Gini Coefficient

A supplement to “Mahler’s Guide to Loss Distributions”

Exam C

prepared by
Howard C. Mahler, FCAS

Howard Mahler
hmahler@mac.com
www.howardmahler.com/Teaching

The Gini Coefficient, Gini Index, or coefficient of concentration is a concept that comes up for example in economics, when looking at the distribution of incomes. I will discuss the Gini coefficient and relate it to the Relative Mean Difference.

The Gini coefficient is a measure of inequality. For example if all of the individuals in a group have the same income, then the Gini coefficient is zero. As incomes of the individuals in a group became more and more unequal, the Gini coefficient would increase towards a value of 1. The Gini coefficient has found application in many different fields of study.

Mean Difference:

Define the mean difference as the average absolute difference between two random draws from a distribution.

Mean Difference = $\int\int |x - y|\ f(x)\ f(y)\ dx\ dy$,

where the double integral is taken over the support of f.

For example, for a uniform distribution from 0 to 10:

Mean Difference = $\int_0^{10}\int_0^{10} |x - y|\ (1/10)\ (1/10)\ dx\ dy$ =

$$(1/100)\int_0^{10}\int_{x=y}^{10} x - y\ dx\ dy + (1/100)\int_0^{10}\int_{x=0}^{y} y - x\ dx\ dy =$$

$$(1/100)\int_0^{10} 50 - y^2/2 + y^2 - 10y\ dy + (1/100)\int_0^{10} y^2 - y^2/2\ dy =$$

(1/100) (500 + 1000/6 - 500) + (1/100)(1000/6) = 10/3.

In a similar manner, in general for the continuous uniform distribution, the mean difference is: (width)/3.[1]

[1] For a sample of size two from a uniform, the expected value of the minimum is the bottom of the interval plus (width)/3, while the expected value of the maximum is the top of the interval - (width)/3. Thus the expected absolute difference is (width)/3. This is discussed in order statistics, on the Syllabus of CAS Exam CAS S.

Exercise: Compute the mean difference for an Exponential Distribution.

[Solution: Mean difference = $\int\int | x - y | \ e^{-x/\theta} / \theta \ dx \ e^{-y/\theta} / \theta \ dy$ =

$$(1/\theta^2) \int_0^\infty e^{-y/\theta} \int_{x=0}^{y} (y - x) \ e^{-x/\theta} \ dx \ dy + (1/\theta^2) \int_0^\infty e^{-y/\theta} \int_{x=y}^{\infty} (x - y) \ e^{-x/\theta} \ dx \ dy =$$

$$(1/\theta) \int_0^\infty e^{-y/\theta} \{y (1 - e^{-y/\theta}) + \theta e^{-y/\theta} + y \ e^{-y/\theta} - \theta\} \ dy +$$

$$(1/\theta) \int_0^\infty e^{-y/\theta} \{\theta e^{-y/\theta} + y \ e^{-y/\theta} - y \ e^{-y/\theta}\} \ dy =$$

$$\int_0^\infty y \ e^{-y/\theta} / \theta + 2 \ e^{-2y/\theta} - e^{-y/\theta} \ dy = \theta + \theta - \theta = \theta.$$

Alternately, by symmetry the contributions from when x > y and when y > x must be equal.

Thus, the mean difference is: $(2) \ (1/\theta^2) \int_0^\infty e^{-y/\theta} \int_{x=0}^{y} (y - x) \ e^{-x/\theta} \ dx \ dy$ =

$$(2/\theta) \int_0^\infty e^{-y/\theta} \{y (1 - e^{-y/\theta}) + \theta e^{-y/\theta} + y \ e^{-y/\theta} - \theta\} \ dy =$$

$$2 \int_0^\infty y \ e^{-y/\theta} / \theta + e^{-2y/\theta} - e^{-y/\theta} \ dy = (2)(\theta + \theta/2 - \theta) = \theta.$$

<u>Comment</u>: $\int x \ e^{-x/\theta} \ dx = -\theta \ (x + \theta) \ e^{-x/\theta}$.

For a sample of size two from an Exponential Distribution, the expected value of the minimum is $\theta/2$, while the expected value of the maximum is $3\theta/2$.
Therefore, the expected value of the difference is θ.]

<u>Mean Relative Difference</u>:

The mean relative difference of a distribution is defined as: $\frac{\text{mean difference}}{\text{mean}}$.

For the uniform distribution, the mean relative difference is: $\frac{(\text{width}) / 3}{(\text{width}) / 2} = 2/3$.

For the Exponential Distribution, the mean relative difference is: $\frac{\theta}{\theta} = 1$.

Exercise: Derive the form of the Mean Relative Difference for a Pareto Distribution.

Hint: $\int \frac{x}{(x+\theta)^a} dx = \frac{-1}{(x+\theta)^{a-1}} \frac{x (a-1) + \theta}{(a-1)(a-2)}$

[Solution: For $\alpha > 1$, $E[X] = \theta/(\alpha-1)$. $f(x) = \frac{\alpha \theta^\alpha}{(\theta + x)^{\alpha+1}}$.

Mean difference = $\int\int |x - y| \frac{\alpha\theta^\alpha}{(\theta+x)^{\alpha+1}} dx \frac{\alpha\theta^\alpha}{(\theta+y)^{\alpha+1}} dy$.

By symmetry the contributions from when $x > y$ and when $y > x$ must be equal.

Therefore, mean difference = $2 \alpha^2 \theta^{2\alpha} \int_{y=0}^{\infty} \int_{x=y}^{\infty} \frac{(x - y)}{(\theta + x)^{\alpha+1}} dx \frac{1}{(\theta + y)^{\alpha+1}} dy$.

Now using the hint: $\int_{x=y}^{\infty} \frac{x}{(\theta + x)^{\alpha+1}} dx = \frac{y\alpha + \theta}{\alpha (\alpha - 1) (\theta + y)^\alpha}$.

$$\int_{x=y}^{\infty} \frac{1}{(\theta + x)^{\alpha+1}} dx = \frac{1}{\alpha (\theta + y)^\alpha}.$$

Therefore, $\int_{x=y}^{\infty} \frac{x - y}{(\theta + x)^{\alpha+1}} dx = \frac{y\alpha + \theta}{\alpha (\alpha - 1) (\theta + y)^\alpha} - \frac{y}{\alpha (\theta + y)^\alpha} = \frac{1}{\alpha (\alpha - 1) (\theta + y)^{\alpha-1}}$.

Thus, mean difference = $\frac{2 \alpha \theta^{2\alpha}}{\alpha - 1} \int_0^{\infty} \frac{1}{(\theta + y)^{2\alpha}} dy = \frac{2 \alpha \theta^{2\alpha}}{\alpha - 1} \frac{1}{(2\alpha - 1) \theta^{2\alpha - 1}} = \frac{2 \alpha \theta}{(\alpha - 1) (2\alpha - 1)}$.

$E[X] = \theta/(\alpha-1)$. Thus, the mean relative difference is: $2 / (2\alpha - 1)$, $\alpha > 1$.]

Lorenz Curve:

Assume that the incomes in a country follow a distribution function F(x).[2]
Then F(x) is the percentage of people with incomes less than x.

The income earned by such people is: $\int_0^x t\, f(t)\, dt = E[X \wedge x] - x\, S(x) = \int_0^x S(t)\, dt$.

