

Arthur Charpentier

charpentier.arthur@gmail.com

[http ://freakonometrics.hypotheses.org/](http://freakonometrics.hypotheses.org/)

Université de Rennes 1, January 2015

Welfare, Inequality & Poverty, # 3

Inequality Comparisons (2-person Economy)

not much to say... any measure of **dispersion** is appropriate

- income gap $x_2 - x_1$
- proportional gap $\frac{x_2}{x_1}$
- any functional of the distance

$$\sqrt{|x_2 - x_1|}$$

graphs are from Amiel & Cowell (1999, ebooks.cambridge.org)

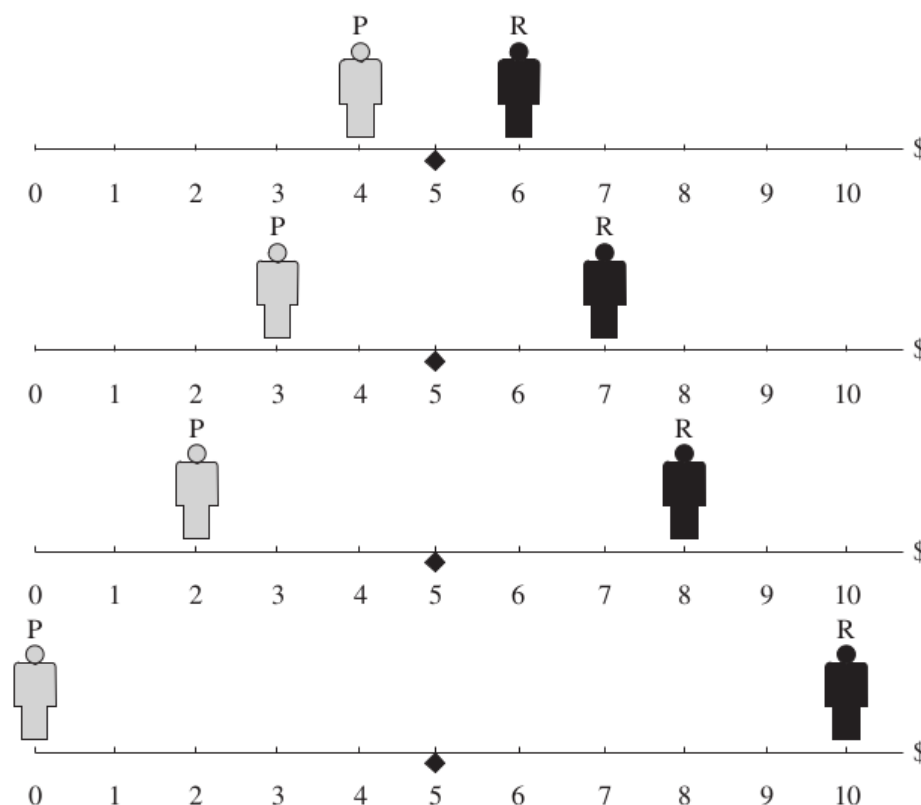
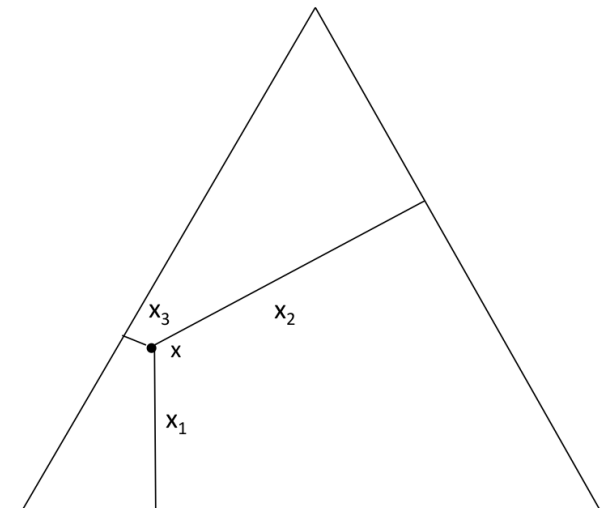
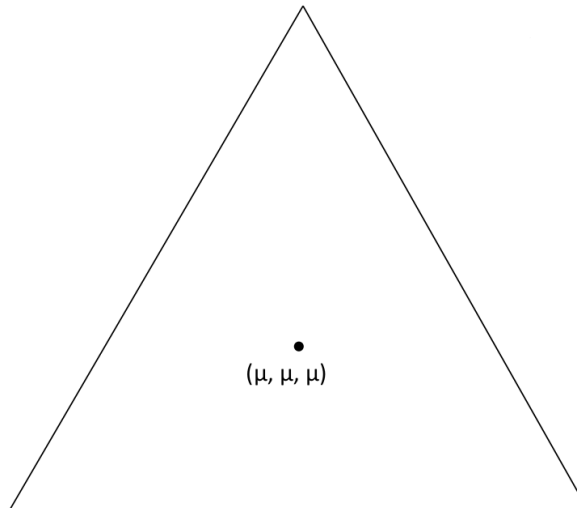
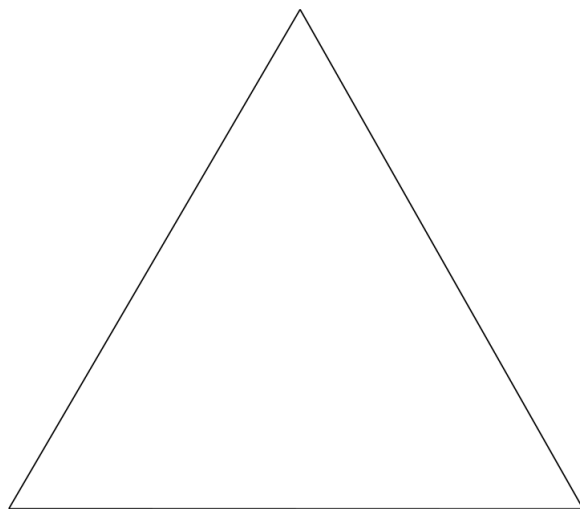


Figure 2.2. Inequality in a two-person world.

Inequality Comparisons (3-person Economy)

Consider any 3-person economy, with incomes $\mathbf{x} = \{x_1, x_2, x_3\}$. This point can be visualized in [Kolm triangle](#).



Inequality Comparisons (3-person Economy)

```
1 kolm=function (p=c(200,300,500)) {  
2   p1=p/sum(p)  
3   y0=p1[2]  
4   x0=(2*p1[1]+y0)/sqrt(3)  
5   plot(0:1,0:1,col="white",xlab="",ylab="",  
6         axes=FALSE,ylim=c(0,1))  
7   polygon(c(0,.5,1,0),c(0,.5*sqrt(3),0,0))  
8   points(x0,y0,pch=19,col="red")}
```

Inequality Comparisons (n -person Economy)

In a n -person economy, comparison are clearly more difficult

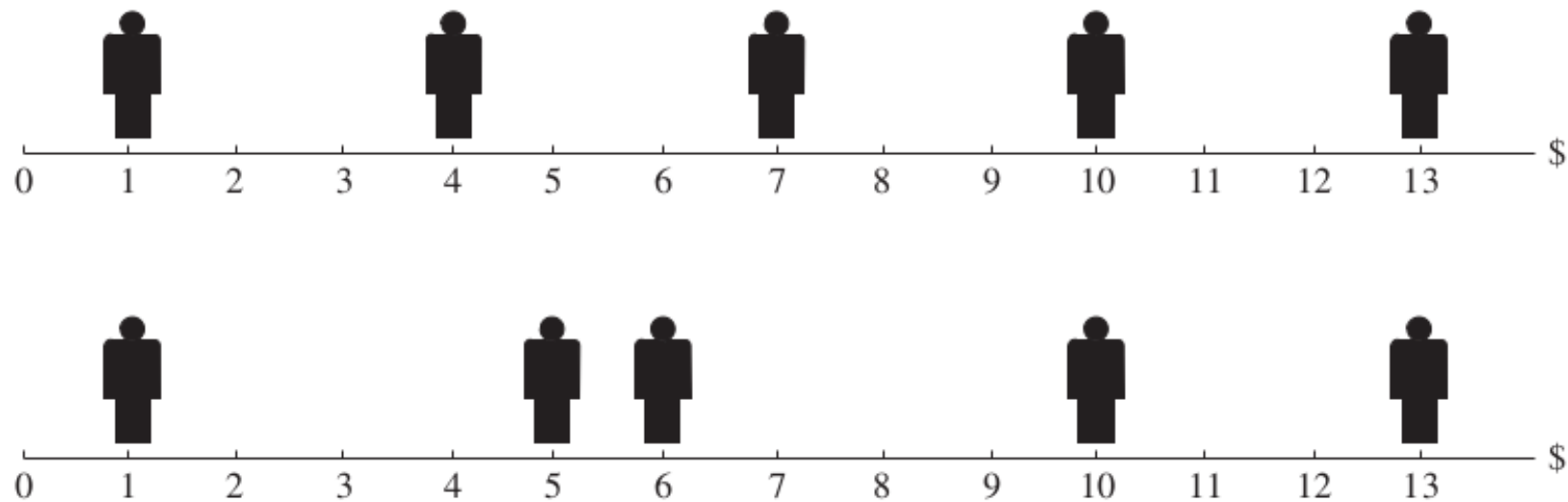


Figure 1.1. A simple distributional experiment.

