

```
---
title: "Lorenz Curve and Gini Coefficient with Piketty Data"
author: "coop711"
date: "2015년 4월 24일"
output: html_document
---
```

Data Description

Thomas Piketty가 쓴 **21세기 자본**에는 자본과 소득의 불평등에 대한 다양한 자료가 등장한다. 특히 점유율(share)이라는 관점에서 수집한 자료들이 매우 흥미를 자아낸다. 이 자료들은 [Thomas Piketty - Capital in the 21st century](http://pikett

작업 파일에는 데이터 뿐 아니라 간단한 구조로 lorenz curve를 그리고, gini계수를 계산하는 함수가 저장되어 있다.
먼저 자료부터 확인하자.

```
```{r}
load("piketty_gini.rda")
ls()
```
```

이 중에서 `piketty.labor`는 노동소득의 점유율, `piketty.co`는 자본소득의 점유율, `piketty.total`은 총소득의 점유율을 나타낸다.

```
```{r}
piketty.labor
```
```

노동소득의 경우 불평등이 낮은 경우(1970~1980년대 스칸디나비아 국가들)는 하위 50%가 전체 노동소득 중 35%를 가져가고 상위 10%는 20%를 가져가는데 그 중에서 상위 1%가 5%를 가져가고 있다. 불평등이 심해질수록 하위 50%에게 돌아가는 몫은 줄어듦고 상위 1%가 가져

```
```{r}
cbind(x = piketty.labor$x, apply(piketty.labor[, 2:5], 2, cumsum))
```
```

자본소득의 집중을 살펴보자

```
```{r}
piketty.co
```
```

자본소득의 경우 불평등이 낮은 경우(존재한 적 없는 이상국가)는 하위 50%가 전체 노동소득 중 25%를 가져가고 상위 10%는 30%를 가져가는 데 그 중에서 상위 1%가 10%를 가져가고 있다. 불평등이 심해질수록 하위 50%에게 돌아가는 몫은 줄어듦고 상위 1%가 가져가는 몫이

```
```{r}
cbind(x = piketty.co$x, apply(piketty.co[, 2:6], 2, cumsum))
```
```

총소득의 집중을 살펴보자.

```
```{r}
piketty.total
```
```

총소득의 불평등이 비교적 낮은 사회(1970~1980년대의 스칸디나비아 국가들)에서는 하위 50%가 자체 총소득의 30%를 가져가고 상위 10%는 25%를 가져가는 데 그 중에서 상위 1%는 7%를 가져간다. 불평등이 매우 심한 경우(2030년 미국 예상) 하위 50%에게 돌아가는 몫은

```
```{r}
cbind(x = piketty.total$x, apply(piketty.total[, 2:5], 2, cumsum))
x<-piketty.total$x
```
```

Plots

이 상황을 그래프로 옮겨보자. x-축에는 소득 순서대로 누적된 사람의 비율을 표시하고 y-축에는 그 사람들이 얼마정도 소득을 축적했는지 누적시켜간다. 따라서 첫 단계는 소득의 누적분포를 구하는 것이다.
우선 노동소득의 `Low` 열에 대하여 `plot()`을 이용하여 누적도표를 그린다. `cumsum(piketty.labor\$Low)` 등이 반복해서 나오게 되므로 별도의 R 객체로 저장한다.

```
```{r fig.width=4.5, fig.height=4.5}
piketty.labor.cum<-apply(piketty.labor[, 2:5], 2, cumsum)
y<-piketty.labor.cum[, 1]
plot(x = x, y = y, type="l", ann=FALSE, xaxt="n", yaxt="n")
```
```

x축과 y축의 눈금을 표시한다.

```

```{r fig.width=6, fig.height=6}
plot(x = x, y = y, type="l", ann=FALSE, xaxt="n", yaxt="n")
axis(side = 1, at = x, labels = paste(x*100))
axis(side = 2, at = y, labels = paste(y*100))
```