The percentage of total income earned by such people is:

$$\frac{\int_0^x y\, f(y)\, dy}{E[X]} = \frac{E[X \wedge x] - x\, S(x)}{E[X]}.$$

Define $G(x) = \dfrac{\int_0^x y\, f(y)\, dy}{E[X]} = \dfrac{E[X \wedge x] - x\, S(x)}{E[X]}$.[3]

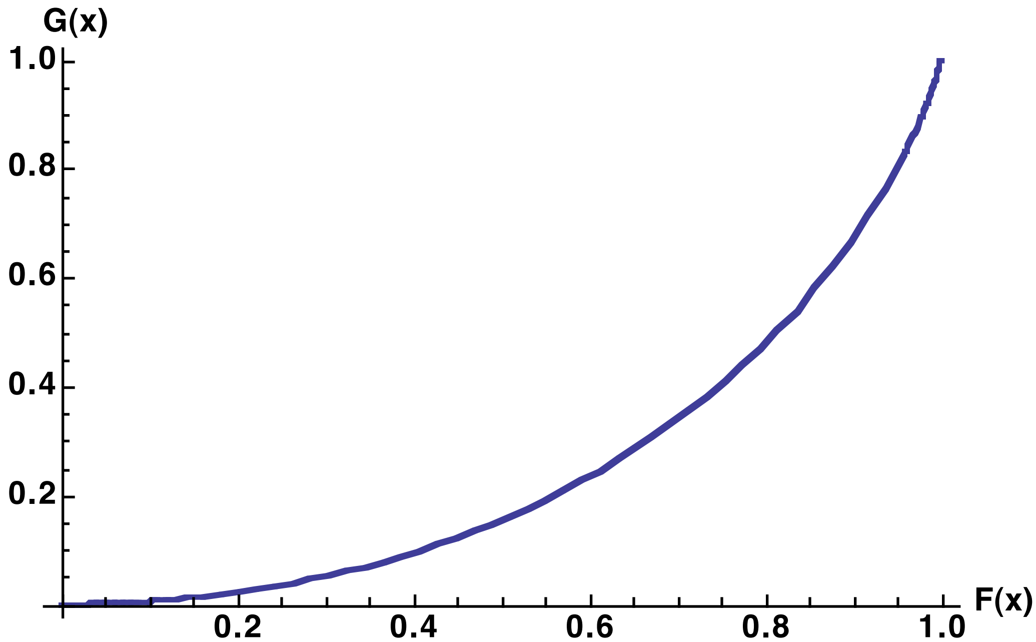
For example, assume an Exponential Distribution.
Then $F(x) = 1 - e^{-x/\theta}$.

$$G(x) = \frac{E[X \wedge x] - x\, S(x)}{E[X]} = \frac{\theta\,(1 - e^{-x/\theta}) - x\, e^{-x/\theta}}{\theta} = 1 - e^{-x/\theta} - (x/\theta)\, e^{-x/\theta}.$$

Let $t = F(x) = 1 - e^{-x/\theta}$. Therefore, $x/\theta = -\ln(1 - t)$.[4]
Then, $G(t) = t - \{-\ln(1-t)\}\,(1-t) = t + (1-t)\ln(1-t)$.

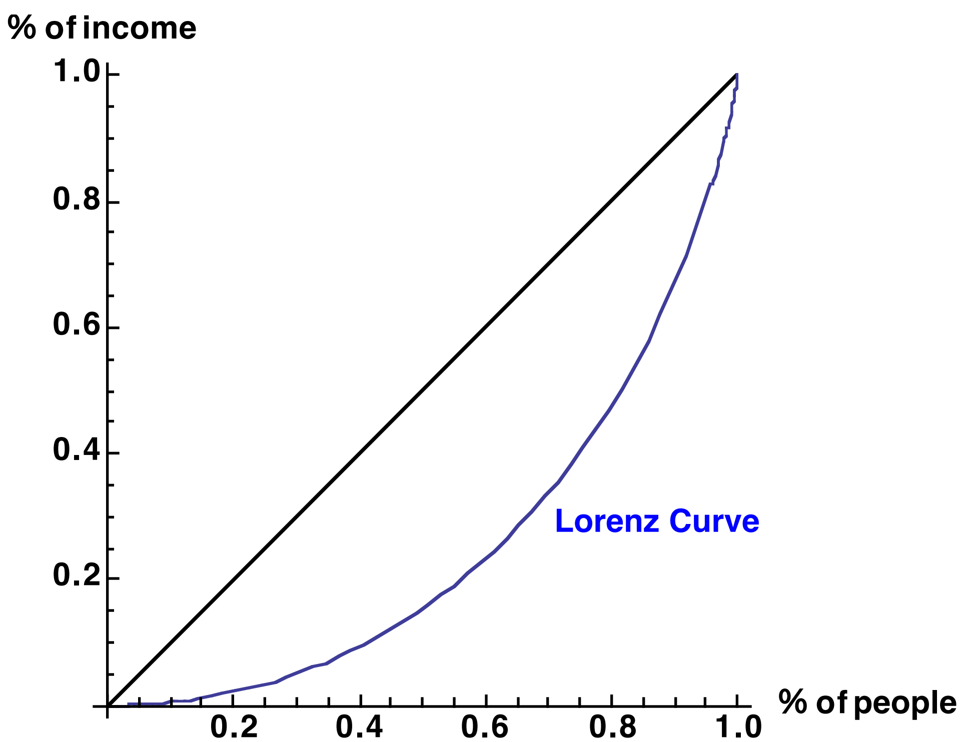
[2] Of course, the mathematics applies regardless of what is being modeled.
The distribution of incomes is just the most common context.
[3] This is not standard notation. I have just used G to have some notation.
[4] This is just the VaR formula for the Exponential Distribution.

Then we can graph G as a function of F:



This curve is referred to as the Lorenz curve or the concentration curve.

Since F(0) = 0 = G(0) and F(∞) = 1 = G(∞), the Lorenz curve passes through the points (0, 0) and (1, 1). Usually one would also include in the graph the 45° reference line connecting (0, 0) and (1, 1), as shown below:



$$G(t) = G[F(x))] = \frac{\int_0^x y\, f(y)\, dy}{E[X]}.$$

$$\frac{dG}{dt} = \frac{dG}{dx} / \frac{dF}{dx} = \frac{x\, f(x)}{E[X]} / f(x) = \frac{x}{E[X]} > 0.$$

$$\frac{d^2G}{dt^2} = \frac{1}{E[X]}\ \frac{dx}{dx} / \frac{dF}{dx} = \frac{1}{E[X]\, f(x)} > 0.$$

Thus, in the above graph, as well as in general, the Lorenz curve is increasing and concave up. The Lorenz curve is below the 45° reference line, except at the endpoints when they are equal.

The vertical distance between the Lorenz curve and the 45° comparison line is: F - G.

Thus, this vertical distance is a maximum when: $0 = \frac{dF}{dF} - \frac{dG}{dF}$.

$\Rightarrow \frac{dG}{dF} = 1. \Rightarrow \frac{x}{E[X]} = 1. \Rightarrow x = E[X].$

Thus the vertical distance between the Lorenz curve and the 45° comparison line is a maximum at the mean income.

Exercise: If incomes follow an Exponential Distribution, what is this maximum vertical distance between the Lorenz curve and the 45° comparison line?

[Solution: The maximum occurs when $x = \theta$.

$F(x) = 1 - e^{-x/\theta}$. From previously, $G(x) = 1 - e^{-x/\theta} - (x/\theta)\, e^{-x/\theta}$.

$F - G = (x/\theta)\, e^{-x/\theta}$. At $x = \theta$, this is: $e^{-1} = 0.3679$.]

Exercise: Determine the form of the Lorenz Curve, if the distribution of incomes follows a Pareto Distribution, with $\alpha > 1$.

[Solution: $F(x) = 1 - \left(\frac{\theta}{\theta + x}\right)^{\alpha}$. $E[X] = \frac{\theta}{\alpha - 1}$. $E[X \wedge x] = \frac{\theta}{\alpha - 1}\left\{1 - \left(\frac{\theta}{\theta + x}\right)^{\alpha - 1}\right\}$.