Inequality Comparisons (n -person Economy)

Why not look at inequality per subgroups,

If we focus at the top of the distribution
(same holds for the bottom),

→ rising inequality

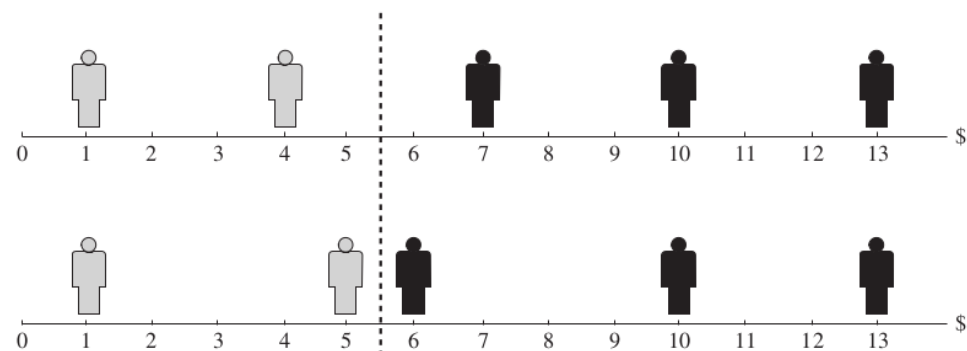


Figure 1.2. A simple distributional experiment: second view.

If we focus at the middle of the distribution,

→ falling inequality

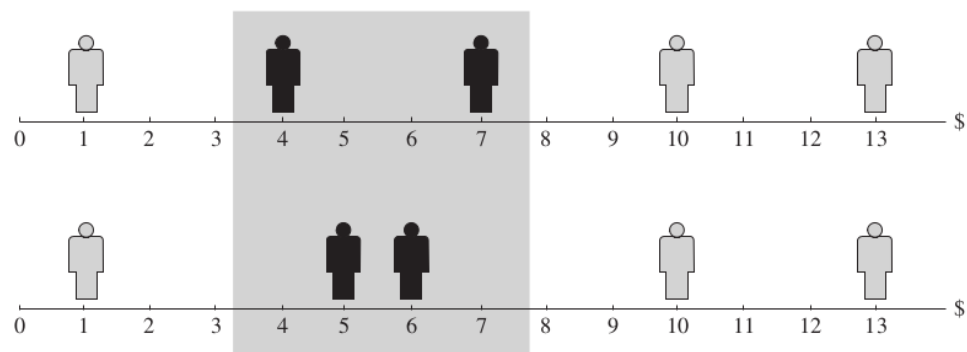


Figure 1.3. A simple distributional experiment: third view.

Inequality Comparisons (n -person Economy)

To measure inequality, we usually

- define ‘equality’ based on some reference point / distribution
- define a distance to the reference point / distribution
- aggregate individual distances

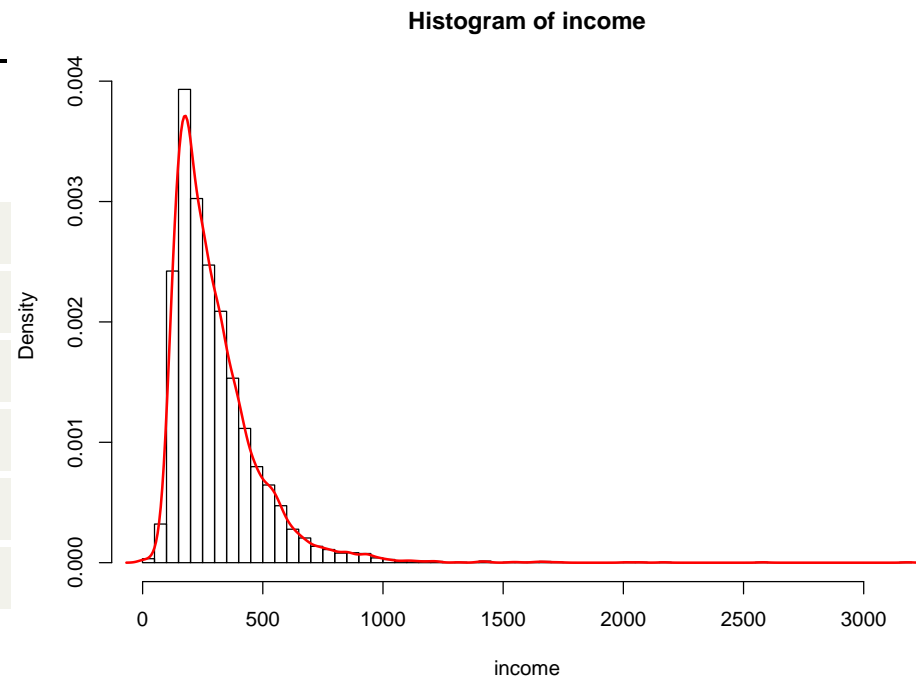
We want to visualize the distribution of incomes

```
1 > income <- read.csv("http://www.vcharite.univ-mrs.fr/pp/lubrano/  
cours/fes96.csv", sep=";", header=FALSE) $V1
```

$$F(x) = \mathbb{P}(X \leq x) = \int_0^x f(t)dt$$

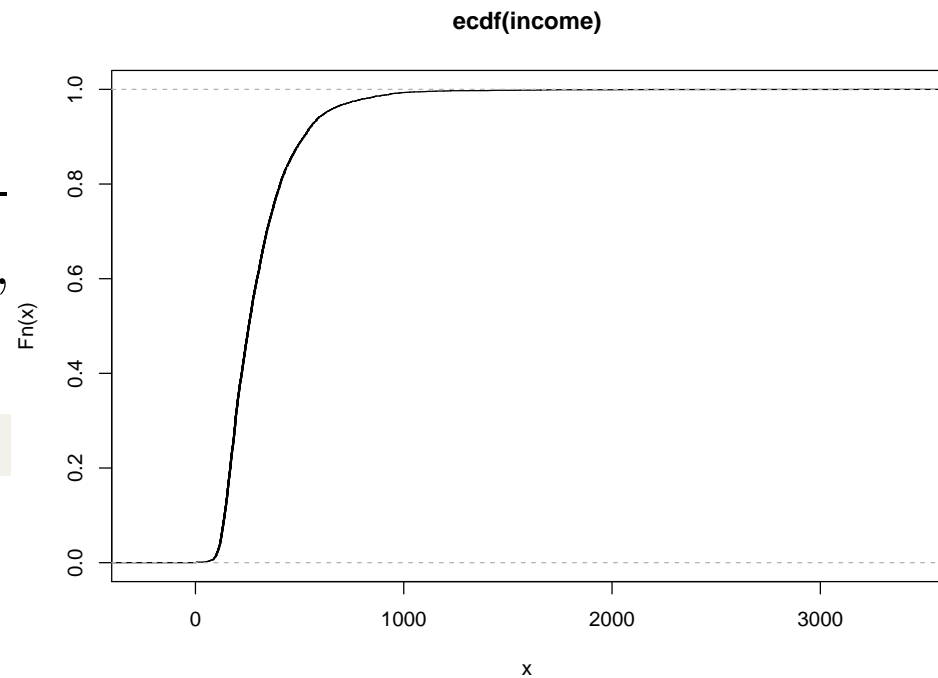
Densities are usually difficult to compare,

```
1 > hist(income ,  
2 + breaks=seq(min(income)-1,max(  
   income)+50,by=50) ,  
3 + probability=TRUE)  
4 > lines(density(income) ,col="red "  
   ,lwd=2)
```



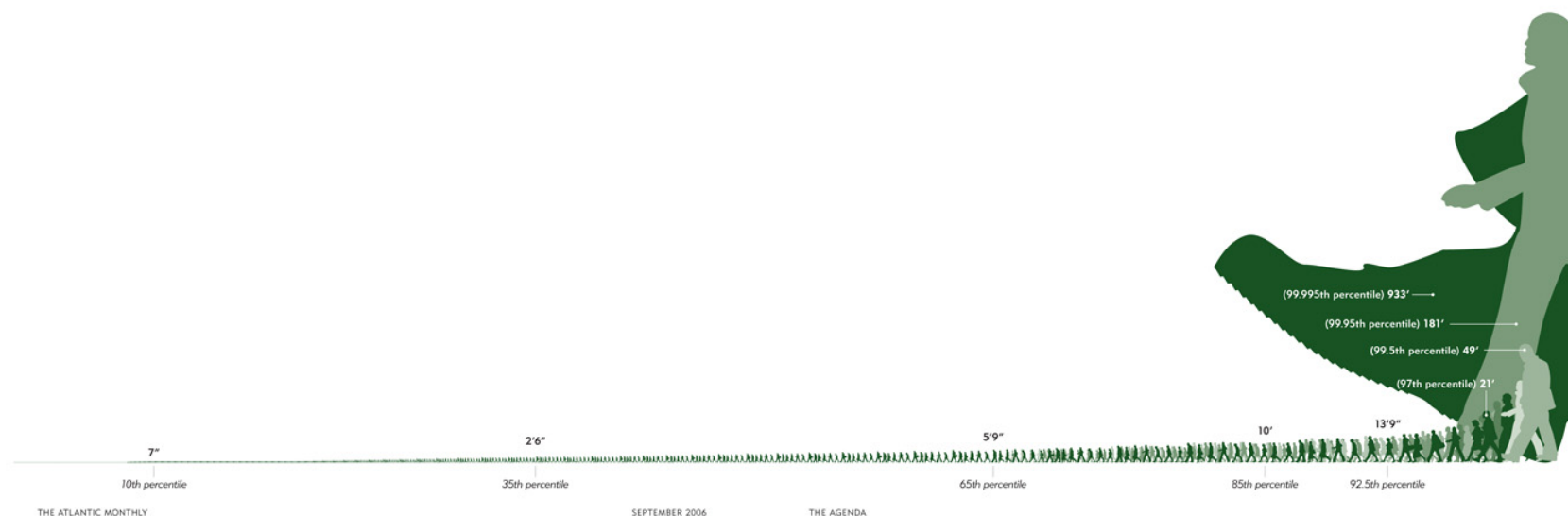
It is more convenient, compare cumulative distribution functions of income, wealth, consumption, grades, etc.

```
1 > plot(ecdf(income))
```