```

완전 평등선이라고 할 수 있는 $y = x$ 를 같은 영역에 그린다.

```

```{r fig.width=6, fig.height=6}
plot(x = x, y = y, type="l", ann=FALSE, xaxt="n", yaxt="n")
axis(side = 1, at = x, labels = paste(x*100))
axis(side = 2, at = y, labels = paste(y*100))
lines(x = x, y = x)
```

```

$x = \text{`r`}$ 과 y 축의 아래 위에 점선으로 된 격자를 설치하여 윤곽을 쉽게 파악하도록 한다.

```

```{r fig.width=6, fig.height=6}
plot(x = x, y = y, type="l", ann=FALSE, xaxt="n", yaxt="n")
axis(side = 1, at = x, labels = paste(x*100))
axis(side = 2, at = y, labels = paste(y*100))
lines(x = x, y = x)
abline(v = x, h = c(0, 1), lty=2)
```

```

``polygon()``을 이용하여 평등선과 소득누적분포의 차이를 빗금으로 표시한다.

```

```{r fig.width=6, fig.height=6}
plot(x = x, y = y, type="l", ann=FALSE, xaxt="n", yaxt="n")
axis(side = 1, at = x, labels = paste(x*100))
axis(side = 2, at = y, labels = paste(y*100))
lines(x = x, y = x)
abline(v = x, h = c(0, 1), lty=2)
polygon(x = c(x, 0), y = c(y, 0), density = 15, angle = 135)
```

```

``lines()``를 이용하여 나머지 소득 누적곡선을 그린다.

```

```{r fig.width=6, fig.height=6}
plot(x = x, y = y, type="l", ann=FALSE, xaxt="n", yaxt="n")
axis(side = 1, at = x, labels = paste(x*100))
axis(side = 2, at = y, labels = paste(y*100))
lines(x = x, y = x)
abline(v = x, h = c(0, 1), lty=2)
polygon(x = c(x, 0), y = c(y, 0), density = 15, angle = 135)
lines(x = x, y = piketty.labor.cum[, 2], col="blue")
lines(x = x, y = piketty.labor.cum[, 3], col="red")
lines(x = x, y = piketty.labor.cum[, 4], col="grey")
```

```

``legend()`` 와 ``title()`` 로 마무리짓는다.

```

```{r fig.width=6, fig.height=6}
plot(x = x, y = y, type="l", ann=FALSE, xaxt="n", yaxt="n")
axis(side = 1, at = x, labels = paste(x*100))
axis(side = 2, at = y, labels = paste(y*100))
lines(x = x, y = x)
abline(v = x, h = c(0, 1), lty=2)
polygon(x = c(x, 0), y = c(y, 0), density = 15, angle = 135)
lines(x = x, y = piketty.labor.cum[, 2], col="blue")
lines(x = x, y = piketty.labor.cum[, 3], col="red")
lines(x = x, y = piketty.labor.cum[, 4], col="cyan")
title(main = "노동 소득의 로렌츠 곡선", xlab = "소득순 사람들의 누적(%)", ylab = "소득의 누적(%)")
legend(x = 0.05, y = 0.95, legend=c("낮음", "중간", "높음", "매우 높음"), lty = 1, col = c("black", "blue", "red", "cyan"), title="소득 불평등")
```

```

지금까지의 과정을 잘 살펴보면 소득순 사람들의 누적분포와 이에 따른 소득의 누적분포를 각각 x 와 y 로 입력하면 로렌츠 곡선을 그리고 평등선과의 차이를 빗금으로 표시할 수도 있다는 것을 알 수 있다. 이를 함수로 나타낸 것이 ``lorenz()``이다. 코드를 살펴보면,

```