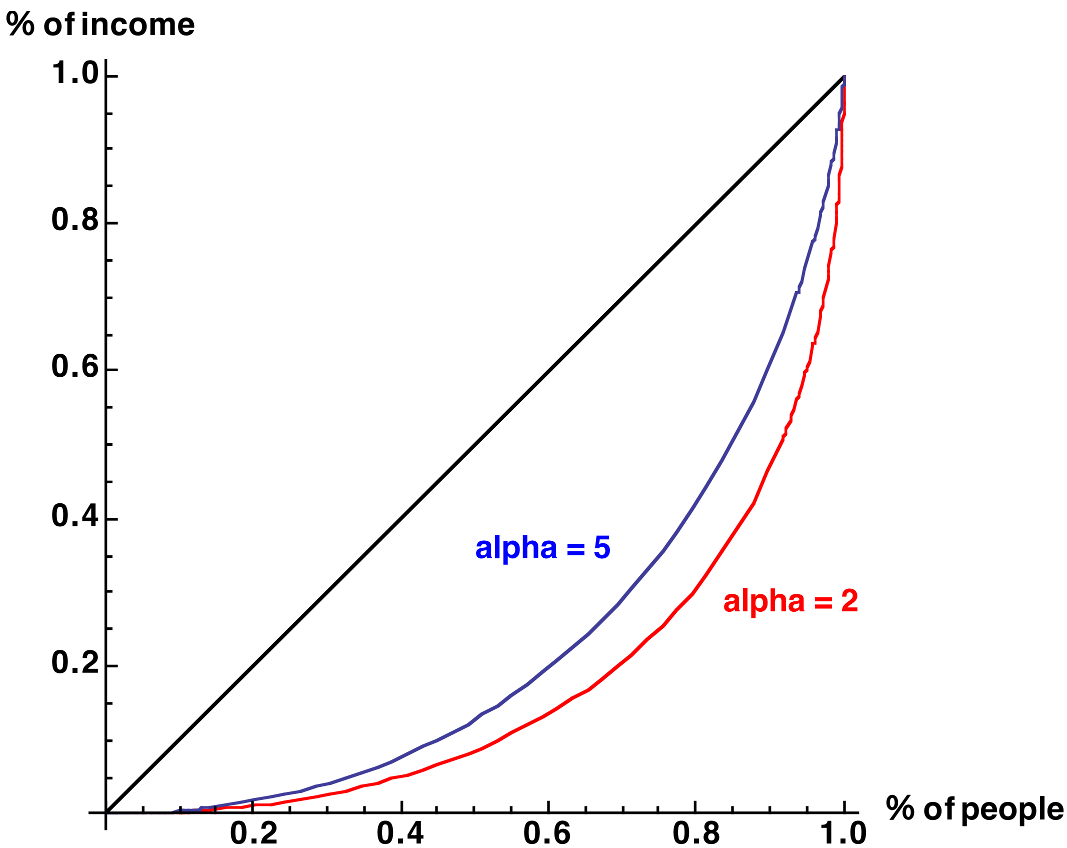
$$G(x) = \frac{E[X \wedge x] - x\,S(x)}{E[X]} = \frac{\frac{\theta}{\alpha - 1}\left\{1 - \left(\frac{\theta}{\theta + x}\right)^{\alpha - 1}\right\} - x\,S(x)}{\theta\ /\ (\alpha - 1)} = 1 - \left(\frac{\theta}{\theta + x}\right)^{\alpha - 1} - (\alpha - 1)\,\frac{x}{\theta}\,S(x).$$

Let $t = F(x) = 1 - \left(\frac{\theta}{\theta + x}\right)^{\alpha}$. $\Rightarrow \left(\frac{\theta}{\theta + x}\right)^{\alpha} = S(x) = 1 - t$. Also, $x/\theta = (1 - t)^{-1/\alpha} - 1$.[5]

Therefore, $G(t) = 1 - (1 - t)^{(\alpha-1)/\alpha} - (\alpha-1)\{(1 - t)^{-1/\alpha} - 1\}\,(1 - t) = t + \alpha - t\alpha - \alpha\,(1-t)^{1-1/\alpha}$, $0 \le t \le 1$.

<u>Comment</u>: $G(0) = \alpha - \alpha = 0$. $G(1) = 1 + \alpha - \alpha - 0 = 1$.]

Here is graph comparing the Lorenz curves for Paretos with $\alpha = 2$ and $\alpha = 5$:



[5] This is just the VaR formula for the Pareto Distribution.

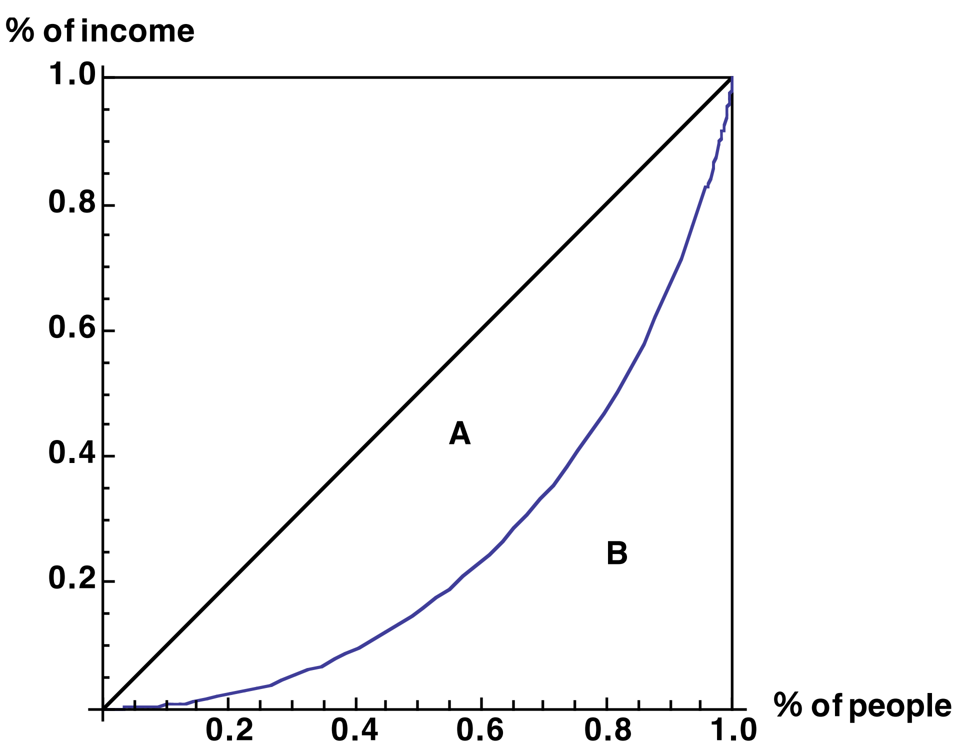
The Pareto with $\alpha = 2$ has a heavier righthand tail than the Pareto with $\alpha = 5$. If incomes follow a Pareto with $\alpha = 2$, then there are more extremely high incomes compared to the mean, than if incomes follow a Pareto with $\alpha = 5$. In other words, if $\alpha = 2$, then income is more concentrated in the high income individuals than if $\alpha = 5$.[6]

The Lorenz curve for $\alpha = 2$ is below that for $\alpha = 5$. In general, the lower curve corresponds to a higher concentration of income. In other words, a higher concentration of income corresponds to a smaller area under the Lorenz curve. Equivalently, a higher concentration of income corresponds to a larger area between the Lorenz curve and the 45° reference line.

Gini Coefficient:

This correspondence between areas on the graph of the Lorenz curve the concentration of income is the idea behind the Gini Coefficient.

Let us label the areas in the graph of a Lorenz Curve, in this case for an Exponential Distribution:



$$\text{Gini Coefficient} = \frac{\text{Area A}}{\text{Area A} + \text{Area B}}.$$

[6] An Exponential Distribution has a lighter righthand tail than either Pareto. Thus if income followed an Exponential, it would less concentrated than if it followed any Pareto.