The Parade of Dwarfs

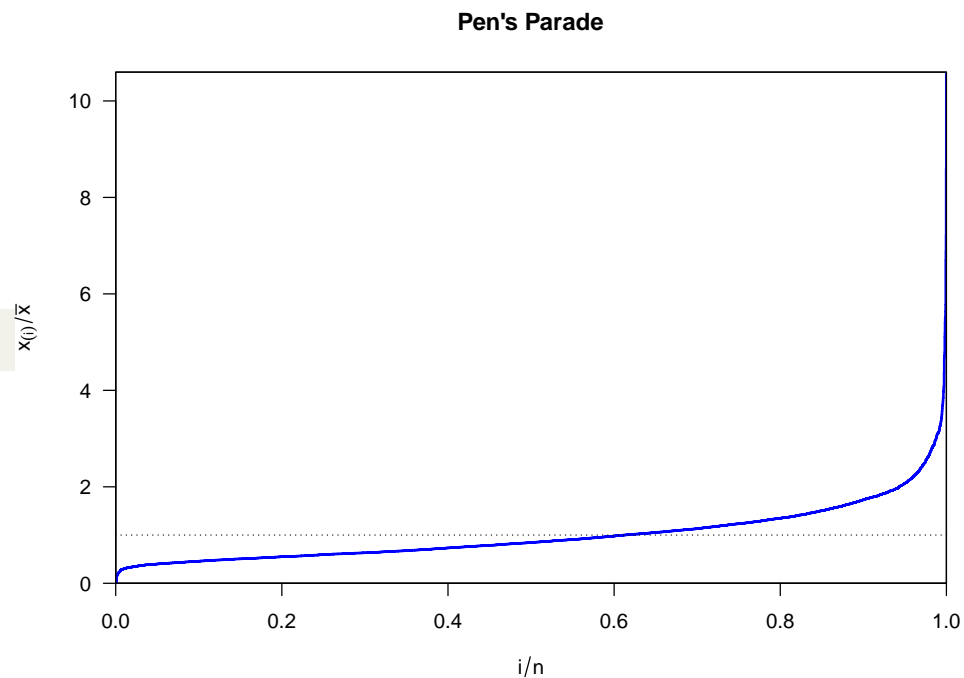
An alternative is to use Pen's parade, also called the parade of dwarfs (and a few giants), “parade van dwergen en een enkele reus”.



The height of each person is stretched in the proportion to his or her income everyone is line up in order of height, shortest (poorest) are on the left and tallest (richest) are on the right let them walk some time, like a procession.

c.d.f., quantiles and Lorenz

1 > Pen (income)



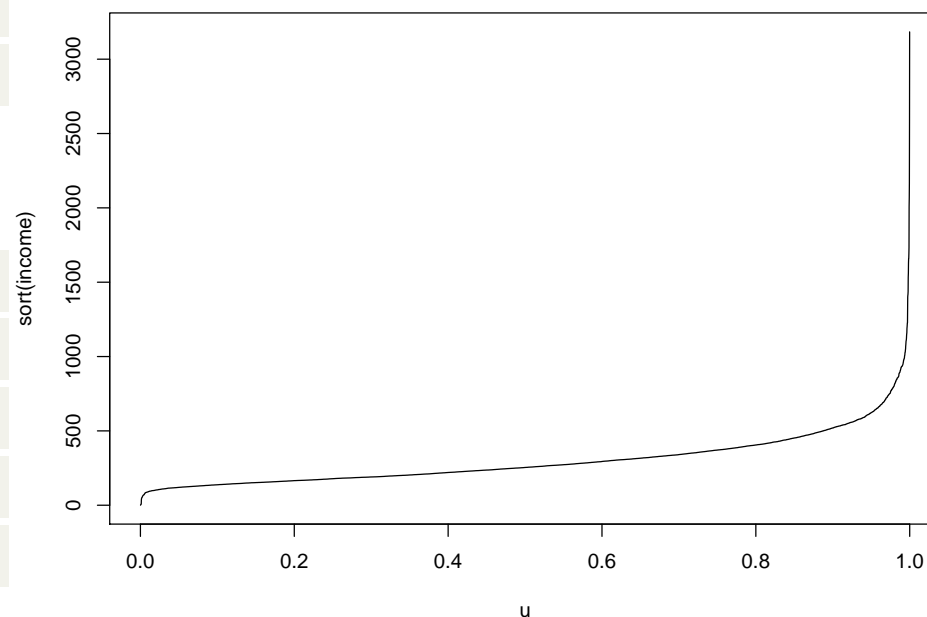
c.d.f., quantiles and Lorenz

This [parade of the Dwarfs](#) function is just the quantile function.

```
1 > q <- function(u) quantile(
  income, u)
```

see also

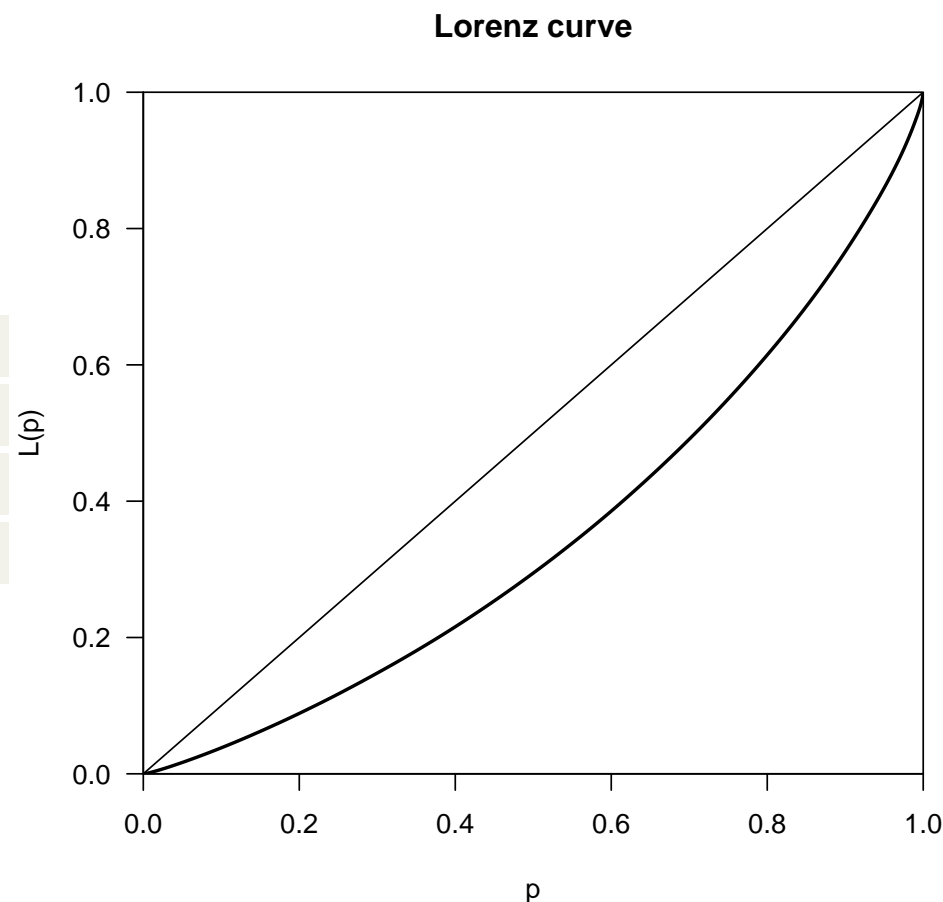
```
1 > n <- length(income)
2 > u <- seq(1 / (2 * n), 1 - 1 / (2 * n),
  length=n)
3 > plot(u, sort(income), type="l")
  plot(ecdf(income))
```



c.d.f., quantiles and Lorenz

To get Lorenz curve, we substitute on the y -axis *proportion of incomes to incomes*.

```
1 > library(ineq)
2 > Lc(income)
3 > L <- function(u) Lc(income)$L[
  round(u*length(income))]
```



c.d.f., quantiles and Lorenz

	x -axis	y -axis
c.d.f.	income	proportion of population
Pen's parade (quantile)	proportion of population	income
Lorenz curve	proportion of population	proportion of income

Standard statistical measure of dispersion

The **variance** for a sample $\mathbf{X} = \{x_1, \dots, x_n\}$ is

$$\text{Var}(\mathbf{X}) = \frac{1}{n} \sum_{i=1}^n [x_i - \bar{x}]^2$$

where the baseline (reference) is $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$.

```
1 > var(income)
```

```
2 [1] 34178.43
```

problem it is a quadratic function, $\text{Var}(\alpha \mathbf{X}) = \alpha^2 \text{Var}(\mathbf{X})$.

Standard statistical measure of dispersion

An alternative is the **coefficient of variation**,

$$cv(\mathbf{X}) = \frac{\sqrt{\text{Var}(\mathbf{X})}}{\bar{x}}$$

But not a good measure to capture inequality overall, very sensitive to very high incomes

```
1 > cv <- function(x) sd(x)/mean(x)
2 > cv(income)
3 [1] 0.6154011
```


Standard statistical measure of dispersion

An alternative is to use a logarithmic transformation. Use the **logarithmic variance**

$$\text{Var}_{\log}(\mathbf{X}) = \frac{1}{n} \sum_{i=1}^n [\log(x_i) - \log(\bar{x})]^2$$

```
1 > var_log <- function(x) var(log(x))  
2 > var_log(income)  
3 [1] 0.2921022
```

Those measures are distances **on the x -axis**.