```{r}

```

```
lorenz
```
```

이제 ``lorenz()`` 함수를 이용하여 자본소득의 로렌츠 곡선들을 그려보자.

```
```{r fig.width=6, fig.height=6}
piketty.co.cum<-apply(piketty.co[, 2:6], 2, cumsum)
piketty.co.cum
y<-piketty.co.cum[, 1]
lorenz(x, y)
lines(x = x, y = piketty.co.cum[, 2], col="blue")
lines(x = x, y = piketty.co.cum[, 3], col="red")
lines(x = x, y = piketty.co.cum[, 4], col="cyan")
lines(x = x, y = piketty.co.cum[, 5], col="green")
title(main = "자본 소득의 로렌츠 곡선", xlab = "소득순 사람들의 누적(%)", ylab = "소득의 누적(%)")
legend(x = 0.05, y = 0.95, legend=c("낮음", "중간", "중상", "높음", "매우 높음"), lty = 1, col = c("black", "blue", "red", "cyan", "green"), title="소득 불평등")
```
```

총소득의 로렌츠 곡선들도 같은 방법으로 그릴 수 있다.

```
```{r fig.width=6, fig.height=6}
piketty.total.cum<-apply(piketty.total[, 2:5], 2, cumsum)
piketty.total.cum
y<-piketty.total.cum[, 1]
lorenz(x, y)
lines(x = x, y = piketty.total.cum[, 2], col="blue")
lines(x = x, y = piketty.total.cum[, 3], col="red")
lines(x = x, y = piketty.total.cum[, 4], col="cyan")
title(main = "총소득의 로렌츠 곡선", xlab = "소득순 사람들의 누적(%)", ylab = "소득의 누적(%)")
legend(x = 0.05, y = 0.95, legend=c("낮음", "중간", "높음", "매우 높음"), lty = 1, col = c("black", "blue", "red", "cyan"), title="소득 불평등")
```
```

평등선과 소득누적곡선과의 차이를 나타내는 gini계수는 결국 면적 계산임을 알 수 있다. gini계수는 $\frac{\text{빈곤선 면적}}{\text{총면적}} = \frac{1}{2} \times \frac{\text{빈곤선 면적}}{\text{총면적}}$ 으로 주어지기 때문에 각 삼각형과 사다리꼴들의 면적을 계산하면서 2로 나누어 줄 필요가 없게 된다.

```
```{r fig.width=6, fig.height=6}
lorenz(x = x, y = cumsum(piketty.labor$Low))
text(x=c(0.3, 0.7, 0.95, 0.995), y=c(0.1, 0.3, 0.5, 0.6), labels=c("A", "B", "C", "D"), col="red")
```
```

A, B, C, D의 면적을 합해서 2배를 해 주면 되기 때문에 2로 나눌 필요 없이 A면적의 2배는 $(y[1] + y[2]) * (x[2] - x[1])$, B면적의 2배는 $(y[2] + y[3]) * (x[3] - x[2])$, C면적의 2배는 $(y[3] + y[4]) * (x[4] - x[3])$, D면적의 2배는 $(y[4] + y[5]) * (x[5] - x[4])$ 이 모두 0이지만 규칙성을 파악하기 쉽도록 집어넣었다.
따라서 x-벡터의 길이보다 하나 적은 갯수의 사다리꼴들의 합을 구하는 문제로 귀착된다. 이 때 i-번째 사다리꼴 면적의 2배는 $(y[i] + y[i + 1]) * (x[i + 1] - x[i])$ 로 주어지게 된다. 이들의 합을 구해서 1에서 빼주면 된다.

따라서 다음 코드가 나온다.

```
```{r}
gini
```
```

``gini()`` 함수를 이용하여 노동소득, 자본소득, 총소득에 나타난 각 경우에 대한 Gini계수는 다음과 같이 구할 수 있다. 노동소득의 각 경우에 지니계수들은,

```
```{r}
apply(piketty.labor.cum, 2, gini, x = x)
```
```

로 계산되고, 자본소득의 각 경우에 지니계수들은,

```
```{r}
apply(piketty.co.cum, 2, gini, x = x)
```
```

이고, 총소득의 각 경우에 지니계수들은,

```
```{r}
apply(piketty.total.cum, 2, gini, x = x)
```
```

로 계산되어 자본소득의 경우 불평등이 매우 심한 경우에는 지니계수가 0.8을 넘기도 한다는 것을 알 수 있다.

뒷 마무리