However, Area A + Area B add up to a triangle with area 1/2.

Therefore, Gini Coefficient = $\dfrac{\text{Area A}}{\text{Area A + Area B}}$ = 2A = 1 - 2B.

For the Exponentials Distribution, the Lorenz curve was: G(t) = t + (1-t) ln(1-t).

Thus, Area B = area under Lorenz curve = $\int_0^1 t + (1-t)\ln(1-t)\,dt = 1/2 + \int_0^1 s\ln(s)\,ds$.

Applying integration by parts,

$$\int_0^1 s\ln(s)\,ds = (s^2/2)\ln(s)\Big]_{s=0}^{s=1} - \int_0^1 (s^2/2)(1/s)\,ds = 0 - 1/4 = -1/4.$$

Thus Area B = 1/2 - 1/4 = 1/4.

Therefore, for the Exponential Distribution, the Gini Coefficient is: 1 - (2)(1/4) = 1/2.

Recall that for the Exponential Distribution, the mean relative difference was 1.
As will be shown subsequently, in general, Gini Coefficient = (mean relative difference)/2.

Therefore, for the Uniform Distribution, the Gini Coefficient is: (1/2)(2/3) = 1/3.
Similarly, for the Pareto Distribution, the Gini Coefficient is: $(1/2)\{2/(2\alpha - 1)\} = 1/(2\alpha - 1)$, $\alpha > 1$.

We note that the Uniform with the lightest righthand tail of the three has the smallest Gini coefficient, while the Pareto with the heaviest righthand tail of the three has the largest Gini coefficient.
Among Pareto Distributions, the smaller alpha, the heavier the righthand tail, and the larger the Gini Coefficient.[7]
The more concentrated the income is among the higher earners, the larger the Gini coefficient.

[7] As alpha approaches one, the Gini coefficient approaches one.

<u>LogNormal Distribution</u>:

For the LogNormal Distribution: $E[X] = \exp[\mu + \sigma^2/2]$.

$$E[X \wedge x] = \exp(\mu + \sigma^2/2)\ \Phi\left[\frac{\ln(x) - \mu - \sigma^2}{\sigma}\right] + x\left\{1 - \Phi\left[\frac{\ln(x) - \mu}{\sigma}\right]\right\}$$

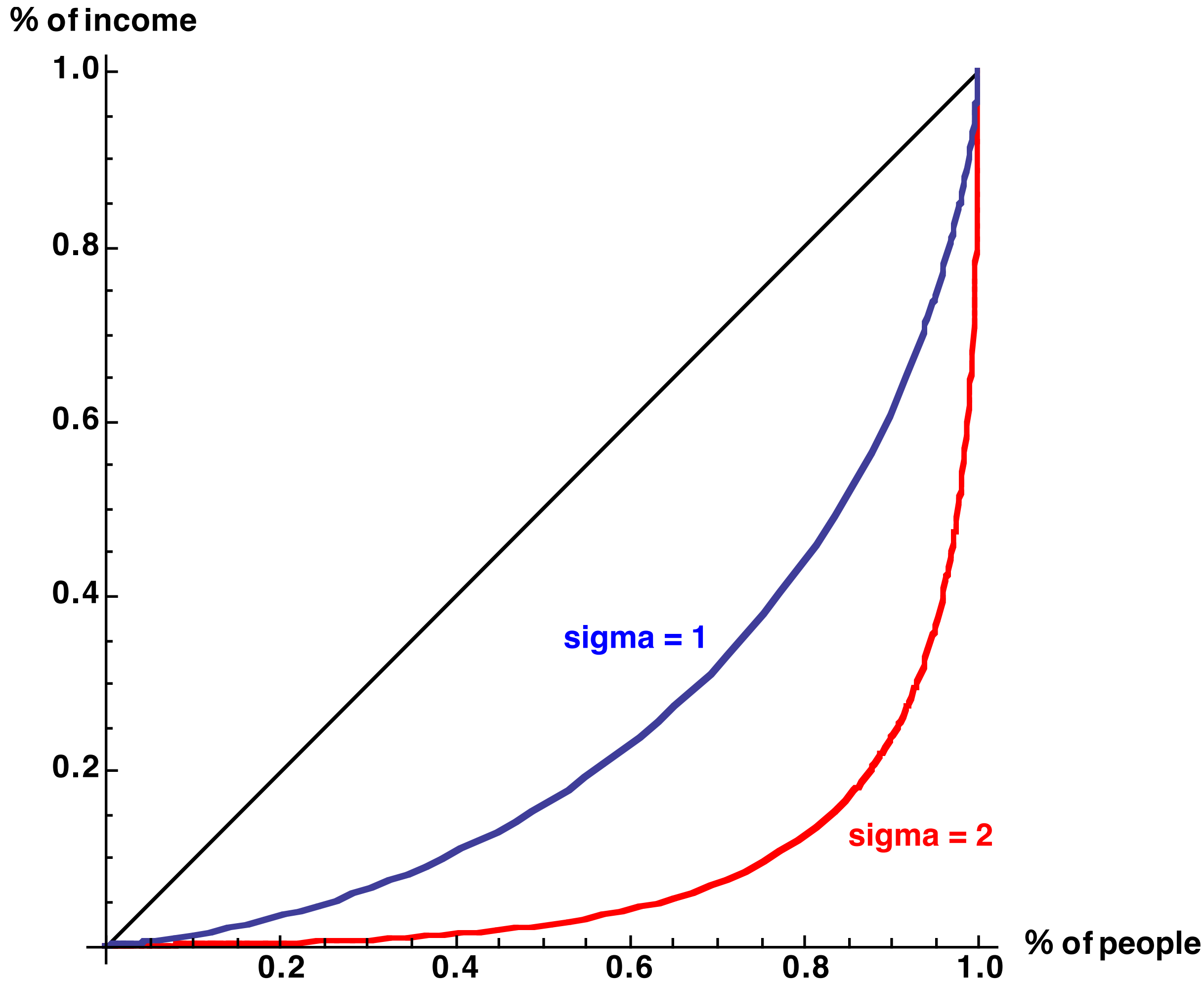
$$= E[X]\ \Phi\left[\frac{\ln(x) - \mu - \sigma^2}{\sigma}\right] + x\ S(x).$$

Therefore, $G(x) = \dfrac{E[X \wedge x] - x\ S(x)}{E[X]} = \Phi\left[\dfrac{\ln(x) - \mu - \sigma^2}{\sigma}\right] = \Phi\left[\dfrac{\ln(x) - \mu}{\sigma} - \sigma\right]$.

Let $t = F(x) = \Phi\left[\dfrac{\ln(x) - \mu}{\sigma}\right]$.