Standard statistical measure of dispersion

Other inequality measures can be derived from Pen's parade of the Dwarfs, where measures are based on distances **on the y-axis**, i.e. distances between quantiles.

$$Q_p = F^{-1}(p) \text{ i.e. } F(Q_p) = p$$

e.g. the **median** is the quantile when $p = 50\%$, the first **quartile** is the quantile when $p = 25\%$, the first **quintile** is the quantile when $p = 20\%$, the first **decile** is the quantile when $p = 10\%$, the first **percentile** is the quantile when $p = 1\%$

1	> quantile (income , c (.1 , .5 , .9 , .99))			
2	10%	50%	90%	99%
3	137.6294	253.9090	519.6887	933.9211

Standard statistical measure of dispersion

Define the quantile ratio as

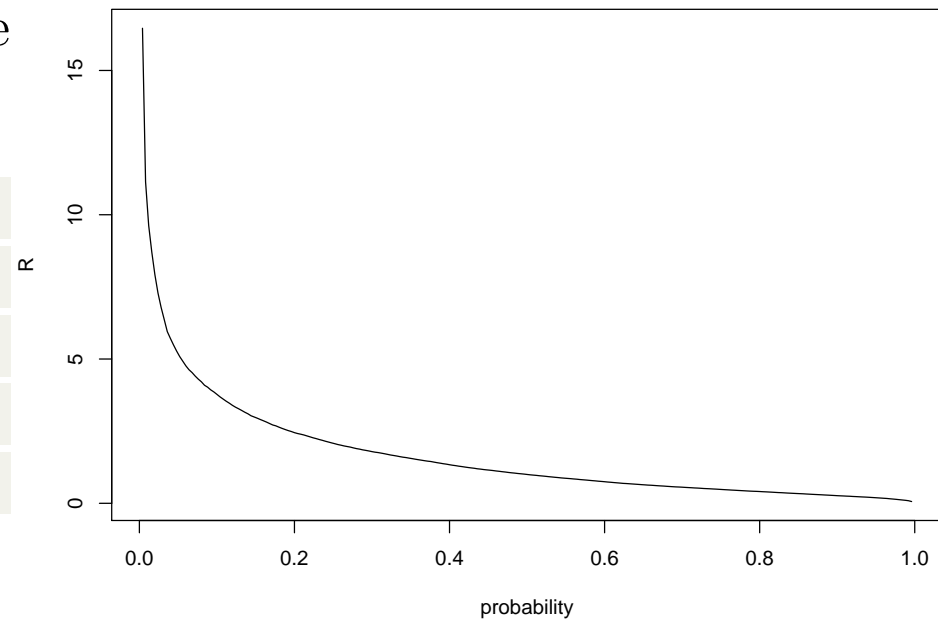
$$R_p = \frac{Q_{1-p}}{Q_p}$$

In case of perfect equality, $R_p = 1$.

The most popular one is probably the 90/10 ratio.

```

1 > R_p <- function(x, p) quantile(x
    , 1-p) / quantile(x, p)
2 > R_p(income, .1)
3 90%
4 3.776
  
```



This index measures the gap between the rich and the poor.

E.g. $R_{0.1} = 10$ means that top 10% incomes are more than 10 times higher than the bottom 10% incomes.

Ignores the distribution (apart from the two points), violates transfer principle.

An alternative measure might be Kuznets Ratio, defined from Lorenz curve as the ratio of the share of income earned by the poorest p share of the population and the richest r share of the population,

$$I(p, r) = \frac{L(p)}{1 - L(1 - r)}$$

But here again, it ignores the distribution between the cutoffs and therefore violates the transfer principle.

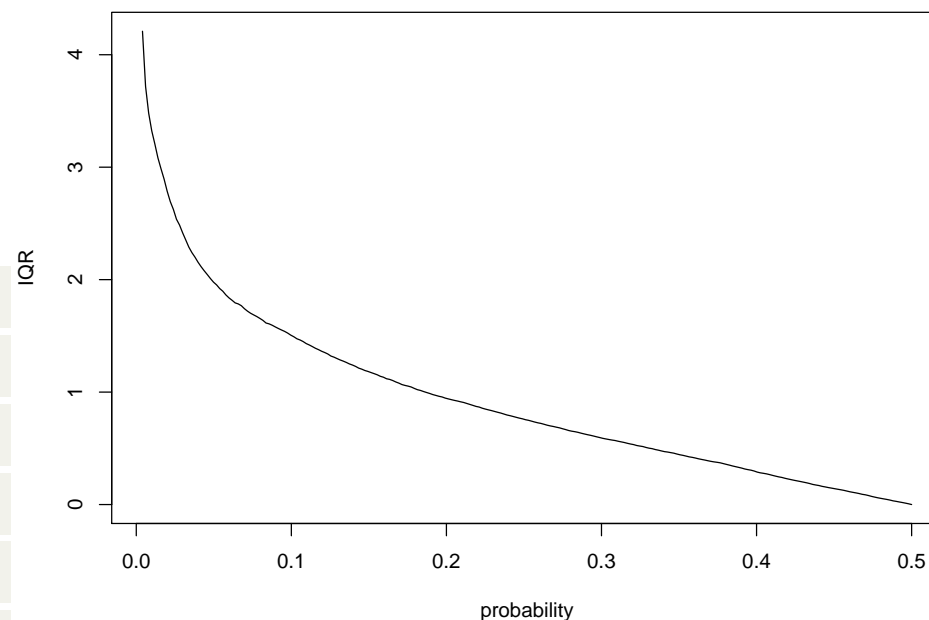
An alternative measure can be the IQR,
interquantile ratio,

$$IQR_p = \frac{Q_{1-p} - Q_p}{Q_{0.5}}$$

```

1 > IQR_p <- function(x, p) (
    quantile(x, 1-p) - quantile(x, p)
    ) / quantile(x, .5)
2 > IQR_p(income, .1)
3     90%
4 1.504709

```



Problem only focuses on top $(1 - p)$ -th and bottom p -th proportion. Does not care about what happens between those quantiles.

Standard statistical measure of dispersion

Pen's parade suggest to measure the green area, for some $p \in (0, 1)$, M_p ,

```

1 > M_p <- function(x, p) {
2   a <- seq(0, p, length=251)
3   b <- seq(p, 1, length=251)
4   ya <- quantile(x, p) - quantile(x,
5     a)
6   a1 <- sum((ya[1:250] + ya[2:251])
7     / 2 * p / 250)
8   yb <- quantile(x, b) - quantile(x,
9     p)
10  b1 <- sum((yb[1:250] + yb[2:251])
11    / 2 * (1 - p) / 250)
12  return(a1 + b1)}

```

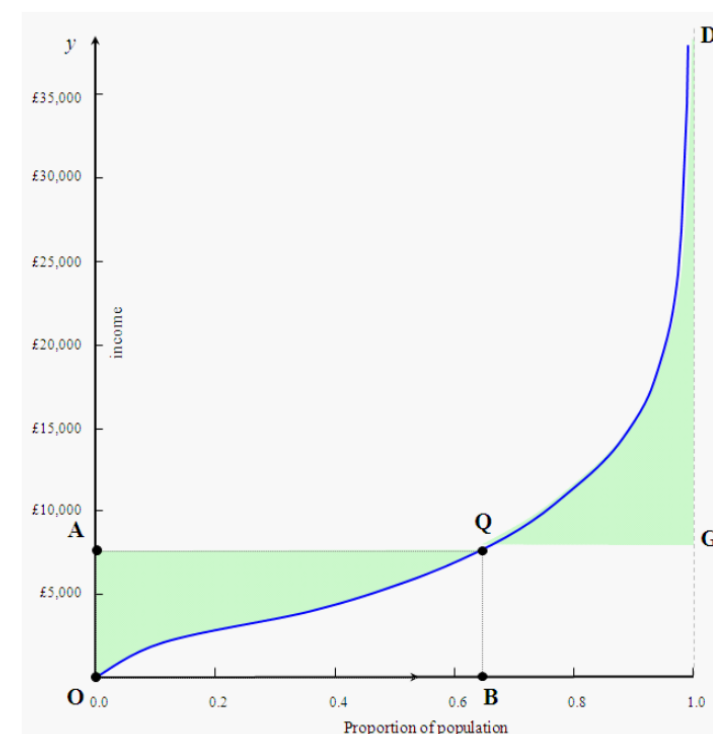


Figure 2.1: The Parade of Dwarfs. UK Income Before Tax, 1984/5. Source: Economic Trends, November 1987

Standard statistical measure of dispersion

Use also the relative mean deviation

$$M(\mathbf{X}) = \frac{1}{n} \sum_{i=1}^n \left| \frac{x_i}{\bar{x}} - 1 \right|$$

```
1 > M <- function(x) mean(abs(x/mean(x) - 1))  
2 > M(income)  
3 [1] 0.429433
```

in case of perfect equality, $M = 0$

Standard statistical measure of dispersion

Finally, why not use Lorenz curve.

It can be defined using order statistics as

$$G = \frac{2}{n(n-1)\bar{x}} \sum_{i=1}^n i \cdot x_{i:n} - \frac{n+1}{n-1}$$

```

1 > n <- length(income)
2 > mu <- mean(income)
3 2*sum((1:n)*sort(income)) / (mu*n*(n-1)) - (n
  +1) / (n-1)
4 [1] 0.2976282

```

Gini index is defined as the area below the first diagonal and above Lorenz curve

Standard statistical measure of dispersion

$$G(\mathbf{X}) = \frac{1}{2n^2\bar{x}} \sum_{i,j=1}^n |x_i - x_j|$$

Perfect equality is obtained when $G = 0$.

