Comments on the report ”Power laws, Pareto distributions and Zipfs law”

Hannes Leskelä

October 17, 2016

Introduction

The introduction gives a clear overview of the concept of power laws, and the histograms really helps to drive the idea through. The formula is also explained in a straight-forward manner with the classic form of the linear equation, albeit with the logarithm of the response and predictor. There is nothing difficult to understand yet, except perhaps why. Why do these patterns emerge in these wide variety of fields? Perhaps it is just one of the many mathematical properties that occur in nature.

I. Measuring Power Laws

In this section we play with a generated data set, and I was surprised about the behavior of the graph at first. But after thinking about it, it seemed normal that some values in the right tail would not occur at all in our generated data, and thus create the erratic log plot. The binning of values also seemed so clever, especially when the point about plotting the data came through. Very often I have been mislead by skewed axis in a graph, so to see that log binning would generate a constant interval size was a relief as the graph would not get skewed.

I had some troubles interpreting the integral describing the number of samples for a given bin, but when I realized it was just the definition of Zipf’s law I understood it better. I did not derive the CDF, but understood it and was exited to see it was in a quite simple form and generated a really good plot of the Pareto distribution.

For the exponent of the Pareto distribution I was surprised to see that a least square fit generated such a poor estimate, but since the estimating sum is so easily calculable with a computer I was glad to see it had an easy solution. Something that I will get back to later is that when I am reading this text I think a lot about how a data scientist would apply these tools. For instance, as mentioned later in the text, the 80/20 rule is often applied to activity of users and similar data related events, something which might be of interest for data driven companies. That is why I really like it when there is an easy way to check for a phenomenon.

A. Example of Power Laws

I did not think much of this section, since I already had a clear view of power laws and where they occur. This is partially because we have briefly discussed the 80/20 rule in a lot of courses in my undergrad, so I was eager to continue to the more information intensive sections.

I was however a bit intrigued by how to pick an x_{min} for which the power distribution holds. To see that it was a result of the better judgment of the analyst to set this value was a bit disappointing, but since I am used to hyper-parameter tuning for various statistical methods in machine learning, it was nothing new.

B. Distributions that do not follow a power law

I wish that this segment was expanded a tiny bit upon. It is clear that there are many distributions that do not follow a power law, but to know when a function has an exponential cutoff and when it is still considered to follow a power law would be interesting to know. For instance, I find figure c on page 8 quite similar to figure b or l on page 6, but these were classified as Pareto distributions.

II. The Mathematics of Power Laws

A. Normalization

I was a bit unsure about why the normalization was defined as it was, but I agreed on the fact that α should be strictly greater than one, since many of the signs would be negated in our formula if not. This would in turn generate equation (9).

B. Moments

I did not understand how the mean value was calculated from the formula, but I suspect I did not give it enough time. The consequences were on the other hand clear and interesting, since $\alpha \leq 2$ did not have a finite mean, and $\alpha \leq 3$. I understand that from a fixed sample one can always find a mean, but to think about an ever increasing mean if the sample size was infinite, was quite astonishing.

C. Largest value

This was the section I had the most trouble understanding, mainly due to the many integrals. I studied calculus four years ago, and am embarrassed to say that I would do best in rehearsing it a lot to fully understand derivations like these. I basically just acknowledged that there is such a property for power laws, but I will try my best to understand it further on my spare time. A lot of the troubles arose due to the fact that I don't remember the gamma and beta distribution, nor know some of the more rare notation used.

D. Top-heavy distributions and the 80/20 rule

I did not have as much trouble following this section as section C above, in fact is was quite easy to understand. Only equation (26) required some slight thinking before I realized it was just the ratio of the two parts one would like to examine. The resulting formula was also fun to play around with, since one could play around to see how wealth is distributed.

E. Scale-free distributions

The scale-free phenomenon was really interesting and I will definitely continue to read section III.E of the paper to see how this can occur in different systems. I would at least expect the multiple to increase linearly as in other distributions when one changed the scale, and I actually had to check that this was not the case by searching on the web. This lead me to papers about scale-free network which are graphs that follow the power law, and is something that I will investigate further.

Conclusion

As a final remark I would like to state that I enjoyed reading this paper very much. It was set on a level that encouraged me to learn more, but not too difficult to understand right from the start. There were also many relevant notes regarding computer science, which meant that this paper felt a lot more relevant than other mathematical papers that I have read in my studies.