Course 2 - Looking at some Distributions

UCLA - Econ 19 - Fall 2018

François Geerolf

Contents

In Course 2 and Course 3, we will stop on a paricularly important remark by Rosen [1983]

> Of particular interest here is an observation, first studied systematically by the great Italian economist Vilfredo Pare to in the late nineteenth century, that the distribution of income contains an unusually large proportion of top earners: that is, among the rich rather than the poor. A visual image will perhaps clarify what is meant by "unusual" in this connection. Imagine a graph plotting IQ scores on t he horizontal and the frequency of scores on the vertical. The result is a familiar bellshaped curve. The peak of the bell occurs at a score arbitrarily scaled at 100 and the curve falls symmetrically on either side o f 100. Now picture a similar graph, except with earnings on the horizontal. The resulting curve is unbalanced and nonsymmetrical-a bell that is definitely out of whack. To the left of the modal (peak) value it appears much like the IQ frequency curve. However, to the right of the mode it does not fall as fast as it does to the left. It looks as if someone ha d stood at the right end of the curve, placed it over his back like a rope, and dragged and stretched it out a very long distance. The upper or right-hand tail of the distribution of income is much thicker than the lower, left-hand tail. The extra weight on the right lends a certain skewness to the distribution of income. What this comes down to is that the distribution of earnings is far from proportionate to the distribution of ability. Amazingly, Pareto's observations have been qualitatively duplicated in virtually every era of every society for which data on income distributions can be found.

1 Bell-Shaped Distributions

1.1 Natural Sciences

```r
height <- read.csv("https://raw.githubusercontent.com/hadley/r4ds/master/data/heights.csv")
```

1.2 Some maths

The Bell shape curve is defined by a density function given by:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{(x-\mu)^2}{\sigma^2}\right).$$

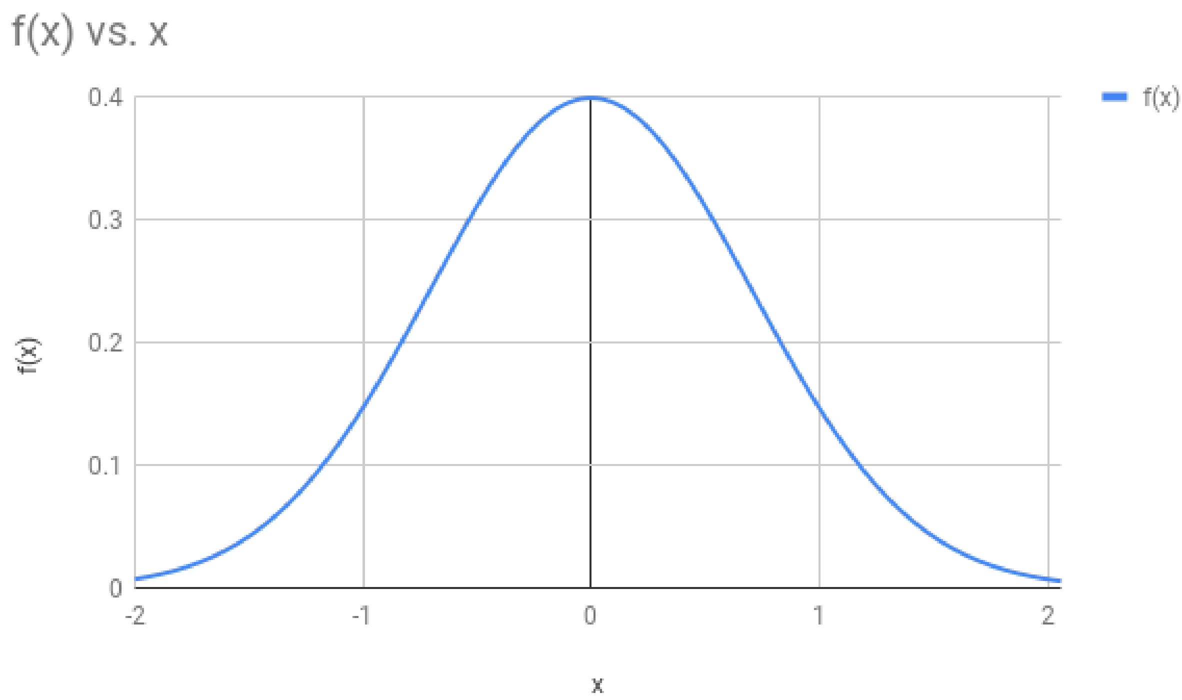

Figure 1: Bell Shaped Curve

1.3 Google Sheets

Here is a link to the Google Sheets that we created in order to look at the Gaussian distribution. In particular, we were able to plot the density function of a Normal Distribution with $\mu = 0$ and $\sigma = 1$, using the formula above.