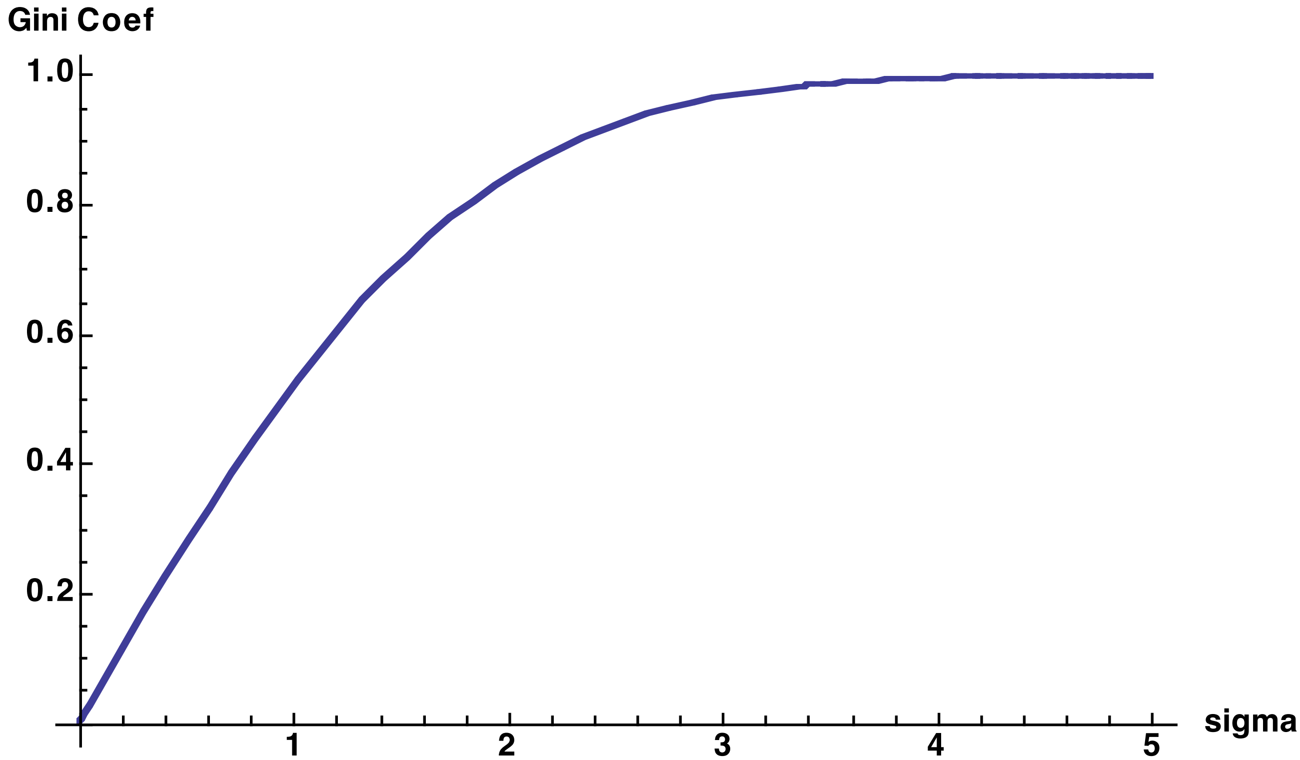
Then the Lorenz Curve is: $G(t) = \Phi[\Phi^{-1}[t] - \sigma]$.

For example, here a graph of the Lorenz curves for LogNormal Distributions with $\sigma = 1$ and $\sigma = 2$:



As derived subsequently, for a LogNormal Distribution, the Gini Coefficient is: $2\Phi[\sigma/\sqrt{2}] - 1$.

Here is a graph of the Gini Coefficient as a function of sigma:

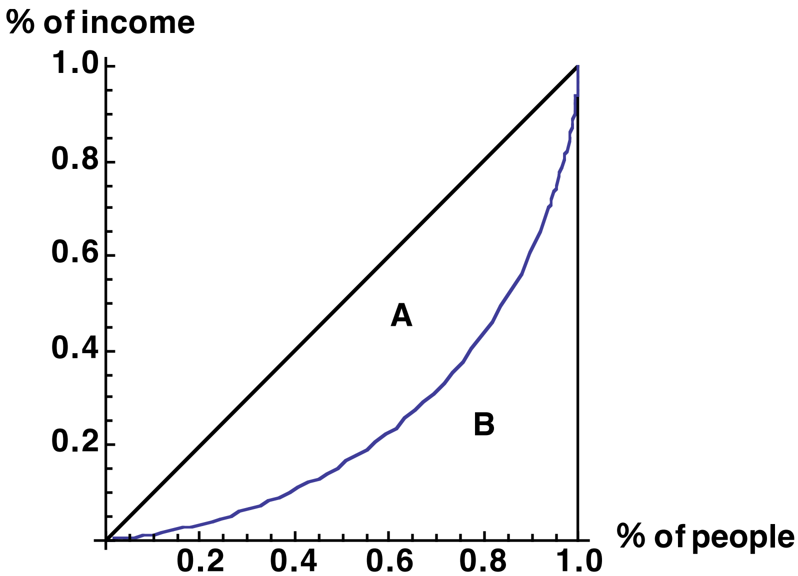


As sigma increases, the LogNormal has a heavier tail, and the Gini Coefficient Increases towards 1.

The mean relative distance is twice the Gini Coefficient: $4\Phi[\sigma/\sqrt{2}] - 2$.

Derivation of the Gini Coefficient for the LogNormal Distribution:

In order to compute the Gini Coefficient, we need to compute area B.



$$B = \int_0^1 G(t)\,dt = \int_0^1 \Phi[\Phi^{-1}[t] - \sigma]\,dt.$$

Let $y = \Phi^{-1}[t]$. Then $t = \Phi[y]$. $dt = \phi[y]\,dy$.

$$B = \int_{-\infty}^{\infty} \Phi[y - \sigma]\,\phi[y]\,dy.$$

Now B is some function of σ.

$$B(\sigma) = \int_{-\infty}^{\infty} \Phi[y - \sigma]\,\phi[y]\,dy.$$

$$B(0) = \int_{-\infty}^{\infty} \Phi[y]\,\phi[y]\,dy = \Phi[y]^2 / 2 \Big]_{y=-\infty}^{y=\infty} = 1/2.$$

$B(\sigma) = \int_{-\infty}^{\infty} \Phi[y - \sigma]\, \phi[y]\, dy$. Taking the derivative of B with respect to sigma:

$$B'(\sigma) = -\int_{-\infty}^{\infty} \phi[y - \sigma]\, \phi[y]\, dy = -\frac{1}{2\pi} \int_{-\infty}^{\infty} \exp[-(y - \sigma)^2 / 2]\, \exp[-y^2 / 2]\, dy$$

$$= -\frac{1}{2\pi}\, \exp[-\sigma^2/2] \int_{-\infty}^{\infty} \exp[-(2y^2 - 2\sigma y) / 2]\, dy$$

$$= -\frac{1}{2\pi}\, \exp[-\sigma^2/4] \int_{-\infty}^{\infty} \exp[-\{(\sqrt{2}\, y)^2 - 2(\sqrt{2}\, y)(\sigma / \sqrt{2}) + (\sigma / \sqrt{2})^2\} / 2]\, dy$$

$$= -\frac{1}{2\pi}\, \exp[-\sigma^2/4] \int_{-\infty}^{\infty} \exp[-\{\sqrt{2}\, y - \sigma / \sqrt{2}\}^2 / 2]\, dy.$$

Let $x = \sqrt{2}\, y - \sigma / \sqrt{2}$. $\Rightarrow$ $dy = dx / \sqrt{2}$.

$$B'(\sigma) = -\frac{1}{2\sqrt{\pi}}\, \exp[-\sigma^2/4] \int_{-\infty}^{\infty} \frac{\exp[-x^2 / 2]}{\sqrt{2\pi}}\, dx = -\frac{1}{2\sqrt{\pi}}\, \exp[-\sigma^2/4].$$ [8]

Now assume that $B(\sigma) = c - \Phi[\sigma/\sqrt{2}]$, for some constant c.

Then $B'(\sigma) = -\phi[\sigma/\sqrt{2}] / \sqrt{2} = -\frac{1}{\sqrt{2\pi}}\, \exp[-(\sigma/\sqrt{2})^2/2] / \sqrt{2} = -\frac{1}{2\sqrt{\pi}}\, \exp[-\sigma^2/4]$, matching above.

Therefore, we have shown that $B(\sigma) = c - \Phi[\sigma/\sqrt{2}]$.

However, $B(0) = 1/2$. $\Rightarrow$ $1/2 = c - 1/2$. $\Rightarrow$ $c = 1$. $\Rightarrow$ $B(\sigma) = 1 - \Phi[\sigma/\sqrt{2}]$. [9]

Thus the Gini Coefficient is: $1 - 2B = 2\Phi[\sigma/\sqrt{2}] - 1$.

