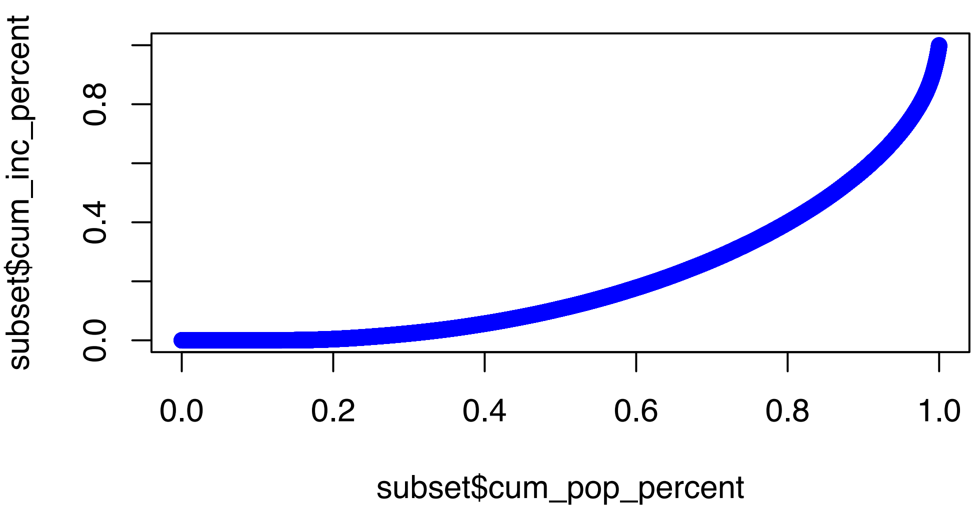
Lorenz Curve and Gini

Rowan, Ari

The Gini coefficient or Gini index is a widely used metric for quantifying income inequality in a given country. It is calculated by finding the area between the Lorenz curve and the line of perfect equality, then dividing that area by the total area under the line of perfect equality.

To calculate the Gini index we first needed to estimate the coefficients of the Lorenz curve. To do this we downloaded a data set from Ipums USA which contained individual, household level, census data about the total yearly income of each household. We cleaned and ordered the data, and then calculated a cumulative population percentage, and a cumulative income percentage variable. When we graph them against each other we can see that classically shaped Lorenz curve emerge:



Next we needed to estimate the coefficients of the model. As we know, the Gini coefficient is given by:

$$G = 1 - 2\int_0^1 l(x)dx$$

Where:

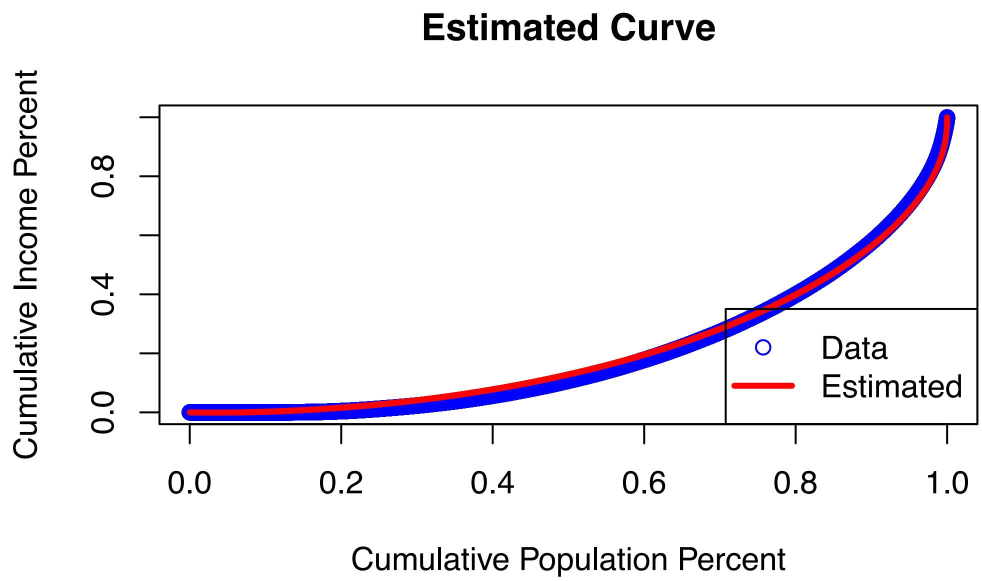
$$l(x) = x - (ax^a(1-x)^b)$$

Therefore we need to use a nonlinear least squares estimation method to estimate the coefficients for a and b. Here you can see the code we ran, and the results for the estimated coefficients for the model. Part of the equation is cut off the edge of the screen due to the pdf rendering method we are using, I couldn't figure it out in time.

```
lorenz_model<- nls(cum_inc_percent ~ cum_pop_percent - (a * cum_pop_percent^a * (1 - cum_p
      data  = subset,
      start = list(a = 0.5, beta = 0.5),  # Starting values for parameters
      lower = c(a = 0, beta = 0),  # Lower bounds to ensure positivity
      upper = c(a = 10, beta = 10), # Upper bounds
      algorithm = "port")  # Port algorithm for bounded fitting

subset$fitted_values_new <- predict(lorenz_model)

print(lorenz_model)
```

```
Nonlinear regression model
  model: cum_inc_percent ~ cum_pop_percent - (a * cum_pop_percent^a *     (1 - cum_pop_percer
   data: subset
     a   beta
0.9816 0.4195
 residual sum-of-squares: 6.69

Algorithm "port", convergence message: relative convergence (4)
```

After estimating the coefficients, we can graph the estimated curve against the actual observed data to check if it visually appears correct.



Finally we can use R to calculate the Gini index by taking the integral of the model, and plugging it into the equation for the Gini index.

```
# Define the fitted Lorenz curve function based on the estimated parameters a and beta
lorenz_function <- function(x, a, beta) {
  x - a * x^a * (1 - x)^beta
}

# Get the estimated parameters from your model
a_estimate <- coef(lorenz_model)["a"]
beta_estimate <- coef(lorenz_model)["beta"]

# Define the function to integrate
gini_integrand <- function(x) {
  x - lorenz_function(x, a_estimate, beta_estimate)
}

# Numerically integrate the function from 0 to 1
gini_result <- integrate(gini_integrand, lower = 0, upper = 1)

# Calculate the Gini coefficient
Gini_coefficient <- 1 - 2 * gini_result$value
```

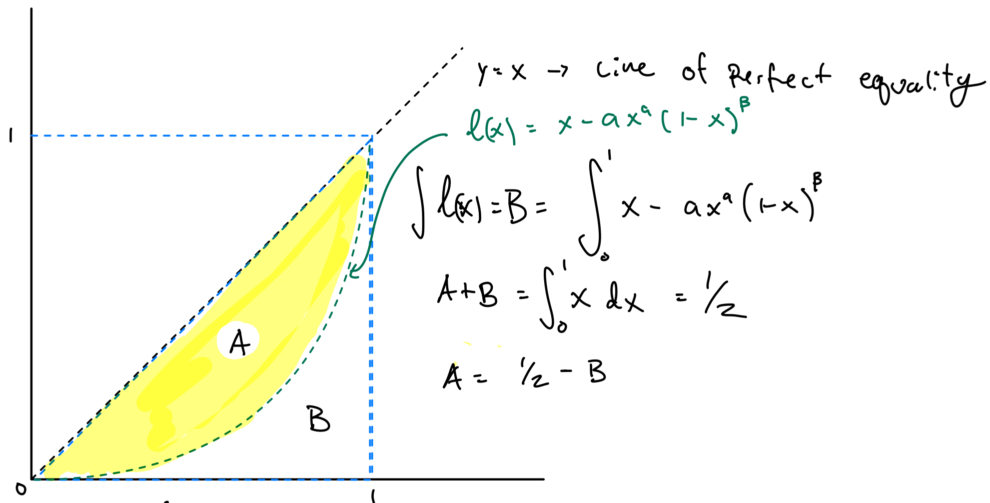
```r
print(Gini_coefficient)
```

```
[1] 0.4214416
```

computed, exact value of Gini-coefficient based on the data

But this is a math class so we are going to show two methods for manually calculating the Gini coefficient here:

Derivation and Computation of the Gini Coefficient



$$\int l(x) = B = \int_0^1 x - ax^{a}(1-x)^{\beta}$$

$$A+B = \int_0^1 x \, dx = \tfrac{1}{2}$$

$$A = \tfrac{1}{2} - B$$

$$G = \frac{A}{A+B} = \frac{\frac{1}{2} - B}{\frac{1}{2}} = 1 - 2B$$

$$B = \int_0^1 x - f(x) \, dx = \int_0^1 \left[x - ax^{a}(1-x)^{b}\right] dx$$

$$B = \tfrac{1}{2} - a\int_0^1 x^{a}(1-x)^{b} \, dx$$

$$G = 1 - 2a\int_0^1 x^{a}(1-x)^{\beta} \, dx$$

$$l(x) = x - a x^a (1-x)^\beta \qquad \text{where} \quad a = 0.978 \quad \beta = 0.4195$$

$$G = 1 - 2a\int_0^1 l(x)\,dx \rightarrow 1 - 2a\int_0^1 x^a (1-x)^\beta\,dx$$

$$\int_0^1 a\, x^a (1-x)^\beta\,dx \longrightarrow$$

$$f(x) = (1-x)^\beta \qquad f(0) = (1-0)^\beta = 1$$

$$f'(x) = -\beta(1-x)^{\beta-1} \qquad f'(0) = -\beta$$

$$f''(x) = \beta(\beta-1)(1-x)^{\beta-2} \qquad f''(0) = \beta(\beta-1)$$

$$f'''(x) = -\beta(\beta-1)(\beta-2)(1-x)^{\beta-3} \qquad f'''(0) = -\beta(\beta-1)(\beta-2)$$

$$f^{4}(x) = \beta(\beta-1)(\beta-2)(\beta-3)(1-x)^{\beta-4} \qquad f^{4}(0) = \beta(\beta-1)(\beta-2)(\beta-3)$$

Taylor series:

$$f(x) = 1 - \beta x + \frac{\beta(\beta-1)}{2!}x^2 - \frac{\beta(\beta-1)(\beta-2)}{3!}x^3 + \frac{\beta(\beta-1)(\beta-2)(\beta-3)}{4!}x^4 + \dots +$$

Plug in $\beta = 0.4195$

$t_0 = 1$

$t_1 = -0.4195(x)$

$$t_2 = \frac{0.4195(0.4195-1)}{2}\cdot x^2 = -0.1217x^2$$

$$t_3 = \frac{0.4195(0.4195-1)(0.4195-2)}{6}x^3 = -0.0641x^3$$

$$t_4 = \frac{0.4195(0.4195-1)(0.4195-2)(0.4195-3)}{24}x^4 = -0.04138x^4$$

$t_5 = -0.0296x^5$

$$\boxed{T_4(x) = t_0 + t_1 + t_2 + t_3 + t_4 \qquad R_4(x) = t_5}$$

$$I(x) = \int_0^1 a\,x^a(1-x)^\beta dx \approx \overbrace{I_4(x) = \int_0^1 a x^a T_4(x)\,dx}^{\text{Estimated value}} \pm \overbrace{\int_0^1 a x^a R_4(x)\,dx}^{\text{Error bound}}$$