Remark Gini index can be related to the variance or the coefficient of variation, since

$$\text{Var}(\mathbf{X}) = \frac{1}{n} \sum_{i=1}^n [x_i - \bar{x}]^2 = \frac{1}{n^2} \sum_{i,j=1}^n (x_i - x_j)^2$$

Here,

$$G(\mathbf{X}) = \frac{\Delta(\mathbf{X})}{2\bar{x}} \text{ with } \Delta(\mathbf{X}) = \frac{1}{n^2} \sum_{i,j=1}^n |x_i - x_j|$$

```
1 > ineq (income , " Gini " )
```

```
2 [1] 0.2975789
```

Axiomatic Approach for Inequality Indices

Need some rules to say if a principle used to divide a cake of fixed size amongst a fixed number of people is fair, or not.

A standard one is the **Anonymity Principle**. Let $\mathbf{X} = \{x_1, \dots, x_n\}$, then

$$I(x_1, x_2, \dots, x_n) = I(x_2, x_1, \dots, x_n)$$

also called **Replication Invariance Principle**

The **Transfert Principle**

for any given income distribution if you take a small amount of income from one person and give it to a richer person then income inequality must increase

Pigou (1912) and Dalton (1920), a transfer from a richer to a poorer person will decrease inequality. Let $\mathbf{X} = \{x_1, \dots, x_n\}$ with $x_1 \leq \dots \leq x_n$, then

$$I(x_1, \dots, x_i, \dots, x_j, \dots, x_n) \geq I(x_1, \dots, x_i + \delta, \dots, x_j - \delta, \dots, x_n)$$

Nevertheless, not easy to compare,
compare e.g. Monday and Tuesday

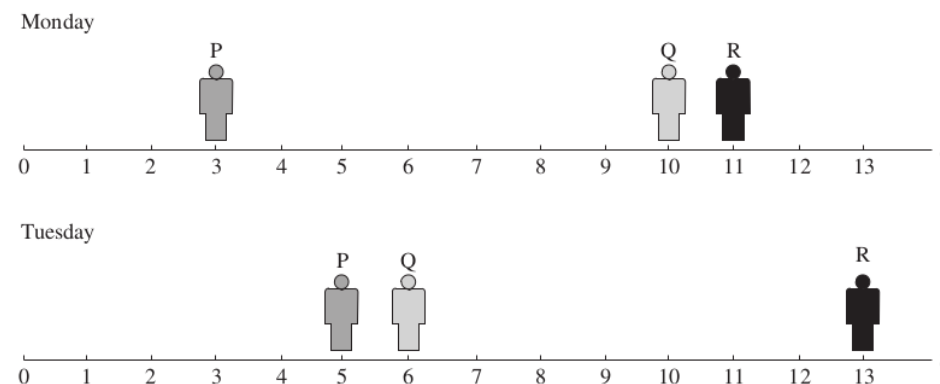


Figure 2.3. Inequality comparisons in a three-person world.

An important concept behind is the idea of **mean preserving spread** : with those $\pm\delta$ preserve the total wealth.

The **Scale Independence Principle**

What if double everyone's income? if standards of living are determined by real income and there is inflation : inequality is unchanged

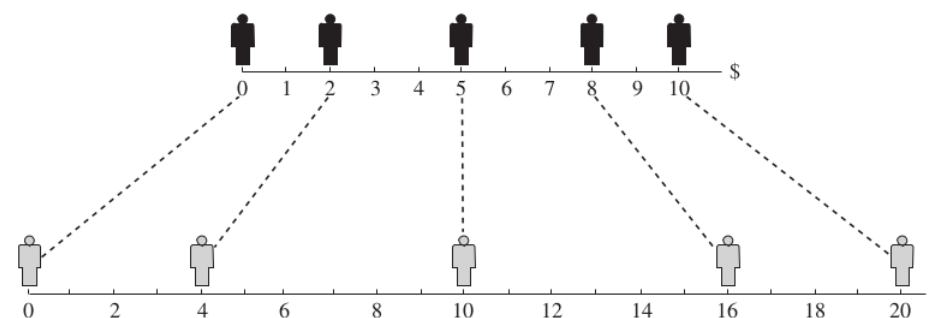


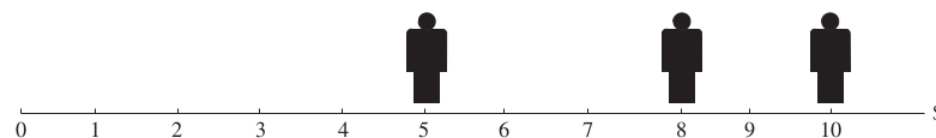
Figure 2.4. Scale independence.

Let $\mathbf{X} = \{x_1, \dots, x_n\}$, then

$$I(\lambda x_1, \dots, \lambda x_n) = I(x_1, \dots, x_n)$$

also called **Zero-Degree Homogeneity** property.

The **Population Principle**



Consider clones of the economy

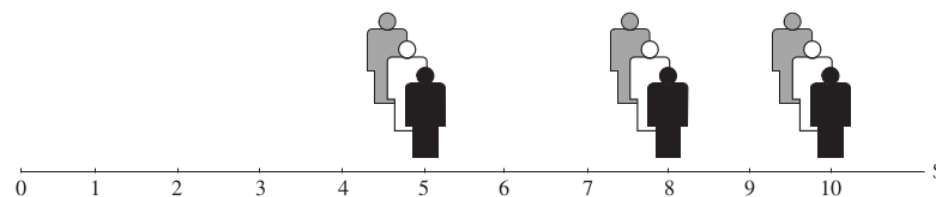


Figure 2.6. The population principle.

$$I(\underbrace{x_1, \dots, x_1}_{k \text{ times}}, \dots, \underbrace{x_n, \dots, x_n}_{k \text{ times}}) = I(x_1, \dots, x_n)$$

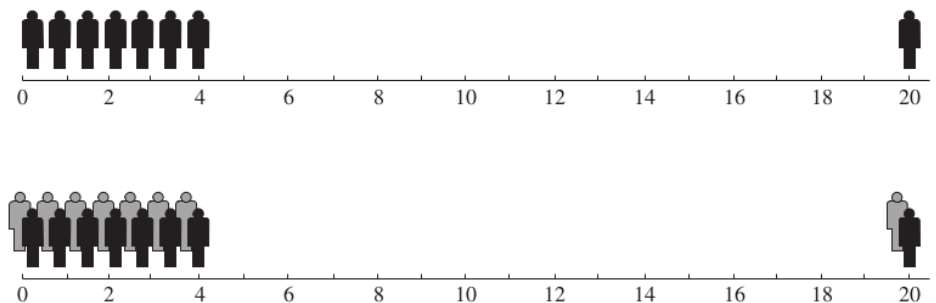


Figure 2.7. Population replication – has inequality fallen?

Is it really that simple?

The Decomposability Principle

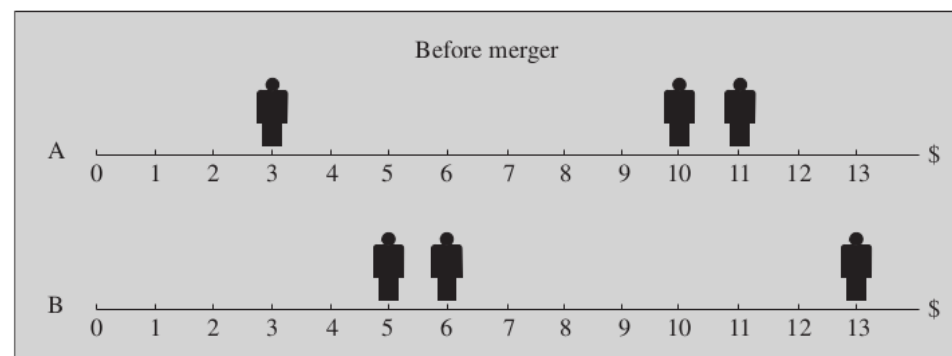
Assume that we can decompose inequality by subgroups (based on gender, race, countries, etc)

According to this principle, if inequality increases in a subgroup, it increases in the whole population, *ceteris paribus*

$$I(x_1, \dots, x_n, y_1, \dots, y_n) \leq I(x_1^*, \dots, x_n^*, y_1, \dots, y_n)$$

as long as $I(x_1, \dots, x_n) \leq I(x_1^*, \dots, x_n^*)$.

Consider two groups, X and X^*



Then add the same subgroup Y to both X and X^*

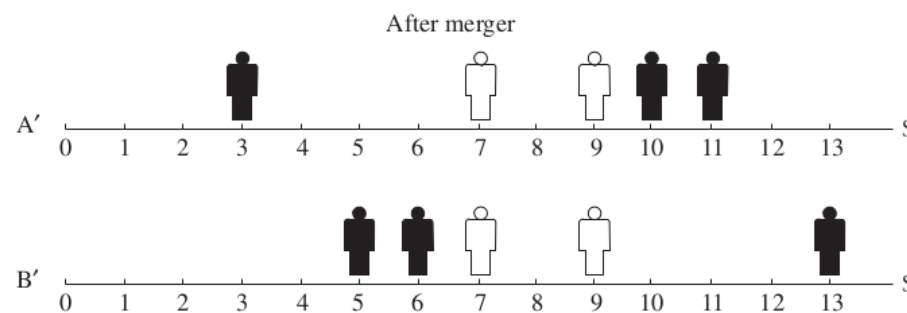


Figure 2.8. Decomposability.