---

[8] Where I have used the fact that the density of the Standard Normal integrates to one over its support from $-\infty$ to ∞.

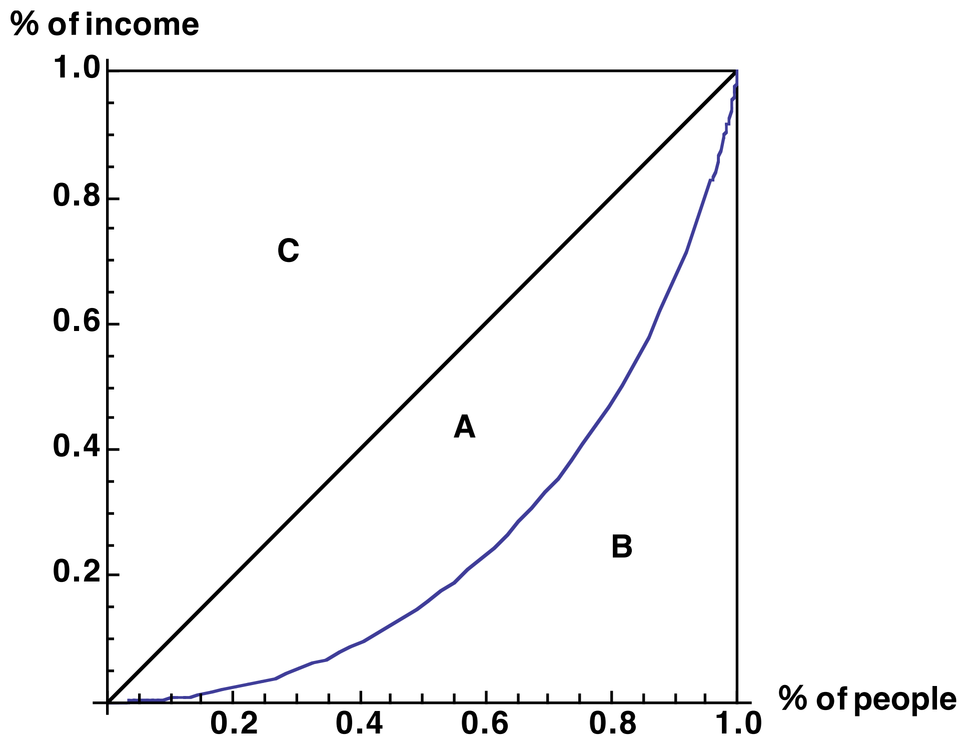
[9] In general, $\int_{-\infty}^{\infty} \Phi[a + by]\, \phi[y]\, dy = \Phi[a / \sqrt{1 + b^2}]$.

For a list of similar integrals, see http://en.wikipedia.org/wiki/List_of_integrals_of_Gaussian_functions

Proof of the Relationship Between the Gini Index and the Mean Relative Difference:[10]

I will prove that: Gini Coefficient = (mean relative difference) / 2.

As a first step, let us look at a graph of the Lorenz Curve with areas labeled:



A + B = 1/2 = C.

B is the area on the Lorenz curve: $\int G \, dF$.

Area B is the area between the Lorenz curve and the horizontal axis.

We can instead look at: C + A = area between the Lorenz curve and the vertical axis = $\int F \, dG$.

Therefore, we have that: $\int F \, dG - \int G \, dF$ = C + A - B = 1/2 + A - (1/2 - A) = 2 A.

$\Rightarrow$ Area A = (1/2) $\{\int F \, dG - \int G \, dF\}$.

$\Rightarrow$ Gini Coefficient = $\dfrac{\text{Area A}}{\text{Area A + Area B}}$ = 2A = $\int F \, dG - \int G \, dF$.

[10] Based on Section 2.25 of Volume I of Kendall's Advanced Theory of Statistics, not on the syllabus.

Recall that $G(x) = \frac{\int_0^x y\ f(y)\ dy}{E[X]}$. $\Rightarrow$ $dG = \frac{x\ f(x)\ dx}{E[X]}$.

Therefore, Gini Coefficient = $\int F\ dG - \int G\ dF = \frac{1}{E[X]} \int_0^\infty F(s)\ s\ f(s)\ ds - \int_0^\infty G(s)\ f(s)\ ds =$

$\frac{1}{E[X]} \int_0^\infty s \int_0^s f(t)\ dt\ f(s)\ ds - \frac{1}{E[X]} \int_0^\infty \int_0^s t\ f(t)\ dt\ f(s)\ ds = \frac{1}{E[X]} \int_0^\infty \int_0^s (s - t)\ f(t)\ dt\ f(s)\ ds$.

$\int_0^\infty \int_0^s (s - t)\ f(t)\ dt\ f(s)\ ds$ is the contribution to the mean distance from when $s > t$.

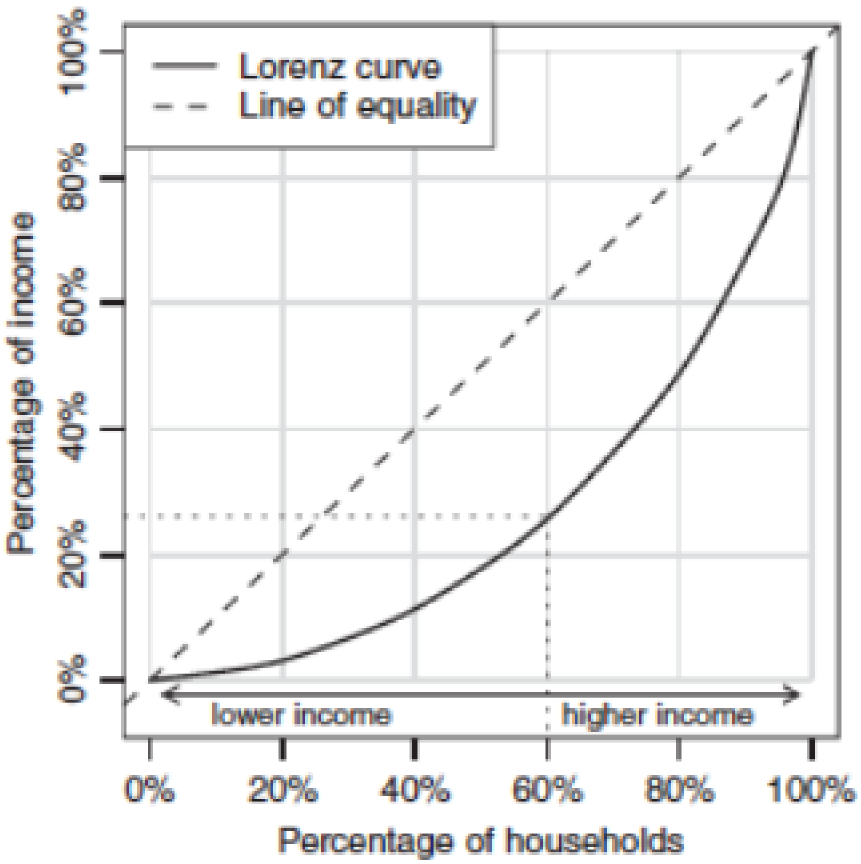
By symmetry it is equal to the contribution to the mean distance from when $t > s$.

Therefore, $2 \int_0^\infty \int_0^s (s - t)\ f(t)\ dt\ f(s)\ ds$ = mean distance.

$\Rightarrow$ Gini Coefficient = $\frac{(\text{mean difference})\ /\ 2}{E[X]}$ = (mean relative difference) / 2.

<u>An Income Example</u>:

Here is a Lorenz Curve for United States 2014 Household Income:[11]



The Gini index is calculated as twice the area between the Lorenz curve and the line of equality. In this case, the Gini index is 48.0%.

