# The Pareto Distribution

Background

Power Function

Consider an arbitrary power function, $x \mapsto k\,x^{\alpha}$ where k is a constant and the exponent α governs the relationship. Note that if $y = k\,x^{\alpha}$, then $\text{Log}[y] = \text{Log}[k] + \alpha\,\text{Log}[x]$. That is, the logarithmic transformation of this power function is linear in $\text{Log}[x]$. Another way to say this is that the elasticity of y with respect to x is constant: $\frac{d\,\text{Log}[y]}{d\,\text{Log}[x]} = \alpha$. Since the elasticity does not depend on the size of x, we sometimes say that the power relationship is scale invariant.

Two Power Functions
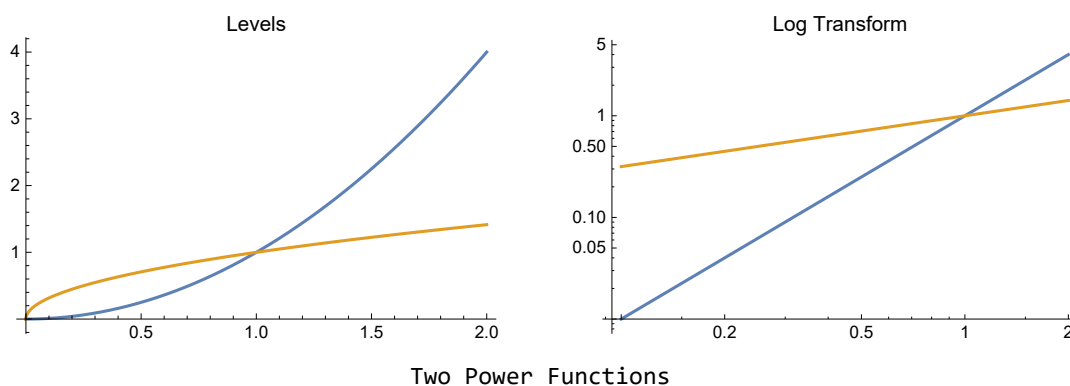
Power Law

A power law is a theoretical or empirical relationship governed by a power function.

- In geometry, the area of a regular polygon is proportional to the square of the length of a side.

- In physics, the gravitational attraction of two objects is inversely proportional to the square of their distance.

- In ecology, Taylor's Law states that the variance of population density is a power-function of mean population density.

- In economics, Gabaix (1999) finds the population of cities follows a power law (with an inequality parameter close to 1 (see below).

- In economics, Luttmer (2007) finds the distribution of employment in US firms follows a power law with inequality parameter close to 1.

■ In economics and business, the Pareto Principle (or 80-20 rule) says that 80% of income accrues to the top 20% of income recipients.

Pareto's Principle

In 1897, Vilfredo Pareto (1848-1923) proposed that the number of people (N_x) with incomes higher than x can be modeled as a power law:

$N_x = A/x^\alpha = A\,x^{-\alpha}$

Let the total population be N_0, and let the minimum income be x_0. Then $N_0 = A\,x_0^\alpha$, and we can write this in proportionate terms:

$N_x/N_0 = (x/x_0)^{-\alpha}$

Normalize $x_0 = 1$. For ease in discussion, we will then call x relative income. Using this notation, we may say that Pareto proposed that $n = x^{-\alpha}$ for $x \geq 1$, where $n = N_x/N_0$ is the proportion of rich in the population (i.e., those with relative income greater than x). Assuming $\alpha > 1$, we can find (by integration) that the area under this curve is $1/(\alpha - 1)$, that the proportion of that area that lies before any point x_0 is $1 - x_0^{1-\alpha}$, and that the proportion of that area that lies past any point $\tilde{x}$ is correspondingly $\tilde{x}^{1-\alpha}$.