2 Pareto Distributions

2.1 Maths of Pareto

A key feature of the Pareto distribution is that the density distribution does not go as fast to 0 as x becomes large. If x is population, his means that there are relatively many cities with a large size. Similarly, there are relatively many incomes that are much larger than the mean. The Pareto Distribution is in fact defined by:

$$f(x) = a\frac{x_m^a}{x^{a+1}}.$$

For the cumulative distribution function, this implies:

$$1 - F(x) = \left(\frac{x_m}{x}\right)^a.$$

2.2 Data on cities using Google Sheets

We have used this Google Sheet in order to plot the city size distribution of cities.

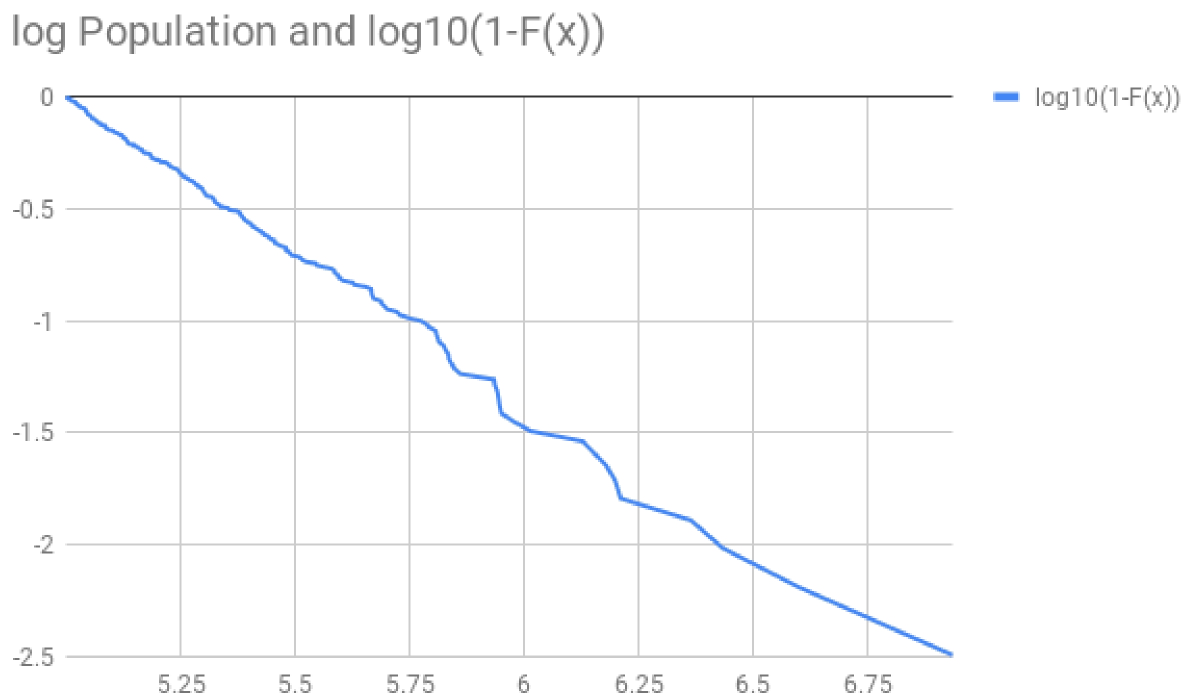

Figure 2: City Size Distribution

We note that the result is something that is close to a linear relationship, suggestive of Pareto like behavior.

2.3 Data on cities using R Statistical software

```
data.on.cities <- "https://en.wikipedia.org/wiki/List_of_United_States_cities_by_population" %>%
  read_html %>%
  html_table(header = TRUE, fill = TRUE)

data.on.cities[[5]][,c(1:4)] %>%
  as.tibble
```

```
## # A tibble: 311 x 4
##    `2017rank` City             `State[5]`   `2017estimate`
##         <int> <chr>            <chr>        <chr>
##  1          1 New York[6]      New York     8,622,698
##  2          2 Los Angeles      California   3,999,759
##  3          3 Chicago          Illinois     2,716,450
##  4          4 Houston[7]       Texas        2,312,717
##  5          5 Phoenix          Arizona      1,626,078
##  6          6 Philadelphia[8] Pennsylvania 1,580,863
##  7          7 San Antonio      Texas        1,511,946
##  8          8 San Diego        California   1,419,516
##  9          9 Dallas           Texas        1,341,075
## 10         10 San Jose         California   1,035,317
## # ... with 301 more rows
```

References

Sherwin Rosen. The Economics of Superstars. *The American Scholar*, 52(4):449–460, 1983. ISSN 0003-0937. URL http://www.jstor.org/stable/41210977.