Review of: "How can we Guess the Distribution of Wealth in Greenville, Mississippi? A Proposed Heuristic for Guessing Distributions"

Zoltán Néda[1]

1 Babes-Bolyai University of Cluj-Napoca

Potential competing interests: No potential competing interests to declare.

Despite of the large amount of electronic databases available nowadays, most of our socio-economic data are far from being complete. Exhaustive data are seldomly available but in the few cases they are, it can offer valuable information on many important socio-economic measures. Distribution of wealth and income is a good example. Most of the data that one can easily collect from the Internet on settlement level are either the mean value over a known population or some rough distribution in a few income/wealth categories. Any method, that would allow a more precise determination of the distribution function from such incomplete information is therefore extremely important. The present work of Neuman and Dayan, offers an interesting methodology in such sense. The main idea is to use some empirical knowledge (mass wisdom) about the nature of the unknown distribution and by using these constraints to maximizing the entropy in the neighborhood of the True Median. With such a methodology one can either select the most probable distribution from a given finite list, or can give constraints on the parameters of a hypothesized distribution kernel.

This idea is nice, however there are some tricky points that the authors could still consider.

First, a major problem we face many times is that even the kernel function for the distribution is unknown. By assuming that many are poor and few are rich allows for many type of distributions (exponential for example), not necessarily the Pareto type. It is true, that nowadays we do have many indications that both wealth and income are distributed with a Pareto-like tail, but even in such cases we do not mean a simple power law distribution with a cutoff. The mathematically correct probability density function we are trying to use is defined on the whole $x > 0$ domain, and the kernel function in such cases are proposed in the form: $\rho(x)\,(1 + a\,x)^{-b}$ or even more realistically as $\rho(x)\,e^{-a/x}x^{-b}$. Knowing the right parameters of such distributions, by using just a few empirical data would be definitely a big hit. The catch for these distributions is however that the true median in not analytical, but the mean value is. So, in such sense the mean value is more handy than the true median.

Maximizing the entropy in the vicinity of the true the median is also not viable for such distributions. Despite all these problems for a really useful quantitative application, I consider the main idea nice and promising for a better estimation of some distributions from incomplete data.