty$) in the Lorenz plot on the right. Specifically, the Gini coefficient is twice the area between the Lorenz curve and the equidistribution line. The Gini coefficient for the Pareto distribution is then calculated (for $\alpha \geq 1$) to be

$$G = 1 - 2\left(\int_0^1 L(F)\,dF\right) = \frac{1}{2\alpha - 1}$$

(see Aaberge 2005).

Computational methods

Random sample generation

Random samples can be generated using inverse transform sampling. Given a random variate U drawn from the uniform distribution on the unit interval (0, 1], the variate T given by

$$T = \frac{x_{\mathrm{m}}}{U^{1/\alpha}}$$

is Pareto-distributed.[34] If U is uniformly distributed on [0, 1), it can be exchanged with (1 − U).

See also

- Bradford's law
- Gutenberg–Richter law
- Matthew effect
- Pareto analysis
- Pareto efficiency
- Pareto interpolation
- Power law probability distributions
- Sturgeon's law
- Traffic generation model

- Zipf's law
- Heavy-tailed distribution

References

1. Amoroso, Luigi (1938). "VILFREDO PARETO". *Econometrica (Pre-1986); Jan 1938; 6, 1; ProQuest*. **6**.
2. "Pareto" (https://www.merriam-webster.com/dictionary/Pareto). *Merriam-Webster Dictionary*. Retrieved 28 July 2019.
3. Pareto, Vilfredo (1898). "Cours d'economie politique" (https://zenodo.org/record/2144014). *Journal of Political Economy*. **6**. doi:10.1086/250536 (https://doi.org/10.1086%2F250536).
4. Pareto, Vilfredo, *Cours d'Économie Politique: Nouvelle édition par G.-H. Bousquet et G. Busino*, Librairie Droz, Geneva, 1964, pp. 299–345. Original book archived (https://web.archive.org/web/20130531151249/http://www.institutcoppet.org/wp-content/uploads/2012/05/Cours-d économie-politique-Tome-II-Vilfredo-Pareto.pdf)
5. VAN MONTFORT, M.A.J. (1986). "The Generalized Pareto distribution applied to rainfall depths" (https://doi.org/10.1080%2F02626668609491037). *Hydrological Sciences Journal*. **31** (2): 151–162. doi:10.1080/02626668609491037 (https://doi.org/10.1080%2F02626668609491037).
6. Oancea, Bogdan (2017). "Income inequality in Romania: The exponential-Pareto distribution". *Physica A: Statistical Mechanics and Its Applications*. **469**: 486–498. Bibcode:2017PhyA..469..486O (https://ui.adsabs.harvard.edu/abs/2017PhyA..469..486O). doi:10.1016/j.physa.2016.11.094 (https://doi.org/10.1016%2Fj.physa.2016.11.094).
7. Barry C. Arnold (1983). *Pareto Distributions*. International Co-operative Publishing House. ISBN 978-0-89974-012-6.
8. S. Hussain, S.H. Bhatti (2018). Parameter estimation of Pareto distribution: Some modified moment estimators (https://www.researchgate.net/publication/322758024_Parameter_estimation_of_Pareto_distribution_Some_modified_moment_estimators). *Maejo International Journal of Science and Technology* 12(1):11-27
9. Eliazar, Iddo (November 2017). "Lindy's Law". *Physica A: Statistical Mechanics and Its Applications*. **486**: 797–805. Bibcode:2017PhyA..486..797E (https://ui.adsabs.harvard.edu/abs/2017PhyA..486..797E). doi:10.1016/j.physa.2017.05.077 (https://doi.org/10.1016%2Fj.physa.2017.05.077).
10. Johnson NL, Kotz S, Balakrishnan N (1994) Continuous univariate distributions Vol 1. Wiley Series in Probability and Statistics.
11. Johnson, Kotz, and Balakrishnan (1994), (20.4).
12. Christian Kleiber & Samuel Kotz (2003). *Statistical Size Distributions in Economics and Actuarial Sciences* (https://books.google.com/books?id=7wLGjyB128IC). Wiley. ISBN 978-0-471-15064-0.
13. Feller, W. (1971). *An Introduction to Probability Theory and its Applications*. **II** (2nd ed.). New York: Wiley. p. 50. "The densities (4.3) are sometimes called after the economist *Pareto*. It was thought (rather naïvely from a modern statistical standpoint) that income distributions should have a tail with a density $\sim Ax^{-a}$ as $x \to \infty$."
14. Lomax, K. S. (1954). "Business failures. Another example of the analysis of failure data". *Journal of the American Statistical Association*. **49** (268): 847–52. doi:10.1080/01621459.1954.10501239 (https://doi.org/10.1080%2F01621459.1954.10501239).
15. Chotikapanich, Duangkamon (16 September 2008). "Chapter 7: Pareto and Generalized Pareto Distributions" (https://books.google.com/books?id=fUJZZLj1kbwC). *Modeling Income Distributions and Lorenz Curves*. pp. 121–22. ISBN 9780387727967.
16. White, Gentry (2006). *Bayesian semiparametric spatial and joint spatio-temporal modeling* (https://mospace.umsystem.edu/xmlui/handle/10355/4450) (Thesis thesis). University of Missouri--Columbia. section 5.3.1
17. http://www.cs.bgu.ac.il/~mps042/invtransnote.htm