$$I_4(x) = a\int_0^1 x^a(1 - 0.4195x - 0.1217x^2 - 0.0641x^3 - 0.04138x^4)\,dx$$

$$= a\left[\overbrace{\int_0^1 x^a dx}^{i_0} - 0.4195\overbrace{\int_0^1 x^{a+1}dx}^{i_1} - 0.1217\overbrace{\int_0^1 x^{a+2}dx}^{i_2} - 0.0641\overbrace{\int_0^1 x^{a+3}dx}^{i_3} - 0.04138\overbrace{\int x^{a+4}dx}^{i_4}\right]$$

$$= a\left[i_0 - 0.4159\,i_1 - 0.1217\,i_2 - 0.0641\,i_3 - 0.04138\,i_4\right]$$

$$\varepsilon < a\int_0^1 x^a\cdot 0.0296x^5\,dx = a\cdot 0.0296\overbrace{\int x^{a+5}dx}^{i_5}$$

Plug in $a = 0.9816$

$$i_0 = \int_0^1 x^a\,dx = \left.\frac{x^{a+1}}{a+1}\right|_0^1 = \frac{1^{a+1}}{a+1} = \frac{1}{a+1} = 0.5046$$ .... Recognize Pattern

$$i_1 = \int_0^1 x^{a+1}\,dx = \frac{1}{a+2} = 0.3354$$

$$i_2 = \int_0^1 x^{a+2}\,dx = \frac{1}{a+3} = 0.2512$$

$$i_3 = \int_0^1 x^{a+3}\,dx = \frac{1}{a+4} = 0.2007$$

$$i_4 = \quad = \frac{1}{a+5} = 0.1672$$

$$i_5 = \quad = \frac{1}{a+6} = 0.1432$$

$$I_4(x) = a\left[i_0 - 0.4159\,i_1 - 0.1217\,i_2 - 0.0641\,i_3 - 0.04138\,i_4\right]$$

$$= a\left[0.5046 - 0.4195\cdot 0.3354 - 0.1217\cdot 0.2512 - 0.0641\cdot 0.2007 - 0.04138\cdot 0.1672\right]$$

$$= a(0.3135) = 0.3077$$

Therefore:

$$I(x) = \int_0^1 a x^a (1-x)^\beta\,dx \approx I_4(x) = 0.3077$$

Where:

$a = 0.9816$, $\beta = 0.4195$

$$\varepsilon = a\cdot 0.0296 \underbrace{\int x^{a+5}\,dx}_{i_5}$$

$$= 0.9816\cdot 0.0296\cdot 0.1432$$

$$\varepsilon = .00416$$

To find G

$$G = 1 - 2(I(x))$$

$$G = 1 - 2(0.3077)$$

$$\boxed{G = 0.384 \pm 0.0083}$$

in terms of Gini:

$$\varepsilon_G = 2\cdot 0.00416$$

$$= 0.00832$$

This is a very poor estimate of the Gini

The real value is $\approx .42$

To get the precise value we will use Gamma and Beta functions

We know:

$$B(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}\,dt$$

$$B(u,v) = \frac{\Gamma(u)\Gamma(v)}{\Gamma(u+v)}$$

So,

$$I = \int_0^1 x^a (1-x)^b\,dx = B(a+1, b+1)$$

$a = 0.9816 \quad b = 0.4195$

$$= \frac{\Gamma(a+1)\Gamma(b+1)}{\Gamma(a+b+2)} = \frac{\Gamma(1.9816)\Gamma(1.4195)}{\Gamma(3.4011)}$$

Using a calculator we find

$$I = \frac{0.9923597 \times 0.8863741}{2.9847166} = 0.294629$$

$$G = 1 - 2a(I)$$

$$2a = 1.9632$$

$$G = 1 - 1.9632 \times 0.294629$$

$$= 1 - 0.578559$$

$$\boxed{G = 0.421441}$$

This is a much more accurate value for the Gini coefficient.

it matches Exactly with the computed value