Axiomatic Approach for Inequality Indices

Any inequality measure that simultaneously satisfies the properties of the principle of transfers, scale independence, population principle and decomposability must be expressible in the form

$$E_{\xi} = \frac{1}{\xi^2 - \xi} \left(\frac{1}{n} \sum_{i=1}^n \left(\frac{x_i}{\bar{x}} \right)^{\xi} - 1 \right)$$

for some $\xi \in \mathbb{R}$. This is the [generalized entropy measure](#).

```

1 > entropy (income ,0)
2 [1] 0.1456604
3 > entropy (income ,.5)
4 [1] 0.1446105
5 > entropy (income ,1)
6 [1] 0.1506973
7 > entropy (income ,2)
8 [1] 0.1893279

```

The higher ξ , the more sensitive to high incomes.

Remark rule of thumb, take $\xi \in [-1, +2]$.

When $\xi = 0$, the **mean logarithmic deviation** (MLD),

$$MLD = E_0 = -\frac{1}{n} \sum_{i=1}^n \log \left(\frac{x_i}{\bar{x}} \right)$$

When $\xi = 1$, the **Theil index**

$$T = E_1 = \frac{1}{n} \sum_{i=1}^n \frac{x_i}{\bar{x}} \log \left(\frac{x_i}{\bar{x}} \right)$$

```
1 > Theil(income)
```

```
2 [1] 0.1506973
```

When $\xi = 2$, the index can be related to the coefficient of variation

$$E_2 = \frac{[\text{coefficient of variation}]^2}{2}$$

In a 3-person economy, it is possible to visualize curve of iso-indices,

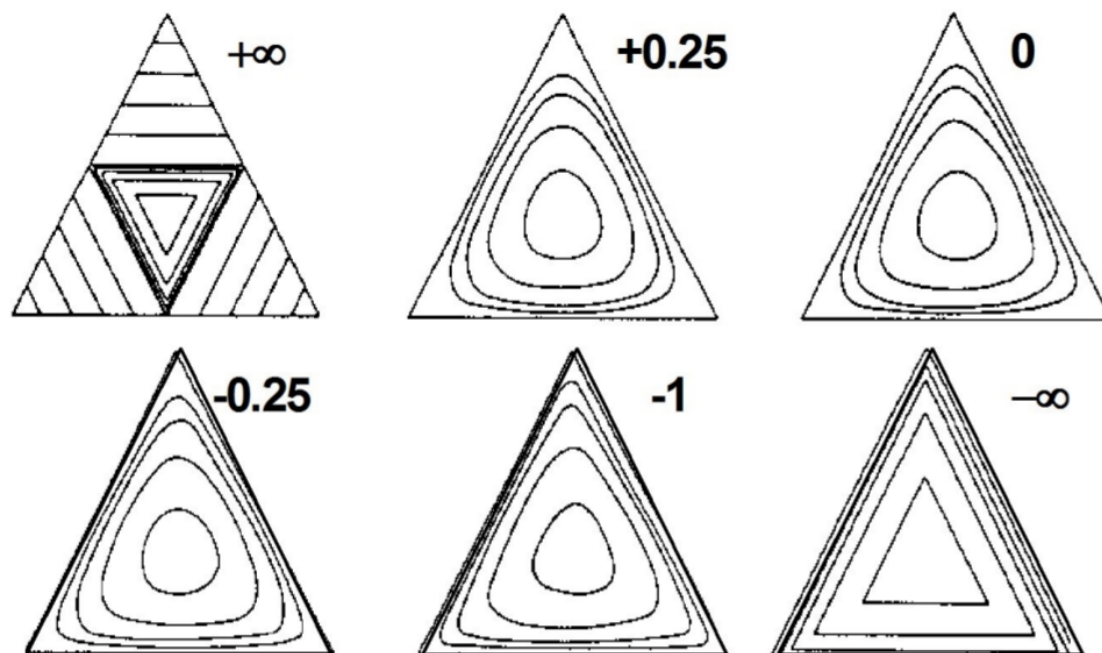


Figure 7: Generalised entropy contours for different values of α

A related index is [Atkinson inequality index](#),

$$A_{\epsilon} = 1 - \left(\frac{1}{n} \sum_{i=1}^n \left(\frac{x_i}{\bar{x}} \right)^{1-\epsilon} \right)^{\frac{1}{1-\epsilon}}$$

with $\epsilon \geq 0$.

```
1 > Atkinson(income, 0.5)
```

```
2 [1] 0.07099824
```

```
3 > Atkinson(income, 1)
```

```
4 [1] 0.1355487
```

In the case where $\epsilon \rightarrow 1$, we obtain

$$A_1 = 1 - \prod_{i=1}^n \left(\frac{x_i}{\bar{x}} \right)^{\frac{1}{n}}$$

ϵ is usually interpreted as an [aversion to inequality](#) index.

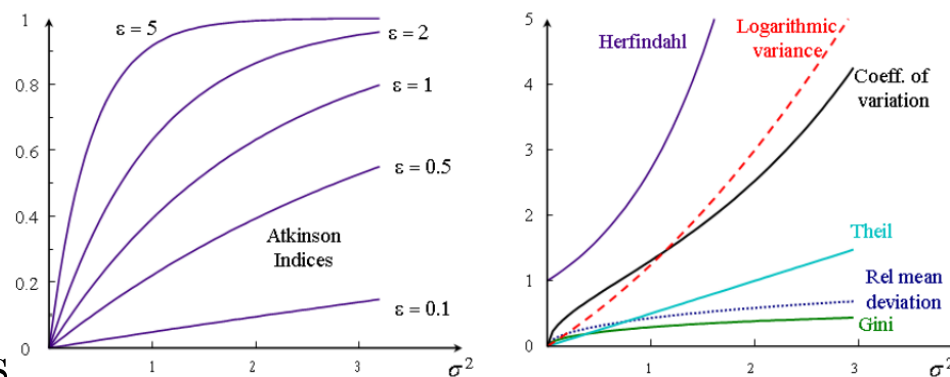
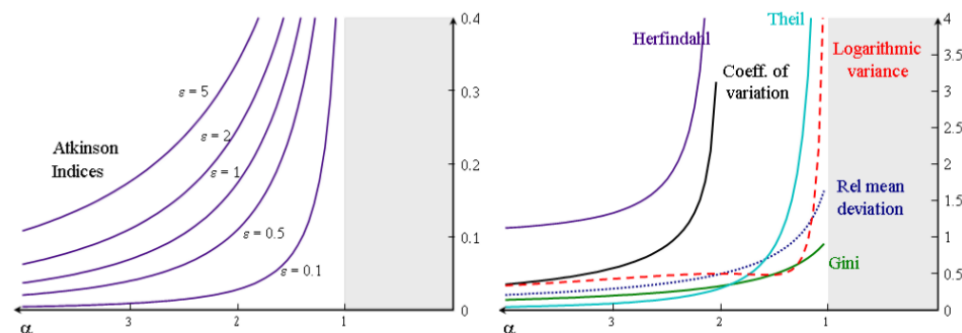
Observe that

$$A_{\epsilon} = 1 - [(\epsilon^2 - \epsilon)E_{1-\epsilon} + 1]^{\frac{1}{1-\epsilon}}$$

and the limiting case $A_1 = 1 - \exp[-E_0]$.

Thus, the Atkinson index is ordinally equivalent to the GE index, since they produce the same ranking of different distributions.

Consider indices obtained when \mathbf{X} is obtained from a $LN(0, \sigma^2)$ distribution and from a $\mathcal{P}(\alpha)$ distribution.

Figure 4.4: Inequality and the Lognormal parameter σ^2 Figure 4.9: Inequality and Pareto's α

Changing the Axioms

Is there an agreement about the axioms ?

For instance, no unanimous agreement on the scale independence axiom,

Why not a **translation independence** axiom ?

Translation Independence Principle : if every incomes are increased by the same amount, the inequality measure is unchanged

Given $\mathbf{X} = (x_1, \dots, x_n)$,

$$I(x_1, \dots, x_n) = I(x_1 + h, \dots, x_n + h)$$

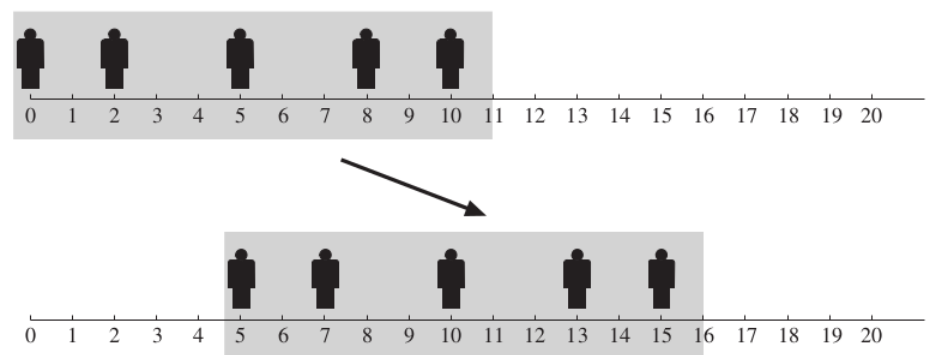


Figure 2.5. Translation independence.

If we change the scale independence principle by this translation independence, we get other indices.

Changing the Axioms

Kolm indices satisfy the principle of transfers, translation independence, population principle and decomposability

$$K_{\theta} = \log \left(\frac{1}{n} \sum_{i=1}^n e^{\theta[x_i - \bar{x}]} \right)$$

```
1 > Kolm(income, 1)
```

```
2 [1] 291.5878
```

```
3 > Kolm(income, .5)
```

```
4 [1] 283.9989
```

From Measuring to Ordering

Over time, between countries, before/after tax, etc.

\mathbf{X} is said to be Lorenz-dominated by \mathbf{Y} if $L_{\mathbf{X}} \leq L_{\mathbf{Y}}$. In that case \mathbf{Y} is more equal, or less unequal.

In such a case, \mathbf{X} can be reached from \mathbf{Y} by a sequence of poorer-to-richer pairwiser income transfers.

In that case, any inequality measure satisfying the population principle, scale independence, anonymity and principle of transfers axioms are consistent with the Lorenz dominance (namely Theil, Gini, MLD, Generalized Entropy and Atkinson).

Remark A regressive transfer will move the Lorenz curve further away from the diagonal. So satisfies transfer principle. And it satisfies also the scale invariance property.

Example if $X_i \sim \mathcal{P}(\alpha_i, x_i)$,

$$L_{\mathbf{X}_1} \leq L_{\mathbf{X}_2} \iff \alpha_1 \leq \alpha_2$$

and if $X_i \sim LN(\mu_i, \sigma_i^2)$,

$$L_{\mathbf{X}_1} \leq L_{\mathbf{X}_2} \iff \sigma_1^2 \geq \sigma_2^2$$

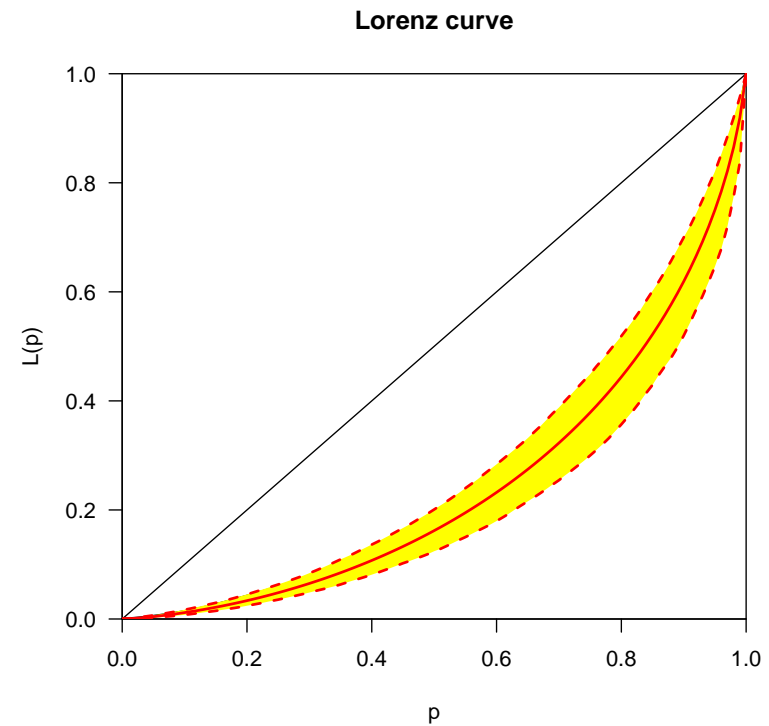
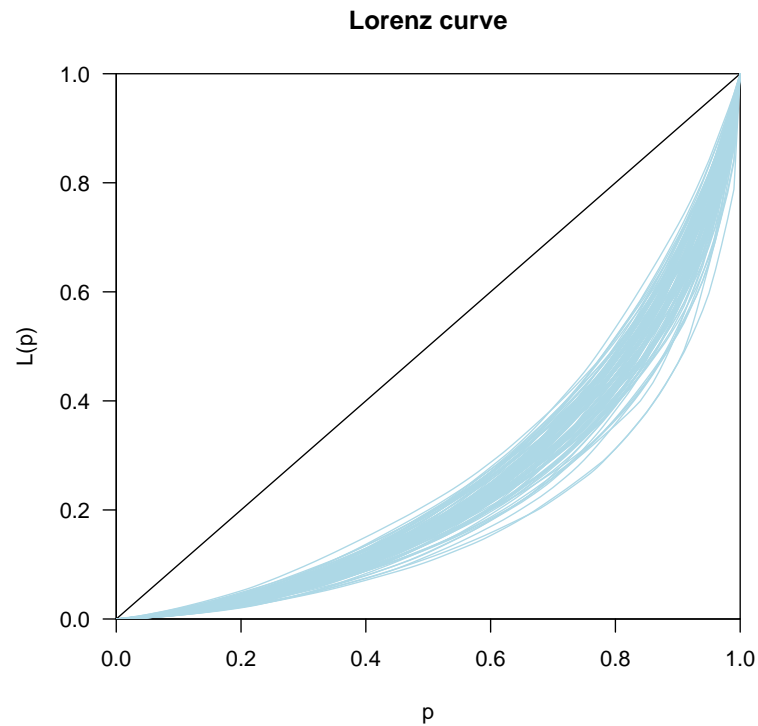
Lorenz dominance is a relation that is incomplete : when Lorenz curves cross, the criterion cannot decide between the two distributions.

→ the ranking is considered unambiguous.

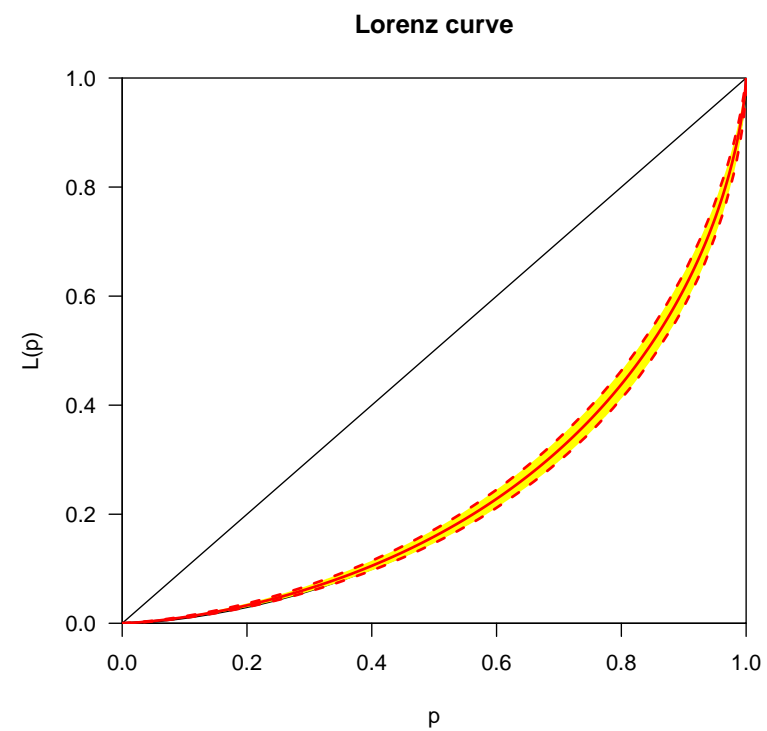
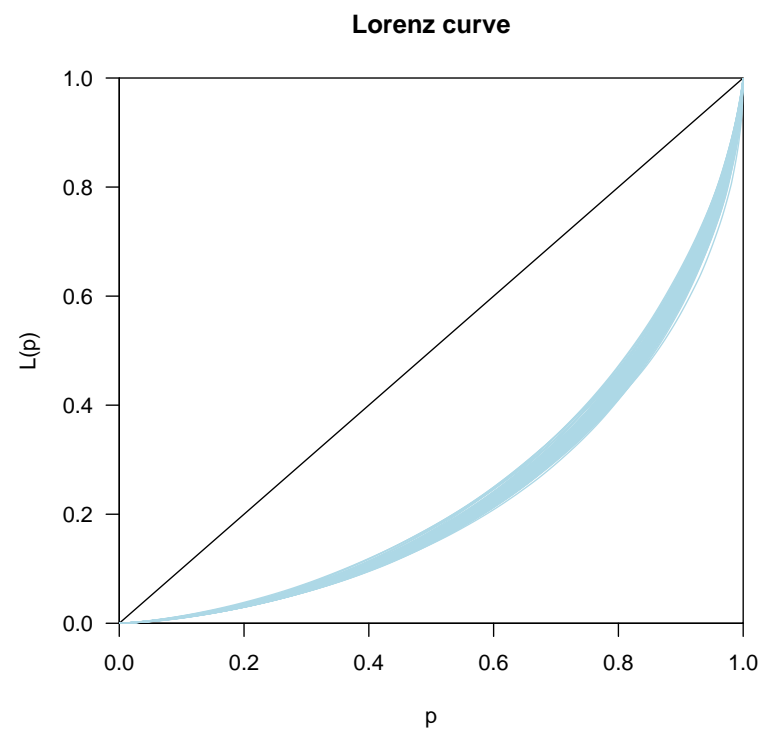
Further, one should take into account possible random noise.

Consider some sample $\{x_1, \dots, x_n\}$ from a $LN(0, 1)$ distribution, with $n = 100$.

The 95% confidence interval is

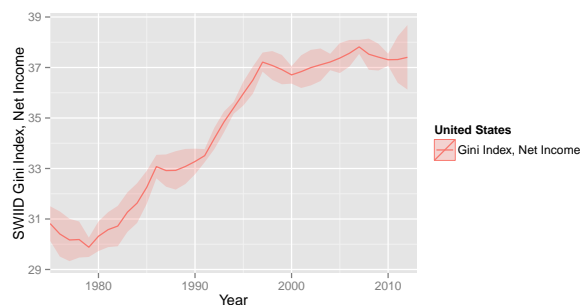


Consider some sample $\{x_1, \dots, x_n\}$ from a $LN(0, 1)$ distribution, with $n = 1,000$. The 95% confidence interval is

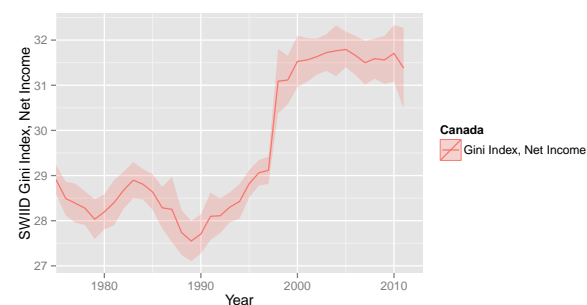


Looking for Confidence

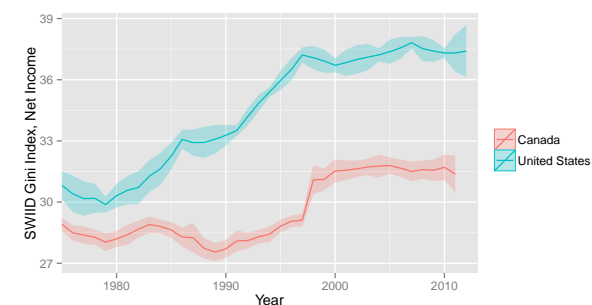
See e.g. <http://myweb.uiowa.edu/fsolt/swiid/>, for the estimation of Gini index over time + over several countries.



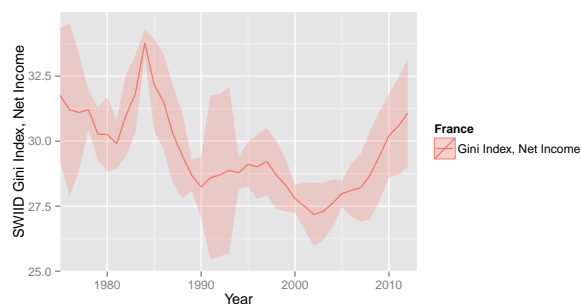
Note: Solid lines indicate mean estimates; shaded regions indicate the associated 95% confidence intervals.
Source: Standardized World Income Inequality Database v5.0 (Solt 2014).



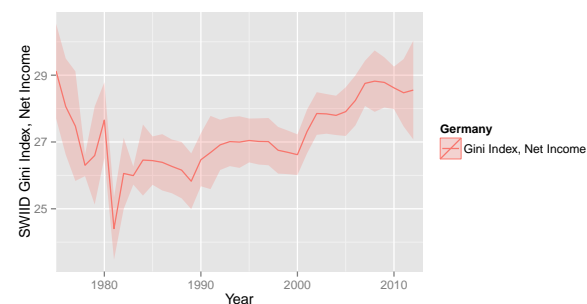
Note: Solid lines indicate mean estimates; shaded regions indicate the associated 95% confidence intervals.
Source: Standardized World Income Inequality Database v5.0 (Solt 2014).



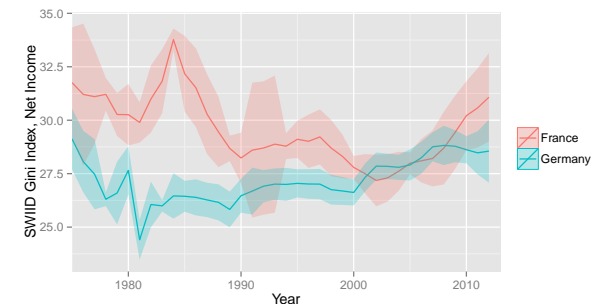
Note: Solid lines indicate mean estimates; shaded regions indicate the associated 95% confidence intervals.
Source: Standardized World Income Inequality Database v5.0 (Solt 2014).



Note: Solid lines indicate mean estimates; shaded regions indicate the associated 95% confidence intervals.
Source: Standardized World Income Inequality Database v5.0 (Solt 2014).



Note: Solid lines indicate mean estimates; shaded regions indicate the associated 95% confidence intervals.
Source: Standardized World Income Inequality Database v5.0 (Solt 2014).



Note: Solid lines indicate mean estimates; shaded regions indicate the associated 95% confidence intervals.
Source: Standardized World Income Inequality Database v5.0 (Solt 2014).

Looking for Confidence

To get confidence interval for indices, use bootstrap techniques (see last week).

The code is simply

```

1 > IC <- function(x, f, n=1000, alpha=.95) {
2   + F=rep(NA, n)
3   + for(i in 1:n) {
4     + F[i]=f(sample(x, size=length(x), replace=TRUE)) }
5   + return(quantile(F, c((1-alpha)/2, 1-(1-alpha)/2))) }

```

For instance,

```

1 > IC(income, Gini)
2      2.5%      97.5%
3 0.2915897 0.3039454

```

(the sample is rather large, $n = 6,043$).

Looking for Confidence

```
1 > IC(income , Gini)
2      2.5%      97.5%
3 0.2915897  0.3039454
4 > IC(income , Theil)
5      2.5%      97.5%
6 0.1421775  0.1595012
7 > IC(income , entropy)
8      2.5%      97.5%
9 0.1377267  0.1517201
```

Back on Gini Index

We've seen Gini index as an area,

$$G = 2 \int_0^1 [p - L(p)] dp = 1 - 2 \int_0^1 L(p) dp$$

Using integration by parts, $u' = 1$ and $v = L(p)$,

$$G = -1 + 2 \int_0^1 p L'(p) dp = \frac{2}{\mu} \left(\int_0^\infty y F(y) f(y) dy - \frac{\mu}{2} \right)$$

using a change of variables, $p = F(y)$ and because $L'(p) = F^{-1}(p)/\mu = y/\mu$.

Thus

$$G = \frac{2}{\mu} \text{cov}(y, F(y))$$

→ Gini index is proportional to the covariance between the income and its rank.

Back on Gini Index

Using integration by parts, one can then write

$$G = \frac{1}{2} \int_0^\infty F(x)[1 - F(x)]dx = 1 - \frac{1}{\mu} \int_0^\infty [1 - F(x)]^2 dx.$$

which can also be written

$$G = \frac{1}{2\mu} \int_{\mathbb{R}_+^2} |x - y| dF(x) dF(y)$$

(see previous discussion on connexions between Gini index and the variance)

Decomposition(s)

When studying inequalities, it might be interesting to discuss possible decompositions either by subgroups, or by sources,

- **subgroups decomposition**, e.g Male/Female, Rural/Urban see FAO (2006, fao.org)
- **source decomposition**, e.g earnings/gvnt benefits/investment/pension, etc, see slide 41 #1 and FAO (2006, fao.org)

For the variance, decomposition per groups is related to ANOVA,

$$\text{Var}(Y) = \underbrace{\mathbb{E}[\text{Var}(Y|X)]}_{\text{within}} + \underbrace{\text{Var}(\mathbb{E}[Y|X])}_{\text{between}}$$

Hence, if $X \in \{x_1, \dots, x_k\}$ (k subgroups),

$$\text{Var}(Y) = \underbrace{\sum_k p_k \text{Var}(Y | \text{group } k)}_{\text{within}} + \underbrace{\text{Var}(\mathbb{E}[Y|X])}_{\text{between}}$$

Decomposition(s)

For Gini index, it is possible to write

$$G(Y) = \underbrace{\sum_k \omega_k G(Y | \text{group } k)}_{\text{within}} + \underbrace{G(\bar{Y})}_{\text{between}} + \text{residual}$$

for some weights ω , where the between term is the Gini index between subgroup means. But the decomposition is not perfect.

More generally, for General Entropy indices,

$$E_\xi(Y) = \underbrace{\sum_k \omega_k E_\xi(Y | \text{group } k)}_{\text{within}} + \underbrace{E_\xi(\bar{Y})}_{\text{between}}$$

where $E_\xi(\bar{Y})$ is the entropy on the subgroup means

$$\omega_k = \left(\frac{\bar{Y}_k}{\bar{Y}} \right)^\xi (p_k)^{1-\xi}$$

Decomposition(s)

Now, a decomposition per source, i.e. $Y_i = Y_{1,i} + \dots + Y_{k,i} + \dots$, among sources.

For Gini index natural decomposition was suggested by Lerman & Yitzhaki (1985, [jstor.org](https://www.jstor.org/stable/2282821))

$$G(Y) = \frac{2}{\bar{Y}} \text{cov}(Y, F(Y)) = \sum_k \underbrace{\frac{2}{\bar{Y}} \text{cov}(Y_{\textcolor{red}{k}}, F(Y))}_{k\text{-th contribution}}$$

thus, it is based on the covariance between the k -th source and the ranks based on cumulated incomes.

Similarly for Theil index,

$$T(Y) = \sum_k \underbrace{\frac{1}{n} \sum_i \left(\frac{Y_{k,i}}{\bar{Y}} \right) \log \left(\frac{Y_i}{\bar{Y}} \right)}_{k\text{-th contribution}}$$

Decomposition(s)

It is possible to use **Shapley value** for decomposition of indices $\mathcal{I}(\cdot)$. Consider m groups, $N = \{1, \dots, m\}$, and define $\mathcal{I}(S) = I(\mathbf{x}_S)$ where $S \subset N$. Then Shapley value yields

$$\phi_k(v) = \sum_{S \subseteq N \setminus \{k\}} \frac{|S|! (m - |S| - 1)!}{m!} (\mathcal{I}(S \cup \{k\}) - \mathcal{I}(S))$$