A Paradox, Induction, and Research Basics - DA2210

October 4, 2021

Philip Salqvist

phisal@kth.se

1992-04-17

Contents

1 Two research studies

1.1

I would argue that we can't draw the conclusion that comments improve the code based on this experiment. First of all, their are to many variables that could contribute to the code being better in group A compared to group B than the variable of commenting or not commenting code. You have to consider the fact that group A could have better programmers than group B, and this is just one example. The fact that the programmers get to choose group between group A and B, could indicate that better programmers would choose the group that is instructed to use comments. If we made a larger sample of groups, and made multiple experiments of comparing groups, we could perhaps derive a statistical correlation between commenting code and writing better code. This would not necessarily mean that commenting code causes better code, but it could at least result in observing a correlation.

To improve the study, I would not just compare two different groups of programmers. Instead, I would look at a large sample of experiments where each experiment contained a group A and a group B. I would not create group A and B by letting programmers choose if they want to be a part of group A or B. I would randomly assign programmers to the two groups, and instruct one of them to use comments, and the other to not use comments. If we look at a large enough sample, then we could derive a statistical correlation between commenting code and getting better results of code.

1.2

We cannot draw the conclusion that functional programming languages give slower code than non functional ones. The only thing that we can conclude in this example is that the program written by group D is faster than the one written by group C. Group C have used a functional programming language, and their code runs slower, but this doesn't tell us that functional programming languages is general are slower. The result could be yielded due to a number of factors. The particular language Haskell might be slower than Java, or the code written by group C might just be of poorer quality in terms of time complexity than the one of group D to name two examples.

To improve the study, I would not organize two groups and have them solve a problem. I would rather determine one particular problem, and the fastest algorithm that is known to solve this problem. We could then implement the algorithm in a number of functional programming languages, and in a number of object object oriented languages and measure the execution time. After doing this we could perhaps derive a result that either functional languages are faster for implementing this particular algorithm, or vice versa. This wouldn't tell us anything about the speed of the two languages in general, but specifically to this type of problem.

2 Zipf's law

a)

I have chosen to work with a data set that describes the wealth of the richest people in the world [1]. Particularly the rank and size of wealth in billions of dollars between these individuals is being compared.

b)

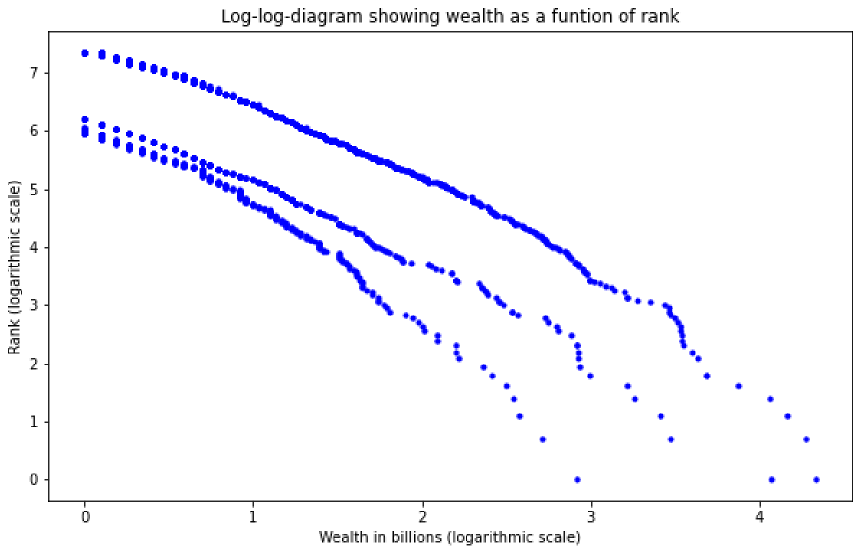

Figure 1: Log-log diagram showing wealth as a function of rank

c)

I've used an OLS regression model to fit the data into a power law on the form of $Cn^{-\alpha}$ using the logarithmic data of wealth and rank. This resulted in the intercept 3.8612, which gives us $C = e^{3.8612} = 47.5223$. The model also resulted in an $\alpha = 0.501$. Using these two coefficients predicted wealth data is created, and is plotted as a function of rank on a logarithmic scale.

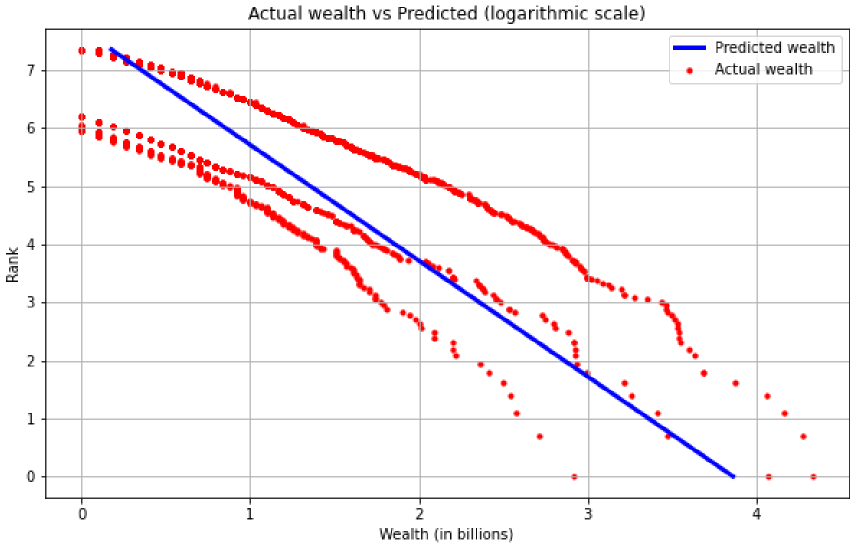

Figure 2: Log-log diagram comparing actual wealth and predicted wealth

d)

In Figure 1 we can see Ogwang's [2] comparison between rank and wealth during 2009. The data set is from Forbes net worth of billionaires, and although the figure only shows data from 2009, we can see similarities between the plot in Figure 3 and the one in Figure 2.

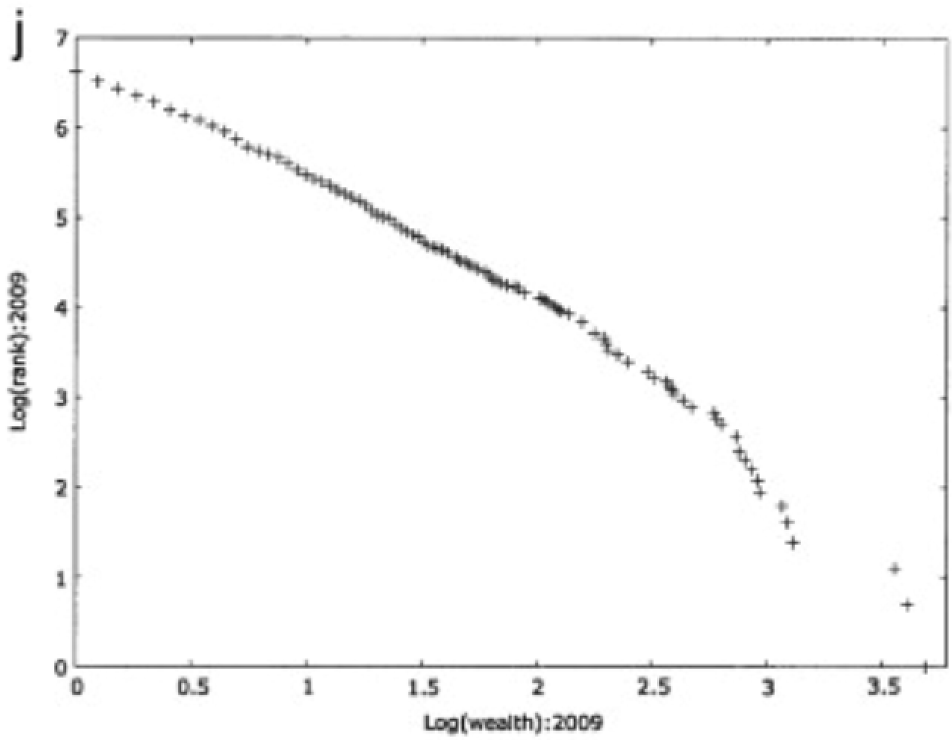

Figure 3: Ogwang [2] rank and wealth on log scale

e)

I would argue that it is an conjecture or perhaps an empirical law. It is a statement that is believed to be true based on our made observations in nature, and to such a degree that it has been made a statistical law. I would argue that it is not a mathematical theorem since it hasn't been proved in the sense of an inductive proof that builds on prior axioms and proved theorems.

References

[1] R. Whitcomb *Billionaires, 2.0.0.* [Dataset] Available: https://corgis-edu.github.io/corgis/csv/billionaires/ [Accessed: October 4, 2021]

[2] T. Ogwang *Is the wealth of the world's billionaires Paretian?*, Physica A: Statistical Mechanics and its Applications, vol 392, no. 4, pp. 757-762, Accessed: October 4, 2021, Available: https://corgis-edu.github.io/corgis/csv/billionaires/