18. Huang, Xiao-dong (2004). "A Multiscale Model for MPEG-4 Varied Bit Rate Video Traffic". *IEEE Transactions on Broadcasting*. **50** (3): 323–334. doi:10.1109/TBC.2004.834013 (https://doi.org/10.1109%2FTBC.2004.834013).
19. Rootzén, Holger; Tajvidi, Nader (2006). "Multivariate generalized Pareto distributions". *Bernoulli*. **12** (5): 917–30. CiteSeerX 10.1.1.145.2991 (https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.145.2991). doi:10.3150/bj/1161614952 (https://doi.org/10.3150%2Fbj%2F1161614952).
20. M. E. J. Newman (2005). "Power laws, Pareto distributions and Zipf's law". *Contemporary Physics*. **46** (5): 323–51. arXiv:cond-mat/0412004 (https://arxiv.org/abs/cond-mat/0412004). Bibcode:2005ConPh..46..323N (https://ui.adsabs.harvard.edu/abs/2005ConPh..46..323N). doi:10.1080/00107510500052444 (https://doi.org/10.1080%2F00107510500052444). S2CID 202719165 (https://api.semanticscholar.org/CorpusID:202719165).
21. H. J. Malik (1970). "Estimation of the Parameters of the Pareto Distribution". *Metrika*. **15**: 126–132. doi:10.1007/BF02613565 (https://doi.org/10.1007%2FBF02613565). S2CID 124007966 (https://api.semanticscholar.org/CorpusID:124007966).
22. For a two-quantile population, where approximately 18% of the population owns 82% of the wealth, the Theil index takes the value 1.
23. Reed, William J.; et al. (2004). "The Double Pareto-Lognormal Distribution – A New Parametric Model for Size Distributions". *Communications in Statistics – Theory and Methods*. **33** (8): 1733–53. CiteSeerX 10.1.1.70.4555 (https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.70.4555). doi:10.1081/sta-120037438 (https://doi.org/10.1081%2Fsta-120037438). S2CID 13906086 (https://api.semanticscholar.org/CorpusID:13906086).
24. Reed, William J. (2002). "On the rank-size distribution for human settlements". *Journal of Regional Science*. **42** (1): 1–17. doi:10.1111/1467-9787.00247 (https://doi.org/10.1111%2F1467-9787.00247). S2CID 154285730 (https://api.semanticscholar.org/CorpusID:154285730).
25. Schroeder, Bianca; Damouras, Sotirios; Gill, Phillipa (2010-02-24). "Understanding latent sector error and how to protect against them" (http://www.usenix.org/event/fast10/tech/full_papers/schroeder.pdf) (PDF). *8th Usenix Conference on File and Storage Technologies (FAST 2010)*. Retrieved 2010-09-10. "We experimented with 5 different distributions (Geometric,Weibull, Rayleigh, Pareto, and Lognormal), that are commonly used in the context of system reliability, and evaluated their fit through the total squared differences between the actual and hypothesized frequencies (χ^2 statistic). We found consistently across all models that the geometric distribution is a poor fit, while the Pareto distribution provides the best fit."
26. Yuji Ijiri; Simon, Herbert A. (May 1975). "Some Distributions Associated with Bose–Einstein Statistics" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC432601). *Proc. Natl. Acad. Sci. USA*. **72** (5): 1654–57. Bibcode:1975PNAS...72.1654I (https://ui.adsabs.harvard.edu/abs/1975PNAS...72.1654I). doi:10.1073/pnas.72.5.1654 (https://doi.org/10.1073%2Fpnas.72.5.1654). PMC 432601 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC432601). PMID 16578724 (https://pubmed.ncbi.nlm.nih.gov/16578724).
27. Harchol-Balter, Mor; Downey, Allen (August 1997). "Exploiting Process Lifetime Distributions for Dynamic Load Balancing" (https://users.soe.ucsc.edu/~scott/courses/Fall11/221/Papers/Sync/harcholbalter-tocs97.pdf) (PDF). *ACM Transactions on Computer Systems*. **15** (3): 253–258. doi:10.1145/263326.263344 (https://doi.org/10.1145%2F263326.263344). S2CID 52861447 (https://api.semanticscholar.org/CorpusID:52861447).
28. Kleiber and Kotz (2003): p. 94.
29. Seal, H. (1980). "Survival probabilities based on Pareto claim distributions" (https://doi.org/10.1017%2FS0515036100006620). *ASTIN Bulletin*. **11**: 61–71. doi:10.1017/S0515036100006620 (https://doi.org/10.1017%2FS0515036100006620).
30. CumFreq, software for cumulative frequency analysis and probability distribution fitting [1] (https://www.waterlog.info/cumfreq.htm)
31. Hardy, Michael (2010). "Pareto's Law". *Mathematical Intelligencer*. **32** (3): 38–43. doi:10.1007/s00283-010-9159-2 (https://doi.org/10.1007%2Fs00283-010-9159-2). S2CID 121797873 (https://api.semanticscholar.org/CorpusID:121797873).

32. "62 people own the same as half the world, reveals Oxfam Davos report" (https://www.oxfam.org/en/pressroom/pressreleases/2016-01-18/62-people-own-same-half-world-reveals-oxfam-davos-report). Oxfam. Jan 2016.
33. "Global Wealth Report 2013" (https://web.archive.org/web/20150214155424/https://publications.credit-suisse.com/tasks/render/file/?fileID=BCDB1364-A105-0560-1332EC9100FF5C83). Credit Suisse. Oct 2013. p. 22. Archived from the original (https://publications.credit-suisse.com/tasks/render/file/?fileID=BCDB1364-A105-0560-1332EC9100FF5C83) on 2015-02-14. Retrieved 2016-01-24.
34. Tanizaki, Hisashi (2004). *Computational Methods in Statistics and Econometrics* (https://books.google.com/books?id=pOGAUcn13fMC). CRC Press. p. 133. ISBN 9780824750886.

Notes

- M. O. Lorenz (1905). "Methods of measuring the concentration of wealth". *Publications of the American Statistical Association*. **9** (70): 209–19. Bibcode:1905PAmSA...9..209L (https://ui.adsabs.harvard.edu/abs/1905PAmSA...9..209L). doi:10.2307/2276207 (https://doi.org/10.2307%2F2276207). JSTOR 2276207 (https://www.jstor.org/stable/2276207).
- Pareto, Vilfredo (1965). Librairie Droz (ed.). *Ecrits sur la courbe de la répartition de la richesse*. Œuvres complètes : T. III. p. 48. ISBN 9782600040211.
- Pareto, Vilfredo (1895). "La legge della domanda". *Giornale Degli Economisti*. **10**: 59–68.
- Pareto, Vilfredo (1896). "Cours d'économie politique". doi:10.1177/000271629700900314 (https://doi.org/10.1177%2F000271629700900314). S2CID 143528002 (https://api.semanticscholar.org/CorpusID:143528002).

External links

- "Pareto distribution" (https://www.encyclopediaofmath.org/index.php?title=Pareto_distribution), *Encyclopedia of Mathematics*, EMS Press, 2001 [1994]
- Weisstein, Eric W. "Pareto distribution" (https://mathworld.wolfram.com/ParetoDistribution.html). *MathWorld*.
- Aabergé, Rolf (May 2005). "Gini's Nuclear Family". *International Conference to Honor Two Eminent Social Scientists* (http://www3.unisi.it/eventi/GiniLorenz05/25%20may%20paper/PAPER_Aaberge.pdf) (PDF).
- Crovella, Mark E.; Bestavros, Azer (December 1997). *Self-Similarity in World Wide Web Traffic: Evidence and Possible Causes* (https://web.archive.org/web/20160304190612/http://www.cs.bu.edu/~crovella/paper-archive/self-sim/journal-version.pdf) (PDF). IEEE/ACM Transactions on Networking. **5**. pp. 835–846. Archived from the original (https://www.cs.bu.edu/~crovella/paper-archive/self-sim/journal-version.pdf) (PDF) on 2016-03-04. Retrieved 2019-02-25.
- syntraf1.c (http://www.csee.usf.edu/~kchriste/tools/syntraf1.c) is a C program to generate synthetic packet traffic with bounded Pareto burst size and exponential interburst time.

Retrieved from "https://en.wikipedia.org/w/index.php?title=Pareto_distribution&oldid=1036538430"

This page was last edited on 1 August 2021, at 05:51 (UTC).

Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.