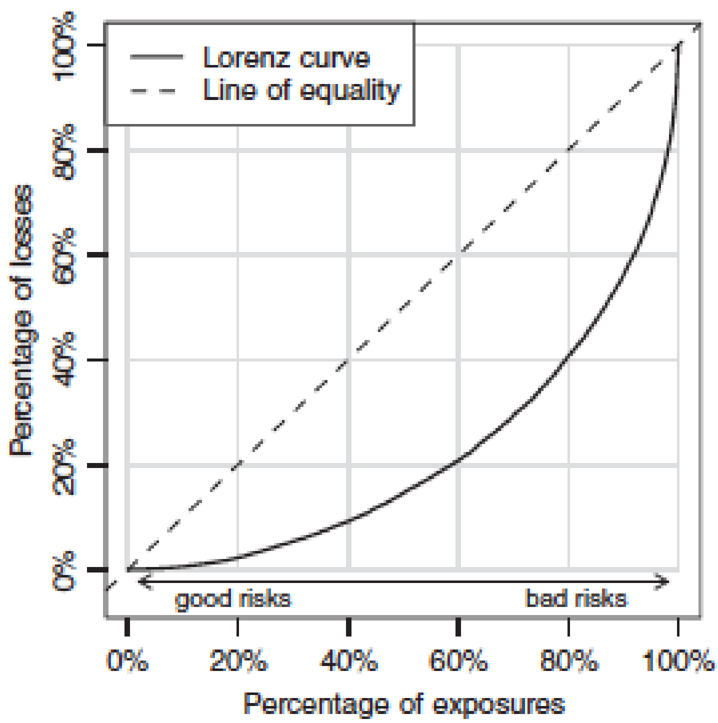
[11] See Figure 21 of of "Generalized Linear Models for Insurance Rating", by Mark Goldburd, Anand Khare, and Dan Tevet, CAS monograph series number 5.

<u>Gini Coefficient and Rating Plans</u>:[12]

The Gini Coefficient can also be used to measure the lift of an insurance rating plan by quantifying its ability to segment the population into the best and worst risks. Assume we have a rating plan. Ideally we would want the model to identify those insureds with higher expected pure premiums.

The Lorenz curve for the rating plan is determined as follows:
1. Sort the dataset based on the model predicted loss cost.[13]
2. On the x-axis, plot the cumulative percentage of exposures.
3. On the y-axis, plot the cumulative percentage of losses.
Draw a 45-degree line connecting (0, 0) and (1, 1), called the line of equality.

Here is an example:[14]



This model identified 60% of exposures which contribute only 20% of the total losses. The Gini Coefficient is twice the area between the Lorenz curve and the line of equality, in this case 56.1%. The higher the Gini Coefficient, the better the model is at identifying risk differences.[15]

[12] See Section 7.2.4 of "Generalized Linear Models for Insurance Rating", by Mark Goldburd, Anand Khare, and Dan Tevet, CAS monograph series number 5.
[13] This should be done on a dataset <u>not</u> used to develop the rating plan.
[14] See Figure 21 of Goldburd, Khare, and Tevet.
[15] A Gini Coefficient does not quantify the profitability of a particular rating plan, but it does quantify the ability of the rating plan to differentiate the best and worst risks. Assuming that an insurer has pricing and/or underwriting flexibility, this will lead to increased profitability.

<u>Problems</u>:

1, (15 points) The distribution of incomes follows a Single Parameter Pareto Distribution, $\alpha > 1$.

a. (3 points) Determine the mean relative distance.

b. (3 points) Determine the form of the Lorenz curve.

c. (3 points) With the aid of a computer, draw and compare the Lorenz curves for $\alpha = 1.5$ and $\alpha = 3$.

d. (3 points) Use the form of the Lorenz curve to compute the Gini coefficient.

e. (3 point) If the Gini coefficient is 0.47, what percent of total income is earned by the top 1% of earners?

2. (5 points) For a Gamma Distribution with $\alpha = 2$, determine the mean relative distance.

Hint: Calculate the contribution to the mean difference from when $x < y$.

$$\int x\, e^{-x/\theta}\, dx = -x\, e^{-x/\theta}\, \theta - e^{-x/\theta}\, \theta^2.$$

$$\int x^2\, e^{-x/\theta}\, dx = -x^2\, e^{-x/\theta}\, \theta - 2x\, e^{-x/\theta}\, \theta^2 - 2e^{-x/\theta}\, \theta^3.$$

<u>Solutions to Problems</u>:

1. a. $f(x) = \frac{\alpha \theta^\alpha}{x^{\alpha+1}}$, $x > \theta$.

The contribution to the relative difference from when $x > y$ is:

$$\int_\theta^\infty \int_0^y (x - y) \frac{\alpha \theta^\alpha}{x^{\alpha+1}} dx \frac{\alpha \theta^\alpha}{y^{\alpha+1}} dy = \alpha^2 \theta^{2\alpha} \int_\theta^\infty \left(\frac{1}{\alpha - 1}\frac{1}{y^{\alpha-1}} + \frac{-1}{\alpha}\frac{y}{y^\alpha}\right) \frac{1}{y^{\alpha+1}} dy =$$

$$\alpha^2 \theta^{2\alpha} \left(\frac{1}{\alpha - 1} - \frac{1}{\alpha}\right) \int_\theta^\infty \frac{1}{y^{2\alpha}} dy = \theta^{2\alpha} \frac{\alpha}{\alpha - 1} \frac{1}{(2\alpha - 1)\, \theta^{2\alpha - 1}} = \frac{\alpha}{(\alpha - 1)(2\alpha - 1)} \theta.$$

By symmetry this is equal to the contribution to the mean distance from when $y > x$.

Therefore, the mean distance is: $2 \frac{\alpha}{(\alpha - 1)(2\alpha - 1)} \theta$.

$E[X] = \frac{\alpha \theta}{\alpha - 1}$, $\alpha > 1$.

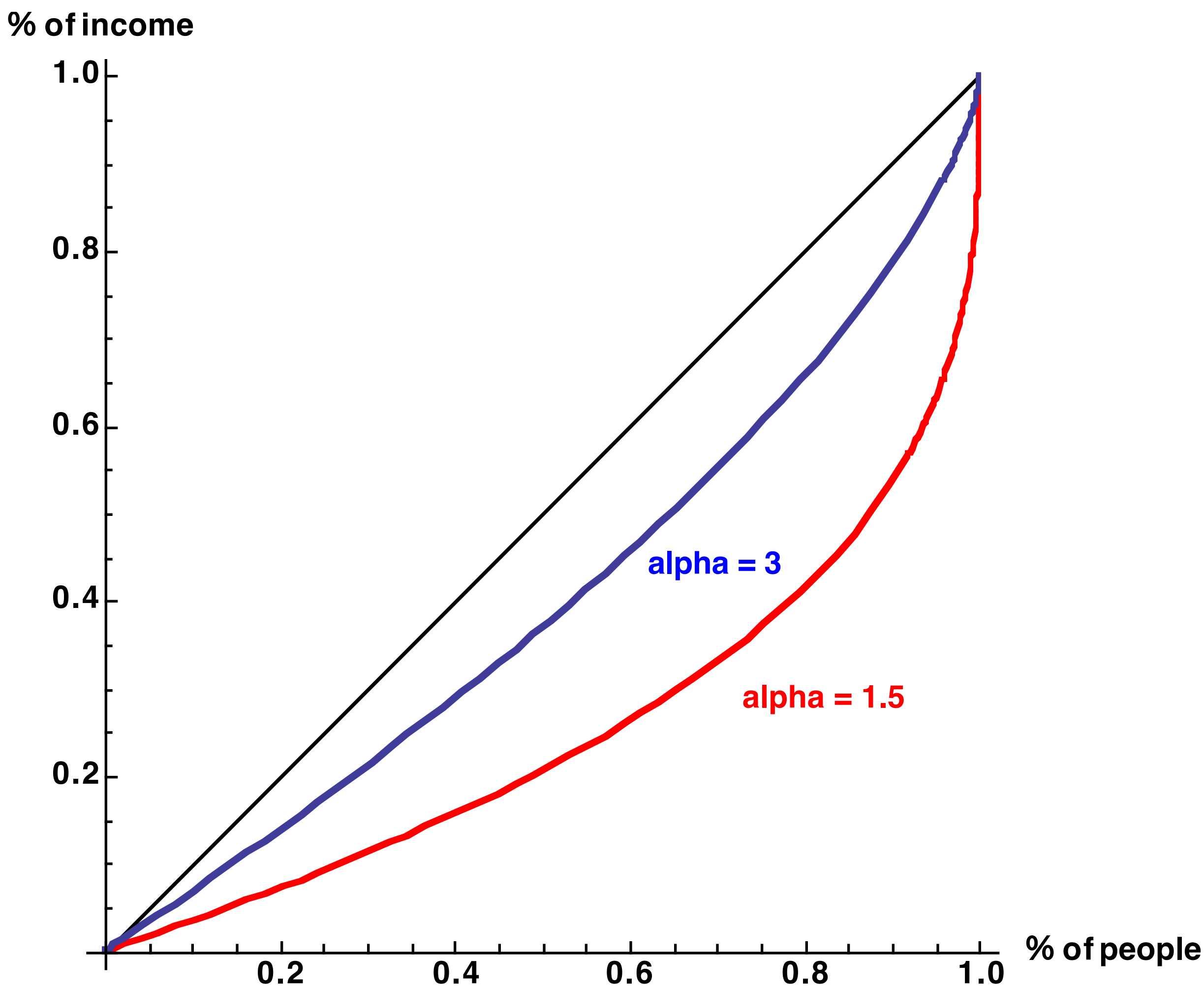
Therefore, the mean relative difference is: $\frac{2}{2\alpha - 1}$, $\alpha > 1$.