Pareto's Principle: Graphical Illustration
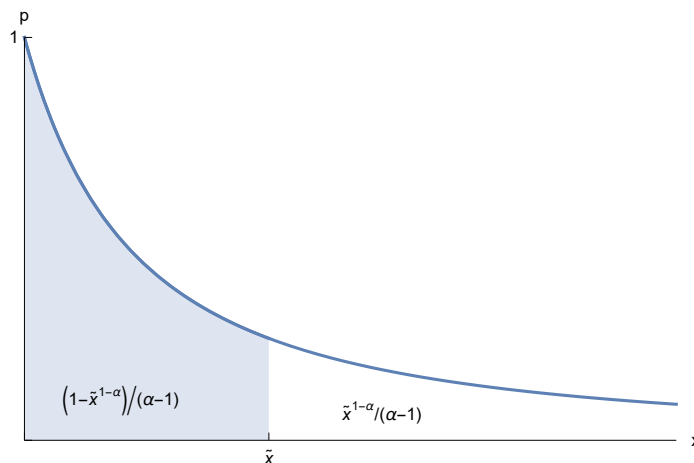
Digression on Integration Details

From the basic principles of integration, we know that the antiderivative of $x^{-\alpha}$ is $x^{1-\alpha}/(1-\alpha)$.

```
Integrate[x^-α, x]
```

$\dfrac{x^{1-\alpha}}{1-\alpha}$

Therefore the definite integral over the interval $[1 .. x_0]$ is $(1 - x_0^{1-\alpha})/(1 - \alpha)$.

```
Simplify[
   1 - x0^(1-α)
   ──────────── == Integrate[x^-α, {x, 1, x0}],
      α - 1
  Assumptions → α > 1 && x0 > 1
]
```

True

Correspondingly, the total area under the curve is $1/(\alpha - 1)$.

```
Integrate[x^-α, {x, 1, ∞}] ~ Simplify ~ (Assumptions → α > 1)
```

$$\frac{1}{-1 + \alpha}$$

The 80-20 Rule

Suppose we are interested in the fraction of total income received by the top 20% of income recipients.

Under the Pareto principle that $n = x^{-\alpha}$, we have seen that the share of total income received by those with relative incomes above x can then be written as $s = x^{1-\alpha}$.

We can also invert the Pareto principle to yield $x = n^{-1/\alpha}$.

Combining these two observations, for a given proportion n of top income earners, we can find the associated top-share of income as $s = (n^{-1/\alpha})^{1-\alpha} = n^{1-1/\alpha}$.

```
Eliminate[n == x^-α && s == x^(1-α), x] // Quiet
Solve[%, α] // PowerExpand // Simplify
```

$$s^{\frac{1}{1-\alpha}} == \left(\frac{1}{n}\right)^{\frac{1}{\alpha}}$$

$$\left\{\left\{\alpha \to \frac{\text{Log}[n]}{\text{Log}[n] - \text{Log}[s]}\right\}\right\}$$

For example, for Pareto's 80-20 rule to hold, we must have $\alpha \approx 1.16$.

```
   Log[n]
─────────────── /. {s → 0.80, n → 0.20}
Log[n] - Log[s]
```

1.16096

Deducing the Income Share of the Richest

Given α we can compute the total income share of the proportion p of the richest income recipients.
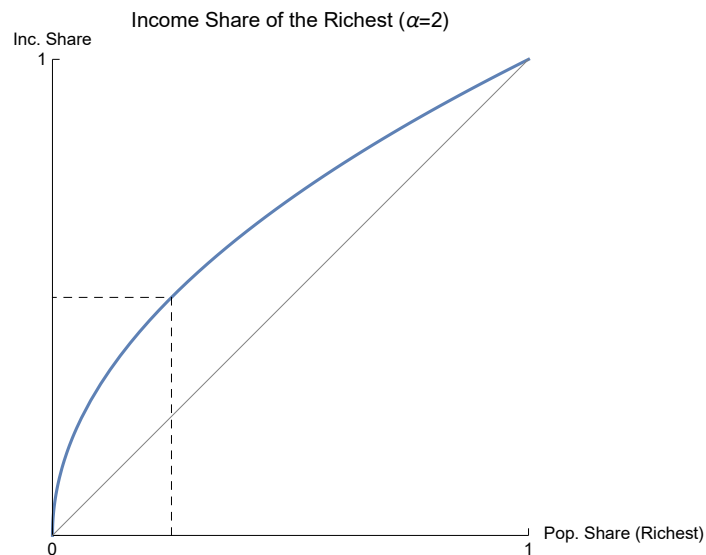
For example, if $\alpha == 2$ then the richest 1% of income recipients receive 10% of total income.

```
p^(1-1/α) /. {p → 0.01, α → 2}
```

```
0.1
```

It naturally follows that the rest of the population (i.e., the poorest 99% of income recipients) receive the rest of income (i.e., 90% of total income).

Income Share of the Richest: Illustrated

Income Share of the Richest ($\alpha=2$)

Income Share of the Poorest (Lorenz Curve)

It is somewhat more common to rotate this plot $180°$ to display the income share of the poorest. The result is usually called a Lorenz curve. (Note however Lorenz (1905) plotted the cumulative population share against the cumulative wealth share. I.e., our Lorenz curve is the inverse of his—a reflection through the $45°$ line.)

Income Share of the Poorest ($\alpha=2$)
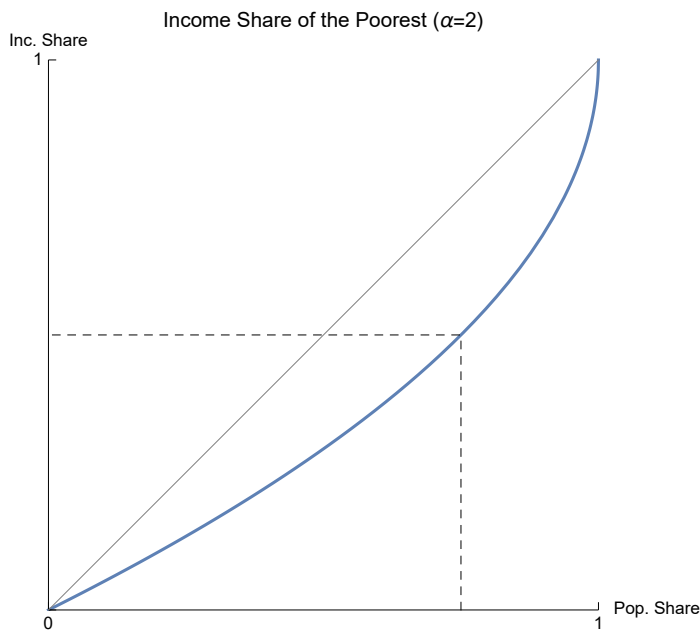
Pareto Inequality

It is common to report $\eta = 1/\alpha$ as the measure of Pareto inequality. Since we generally think $\alpha \in (1, 2)$, we should find $\eta \in (0.5, 1)$. Jones (2015) claims that in the contemporary US, $\eta \approx 0.6$ for the distribution of income. This should correspond to a top 1% share of around 16%.

```
p^(1-1/α) /. {p → 0.01, α → 10 / 6}
```

```
0.158489
```

Power-Law Distributions

Continuous Power-Law Probability Distribution

A continuous power-law distribution with shape parameter $\alpha>0$ and size parameter $x_0 > 0$ has probability distribution function $p(x) = k\,x^{-(1+\alpha)}$ for $x > x_0$. The antiderivative is

$$-\frac{k\,x^{-\alpha}}{\alpha}$$

We can therefore compute the (improper) integral over $[x_0 \,..\, \infty]$ to be

$$\frac{k\,x0^{-\alpha}}{\alpha}$$

The constant k must be chosen to satisfy normalization (unitarity), since the total area under the PDF must equal 1. This means that k must be

$\mathrm{x0}^{\alpha}\,\alpha$

Plugging in our solution for the constant of integration back into our PDF, we fully characterize of our power-law distribution in terms of two parameters: the shape parameter (α) and the size parameter (x_0).

$\mathrm{x}^{-1-\alpha}\,\mathrm{x0}^{\alpha}\,\alpha$

This distribution is usually known as the Pareto distribution, and we will soon relate it to the Pareto principle. (It is sometimes known as the Bradford distribution, after Bradford (1934), but this term also refers to a related truncated distribution.)

The Pareto Distribution

The social sciences have found that the Pareto distribution embodies a useful power law. The Pareto Distribution is most often presented in terms of its survival function, which gives the probability of seeing larger values than x. (This is often known as the complementary CDF, since it is 1-CDF. It is sometimes called the reliability function or the tail function.) The survival function of a Pareto distribution for $x \in [x_0 \, .. \, \infty]$ is

$$\left(\frac{\mathrm{x}}{\mathrm{x0}}\right)^{-\alpha}$$

This value of this survival function is initially 1 and declines to 0 as x increases. It defines a continuous probability distribution on $[x_0 \, .. \, \infty]$.

We are only interested in $x > x_0$, and we are usually interested in $\alpha > 1$ (which is required for finite mean value). We call $x_0 > 0$ the location parameter; we call $\alpha > 0$ the shape parameter (or slope parameter, or Pareto index); and we say the distribution is Pareto$[x_0, \alpha]$.

Survival: Intuition

Consider a population of households and suppose sampling household incomes is like sampling from a Pareto[10000,2].
What proportion of people earn more than \$100000 (i.e., ten times the minimum)? From the form of the survival function $(x/x_0)^{-2}$, it should be obvious that the answer is 10^{-2}: only 1 in 100 households earn more than \$100000.

Elaborating, we see that for $\alpha = 2$ and any x_0, we find that 1% of the population has income greater than $10 * x_0$. This is one way in which the Pareto distribution (along with other power law distributions) is *scale free*.

```
Simplify[
 SurvivalFunction[ParetoDistribution[x0, 2], 10 * x0],
 Assumptions → x0 > 0]
```

$$\frac{1}{100}$$

What is more, this scale-free relationship holds as well for subgroups: only 1% of the top 1% will

have incomes that are another ten times higher.

```
Simplify[
 SurvivalFunction[ParetoDistribution[x0, 2], 10 * 10 * x0],
 Assumptions → x0 > 0]
```

$$\frac{1}{10\,000}$$

Survival and the Pareto Index

To simplify comparisons, let us work with a normalized Pareto distribution: $S[x] = x^{-\alpha}$ for $x \geq 1$. That is, we normalize $x_0 = 1$. This lets us work with a single parameter: α, the shape parameter of the Pareto distribution.

Pareto Survival Functions
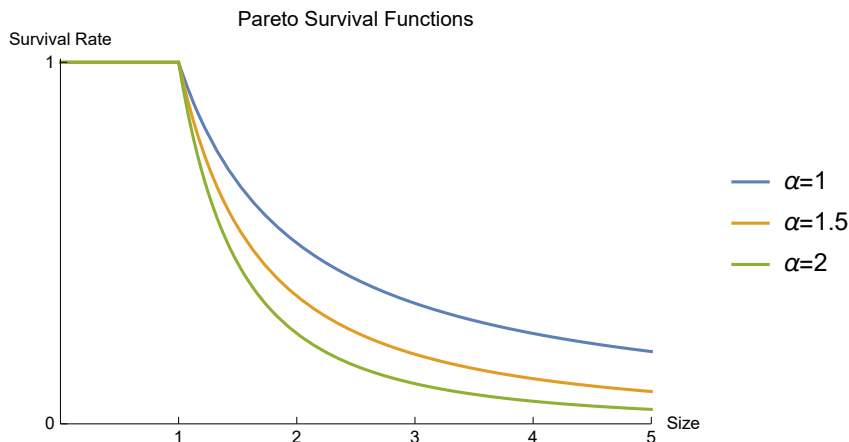
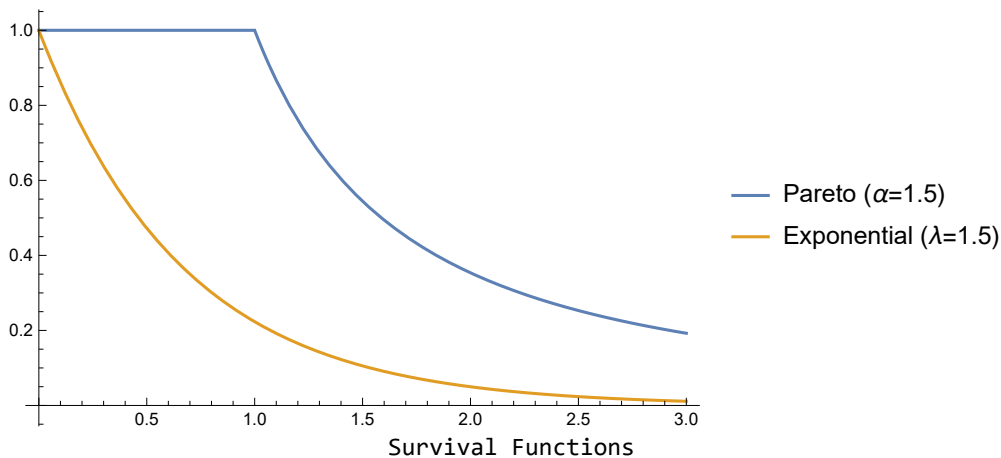## Contrast with the Exponential Distribution

The Exponential Distribution

Let us briefly compare the Pareto distribution to the exponential distribution, which may initially seem similar. The survival function of the exponential distribution is $S[x] = e^{-\lambda x}$ for $x \geq 0$, where $\lambda > 0$ is the shape parameter of the distribution. (Correspondingly the CDF is $F[x] = 1 - e^{-\lambda x}$ and the PDF is $f[x] = \lambda e^{-\lambda x}$.)

Survival Rate

Survival Function (Exponential[1.5])
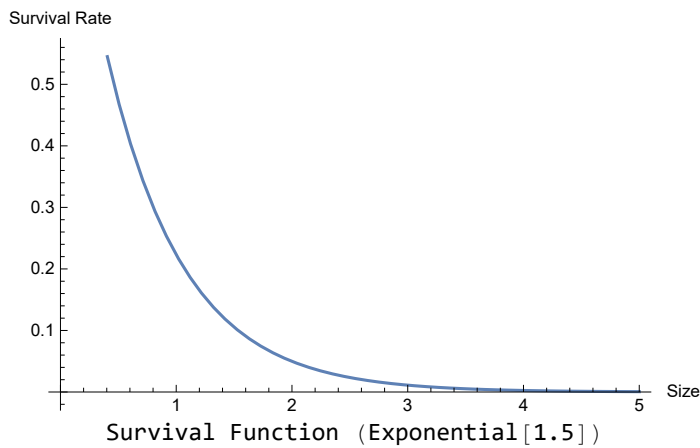
Contrast with the Exponential Distribution

Because the survival rate in the tail is higher for the Pareto distribution than for the exponential, we say that the Pareto distribution has a fat tail. We can begin to see the difference by plotting the survival functions.

Pareto (α=1.5)

Exponential (λ=1.5)

Survival Functions

Further Contrast with the Exponential Distribution

Recall that the survival function of the exponential distribution is $S[x] = e^{-\lambda x}$. At first sight the Pareto distribution may seem to have much in common with the exponential distribution. However, the survival rate of the Pareto distribution declines much more slowly. Here is a way to consider that contrast: for x1, x2 > x_0 and associated N1, N2, the Pareto distribution implies

$\log (N_1 / N_2) = -\alpha \log (x_1 / x_2)$

whereas for the exponential distribution

$\log (N_1 / N_2) = -\lambda (x_1 - x_2)$

Under a Pareto distribution, relative survival depends only on the ratio (x_1 / x_2), so the same relationship holds anywhere tail of the income distribution, no matter how far out. If the top 20%

of people receive 80% of income, then the top 4% (20% of 20%) receive 64% (80% of 80%) of income, and so on.

Fat-Tailed Distributions

Most popular probability distributions have well defined means, variances, and higher-order moments. For example, the exponential distribution with parameter $\lambda > 0$ has a mean of $1/\lambda$ and a variance of $1/\lambda^2$. For such distributions, outcomes far from the mean are very rare. Other distributions have "fat" tails: outcomes far from the mean are less rare. For example, the Pareto distribution has infinite variance if $\alpha \le 2$.

A probability distribution is said to be *fat-tailed* if eventually (i.e., as x gets big) the PDF is proportional to a power function of the form $x \mapsto x^{-(1+\alpha)}$ where $\alpha > 0$.

Equivalently, the survival function $P[X > x]$ is eventually proportional to a power function of the form $x \mapsto x^{-\alpha}$ where $\alpha > 0$. Contrast this with the survival function for an exponential distribution: $e^{-\lambda x}$. No matter how small we make $\lambda > 0$, we will find $e^{-\lambda x}/x^{-\alpha}$ is eventually tiny. Any power law distribution eventually has a much bigger tail than any exponential distribution.

```
Assuming[λ > 0 && α > 0,
 Limit[Exp[-λ x] / x^-α, x → ∞]
]
```
0

Algebraic Details (Fat Tail)

We can show more formally that survival function declines more rapidly for the exponential than for the Pareto distribution.

$\lim_{x \to \infty} e^{-\lambda x + \alpha \ln x} = \text{Exp}[\lim_{x \to \infty} -\lambda x + \alpha \ln x]$ by continuity, and we can write the latter as $\text{Exp}[\lim_{x \to \infty} (-\lambda x/\ln x + \alpha) \ln x]$. Recall that $\infty == \lim_{x \to \infty} (x/\ln x)$, or apply L'Hospital's rule to show it.

```
Assuming[λ > 0 && α > 0,
 Exp[Limit[(-λ x + α) Log[x], x → ∞]]
]
```
0

Intimate Relation of Pareto to Exponential

There is an intimate relationship between the Pareto and exponential distributions. Recall that the survival function of the exponential distribution is $e^{-\lambda x}$.

Let Y be exponentially distributed, with survival function $P[Y > y] = (e^y)^{-\alpha}$. (For convenience we

have changed the name of the shape parameter.) Define a new random variable by $X = x_0 \, e^Y$. Then compute the survival probability

$P[X>x]=P[x0 \, e^Y > x] = P[e^Y > x/x0] = P[Y > \text{Log}[x/x0]] = (x/x0)^{-\alpha}$

Comparing to our survival function for the Pareto distribution, we see that X has a Pareto distribution.

Some Pareto Details

Pareto Distribution: CDF and PDF

Recall the that the Pareto survival function for values greater than the minimum is $(x/x_0)^{-\alpha}$. (Recall that x_0 is a minimum value, called the location parameter.) Correspondingly, the cumulative distribution function (CDF) gives the probability of seeing a given size or lower.

```
cdfPareto = Simplify[
   CDF[ParetoDistribution[x0, α], x],
   x > x0 > 0]
```

$$1 - \left(\frac{x0}{x} \right)^{\alpha}$$

As is evident by inspection, this is just the complement of the survival function.

```
Simplify[
 cdfPareto == 1 - SurvivalFunction[ParetoDistribution[x0, α], x],
 x > x0 > 0]
```

True

CDF vs Survival Function of Pareto Distributions

Recall the that the Pareto survival function for values greater than the minimum is $(x/x_0)^{-\alpha}$. Normalize $x_0 == 1$ for presentational simplicity, so that we have $n == x^{-\alpha}$.

Pareto Distribution (α=1.5)
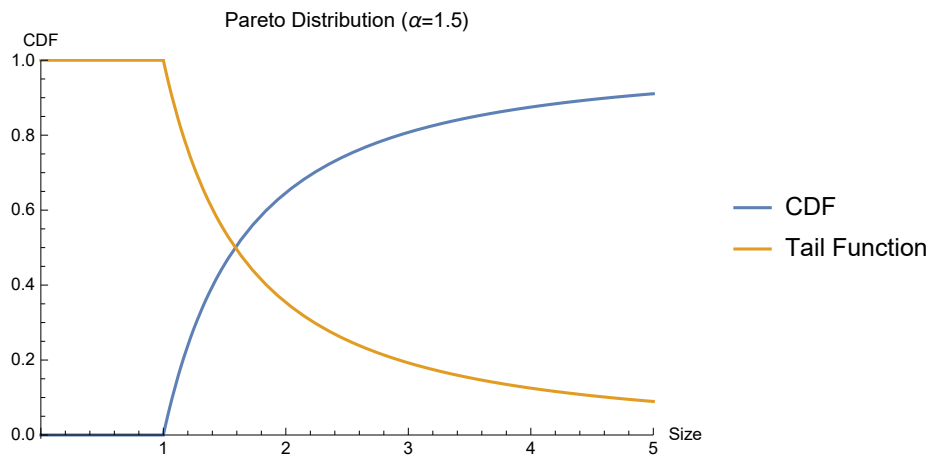
PDF of Pareto Distributions

The probability distribution function (PDF) gives the likelihood of an outcome occurring near any particular possible value.

```
pdfPareto = Simplify[
   PDF[ParetoDistribution[x0, α], x],
   x > x0 > 0]
```

$x^{-1-\alpha} \, x0^{\alpha} \, \alpha$

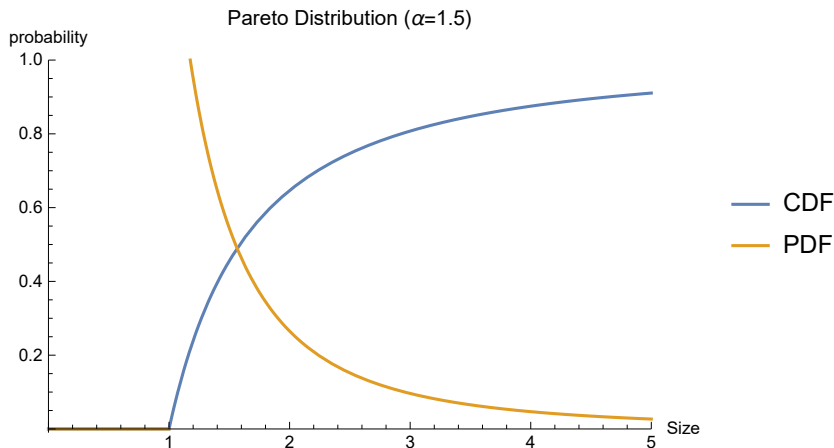Recall that the PDF is the derivative of the CDF.

```
pdfPareto == D[cdfPareto, x] // PowerExpand
```

True

We may say the the PDF gives us the marginal rate at which the cumulative probability (given by the CDF) is increasing.

CDF and PDF of Pareto Distributions: Illustration

Pareto Distribution (α=1.5)

Log-Linear Survival

Recall that the survival function of the normalized Pareto can be characterized by $n == x^{-\alpha}$. Take the logarithm of both sides:

```
Log /@ (n == x⁻ᵅ) // PowerExpand  (* 11.3+, use ApplySides instead of Map *)
```
$$\text{Log}[n] == -\alpha \, \text{Log}[x]$$

Note that we have a linear relationship in logs. We can therefore say that the size elasticity of the survival rate is the constant α. (We will return to this.)

Similarly, since the CDF is $1 - x^{-\alpha}$, the PDF is $\alpha x^{-(\alpha+1)}$, which a log transformation again linearizes.

```
Simplify[
   Log[α x⁻⁽ᵅ⁺¹⁾],
   Assumptions → α > 0 && x > 0] // PowerExpand
```
$$(-1 - \alpha) \, \text{Log}[x] + \text{Log}[\alpha]$$

PDF of Pareto Distributions: Loglinearity

PDF of Pareto[1,1.5]
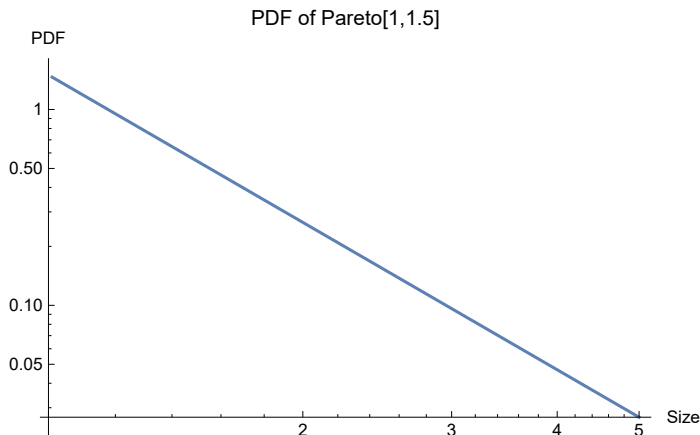
Average Size

The PDF of the Pareto[x0,α] distribution is $(\alpha/x_0)(x/x_0)^{-(\alpha+1)}$.

Use this PDF to compute the average size of a draw from the Pareto distribution as $\int_{x0}^{\infty} x\, p(x)\, dx == \int_{x0}^{\infty} \alpha \left(\frac{x}{x_0}\right)^{-\alpha} dx$. This integral only exists for $\alpha > 1$, in which case it is

$$\frac{x0\,\alpha}{-1+\alpha}$$

Here is a computation of this result using WL's builtin functions.

```
Simplify[
 Mean@ParetoDistribution[x0, α],
 x0 > 0 && α > 1]
```

$$\frac{x0\,\alpha}{-1+\alpha}$$

Proportion of Total Income

If income follows a Pareto[x0,α] distribution, then the weighted sum of all incomes less than or equal to *t* is

$$\int_{x0}^{t} x\, p[x]\, dx$$

$$\frac{x0\,\alpha}{-1+\alpha} - \frac{t^{1-\alpha}\,x0^{\alpha}\,\alpha}{-1+\alpha}$$

Dividing by the mean (i.e., the probability weighted sum of all incomes) produces an expression for the proportion of total income constituted by incomes of *t* or less: $1 - (t/x0)^{1-\alpha}$.

```
cumShare = cumSize / meanPareto // Simplify
```
$1 - t^{1-\alpha} x0^{-1+\alpha}$

Note that this expression only makes sense for $\alpha > 1$, the case in which the mean exists. Also note that the value of the result $1 - \left(\frac{t}{x_0}\right)^{1-\alpha}$ depends only on the value of the ratio t/x_0.

Lorenz Curve and Gini Coefficient
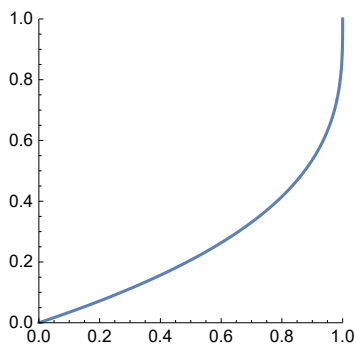
Lorenz Curve: Parametric Plot

A Lorenz curve for x (e.g., income or wealth) plots the cumulative share of x vs the cumulative share of the population possessing it.

We have just found that the cumulative share of income for incomes up to t can be written as $1 - (t/x0)^{1-\alpha}$.

Additionally, we have see that the CDF at income t (which gives the proportion of the population earning no more than t) is $1 - (t/x0)^{-\alpha}$.

So for any alpha, we can make a parametric plot of the Lorenz curve. Defining $m = x0/t$ we can write:
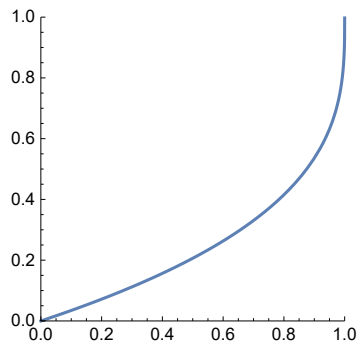
```
With[{α = 1.5},
 ParametricPlot[{1 - mᵅ, 1 - mᵅ⁻¹}, {m, 0, 1},
  PlotRange → {{0, 1}, {0, 1}}, AspectRatio → 1, ImageSize → Small]]
```


Another Approach

Of course, from the parametric formulation we can solve for the income share as a function of the population share.

```
With[{α = 1.5},
  Plot[1 - (1 - p) ^ (1 - 1/α), {p, 0, 1},
    PlotRange → {{0, 1}, {0, 1}}, AspectRatio → 1, ImageSize → Small]]
```


Lorenz Curve

We can transform the above parametric representation of the Lorenz curve to express the cumulative share of income as a function of the cumulative share of the population earning it. Recall that the CDF at income t gives the share of the population (say, $s(t)$) earning less than t. Since this CDF is strictly increasing, we can produce the inverse function $t(s)$,

```
cdfPareto /. {x → t}
ts = Solve[s == %, t] // Flatten (* inverse of CDF *)
```

$$1 - \left(\frac{x0}{t}\right)^{\alpha}$$

Solve::ifun : Inverse functions are being used by Solve, so some solutions may not be found; use Reduce for complete solution information. ≫

$$\left\{t \to (1 - s)^{-1/\alpha} x0\right\}$$

which we can then substitute into our expression of the cumulative share of income as a function of the income level to get the cumulative share of income as a function of the cumulative share of the population.