b. $$G(x) = \frac{\int_\theta^x y\, f(y)\, dy}{E[X]} = \frac{\int_\theta^x y \frac{\alpha \theta^\alpha}{y^{\alpha+1}} dy}{\frac{\alpha \theta}{\alpha - 1}} = (\alpha-1)\, \theta^{\alpha-1} \int_\theta^x \frac{1}{y^\alpha} dy = 1 - \frac{\theta^{\alpha - 1}}{x^{\alpha - 1}},\ x > \theta.$$

Now let $t = F(x) = 1 - \frac{\theta^\alpha}{x^\alpha}$, $x > \theta$. $\Rightarrow \theta/x = (1-t)^{1/\alpha}$.

Then $G(t) = 1 - (1-t)^{1-1/\alpha}$, $0 \le t \le 1$.

c. For $\alpha = 1.5$, $G(t) = 1 - (1-t)^{1/3}$, $0 \le t \le 1$. For $\alpha = 3$, $G(t) = 1 - (1-t)^{2/3}$, $0 \le t \le 1$.
Here is a graph of these two Lorenz curves:



The Lorenz curve for $\alpha = 1.5$ is below that for $\alpha = 3$.
The incomes are more concentrated for $\alpha = 1.5$ than for $\alpha = 3$.

d. The Lorenz curve is: $G(t) = 1 - (1-t)^{1-1/\alpha}$, $0 \le t \le 1$.
Integrating, the area under the Lorenz curve is: $B = 1 - 1/(2 - 1/\alpha) = 1 - \alpha/(2\alpha-1) = (\alpha-1)/(2\alpha-1)$.

Gini coefficient is: $1 - 2B = 1 - 2(\alpha-1)/(2\alpha-1) = \frac{1}{2\alpha - 1}$, $\alpha > 1$.

Note that the Gini Coefficient = (mean relative difference) / 2 = $(1/2)\ \frac{2}{2\alpha - 1} = \frac{1}{2\alpha - 1}$.

e. $0.47 = \frac{1}{2\alpha - 1}$. $\Rightarrow \alpha = 1.564$. $E[X] = \theta\ \alpha / (\alpha-1) = 2.773\ \theta$.

The 99th percentile is: $\theta\ (1 - 0.99)^{-1/1.564} = 19.00\ \theta$.

The income earned by the top 1% is:

$$\int_{19\theta}^{\infty} x\ \frac{1.564\ \theta^{1.564}}{x^{2.564}}\ dx = (1.564/0.564)\ \theta^{1.564} / (19\theta)^{0.564} = 0.527\ \theta.$$

Thus the percentage of total income earned by the top 1% is: $0.527\ \theta / (2.773\theta)$ = **19.0%.**

Comment: The mean relative distance and the Gini coefficient have the same form as for the two-parameter Pareto Distribution.

The distribution of incomes in the United States has a Gini coefficient of about 0.47.

For a sample of size two from a Single Parameter Pareto Distribution with $\alpha > 1$, it turns out that:

$E[\text{Min}] = \frac{2\alpha\theta}{2\alpha - 1}$. $E[\text{Max}] = \frac{2\alpha^2\theta}{(\alpha - 1)(2\alpha - 1)}$.

Therefore, the mean difference is: $\frac{2\alpha^2\theta}{(\alpha - 1)(2\alpha - 1)} - \frac{2\alpha\theta}{2\alpha - 1} = 2\theta\ \frac{\alpha}{(\alpha - 1)\ (2\alpha - 1)}$.

Since, $E[X] = \frac{\alpha\ \theta}{\alpha - 1}$, the mean relative distance is $\frac{2}{2\alpha - 1}$.

2. $f(x) = \frac{x^{\alpha-1}\ e^{-x/\theta}}{\theta^{\alpha}\,\Gamma(\alpha)} = x\ e^{-x/\theta} / \theta^2.$

The contribution to the relative difference from when x < y is:

$$(1/\theta^4) \int_0^{\infty}\int_y^{\infty} (y - x)\ x\ e^{-x/\theta}\ dx\ y\ e^{-y/\theta}\ dy = (1/\theta^4) \int_0^{\infty} \left(y \int_y^{\infty} x\ e^{-x/\theta}\ dx\ -\ \int_y^{\infty} x^2\ e^{-x/\theta}\ dx \right) y\ e^{-y/\theta}\ dy$$

$$= (1/\theta^4) \int_0^{\infty} \{y(-ye^{-y/\theta}\ \theta\ -\ e^{-y/\theta}\ \theta^2)\ +\ y^2 e^{-y/\theta}\ \theta\ +\ 2ye^{-y/\theta}\ \theta^2\ +\ 2e^{-y/\theta}\ \theta^3\}\ y\ e^{-y/\theta}\ dy =$$

$$(1/\theta^4) \int_0^{\infty} y^2\ e^{-2y/\theta}\ \theta^2\ +\ 2y\ e^{-2y/\theta}\ \theta^3\ dy\ = (1/\theta^4)\ \{\theta^2\ 2(\theta/2)^3 + 2\theta^3\ (\theta/2)^2\} = \theta\ 3/4.$$

By symmetry this is equal to the contribution to the mean distance from when x > y.
Therefore, the mean distance is: θ 3/2. $E[X] = \alpha\theta = 2\ \theta$.

Therefore, the mean relative difference is: $\frac{\theta\ 3/2}{2\theta}$ = **3/4**.

<u>Comment</u>: The Gini Coefficient is half the mean relative difference or 3/8.
One can show in general that for the Gamma the mean relative distance is 2 - 4 $\beta(\alpha+1, \alpha; 1/2)$.

Then in turn it can be shown that for alpha integer, the mean relative distance is: $\binom{2\alpha}{\alpha} / 2^{2\alpha-1}$.

For example, for α = 4, the mean relative difference is: $\binom{8}{4} / 2^7 = 70/128 = 35/64$.

The Gini Coefficient is half the mean relative difference, and is graphed below as a function of alpha: